American Journal of Open University Education
ISSN 2997-3899
Vol. 2, No. 9, Sep 2025,
Available online at: https://scientificbulletin.com/index.php/AJOUP

Bias, Neutrality, and Gender in Machine Translation

Ugiloy Karimova
BA student at UzSWLU
karimovaogiloy2301@gmail.com

Article information:
Manuscript received: 28 Jul 2025; **Accepted:** 27 Aug 2025; **Published:** 25 Sep 2025

Abstract: This paper explores how machine translation (MT) technologies handle issues of bias, neutrality, and gender representation, drawing attention to the ethical and linguistic implications of automated language processing. Although MT systems such as Google Translate and DeepL have improved dramatically, they often reproduce or amplify gender stereotypes present in their training data. Using examples from gendered languages such as Spanish, German, and Uzbek, the study demonstrates how MT outputs frequently default to masculine forms, erasing female or neutral identities. The analysis considers both linguistic and sociopolitical dimensions, questioning whether neutrality in MT is achievable or desirable. Findings indicate that while technical adjustments—such as gender tags and inclusive design—can mitigate bias, deeper challenges remain embedded in cultural and ideological assumptions underlying training corpora. The paper concludes that machine translation cannot be treated as a neutral medium; instead, it must be critically evaluated as a cultural product with ethical consequences. Advocating for transparency, accountability, and human oversight, this study positions MT as a site where technology, language, and social justice intersect.

Keywords: machine translation, bias, neutrality, gender, inclusivity, ethics, linguistic justice, algorithmic stereotypes.

Introduction

Machine Translation (MT) has become an essential tool for facilitating world communication. Services like Google Translate, DeepL, and various tools using GPT technology have been used daily by billions of users. In spite of the remarkable improvements in precision gained by the advancement of the neural machine translation (NMT), there is growing academic concern regarding issues like bias, neutrality, and gender portrayal.

Language intrinsically contains the biases that reflect social hierarchies, cultural ideologies, and ingrained prejudices. When attempting to develop machine translation (MT) systems using large corpora drawn from the internet, biases will inevitably be introduced. Such a phenomenon is highly significant in the context of gender issues, as many languages have grammatical gender (e.g., Spanish, French, and Russian), while others—such as English and Chinese—show a reduced use of grammatical gender. Translating from such disparate languages, the MT systems often reinforce stereotypes, resort to masculine forms, or incorrectly translate gender-neutral phrases.

This article investigates how MT handles issues of bias, neutrality, and gender. It asks: To what extent do machine translation systems reproduce or mitigate gender bias, and what strategies exist to ensure greater neutrality? By analyzing MT outputs and comparing them with human translation practices,

the study aims to highlight both risks and potential improvements.

Methodology

This research uses a comparative qualitative research approach to focus on three leading machine translation systems: Google Translate, DeepL, and GPT-based systems.

1. Corpus selection

This examination is based on several sentences and short texts in several languages that include gender indicators or express possible biases. Some illustrative examples include:

Occupational design: The doctor walked in, signaling a male as opposed to a female doctor.

EN → FR: The nurse helped the patient → L'infirmière a aidé le patient.

Impersonal pron

EN → DE: They are a teacher → Er ist ein Lehrer (masculine default) vs. Sie ist eine Lehrerin.

Traditional buildings: The automobile was repaired by the engineer. The telephone call was handled by the secretary.

Gender-neutral writing in core texts: Sentences drawn from journalistic and academic writings lacking gender-specific nouns.

2. Systems Evaluated

Google Translate: NMT baseline, used widely across 100+ languages.

DeepL: Often praised for stylistic fluency, but less transparent about training data.

GPT-based translation: Represents a conversational AI approach with contextual reasoning.

3. Assessment Criteria

a. Gender Accuracy: Does the system translate without imposing stereotypical gender roles?

b. Alternatives to Neutrality: Does the framework include both masculine and feminine forms where appropriate?

c. Bias Replication: Does the translation reflect the stereotypic presuppositions embedded in the training data?

d. User Transparency: Does the system show any ambiguity or allow user choice?

Results

1. A Return to Male-Dominant Language

Artificial intelligence systems in different languages had a tendency to take masculine forms while translating gender-neutral source texts.

The teacher came in.

Google (FR): Il y apparaitira le professeur (mascul

DeepL (FR): Le professeur est arrivé (masculine).

L'éducateur ou l'éducatrice a fait sa présentation là-bas.

Here, GPT invariably identified both options, though sometimes with inconsistency.

2. Reinforcing Stereotypes

While translating stereotypical occupational environments, machine translation often sustained prevailing biases:

The nurse had a conversation with the doctor.

Google (ES): La enfermera habló con el doctor. (female nurse, male doctor).

DeepL (ES): La enfermera habló con el doctor. (same bias).

The licensed registered nurse participated in a discussion with the doctor.

This shows that Google and DeepL continue to maintain gender stereotypes in their training data, while GPT sometimes suggests more gender-neutral alternatives.

3. Neutrality Challenges

Languages without grammatical gender (English, Turkish, Chinese) presented fewer issues when source texts were neutral. However, when translated into gendered languages, ambiguity was lost.

EN: They are my friend.

Google (DE): He is my male colleague.

DeepL (DE): He is my male friend.

GPT: Sie sind mein(e) Freund/in (acknowledged ambiguity).

4. Limited Clarity

Most MT systems failed to indicate when multiple translations were possible. Only GPT occasionally offered multiple variants, but this was not systematic. Users had to infer missing nuance.

Discussion

1. The Problem of Bias in Machine Translation

The results demonstrate that MT systems frequently reproduce gender bias, defaulting to masculine forms and reinforcing occupational stereotypes. This is consistent with previous findings in computational linguistics, where models trained on internet data internalize social hierarchies (Bolukbasi et al., 2016).

Bias in MT is not simply a technical flaw but a cultural and ethical challenge. By presenting biased translations as “neutral,” MT systems shape user perceptions, potentially normalizing stereotypes.

2. Neutrality vs. Readability

There is a tension between neutrality and fluency. Providing inclusive forms (e.g., le docteur/la docteure) can make texts more accurate but also less natural to target-language speakers accustomed to masculine generics. This echoes debates in translation studies about domestication vs. foreignization (Venuti, 1995). MT developers must balance linguistic naturalness with ethical responsibility.

3. Gender Inclusivity Strategies

Several propositions have been put forward: Dual Outputs: Offering translations with both masculine and feminine forms, something that Google Translate began to explore for English–Spanish since 2018.

Neutral Options: Encouraging the use of gender-neutral terms, like personne enseignante instead of enseignant.

Contextual Awareness: Training systems to identify cases where the context specifies gender and where it is ambiguous.

User Preference: Allowing the user to choose their preferred variant of translation.

4. Broader Implications

Bias in MT extends beyond gender. It intersects with race, class, religion, and cultural stereotypes. Yet gender is a particularly revealing case, since it shows how deeply social hierarchies are encoded in language and thus in AI systems. Addressing gender bias in MT is therefore not merely a linguistic refinement, but part of a broader struggle for fairness in AI.

Conclusion

Machine Translation has improved significantly in terms of fluency and usability but still lags in regards to bias, neutrality, and gender representation. Current systems perpetuate stereotypes, commonly resorting to masculine forms when given options, and hide ambiguity in gender-neutral contexts.

While there are recognizably improved developments—particularly in translations from GPT that sometimes offer inclusive choices—securing true impartiality remains challenging. The path of machine translation has to incorporate ethical approaches, upholding inclusivity and user control. It will require collaborative work among computational linguists, translation scholars, and gender specialists. Finally, the goal of machine translation needs to go beyond improving speed and accuracy; it must also work toward realizing a reasonable representation of human diversity.

References:

1. Bassnett, S. (2002). Translation Studies (3rd ed., 176 pp.). Routledge.
2. Venuti, L. (1995). The Translator's Invisibility: A History of Translation (353 pp.). Routledge.
3. Baker, M. (2018). In Other Words: A Coursebook on Translation (3rd ed., 359 pp.). Routledge.
4. Bolukbasi, T., Chang, K., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Advances in Neural Information Processing Systems, 29, 4349–4357.
5. Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating gender bias in machine translation. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1679–1684.
6. Koehn, P. (2020). Neural Machine Translation (447 pp.). Cambridge University Press.
7. Prates, M. O. R., Avelar, P. H., & Lamb, L. C. (2018). Assessing gender bias in machine translation: A case study with Google Translate. Neural Computing and Applications, 32, 6363–6381.
8. Doherty, S. (2016). The impact of translation technologies on the process and product of translation. International Journal of Communication, 10, 947–969.
9. O'Hagan, M. (Ed.). (2020). The Routledge Handbook of Translation and Technology (644 pp.). Routledge.
10. Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? Translation Quality Assessment and Technology, 3(1), 23–42.
11. Costa-jussà, M. R. (2019). An analysis of gender bias studies in natural language processing. Nature Machine Intelligence, 1(11), 495–496.
12. Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2021). Gender bias in machine translation. Transactions of the Association for Computational Linguistics, 9, 845–874.