Towards Better Understanding of Cybercrime The Role of Fine-Tuned LLMs in Translation

DECISION: Accept

AUTHORS: Veronica Valeros, Anna Širokova, Carlos Catania and Sebastián García.

Summary of Reviews

- Review 1: 1 (2)
- Review 2: 1 (3)

Reviews

Review 1

Total score: 1

Overall evaluation: 1

Reviewer's confidence: 2

I enjoyed reading this paper, which is a proof of concept for an economically affordable (when one does not factor in environmental externalities) method to perform almost real-time translation of hacker fora.

The paper does a good job of explaining the method to non-LLM specialists, such as myself. Yet, considering WACCO's interdisciplinary audience, authors could further help by explaining jargo (eg BLEU and other metrics), provide a list of acronyms and add tag steps in figure 2 to align with description at section 3. Puzzling to a non-specialist audience is also how 130 lines of messages are sufficient to train the LLM system: can you explain how this is so? If humans used to fine-tune the LLM and carry out the validation were not qualified translators, please avoid the use of this term, sticking to the periprhases you use throughout.

At s 3.4 'the privacy disclosure' : do you mean data protection / privacy information notice? informed consent? this is not very clear.

On a substantive level, authors should explain in the introduction the role of NoName057(16) in the conflict against Ukraine. The introduction should be revised to account for the fact that the paper does not engage in a substantive analysis of the hacktivist fora active in the conflict, but rather in a proof of concept, which is the reason why the evaluation is one of weak accept.

Review 2

Total score: 1

Overall evaluation: 1

Reviewer's confidence: 3

In a sense, this is another paper about "machines do it better than humans" and while the results that the LLM is, indeed, more efficient, this is still an important (if inevitable) finding. Presumably the methods could be replicated to other languages just as easily, too, I imagine (which makes this initial work more foundational).

The authors write, "Respondents also mentioned they were "irritated" and "triggered" by the messages they were asked to review, that they had to "take breaks"" -- I'm intrigued by this statement. What was it about the text that triggered the translators?

In fact, if it's true the for some reason, translating these kinds of texts was psychologically burdensome for humans, than this provides even more justification for automated techniques -- an insight which the authors could include in their initial list of justifications at the start of the paper.