

# Artificial Intelligence and Language Translation in Scientific Publishing

Sarah Frances Gordon

## Introduction

Language translation is an important part of the scholarly and scientific publishing profession, as it allows knowledge to be communicated across the globe, transcending language barriers. Through translating research texts, authors can expand their readership and ensure that valuable scientific knowledge is widely accessible. Translation services are key to the scholarly and scientific research community because they help improve global collaboration. However, with the recent introduction of artificial intelligence (AI), the landscape of language translation has changed.

## Neural Machine Translation

Neural machine translation (NMT) systems such as Google Translate, DeepL Translator, and Microsoft Translator are just a few of the NMT systems that translation companies and professionals often use. NMT aims to create algorithms that can translate text between different languages. Neural networks, which are computer systems inspired by the human brain, constitute the foundation of NMT models. These models learn to provide translations by spotting patterns in the massive volumes of text in several languages they are trained on. Another key feature of NMT models is that text translation can be learned without the need of mathematical models or explicit rules. However, more recently, ChatGPT has become one of the most common AI tools used by professional translators, as it has an easy-to-use interface and is suited to both individual and wide-scale use.

Sarah Frances Gordon (<https://orcid.org/0000-0001-5131-8519>) is with Universidad Iberoamericana, Mexico City, Mexico.

Opinions expressed are those of the authors and do not necessarily reflect the opinions or policies of the Council of Science Editors or the Editorial Board of Science Editor.

<https://doi.org/10.36591/SE-4701-05>

## AI Changes How We Work as Translators

With the technological advances of AI in translation, many professional translators perceive AI and automation as threats to the profession and distrust the recent advances in AI.<sup>1,2</sup> The fear is that AI will change the nature of a translator's job, leaving professional translators to predominantly edit machine-translated texts and train AI to perform machine translations.<sup>1,2</sup> In a way, this fear is justified because it is true that AI is changing the way we perform our jobs as translators. In many scientific journals, AI technology has been integrated into the editorial workflow and is built into the interface that handles the submissions, peer review, corrections, and editorial and production processes involved in publishing a journal article. However, machine-generated translations are often checked by human translators or another AI tool. This has been shown to enhance the editorial process. Also, many of us already double-check our translations using technology such as GPT-4 and use applications such as Grammarly<sup>3</sup> and Paperpal<sup>4</sup> to check our grammar. The fear is that the creativity required by translation and linguistic skills will be lost amidst all the new technological advances. In scholarly publishing, where objectivity is vital, we must examine the AI tools we are using in our work and critically reflect on the potential risks associated with them.

## Inherent Linguistic Bias In AI

NMTs and other large language models (LLMs) have a significant imbalance in their coverage of languages, and these systems tend to perform better with high-resource languages such as English, Spanish, Chinese, and French. Even advanced LLMs, like ChatGPT, have imbalances because they are primarily designed to work more effectively in English than any other language. This imbalance means that texts or translations in languages other than English will not be as accurate or as culturally relevant as they should be.<sup>5</sup> In addition, AI technology, such as GPT-4, seems to be able to translate many languages into English, but they start to experience problems when they try to translate English into any other language, especially those with non-Latin alphabets, such as Korean.<sup>6</sup>

## CONTINUED

The prioritization of English as the lingua franca in the scientific world is not a new phenomenon and stems from a history of colonization. Therefore, it is logical that this bias should exist in the AI tech space, and it is no surprise that most NMTs and LLMs struggle to capture the context and richness of languages that are not English.<sup>5</sup> Many languages spoken by smaller populations in underrepresented regions or with a smaller online presence are underrepresented in the development of NMTs and LLMs. Critics argue that AI technology may help us translate dominant languages in the Western world such as English, Spanish, and French. Still, the same models and systems struggle to do the same for languages considered “low-resource” such as Bengali, Swahili, isiXhosa, Tigrinya, Tamil, or Amharic.<sup>7</sup> It is no surprise that low-resource languages are often from developing countries with histories of colonization and oppression.<sup>8</sup> Although some organizations and researchers are actively working on the development of machine translation models for low-resource languages to make language technology more inclusive and accessible, these biases still exist. Therefore, it is our responsibility in the scientific and scholarly publishing community to be aware of these potential biases and take measures to combat them.

## A Way Forward

Researchers and scholars all benefit from more accurate translations; therefore, we must be aware of the potential errors and biases in LLMs and NMTs. While human translators are considered expensive, they are accurate and when trained well, they are at much less risk of making errors in their translation work than AI technology. The human touch is invaluable because translation work is not an exact science. Translation is also about preserving the author’s voice and keeping the cultural nuances and tone of text intact across languages. This requires a splash of creativity that is difficult to program into an AI model.

However, when debating the values of AI technology, we should not be so quick to “throw the baby out with the bath water” (an English idiom, originally translated from the German, “das kind mit dem bade ausschütten”) as AI does have some valuable contributions to make. Despite the current challenges, I argue that AI technology should be used to complement the work of translators in the scientific and scholarly fields. I believe that the way forward in translation work is a hybrid model that draws on AI technology and human review.

We should call on technological companies in the AI arena to be more inclusive and engage in constant monitoring, refining algorithms, and incorporating diverse datasets to ensure translations are accurate and do not contain cultural or contextual biases. It is important that all languages are represented in the digital space, and global linguistic diversity is maintained. On the side of translation companies, they need to be mindful of the technology with which they are engaging and incorporate AI technology with human review because a collaborative approach between human expertise and AI will ensure the highest quality of translated content.

## Concluding Thoughts

The future of language translation should be a hybrid model that integrates AI technologies. Continued collaborations between researchers, linguists, and AI experts will lead to more sophisticated models capable of handling different languages, which are hopefully able to capture cultural nuances. As AI continues to evolve, researchers, authors, and publishers should navigate the ethical considerations associated with bias and ensure that the human touch remains integral in the translation process. The future holds exciting possibilities for the translation field in the scholarly publishing and scientific community, and these advancements will help ensure the dissemination of knowledge across language barriers.

## References and Links

1. Kirov V, Malamin B. Are translators afraid of artificial intelligence? *Societies*. 2022;12:70. <https://doi.org/10.3390/soc12020070>.
2. Tavares C, Oliveira L, Duarte P, da Silva MM. Artificial intelligence: a blessing or a threat for language service providers in Portugal. *Informatics*. 2023;10:81. <https://doi.org/10.3390/informatics10040081>.
3. <https://www.grammarly.com/>
4. <https://paperpal.com/>
5. Nicholas G, Bhatia A. Lost in translation large language models in non-English content analysis. *The Center for Democracy & Technology*, 2023. <https://doi.org/10.48550/arXiv.2306.07377>.
6. Dave P. ChatGPT is cutting non-English languages out of the AI revolution. [accessed February 5, 2024]. *Wired*, 2023. <https://www.wired.com/story/chatgpt-non-english-languages-ai-revolution/>.
7. Deck A. We tested ChatGPT in Bengali, Kurdish, and Tamil. It failed. [accessed February 5, 2024]. *Rest of World*, 2023. <https://restofworld.org/2023/chatgpt-problems-global-language-testing/>.
8. Expert insights on the use of MT for low-resource languages. [accessed February 5, 2023]. *MachineTranslation.com*, 2023. <https://www.machinetranslation.com/blog/mt-for-low-resource-languages>.