Is AI Better at Translating Specialized Language than NMT? A Case-Study of English, French and Croatian Cybersecurity Terminology *

Marta Richter[1,*,†] and Dalibor Vrgoč[2,†]

[1] Faculty of Humanities and Social Sciences, University of Zagreb, Ivana Lučića 3, 10000 Zagreb, Croatia
[2] Institute of the Croatian Language, Republike Austrije 16, 10000 Zagreb, Croatia

Abstract

In this paper[1] we compare the performance of two MT tools, Google Translate (GT), an NMT system, with GPT-54, an LLM, in domain-focused terminology-sensitive translation. We investigate the translation of specialised texts related to cybersecurity, one of the core components of modern military operations. Our study includes two large world languages, English and French, as sources, and a smaller language, Croatian, as the target. The analysis is based on automatic evaluation of MT outputs, including four metrics: BLEU, COMET, BERTscore and TER, as well as on qualitative human evaluation. The automatic evaluation results demonstrate that GPT-54 slightly outperforms GT in almost all cases, with French-Croatian translations exhibiting lower quality than English-French ones, although the observed differences are not significant. Human evaluation corroborates the automatic one, revealing the highest proportion of errors in the Fluency category, which would not have a negative impact on the use of the translation. Terminological errors occur more frequently in GT's outputs, which also displays inconsistent term usage. It can generally be concluded that GPT outperforms GT in translation into Croatian; while it does better with English source texts than those in French, and while it does not exhibit terminological inconsistency, it still does not achieve human parity, and its outputs should be post-edited by an expert translator.

Keywords

NMT, AI, LLMs, translation quality assessment (TQA), specialised translation, cybersecurity terminology

1. Introduction: current state of technologies in the language industry

Large language models (LLMs), powerful AI systems trained on massive datasets to understand, generate, and process human language, became publicly available three years ago, but they are already marking a huge impact on the language industry. Terminology is no exception, and a growing number of studies are devoted to the various applications of LLMs in different terminology-related purposes. From a linguistic perspective, one the greatest advantages of LLMs is their ability to consider a broader context, i.e. analysing how a word relates to all other words within a given context.

Despite the widespread excitement sparked by LLMs, various AI offshoots have been around in the language industry for decades. One of the most important subdomains of AI is NLP, natural language processing, which includes the large field of machine translation (MT).

MT has seen rapid evolution, spanning three generations of systems based on: 1) rules (RBMT, rule-based machine translation), 2) statistical probabilities (SMT, statistical machine translation), and finally, 3) neural networks (NMT, neural machine translation), its latest development. Technical advances in machine translation have been accompanied by ever-improving results and a growing number of languages processed, and some MT service providers have already approached human-level accuracy [1].

* *5th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.*
[1,*] Corresponding author.
mvlaic@m.ffzg.hr (M. Richter); dvrgoc@ihjj.hr (D. Vrgoč)
0009-0003-0516-6528 (M. Richter); 0000-0002-6259-0187 (D. Vrgoč)

[1] The authors would like to thank three anonymous reviewers for their valuable comments and suggestions.

Given that LLMs have caused a major disruption in the language industry, the main question now is whether these systems perform better at language-related tasks, such as translation and terminology management, than state-of-the-art NMT models. This is precisely our goal in this paper: to analyse if translations produced by LLMs outperform those by NMT. In order to do so, we shall focus on translations of specialised texts from the field of cybersecurity. More precisely, we shall analyse machine translations generated by one of the most widely used LLMs, ChatGPT, and one of the most common NMT tools, Google Translate, and compare them with human translations.

It will also be interesting to observe the performance of ChatGPT in a smaller language like Croatian because most research to date has confirmed the strong ability of LLMs to translate resource-rich languages, such as English or German, while studies on resource-poor languages remain under-represented [2]. Similarly, [3] found that ChatGPT's performance is comparable to that of machine translation services, such as Google Translate, for major European languages, but that it struggles with resource-poor or typologically very different languages. This is precisely why our analysis will combine two large world languages, English and French, and their translation into Croatian, a moderately resourced language (cf. [4]).

The rest of the paper is organised as follows: in Section 2, we describe the domain of cybersecurity, as well as the texts chosen for translation; in Section 3, we present our methodology; Section 4 brings the results of our analysis, in Section 5 we present a discussion of the results and we finish off with concluding remarks in Section 6.

2. The cybersecurity domain

In recent years, cybersecurity has risen in prominence, and now occupies a pivotal role in modern security studies [5], driven by the widespread digitisation of critical societal and defence infrastructures. In brief, contemporary military operations depend on networked information infrastructures for command and control, intelligence, logistics, and weapons systems, making them highly vulnerable to cyber threats. As a result, both malign state and non-state actors increasingly employ cyber capabilities alongside conventional and hybrid warfare to disrupt critical infrastructure, compromise sensitive data, and influence decision-making processes. Cyberspace has thus emerged as a distinct and prominent operational domain,[2] capable of generating strategic effects comparable to kinetic actions, often characterised by high speed, asymmetry, and plausible deniability.

As cyberwarfare has emerged as a distinct operational domain, it has given rise to specialised terminology that shapes language and conceptual understanding. The emergence of new cyberwarfare terminology influences language by introducing rapidly evolving, domain-specific concepts that blur traditional distinctions between civilian, military, and technical vocabularies. Consequently, for LLMs, accurately interpreting context, intent, and legal or strategic meaning is challenging, particularly as terms such as "attack," "operation," or "deterrence" today can acquire non-kinetic and ambiguous definitions which, alongside the proliferation of synonyms, complicates translation across languages in this new domain [6].

The emergence of terms such as "cyber persistent engagement," "grey-zone cyber operations," "hacktivism-as-a-service," "cognitive cyber operations," and "cyber resilience" reflects a shift from older, technically focused cybersecurity vocabulary toward concepts that capture strategic, operational, and human-centric dimensions of cyber conflict. For LLMs, this evolution poses concrete challenges: accurately interpreting these terms requires distinguishing between kinetic and non-kinetic operations, identifying state versus proxy or non-state actors, and understanding context-specific doctrinal or legal implications. Furthermore, the proliferation of synonyms and nuanced differences in meaning – especially across languages and translations – can lead to

[2] To be more specific, recent conflicts, or more accurately wars in Ukraine and the Middle East, illustrate the growing operational significance of cyber warfare. In the ongoing Russo–Ukrainian war, sustained cyber campaigns have repeatedly targeted telecommunications, government services, and energy infrastructure, exemplified by the 2023 attack on Ukraine's largest telecom provider.

misinterpretation or oversimplification, making precise comprehension and generation of outputs dependent on specialised, domain-aware training.

While incorporating insights from NATO's recent strategies, policies, and official reports on cybersecurity and cyberwarfare [7, 8], this research focuses primarily on the NATO document *Cybersecurity: A Generic Reference Curriculum* as one of doctrinal educational frameworks.[3]

3. Research methodology

As has already been mentioned, two MT tools will be used: ChatGPT, an LLM, and Google Translate, an NMT tool. ChatGPT is based on the Generative Pre-Trained Transformer (GPT) architecture, it was developed in 2018 by the American company OpenAI [9] and made publicly available in December 2022. It is a web chatbot trained to respond to user queries that caused a global sensation and experienced the fastest growth in history [10].[4] Like other LLMs, ChatGPT was not designed exclusively for translation, but it has demonstrated sufficient technical capability to produce fluent and consistent translations, rivalling or even surpassing machine translation services such as Google Translate and DeepL [11]. It is a highly intelligent language model, capable of improving its learning through user feedback or training on new data [10]. In this study, we have used GPT-54, the newest publicly available model launched by OpenAI, which can currently not only translate "across 50+ languages",[5] but also use various inputs (text, voice, even images) to produce translations.

Google Translate (GT), Google's free machine translation service, one of the most popular worldwide, added 110 new languages in June 2024, bringing the total number of supported languages to 249, according to some sources.[6]

The texts chosen for translation were originally written in English and French, and are part of the previously described document on cybersecurity published by NATO. Given that NATO's two official languages are English and French,[7] its documents were a logical choice of source texts, as they can be considered as two equally authentic originals. The source texts contain 10,655 characters with spaces (English), and 12,771 characters with spaces (French), respectively. They were first translated by two experienced professional translators, and then reviewed by a terminologist specialising in military terminology.

After that, both texts were translated by Google Translate and GPT-54 into Croatian. GPT-54 was only given the instruction (or prompt[8]) to translate the text from English/French into Croatian, with no further details provided. The MT outputs were then first analysed on the basis of four standard metrics for automatic MT evaluation (TER, BLEU, BERTScore and Comet[9]), with the reference human translations used as a gold standard.[10] A human evaluation phase followed,

[3] The document is available here: www.nato.it/content/dam/nato/webready/documents/deep/deep-cyber-curr-en.pdf (access in April 2026).

[4] It had more than 100 million monthly users before January 2023 [10]. According to some sources, it has more than 700 million weekly active users, or about 10% of the world's adult population, with 18 billion messages exchanged each week in 2025 (https://www.usine-digitale.fr/article/chatgpt-entre-les-mains-du-grand-public-ca-donne-quoi.N2237822).

[5] See: https://chatgpt.com//translate/ (access in January 2026).

[6] See: https://blog.google/products/translate/google-translate-new-languages-2024/ and https://en.wikipedia.org/wiki/Google_Translate (access in January 2026).

[7] See: https://www.nato.int/en/work-with-us/careers/freelance-interpreters (access in January 2026).

[8] Unlike MT tools, that produce translations without human intervention, ChatGPT requires an instruction or prompt to generate the desired responses. Prompts play a crucial role in optimizing results: while a well-designed prompt can significantly improve the quality of ChatGPT's output, a poorly written one can lead to erroneous responses. The importance of prompts is such that a whole new discipline called prompt engineering has emerged, referring to "the science and art of designing, formatting and optimizing conversational prompts to better guide the discourse with AI or machine learning models" [12].

[9] TER and BLEU are metrics that operate at the surface level of the sentence, while BERTscore and COMET are neural metrics, which rely on pre-trained language models and use machine learning [13]. Higher scores indicate better-quality translations in all cases except in the case of TER, where lower scores are an indication of higher quality due to the fact that this metric calculates the number of edits needed for the MT output to approach the quality of the human reference translation.

[10] In MT evaluation, translations produced by professional translators used as reference translations are often called the *gold standard* [14].

combining a qualitative, error-based analysis according to a classification system developed specifically for Croatian [15] with MQM[11] error typology.

4. Results of automatic and human evaluation

In this chapter, we first bring the results of automatic MT quality evaluation, which will be followed by human evaluation.

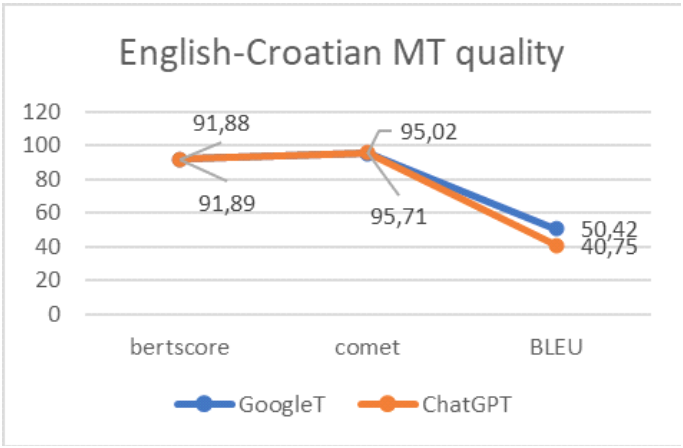
4.1. Automatic evaluation

Below is Table 1 providing MT quality scores through all four automated metrics. The bolded scores indicate the best results per metric, while the highlighted ones indicate high-quality results.

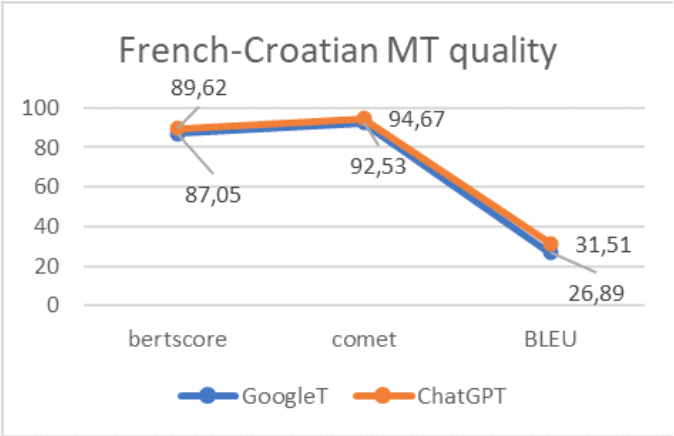
Table 1
MT quality scores according to four automated metrics

MT system	BERTscore	Comet	BLEU	TER
GoogleT Eng-Cro	91,88	95,02	**50,42**	**40,78**
GoogleT Fre-Cro	87,05	92,53	26,89	62,81
ChatGPT Eng-Cro	**91,89**	**95,71**	40,75	48,02
ChatGPT Fre-Cro	89,62	94,67	31,51	57,06

It can be concluded that the highest quality MT output is the one generated by GPT-54 from English into Croatian according to BERTscore and Comet, while it has a lower BLEU score. The only high-quality MT according to BLEU is GT from English into Croatian. In other words, across three metrics, English-Croatian MTs exhibit higher quality than the French-Croatian ones. The quality gap between translations from French and English can be observed by contrasting the two graphs below, where English performs better on three metrics (BERTscore, Comet, BLEU). It must be emphasised, however, that the differences between English-Croatian and French-Croatian translation quality are not significant (see Figures 1 and 2 below).

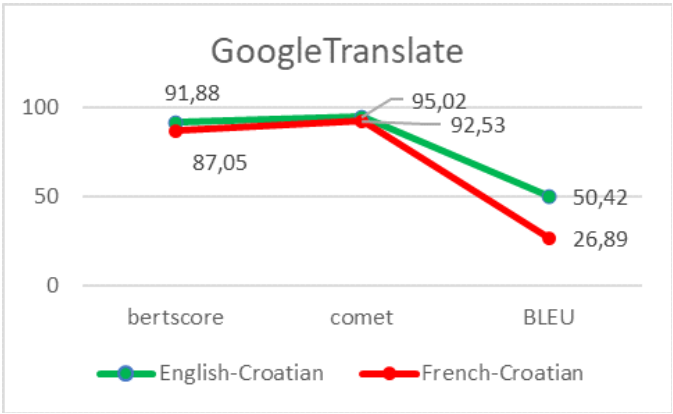


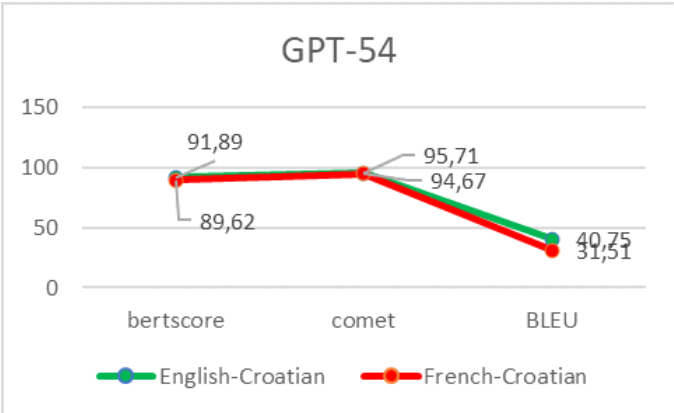
[11] MQM (Multidimensional Quality Metrics) is a framework for analytic translation quality evaluation which has been "widely used by practitioners in the translation and localization industry" ever since its introduction in 2014 [15].



Figures 1 and 2: Comparison of English-Croatian MT quality and French-Croatian MT quality

As to the difference in quality between the two MT tools, GPT-54 has obtained slightly better results than GT according to BERTscore and Comet (see Figures 3 and 4 below). However, the difference between their scores is not significant, except in the case of BLEU score for the English-Croatian combination, in which GT has a 9.7-point advantage over GPT-54. Both MT tools generally produced slightly lower quality results in translations from French, and only GT's English-Croatian output is above BLEU quality threshold (50).





Figures 3 and 4: Comparison of GT and GPT-54 output quality across three metrics

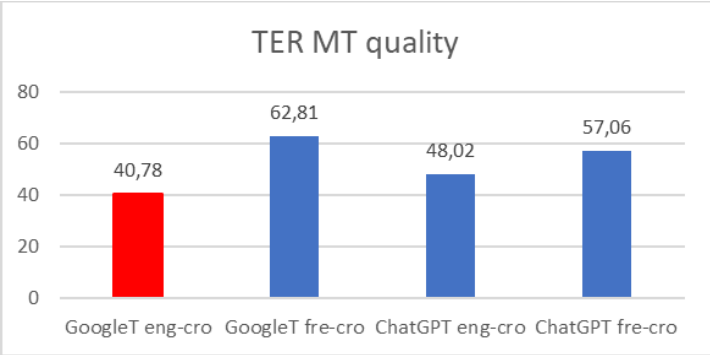


Figure 5: MT outputs according to TER

With respect to the TER metric (Figure 5 above), only the GT translation from English into Croatian can be considered moderate-quality. The English-Croatian MT produced by ChatGPT is close in score, but both French-Croatian translations are of lower quality.

4.2. Human evaluation

It is known that human evaluation, especially the one done by expert translators, is the most adequate assessment of the quality of a translated text (e.g. [16]). In spite of this fact, a majority of studies dealing with MT quality are entirely based on automated metrics [17], due primarily to the fact that human evaluation is a time consuming and costly process. Human and automated quality assessment may differ largely, because human evaluators tend to focus on somewhat different phenomena when assessing the quality of a translated text [17]. Bearing in mind that human judgement is vital for fine-grained MT performance evaluation, in this paragraph we introduce the results of a human quality assessment of the produced machine translations which was performed by an experienced translator[12] and verified by a terminologist familiar with cybersecurity terminology.

In the following part, we bring an overview of all error types found in various machine translations. Our methodology is a combination of MQM criteria and those introduced by [18]. We have opted for a combined error analysis due to the specificities of the languages involved, especially with respect to the target language, Croatian, which is in the focus of this study. As a typical Slavic language rich in inflectional morphology, it required an extended version of the Grammar error category, with verbal tense or form being one of the typical errors due to the intricacies of the Slavic verbal system, and a general "Grammatical form" category which included erroneous inflectional forms of all other parts of speech. In the second category, the most problematic words were almost exclusively pronouns, which are often implied in Croatian[13] and do not need to be explicated because this renders the text less natural.

Table 2
Types and sub-types of errors in MT outputs based on human evaluation

			GT ENG>CRO	GPT ENG>CRO	GT FRE>CRO	GPT FRE>CRO
Fluency	Orthography	Spelling & capital letters	4	9	11	13
		Punctuation	18	20	13	9
	Grammar	Verbal form/tense	10	2	6	2
		Grammatical	7	10	19	5

[12] The translator specializes in English-Croatian-English and French-Croatian-French translation. Considering the fact that there was only one evaluator, there was no need to assess inter-evaluator agreement.
[13] Croatian is a pro-drop language (e.g. [19]).

	form				
	Calqued phrase	6	7	3	6
Accuracy	Omission	5	5	7	11
	Lexical choice	13	7	12	6
Terminology	Wrong term	16	11	23	20
Style	Awkward	1	-	2	2
Total		80	71	96	74
Total (without Fluency errors)		35	30	44	39

A closer look at the error categories in Table 2 above shows that the highest number of errors can be found in the *Fluency* category, across all MT engines and language combinations. These, however, are not major errors, as they would not have a negative impact on the use of the translation or its understanding. As for Fluency error subtypes, the highest number of errors in capital letters was found in translations from French, due to the capitalisation of *Internet* in the French language, which is not capitalised in Croatian. When it comes to punctuation, the errors primarily concerned the retention of semicolons in bulleted list formatting, which should be left out in Croatian;[14] this was a very frequent error as the original documents contained a number of bulleted lists.

In the *Accuracy* category, *omissions* refer in a large number of cases to the lack of explanations of abbreviations borrowed from English, such as in the case of the term *napredne trajne prijetnje* (*advanced persistent threats/menaces persistantes avancées*), for which the Croatian abbreviation remains the same as in English (APT), so it should be explicated, at least the first time the term appears in the text. In other frequent cases, the name of an authority, such as IANA,[15] was not translated into Croatian. The *Lexical choice* subcategory mostly contains cases of collocational misuse.

With respect to cybersecurity terminology, there appeared more errors in GT's translation output than in the one produced by GPT, and more in translations from the French source than from English. It should be pointed out, however, that some of the inaccurate terms were very similar to the official ones (**sigurnost informacija* instead of *informacijska sigurnost* for *information security/sécurité de l'information*; **ciljevi sigurnosti* instead of *sigurnosni ciljevi* for *security objectives/objectifs de sécurité*, etc.). This category also included inconsistent terminology use, which was only recorded in GT's outputs, both English-Croatian and French-Croatian.

On the basis of the results presented in Table 2, it can be concluded that French-Croatian translations contain a higher number of errors, perhaps owing to a reduced volume of training data available for both Croatian and French with respect to English.

As a general conclusion, taking into account the error statistics (and leaving out the Fluency category as not highly relevant), the best translation is the one generated by GPT from English into Croatian, followed by GT's translation with the same language combination. These results corroborate the findings obtained through automatic metrics, where GPT's English-Croatian translation scored best according to BERTscore and Comet, and GT with the same combination scored best based on BLEU. While French-Croatian translations are generally lower-quality with respect to the English-Croatian ones, GPT scored better than GT in these cases as well.

5. Discussion

In spite of its limited scope resulting from a rather small dataset, our study sheds some more light on the quality of MT outputs that combine English, as a well-resourced language, with French and Croatian, two more moderately resourced languages from different branches of the Indo-European

[14] According to Croatian spelling rules, there is no punctuation within list items, cf. https://pravopis.hr/pravilo/zarez/60/.
[15] IANA stands for the *Internet Assigned Numbers Authority*, whose Croatian equivalent is *Međunarodna organizacija za dodjelu brojeva*.

family. Our automatic evaluation has demonstrated that GPT-54 generally outperforms GT, a more traditional model based on NMT, both according to automatic and human evaluation. While automatic metrics show smaller differences between GT's and GPT's outputs, the results of the human evaluation are more nuanced. More precisely, human evaluation has not only identified the weaker areas with a higher error rate (punctuation, grammatical form and terminology), but it has also helped us to refine the results of the automatic evaluation. We can conclude that our results show a high correlation of automatic evaluation with human judgment. Human evaluation has confirmed that GPT outperforms GT in both language combinations (English-Croatian and French-Croatian), as well as the fact that translations from French exhibit a higher proportion of errors than the English-Croatian ones. In addition, human evaluation has demonstrated that while both GT and GPT-54 are partially challenged by cybersecurity terminology, GPT demonstrates superior performance than GT in this field. Moreover, GT's outputs exhibit instances of inconsistent terminology use.

6. Concluding remarks

LLMs are profoundly reshaping the translation industry as one of the most popular AI applications. Since their public release, these new powerful tools have challenged the more traditional ones that have been developing in the language industry for decades, with NMT as its latest and most successful development. This paper contributes to a better understanding of whether LLMs outperform NMT tools in specialized translation. We have focused on two languages pairs: English-Croatian and French-Croatian, and two MT tools, Google Translate and GPT-54. On the basis of automatic evaluation metrics (BLEU, TER, Comet and BERTscore), the quality of the translations was compared to a human gold standard. Our results indicate that GPT-54 outperforms GT across both language combinations, although the performance gap remains modest. English-Croatian MT consistently yields outputs of marginally higher quality compared to French-Croatian. In order to refine our results, we have also incorporated human evaluation by an expert translator and a terminologist. According to human evaluation, categories with higher error rates are Fluency and Terminology. While Fluency is not very relevant for the use of the translations, when it comes to Terminology, GPT has demonstrated a better command of cybersecurity terminology than GT, which also exhibits inconsistencies in its use. Overall, while both MT engines produce translations of relatively high quality in the domain of cybersecurity, especially with English as a source text, GPT-54 performs slightly better than GT. Nevertheless, neither system has yet attained human parity.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-54 in order to generate machine translation outputs that were analyzed in their study.

References

[1] S. Chen, C. Liu, M. Hague, Z. Song, W. Yang, NMTSloth: understanding and testing efficiency degradation of neural machine translation systems, in: A. Choudhoury, C. Cadar, M. Kim (Eds.), Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022), Communications of the ACM 50, Association for Computing Machinery, New York, 2022, pp. 36–44. doi:10.1145/3540250.3549102.

[2] Z. Jiang, Z. Zhang, Can ChatGPT rival neural machine translation? A comparative study, arXiv:2401.05176v1 (2024) 1–20.

[3] W. Jiao, W. Wenxuan, J. Huang, X. Wang, S. Shi, Z. Tu, Is ChatGPT a Good Translator? Yes With GPT-4 as the Engine, arXiv:2301.08745v4 (2023). doi: 10.48550/arXiv.2301.08745.

[4] V. Štefanec, D. Farkaš, G. Thakkar, M. Tadić, Building a Large Language Model for Croatian, in: C. Orasan, T. Ranasinghe, G. Corpas Pastor, R. Mitkov, M. Kunilovskaya, V. Sosoni, M. Escribe (Eds.), Proceedings of New Trends in Translation and Technology Conference, NeTTT 2024, Incoma Ltd, Varna, 2024, pp. 204–209. doi: 10.26615/issn.2815-4711.2024_017.
[5] F. Jimmy, Emerging Threats: The Latest Cybersecurity Risks and the Role of Artificial Intelligence in Enhancing Cybersecurity Defense, International Journal of Scientific Research and Management (IJSRM) 9 (2021) 564–574. doi:10.18535/ijsrm/v9i2.ec01.
[6] D. Vrgoč, Postmodernizam, jezična politika i ratovanje na Bliskome istoku: odraz neoteričnoga eksperimentiranja u vojnoj terminologiji, Anali Hrvatskog politološkog društva (2021) 215–232. doi:10.20901/AN.18.06.
[7] NATO 2022 Strategic Concept, 2022. URL: https://www.nato.int/content/dam/nato/webready/documents/publications-and-reports/strategic-concepts/2022/290622-strategic-concept.pdf.
[8] M. N. Schmitt (Ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations United States Naval War College, Newport, Rhode Island, 2017.
[9] K. I. Roumeliotis, N. D. Tselikas, ChatGPT and Open-AI Models: A Preliminary Review, Future Internet 15 (2023) 1–24. doi:10.3390/fi15060192.
[10] U. Kamath, K. Keenan, G. Somers, S. Sorenson, Large Language Models: A Deep Dive. Bridging Theory and Practice, Springer, Berlin, 2024, doi:10.1007/978-3-031-65647-7.
[11] T. K. Lee, Artificial intelligence and posthumanist translation: ChatGPT versus the translator, Applied Linguistics Review 15 (2023) 1–22. doi:10.1515/applirev-2023-0122.
[12] D. C. Schmidt, J. Spencer-Smith, Q. Fu, J. White, Towards a Catalog of Prompt Patterns to Enhance the Discipline of Prompt Engineering, Ada Letters 43 (2024), 43–51. doi:10.1145/3672359.3672364.
[13] M. Nakhlé, L'évaluation de la traduction automatique du caractère au document : un état de l'art, in: M. Candito, T. Gerald, J. G. Moreno (Eds.), Actes des 16e Rencontres Jeunes Chercheurs en RI (RJCRI) et 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL), ATALA, Paris, 2023, pp.143–159.
[14] B. Babych, Automated MT evaluation metrics and their limitations, Revista Tradumàtica: tecnologies de la traducció 12 (2014) 464–470. doi:10.5565/rev/tradumatica.70.
[15] A. Lommel, S. Gladkoff, A. Melby, S. E. Wright, I. Strandvik, K. Gasova, A. Vaasa, A. Benzo, R. Marazzato Sparano, M. Foresi, J. Innis, L. Han, G. Nenadic, The Multi-Range Theory of Translation Quality Measurement: MQM scoring models and Statistical Quality Control, in: R. Knowles, A. Eriguchi, S. Goel (Eds.), Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 2: Presentations), Association for Machine Translation in the Americas, Chicago, 2024, pp. 75–94. doi:10.48550/arXiv.2405.16969.
[16] E. Chatzikoumi, How to evaluate machine translation: A review of automated and human metrics, Natural Language Engineering 26 (2020) 137–161. doi:10.1017/S1351324919000469.
[17] Z. Jiang, Q. Lv, Z. Zhang, L. Lei, Convergences and Divergences between Automatic Assessment and Human Evaluation: Insights from Comparing ChatGPT-Generated Translation and Neural Machine Translation, arXiv:2401.05176 (2024), doi:10.48550/arXiv.2401.05176.
[18] N. Pavlović, Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka, in: D. Stolac, A. Vlastelić (Eds.), Jezik kao predmet proučavanja i jezik kao predmet poučavanja, Srednja Europa/Hrvatsko društvo za primijenjenu lingvistiku, Zagreb, 2016, pp. 279–295.
[19] P. Fabijanić, M. Palmović, Pronoun reference resolution in a pro-drop language, in: M. Racsmany (Ed.), Learning and Perception, Wolters Kluwer, Budapest, 2013, pp. 38–39.