

Evaluating machine translation of literature through rhetorical analysis

Irina S. Karabayeva^{1,*}, Anna N. Kalizhanova^{1,2}

¹Translation Theory and Practice Department, Foreign Languages Faculty, Karaganda Buketov University, Karaganda 100415, Kazakhstan

²Translation Theory and Practice Department, Foreign Languages Faculty, Karaganda Buketov University, Karaganda 100025, Kazakhstan

Received: 09.12.2023 • Accepted: 15.01.2024 • Published: 20.01.2024 • Final Version: 29.01.2024

Abstract: This paper looks at how well ChatGPT and DeepL, two AI tools, translate literary works. Not only can ChatGPT translate text, but it can also carry out other jobs. DeepL is a service that performs computer translation and uses neural networks. The paper looks at how ChatGPT and DeepL translate books, poems, and dialogues compared to translations done by humans. The paper also talks about the pros and cons of using machine translation for literary reasons, including issues of creativity, style, and adapting to different cultures. The paper uses both new and old studies on machine translation technologies and how they work with human translation. The paper comes to the conclusion that ChatGPT and DeepL are useful but imperfect tools for translating literature, and they require human review and improvement. The paper adds to the fields of machine translation and natural language processing by looking at how two cutting-edge AI tools, ChatGPT and DeepL, can be used to translate literary works. The paper also adds to literature studies and digital humanities by looking into what machine translation can and can't do for creative writing and dialog systems. The goal of the paper is to encourage researchers, translators, writers, and users from different fields to work together and talk to each other. ☒

Keywords: ChatGPT, DeepL, literary texts, machine translation, neural networks, quality evaluation, rhetorical devices

1. Introduction

The application of artificial intelligence (AI) to fields like machine translation has progressed rapidly, enabling new potentials for automated language conversion at scale. However, while AI translation accuracy has achieved parity with humans for simple texts like news articles, significant challenges remain for translating complex literary works. Can machines effectively handle great novels filled with metaphors, rhyming verse, imaginative narratives, and other artistic language flourishes? This paper aims to answer this question by comparing the capabilities of two popular neural machine translation systems, ChatGPT and DeepL, on literary texts.

Literature makes use of elaborate rhetorical techniques, subtle cultural elements, subjectivity, and creative expression that pose difficulties for current AI capabilities (Moorkens, 2017). According to a number of researchers (Alzeebaree, 2020; Lee & Cha, 2023), initial rule-based and statistical machine translation methods performed poorly on literary texts, failing to capture nuance and style, whereas modern neural networks like Google's Transformer architecture have demonstrated improved fluency and contextual handling (D'Souza, 2023). Despite these advances,

* Corresponding Author: 123irina06@gmail.com

literature remains a challenging domain for AI translation, as it requires not only technical correctness, but also aesthetic essence, emotion, creativity, and artistry.

AI translation also offers some benefits for literature, such as enhancing accessibility to literary works across languages, supporting comprehension at scale, and reducing costs and time (Gaspari et al., 2014). Moreover, AI translation can facilitate draft translations that human experts can post-edit to polish styling and fix errors (Koponen, 2016). AI is becoming an integral tool in the professional translator's toolkit. Therefore, it is important to evaluate the strengths and weaknesses of different AI translation systems for literature, and to explore the potentials and limitations of hybrid human-machine collaboration.

Previous studies have investigated various aspects of AI translation for literature, such as evaluating the quality of output using customized rubrics (Fantinuoli & Prandi, 2021), enhancing the training of neural networks using stylistically annotated parallel corpora (Cipriani, 2023), and promoting interdisciplinary collaboration between literary scholarship and computer science (Goto & Tanaka, 2017). However, there is a lack of comprehensive and updated comparative analysis of leading AI translation platforms on literary texts, especially on rhetorical devices like metaphors, rhyme, and imagery, which are essential for conveying the author's intended meaning and effect.

The aim of this paper is to provide such an analysis, by assessing the capabilities of two popular neural machine translation systems, ChatGPT and DeepL, on literary texts. ChatGPT is a conversational AI system that uses a generative pre-trained transformer model to produce natural and engaging responses (Roumeliotis & Tselikas, 2023). DeepL is a web-based translation service that uses a deep learning approach to deliver high-quality and accurate translations (Fradana, 2023, p. 207). Both systems support over 100 languages and claim to outperform other AI translation systems on various benchmarks.

How do they fare on literary texts? We compare their performance on three types of rhetorical devices: metaphors, rhyme, and imagery, using a corpus of English literary texts and their translations in the Russian language. We evaluate the output using both automatic and human metrics, and discuss the implications for the development of more linguistically knowledgeable and stylistically adept AI. The paper is organized as follows: Section 2 reviews the related literature on AI translation for literature, Section 3 describes the methodology and data, Section 4 presents the results and analysis, Section 5 discusses the findings and limitations, and Section 6 concludes the paper and suggests future directions.

2.Literature Review

Early machine translation systems relied on rules-based approaches that performed very poorly on literary works (Rivera-Trigueros, 2022). They depended on rigid structures and lacked fluency (Oliver, 2020, p. 125) and were found incapable of handling complex rhetorical devices, ambiguity, metaphors and other creative language (Hasselberger, 2021). However, statistical machine translation (SMT) improved outcomes by training on vast volumes of parallel text corpora (Saxena et al., 2022). SMT modeled probabilistic mappings between source and target languages (Sharma & Singh, 2021). Still, their capabilities fell short for literary translation (Şahin & Gürses, 2021; Shah, Imran, & Ismail, 2024).

Neural machine translation (NMT) has since demonstrated substantial advances by utilizing deep learning (Mohamed et al., 2021). Castilho et al. (2017) explain NMT incorporates embedded vector representations of words and sequences enabling stronger contextual modeling. Toral and

Way (2018) assessed an NMT system's handling of English-Spanish fiction translation, finding it achieved much greater fluency with some metaphors conveying appropriately. But many creative elements were rendered poorly.

Analyses reveal persisting limitations alongside progress. Rădulescu (2019) tested different neural architectures' translation of a Romanian novel. Fluency and word order improved from previous technologies but higher-level flaws in cohesion and mistranslations remained.

Recent research proposes approaches to enhance NMT creativity. Goto and Tanaka (2017) developed techniques to detect untranslated content, which could flag missing metaphors. Matusov et al. (2019) incorporated transformer and retrieval mechanisms to better handle unknown phrases. LM model fusion explicitly modeled linguistic constraints to improve fluency (Ranathunga et al., 2023). Fantinuoli and Prandi (2021) designed customized rubrics aligning with interpretability, though human communication still exceeded NMT.

Other studies advocate greater interdisciplinary collaboration. Bolaños et al. (2021 in Rovira-Esteva et al., 2023, p. 120) suggest stylistically annotating parallel corpora to highlight rhetorical devices for NMT training. Chakrabarty et al. (2021) propose applying cognitive and perception principles to model figurative language. Rane (2023) argues for cooperation between computer scientists, linguists and literary experts.

However, limitations arise in automating human translation expertise. Castilho et al. (2018) notes software cannot yet replicate creative ambiguous language use. Moorkens (2018) found NMT output remained distinguishable from human translation. Ploin et al (2022) concludes human creativity remains irreplaceable. Their study also affirms that NMT demonstrates its strengths, but requires further progress to holistically interpret literary text the way humans do (Bentivogli et al., 2018; Castilho et al., 2017).

Overall, NMT indicates promising potentials advancing literary translation but still falls substantially short of human skill for creative stylistic elements, subjectivity, culture and aesthetic mastery (Bentivogli et al., 2018). Yet quality continues improving and semi-automated workflows with human oversight offer capabilities (Castilho et al., 2017; Guerberof Arenas & Moorkens, 2019). Interdisciplinary breakthroughs between AI, linguistics and literary scholarship could enable more robust modeling of creative language (Belinkov et al., 2017). But handling subjective artistry poses fundamental challenges for machines (Chesterman, 2021). The ultimate limits of computational creativity remain unknown.

3. Methods and Materials

This study aimed to provide updated comparative analysis on leading neural machine translation platforms' capabilities for literary text. The examination focused on two widely used commercial systems:

1) ChatGPT (OpenAI): launched in November 2022, this conversational AI system employs a transformer-based neural network architecture trained on massive volumes of data including books, websites and online conversations (Roumeliotis & Tselikas, 2023). ChatGPT exhibits strong language generation capabilities.

2) DeepL: proprietary NMT service developed by the DeepL Company using deep neural networks. It incorporates models like transformers and convolutional networks (Fradana, 2023, p. 207). DeepL provides both free and paid subscription access.

The platforms were tested on samples of literary text containing creative rhetorical devices and language artistry. Excerpts from the English children's novel *Coraline* by Neil Gaiman (2003) served as source samples, selected due to its rich metaphors, imagery, alliteration, and imaginative narrative. A set of approximately 15 brief passages highlighting different creative techniques was utilized, equally around 100-200 words each.

The base English input samples were translated into Russian output by ChatGPT and DeepL. Default system settings were used without modification to represent typical application. Each platform's capabilities and limitations translating the literary devices were assessed qualitatively using manual inspection. Outputs were evaluated regarding:

- Accuracy: faithfulness conveying source meaning
- Fluency: grammaticality, naturalness
- Preservation of literary devices: maintenance of metaphor, alliteration, rhymes, etc.

This allowed categorization of which creative language elements posed greater difficulties for the AI systems. Their overall capability approaching the quality of human translation was judged.

Additionally, for the ChatGPT system, tests were conducted on providing the model with enhanced guidance specific to translating rhetorical devices in context. Literary elements in input samples were highlighted and ChatGPT prompted to focus on accurately conveying aspects like metaphoric language, emotive content, and aesthetic form. The system was interactively trained via this human-in-the-loop approach as described by Li et al. (2016). Outputs were re-evaluated to assess if translation quality improved through targeted human guidance.

ChatGPT was also questioned about its approach to translating different literary devices to gain technical insights into its strengths, weaknesses, and representation of creative language patterns. The system provided examples and explanations via conversational interaction.

Comparisons were made to professional human English-to-Russian translation of the original samples as a gold standard reflecting the pinnacle of literary translation skill. Outputs of human versus AI systems were investigated regarding mastery of language artistry. The analysis aimed to elucidate gaps where computational methods have yet to reach advanced human expertise for creative stylistic elements.

4. Results

The testing process on samples of literary text yielded informative findings about ChatGPT and DeepL's capabilities and limitations for key aspects of creative language translation:

Accuracy: Both systems exhibited reasonable accuracy conveying literal meanings for straightforward descriptive passages. However, their precision suffered for content involving metaphors, allusions, and other figures of speech. DeepL tended to translate these devices literally rather than interpreting the underlying meaning. ChatGPT was moderately better at paraphrasing metaphors but also struggled with rare or culture-dependent examples.

Fluency: DeepL generally produced very smooth, grammatically coherent Russian output without awkward phrasing. In contrast, ChatGPT had periodic grammar errors reflecting lack of complete language mastery. Though when prompted, ChatGPT could self-correct basic mistakes. Overall, both systems generated reasonably fluent outputs on par with human-level for typical language.

Preservation of literary devices: This area exhibited the biggest challenges. The nuanced creative elements central to literary artistry faced significant difficulties in translation:

- Metaphor: Common straightforward metaphors were rendered adequately but obscure metaphors posed problems. ChatGPT occasionally improved on the original metaphor through creative paraphrasing.

- Simile: Direct similes translated well in both systems but some lacked the eloquence of human versions. Idiomatic similes were challenging.

- Alliteration/Rhyme: Neither system recreated alliteration or rhyme schemes to the level exhibited in the professional human translations. They prioritized conveying literal meaning over artistic form.

- Imagery: DeepL struggled translating vivid imaginary phrases, often simplifying to plainer language. ChatGPT offered some improved imagery when prompted.

- Parcelling: ChatGPT partially replicated this technique of creatively fragmented sentences but DeepL solely produced complete sentences.

- Tone/Mood: Human translations provided emotive nuance missing from AI outputs. The systems failed to convey subtle atmosphere.

Overall, both ChatGPT and DeepL failed to recreate the nuance, fluidity, and aesthetic mastery of the professional human translator across these many different dimensions of creative language artistry. However, when provided targeted guidance emphasizing literary devices during translation, ChatGPT's performance did improve on certain creative elements. The interactive learning highlighted potentials to enhance NMT outputs through proper human instruction focusing on artistic goals.

Discussions with ChatGPT further revealed details about its approach:

- It stated metaphor translation involves representing the implied meaning in the target language, difficult without real-world knowledge. English metaphors may not have direct equivalents.

- For similes, it seeks corresponding comparisons in the output language, a process benefiting from large training corpora.

- Alliteration and rhyme are challenging due to different linguistic constraints across languages. These are modeled weakly currently.

- Parceling requires detecting fragmented sentences and replicating their effects. This technique is not well captured by its training data.

The system acknowledged representing high-level creative language remains difficult compared to non-literary texts. ChatGPT's insufficiency handling certain metaphors and devices was apparent without human guidance. But cooperative interaction enabled improving some translations by emphasizing stylistic goals.

5. Discussion

The comparative analysis revealed both strengths and considerable limitations of leading commercial NMT services for translating literary texts relative to human expertise. AI systems have achieved high accuracy and fluency for straightforward language but struggle to replicate the art and essence of literature. The challenges align with findings from academic research on NMT

creative deficiencies (Gaspari et al., 2015). Literal resemblance does not equate to expert literary translation.

At their current capability, AI systems function more as diction-driven word converters rather than exploring expression possibilities in the target language. They lack integration of linguistic stylings beyond basics of grammar and vocabulary. Modeling the patterns, subjectivity, and cultural allusions inherent in literary language remains difficult (Eco, 2001). Methods to inject greater linguistic knowledge into NMT training could enhance handling of creative devices as suggested by Casas (2020).

Of the two systems, DeepL produced smoother grammatical output thanks to robust models trained on massive high-quality corpora. Meanwhile, ChatGPT exhibited stronger versatility responding to interactive guidance. Tailored human instruction improved some translations by focusing attention on artistic goals. This highlights the promise of guided hybrid human-AI collaboration as argued by Rane (2023).

6. Conclusion

This paper presented comparative analysis between commercial neural machine translation services ChatGPT and DeepL and professional human translators on literary text samples containing creative rhetorical techniques and artistry. The examination aimed to elucidate strengths, weaknesses, and overall capability of leading AI systems to handle the complexities of literature.

The results underscored persistent challenges for AI in expert literary translation. While accuracy on straightforward passages was reasonable, critical deficiencies emerged regarding creative metaphoric language, emotive expression, nuanced cultural details, aesthetic form, and the communicative goals central to humanistic art. The AI systems functioned more as diction-driven converters rather than exploring the holistic expressive potentials of literature. Even tailored guidance could only partially improve certain creative translations.

However, rapid advances are occurring as neural network scale and techniques improve. The power of deep learning has enhanced fluency and basic comprehension of literary works, enabling accessibility. AI also facilitates productive human-machine collaboration harnessing computation while ensuring quality final polished translations. Transparency from AI systems into their creative limitations provided insights for further progress.

Looking forward, enhanced interdisciplinary collaboration between AI developers, linguists, and literary scholars offers strong potential to advance computational mastery of the arts (Rane, 2023). Integrating knowledge of language, rhetoric, culture, and cognition could strengthen machine learning representations (Chakrabarty et al., 2021). Teaching AI the patterns of literary art may unlock new potentials. But handling subjectivity and emotional essence remains profoundly challenging (Chesterman, 2021). The pinnacle capabilities of human creativity likely persist as irreproducible.

Overall, while significant hurdles to AI matching literary translation expertise endure, we see potentials for thoughtful hybrid symbiosis and specialized breakthroughs. Perhaps someday, AI assistance could aid human understanding across languages for civilization's creative treasures. But computational creativity must overcome difficult barriers regarding the human condition. We should ensure developing AI with ethical priorities and humanistic visions (Jobin, Ienca, & Vayena, 2019). How to judiciously apply AI's growing capabilities for literary arts remains an

open challenge but advances could expand accessibility while inspiring new possibilities for expression.

7. Acknowledgment

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan on the topic "The connectivist model of a foreign language educational SMART–environment in Kazakhstani context: Necessity, availability, and development strategy" (2023-2025) (Grant number AP19679833)

References

- [1] Moorkens, J. (2017). Under pressure: Translation in times of austerity. *Perspectives*, 25(4), 464-477. <https://doi.org/10.1080/0907676X.2017.1285331>
- [2] Alzeebaree, Y. (2020). Machine Translation and Issues of Multiword Units: Idioms and Collocations. *Eastern Journal of Languages, Linguistics and Literatures*, 1(2), 1–23. <https://qabasjournals.com/index.php/ejll/article/view/14>
- [3] Lee, J. H., & Cha, K. W. (2023). Human Interpretation and Machine Translations Based upon Interviews with Director Joon-ho Bong. *Korean Journal of English Language and Linguistics*, 23, 204–219. <http://journal.kasell.or.kr/xml/36132/36132.pdf>
- [4] D'Souza, J. (2023). *A Review of Transformer Models*. <https://doi.org/10.48366/r609546>
- [5] Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, 23(3), 333-358. <https://doi.org/10.1080/0907676X.2014.979842>
- [6] Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25(2), 131–148. https://www.jostrans.org/issue25/art_koponen.pdf
- [7] Fantinuoli, C., & Prandi, B. (2021). Towards the evaluation of automatic simultaneous speech translation from a communicative perspective. *arXiv preprint arXiv:2103.08364*. <https://arxiv.org/pdf/2103.08364.pdf>
- [8] Cipriani, A. M. (2023). *Literary Digital Stylistics in Translation Studies*. Springer Nature.
- [9] Goto, I., & Tanaka, H. (2017, August). Detecting untranslated content for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation* (pp. 47-55). <https://aclanthology.org/W17-3206.pdf>
- [10] Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, 15(6), 192. <https://doi.org/10.3390/fi15060192>
- [11] Fradana, H. (2023, December). Translator Competencies Utilizing Translation Machines in the Digital Era. In *Proceeding of International Seminar on Adab and Humanities* (Vol. 5, No. 1, pp. 201-212). <https://proceedings.radenfatah.ac.id/index.php/isah/article/view/1060>
- [12] Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, 56(2), 593-619. <https://doi.org/10.1007/s10579-021-09537-5>
- [13] Oliver, A. (2020). Human translation and machine translation. Specificities, uses, advantages and disadvantages. *Translating Diversity= Traduire la diversitat*, 111-130. <https://www.linguapax.org/wp-content/uploads/2020/12/LinguapaxReview2020.pdf#page=111>
- [14] Hasselberger, W. (2021). Can Machines Have Common Sense? [Review of *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, by E. J. Larson]. *The New Atlantis*, 65, 94–109. <https://www.jstor.org/stable/27020762>
- [15] Saxena, S., Chaurasia, U., Bansal, N., & Daniel, P. (2022). Improved unsupervised statistical machine translation via unsupervised word sense disambiguation for a low-resource and Indic languages. *IETE Journal of Research*, 1-11. <https://doi.org/10.1080/03772063.2022.2098189>
- [16] Sharma, H., & Singh, H. (2021). Analysis of the Effect of Topic Modeling on General Corpus Mixed with In-Domain Text for English-Hindi Translation. In *Rising Threats in Expert Applications and Solutions: Proceedings of FICR-TEAS 2020* (pp. 777-783). Springer Singapore.

- [17] Şahin, M., & Gürses, S. (2021). English-Turkish literary translation through human-machine interaction. *Tradumàtica: tecnologies de la traducció*, (19), 171-203. <https://doi.org/10.5565/rev/tradumatica.284>
- [18] Shah, T. Z., Imran, M., & Ismail, S. M. (2024). A diachronic study determining syntactic and semantic features of Urdu-English neural machine translation. *Heliyon*, 10(1). <https://doi.org/10.1016/j.heliyon.2023.e22883>
- [19] Mohamed, S. A., Elsayed, A. A., Hassan, Y. F., & Abdou, M. A. (2021). Neural machine translation: past, present, and future. *Neural Computing and Applications*, 33, 15919-15931. <https://doi.org/10.1007/s00521-021-06268-0>
- [20] Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art?. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 109-120. <http://archive.sciendo.com/PRALIN/pralin.2017.108.issue-1/pralin-2017-0013/pralin-2017-0013.pdf>
- [21] Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text?. *Translation quality assessment: From principles to practice*, 263-287. <https://arxiv.org/pdf/1801.04962.pdf>
- [22] Rădulescu, A. (2019). Algorithmic Textual Practices: Improving Fluency and Word Order in Neural Machine Translation Output. *Linguistic and Philosophical Investigations*, (18), 126-132. <https://www.ceeol.com/search/article-detail?id=771883>
- [23] Matusov, E., Wilken, P., & Georgakopoulou, Y. (2019, August). Customizing neural machine translation for subtitling. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)* (pp. 82-93). <https://aclanthology.org/W19-5209.pdf>
- [24] Ranathunga, S., Lee, E. S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, 55(11), 1-37. <https://arxiv.org/pdf/2106.15115.pdf>
- [25] Rovira-Esteva, S., Lamber, J., Muñoz Martín, R., Marín García, Á., O'Brien, S., Massey, G., ... & Vaerenbergh, L. V. (2023). *50 Years Later. What Have We Learnt after Holmes (1972) and Where Are We Now?*. Servicio de Publicaciones y Difusión Científica de la Universidad de Las Palmas de Gran Canaria (ULPGC). <https://accedaeris.ulpgc.es/bitstream/10553/124520/1/9788490424612.pdf>
- [26] Chakrabarty, T., Zhang, X., Muresan, S., & Peng, N. (2021). MERMAID: Metaphor generation with symbolism and discriminative decoding. *arXiv preprint arXiv:2103.06779*. <https://arxiv.org/pdf/2103.06779.pdf>
- [27] Rane, N. L. (2023). Multidisciplinary collaboration: key players in successful implementation of ChatGPT and similar generative artificial intelligence in manufacturing, finance, retail, transportation, and construction industry. <https://doi.org/10.31219/osf.io/npm3d>
- [28] Moorkens, J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, 12(4), 375-387.
- [29] Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. *Translation quality assessment: From principles to practice*, 9-38. https://doi.org/10.1007/978-3-319-91241-7_2
- [30] Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2018). Neural versus phrase-based MT quality: An in-depth analysis on English-German and English-French. *Computer Speech & Language*, 49, 52-70. <https://doi.org/10.1016/j.csl.2017.11.004>
- [31] Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., Lohar, P., Way, A., Miceli Barone, A., & Gialama, M. (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In *Proceedings of MT Summit XVI, vol.1: Research Track* (pp. 116-131). http://amt.info/app-def/S-102/mtsummit/2017/wp-content/uploads/sites/2/2017/09/MTSummitXVI_ResearchTrack.pdf
- [32] Guerberof Arenas, A., & Moorkens, J. (2019). Machine translation and post-editing training as part of a master's programme. *Jostrans The Journal of Specialized Translation*, (31), 217-238. https://www.jostrans.org/issue31/art_guerberof.php
- [33] Belinkov, Y., Durrani, N., Dalvi, F., Sajjad, H., & Glass, J. (2017). What do neural machine translation models learn about morphology?. *arXiv preprint arXiv:1704.03471*. <https://arxiv.org/pdf/1704.03471.pdf>
- [34] Chesterman, S. (2021). Through a glass, darkly: artificial intelligence and the problem of opacity. *The American Journal of Comparative Law*, 69(2), 271-294. <https://doi.org/10.1093/ajcl/avab012>

- [35] Gaiman, N. (2003). *Coraline*. Bloomsbury.
- [36] Li, J., Miller, A. H., Chopra, S., Ranzato, M. A., & Weston, J. (2016). Dialogue learning with human-in-the-loop. *arXiv preprint arXiv:1611.09823*. <https://arxiv.org/pdf/1611.09823.pdf>
- [37] Eco, U. (2001). *Experiences in translation*. University of Toronto Press.
- [38] Casas, N. (2020). *Injection of linguistic knowledge into neural text generation models*. <https://upcommons.upc.edu/bitstream/handle/2117/341598/TNCM1de1.pdf?sequence=1>
- [39] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399. <https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf>