Impacts of Artificial Intelligence on Language Learning

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Mitchell Taylor

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

Introduction

On November 30, 2022, OpenAI released one of the most polarizing new technologies in the last decade, ChatGPT. As an Artificial Intelligence (AI), ChatGPT is a large language model designed to achieve language generation and understanding. This new AI with language capabilities has had a profound impact on multiple fields ranging from finance to computer science. Due to these wide-reaching impacts, the public response to AI has varied drastically. On one side, there are many who consider AI as a serious threat to job markets and economic security. On the other hand, there are others who hail AI as the new Industrial Revolution with the capability to completely alter the structure of working society. In this sense, the growth of AI is like the introduction of the steam engine in the late 18th century. AI does not require compensation to perform a professional task like a human being would, and oftentimes it could perform the task more efficiently. The major difference lies in the nature of the task, with AI taking over processing tasks so complex they have been limited to human beings. For instance, the field of computer science mentioned earlier has been seriously altered by the introduction of large language model AI with programming capabilities. As AI has gained the ability to write programs that were once dependent on the efforts of computer scientists, computer science is forced to reconcile with the ever-growing effects of AI. With an understanding of these effects of AI, especially a language oriented one such as ChatGPT, one potential use consideration for AI is translation. Commonplace machine translators often fail to fully convey the significance of a statement through their reliance on formulaic translation that lacks the fluidity of human speech and the capacity to adjust to varying linguistic differences. The cultural variations and colloquialisms that frequently appear in human language cannot be captured using a structured

translator that provides word-by-word translation. These are issues which can and have been solved by AI, bringing forth a revolutionary tool for communication with near human capabilities. This research paper will seek to explore the impacts that newly developed AI has had on language learning through the increased availability of translation and the restructuring of institutions around AI.

Methods

To analyze the data on this topic, several different methods will be used. Primarily, data collected from academic papers centered on the topic will provide the basis for research. Using this scientific evidence, a detailed analysis of the history of translation technology as well as its current state will be presented to create a foundational understanding for the reader. Following this contextual introduction, the research will focus on the differences between common machine translators and AI translation to create a platform for continued discussion. Furthermore, the effects AI can have on language acquisition, mainly the speed and quality of acquisition in comparison with typical methods, will be investigated following the research question posed in the introduction. With this established through academic research, then history and case studies on the use of Artificial Intelligence in education will be utilized for the analysis of the impacts that it has on the learning process. This method will provide context for the overall situation, while case studies will be critical for my methodology, allowing me to present clear examples of AI in language learning and strengthen the validity of discussion. Throughout this process of investigation, the thesis will examine the results through the framework of Social Construction of Technology (SCOT). As described in *The Social Construction of Technological Systems* by

Bijker, Hughes, and Pinch, SCOT is a theory which states that the development and interactions of technology do not determine human society, but that human behavior is a driving force in the shaping of technology. According to the authors who first defined SCOT, "In SCOT… technological artifacts are culturally constructed and interpreted; in other words, the interpretative flexibility of a technological artifact must be shown. By this we mean not only that there is flexibility in how people think of or interpret artifacts but also that there is flexibility in how artifacts are designed" (Bijker, et. al, 1987, 34). To fully understand how a technology functions and develops, it is critical to understand how it is applied in a social context. In the case of AI for language learning, it is applied almost entirely for social purposes. Access to simple, accurate translation and an educational tool provides not only the capability to comprehend content in another language, but also generates opportunities for communication between previously disparate groups. This degree of social influence should have a clear effect on the development of AI for language acquisition purposes which will be observed in the results and analysis sections. Using this framework, the manner that AI interacts with the public and how these interactions can further shape AI into a better language tool will be analyzed.

Machine Translation

To begin, an understanding of current machine translation must be established. The most accurate form of machine translation currently available without a form of AI is statistical machine translation. According to The Routledge Encyclopedia of Translation Technology published in 2014, "Stages of translation in statistical machine translation (SMT) systems are: first, alignment of bilingual corpora (i.e. texts in original language and texts in target language,

or texts in comparable corpora which are not directly alignable), either by word or phrase; then, frequency matching of input words against words in the corpus, extraction of most probable equivalents in the target language ('decoding'); reordering of the output according to most common word sequences using a 'language model', a monolingual corpus providing word frequencies of the TL; and finally production of the output in the target language." (Chan, 2014, 128). In essence, the machine translator creates a probabilistic model of the word frequency in each language and matches the two together to create the translated output. As could be expected, this form of translation struggles immensely with statistically ambiguous concepts such as grammatical gender. Additionally, the reduction of language to a simple model completely neglects the underlying contextual sentiments of a statement. This issue is referred to as the commonsense knowledge problem, where a machine is required to make an evaluation based on evidence that would be common sense to the average human being, but completely unknown to the machine.

To further the point of machine translation inaccuracy, a study done in 2009 by the University of Mississippi compared the four most used online translation software at the time (Google Translate, Yahoo SYSTRAN, AppliedLanguage, and x10) with the goal of determining the software with the greatest accuracy. A series of statements and paragraphs were translated between English, German, and Spanish using each of the software and the results were compared by native speakers of the languages. Google Translate was determined to be the best form of translation, but according to the researchers, "For example, determining the gender of the subject was difficult in some cases. Not recognizing that "Daisy" is a female name, both languages had problems. The word "sie" can be "she" or "they" in German, and one German word ("unbändigem" - unrestrained) was not translated" (Aiken et al, 2009). As explained by Aiken,

the Google translations were routinely inaccurate, often swapping gender and misinterpreting phrases in the translation process.

Having established an understanding of machine translation, the impact of machine translation on language learning can be analyzed to highlight the new effects created by AI for machine translation. In a study written by researcher Yanxia Yang in early 2024 involving 500 Chinese EFL learners, machine translation plays a positive role in language learning through increased convenience of multilingual resources without promoting reliance on the software. A key finding from this study is the importance of "machine translation fit", the notion that machine translators are best applied in certain circumstances. The study determined that "This finding revealed that machine translation fit functioned as a critical determinant of individuals' performance in tasks related to machine translation, which was in line with the prior relevant studies and further supported the notion that technology fit played a crucial role in language learning" (Yang, 2024, 11). This result supports the use of machine translation for language learning through the application of the technology's strengths and weaknesses. Applying machine translation for relevant tasks creates a positive guide for future translations while poor machine translations are examples of mistakes to avoid when translating.

As AI evolves, its use application for translation purposes continues to grow as well. AI translation differs from other forms of machine translation using neural networks to implement training and iterative algorithms which permit AI models to create better results with each iteration. As written by Mondal et al. in a 2023 study of machine translation, Neural Machine Translation (NMT) has been proven to outperform SMT in almost every case. When comparing translations between multiple languages for different types of machine translators, the NMT consistently provided the most accurate translation apart from decreased accuracy seen for

sentences longer than 35 words. This weakness in NMT has been addressed with newly developed neural network structures. The transformer model, the most advanced neural network used for large language models like ChatGPT-4, processes all input words in a sentence simultaneously and assigns them weights based on relevance to other words (Mondal et al., 2023). Further proof that NMT consistently outperforms SMT can be found in a 2019 study during which SMT and NMT results were compared and edited by official translators for the Swiss Post's Language Service. The study found that between the two forms of machine translation, the SMT was much more likely to require edits, either substitutions or deletions. The authors go on to say that "NMT edits are more often considered to be non-significant from a post-editing point of view (14.68%), as opposed to SMT edits (68.53%); (iv) translators have more difficulties stating whether a sentence has to be modified with NMT than with SMT" (Mutal et al., 2019, 80). These results are consistent with the claims made by Mondal and other sources on machine translation, that NMT provides more rigorous translation when compared to SMT. While neither technology is perfect, the NMT reduced the quantity of significant human edits by nearly a factor of five. This significant improvement clearly demonstrates the superiority of AI in the field of machine translation.

AI Performance

For learning applications, AI is a tool that is slowly being integrated into the educational process due to its sudden growth and concerns of human replacement. In "AI: A Review of its Uses in Language Teaching and Learning", the author Zuraina Ali highlights several advantages of AI for language education as of 2020. According to Ali, the major advantage of AI is the

ability to recognize patterns in natural human speech and respond accordingly. Based on a psycholinguistic study from 2018, this theme focuses on the ability of AI to observe common mistakes made by learners under stress. Following the experimental results, it was determined that "individuals who tend to be anxious and emotional in a stressful situation may become less "talkative" during isolation (quantitative change). They may also be distinguished by making similar grammar mistakes in English, regardless their mother tongue (qualitative change)" (Anikushina et al., 2018, 27). The paper goes on to mention the applications of this finding, that AI could be used to identify linguistic errors made under stress and the quality of human-machine interaction would become quicker and easier. It is possible that AI could be "tailored" to fit the individual, using the recognition of stressful responses to alter the style of interaction to reduce stress and increase performance. In the same vein of AI-to-individual ease of interaction, AI has been found to ameliorate language skills in lower-level learners. One study performed in 2022 comparing interactions between ESL students and AI or native speakers found that "These results revealed that the two different types of interlocutors positively influenced low-level participants to improve their speaking skills, but low-level learners would benefit more from AI-based interactions" (Kang, 2022, 28). According to the study, learners of high and low levels benefitted from interactions with native speakers and AI, but the low-level learners obtained better results when interacting with the AI. This is hypothesized to come from anxiety induced in new learners when conversing with experienced speakers. This is further supported by an observation by the author, "The findings suggested that interactions with NSs might be more beneficial to high-level NNSs than interactions with AI interlocutors" (Kang, 2022, 28). The more experienced learners, less anxious about speaking, benefitted more from human interaction

than AI interaction. Using AI as a method to reduce stress in new learners, the technology could be applied to foster a positive learning environment at a critical stage of learning.

Another major advantage of AI is the increased effectiveness and efficiency of learning through the learning process. AI can automate the learning process for educators, creating learning activities and content for students with a limited degree of demand from the educator. A prime example of this implementation on a broad scale is demonstrated in research published by Duolingo, a leader in free online language education. In "Machine Learning-Driven Language Assessment", the researchers state, "Our approach is the first to use machine learning and natural language processing to induce proficiency scales based on a given standard, and then use linguistic models to estimate item difficulty directly for computer-adaptive testing" (Settles et al., 2020, 247). The process described in this paper, using machine learning to create proficiency standards and test them with AI language models, was utilized to create an English proficiency test that has been utilized by the company for assessing English skill levels of their users. A second, more specific example of using AI for generating educational content comes from an academic study done on creating resources for Icelandic learning. "The results indicate a practical benefit in using generative AI tools for creating unique content suitable for language learning at different learner levels and age groups… Future endeavors should also focus on enriching the stories with cultural information and possibly amusing content to foster a more immersive learning experience" (Simonsen, 2023, 282-283). Utilizing ChatGPT and DALL-E 2 to create a variety of educational resources, from short stories and simple exercises to images, this study found that the tools could produce resources benefiting the learning process. After practicing with these resources, the participants showed improvement in reading skills and expanded vocabulary. The content generated by these AI tools functioned well for educational

purposes, but it exhibited the telltale lack of cultural context necessary for a well-rounded language learning experience. Prompts were often twisted to focus on American characters practicing American customs as opposed to Icelandic stories written in Icelandic. The cultural inadequacies of AI should be considered when designing curriculum with these tools.

One significant disadvantage of AI can be seen in Low Resource Languages (LRLs). An LRL is a language that lacks a large pool of available resources and are usually has a small quantity of speakers, like Nepali or Zulu for example. In terms of AI translation, a language with an inadequate parallel corpus (sentence pairs spanning both target languages) is considered an LRL. Due to the heavy reliance on large data sets for AI training, AI translation for LRLs is a noticeable area of weakness. Despite this weakness, a 2023 survey which conducted an in-depth literature review of the prevailing papers on NMT for LRLs concluded that "Due to the recent advancements in the field, NMT is no longer an unattainable goal for LRLs" (Ranathunga, 2023, 27). Based on the sheer volume of parallel research taking place regarding the relationship between NMT and LRLs, the researchers were able to conclude that while currently underperforming, it is possible for the LRL-NMT landscape to shift in a positive direction following the varying successes in the field. At the end of the paper, the authors recommend a course of action to increase the performance of NMT for LRLs, "Based on our findings, we provide the following set of recommendations to advance the LRL-NMT solutions: (1) create more LRL resources (datasets and tools), (2) make computational resources and trained models publicly available, and (3) involve research communities at a regional level" (Ranathunga, 2023, 27). It is important to note that these recommendations, while being suggestions that would increase the quality of LRL-NMT interactions, do not serve to advance NMT technologies but rather attempt to directly address LRL by gathering and generating more resources for LRLs.

This illustrates how regardless of new research and methods for NMT, the technology struggles with LRLs and will likely continue to struggle.

Analysis

Advancing closely with the growing developments in computational power, artificial intelligence will continue to expand beyond the current limitations. Based on the data collected, AI has become a valuable instrument for the language learning process. In comparison with typical machine translation, AI has developed beyond the capabilities of any rule-based or statistical model. Evidence gathered points to a noticeable disparity in performance between NMT and any other form of machine translation, a divide which has only existed for roughly two decades due to the stagnation of the traditional approaches and the explosive development of AI technology in recent years. With AI experiencing such substantial growth in a relatively small period, the expanding trend of AI research illuminates the path forward for translation technologies. From a curriculum perspective, the increasing prevalence of AI warrants its use in language learning, if only to promote up to date technological literacy amongst learners. Beyond that simple argument, the degree of consistency available in modern machine translation courtesy of AI development has fostered a quality of linguistic resources conducive to language learning.

As demonstrated by the research collected above, AI can be used to generate content for educational purposes. Companies ranging from Duolingo to Babbel use AI for their content creation, as well as teachers and researchers on the order of a single classroom. The resources created by AI are considered generally sound, but the educator's attention is still required to

verify the accuracy of the AI. Despite that drawback, AI resource generation saves time by providing a framework for exercises and activities. Not only can AI expedite the learning process for the educator, but it can also ease the learning process for learners. New learners have been shown to respond positively to interactions with AI and this could be used to form a solid foundation for new learners. Informing new learners about how best to use AI and providing them opportunities to utilize the technology to complete introductory written and oral assignments. Additionally, AI could be trained to recognize the emotional state of the learner and respond accordingly which would reduce stress in users and increase motivation.

An important social aspect present in AI translation is the use of gender. Gender is a highly nuanced concept and AI translation is based on preexisting cultural contexts that are observed through training. As mentioned earlier, machine translators struggle to correctly identify gender in provided phrases. This cultural bias present regarding gender becomes clear when investigating the translation of certain verbs and occupations. AI translation has demonstrated low performance when changing between languages with differing grammatical gender. In one instance, the translation swapped the gender of a doctor from female to male and assigned female pronouns to a nurse of an unspecified gender. Following SCOT, this is a direct result of the social biases of language which AI translation relies on during operation. During translation, the AI assumes stereotypes because that inherent assumption was both present in the data that trained it as well as overlooked by those responsible for its creation. This is perhaps the biggest pitfall of machine translation as a whole and should be addressed with a renewed focus on gender unbiasing data sets. As the broader cultural perception on gender continues to shift as it has in the past century, AI must become more adept at handling gender if it is to be used

professionally. These faults do still provide some educational benefits, allowing educators to highlight mistakes made during translation and explain how to achieve a correct result.

Cultural contexts represent a key part of human communication. A lighthearted phrase in one culture may appear offensive to another. Understanding a new culture is one of the most difficult barriers in language acquisition. Language learning curriculums focus on sharing these cultural differences, from new foods to traditional holidays. Culture becomes especially important when using AI for language education or translation. AI struggles to carry meaning across cultural divides, often opting to frame response through the culture it was trained upon. SCOT applies well in this case, with the technology being so closely tied to its culture of origin. The AI has been trained to behave in a particular way biased towards the majority culture rather than the target culture. From names and locations to activities and phrases, AI will routinely develop content based on its origin. When using AI for language education, this weakness in cultural understanding must be addressed and at times remedied to promote the best educational experience.

As discussed previously, AI requires large quantities of data to create effective language models. This data is typically acquired either from massive databases or scraped directly from the Internet. This tremendous scale of data has restricted the scope of AI to commonly used and studied languages like English or Japanese. This presents a serious issue when it comes to LRLs. Languages with few resources and attention will struggle to implement AI effectively, whether that be direct translation or interaction with an AI in to complete a specific task. This unfortunate development can be easily explained through the lens of SCOT. The technology is developed by a large body of distinct, unrelated researchers which necessitates research to be done in a common language to expedite the sharing of knowledge. Moreover, AI benefits from training on

enormous data sets as discussed previously which promotes training for the languages most frequently used to generate content. As could be expected, the behavior of this technology is shaped by society and the optimal functionality of the technology is restricted to the majority that created it.

Conclusion

Considering the benefits and harms of AI for language learning, AI could serve as a versatile tool for language education in the coming years. For human history, translation has been reliant on a group of individuals fluent in multiple languages to bridge the gap between cultures. To learn a new language, the learner was dependent on multilingual individuals, whether that be through educational resources they created or through direct interactions. With the rise of AI for translation, that dependency has been removed for the everyday learner. Translations traditionally reserved for human beings can now be performed in seconds by anyone. This widespread availability of free translation is unprecedented on the global scale; such a growth for the spread of information being comparable to the invention of printing press. How humanity once suddenly gained access to the whole of information in their native tongues, now scientific progress has brought life to a system to provide access to information in all languages.

Notwithstanding, it is critical to highlight the use of AI as a tool for education, not a crutch or machine to perform instead of the user. A tool is designed to aid a person for a specific purpose, not to replace them. For AI to truly thrive in pedagogy, it must be implemented in a manner that highlights the process rather than the result. At its best, AI can guide the user to

further understanding with convenient translations and simple explanations of grammatical contexts. Any errors made by the tool can be noted and explained to learners, prompting discussions on difficult grammatical topics or even social topics such as gender, privilege, and bias. At its worst, AI can generate content that could promote poor habits in new language learners, fostering a dependence on the tool rather than a positive instructive relationship. In the coming years, AI will be a powerful instrument for change, and it is crucial to consider how it will be applied in the future.

Bibliography

Aiken, M., Ghosh, K., Wee, J. Z., & Vanjani, M. (2014). An evaluation of the accuracy of online translation systems. *Communications of the IIMA*, *9*(4). https://doi.org/10.58729/1941-6687.1122

Anikushina, V., Taratukhin, V., & Von Stutterheim, C. (2018). Natural language Oral communication in humans under stress. Linguistic Cognitive Coping Strategies for enrichment of Artificial Intelligence. *Procedia Computer Science*, *123*, 24–28. https://doi.org/10.1016/j.procs.2018.01.005

Bijker, W. E., Hughes, T. P., & Pinch, T. (1987). *The Social Construction of Technological Systems: New directions in the Sociology and History of Technology*. http://ci.nii.ac.jp/ncid/BA12006380

Chan, S. (2014). Routledge Encyclopedia of Translation Technology. In *Routledge eBooks*. https://doi.org/10.4324/9781315749129

Kang, Hanna. (2022). Effects of Artificial Intelligence (AI) and native speaker interlocutors on ESL learners' speaking ability and affective aspects. *Multimedia-Assisted Language Learning*, 25(2), 9-43.

Mondal, S. K., Zhang, H., Kabir, H. M. D., Ni, K., & Dai, H. (2023). Machine translation and its evaluation: a study. *Artificial Intelligence Review*, *56*(9), 10137–10226. https://doi.org/10.1007/s10462-023-10423-5

Mutal, J., Volkart, L., Bouillon, P., Girletti, S., & Estrella, P. (2019). Differences between SMT and NMT Output - a Translators′ Point of View. In *Proceedings of the Human-Informed*

Translation and Interpreting Technology Workshop (HiT-IT 2019) (pp. 75–81). Incoma Ltd., Shoumen, Bulgaria.

Ranathunga, S., Lee, E. A., Skënduli, M. P., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural Machine Translation for Low-resource Languages: A survey. *ACM Computing Surveys*, *55*(11), 1–37. https://doi.org/10.1145/3567592

Settles, B., LaFlair, G. T., & Hagiwara, M. (2020). Machine Learning–Driven Language Assessment. *Transactions of the Association for Computational Linguistics*, *8*, 247–263. https://doi.org/10.1162/tacl_a_00310

Simonsen, A.; Bédi, B. (2023). Using Generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic. In CALL for all Languages - EUROCALL 2023 Short Papers. 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16994

Yang, Y. (2024). Understanding machine translation fit for language learning: The mediating effect of machine translation literacy. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12650-x

Zuraina Ali. (2020). Artificial Intelligence (AI): A Review of its Uses in Language Teaching and Learning. *IOP Conf. Ser.: Mater. Sci. Eng.* https://doi.org/ 10.1088/1757-899X/769/1/012043