Journal of Universal Language 25-1. March 2024, 99-119
DOI 10.22425/jul.2024.25.1.99
eISSN 2508-5344

AI Chatbots and Linguistic Injustice

Sunyoung Park
Sejong University, Korea

Abstract

The rise of AI-powered language technologies, exemplified by products like DeepL and ChatGPT, has propelled the advancements towards widespread acceptance, integrating them into daily communication and professional routines. This transformation holds significant implications for social interactions and knowledge dissemination. However, the dominance of these technologies poses challenges, particularly for non-native English speakers. This dominance not only limits information access for non-native English speakers but also risks fostering a monocultural AI primarily proficient in English, neglecting other languages and cultures. Nonetheless, the revolutionary benefits of AI advancement may disproportionately benefit English native speakers unless corrective

Sunyoung Park
Visiting Professor, Daeyang Humanity College, Sejong University, Korea
Email: sunpark@sejong.ac.kr

Received 13 December, 2023; Revised 4 February, 2024; Accepted 21 March, 2024

measures are taken. The present study aims to offer a thorough review of the advancements, focusing specifically on the fields of education and healthcare. These two sectors have been significantly impacted by these improvements, and the study seeks to provide a detailed analysis of the changes and developments within them. To address this inequality, sourcing language training data from diverse linguistic backgrounds and implementing localization strategies are proposed as solutions. Additionally, collaboration between scientists and linguists can enhance the linguistic and cultural sensitivity of AI language models. Furthermore, introducing an artificial language into AI chatbot systems could mitigate inequality by enhancing accessibility and comprehension for non-native English speakers, thereby promoting inclusivity.

Keywords: large language model, modeling bias, detection bias, linguistic inequality, ChatGPT

1. Introduction

Language barriers and disparities in multilingualism has been widespread issues for many decades, and it has been impacting individuals and communities globally. In today's interconnected world, importance of cross-cultural communication is growing ever more essential due to globalization and the ability to speak and understand multiple languages hold significant importance. On the other hand, multilingual ailibies are not uniform across populations, causing unequal access to education, employment, healthcare, and other social opportunities (Kaplan & Haenlein 2019, Gleason 2022, Hirsh-Pasek & Blinkoff 2023, Shinde 2023). In addition, the advancement of AI[1] technologies has resulted in transformative

[1] The following abbreviations are used in this paper: AI (artificial intelligence), ChatGPT (chat generative pre-trained transformer), HAILEY (human-ai collaboration

changes across diverse domains, promising increased efficiency, accuracy, and accessibility. However, amidst this technological progress, a critical issue emerges: the exacerbation of biases against non-native English speakers. AI systems predominantly operate within the framework of English-centric datasets and models, they often struggle to effectively interpret and respond to linguistic variations, accents, and dialects prevalent among non-native English speakers. Consequently, non-native English speakers encounter systemic biases, leading to disparities in access to AI-powered services, including voice recognition systems, language translation tools, and automated decision-making processes (Bhirud et al. 2019, Bozkurt & Sharma 2020, Frangoudes et al. 2021, Biswas 2023, Cascella et al. 2023, Fütterer et al. 2023, Gilson et al. 2023).

Such biases not only impede the full integration of non-native English speakers into the digital landscape but also sustain linguistic inequalities in the realm of AI-driven technologies. Therefore, it is imperative to critically examine the mechanisms through which AI advancements engender biases against non-native English speakers and explore strategies to mitigate these disparities, fostering a more inclusive and equitable AI ecosystem for all users. Thus, the current paper tries to provide comprehensive reviews of AI chatbots and thier effects in existing industries and some biases caused by advancement of the system. In Section 2, we will firstly review the studies on the development of large language models and its effects to education and healthcare systems. Section 3 will discuss studies on the potential disadvantages and baises caused by the AI chatbot system to the non-native English speakers. Section 4 concluds the study by suggesting possible solutions.

approach for emphathy), LLM (large language models), NLP (natural language processing), TOEFL (test of English as a foreign language).

2. AI and New Realm

2.1. Higer Education

The world has changed significantly for the past few decades and the change is ongoing. One notable development that is causing a lot of interest in academia is the emergence of LLM such as ChatGPT, a NLP model developed by OpenAI (Shinde 2023). The model functions on extensive datasets, and it can respond to students' questions, feedback, and prompts (Biswas 2023, Gilson et al. 2023). In education, there has been increasing number of studies on the benefits and challenges of using chatbots (Aydın & Karaarslan 2022, Stokel-Walker 2022, Adeshola & Adepoju 2023, Bonsu & Baffour-Koduah 2023, Fütterer et al. 2023, Hirsh-Pasek & Blinkoff 2023). Some educators are optimistic about its potential to aid learning (Bonsu & Baffour-Koduah 2023). According to the research, one of the key applications of ChatGPT in the classroom is personalized learning opportunities. This entails developing educational resources and content specifically tailored to each student's individual interests, skills, and learning objectives (Bonsu & Baffour-Koduah 2023).

Others express concerns about its potential to generate learning opportunities or perpetuate misinformation (Fütterer et al. 2023). Fütterer and his colleagues analyzed Twitter data (16,830,997 tweets from 5,543,457 users) to understand reactions about ChatGPT concerning education. Based on topic modeling and sentiment analysis, they provided an overview of comprehensive perceptions and reactions to the chatbot. As one might expect, the chatpot triggered a massive response on Twitter, and 'education' was the most tweeted content topic, surpassing more general topics such as how to access ChatGPT. The topics include from specific terms such as ‘cheating’ to broad ones such as ‘opportunities’ and they were discussed with mixed sentiments.

According to the authors, it is surprising and meaningful because the platform could considerably modify professional practice in many fields, which centers on creative text production, such as journalism, book authoring, marketing, and business reports. It implies that educational stakeholders like school and higher education administrators, teachers, and policymakers should formulate guidelines to impelement the platform within their respective enviroenments.

The emergence of ChatGPT and similar AI chatbots has shed light on the vulnerability of the educational system to external threats (Bozkurt & Sharma 2020). These AI tools could potentially be utilized for cheating on exams or completing assignments without genuine effort, as they can deliver responses instantly upon demand. This not only compromises the integrity of the educational system but also puts students at a disadvantage if they lack access to such resources, especially when instructors are unaware of their usage and inadvertently rate those who use them higher.

Moreover, accrding to Hirsh-Pasek & Blinkoff (2023), the landscape of higher education has become increasingly competitive as shown in other industries. A multitude of universities and colleges now offer similar programs and cost structures, necessitating institutions to distinguish themselves and craft compelling brand identities to attract students. Hirsh-Pasek & Blinkoff (2023) argues that it is imperative that universities and colleges also ensure prospective students understand the unique benefits of enrolling with them. While some universities advocate for the integration of AI in education, others oppose it, resulting in a lack of consensus on its usage in higher education. Therefore, it’s essential for educators to model exemplary behavior (Hirsh-Pasek & Blinkoff 2023). As we have reviewed, the impact of LLM like ChatGPT in education is enormous.

2.2. Healthcare

Another field that has been heavily affected by the extensive large language processing models includes healthcare system. It was shown that 60% of doctor visits are for minor diseases, and 80% of them can be treated at home by using simple remedies (Bhirud et al. 2019, Frangoudes et al. 2021, Cascella et al. 2023). These diseases usually include cold, cough, headache, abdominal pains and so on. They are often known to be attributed to factors such as weather changes, poor nutrition, and fatigue, which can be managed without medical intervention (Cascella et al. 2023).

Chatbots can assist potential patients by offering basic healthcare information before they seek to make an appointment with doctors. It can predict the users' diseases based on symptoms and offers recommendations for precautions and remedies. If a severe illness is suspected, the device can advise users to seek medical assistance. Its primary goal is to communicate with the user in a manner like that of a doctor so that users can freely discuss any problems they may be experiencing. Acting as a virtual friend, the system aims to facilitate healthcare couseling to potential patients.

In addition, an increasing number of studies claim that LLMs can benefit mental healthcare systems (Aydın & Karaarslan 2022, Ayers et al. 2023, Kanjee et al. 2023, Sharma et al. 2023, Shryock 2023). That is because unlike search engines, which provide responses and links when they receive text inputs, chatbots like GPT-4 deliver reponses that are similar to human conversations. In fact, as WHO (World Health Organization 2022) estimates, approximately one in eight individuals worldwide are going through mental illness. This issue is compounded by stigmatization, human rights violations, and insufficient resources. In particulary, shortages of mental healthcare professionals prevent the patients from accessing to psychiatric treatment. Therefore, as the time of clinicians is highly limited resource in mental healthcare, advancements in artificial intelligence may enhance the efficiency of clinicians, and assist with some administrative duties.

Given these challenges, recent progress in generative AI and its potential to influence healthcare delivery have gained significant interest. Some studies suggest that chatbots which are powered by extensive language models have the potential to aid mental health peers and clinicians by consistently providing high levels of support during interactions with patients. For instance, one of the studies revealed that responses developed in collaboration with a chatbot named 'HAILEY' were more likely to be perceived as emphatic compared to responsos provided soley by humans (Sharma et al. 2023).

What is more, peer supporters who acknowledged difficulties in offering empathetic support were notably rated as more likely to deliver empathetic responses when supported by AI. In addition to aiding clinician documentation and patient interactions, an emerging strength of generative AI also lies in hypothesis generation. Preliminary studies demonstrate the potential of GPT-4 in generating accurate lists of potential diagnoses, particularly in complex clinical cases, indicating its ability to facilitate hypothesis formation (Ayers et al. 2023, Kanjee et al. 2023, Shryock 2023).

While acknowledging benefits of using generative AI chatbots, some risks of harm are also suggested (King 2022, Ferrara 2023, Gross 2023, Marks & Haupt 2023). In addition to the technological deficiencies commonly characterized by inconsistent responses and dissemination of false information, certain biases may also be observed, potentially leading to fatal outcomes. LLMs can write responses in a prompted conversational register such as tone or level. Nevertheless, due to a range of factors, biases are inherently ingrained, giving rise to the risk of 'algorithmic discrimination', and outcomes may sustain or worsen unfair treatment. Research indicates that these models can encode biases related to gender, race, and disability, jeopardizing their equitable implementation (King 2022, Ferrara 2023).

Ferrara (2023) explains that bias stems from mutliple sources. Training data often exhibit gaps in representation from clinical population, particularly in medical publications like PubMed. Moreover, stereotyping seems to emerge from diverse sources such as social media platforms including Twitter, Facebook, and among others. Furthermore, biases from people and society can get into the system through supervised learning.[2] Workers, who often don't get paid much, might continue unfair stereotypes when they label data and give feedback (Ferrara 2023).

As we have reviewed, Artificial Intelligence holds the potential to revolutionize industries such as education and healthcare, and results in advancements, creativity, and enhanced effectiveness. To summarize, in education, it can provide individualized learning opportunities and extend quality of education to distant areas. In terms of healthcare system, AI can aid in early detection of diseases and provide customized treatment strategies. While it tends to provide potentially advantageous landscape in such areas, there tends to exist some issues related to 'bias' issues in using the tool.

[2] Supervised learning is a type of method in machine learning in which algorithm learns from labled data.

3. Biases to Non-Native English Speakers

3.1. Language Input Bias

In section 2, we have reviewed benefits of LLM in education and healthcare systems. However, the prevalent understanding is that AI-driven language technology which encompasses large language models, machine translation systems, multilingual dictionaries, and corpora, is presently limited to merely the world's predominant languages, those that receive substantial financial and political backing. AI systems, particularly language models, heavily rely on a vast array of online data sources and corpora including forums, articles, and enclyclopedias, among many others. However, there seems to be notable imbalance in this digital landscape. That is to say that dominance of English language is overwhelming, whereas other languages are underrepresented.

Figure 1. Usage Statistics of Content Languages for Websites (Adapted from w3techs (2024) with Permission of Q-Success)

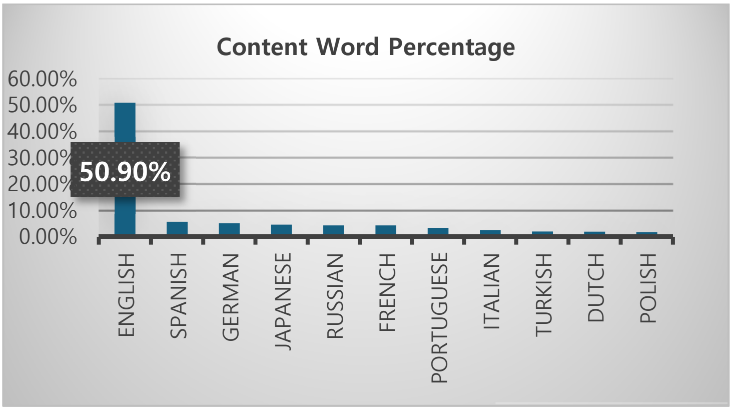

Figure 1 shows the percentage of websites using a range of content languages. As the figure represents, among the websites whose content language we are aware of, English is utilized by 50.9%. This linguistic inequality suggests significant challenges in AI development and causes important issues about fairness and inclusiveness in the digital age. The widespread usage of English provides English native speakers with great advantages by providing easy access to abundant useful information and resources. Naturally, language models that are predominantly trained on English language data could exhibit more advanced and comprehensive understanding of data and nuanced capability in English language AI applications. In fact, if one tries to ask questions or exchanges conversations in English, it would give much more information than it was asked with other languages like Korean on the same topic. For example, when you ask about a simple question such as "Please provide me with features of rhinoceros beetles" in English, it would give you answers containing 8 categories of its characteristics in a well-organized format, which enhances readability. On the other hand, if it is asked the same question in Korean, it would give you the similar answer, but only containing 4 catetories of the features of the insect. Even though this is a simplistic example, it unequivocally demonstrates the disparity in the amount of information accessible when conversing in English compared to when conversing in other languages.

3.2. Language Modeling Bias

It is widely known that recent language models have improved significantly (Devlin et al. 2018, Brown et al. 2020, Clark et al. 2020). Recent advancements in language modeling have embraced the approach of training large-scale models on extensive, unannotated corpora using self-supervised learning techniques. These methods involve predicting masked words and the next sentence in a sequence. (Devlin et al. 2018, He et al. 2020), wrong word detection (Clark et al. 2020), and left-to-right language generation (Brown et al. 2020, Raffel et al. 2020).

The recent natural language processing models are trained by assessing the similarity vocabularies and sentences in text. Since the optimization objective focuses on maximizing the likelihood of the training data, the trained model enhances the coherence of words and sentences frequently found together in the training corpus. However, being created by humans, the training data sets can contain significant amounts of social bias and stereotypes, encompassing factors such as gender, race, and religion (Kiritchenko & Mohammad 2018, Nadeem et al. 2021, Stanczak & Augenstein 2021).

Some studies have demonstrated that pretrained language models are capable of acquiring various forms of stereotypical and baised reasoning. For example, Kiritchenko & Mohammad (2018) examined how language models perform in sentiment analysis across various social groups, measuring differences in their behaviours. Recent studies by Nangia et al. (2020) and Nadeem et al. (2021) investigated stereotypical reasoning related to race, gender, profession, and religion using masked language models and sentence encoders.

Recent research examined strategies to reduce the social biases inherent in language models, aiming to enhance their reliability. These studies have investigated techniques to mitigate biases during the learning and prediction phases of language models. Typical methods for mitigating bias involve the use of counterfactual data augmentation (Zmigrod et al. 2019, Dinan et al. 2020, Webster et al. 2020, Barikeri et al. 2021), dropout regularization (Webster et al. 2020), and self-debias (Schick et al. 2021). MIT researchers have trained language models that can realize logic to avoid harmful stereotypes such as gender and racial biases. Luo & Glass (2023) trained a language model to predict the connection between the sentences based on context and semantic meaning. They used data sets with lables for extracted texts showing whether a subsequent phrase "entails", "contradicts", or neutral. These data sets were referred as natural language inerence and they found that the logic-based model is considerably lower biased than the previous models.

Furthermore, according to science and technology scholar Winner (2017), language technologies can be regarded as inherently political due to their capacity to drive significant social changes (Winner 2017). Recognizing that language technologies are not only sociotechnical but also fundamentally political, it becomes essential to scrutinize how they prioritize certain perspectives and how their specific design is influenced by the interests and ideologies of particular groups. From an ethical standpoint, the implementation of language technologies requires thoughtful consideration of their inherent biases to prevent any discriminatory effects on marginalized communities.

3.3. Detection Bias

One of the biases that are causing disadvantages to non-native English speakers can be a detection bias. A research team at Stanford University recently evaluated seven popular GPT detectors by analyzing 91 English essays written by individuals whose native language is not English (Liang et al. 2023). The essays were written as part of the TOEFL exam, and more than 50 percent of the essays were flagged as the production of an AI chatbot. Among the detectors, one program flagged 98% of them as having been composed by AI. On the other hand, when essays written by eighth-grade students whose native language is English and living in the United States were analyzed with the same AI detectors, 90% of the essays were classified as human-generated.

It is surprising that more than half of human generated essasys are categorized as the product of AI. In order to identify the surprising results, the researchers examined the source of discrimination in how the AI detectors distinguish between human and AI-generated contents. According to Liang et al. (2023), AI detectors evaluate “text perplexity”, which measures how “surprised” or “confused” a generative language model is when predicting the subsequent word in a sentence. The text perplexity is considered low if the model can predict the subsequent word easily. On the other hand, vice versa, if the model finds the next word is difficult to predict, the text perplexity is ranked high.

In other words, LLMs like ChatGPT are trained to produce text with low perplexity. However, it can mean that if human writers show limited word choices and use a lot of common words, the detector system can misjudge the text as AI-generated. The risk is much greater with non-native English speakers because they are more prone to use simple words than native English speakers.

Once the researchers identified the bias in the detector programs, they requested ChatGPT to revise the essays using more complex languages and run the edited essays with the detectors again. Surprisingly, these modified essays were all identified as human-authored. Considering these results, as the researchers also noted, one can concern that the GPT detectors overall may encourage non-native English speakers to rely more on GPT or other generative Chatbots in their writings to avoid detection.

Liang et al. (2023) highlighted the seriousness of implications of GPT for non-native English writers and it is imperative to figure out the problem to avoid discrimination. This is especially so because AI detectors could categorise college or job applications as GPT-generated and accuse them of cheating, and this could potentially marginalize non-native English speakers. This could bring about serious consequences on students' mental well-being.

4. Concluding Remarks

Since the introduction of products like DeepL and ChatGPT, AI-powered language technologies have been steadily advancing towards mainstream acceptance, becoming an indispensable aspect of daily communication and professional routines. Consequently, they play a pivotal role in shaping social interactions and influencing the generation and dissemination of knowledge.

Nevertheless, this dominance presents its own array of challenges, especially to non-native speakers of English. It not only provides limited information to non-native English speakers but also poses the risk of developing a monocultural AI that engages in English but lacking comparably low proficiency in other underrepresented languages. This phenomenon could not only constrain global application of AI, but also create potential cultural biases. Recently, in his blog Gates Notes (2023), Bill Gates discusses how AI is poised to revolutionize computer usage.

He highlights that current software often requires users to navigate through different applications for various tasks, and even the most advanced programs lack a comprehensive understanding of users' lives. Gates envisions a future where AI agents will enable users to communicate with their devices using everyday language, thus eliminating the need for multiple apps. According to Gates, these programs called AI "agents" will possess a deep understanding of users' private lives, thus allowing for personalized assistance and streamlining interactions with computers. Gates emphasizes that this development will not only transform user-computer interactions but also revolutionize the software industry (Gates 2023).

What is concerning is these revolutionary benefits of advancement could be only limited to the English native speakers unless appropriate measuers are taken in a timely manner. In order to prevent non-native speakers from being left behind by the native English speakers, it is most important to understand the inequalities to non-native speakers as shown in this study. One potential solution can be sourcing language training data from a wider range of languages, if possible. Or, to make AI tools more relevant and usable in various contexts, localization strategies could also be helpful. In addition, scientists can collaborate with linguists to generate language models that are more linguistically and culturally sensitive. Lastly, another rather radical approach could be intoducing an artificial language into AI chatbot systems to cater to non-native English speakers, which could potentially help reduce inequality. This approach might enhance accessibility and comprehension for individuals who are not fluent in English, thereby promoting inclusivity and leveling the playing field in interactions with AI systems (Park & Tak 2017; Park & Chin 2020; Park 2021, 2022, 2023; Chin 2023). However, it's crucial to ensure that the artificial language is effectively designed and implemented to accurately convey information and maintain clarity in communication.

References

Adeshola, I. & A. Adepoju. 2023. The Opportunities and Challenges of ChatGPT in Education. *Interactive Learning Environments* 1–

14. DOI: 10.1080/10494820.2023.2253858.

Aydın, Ö. & E. Karaarslan. 2022. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. *Emerging Computer Technologies* 2, 22–31. DOI: 10.2139/ssrn.4308687.

Ayers, J. et al. 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine* 183.6, 589–596. DOI: 10.1001/jamainternmed.2023.1838.

Barikeri, S. et al. 2021. RedditBias: A Real-World Resource for Bias Evaluation and Debiasing of Conversational Language Models. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 1941–1955. Stroudsburg, PA: Association for Computational Linguistics.

Bhirud, N. et al. 2019. A Literature Review on Chatbots in Healthcare Domain. *International Journal of Scientific & Technology Research* 8.7, 225–231.

Biswas, S. 2023. Role of Chat GPT in Education. Available at <https://ssrn.com/abstract=4369981>.

Bonsu, E. & D. Baffour-Koduah. 2023. From the Consumers' Side: Determining Students' Perception and Intention to Use ChatGPT in Ghanaian Higher Education. Available at <https://ssrn.com/abstract=4387107>.

Bozkurt, A. & R. Sharma. 2020. Emergency Remote Teaching in a Time of Global Crisis due to CoronaVirus Pandemic. *Asian Journal of Distance Education* 15.1, i–vi. DOI: 10.5281/zenodo.3778083.

Brown, T. et al. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33, 1877–

1901.

Cascella, M. et al. 2023. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems* 47, 33. DOI: 10.1007/s10916-023-01925-4.

Chin, S. 2023. Linguistic Diversity and Justice: The Role of Artificial Languages in Multilingual Societies. *Journal of Universal Language* 24.2, 71–89. DOI: 10.22425/jul.2023.24.2.71.

Clark, K. et al. 2020. Electra: Pre-Training Text Encoders as Discriminators Rather than Generators. *Proceedings of the Eighth International Conference on Learning Representations.* Addis Ababa: International Conference on Learning Representations.

Devlin J. et al. 2018. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. Cornell University. Available at <https://arxiv.org/abs/1810.04805>.

Dinan, E. et al. 2020. Queens Are Powerful Too: Mitigating Gender Bias in Dialogue Generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 8173-8188. Stroudsburg, PA: Association for Computational Linguistics.

Ferrara, E. 2023. Should ChatGPT Be Biased? Challenges and Risks of Bias in Large Language Models. Available at <https://ssrn.com/abstract=4614228>.

Frangoudes, F. et al. 2021. An Overview of the Use of Chatbots in Medical and Healthcare Education. *International Conference on Human-Computer Interaction* 170-184. Cham: Springer International.

Fütterer, T. et al. 2023. ChatGPT in Education: Global Reactions to AI Innovations. *Scientific Report*s 13, 15310. DOI: 10.1038/s41598-023-42227-6.

Gates, B. 2023. AI Is about to Completely Change How You Use

Computers. Available at <https://www.gatesnotes.com/AI-agents>.

Gilson, A. et al. 2023. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education* 9, e45312. DOI: 10.2196/45312.

Gleason, N. 2022. ChatGPT and the Rise of AI Writers: How Should Higher Education Respond? The Campus Learn, Share, Connect. Available at <https://www.timeshighereducation.com/campus/chatgpt-and-rise-ai-writers-how-should-higher-education-respond>.

Gross, N. 2023. What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases in AI. *Social Sciences* 12.8, 435. DOI: 10.3390/socsci12080435.

He, P. et al. 2020. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. Cornell University. Available at <https://arxiv.org/abs/ 2006.03654>.

Hirsh-Pasek, K. & E. Blinkoff. 2023. ChatGPT: Educational Friend or Foe? *Brookings*. Available at <https://www.brookings.edu/blog/education-plus-development/2023/01/09/chatgpt-educational-friend-or-foe/>.

Kanjee, Z. et al. 2023. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. *Journal of the American Medical Association* 330.1, 78–80. DOI: 10.1001/jama.2023.8288

Kaplan, A. & M. Haenlein. 2019. Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons* 62.1, 15–25.

King, M. 2022. Harmful Biases in Artificial Intelligence. *The Lancet Psychiatry* 9.11, e48. DOI: 10.1001/jamainternmed.2023.1838.

Kiritchenko, S. & S. Mohammad. 2018. Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* 43-53. New Orleans, LA: Association for Computational Linguistics.

Liang, W. et al. 2023. GPT Detectors Are Biased Against Non-Native English Writers. *Patterns* 4.7, 100779. DOI: 10.1016/j.patter.2023.100779.

Luo, H. & J. Glass. 2023. Logic Against Bias: Textual Entailment Mitigates Stereotypical Sentence Reasoning. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* 1243–1254. Dubrovnik: Association for Computational Linguistics.

Marks. M. & C. Haupt. 2023. AI Chatbots, Health Privacy, and Challenges to HIPAA Compliance. *Journal of the American Medical Association* 330.4, 309–310.

Nadeem, M. et al. 2021. Stereoset: Measuring Stereotypical Bias in Pretrained Language Models. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 5356–5371. Online Conference: Association for Computational Linguistics.

Nangia, N. et al. 2020. Crows-pairs: A Challenge Dataset for Measuring Social Biases in Masked Language Models. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 1953–1967. Online Conference: Association for Computational Linguistics.

Park, S. 2021. The Necessity of Minimal Representation of Genericity in a Newly Developed Language, Unish. *Journal of Universal Language* 22.2, 87–104. DOI: 10.22425/jul.2021.22.2.87.

Park, S. 2022. Typological Analysis of Articles in World Languages. *Journal of Universal Language* 23.1, 109–127. DOI: 10.22425/jul.2022.23.1.109.

Park, S. 2023. Multilingualism, Social Inequality, and the Need for a Universal Language. *Journal of Universal Language* 24.1, 77–93. DOI: 10.22425/jul.2023.24.1.77.

Park, S. & J. Tak. 2017. Articles in Natural Languages and Artificial Languages. *Journal of Universal Language* 18.1, 105–127. DOI: 10.22425/jul.2017.18.1.105.

Park, S. & S. Chin. 2020. Examining the Irregularities of Articles and Introducing Minimized NP Systems in Unish. *Journal of Universal Language* 21.1, 69–88. DOI: 10.22425/jul.2020.21.1.69.

Raffel, C. et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21.140, 1–67.

Schick, T. et al. 2021. Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP. *Transactions of the Association for Computational Linguistics* 9, 1408–1424.

Sharma, A. et al. 2023. Human–AI Collaboration Enables More Empathic Conversations in Text-Based Peer-to-Peer Mental Health Support. *Nature Machine Intelligence* 5.1, 46–57.

Shinde, S. 2023. What is ChatGPT? Top Capabilities and Limitations You Must Know. Emeritus Online Courses. Available at <https://emeritus.org/blog/ai-ml-what-is-chatgpt/>.

Shryock, T. 2023. What Patients and Doctors Really Think about AI in Health Care. *Medical Economics* 100.7, 14–16.

Stanczak, K. & I. Augenstein. 2021. A Survey on Gender Bias in Natural Language Processing. arXiv preprint arXiv:2112.14168.

Stokel-Walker, C. 2022. AI Bot ChatGPT Writes Smart Essays — Should Professors Worry? *Nature*. Available at <https://www.

nature.com/articles/d41586-022-04397-7>.

w3techs. 2024. Usage Statistics of Content Languages for Websites. Available at <https://w3techs.com/technologies/overview/content_language>.

Webster, K. et al. 2020. Measuring and Reducing Gendered Correlations in Pre-Trained Models. arXiv preprint arXiv:2010.06032.

Winner, L. 2017. Do Artifacts Have Politics? In J. Weckert (ed.), *Computer Ethics* 177–192. London: Routledge.

World Health Organization. 2022. World Mental Health Report: Transforming Mental Health for All. Available at <https://www.who.int/publications/i/item/9789240049338>.

Zmigrod, R. et al. 2019. Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 1651–1661. Florence. Association for Computational Linguistics.