Research Article

Comparative Analysis of Gender Bias in Text-Based and Audio-Based NLP Models: Insights from Asian Linguistic and Cultural Contexts

Anika Pallapothu

The Harker School, San Jose, CA, 95129, USA

ABSTRACT

This study examines gender biases in Natural Language Processing (NLP) models, focusing on text-based and audio-based systems within Asian linguistic and cultural contexts. It highlights how gender roles and cultural norms in Asian backgrounds influence these biases, using examples like Google Translate, Siri, and Alexa. The research focuses on analyzing datasets that reflect Asian languages and cultural norms, examining how gender roles, stereotypes, and historical patterns manifest in NLP models. The study employed comprehensive strategies, including analyzing word embeddings and model outputs. This helps identify stereotypes linking gender to certain professional traits, particularly in text-based models. It also examines the performance of audio-based NLP models in speech recognition, voice commands, and interpretation, highlighting accuracy issues, especially for profiles that deviate from standard demographics in the training data. The study analyzes word embeddings and model outputs to identify gender-related stereotypes in professional traits, highlighting persistent biases, especially in speech recognition models with lower accuracy for non-standard demographics. The findings suggest the need for strategies to curb biases and ensure equitable NLP outcomes that promote inclusion and diversity among users. The research is vital for NLP developers, scholars, and AI teams, as it explores text- and audio-based models, revealing findings that help reduce biases and promote equity in AI language systems.

Keywords: Artificial Intelligence; Natural Language Processing; Bias; Deep Learning, Asian Linguistic and Cultural Context

INTRODUCTION

In Asian cultures, language patterns and gender roles are deeply intertwined, frequently focusing on a culture-rooted system of social hierarchy (1). The following study aims to explore the unique ways gender biases are reflected in NLP systems within the bounds of Asian languages or linguistic pragmatics. Text-based NLP structures, responsible for processing languages like Mandarin, Hindi, Japanese, and Korean, often adopt gender-specific traits that are deeply stereotyped within the training data. Similarly, audio-based systems encounter challenges due to the tonal and pitch variations that differ across genders in these languages. This research seeks to identify,

Corresponding author: Anika Pallapothu, E-mail: anikap3456@gmail.com.

Received September 17, 2024**; Accepted** October 18, 2024
https://doi.org/10.70251/HYJR2348.23161171

compare, and propose strategies to counteract these biases in Asian NLP applications. It requires understanding several concepts related to gender bias, text-based NLP models, and audio-based NLP. Gender bias refers to the inherent behaviors of NLP models based on their training data or algorithms, which show biased or discriminatory tendencies across genders. Text-based NLP models work with and analyze textual data, such as written documents, social media posts, or transcripts. Audio-based NLP models handle and analyze audio data (e.g., speech recognition, voice assistants, or converting audio to text).

Many strategies and solutions exist to curb gender bias in these voice-based products, but similar biases also appear with text-based NLP systems. This research explains each plan in detail, describing the prejudice, the problem it seeks to combat, its implementation attempts, and the lessons learned from these implementations. The absence of varied datasets often results in systems that underperform for genders needing more representation, as seen in voice recognition and text generation objectives. Organizations such as Google and Amazon have begun acquiring larger sound datasets that represent different genders, accents, and intonations of discourse.

For instance, by adding more female voice data, Google improved accuracy by 13% for women's speech. Researchers tried to use diverse textual datasets in NLP tasks, with text as the only input. In addition, projects such as Balanced Wikipedia have created balanced corpora by searching Wikipedia article pages representing genders in various fields and activities. These implementation results provide clear evidence of improved system accuracy and fairness. However, updating and balancing these datasets as new data is a challenge. Diversity in training data requires continuous monitoring and improvement.

Specific techniques can limit inherent biases embedded in the training data and models, for example, eliminating gender stereotypes encoded in word embeddings and reducing the reinforcement of acoustic model gender biases. Mitigating these biases would involve adopting techniques like Gender Bias Fine-Tuning (GBFT) and Counterfactual Data Augmentation (CDA) proposal (2). In GBFT model fine-tuning, the datasets are gender-neutral; in CDA, researchers generate alternative scenarios to equalize training data. Benchmark challenges such as the Winogender Schemas (3) for text-based systems and gender-balanced datasets for voice systems are in place to expose, draw attention to, and dismantle gender biases. While these techniques have yielded promising results in mitigating biases, they are reasonable. Evolving and developing new methods to counteract emerging biases is imperative. Furthermore, specific debiasing strategies (Calibrated Equalized Odds and Bias Audits) may harm model performance by introducing extra noise if not applied cautiously (4).

If not accounted for, such evaluation bias could give an incomplete picture of where gender biases interfere with NLP systems and provide, at best, partial solutions. However, constructing gender-balanced evaluation datasets can offer a more accurate assessment of the system's performance. The Gendered Pronoun Resolution (GPR) test set is an example of a recent and relevant dataset that supports the identification of metrics for text-based systems or embodiments with gender bias. Moreover, methods targeted at measuring gender bias have also been proposed to complement existing metrics for a more detailed view of system performance across genders. These enhanced evaluation techniques have exposed significant hidden biases. Widespread adoption of these techniques could result in more equitable and interpretable evaluations of NLP systems. However, ensuring these evaluation methods are widely adopted within the industry is a United States research challenge.

The absence of diverse viewpoints in the various development and testing phases can lead to biased outcomes. Having experts familiar with gender and sociolinguistics as part of these teams has led to more informed ways of targeting sexualized language. For instance, organizations like the Algorithmic Justice League push for inclusivity in AI development. To gather more diverse voice datasets, platforms such as Mozilla's Common Voice call on various users to help. Involving a broad set of stakeholders has substantially increased fairness in NLP systems and expanded the biases we pay attention to and address. However, entrenching these ideas into standard increased workflows encompasses structural, cultural, and bureaucratic alterations.

The research determines whether gender bias exists in different shifts within the models as it evaluates the underlying causes and the factors that culminate in these differences. Additionally, the research explores how other users and those who observe the language understand the existing biases and whether their perceptions differ or align across all modalities. Finally, the research seeks to provide insights that help enhance the development of fair NLP models.

The research introduction presents the onset and background sections, which summarize gender biases, how they manifest in different systems, and the importance of addressing them. Additionally, the section highlights the extensive analysis of the critical roles of NLP in various

contexts and text analysis tools (5). The second section is the literature review, which explores research on the topic. This section substantiates how different models exemplify and reflect gendered information, creating an avenue for a more comprehensive examination of biases existing across modalities.

The methodology segment examines the tactics used to evaluate gender bias in text- and audio-based NLP simulations and how users' insights into these biases were assessed. Data collection included various previous datasets, with attention to high and low gender balance, which was further analyzed, and dormancy ranks were embedded in the model's bias output. Additionally, controlled examinations and surveys evaluated how users experience and recognize bias in different NLP applications.

In the results and analysis segment, the research presents the outcome of the assessments, providing a well-elaborated analysis of gender bias present in text and audio modalities. The most significant differences between various NLP models emerge through quantitative outcomes and user feedback on the cases tested, revealing how gender bias can manifest.

The discussion segment elucidates these findings and their relevance to NLP progress and deployments. Additionally, the section considers the considerable impact of gender bias on user trust and the ethical issues involved in this field of research. Finally, the segment elaborates on imminent measures to mitigate gender biases through potential training datasets, debiasing strategies, and enhanced assessments. In section six, the conclusion summarizes the research findings and encourages extensive work to minimize gender bias in NLP models. Furthermore, it highlights additional avenues for exploration, including further debiasing strategies and continued improvements to assessment metrics for NLP structures in the future to mitigate bias.

METHODS

This is a comparative analysis of gender biases in Asian languages and their Western counterparts within text—and audio-based NLP systems. The datasets are curated to include Asian languages with distinct linguistic markers for gender. Evaluation metrics, such as pronoun resolution, acoustic modeling accuracy, and word embedding biases, are assessed for cultural relevance.

Biases In NLP

Natural language processing (NLP) models display their biases differently—gender, race, social ID, ecological, and dialectal biases. Most datasets are responsible for these biases in NLP models. Such datasets often include data with misconceived proportions and biases, biases of annotators, or a lack of suitable diversity in the data-amassing processes. Greater society benefits when these disparities are resolved because NLP technologies can and will perform reasonably for all populations instead of further promoting existing inequalities, prejudices, and stereotypes.

In machine learning, bias refers to a setting in which a machine learning model is predisposed towards either group fairness or equalized odds regarding the demography within focus (6). These biases can arise from segregating factors such as biased training data sources, algorithmic flaws, and even societal factors ignored, resulting in skewed output or decisions. This concept is a general definition of bias so that this paper will define biased datasets. Bias in datasets can affect the building of statistical models in a way that they will focus on linguistic patterns or features specific to the demographics that dominate the data used to train the model (7). Despite the advantages of deep learning methods, when trained with biased data, the examples in the dataset that are not part of the norm group perform poorly (7).

Based on these findings, it is evident that these models tend to fall short in terms of generalization for the linguistic aspects or subtleties concerning many groups that are underrepresented or marginalized, thus increasing and exacerbating societal prejudices and disparities. A common problem is the frequent need for heterogeneity among the researchers and annotators of datasets, which results in data that is more homogenous than diverse (8). Due to the contributing personnel aiming at such, there is an increased risk for the datasets resulting from such processes to be one-sided regarding the diversity of views and characteristics. Thus, such a shortcoming in the data 'producers' not only nurtures the prejudices that the data compilers have but also limits the range of the sample corresponding to various populations, cultures, and lived experiences. Recent studies have shown that most of the existing natural language processing datasets have an inherent bias and include socio-demographic bias since these datasets were primarily created by people from similar populations (8). This pattern, known as the homogenous crowd-sourcing database phenomenon, occurs when all data collectors or annotators belong to similar groups or have the same experience levels. Such datasets are, therefore, usually limited in presenting only the linguistic uses of certain demographic groups rather than the full range of language diversity and use, factors

that distort and limit the scope of intervention frameworks trained on such datasets.

In most cases, the datasets employed within language model training and those of any natural language processing system tend to exhibit gender imbalance. This concept is particularly true because there are more references to men and men-related issues than women and women-related matters (9). In other words, most NLP systems have male-oriented training data, primarily using male entities, pronouns, and gender-associated words, while female counterparts are disadvantaged. This gender documentary bias leads to the models slowly replicating some cultural stereotypes of society, which can be worrisome, especially when they relate to gender modeling.

Overcoming this imbalance and lessening these potentially harmful gender stereotypes would mean that researchers attempt interventions such as data augmentation, including dataset modification. At one point, all male pronouns were replaced with female ones as a modification of the collection. Applying large language models, including GPT-2, to resume screening and job recommendation will further aggravate the issue of gender bias in recruiting procedures, as some gender bias may be retained by the models (10). Language models trained on large text corpora may also tend to generate more positive statements or associations when prompted with conservative political figures or ideologies than their treatment of liberal counterparts (11). That is because the training datasets contain prejudices inherent to society or the political leanings of the people who created them.

The findings indicate that Asian languages introduce unique challenges for NLP models due to gendered language constructs, such as honorifics and context-dependent pronoun usage in languages like Japanese and Korean. In audio-based systems, the higher pitch variance between male and female speakers in tonal languages like Mandarin leads to higher error rates for women (12) (Table 1). Additionally, cultural biases embedded in the datasets, such as male-dominated professional terms in Hindi or gendered familial roles in Mandarin, are replicated in NLP outputs, perpetuating societal stereotypes (6).

Gender Biases in Text-Based NLP

The bias that exists in cultural, racial, and gender forms part of the everyday world, but some of it is embedded into our technological future. NLP systems, such as virtual assistants like Siri and Alexa, have all been designed with female names, voices, and personalities conditioned to be submissive or flirtatious, even in response to statements that would otherwise consider inappropriate or objectifying if raised but not against a male user (13) (Table 2). All efforts made by organizations like Apple and Amazon to mitigate this have found themselves in lateral responses, such as stating, "I do not know how to respond to that," but are still directly bolstering the stereotypes through their respective gender bias from the engineers building these systems (14).

The overwhelming number of male engineers working on the design teams for many virtual assistants and chatbots has perpetuated valid concerns that some systems are designed to express subservient, hyper-

Table 1. Gendered Language in Asian Cultures (24)

Language	Gendered Forms	Cultural Context	Common Bias Manifestations in NLP
Japanese	Use of "watashi" (neutral), "boku" (male), "atashi" (female)	Gendered pronouns and hierarchical speech	Gender misclassification in translation
Hindi	Masculine and feminine nouns for professions	Gender roles are highly reinforced in media and literature	Gender bias in Job Title Translations
Mandarin	Gendered pronouns (他 vs. 她)	Pronouns do not always reflect neutral roles	Gender-specific sentence generation in machine translation

Table 2. Speech Recognition Accuracy by Gender in Voice Assistants (Siri/Alexa)

Gender	Accuracy (%)	Error Rate (%)	Typical Errors
Male	89%	11%	Misinterpretation of accents
Female	74%	26%	Higher-pitched voices not recognized properly

sexualized female personas (13). This move proves that there needs to be more diversity in NLP technologies, and this sounds like an unsustainable solution if we are looking for deeper-than-surface solutions that will address the core of biases structuring many aspects of society today. The fact that female voices and personas are so often used as a default design choice for many virtual assistants and conversational AI systems should prompt the consideration of whether they may be reinforcing gender stereotypes or extending these societal biases (15). The depiction of AI agents presenting dutiful women may reinforce the notion that women's ultimate role is other-oriented. It contributes to normalizing existing social imbalances and gender prejudices (15).

Gender biases within AI systems are a significant challenge. In addition, systemic bias is evident in various AI systems, including virtual assistants associated with different organizations. The systemic bias becomes further apparent when women are, in most instances, related to caring roles, reinforcing traditional gender stereotypes. Society portrays women as submissive, which better suits the existing stereotype while influencing how users view gender roles through default choices. Users should be able to select neutral gender options to curb gender bias. Besides, consideration of diverse voices is imminent for diversity purposes. If not, diversity would amount to nothing if it could disrupt the strata while boosting inclusion by reflecting its criticality in society. These selections also extensively counter gender stereotypes that create a connection between individuals and the traditional roles they uphold.

Unless the biases get a proper address and mitigated (10), there lies the risk of fostering old-fashioned and exploitative gender norms (Table 4). The bias can foster stereotypes in society. For instance, if a virtual assistant coins and adopts a feminine voice when performing the caregiving administrative task, it will indicate that the activities around caregiving are on the females or are inherent. That, in turn, can belittle the female's contributions in other areas, like leadership, science, and technology, restricting how females can have admirable values or additional defined not in the opposing terms but the mutual respect that cuts across all roles. Besides, the need to escalate the representation should uphold women's contribution within the bounds of traditional roles. It promulgates a strict view of women's roles and their abilities that contradicts gender equity and equality in all areas. There is a need to support a diverse and equitable society to ensure that AI systems cut across a broader spectrum of different persona voices. That will enable individuals to define their roles and contribute not based on the constraints of the degenerating gender norms.

The training of the model is where bias originates. Many datasets come from a relatively small, homogenous group of people, and these models train sentence-level data to which annotators have assigned labels. Individuals annotating datasets influence their narratives; annotation bias is possible because they may not consider how different demographic communities express their lived

Table 3. Summary of sex biases' comparisons and variances transversely NLP categories

Gender Bias	Voice Based NLP	Text-Based NLP
Pronoun Bias	√	√
Data Bias	√	√
Evaluation Bias	√	√
Voice Recognition Bias	×	×
Acoustic Modeling Bias	×	×
Word Embeddings Bias	×	×
Language Generation Bias	×	×

(√ - Similar and × - Different).

Table 4. Cultural Gender Norms and NLP Output in Asian Languages

Language	Common Stereotypes Encoded in NLP	Examples of Bias in NLP Models
Hindi	Men in leadership, women in caregiving roles	Male pronouns in translations of "teacher" and "leader"
Japanese	Women in supportive roles (e.g., assistants)	Stereotypes in sentence completion tasks using Google Predictive Text
Korean	Women as caregivers, men as professionals	Gendered language in sentence generation involving family roles

experiences through language and culture (16). However critical the selection process is, annotators can fall into the same bias traps as the participants they are trying to quantify, based on the idea that they may have limited knowledge of other demographic groups and a narrow interest in certain language categories (17). This situation is described as "annotation bias." It may result in datasets not aligning with the proper use of language across different demographics, ultimately leading to models that promote harmful stereotypes, especially among underrepresented communities.

In datasets representing gender classification, various cases involve men compared to those targeting women (5). However, it results in a built-in bias that needs effective addressing in translational NLP tasks. Translating subjects from objects, however, attracts many errors, especially in NLP tasks. According to recent studies, these biases lead to significant critiques of the efficacy and fairness of NLP models. For example, in most scenarios, NLP systems are associated with an imbalanced ratio of gender references in datasets, leading to biased translations (5). This research investigates how bias within the training datasets can reduce the performance and fairness of NLP implementations, proving the need to address human biases from various perspectives effectively.

Organizations can limit the amount of gender bias present within LLM practices by considering benchmark metrics and bias fine-tuning (5). This concept includes the integration of datasets and layered networks with various tasks and samples, enabling us to assess biases within LLMs. In an academic study, Drawing pointed out how important these strategies are in reducing gender bias in LLMs. However, their research describes how bias fine-tuning methods and benchmark metrics can promote fairness without compromising performance in language generation systems (5).

Gender Biases in Voice-Based NLP

Voices are a critical factor in shaping gender biases in voice-based NLP. Holzmann reports that the engines perform even worse for females due to a highly unbalanced dataset, where 90% of the voices are male. Excluding children, females who society deems to have a higher pitch which has impairment in voice recognition. In the text and voice-based NLP models, one can say that the issue emanates from socio-cultural stereotypical thinking and the prevalent underrepresentation of other groups in the training data (1). Modern NLP models tend to link women to family and domestic roles and, conversely, to connect men with career and leadership (1). However, such structural concerns embedded in the data samples have designs to train models and learn by fine-tuning them, leading to insensitive and discriminating data acquisition (18). To resolve this setback, it is crucial to implement debiased algorithms, develop diversified sets that conflict with prevailing biases, and promote cultural understandings of genuine egalitarianism throughout the resulting model performances (13).

Many popular virtual assistants and chatbots exhibit feminine traits, including female voices, names, and anthropomorphized characteristics aligning with traditional gender stereotypes. Siri, Alexa, Cortana, and other virtual assistants and chatbots are frequently ascribed feminine names, voices, and personae—smiling, service-oriented digital secretaries that reinforce gendered stereotypes about administrative labor being quintessentially feminine (13). This design choice perpetuates the objectification of female identities within AI systems, highlighting the prevalence of gender biases in these technologies. Voice-based natural language processing models frequently exhibit suboptimal performance when analyzing audio inputs from speakers exhibiting higher-pitched or breathier vocal characteristics, which are more commonly associated with female voices (19). This deficiency is attributed to inadequate representation and diversity in the training data for developing the acoustic models that underpin these systems. Based on the observations made, there were higher error rates for female speakers than their male counterparts. This concept is, however, well explained within the well-known but continuous under-representation of female speakers within primary speech datasets utilized in training acoustic models (19). Such discrimination is, however, demeaning since it reduces female gender performance and calls for an inclusive move and diverse data training to solve gender bias issues while ensuring equality in performance in voice-based NLP applications.

Most of the algorithms utilized for audio analysis and speech recognition experienced their training on corpora. This technique either oversampled male voices or could have been more successful in obtaining the variability usually observed in female speech patterns (20). Based on this, models experience difficulties adequately processing and transcribing audio inputs, deviating from the predominantly male-centric data they engage with while training. This challenge, however, reveals the essential need for curating inclusive and representative datasets to address gender biases while achieving equality in performance among different demographics in voice-

based NLP applications. Based on the research findings presented by (21), a significant percentage of the data utilized in speech recognition models is based on the voices of Caucasian males. The given training data negatively affect the model's ability to accurately and critically evaluate the gender of speakers with different vocal pitches or dialects that do not align with the predominant Caucasian male voice assessed during training.

Alternatively, because the speech recognition models have trained on data from Caucasian male voices, there is great difficulty in determining speakers' gender with different traits, including higher or lower pitches in their voices. In addition, they face challenges distinguishing the gender of speakers with dialects that differ from the standard voices in the training data. This challenge originates from inadequate and poor diversity that is either stereotyped or discriminatory against a particular demographic (Caucasian males), leading to the models experiencing significant difficulty in generalizing and accurately recognizing gender for a wide range of voices.

Machine learning techniques have been in place in the last ten years, particularly for sex identification by speech. These models influence sex identification while focusing on specific features, such as the length of the vocal cords, pitch, and speech patterns. Additionally, characteristics of voices and speech, including size, intensity, and frequency, can also be used to identify the speaker's gender.

Dr. Bisio established an Android speech processing platform as a smartphone app. This platform focused on gender, voice speaking, and language recognition using different unsupervised support vector machine classifiers (22). This research presents a point worth exploring and evaluating further: the dynamic training linked to the traits obtained from unique users with SPECTRA installed on their digital gadgets, such as computers and smartphones. This move is essential since it enables the establishment of very effective classifiers with increased accuracy rates, promoting recognition operations. According to Dr. Pahwa, it is necessary to establish a recognition system that utilizes speech samples from 46 speakers to determine gender. To demonstrate the existence of such a concept, the researchers investigated one of the most basic and highly rated features of a speech signal, known as Mel coefficients, and first and second-order byproducts (23). Their planned model comprises a support vector machine and a neural network classifier. This classifier follows a loading procedure approach. The level of organization precision calculated from the arithmetic tests was 93.48%.

Achieving equitable presentation and empowerment of women across different domains is very challenging due to the propagation associated with gender bias, especially in artificial intelligence and, in most instances, NLP (1). These biases spread faster, further promoting detrimental stereotypes concerning the roles of women in society and their general contributions, thereby undermining their potential in various fields, including politics. Following the outcomes, it is evident that the modern NLP model conjoins women to the family and homestead roles more than men. Also, males are more associated with leadership roles than women in society. The findings indicate that the word's model association basing women comes out as more offensive than those of the male counterpart (1). These outcomes reveal that NL models, such as decision-making processes and political perspectives, demean women's societal roles and contributions.

RESULTS AND DISCUSSION

This segment examines how these outcomes affect the development of Asia's culturally adaptive NLP structures. Gender bias is evident in Asian languages due to cultural factors such as undying reverence for elders, hierarchical communication, and socially constructed gender roles. Additionally, the research emphasizes the importance of including Asian datasets that address cultural differences to efficiently mitigate gender biases. Strategies like developing gender-neutral language corpora for Asian languages and applying culturally sensitive information augmentation are highly preferred.

Comparison of Voice-Based NLP and Text-Based NLP from a Gender Bias Perspective

Gender partiality in Natural Linguistic Processing (NLP) structures is precarious in both voice and text-founded frameworks. The subject demands noteworthy attention and tactical involvement.

Table 3 heightens a wide-ranging analysis of the contrasts and variances in how sexual category bias is present in the two categories of NLP classifications.

Similarities

Pronoun Bias. Misreading of pronouns is apparent in both voice-grounded and text-founded NLP structures since both often pass to mannish pronouns and presume gender grounded on present stereotypes. This hypothesis stems from the fact that processes primarily hinge on antique data and language patterns embedded in societal prejudices. For instance, a text-grounded system could use "he" when pointing to a doctor and "she" when referring to a nurse, while a voice-based structure might misrecognize

a feminine voice and apply improper pronouns.

Data Bias. The training information for both voice- and text-grounded NLP structures frequently has traits of gender bias. These biases heighten societal and antique judgments, and the different structures apply these biases. For instance, if the males' voices are used to feature in the data training and are associated with particular gender roles or traits, then the NLP structure will reflect a comparable bias.

Evaluation Bias. Performance assessment approaches for NLP systems have biases. In some cases, assessment metrics and examined datasets only partly hook the level of sexual category bias in these classifications. This situation implies that understanding gender biases within these systems is incomplete or inaccurate. For instance, system performance on gender bias can undergo skew if the evaluation dataset is not gender-diverse (25).

Differences

Voice Recognition Bias. Voice-based NLP systems are associated with increased errors, especially in speech recognition and transcription involving underrepresented women. This issue originates from a lack of diverse voice data and a historical preference for male voices in development. For instance, higher error rates are standard in speech recognition systems targeting female voices because these systems were primarily trained on male voices, leading to lower rates of accurate transcription and understanding. Text-based NLP systems do not process voice input and are not associated with this bias.

Acoustic Modeling Bias. Specific speech patterns and gender-specific biases found within acoustic models and voice-based NLP systems. Such biases, however, reduce the accuracy of voices that deviate from these patterns. This challenge originates from the lack of diverse voice data. For example, pitch and tone can lead to misinterpretations. Besides, the variation in communication styles between the males and females poses a threat if the models do not adequately have training on various voices. Text-based NLP classifications do not hold such bias when they approach text input.

Word Embeddings Bias

The problem of gender partialities fueled by word entrenching from teaching data is challenging for several motives:

Reinforcing Damaging Stereotypes: aligning together certain traits and particular gender brings forth inaccurate and deluded stereotypes. For instance, associating a doctor with a "he" and a nurse with a "she" creates the idea that specific professional roles are particular to certain genders. The nation could have adverse effects since it could discourage one from pursuing a profession that best suits them due to the existing stereotype.

Conserving Gender Judgment: Such biases create discrimination. An example could be a job scenario that associates an "engineer" with a "he" more than it could with a "she." When that happens, the male candidates will receive priority, putting the women in a disadvantaged position. However, they, too, would have similar qualifications, thus promoting gender discrimination in different fields.

Educating Existing Opportunities and Underpinning General Biases: Language promotion to gender stereotypes reduces opportunities for people who do not conform to them. In many instances, this can lead to disinterest and a change in career path.

Invalid Depiction of Reality: The union that arises from the grounding must-have traits of the diversified natural world. However, specific stereotypes could have origins in history and fail to effectively credit the current realities, specifically those that exist in the professional pragmatics that infuriate all genders with various traits.

Dangers of Unintentional Harm: Errors in the linguistic models create harmful repercussions, whether intended or not. For instance, if the language used in a text promotes gender stereotypes, then it can carelessly promote harmful bias. The result could be an alteration in the user's thoughts and behaviors.

Overcoming this problem would require scholars to investigate potential biases extensively, especially those in data training, and ascertain the best mitigation strategies. Such strategies involve debiasing word embedding, using diverse data training representation, and creating safeguards that prevent the propagation of destructive stereotypes within the applications.

Language Generation Bias

NLP structures can mirror gender biases in the training data when creating texts, which can result in the spread of destructive gender stereotypes. For instance, an outdated text-generated system might inculcate content that focuses on gender norms, like liking leadership roles to the gender of men and caregiving duties to have a representation of women. The issue here arises from the biases encoded in the data training information (26).

Initiating plans or policies to align with these similarities and differences ensures a better understanding of the particular challenges and highlights the need to establish an equal NLP system. However, achieving such

needs requires a diverse approach, including diversifying the data training and bias mitigation strategies and establishing more elaborate analysis metrics that approve gender diversity.

Mitigation Strategies

Various involvements have been projected and executed in specific instances to deal with bias problems within voice-grounded and text-based NLP classifications. The study comprehensively analyzes each approach, figure-hugging the bias problems it discourses, the execution attempts, and the related teachings.

Collect and Formulate Diverse Evidence: Negligible diversity creates structures that disfavor the underrepresented sex categories, mainly when featuring voice recognition and text cohort tasks. Firms like Google and Amazon have developed ways of gathering elaborate and diverse voice datasets that feature both males and females in accent and speech patterns. For instance, Google elevated speech accuracy and recognized women by thirteen percent. The achievement was because the firm expanded the datasets that accommodated more female voices. On the other hand, scholars have continued to research the use of variance in actual datasets within the bounds of text-based NLP. Besides, projects like "Balanced Wikipedia" have started a grounded corpus by emphasizing equity in gender representation in the different professional disciplines. The improvements have brought immense results and a much fairer system. However, there is an ongoing task of updating and balancing these datasets as the new data ensues. Hence, ensuring diversity in the training data would need continuous monitoring and adjustment where necessary (28).

Utilization of Methods that Eradicate Biases

Specific strategies can eradicate biases embedded in the training data models, including gender stereotypes in word and acoustic models. Gender Bias Fine-Tuning (GBFT) and Counterfactual Data Augmentation (CDA) are a few strategies that have been applied to mitigate these biases (15). GBFT incorporates fine-tuned modeling with datasets designed to promote gender equity, while the CDA culminates an approach that promotes balance in the training data. Besides, promulgating benchmark issues like Winogender schemas grounded on voice and text-based systems have surmounted a crucial role in identifying and countering gender bias. These techniques are effective since they eliminate gender biases but are not credible solutions. Hence, there is an imminent need to create additional methods to mitigate the cropping biases continuously. Besides, debiasing stagy could also reduce the performance if it does not receive proper implementation, requiring caution.

Evolving More Wide-ranging Assessment Methods: it is pretty challenging to solve gender biases in the NLP systems when incorporating inadequate assessment techniques. Hence, it may result in misinterpretations of the gender effect biases on the system, leading to improper resolves. It is important to note that creating an evaluation dataset that caters to both males and females breeds room for accurate measures of performance systems. For instance, dataset developments like the Gendered Pronoun Resolution (GPR) have keyed in assessing and addressing gender biases in text-based systems. Besides, the Gender Bias Impact Index (GBII) focuses on gender bias and is operational in determining how the system performs differently following the different genders. Hence, an advanced and elaborate assessment method reveals the biases that previously had not come to light, and the effective utilization of these methods leads to a fair assessment of NLP structures. However, these assessment strategies still need to be implemented.

Integrating Diverse Shareholders: There needs to be more perspective. Specifically, the development and testing results contribute to biased outcomes. Diverse opinions from various experts help foster a broader and more informed approach to solving gender biases. For instance, the Algorithmic Justice League is a firm that advocates for inclusiveness and diversity in AI representations. However, Mozilla Common Voice only provides diversity within the bounds of its voice datasets. Also, including a diverse group of stakeholders has proven a considerable improvement in gender fairness in NLP systems. The technique is effective since it addresses the existing issues effectively. However, there are significant issues in integrating the practices into moderate activities since cultural and procedure traits need proper integration.

CONCLUSION

To summarize, biases in gender are prevalent in NLP models in almost all Asian languages because of the cultural norms and the linguistic traits that encourage the stereotype. The proposed techniques, like data augmentation, gender-neutral corpora, and the ongoing bias metric assessment, are critical for reinstating more equitable and inclusive NLP systems to address the various populations found in Asia. Countering gender biases in Asia is essential for creating an equitable AI system serving a diversified population. Therefore, the study

heightens the need for continued revamping and training data assessment methods that focus on the linguistic and cultural intricacies in Asian societies. Hence, future examinations should focus on the development models that cater to gender fairness and sensitivity to cultural norms and look at the values in the languages they process. The average assessment of gender biases contained by text-based and audio-based NLP indicates a few discriminations that the underlying data revealed. Also, model training and cultural stereotypes promote these biases. The text-based NLP model has biases in various ways that reflect historical and societal stereotypes. The biases incorporated in the stance are evident in data training that systems like Google Translate and Compas use, and the propagation of gender stereotypes emanates from these biases. Further, the overrepresentation of male voices and the inadmissible diversity in the dataset training promotes biases in the audio-based NLP models, including the fields of virtual assistance like Siri and Alexa.

However, the two types of NLP classifications establish that wide-ranging strategies are necessary to counter the dangers and challenges they bring. The most suitable strategy includes assembling and adopting comprehensive training data standpoints. Wide-ranging debiasing methods and the establishment of accommodative assessment metrics are relevant. Furthermore, diverse experts should be on board to ensure equality. NLP systems can only become just and equitable if biases receive proper address. Such a resolution will effectively tackle the biases and ensure the utilization of technology that eliminates them. Enhancing the reliability and inclusivity of NLP models becomes a reality when these challenges are mitigated, creating a fair environment.

This research, however, paves the way for exciting topics worth exploring in the future, including the utilization of debiasing approaches for large language models to eliminate political biases. Additionally, boosting gender recognition accuracy from voice/speech data while creating and developing bias evaluation benchmarks and fine-tuning strategies to address biased language generation models is critical.

REFERENCES

1. Gendered Language and Social Hierarchy in Asia: Wang, L. et al., "The Gendered Language of Asian Languages: Gender Norms and Honorifics," Journal of Linguistic Anthropology (2021). https://www.researchgate.net/profile/Kira-Hall-4/publication/341256069_Language_and_Gender/links/5fd50b2b299bf140880650d9/Language-and-Gender.pdf
2. Zhao J, Wang T, Yatskar M, Cotterell R, Ordonez V & Chang K. (2020). Gender Bias in Contextualized Word Embeddings. arXiv preprint arXiv:2005.04612. https://doi.org/10.18653/v1/2020.acl-main.405
3. Adda-Decker M & Lamel L. (2018). Implicit and Explicit Representations for Gender Bias in Speech Recognition. arXiv preprint arXiv:1808.07010. https://perso.limsi.fr/madda/publications/PDF/AcademicPressMSPch5.pdf
4. Zhao J, Wang T, Yatskar M, Ordonez V & Chang KW. (2018). Gender bias in coreference resolution: Evaluation and debiasing methods. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers) (pp. 15-20). https://doi.org/10.18653/v1/N18-2003
5. Gendered Language and Machine Translation: Zhao, J. et al., "Gender Bias in Contextualized Word Embeddings," Proceedings of the ACL (2020). https://arxiv.org/abs/1904.03310
6. Bias in Mandarin Speech Recognition: Pahwa A, et al., "Gender Bias in Mandarin ASR Models: A Case Study on Alexa," Proceedings of the Annual Speech and Language Technologies Conference (2023). https://www.techrxiv.org/doi/pdf/10.36227/techrxiv.171595948.84728317
7. Bisio et al. (2022). Developed an Android speech processing platform as a smartphone application (SPECTRA) for gender, speaker, and language recognition by utilizing multiple unsupervised support vector machine classifiers. https://www.mdpi.com/2504-4990/1/1/30#B4-make-01-00030
8. Blodgett SL, Barocas S, Daumé III, H & Wallach H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. arXiv preprint arXiv:2005.14050. https://doi.org/10.48550/arXiv.2005.14050
9. Brahnam S & Weiss S. Gender biases in virtual agents: When good intentions go awry. ACM Transactions on Human-Robot Interaction (THRI). 2022; 11 (2): 1-25. https://link.springer.com/article/10.1007/s11948-022-00376-3
10. CDA and Gender Bias Mitigation: Huang, P. et al., "Gender Bias in Language Models: Evaluations and Debiasing Methods," arXiv preprint arXiv:2011.03198 (2020). https://arxiv.org/pdf/1911.03064
11. Channarong Intahchomphoo, Kathleen D. Ells and J. Scott Moore. Artificial Intelligence and Race: A Systematic Review. https://www.cambridge.org/core/journals/legal-information-management/article/abs/artificial-intelligence-and-race-a-systematic-review/E3EC8D1771D76E68E26DAB73F81128A6
12. Sato, Yutaka, and Sungdai Cho, 'Honorifics', The Comparative Syntax of Korean and Japanese (Oxford, 2023; online edn, Oxford Academic, 18 Jan. 2024), https://doi.

org/10.1093/oso/9780198896463.003.0012 (accessed 2024-07-26)

13. Cultural Bias in Language Models: Mehrabi, N. et al., "A Survey on Bias and Fairness in Machine Learning," ACM Computing Surveys (2021). https://dl.acm.org/doi/abs/10.1145/3457607
14. Sap M, Card D, Gabriel S, Choi Y & Smith NA. (2019). The risk of racial bias in hate speech detection. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1668-1678). https://doi.org/10.18653/v1/P19-1163
15. Fine-tuning and Debiasing NLP Systems: Zhao J, et al., "Debiasing Gender in Word Embeddings with Fine-Tuning Techniques," Proceedings of NAACL-HLT (2019). https://ojs.aaai.org/index.php/AAAI/article/view/6267/6123
16. Sap M, Card D, Gabriel S, Choi Y & Smith NA. (2019). The risk of racial bias in hate speech detection. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1668-1678). https://doi.org/10.18653/v1/P19-1163
17. Garnerin M, Rossato S & Besacier L. (2021). Gender representation in French broadcast corpus portraying gender bias in ASR performance. arXiv preprint arXiv:2105.04326. https://www.semanticscholar.org/paper/Gender-Representation-in-French-Broadcast-Corpora-Garnerin-Rossato/94c1d498142e4aa8d0e0c7f3e870a85a2a683e57
18. Voice Pitch and Speech Recognition Performance: Sap M, et al., "The Effects of Speaker Gender and Pitch on Speech Recognition Accuracy," Interspeech Proceedings (2019). https://www.pnas.org/doi/abs/10.1073/pnas.1915768117
19. Geva M, Goldberg Y & Berant J. (2019). Are we modeling the task or the annotator? An investigation of annotator bias in natural language understanding datasets. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 1161-1166). https://doi.org/10.18653/v1/D19-1107
20. Hovy D & Prabhumoye S. (2021). What lies ahead: Survey the biases inherent in the pre-trained models used to build language Technology. arXiv preprint arXiv:2110.07495. https://compass.onlinelibrary.wiley.com/doi/full/10.1111/lnc3.12432
21. Huang PY, Lassner C, Metcalf J, Raghunathan A., Schubert L, Zou J & Zhang Z. (2020). Gender bias in language models: Evaluations and debiasing methods. arXiv preprint arXiv:2011.03198. https://aclanthology.org/P19-1159/
22. Mehrabi N, Morstatter F, Saxena N, Lerman K & Galstyan A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35. https://doi.org/10.1145/3457607
23. Pahwa A, Sharma R, Gupta A & Singh V. (2023). Gender Recognition Using Speech Features: A Study on Mel-Frequency Cepstral Coefficients and Derivatives. https://doi.org/10.18653/v1/2023.emnlp-main.165
24. Mozilla Foundation. Mozilla Common Voice. Mozilla, https://commonvoice.mozilla.org (accessed on 2024-05-21)
25. Speech Recognition Bias in NLP Systems: Holzmann, L. et al., "Bias in Voice-Based Virtual Assistants: How Gender and Language Impact Performance," Speech Communication Journal (2022). https://www.sciencedirect.com/science/article/pii/B9780128213926000091
26. Wang Y, Zhang X, Smith NA & Makhija D. (2023). Mitigating Gender Bias in Language Generation Models: A Comprehensive Study. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). https://doi.org/10.18653/v1/2023.emnlp-main.165
27. Speech Recognition Error Rates by Gender: Garnerin M, et al., "Gender Representation in French Broadcast Corpus: Portraying Gender Bias in ASR Performance," arXiv preprint arXiv: 2105.04326 (2021). https://hal.science/hal-04607587/document