

# Technical-Term Mistranslation by LLMs and Expert Responsibility:

## Deferentialism on Terminological Translation and Wiktionary

Haruto Aoki

Institute for the Delegation of Blame

*Associate Professor of Applied Blame Allocation*

haruto.aoki@deference.ac.jp

September 12, 2026

### Abstract

This paper examines moral responsibility for mistranslations of technical terms by large language models (LLMs) in light of expert contributions to public lexical resources. It argues that providing normatively appropriate translations in Wikipedia and Wiktionary may improve translation by LLMs that use these resources, and proposes the notion of deferentialism on terminological translation. It also questions the exclusive attribution of moral responsibility for mistranslation to model developers.

**Keywords:** large language models; machine translation; Wiktionary; semantic deference; expert responsibility

“The opening hours are wrong. Get your act together, Google!”

An acquaintance, opening a shop’s listing in Google Maps

## 1 Introduction

Large language models (LLMs) can be used for machine translation (MT). For example, uploading a PDF of a paper to OpenAI’s ChatGPT and entering a prompt requesting a translation will produce a reasonably serviceable translation. Similarly, supplying Google’s NotebookLM with academic materials in a foreign language will elicit an account of their contents in Japanese. A number of researchers must already be using LLMs as machine translators.

Yet LLMs frequently mistranslate. They get technical terms wrong. For example, they may translate “realism” as 現実主義 (“genjitsushugi”) *rather than* 実在論 (“jitsuzairon”); “inflationism” as 膨張主義 (“bōchōshugi”) *rather than* インフレ主義 (“infureshugi”); “Cartesian dualism” as カルテジアン二元論 (“karutejian nigenron”) *rather than* デカルト的二元論 (“dekarutoteki nigenron”); and “foundationalism” as 基礎主義 (“kisoshugi”) *rather than* 基礎付け主義 (“kisozukeshugi”). They may also translate “idealism” as 理想主義 (“risōshugi”). In these ways, present-day LLMs sometimes produce translations that specialists find unacceptable. These errors can lead specialists to question LLMs’ translation abilities: “They are still terrible at the technical terminology in my field.”

In what follows, I identify a flaw in this reasoning and argue that LLM mistranslations are partly attributable to experts’ negligence. More specifically, I argue that mistranslations by conversational large

language models with particular architectures may be corrected through experts' provision of normatively appropriate translations in Wiktionary, an open-content system.

## 2 Why LLM Mistranslations Occur

Let us first consider why LLMs mistranslate. An LLM's output causally depends on its training data and post-training (reinforcement learning). For example, an LLM trained on data stating that "apples are blue" would presumably come to answer "blue" when a user asks, "What color are apples?" As this example illustrates, an LLM's output can be described as a function of its training data.

Of particular relevance to mistranslation is *the weighting of particular data during training*. Common models such as Gemini, Llama, and GPT-4 are known to give weight to Wikipedia data, which are regarded as high-quality data, during training (Brazilek et al., 2026). The original LLaMA report, for example, describes repeated exposure to Wikipedia data (Touvron et al., 2023). Textual data on Wikipedia thus have a substantial influence on the outputs of these conversational LLMs.

Editing Wikipedia articles to conform to appropriate standards is an effective way to prevent mistranslations. For example, if a Japanese translation of the English Wikipedia article "Truthmaker theory" (Wikipedia contributors, n.d.-a) were published on Japanese Wikipedia under the title 真にするものの理論 ("shin ni suru mono no riron"), an LLM would presumably become more likely to translate "truthmaker" as 真にするもの ("shin ni suru mono").<sup>1</sup> If a Japanese article entitled ポール・サガード ("Pōru Sagādo") were connected by an interlanguage link to the English article "Paul Thagard" (Wikipedia contributors, n.d.-b), even an LLM that struggles with proper names would presumably become more likely to render "Paul Thagard" correctly as ポール・サガード ("Pōru Sagādo").<sup>2</sup> Correctly editing Japanese Wikipedia articles can therefore increase the likelihood of improving the correctness of outputs from LLMs that give weight to these data.

## 3 Wiktionary's Influence on Technical-Term Translation

Wiktionary (Wiktionary contributors, n.d.-h) is the dictionary counterpart of Wikipedia. Unlike Wikipedia, it allows translations of terms to be specified in detail. Like Wikipedia, Wiktionary is presumably assigned weight on the basis of its perceived value during LLM training. It is therefore plausible that correctly editing Wiktionary can improve the correctness of LLM translations.

Indeed, while using LLMs to translate literature, I have encountered translations of "Brandom" as ブランドン ("Burandon") rather than ブランドム ("Burandamu"), and of "property" as 特性 ("tokusei"). I prefer 性質 ("seishitsu") for the latter, since 特性 ("tokusei") seems liable to invite confusion with "thisness" (*haecceitas*).<sup>3</sup> These mistranslations appear to stem from the absence of appropriate Japanese translations in the entries for "Brandom" (Wiktionary contributors, n.d.-b) and "property." (Wiktionary contributors, n.d.-d) In the case of "property," the Translations section lists 属性 ("zokusei"), 特性 ("tokusei"), and 特徴 ("tokuchō"), and I suspect that LLMs are being drawn toward those renderings.

<sup>1</sup> 真理の制作者 ("shinri no seisakusha") is another translation of "truthmaker." See, for example, Hashimoto (n.d.). From the standpoint of deference to experts (deferentialism), 真理の制作者 ("shinri no seisakusha") is also a normatively appropriate translation. The rendering 真にするものの理論 ("shin ni suru mono no riron") follows Akiba (2014).

<sup>2</sup> I suspect that current LLMs' difficulty in translating proper names is due to their tokenization mechanisms. In byte-level byte-pair encoding (BBPE; Wang et al., 2020), for example, individual ASCII characters are not simply vectorized as they stand. Because BBPE is a tokenization method that assumes the principle of compositionality, it may be useful for compositional elements such as morphemes while sometimes producing undesirable tokenizations of proper names, including personal names.

<sup>3</sup> Although the Japanese Wikipedia article 性質 ("seishitsu") (Wikipedia contributors, n.d.-c) explicitly identifies the term as 性質 ("seishitsu") (English: "property") LLMs sometimes translate "property" as 特性 ("tokusei"). I suspect that this is because they prioritize Wiktionary's translation equivalents.

Conversely, correcting these translations can contribute to improving the quality of LLM translation. Wiktionary is an open-content platform, so anyone with adequate access to information resources can edit it.<sup>4</sup> The appendix explains how to add a translation to an entry.

## 4 Expert Responsibility and Deferentialism on Terminological Translation

Specialists in the humanities sometimes mistakenly infer, out of ignorance of how LLMs work, that responsibility for LLM mistranslations can be attributed exclusively to their developers, such as OpenAI, Google, or Anthropic. Admittedly, that inference seems justifiable to some extent; I omit the argument here. As noted above, however, LLM outputs supervene not only on their architectures but also on their training data and the data supplied during reinforcement learning. These training data may be *public*. Failing to provide normatively appropriate translations of technical terms in the public resource Wiktionary can therefore result in LLM mistranslations.<sup>5</sup> Attributing LLM mistranslations to developers or to model architectures resembles the error of holding Google responsible for inaccuracies in user-generated information on Google Maps: the metadata displayed there are merely folk contributions submitted by users.

Let us call the view that experts constitute the normative authority for technical-term translations in LLM translation *deferentialism on terminological translation*. This concept focuses semantic deferentialism (Burge, 1979; Putnam, 1975) on the translation of terminology. Unlike ordinary speakers, LLMs are semantically deferential entities (Borg, 2026). Accordingly, normative authority over the translation of technical terms is also deferred to experts. This deference is realized through LLM architectures that assign weight, on the basis of perceived value, to data from Wiktionary and Wikipedia during training.

## 5 Conclusion

This paper has argued that responsibility for LLM mistranslations of technical terms may be partly attributable to experts' negligence in failing to publish correct translations of those terms in Wiktionary. It has also argued that deferentialism on terminological translation offers a conceptual description of how expert knowledge has authority over, and causally influences, technical-term translation by LLMs. This concept may help us understand the mechanisms underlying LLM mistranslation and show that means of reducing such errors are available.

## 6 Disclaimer

To establish the paper's central claim that "LLM mistranslations are also attributable to your negligence as an expert," a specific argument concerning the logical relationship between moral responsibility and negative causation (causation by omission) is required. Such an argument has not been provided here.

---

<sup>4</sup>Some architectures, such as Immersive Translate (Immersive Translate, n.d.) and Readable (Readable, n.d.), instead inject normatively appropriate translations into the prompt. During inference, however, such prompts are often ignored.

<sup>5</sup>From a position that endorses piracy, one possible strategy would be to enable correct translation of terminology by uploading Japanese books to shadow libraries. Such a strategy might be morally justifiable. However, this discussion appears to require attention to the difference in digitization between Japanese and foreign-language books. Japanese books are less extensively digitized. For example, of the 23 books in Shunjusha's series 現代哲学への招待 ("gendai tetsugaku e no shōtai") ("An Invitation to Contemporary Philosophy"), only Stroud's 君はいま夢を見ていないとどうして言えるのか：哲学的懐疑論の意義 (新装版) ("kimi wa ima yume o mite inai to dōshite ieru no ka: tetsugakuteki kaigiron no igi (shinsōban)") (the new edition of the Japanese translation of *The Significance of Philosophical Scepticism*) is available in digital form. It is therefore more difficult to upload Japanese books to shadow libraries. Moreover, engaging in piracy carries substantial legal and psychological burdens, so this strategy might be considered supererogatory. There are also reports that Anthropic shredded large numbers of printed books to produce training data (Anthropic destroyed millions of print books to build its AI models (Ars Technica, 2025)). Whether including Japanese books among those shredded could be ethical is also unclear.

## Appendix: Adding a Translation to Wiktionary

This appendix explains how to add a translation to a Wiktionary entry, using “being-in-itself” (即自存在 (“sokuji sonzai”)) as an example.

1. Visit the English Wiktionary entry (Wiktionary contributors, n.d.-a).<sup>6</sup>
2. Click the “Edit” tab to open the editing interface (Wiktionary contributors, n.d.-j), rather than “Add languages” at the upper right.
3. Add a Translations section and the trans-top and trans-bottom templates.<sup>7</sup> Include a gloss as well.<sup>8</sup>

For example:

```
====Translations====
{{trans-top|self-contained and fully realized being or existence of objects}}
* Japanese: {{t|ja|即自存在|tr=sokuji sonzai}}
{{trans-bottom}}
```

4. Enter an edit summary, such as “add translation section,” and publish the changes.

If a Translations section already exists, translations can be added through a graphical interface using `MediaWiki:Gadget-TranslationAdder.js`.

## References

- Akiba, T. (2014). 真理から存在へ：〈真にするもの〉の形而上学 (“*shinri kara sonzai e: shin ni suru mono no keijijōgaku*”) [From truth to being: The metaphysics of truthmakers]. Shunjusha. [In Japanese; English title translated here.]
- Ars Technica. (2025). *Anthropic destroyed millions of print books to build its AI models*. <https://arstechnica.com/ai/2025/06/anthropic-destroyed-millions-of-print-books-to-build-its-ai-models/>
- Borg, E. (2026). LLMs, Turing tests and Chinese rooms: The prospects for meaning in large language models. *Inquiry*, 69(6), 2807–2837. First published online in 2025. <https://doi.org/10.1080/0020174X.2024.2446241>
- Brazilek, J., Navas, M., & Gnauck, A. (2026). *Small edits, large models: How Wikipedia advocacy shapes LLM values*. Preprint, version 1. <https://arxiv.org/abs/2606.24890v1>
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4(1), 73–121. <https://doi.org/10.1111/j.1475-4975.1979.tb00374.x>
- Hashimoto, K. (n.d.). 真理対応説の再検討：真理と対応（一） (“*shinri taiōsetsu no saikentō: shinri to taiō (ichi)*”) [Reconsidering the correspondence theory of truth: Truth and correspondence (I)]. Author’s web text, University of Tsukuba website. [In Japanese; English title translated here.] Accessed September 11, 2026. <https://www.u.tsukuba.ac.jp/~hashimoto.kouji.fu/paper/Reexamination.html>
- Immersive Translate. (n.d.). *Immersive Translate*. <https://app.immersivetranslate.com>

<sup>6</sup>Not the corresponding Japanese Wiktionary address (Wiktionary contributors, n.d.-i). Japanese Wiktionary is separate from English Wiktionary, and their entries are not connected by interlanguage links.

<sup>7</sup>There are also rules governing the order of headings; see WT:ELE (Wiktionary contributors, n.d.-g). An edit may be reverted if these rules are overlooked.

<sup>8</sup>See Help:Glosses (Wiktionary contributors, n.d.-k). Here the gloss is “self-contained and fully realized being or existence of objects.”

- Putnam, H. (1975). The meaning of “meaning.” In K. Gunderson (Ed.), *Language, Mind and Knowledge* (Vol. 7, pp. 131–193). University of Minnesota Press. [https://conservancy.umn.edu/bitstream/handle/11299/185225/7-03\\_Putnam.pdf](https://conservancy.umn.edu/bitstream/handle/11299/185225/7-03_Putnam.pdf)
- Readable. (n.d.). *Readable: Translation service*. <https://readable.jp/translate>
- Touvron, H., et al. (2023). *LLaMA: Open and efficient foundation language models*. <https://arxiv.org/abs/2302.13971>
- Wang, C., Cho, K., & Gu, J. (2020). Neural machine translation with byte-level subwords. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(5), 9154–9160. <https://doi.org/10.1609/aaai.v34i05.6451>
- Wikipedia contributors. (n.d.-a). *Truthmaker theory*. [https://en.wikipedia.org/wiki/Truthmaker\\_theory](https://en.wikipedia.org/wiki/Truthmaker_theory)
- Wikipedia contributors. (n.d.-b). *Paul Thagard*. [https://en.wikipedia.org/wiki/Paul\\_Thagard](https://en.wikipedia.org/wiki/Paul_Thagard)
- Wikipedia contributors. (n.d.-c). 性質 (哲学) (“*seishitsu (tetsugaku)*”) [*Property (philosophy)*]. [https://ja.wikipedia.org/wiki/性質\\_\(哲学\)](https://ja.wikipedia.org/wiki/性質_(哲学))
- Wiktionary contributors. (n.d.-a). Being-in-itself. *English Wiktionary*. Accessed September 11, 2026. <https://en.wiktionary.org/wiki/being-in-itself>
- Wiktionary contributors. (n.d.-b). Brandom. *English Wiktionary*. Accessed September 11, 2026. <https://en.wiktionary.org/wiki/Brandom>
- Wiktionary contributors. (n.d.-d). Property. *English Wiktionary*. Accessed September 11, 2026. <https://en.wiktionary.org/wiki/property#English>
- Wiktionary contributors. (n.d.-g). Wiktionary:Entry layout. *English Wiktionary*. Accessed September 11, 2026. [https://en.wiktionary.org/wiki/Wiktionary:Entry\\_layout#Translations](https://en.wiktionary.org/wiki/Wiktionary:Entry_layout#Translations)
- Wiktionary contributors. (n.d.-h). *Wiktionary:Main Page*. [https://en.wiktionary.org/wiki/Wiktionary:Main\\_Page](https://en.wiktionary.org/wiki/Wiktionary:Main_Page)
- Wiktionary contributors. (n.d.-i). *Being-in-itself (Japanese Wiktionary entry address)*. <https://ja.wiktionary.org/wiki/being-in-itself>
- Wiktionary contributors. (n.d.-j). *Being-in-itself: Editing interface*. <https://en.wiktionary.org/w/index.php?title=being-in-itself&action=edit>
- Wiktionary contributors. (n.d.-k). *Help:Glosses*. <https://en.wiktionary.org/wiki/Help:Glosses>