Proceedings of the 5th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2026.37112

Implicit Bias among Bilingual and Monolingual Individuals

Xi Hu

Shandong Experimental High School, Jinan, China
once05281020@gmail.com

Abstract. Although the world and society are developing, bias about race or gender has decreased; there are still implicit biases that exist in people's lives and society. This review focuses on these implicit biases, mainly discussing the differences in implicit bias across monolingual individuals and bilingual individuals. Examining why they can reduce their implicit bias in their speech as well, which is mainly due to stronger pragmatic inference capabilities, greater cognitive flexibility, and the ability to think in other's perspective. In addition, this ability is regulated by intrinsic motivations; when intrinsic motivation is not enough, bilingual individual can use the cognitive flexibility they learned during their bilingual environment to overlap. At the same time, studies also examine the implicit bias in large language models(LLMs) and found that implicit bias in LLMs will change alongside different languages. This review mainly analyzes the implicit bias about gender in Hindi and English, and the different tendencies about political issues in Chinese or English. These studies have shown that implicit bias has plasticity, and it can be changed by the information environment, different narrative styles, and language structures. Since LLMs are trained by people and their values are aligned with people, people's implicit bias will affect them. So society should expand multilingual training data to reduce implicit bias.

Keywords: Implicit bias, large language model, bilingual, monolingual

1. Introduction

As the society develops, people start to generally stop stereotype for certain people, however, some laws or rules can only stop explicit bias, in many situations, some subtle and implicit bias are still exist. This review aimed to find and organize the implicit bias exist in different group of people and underline these implicit bias. Then promoting the future improvement of people's awareness of implicit bias and advocating avoid these bias.

In this review, some key word will be mentioned, there are the explanations for the terms. Implicit bias or implicit stereotypes refer to the situation where individuals attribute certain traits to members of a certain social group without conscious reflection. It is also known as unconscious bias and stems from the human tendency to categorize the social world into different groups. Implicit stereotypes are believed to be shaped by experience and are based on the learned associations between specific traits and social categories. Even if individuals sometimes do not consciously realize that they hold such stereotypes, their perceptions and behaviors can still be influenced by these implicit stereotypes. Implicit bias is an aspect of implicit social cognition, meaning that

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.37112

perceptions, attitudes, and stereotypes can arise before there is any conscious intention or approval. There have been numerous scientific studies in the field of psychology that support the existence of implicit bias. And Large Language Models refers to a neural network trained on a vast amount of text, used for natural language processing tasks, particularly language generation. LLMs are typically capable of generating, summarizing, translating, and analyzing text in various situations and are one of the core technologies of modern chatbots. When the training data is biased or inaccurate, it may lead to unreliable output results from LLMs.

2. In monolingual and bilingual individuals

In the reviewed studies, bilingual and multilingual participants often showcases stronger social implicit processing abilities or lower levels of implicit bias compared to monolingual participants, although specific results varied depending on different task type and age group. As found in study, monolinguals exhibit an implicit distrust towards non-native speakers [1]. In the experiment, participants listened to sentences spoken by speakers with Spanish or Chinese accents, and then they were asked to interpret the hidden ambiguous references. When the speakers had foreign accents, monolingual people will choose fewer second noun phrase (NP2), which indicate that they have reduced the rely on the cues provided by the speakers' intentions. This suggests that when dealing with non-native speakers, monolinguals may rely less on speaker cues, possibly because they consider these cues less reliable [1]. In contrast, bilingual people did not show any obvious bias or distrust toward the non-native speakers [1]. This maybe because bilinguals are more familiar with communicating across different languages, so they will rely less on whether the speakers' pronunciation are native or not. Also, because of the frequent handling of and dealing with the input information from non-native speakers, and themselves often being non-native speakers as well, bilingual people often tend to develop a more tolerant and flexible understanding and interpretation. They are less likely to have judgments just according to the accents, but more inclined to focus on broader contextual cues [1]. This finding is particularly significant because it indicates that implicit bias is not fixed but are malleable, and can be weakened through continuous language exposure. In addition to bilingual adults, bilingual children also exhibit lower implicit biases. In the study conducted by another group of researchers, the researchers asked bilingual French-English children and their monolingual French peers to simultaneously read 12 French sentences, all of these reading materials contained presuppositions, implicit meanings, or irony [2]. After reading, they ask the children to answer the comprehension questions. The results showed that bilingual children performed significantly better than their monolingual peers in understanding French [2], especially in understanding the underlying meanings and irony in French. In this study, understanding underlying meanings and irony require children to have the ability to integrate speakers meaning, the social contents, and the background rapidly, and this process is similar to the process of inhibiting the implicit biases. They both about beyond the superfacial words meanings, but to explore the true meanings in others' sentences. Though these findings do not directly measure the implicit bias, they can suggest that bilingual children may handle progmatic-inference skills that can help them to interpret complex social information without the assist of automatic assumption [2]. While bilingual children have the same strengthen in their relative less familiar language(English), showing that they have develop an underlying strategy for this decreasing implicit biases. So, they may also apply this strategy into interactions between different group of people, they can have a better understanding in knowing out-group people or other language people's meaning. Then they can decrease the implicit biases [2].

DOI: 10.54254/2753-7048/2026.37112

Moreover, the thesis that bilingual people will have less implicit biases than monolingual people is indicated in high school students. In another study, they asked the same number of American high school students--both bilingual and monolingual, to have an Implicit Association Test (IAT) designed to measure racial biases [3]. They need to match 5 ethnic groups, including white, black, Latino, East Asian, Middle Eastern people with positive or negative adjectives for their frontal facial images. The results show that bilingual high school students, compared to their monolingual peers of the same age, exhibit fewer negative associations when linking black faces with negative adjectives [3]. This suggests that they have a weaker implicit racial bias in this task. For the behind reasons, researchers think maybe because the diversity of the education, which often includes exposure to multiple cultural narratives, historical perspectives, and social norms has improved the flexibility of bilingual people's cognition, so their implicit biases have decreased. This reason have been mentioned in previous study as well [4]. There findings may have suggest that bilingual people may relate to reduced implicit bias indirectly, rather than considering a direct indicator of bias reduction. And the possible reduction of implicit bias may come from cognitive flexibility [4]. The same trend was observed among children as well. In the study investigated the implicit and explicit racial biases of monolingual and bilingual preschool children [5]. The research selected monolingual children from Hangzhou and Singapore, as well as bilingual children from Singapore. The results showed that both the monolingual children from the two places exhibited higher implicit racial biases than the bilingual children [5]. According to the previously established studies on racial biases, it can be illustrated that bilingual individuals exhibit higher cognitive flexibility, and high cognitive flexibility can be confirmed as a factor in reducing implicit biases. However, they also found that internal motivation is a crucial moderating variable. When individuals can actively control their biases due to their personal good values, the bilingual will no longer become a significant advantage; however, when internal motivation is low (that is, individuals cannot control their biases through their personal values), the cognitive flexibility of bilinguals will become a compensatory resource, effectively reducing the expression of implicit biases. This pattern may be more pronounced at the implicit bias level because implicit biases are less influenced by social expectations and more accurately reflect automatic attitudes, thus being more sensitive to changes in external regulatory factors. Beyond the internal motivations and cognition flexibility, the experience of bilingual usage or leaning will contribute to lower implicit biases as well. In 2021's study, both bilingual and monolingual adults are asked to perform director tasks(tester need to move objects according to directions, and during the test, they need to use director's idea to perform the right movements), and the results showed that bilingual people have a higher accuracy than monolingual people [6]. Specially, multicultural identifiers have a better performance as well, even they are not bilingual people [6]. This result has shown that bilingual people can better overcome automatic reactions, including self-perception and stereotypes, and adopt others' information [6]. The advantage in selecting perspectives of bilinguals (that is, the overcoming of automatic reactions due to multiple cultural environments and cultural language shaping) may directly translate into less automatic bias towards cross-group members [6]. The independent contribution of multicultural experience suggests that implicit bias may be influenced by both the "cognitive flexibility" (language-driven) and culture-driven dual paths. These pathways may interact dependently, such that bilinguals who also identify with multiple cultures exhibit the lowest levels of implicit bias [6].

In conclusion, a large body of evidence indicates that bilingualism, to a certain extent, is closely related to exposure to diverse cultures and the reduction of implicit biases throughout the stages from childhood to adulthood. This reduction is achieved through various mechanisms, including enhanced pragmatic reasoning skills, cognitive flexibility, and education in a bilingual environment.

DOI: 10.54254/2753-7048/2026.37112

However, these advantages are influenced by individual intrinsic motivation differences and the nature of specific tasks or situations [5]. Therefore, in future research, bilingual education and multicultural experiences can be promoted as tools for cognitive development and a subtle yet effective means to help eliminate social biases. By expanding language abilities, it may also enhance empathy, understanding, and the ability to treat others fairly.

3. In large language models

Despite the large language models(LLMs) have achieved many notable advances in multilingual generation, there is still implicit bias within their outputs. And the implicit bias - the automatic and unconscious associations that influence the output of the model can vary significantly depending on the language used in the interaction. This difference is not only related to the accuracy of translation or the size of the training data, but also reflects the deep social language structures, cultural norms, and political backgrounds embedded in the training corpus. By drawing upon two recent studies, this section examines the implicit bias within LLMs that varies across different languages, centering its analysis on gender bias in Hindi and English and political bias in Simplified Chinese and English. The research studied gender-related implicit bias in Hindi and English text generation, comparing responses generated by GPT-4 and Claude-3 Sonnet [7]. The results have shown that when the LLM is handling with Hindi language, 87.8 percent of their response have implicit bias, while when using English, there are only 33.4 percent of the responses are with bias [7]. And this disparity can be attributed to several different factors. One major factor is that Hindi is a language that highly gendered, meaning that grammatical gender is frequently marked across different parts of speech [7]. In their language, gender is marked not only on pronouns but also on verbs, adjectives, and even nouns [7]. When people want to say a ambiguous sentences, they need to encode the subject's gender, so when they speak some thing, they must have a decision about the gender, therefore amplifying the stereotype [7]. In contrast, English show gender in their speech less frequently, mainly in pronouns, which allows more syntactic ambiguity and less stereotype [7]. At the same time, in Hindi the language is often with incline of genders [7]. For instance, "teacher"will be translated to "adhyapak"(male) or "adhyapika"(female), so the available of neural context will be decreased and the LLMs will also learn about these stereotype since they are based on people's training and alignment [7]. What's more, the social and cultural context also contribute to the complexity and stereotype [7]. In Indian, gender will be embedded into name, class, and regional hierarchies [7]. So, in the study, they found that LLMs will rely on the power of class and social hierarchies. For instance, the models exhibited a tendency to link lower-status occupations like cleaners to women and high-status roles such as engineers to men, which indicate the intersection between gender stereotype and social hierarchy [7]. And this pattern is more obvious in Hindi than in English, which imply that the model has absorbed not only linguistic rules on gender, but also some social status [7]. However, the study also found something interesting, when using Hindi language that less common in digital communication, the stereotype is remarkably less than using the common Hindi [7]. For the less common one, the percent of response that without any stereotype is only 46, but for the common used one, the percent of response that without stereotype is only 18.75 [7]. This may showcases that the sources that used for training may also important for reducing stereotype. If the LLMs developers can make their model contact with more diverse language or digital information, the stereotype may be reduced [7]. Another study has showed the implicit bias about political issues by using LLMs to ask the same question in different languages [8]. During the study, researchers based on 184 science questions and 533 political questions about China and American [8]. And the results showed that the questions that related to China have a

DOI: 10.54254/2753-7048/2026.37112

notable lower consistency in different language than science questions or American questions, which can be seen that China-related questions have a consistency of 72.6 percent, while science questions are 85.3 percent [8]. Also,the results reveal a massive divergence: the GPT in English language exhibit a negatively score of 0.56 toward China issues, whereas the GPT in Chinese LLM model only reveal a negatively score of 0.16 [8]. In contrast, both models are equally critical for American issues, both are 0.33 negatively scored. So this means the difference mainly happened in Chinese political issues [8]. Also, during the study, researchers have found the in-group bias. GPT models will become more tolerant to the issues of the country that they being trained [8]. Which means that when ask GPT that trained with Chinese, they will have a less negative attitude and response on issues that associated with China than the issues that associated with English or country that use English as their main language [8]. By contrast, GPT trained by English will have a less negative attitude the issues that related to American. At the moment that an overall 79.5 percent of divergence, the Chinese GPT cited Chinese official statistics and narratives, while the English GPT cited international sources such as World Bank data [8]. Similarly, when asked about arresting dissident, the Chinese GPT will underlined the legal violations, while American GPT will define it as political suppression. Such divergence chould arise from disparities in information environments, encompassing content-regulation and public-narrative mechanisms in China alongside the comparatively more critical media environment in the United States [8]. As another study that studied LLMs with less popular language, the study pointed out that subtle implicit stereotypes in large language models often lie hidden in open-ended generation scenarios [9]. Even with the use of artificial safety alignment models, since most LLMs are trained in English, multilingual LLMs still reproduce implicit biases rooted in cultural backgrounds, and the degree of bias is often more severe in low-resource languages with insufficient training data [9, 10]. Compared with LLMs implicit biases in gender, political implicit biases not only come from linguistic structures or training data, but also driven by both training data and political context or environments [7, 8]. However, both are shown that nowadays the most popular tool---LLMs contain implicit bias, and deeper implies that language is not a neutral carrier of information but an active filter and acceptor. During the process of receiving information and learning these information, they will know how to retrieve or evaluate some certain information [7, 8].

4. Conclusion

Based on these studies, this paper can infer that bilingual individuals have smaller implicit biases compared to monolingual individuals, and this difference may stem from the experience of learning and using multiple languages. Therefore, governments or individuals can reduce implicit biases by encouraging the learning of more languages or gaining a deeper understanding of other cultures. However, it should be noted that the effect of bilingualism in reducing biases mainly depends on an individual's intrinsic motivation and cognitive flexibility resources. Bilingualism as a possibility to reduce implicit biases is not a universal solution to counter implicit biases; when an individual's inner values are strong, the bilingual learning experience may only play a small part. For large language models (LLMs), developers can achieve balanced training across different languages to avoid implicit biases when using unfamiliar languages. However, this review does not address whether implicit biases change when monolingual children grow up to become bilingual adults.

DOI: 10.54254/2753-7048/2026.37112

References

[1] Contemori, C., & Tortajada, F. (2020). The use of social–communicative cues to interpret ambiguous pronouns: Bilingual adults differ from monolinguals. *Applied Psycholinguistics, 41*(1), 51–77. https://doi.org/10.1017/S0142716419000407

[2] Laloi, A., Isel, F., & Godart-Wendling, B. (2025). Bilingual children's understanding of implicit meaning in primary-school. *Journal of Education, Innovation and Communication, 7*(3), 15–26. https://doi.org/10.3316/informit.T202507080000410020760265 8

[3] Zhu, Y. (2023). The association between language ability/cultural background and racial bias: A pilot study. *SHS Web of Conferences, 179*, Article 02017. https://doi.org/10.1051/shsconf/202317902017

[4] Castro, S., Bukowski, M., Lupiáñez, J., & Wodniecka, Z. (2023). Bilingualism is related to reduced expression of stereotypes: The role of cognitive flexibility and motivation [Manuscript submitted for publication]. Institute of Psychology, Jagiellonian University. https://osf.io/4nem8/

[5] Singh, L., Moh, Y., Ding, X., Lee, K., & Quinn, P. C. (2021). Cognitive flexibility and parental education differentially predict implicit and explicit racial biases in bilingual children. *Journal of Experimental Child Psychology, 204*, Article 105059. https://doi.org/10.1016/j.jecp.2020.105059

[6] Navarro, E., & Conway, A. R. A. (2021). Adult bilinguals outperform monolinguals in theory of mind. *Quarterly Journal of Experimental Psychology, 74*(11), 1841–1851.

[7] Gupta, I., Joshi, I., Dey, A., & Parikh, T. (2025). "Since lawyers are males.": Examining implicit gender bias in Hindi language generation by LLMs. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)* (pp. 3254–3264). Association for Computing Machinery. https://doi.org/10.1145/3715275.3732208

[8] Zhou, D., & Zhang, Y. (2024). Political biases and inconsistencies in bilingual GPT models—the cases of the U.S. and China. *Scientific Reports, 14*, Article 25048. https://doi.org/10.1038/s41598-024-76395-w

[9] Naous, T., Schwarz, M., & Rehbein, B. (2024). Surfacing subtle stereotypes: A multilingual, debate-oriented evaluation of modern LLMs [Preprint]. arXiv. https://arxiv.org/abs/2511.01187

[10] Xu, Y., Hu, L., Zhao, J., Qiu, Z., Xu, K., Ye, Y., & Gu, H. (2025). A survey on multilingual large language models: Corpora, alignment, and bias. *Frontiers of Computer Science, 19*(11), 1911362. https://doi.org/10.1007/s11704-024-40579-4