

TO: CAIDP AISCI (AI Contract Index) Team
FROM: Marc
SUBJECT: Use of Google Translate and Other Machine Translation Tools
DATE: October 12, 2020

The question has been raised whether to use Google Translate (GT) to convert non-English passages from relevant documents into English for inclusion in the AISCI. The arguments in favor include (1) the original text may not be available in an authoritative English translation, (2) GT provides a reasonably accurate translation. There is no dispute as to the treatment, for citation purposes, of the titles of documents – they should appear in the original language.

There are, at least, two considerations to answer this question: (1) does this technique ensure the authority of the central text? (2) would we violate one of the key AI norms, i.e. to ensure human review of automated decisions?

On point one, government agencies and courts have both said that machine translation is not authoritative. For example the US Department of Justice and the Department of Education wrote in 2015:

“Some school districts have used web-based automated translation to translate documents. Utilization of such services is appropriate only if the translated document accurately conveys the meaning of the source document, including accurately translating technical vocabulary. The [U.S.] Departments [of Justice and Education] caution against the use of web-based automated translations; translations that are inaccurate are inconsistent with the school district’s obligation to communicate effectively with LEP parents. . . .

Thus, to ensure that essential information has been accurately translated and conveys the meaning of the source document, the school district would need to have a machine translation reviewed, and edited as needed, by an individual qualified to do so.”¹

In a study published in 2019 by the Journal of the American Medical Association, researchers analyzed 100 sets of emergency discharge instructions translated by Google’s machine learning algorithm.² They found the algorithm was 92% accurate for Spanish and 81% accurate for Chinese. Google Translate, according to the study, had difficulty when physicians used colloquial terms such as “skip a meal,” a phrase that Google translated into Chinese as “jump over” a meal. In addition, when a physician told a patient to “hold the kidney medicine,” meaning to stop taking it, Google translated the statement in Spanish as, “keep the medication” and in Chinese as, “keep taking” the medication.

¹ *Joint Letter from U.S. Department of Justice and U.S. Department of Education* (Jan. 7, 2015), <https://www2.ed.gov/about/offices/list/ocr/letters/colleague-el-201501.pdf>

² Bromberg Translations, *To Google Translate or Not (That is the question)* (June 5, 2019) <https://www.brombergtranslations.com/to-google-translate-or-not-that-is-the-question/>

A recent study Information, Communication & Society on Machine Translation (MT) presented “the first structured literature review of the implications of misusing MT as a communication tool in medical and legal setting.”³ From the study:

Particularly in high-stakes settings, misuse of MT can have serious consequences. In one recent case, evidence was dismissed in court because consent to perform a police search had been obtained with Google Translate, which raised concerns about the consent’s validity (Grosdidier, [2019](#)). In a medical setting, an evaluation of errors that could be caused by MT revealed that the sentence ‘your child is fitting’ would in one case have been translated to Swahili as ‘your child is dead’ (Patil & Davies, [2014](#)).

On point two, almost all AI frameworks endorse “human review” as a key element of AI accountability. We run the risk of ignoring a basic AI tenant if we adopt machine translation, without human review, to determine the meaning of essential texts, such as a National AI Strategy.

Related Services

Much of the MT focuses on GT but there are related services that may be preferable. For example the European Commission has established the Connecting Europe Facility (CEF) that provides eTranslation services.⁴ CEF eTranslation:

is a machine translation tool, which draws upon decades worth of work by EU translators (over 1 billion sentences in the official languages of the Union) and is designed to **retain the format of structured documents** during translation. It can **translate multiple documents to multiple languages** at once and **guarantees continuous service** of high quality, with due consideration for the **confidentiality and security of data** during the translation process.⁵

CONCLUSION

The problem with machine translation is similar to the problem with secondary sources. Both rely on the underlying text and both interpret the underlying text, but neither presents the actual underlying text and therefore neither can be considered authoritative. There are also privacy and security concerns associated with the use of unregulated Internet services.

³ Lucas Nunes Vieira, Minako O’Hagan & Carol O’Sullivan (2020) *Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases*, Information, Communication & Society, DOI: [10.1080/1369118X.2020.1776370](https://doi.org/10.1080/1369118X.2020.1776370)

⁴ CEF Digital, <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/>

⁵ CEF Digital, *CEF eTranslation really is one in a million* (Oct. 11, 2019) <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2019/10/11/CEF+eTranslation+really+is+one+in+a+million>

We should use Google Translate and other machine translation tools, such as CEF eTranslation, to help interpret foreign language text, but the country investigator remains responsible for the accuracy of the text and should not rely solely on machine translation. In the rare instance that we can obtain only a Google Translate or other Machine Translate version of a text that is excerpted in the AISCI, the notation “(GT),” “(MT),” or “(ET)” should be included in the citation to indicate that the text was produced by Google Translate or other Machine Translate technique.