Teaching Linguistic Prompt Control for LLM-Based Translation: A Classroom Approach to Developing Critical and Responsible AI Literacy

Katrin Menzel
Leipzig University / Institut für Angewandte Linguistik und Translatologie
katrin.menzel@uni-leipzig.de

Abstract

As Large Language Models (LLMs) are increasingly integrated into professional translation, there is a growing need for pedagogical frameworks that move beyond simple trial-and-error prompting across any available tools used for translation tasks. This paper presents a framework developed in a university MA seminar on translation to foster responsible AI literacy as well as linguistically grounded prompt control and structured prompting. The approach also emphasises that AI is understood as a tool for suggesting possible translation solutions, while human translators remain the final decision makers. Integrating translation-oriented text analysis and corpus-informed feature extraction from parallel and comparable data, the approach teaches students to develop register-specific and model-interpretable instructions when translating specialised texts with the assistance of generative AI tools. During a one-semester course, students familiarised with prompting strategies and different tools, including commercial, freely available models, GDPR-compliant institutional infrastructure and local models. These steps and the comparative evaluation of these tools allowed the students to identify optimal configurations that conform best to professional quality standards for specialised texts, and they ultimately led to highly improved translation output compared to unsteered model results.

1 Introduction

1.1 Overview of the growing role of LLMs in translation

As Large Language Models (LLMs) become increasingly common tools, professional translators and translation students are using them more frequently to support the translation process, e.g., by generating draft translations and revising target texts. The growing presence of these tools has important pedagogical implications. While LLMs can produce fluent translations, their output is shaped by statistical tendencies inherited from their training datasets. The output can vary across runs and lead to apparently random or inconsistent results for the same input. Moreover, particularly in unsteered baseline translations produced without specific instructions, LLM-generated translations may exhibit certain unwanted characteristics that set them apart from both traditional machine translation and human translation, including, for instance, differences in stylistic variation, lexical choice and degrees of explicitness (cf. Sizov et al., 2024).

1.2 Related research on AI-supported translation and pedagogical implications

Research on AI-supported translation tasks and translator education emphasises the importance of maintaining human-centered decision making and the development of the capacity to analyse and manage model behaviour rather than treating

LLMs as opaque drafting tools or black boxes. Recent publications in the field include work by Angelone (2026), Ayvazyan and colleagues (2024, 2025), Cui et al. (2025), Gao et al. (2023), He (2024), Ren and Wang (2025), Siu (2024), Yamada (2023, 2026), Yang et al. (2024) and Zhang et al. (2023). Existing studies highlight the need for structured training that equips translation students and professional translators with the skills to engage critically with LLM outputs and to control and interpret model behaviour in translation processes. Such training may draw on research on efficient prompt engineering and general prompting frameworks that provide systematic approaches for designing prompts. For instance, the CO-STAR framework structures prompts around the dimensions of context, objective, style, tone, audience and response (cf. Teo, 2023). The CLEAR framework emphasises prompts that are concise, logical, explicit, adaptive and reflective (Lo, 2023). Such frameworks offer general principles applicable to a wide range of LLM tasks. In addition to such general frameworks, various model- and task-specific prompt templates have been suggested. For example, TranslateGemma, an open variant of the Gemma 3 model specifically enhanced for translation has been reported to work well with a prompt that instructs the model to produce a faithful and culturally sensitive translation as follows: *"You are a professional {source_lang} ({src_lang_code}) to {target_lang} ({tgt_lang_code}) translator. Your goal is to accurately convey the meaning and nuances of the original {source_lang} text while adhering to {target_lang} grammar, vocabulary, and cultural sensitivities. Produce only the {target_lang} translation, without any additional explanations or commentary. Please translate the following {source_lang} text into {target_lang}:\n\n\n{text}"* (Finkelstein et al., 2026).

While the above-mentioned frameworks and such systematic translation-related prompts demonstrate that structured prompting can reduce unwanted variability in LLM outputs, they serve only as a general starting point for professional translation tasks due to their broad applicability. Translators and educators can further refine them to address the specific requirements of different registers and specialised domains when working with selected tools and language pairs.

1.3 Characteristics of LLM-generated translations

In specialised translation contexts, attentive readers and language experts, including professional translators, can identify recurring AI fingerprints in LLM translation outputs that deviate from established professional standards for the target texts. These tendencies may include repeated and unwanted syntactic structures – such as a disproportionate overuse of non-finite clauses in English texts – and a general inclination toward stylistically flattened, register-neutral formulations. Such patterns may be unproblematic in some contexts, but they can become limitations in domains where specific register features and specific lexico-grammatical and stylistic choices are essential for the text's communicative function. For instance, in an informational specialised text, an LLM-generated translation may fail to maintain the high density of nominal structures and impersonal expressions that would be typically desired by the producer of the target language text to convey institutional neutrality and compact informational content. It may tend to default to more general language or conversational patterns.

1.4 Considerations in model selection

The challenge for translators in navigating AI tools is exacerbated by the vast and heterogeneous landscape of available technologies and the rapid pace of ongoing developments in the field. AI systems differ in how well they handle multilingual tasks including translation. This performance gap is particularly evident when working with different language pairs. While models typically achieve higher accuracy for high-resource European languages that are typologically close, their effectiveness declines for low-resource languages and typologically distant combinations. To navigate this complexity, practitioners can rely on their own linguistic expertise to assess translation quality. They may also consult external evaluation frameworks from academic and industry contexts. Such frameworks provide relatively standardised comparisons of model performance and practice-oriented rankings of system capabilities, some focusing specifically on translation and others addressing translation in addition to other multilingual and creative language tasks (cf., for instance, Burkert, 2025 and Spiridonov, 2026). However, such benchmarks capture only limited aspects of translation quality, and it might be difficult for translators to interpret the scores or understand

how they were computed. Moreover, such rankings may not generalise to specialised translation contexts.

Furthermore, the use of AI in translation tasks and translator training is complicated by GDPR-related concerns, alongside practical limitations such as the requirement for specialised infrastructure and access to cost-effective AI services. The use of commercial closed models, i.e. proprietary systems that cannot be inspected or modified by users and are typically accessed via an API, has become increasingly common in translation workflows. In such setups, user input is processed on external servers, which raises concerns regarding data protection when handling confidential material. This contrasts with locally deployed models, which can be run entirely on the user's own hardware without transmitting data externally and may in some cases be adapted or fine-tuned by users.

These considerations have stimulated growing interest among translators (who are typically not trained in computer science) in local deployment solutions that do not require substantial hardware investment or advanced technical expertise. There is also increasing interest in secure, non-commercial, institutionally managed cloud infrastructures. However, such solutions are still under active development. Many tools remain experimental and sometimes provide less user-friendly interfaces and fewer features than the leading commercial systems, for example, limiting the amount of text that can be uploaded at one time. For enterprise users, individual cloud-based translation solutions are available, such as Language Weaver,[1] which is described by its developers as a secure, scalable AI translation platform. It combines neural machine translation and generative AI and was developed with confidentiality in mind.

1.5 AI-supported translator training in practice

Building on the observations outlined above and the prior research discussed, this paper proposes a pedagogical framework that integrates linguistic analysis, structured prompt design and reflective evaluation into a unified teaching sequence. It summarises tasks from a one-semester seminar at Leipzig University intended to guide MA translatology students toward becoming informed designers of AI-supported translation workflows. Integrating translation-oriented text analysis based on Systemic Functional Linguistics (SFL) and corpus-informed feature extraction from parallel and comparable data, the approach helped students to develop register-specific and model-interpretable instructions when translating specialised texts with the assistance of generative AI tools. The students utilised a diverse ecosystem over the course, including commercial and freely available models, GDPR-compliant institutional infrastructure and local models. The seminar also benefited from input by Martin Czygan from the Leipzig University Library, who shared his expertise on AI tools and workflows in a workshop held within the translation seminar.

One of the key outcomes of the seminar was a reduction in inefficient trial-and-error prompting and unnecessary iterations, as well as the mitigation of unacceptable translation outputs suggested by individual models. This was achieved by guiding students to employ a structured prompt architecture comprising a configuration prompt, a task prompt and a revision step and by supporting informed model selection based on quality considerations.

2 Pedagogical framework

2.1 Baseline analysis of LLM output

The pedagogical framework was implemented within an MA-level, cross-linguistic methodological seminar, open to students of translation and interpreting studies who work with a range of different language pairs in their respective specialisations. While most students in the seminar had substantial experience with English-German translation, some also had considerable experience with other language combinations involving, for instance, French, Spanish or Arabic as one of their main working languages.

The process began with students producing an unsteered baseline translation of an informational English text from an EU website on a topic of current relevance, such as antimicrobial resistance, using a minimal prompt (e.g., *"Translate this text into German"*) with a model of their choice. The text selected for this task[2] was approximately 1,200 words in length, with students initially working on shorter extracts before progressing to larger sections and, ultimately, the full text. It can be classified as a specialised text targeting a non-specialist but informed readership, including

[1] https://www.rws.com/language-weaver/

[2] https://health.ec.europa.eu/antimicrobial-resistance/eu-action-antimicrobial-resistance_en

stakeholders and interested members of the public. In this initial phase, students primarily relied on widely available commercial platforms that provide access to LLMs, such as those underlying OpenAI's ChatGPT, Microsoft Copilot (institutionally licensed) or Google Gemini. These tools produced translations that the students used to discuss default AI translation patterns and recurring linguistic features, based on their prior linguistic knowledge.

For the majority of the cohort, the English-German language pair served as the primary framework, with English-French provided as an option for those who were not native speakers of German. However, since several students primarily work with other language combinations in their daily practice and academic curriculum, e.g., involving Arabic, a specific challenge emerged in identifying institutional text resources with comparable stylistic consistency and parallel availability. While informational EU texts, which are available in various languages, served as suitable benchmarks for observing typical register-specific features and for comparing officially published versions with LLM-generated outputs, United Nations texts were proposed as alternative resources for students working with Arabic, given their availability as comparable institutional web texts on specialised topics in a public-facing communicative context.

2.2 Translation-oriented text analysis, corpus-based exploration and terminology work for prompt preparation

The next step involved deepening the students' understanding of the same informational texts they had used in the initial task by conducting a register analysis grounded in SFL (Halliday & Matthiessen 2013). The students examined how ideational, interpersonal, and textual meanings are realised, and they identified linguistic features in the texts that contribute to the thematic domain (Field), the construction of social roles and relations of expertise (Tenor) and the mode of communication in terms of medium, information structure and cohesive organisation (Mode).

The students then consulted large-scale corpus data in order to validate their observations derived from individual texts and to establish empirical frequencies of register-specific linguistic features. The corpora were selected for their register proximity to the analysed text type; however, no exact matching corpus of informational institutional website texts was available, so functionally and thematically comparable corpora were used instead. Europarl corpora, accessible for instance via Sketch Engine[3] and via OPUS,[4] were used to verify whether patterns identified in individual institutional web texts (such as high nominal density, frequent passive constructions and specific hedging strategies) constitute systematic features of the broader institutional register. The United Nations Parallel Corpus available via Sketch Engine was additionally used to ensure access to comparable institutional large datasets for the less commonly represented language combinations in the seminar.

The students observed that the analysed EU informational web texts and EU parliamentary debates in Europarl share core features of institutional discourse. These include a formal tone, the use of domain-specific and relatively standardised terminology, frequent quantification through measurable indicators and statistical data, explicit references to data sources, the use of comparative markers and references to procedural and institutional authority and to collective actors. Such characteristics provide important contextual information that can be integrated into translation prompts to guide AI tools toward more contextually appropriate and institutionally authentic outputs when processing similar texts.

The students further noted that the analysed EU web texts were typically more information-oriented and characterised by higher nominal density and a more impersonal style. In contrast, the Europarl data, which consist of edited and published plenary debates, naturally display more features of spoken political discourse, including increased personal reference and evaluative language. These characteristics are specific to debate-based data and should not be transferred to specialised informational web texts. Instead, translation prompts in this case should prioritise features that occur across different forms of EU discourse, together with the structural and stylistic conventions that are typical of institutional informational writing which is addressed directly to civic audiences.

Furthermore, the students utilised Sketch Engine's OneClick Terms tool[5] to identify terminology and multi-word expressions from existing comparable

[3] https://www.sketchengine.eu/
[4] https://opus.nlpl.eu/
[5] https://terms.sketchengine.eu/

institutional texts and parallel text pairs. In addition to EU and UN materials, they also worked with texts from other specialised domains, particularly in scientific and legal domains, to apply the tasks across different types of specialised discourse. This workflow enabled the students to perform structural alignment of existing reference texts and to create bilingual glossaries that could also serve as input for prompts. The students also learned to preprocess texts to improve structural alignment and to evaluate statistically derived term candidates before integrating them into their translation workflows.

2.3 Model selection and exploration of GDPR-compliant options

In early phases, the students experimented with commercial models, as outlined above, and compared their outputs in class discussions. After some practical experience in evaluating models from a qualitative linguistic standpoint, benchmarking approaches were additionally discussed in a dedicated workshop with Martin Czygan, a software engineer working at the Leipzig University Library (Digital Services). The workshop provided the students with further insights into evaluation criteria and comparative assessment strategies for model outputs in translation tasks. In this workshop, the seminar also expanded the previously used ecosystem of tools and platforms to include demonstrations of local models such as TranslateGemma, run autonomously using the Jan.AI tool,[6] as well as GDPR-compliant institutional infrastructure, in particular the Academic Cloud.[7]

The Academic Cloud was developed and coordinated within German higher education networks with significant involvement of the Göttingen State and University Library and is technically operated by the Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG). It was tried out by the students in the seminar as it provides a GDPR-compliant environment for research and teaching, including AI-based applications. Despite not being intended for professional translators as a target group, the Academic Cloud ChatAI service[8] has proven to be a valuable experimental tool in our academic translator-training context. It provides access to a curated selection of LLMs, including institutionally hosted open-source models such as Intel Neural Chat 7B, Mixtral 8x7B Instruct, Qwen1.5 72B Chat and Meta LLaMA 3 70B Instruct as well as externally hosted models such as OpenAI GPT-4. Internally hosted models are operated on infrastructure managed within the Academic Cloud/GWDG environment, whereas externally hosted models are accessed through integrated interfaces to third-party providers under GDPR-compliant data-processing arrangements. Such infrastructures are particularly useful in educational contexts, as they enable students and instructors to experiment with different LLMs through a unified academic platform.

2.4 Prompt architecture: configuration, task and revision prompts

Generally, after selecting an LLM interface and, where applicable, a specific model, the students structured their prompting into three components: a general configuration prompt, a more specific task prompt and a revision prompt.

Configuration prompt: This prompt involves a set of role-based instructions and general translational principles formulated in operational terms that was developed collaboratively with the students. This provides stable, high-level behavioural guidelines for the model and reduces the need to address individual translation issues in this step. In the present seminar context, this involved assigning the model a translator role for specialised texts, e.g., EU health-related informational websites translated from English into German or vice versa, and specifying consistent translational principles. These include fidelity to source meaning without unwarranted additions or omissions, register consistency in maintaining a formal institutional, information-oriented style for a non-expert but informed readership, and terminological precision aligned with EU domain-specific usage. In addition, this prompt may instruct the model to produce only the translation itself, as in the preferred TranslateGemma prompt discussed above, to ensure that no unintended content is added. Where supported by the chosen tool, the configuration prompt could be implemented as an assistant- or agent-style prompt.

Task prompt: The task prompt incorporates language-pair-specific considerations regarding systematic differences (e.g., between English and German) in the given register. As the students have identified a wide range of features in previous steps, their individual prompts will ultimately prioritise those aspects most relevant to them

[6] https://www.jan.ai/

[7] https://academiccloud.de/de/

[8] https://academiccloud.de/services/chatai/

based on their text analysis and corpus-driven findings. These considerations may include differences in word-formation preferences in institutional discourse, e.g., the greater tendency towards closed nominal compounds in German. Another example that some students may consider worth including, in light of insights from a specific text and corpus analysis, is the avoidance of premodifying hyphenated structures in German as translation equivalents for comparable constructions in English. For instance, where English may use forms such as "*bacterium-antibiotic resistance combinations"*, German institutional texts might prefer to encode relations more explicitly and in a different order through postmodifying prepositional phrases rather than premodifying attributive compound structures, e.g., *"Resistenzkombinationen zwischen Bakterien und Antibiotika"*, but LLM baseline outputs sometimes translate such structures too literally. The prompt can also address issues of grammatical gender representation, that is, the need to make gender inclusivity explicit in German where English uses gender-neutral expressions (e.g., *government experts* vs. *Regierungsexpertinnen und -experten*). Furthermore, the task prompt should address the handling of institutional terminology, including names, policy labels and abbreviations, e.g., in cases where full and short forms do not correspond directly across languages (e.g., *European Centre for Disease Prevention and Control (ECDC) / Europäisches Zentrum für die Prävention und die Kontrolle von Krankheiten (ECDC)*), since German institutional discourse may avoid repeated use of initialisms derived from foreign-language full forms.

A key pedagogical insight emerged during this step from the comparison between human-oriented and LLM-oriented instructions. When the students attempted to brief an LLM as they would a human translator, they observed that the model sometimes followed individual instructions inconsistently, even when these appeared straightforward and unambiguous. In such cases, the students attempted to include more examples and contextual information in their prompts in order to translate linguistic constraints into more explicit, model-friendly instructions. More explicit commands with examples usually worked better, especially when a revision step followed the task prompt.

However, rather than attempting to account for every possible linguistic construction, the task prompts generally combined rule-based guidance with some illustrative examples and provided guidance without aiming to be exhaustive. In this step, the students also need to consider model-specific capacities and limitations, including differences in context-window size, maximum prompt length and the number and types of possible supplementary files (e.g., glossaries, reference translations or other background documents) that can be integrated into the prompting workflow.

Revision prompt: This prompt is applied after the initial translation has been generated during the preceding steps. It instructs the model to assess its output relative to the specified constraints and produce a revised version accordingly. In practice, the model does not perform a human-like review; rather, it enters a second, constraint-aware generation phase in which it negotiates potentially competing requirements to improve alignment with the defined translation objectives. When multiple constraints conflict, the model may prioritise higher-level constraints, and it can be instructed in this step to flag any constraints that could not be fully satisfied and to provide a brief justification.

This refinement step enables iterative self-correction and reduces the need for repeated user-driven prompt modifications. To a certain extent, it enhances transparency in the model's decision-making process. Preliminary observations indicate that this self-evaluation mechanism improves adherence to register-specific constraints and leads to a higher quality of the translation output as well as improved compliance with the given instructions. It also helps the students to identify areas where LLMs systematically struggle to comply with or follow specific instructions.

Overall, this architecture provides the students with a reusable framework for controlling LLM output while keeping the prompting process structured and manageable. Over the course, the students worked with texts representing different specialised registers, primarily within the English-German language pair (in both directions). They were then free to choose their own case study and apply the methods and knowledge they had acquired during the semester. The students selected specialised texts for translation and carried out individual translation projects which concluded with a final presentation of their work and findings.

3 Discussion

In conclusion, the paper illustrated how a structured pedagogical framework combining linguistic analysis, corpus-informed insights and staged prompting can help translation students become

informed designers of AI-assisted translation workflows. The teaching framework that was presented in this paper also addresses a central challenge of LLM-assisted translation: the tension between linguistic specificity and prompt complexity. The approach encourages critical reflection on the limitations of different model types and their appropriate applications.

One limitation remained evident even when detailed and example-rich task prompts were provided: LLMs did not always consistently monitor or apply stylistic and linguistic constraints in the way human translators typically do. Human translators naturally monitor a wide range of features and vary their choices depending on the register; they rarely need to be told, for example, to avoid producing the same syntactic structure repeatedly. LLMs, however, sometimes exhibit recurrent patterns that would be unusual in human translation. Users may find themselves giving instructions that would be less necessary for human translators, such as *"Avoid using participle clauses repeatedly across consecutive sentences or stacked in a single sentence in this English translation"*, because such patterns can emerge as unwanted regularities in a model's output. Instructions that may seem straightforward for humans may not always produce predictable effects on model behaviour. In this sense, LLMs need to be guided to attend to aspects of the text differently than human translators and will often require additional concrete examples rather than relying solely on abstract linguistic instructions.

More explicit commands as the following worked much better, but were still sometimes followed inconsistently: *"Prefer full clauses with clear subjects and finite verbs. Participle constructions can be used where appropriate, but sparingly. Avoid using participle constructions such as 'emerging as…' or 'driven by…' repeatedly across consecutive sentences or stacking them in a single sentence as in the following non-preferred example: 'Antimicrobial resistance, emerging as a major public health concern and driven by the overuse of antibiotics, occurs when microorganisms, exposed to repeated treatments and adapting to survive hostile environments, evolve to resist the effects of medicines designed to kill them.'"*

Inconsistencies in following such prompt instructions with linguistic specifications do not necessarily arise from ambiguous instructions; rather, they reflect the fact that LLMs cannot reason about or systematically apply human-defined linguistic rules for translation purposes. Translation pedagogy can leverage this challenge by helping students learn to reframe linguistic constraints as prompts that LLMs are most likely to interpret reliably. This effect appears to be especially pronounced when the model is given a dedicated revision step after the initial task prompt. At the same time, instances where LLMs fail to follow instructions may serve as opportunities for cultivating critical AI literacy.

Overall, the proposed framework led to improved translation outputs, reduced reliance on trial-and-error prompting and heightened student awareness of model-specific strengths and limitations. It fosters more responsible and effective use of AI tools in translation tasks and can be adapted to similar translation contexts or other domains that require context-sensitive language production. Far from diminishing the role of linguistic expertise, the use of LLMs highlights the necessity of human knowledge and the capacity to craft precise, contextually informed instructions. Students learn to engage with LLMs as systems that require active, linguistically grounded guidance. More broadly, this approach provides a sustainable and pedagogically meaningful framework for integrating LLMs into translation training while reinforcing core disciplinary competencies and human decision making as central to the process.

4 Carbon impact statement

This study involved classroom-based translation tasks in which students used existing AI-powered tools to translate texts of different lengths to refine their prompting strategies. The computational resources required for the study were not directly measurable. The research also aimed to explore efficient uses of AI in translation workflows and training, for example by reducing unnecessary iterations and limiting the number of prompts needed to achieve satisfactory results, thereby contributing to more resource-efficient practices in the future.

Acknowledgements

The author would like to thank Martin Czygan from the Leipzig University Library for sharing his expertise in a workshop held within the discussed translation seminar. During the workshop, he introduced local LLMs, demonstrated the functions of, for instance, the Academic Cloud, Jan.AI and TranslateGemma, provided an overview of benchmarks and the

possibilities of different LLMs and supplied a repository of documentation.

References

Angelone, Erik. 2026. Generative AI as a facilitator of deliberate practice in translator training. In Penet, J. C., Joss Moorkens, and Masaru Yamada, editors, *Teaching Translation in the Age of Generative AI: New Paradigm, New Learning?*, pages 27–47. Language Science Press, Berlin.

Ayvazyan, Nune, Yu Hao, and Anthony Pym. 2024. Things to do in the translation class when technologies change: The case of generative AI. In Peng, Yuhong, Huihui Huang, and Defeng Li, editors, *New Advances in Translation Technology: Applications and Pedagogy*, pages 219–238. Springer, Singapore.

Ayvazyan, Nune. 2025. Human-centered pedagogies in the age of generative AI: Examining student perceptions of augmentation, progress and agency. *InContext: Studies in Translation and Interculturalism*, 5(1):167–193.

Burkert, Tomáš. 2025. LLM benchmarking leaderboard: Languages and creativity tasks. *RWS Blog.*

Cui, Feng, Defeng Li, and Chiyuan Zhuang. 2025. Introduction: Transforming translation education through Artificial Intelligence. *The Interpreter and Translator Trainer*, 19(3-4):227–233.

Finkelstein, Mara, Isaac Caswell, Tobias Domhan, Jan-Thorsten Peter, Juraj Juraska, Parker Riley, Daniel Deutsch, Geza Kovacs, Cole Dilanni, Colin Cherry, Eleftheria Briakou, Elizabeth Nielsen, Jiaming Luo, Kat Black, Ryan Mullins, Sweta Agrawal, Wenda Xu, Erin Kats, Stephane Jaskiewicz, Markus Freitag, and David Vilar. 2026. *TranslateGemma technical report*.

Gao, Yuan, Ruili Wang, and Feng Hou. 2023. How to design translation prompts for ChatGPT: An empirical study. *arXiv*.

Halliday, Michael A. K. and Christian M. I. M. Matthiessen. 2013. *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge, London/New York.

He, Sui. 2024. Prompting ChatGPT for translation: a comparative analysis of translation brief and persona prompts. *arXiv*.

Leo S. Lo. 2023. The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship,* 49(4):1–3.

Pym, Anthony and Yu Hao. 2024. *How to Augment Language Skills: Generative AI and Machine Translation in Language Learning and Translator Training*. Routledge, London/New York.

Ren Xiaobin and Wang Ruoxuan. 2025. Integrating human-AI collaboration into translation education: A comprehensive protocol for assessment, diagnosis and strategy development. *PLoS One*, 20(12):1–8.

Siu, Sai Cheong. 2024. Revolutionising translation with AI: Unravelling Neural Machine Translation and generative pre-trained Large Language Models. In Peng, Yuhong, Huihui Huang, and Defeng Li, editors, *New Advances in Translation Technology. Applications and Pedagogy*, pages 29–54. Springer, Singapore.

Sizov, Fedor, Cristina España-Bonet, Josef Van Genabith, Roy Xie, and Koel Dutta Chowdhury. 2024. Analysing translation artifacts: A comparative study of LLMs, NMTs, and human translations. In *Proceedings of the 9th Conference on Machine Translation*, pages 1183–1199, Miami. Association for Computational Linguistics.

Spiridonov, Ilya. 2026. Best LLM for translation in 2026: A data-driven engine scoreboard. *Alconost Blog.*

Teo, Sheila. 2023. How I won Singapore's GPT-4 prompt engineering competition. *Towards Data Science.*

Yamada, Masaru. 2023. Optimizing machine translation through prompt engineering: An investigation into ChatGPT's customizability. In *Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track*, pages 195–204, Macau. Asia-Pacific Association for Machine Translation.

Yamada, Masaru. 2026. Teaching translation with AI: Bridging theory and practice through prompt engineering. In Penet, J. C., Joss Moorkens, and Masaru Yamada, editors, *Teaching Translation in the Age of Generative AI: New Paradigm, New Learning?*, pages 87–104. Language Science Press, Berlin.

Yang, Zhishen, Tosho Hirasawa, Edison Marrese-Taylor, and Naoaki Okazaki. 2024. Large language models as manga translators: A case study. In *Proceedings of the 30th Annual Meeting of the Association for Natural Language Processing*, pages 2012–2017, Tokyo, Association for Natural Language Processing.

Zhang, Biao, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. In *Proceedings of the 40th International Conference on Machine Learning*, pages 41092–41110, Honolulu. Journal of Machine Learning Research.