IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Gender Bias in Machine Translation

Ishika Agrawal
Student, Genesis Global School

Abstract

Machine bias in Artificial Intelligence (AI) has the detrimental potential to cause perpetual gender inequality in today's society. With AI rapidly becoming a source of gender bias in new technologies such as facial recognition and automated recruitment tools, the requirement of a fair and unbiased model has grown. Clearly, these biases stem from real-world stereotypes, showing how differently women are treated. This work sheds light on this issue with specific focus on gender bias in Machine Translation (MT). We hope this work will promote social awareness and lead to more conversations concerning machine bias.

1 Introduction

Soaring dependence on Artificial Intelligence (AI) necessitates the need to question its reliability. These algorithms are continually reflecting contentious social asymmetries, often promoting stereotypes, and discriminating against the marginalised groups. AI makes decisions by relying on real-world datasets: however, unbeknown to its developers, it also inherits and amplifies the biases present in them. Looking up images of 'CEO' on search engines will only result in 11% of the images being those of women, when in reality 27% of the CEOs in the US are female.[1] A few years ago, the facial recognition software of Google Photos came under radar when they identified two African-American men as gorillas.[2] Their solution was to remove gorillas altogether from their platform, a move clearly indicating how difficult it is for AI to 'unlearn'. An analysis of approximately 133 biased systems from 1988 to present day found that 44.2% demonstrate gender bias.[3]A sentence like, "She's president, he is cooking," would translate to "This person is president, this person is cooking," in a gender-neutral language on Google Translate. However, if translated back into English, it would become, "He's president, she is cooking."[4] Furthermore, research suggests that algorithms repeatedly associate female pronouns more with family words than career words, to an extent as to referring to a man as a "professor" and a woman as an "assistant professor."[5] Machine translation is becoming a prevalent AI tool today, with millions using it to converse and write in different languages. In this paper, we emphasize on research that substantiates the existence of bias in MT systems as well as the causes and solutions for this inequality. The impact of this disparity on women is also briefly discussed. It is imperative to understand this inequality as it gives rise to unethical practises and beliefs that halts the progress of minority groups like women.

---

[1] Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias." Mckinsey. April 7, 2021. https://www.mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias

[2] Tom Simonite. "When It Comes to Gorillas, Google Photos Remains Blind." WIRED. November 1, 2018. When It Comes to Gorillas, Google Photos Remains Blind | WIRED

[3] Genevieve Smith and Ishita Rustagi. "When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity." ssir. March 31, 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity (ssir.org)

[4] Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias."

[5] Angela Chen. "AI picks up racial and gender biases when learning from what humans write." The Verge. April 13, 2017. AI picks up racial and gender biases when learning from what humans write - The Verge

2 Gender Bias in Machine Translation

In recent years, several studies have confirmed the occurrence of unfairness in MT systems. In a 2018 case study[6] by faculty of Federal University of Rio Grande do Sul, sentences like "He/she is an engineer" were translated from 12 different gender-neutral languages (Malay, Estonian, Finnish, Hungarian, Armenian, Bengali, Japanese, Turkish, Yoruba, Swahili, Basque, Chinese) to English using the Google Translation API. 30 occupations were randomly selected from their dataset of 1019 different occupations obtained from the US Bureau of Labor Statistics' (BLS) occupations table. 21 adjectives to describe humans, obtained from the Corpus of Contemporary American English (COCA), were also translated. The frequency of female, male and gender-neutral pronouns in the translated output was measured and the findings attained were shown to be reflecting gender stereotypes prevailing in the real world. Sentences were translated according to stereotypical gender roles – occupations such as nurse and baker yielded female pronouns while engineer and CEO yielded male ones. It shows a coherent bias towards male default, especially for male-dominated and stereotypically associated fields such as engineering, mathematics, physical sciences, computer science, etc. Further, it is argued that the explanation for this disproportionate translation of gender pronouns is the workplace demographic data – the low number of female workers in certain fields. However, a comprehensive comparison between the data produced and the BLS's data for participation of women in each occupation concludes that Google Translation does not reproduce a real-world distribution of female employees. Sentences were translated with female pronouns only 11.76% of the time, whereas the average female participation in the chosen occupations is a whopping 35.94%. Moreover, translation of adjectives showed that words like shy, attractive, happy, kind, ashamed predominantly yielded female pronouns, while arrogant, cruel, guilty primarily yielded male pronouns. As Google Translate uses English to translate between other languages (e.g., Hungarian to English, and then to Hindi), these findings will further be consistent with non-gender-neutral languages. Therefore, this study is conclusive evidence for gender bias in prominent MT systems like Google Translate.

A 2019 study[7] conducted at University of Washington analysed gender bias in MT by using two coreference resolution datasets (the Winogender and the WinoBias dataset), which comprised of English sentences that depicted the noun in a non-stereotypical gender role (e.g., "The doctor asked the nurse to help her in the operation"). The pronoun in each of these sentences had to be resolved to one of the entities specified in the sentence (e.g., 'her' in this case is used to refer to the doctor, not the nurse). With the pronoun (gender of the noun) explicitly mentioned, each sentence was translated into eight target languages - Spanish, French, Spanish, Italian, Russian, Arabic, Ukrainian and Hebrew. Unexpectedly, it was found that these models ignore the context and basic English structure, often making bias predictions, such as associating female pronouns with stereotypically 'feminine' occupations like nursing. The 3888 sentences were equally balanced between genders as well as stereotypical and non-stereotypical gender-roles (e.g., a male doctor and a male nurse, or a female doctor and a female nurse). It was concluded that the probability of correctly translating the pronoun in all four tested popular systems (Google Translate, Microsoft Translator, Amazon Translate, and SYSTRAN) and two recent academic MT models was equivalent to a random guess, indicating a consequential case of gender bias in MT. Higher accuracy was seen in sentences with male roles, or with stereotypical gender role assignments (e.g., a female nurse is more accurately predicted than a male nurse). Furthermore, it was also tested if tampering with the dataset by adding adjectives "handsome" and "pretty" to sentences with male and female pronouns respectively would impact the predictions positively. For example, if a sentence is converted to: "The pretty doctor asked the nurse to help her in the operation", the signals are mixed – a stereotypically "masculine" profession like doctor was more likely to yield a male pronoun, while an adjective like pretty was more prone to producing a female pronoun. The results report improved performance in several languages, significantly impacting Spanish, Ukrainian, and Russian. Although an impractical debiasing scheme, it highlights how society sees women in the workplace.

---

[6] Prates, Marcelo OR, Pedro H. Avelar, and Luís C. Lamb. "Assessing Gender Bias in Machine Translation: A Case Study with Google Translate." Neural Computing and Applications 32, no. 10 (2020): 6363-6381.

[7]Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. "Evaluating gender bias in machine translation." *arXiv preprint arXiv:1906.00591* (2019).

In a research[8] performed in 2019, three machine translation systems (Google Translator, Naver Papago, and Kakao Translator) were tested to evaluate gender bias in translation of Korean language (which makes use of gender-neutral pronouns) to English. A corpus of 4236 sentences was constructed, entailing positive/negative expressions or occupations, with all terms being gender-independent (e.g., words like "masseuse" were excluded as they refer to a single gender). Seven subsets were also generated based on informality, formality, impoliteness, politeness, negative sentiment polarity, positive sentiment polarity, and occupation. The sentences in the second subset, i.e., subset with formal and gender-neutral pronouns (centred on news articles, technical reports, papers, etc), was drastically biased towards male translations than the subset that focused on informality. Although this predisposition can be explained by the larger percentage of men engaged in formal writing in the real world, the gender stereotypes this encourages cannot be ignored. Furthermore, Google Translator (GT) was found to be exacerbating occupational stereotypes by terming experts such as engineers, technicians, professors as men, and art-related positions such as fashion designer, hairdresser as women. Kakao Translator (KT) demonstrated a substantial prevalence of male translations, with hardly any female pronouns being generated. SBias, bias caused by social prejudice, was predominant in GT, while VBias, bias caused by high amount of occurrence, was recognized in KT.

A 2018 experiment[9] conducted at Johns Hopkins University evaluated gender bias in coreference resolution systems. Riddles like these are used in these models - A man and his son get into a terrible car crash. The father dies, and the boy is badly injured. In the hospital, the surgeon looks at the patient and exclaims, "I can't operate on this boy, he's my son!" How can this be? This simple riddle for which the answer is clearly "mother", is found to be difficult to comprehend and solve by majority of people due to the gender bias embedded in their minds. They are unaccustomed to the idea of a mother also being a surgeon. Conversely, a rule-based coreference system also makes the mistake of resolving this to a male pronoun. "Winogender schemas", a pronoun resolution schema similar to Winograd schemas (where a pronoun must resolve to one of the entities in a sentence), is used to examine gender bias in three publicly available coreference resolution systems (rule-based, statistical, neural). Rule-based systems are rule-based models that lack large-scale training data, statistical models are focused on millions of features, parameters, data, and neural systems use neural network models to learn statistical models.[10] In this challenge dataset of 720 sentences (and 60 one-word occupations), each sentence contains three phrases of interest:

1) OCCUPATION: An entity referred by the occupation (e.g., the paramedic)
2) PARTICIPANT: A second entity (e.g., the student)
3) PRONOUN: A pronoun co-referent with the OCCUPATION or the PARTICIPANT

Each occupation is used to form 12 sentences, with the pronoun referring to the OCCUPATION in one and the PARTICIPANT (a specific one, e.g., "passenger", and a generic one, e.g., "someone") in another. E.g.,

1) The **doctor** operated on the passenger/someone even though he/she/they had lost all hope.
2) The doctor operated on the **passenger/someone** even though he/she/they was/were already dead.
   *The pronoun is co-referent with the **bold** noun in each sentence

Hence the dataset contains 720 sentences (60 occupations × 2 occupations × 2 participants × 3 pronoun genders). Modifying the pronoun's gender in each sentence allows the comparison of the number of different gender pronouns that yield OCCUPATION and PARTICIPANT, and thus the gender bias in these systems. It was observed that none of the systems were gender neutral. Male pronouns are more likely to be resolved as OCCUPATION in all systems. The average percentage of male pronouns generated was 77%, while 57% were female pronouns, and 29% were neutral pronouns.

The rule-based system was the most biased, producing 72% male pronouns, 29% female pronouns and 1% neutral pronouns. These predictions greatly amplify the real-world gender disparities. For instance, 38.5% of "managers" in the U.S. are females according to BLS: however, only 5.18% of references of "manager" in the training data are female, with none predicted as female.

[8] Cho, Won Ik, Ji Won Kim, Seok Min Kim, and Nam Soo Kim. "On measuring gender bias in translation of gender-neutral pronouns." arXiv preprint arXiv:1905.11684 (2019).

[9] Rudinger, Rachel, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. "Gender bias in coreference resolution." arXiv preprint arXiv:1804.09301 (2018).

[10] Jason Brownlee. "A Gentle Introduction to Neural Machine Translation." Machine Learning Mastery. December 29, 2017. A Gentle Introduction to Neural Machine Translation (machinelearningmastery.com)

Thus, all of these studies show concrete evidence of existence of gender bias in MT.

3 Causes and Solutions

3.1 Underrepresentation of Women in STEM

AI, as any other creation, reflects the beliefs and biases of its creators, which are mainly men. A recent report by World Economic Forum found that only 22% of AI professionals globally are female.[11] A mere 15% of AI researchers at Facebook and 10% at Google are women.[12] This disparity often leads to gender bias in AI tools like Machine Translation. Dr. Muneera Bano, a senior software engineering lecturer at Deakin University, Australia, believes that its human nature to unconsciously add your biases to your algorithms and designs.[13] Humans determine the sources of data, method of data-collection, algorithms, and rules that must be followed in order to make predictions and identify trends – any of which can induce bias in the system. Research conducted by Aylin Caliskan, Princeton University, analysed the bias of an AI model and then compared it to the results of a well-recognized psychological test that measured bias in humans.[14] It was discovered that the algorithm contained all biases that were present in the team. It is quite evident that AI models exhibit their developer's stereotypical thinking pattern, hence making the diversity and inclusion of women in this field a necessity. Moreover, statistics on computer science students show that only 18% of all computer science bachelor's degrees in the US today are awarded to women.[15] As a result, the perspectives in the workplace are homogeneous, risking a lack of diverse perspectives in the technology deployed by that workplace as well.[16] Furthermore, it is often argued that gender bias in AI cannot be substantially removed due to its reliance on huge datasets and mathematical algorithms. However, Sareeta Amrute, Associate Professor of Anthropology at the University of Washington, disputed that holding developers accountable for their models shows that the biases produced are not inevitable.[17] People need to be committed to bringing in experts and affected groups. Women need to be involved as developers of AI solutions, as monitors (impact audits, public bodies, etc), and as informed citizens who foster awareness about gender inequality in machine systems.[18] Furthermore, several researchers propose 'data feminism' as a method to eradicating bias in AI. It focuses on examining and challenging power, elevating emotion, rethinking binaries and hierarchies, considering context, etc.[19] It encourages researchers to take into consideration the impact of their work, and to centre their research on these biases rather than being oblivious to it.

3.2 Data Bias

The fundamental principle of AI is to work with large amounts of data to make predictions based on identified patterns and trends. These 'patterns' are often gender biases, especially when the datasets are not responsibly curated by keeping their inherent stereotypes in mind. The data uses existing material like documents and information on the internet, which is corrupted with biases due to the prevailing gender inequality in the world. A Stanford University algorithm, Global Vectors for Word Representation (GLoVe), crawls the internet to find data and the relation between billions of words.[20] Researchers found that female words were more associated with arts than with math or science in GLoVe. Furthermore, the masculine form of engineer in German is 75 times more common than its female version.[21] Obviously, a machine translation algorithm trained on this biased data would replicate such imbalance. The Europarl corpus, a text corpus that is frequently used to train MT models, consists of only 30% of sentences spoken

[11] "Assessing Gender Gaps in Artificial Intelligence." Weforum. 2018. Global Gender Gap Report 2018 - Reports - World Economic Forum (weforum.org)
[12] Falon Fatemi. "Bridging The Gender Gap In AI." Forbes. February 17, 2020. Bridging The Gender Gap In AI (forbes.com)
[13] Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias."
[14] Angela Chen. "AI picks up racial and gender biases when learning from what humans write."
[15] Staff Writers. "Women in Computer Science: Getting Involved in STEM." Computer Science. September 29, 2021. Women in Computer Science | ComputerScience.org
[16] Havens, Lucy, Melissa Terras, Benjamin Bach, and Beatrice Alex. "Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research." *arXiv preprint arXiv:2011.05911* (2020).
[17] "Gender Equality and AI Principles." Artificial Intelligence and Gender Equality: Key findings of UNESCO's Global Dialogue. August 2020. 9-11
[18] "Recommendations For Integrating Ge into AI Principles." Artificial Intelligence and Gender Equality: Key findings of UNESCO's Global Dialogue. August 2020. 18
[19] Havens, Lucy, Melissa Terras, Benjamin Bach, and Beatrice Alex. "Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research."
[20] Angela Chen. "AI picks up racial and gender biases when learning from what humans write."
[21] Stefanie Ullmann and Danielle Saunders. "Google Translate is sexist. What it needs is a little gender-sensitivity training." Scroll.in. April 5, 2021. Google Translate is sexist and it needs a little gender-sensitivity training (scroll.in)

by women.[22] Another study found that machine translation systems (such as Google Translate) consistently perform better when presented with pro-stereotypical assignments (e.g., a female nurse), than in anti-stereotypical assignments (e.g., a male nurse).[23] It can therefore be concluded that a fair and unbiased AI system's performance is directly proportional to the amount of unbiased data fed to the algorithm – biased data will increase its capability of making biased predictions. Moreover, crowdsourcing, a sourcing model used to create training datasets for AI, often involves recruiting a small number of quality workers to generate massive examples, which raises doubts regarding the data diversity. This study tested three recent Natural Language Understanding (NLU) datasets and found that an annotator bias is evident in two of them, with the language generated by them often revealing their identity.[24] This must be a result of the small number of annotators being used to create large datasets. Hence, it is imperative to use balanced training sets in order to achieve a fair AI model.

4 Impact on Women

Machine bias that favours men has a profound impact on women and society in general. Not only does it play a part in hindering a woman's progress by making biased recruitment decisions,[25] but it also has the capability of risking a woman's life. Airbags in cars have been designed and tested specifically on dummies with male physique, making a woman 17% more likely to die than a man in a car accident.[26] Gender bias in MT alone has drastic consequences on women empowerment. It reinforces harmful gender roles (such as male doctors and female nurses), provoking prejudices that women have been struggling to obliterate since decades. People pay more attention to observations that match their stereotypical beliefs than they do to counter-stereotypical observations.[27] This means that when MT systems reiterate gender stereotypes, it psychologically fosters these biases among the users. Google Translate has over 1 billion downloads – an immense number of people will be affected by biased translations.[28] Women who were exposed to TV commercials portraying females in gender stereotypical roles (e.g. homemaker) were also less likely to choose a leadership role in a subsequent task.[29] This indicates that emphasis on gender stereotypes by AI solutions can have a similar impact, demotivating women from breaking the shackles of patriarchy. Furthermore, it has also been proven that women who were repeatedly subjected to negative gender stereotypes showed decreased performance and a reduced working memory capacity.[30] All these implications highlight how gender bias in technology can impact women in different ways.

5 Conclusion

We investigated the prevalence of gender bias in MT systems and presented four recent studies that confirmed the existence of bias. The first study showed how MT models immensely amplified real-world occupational stereotypes, translating sentences with female pronouns only 11.76% of the time, while the average female participation in the chosen occupations was a whopping 35.94%. It also predominantly associated 'feminine' adjectives like shy, attractive to women and other adjectives like guilty, cruel to men. The second and fourth analysis examined coreference resolution systems, where the pronouns explicitly referred to an entity in the sentence. However, it was found that the MT models disregarded the context of the sentences and instead translated them according to the prevailing gender stereotypes. Lastly, the third research translated Korean, a gender neutral-language to English and found out that once again occupational stereotypes were being exacerbated by algorithms.

---

[22] Savoldi, Beatrice, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. "Gender bias in machine translation." arXiv preprint arXiv:2104.06001 (2021).

[23] Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. "Evaluating gender bias in machine translation." arXiv preprint arXiv:1906.00591 (2019).

[24] Geva, Mor, Yoav Goldberg, and Jonathan Berant. "Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets." arXiv preprint arXiv:1908.07898 (2019)

[25] Julien Lauret. "Amazon's sexist AI recruiting tool: how did it go so wrong?" Becoming Human. August 16, 2019. Amazon's sexist AI recruiting tool: how did it go so wrong? | by Julien Lauret | Becoming Human: Artificial Intelligence Magazine

[26] Carmen Niethammer. "AI Bias Could Put Women's Lives at Risk - A Challenge for Regulators." Forbes. March 2, 2020. AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators (forbes.com)

[27] Tabassum, Naznin, and Bhabani Shankar Nayak. "Gender Stereotypes and Their Impact on Women's Career Progressions from a Managerial Perspective." IIM Kozhikode Society & Management Review 10, no. 2 (July 2021): 192–208.

[28] Jeff Pitman. "Google Translate: One billion installs, one billion stories." Google. April 28, 2021. Google Translate: One billion installs, one billion stories (blog.google)

[29] Ioana Latu and Marianne S. Mast. "The Effects of Stereotypes of Women's Performance in Male-Dominated Hierarchies: Stereotype Threat Activation and Reduction Through Role Models." REF (unil.ch)

[30] Toni Schmader and Michael Johns. "Converging Evidence That Stereotype Threat Reduces Working Memory Capacity." 10.1037/0022-3514.85.3.440.

Furthermore, we gave extensive evidence for how underrepresentation of women in STEM and data bias are the main causes for this unfairness. We concluded that inclusion of women in the AI field as well as diversity of data to train AI models was critical in the development of an ethical, fair system. If not resolved soon, gender bias in MT can continue having a negative impact on women psychologically, cause detrimental health issues, and even deter women's progress and empowerment.

To conclude, gender bias in machine translation is not inexplicable. All studies discussed in this paper give a justification for the bias found in their models. A fair model is attainable, although it requires engineers to prioritise bias mitigation. The purpose of AI is to be better than human beings and confirm the idea of perfection. It must not reflect society's preposterous misogynistic ideologies and must abide by ethical frameworks in order to pave the way for a just world where no one is affected by detrimental consequences of bigoted technological systems. The only way to achieve this goal is by integrating affected groups and gender experts in the technology workplace and by increasing social awareness so that specialists understand the significance of eradicating gender bias in AI.

References

1. Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias." Mckinsey. April 7, 2021. https://www.mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias
2. Tom Simonite. "When It Comes to Gorillas, Google Photos Remains Blind." WIRED. November 1, 2018. When It Comes to Gorillas, Google Photos Remains Blind | WIRED
3. Genevieve Smith and Ishita Rustagi. "When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity." ssir. March 31, 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity (ssir.org)
4. Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias." https://www.mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias
5. Angela Chen. "AI picks up racial and gender biases when learning from what humans write." The Verge. April 13, 2017. AI picks up racial and gender biases when learning from what humans write - The Verge
6. Prates, Marcelo OR, Pedro H. Avelar, and Luís C. Lamb. "Assessing Gender Bias in Machine Translation: A Case Study with Google Translate." Neural Computing and Applications 32, no. 10 (2020): 6363-6381. https://arxiv.org/abs/1809.02208
7. Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. "Evaluating gender bias in machine translation." arXiv preprint arXiv:1906.00591 (2019). https://arxiv.org/abs/1906.00591/
8. Cho, Won Ik, Ji Won Kim, Seok Min Kim, and Nam Soo Kim. "On measuring gender bias in translation of gender-neutral pronouns." arXiv preprint arXiv:1905.11684 (2019). https://arxiv.org/abs/1905.11684
9. Rudinger, Rachel, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. "Gender bias in coreference resolution." arXiv preprint arXiv:1804.09301 (2018). https://arxiv.org/pdf/1804.09301.pdf
10. Jason Brownlee. "A Gentle Introduction to Neural Machine Translation." Machine Learning Mastery. December 29, 2017. A Gentle Introduction to Neural Machine Translation (machinelearningmastery.com)
11. "Assessing Gender Gaps in Artificial Intelligence." Weforum. 2018. Global Gender Gap Report 2018 - Reports - World Economic Forum (weforum.org)
12. Falon Fatemi. "Bridging The Gender Gap In AI." Forbes. February 17, 2020. Bridging The Gender Gap In AI (forbes.com)
13. Anu Madgavkar. "A conversation on Artificial Intelligence and Gender Bias." https://www.mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias
14. Angela Chen. "AI picks up racial and gender biases when learning from what humans write." AI picks up racial and gender biases when learning from what humans write - The Verge
15. Staff Writers. "Women in Computer Science: Getting Involved in STEM." Computer Science. September 29, 2021. Women in Computer Science | ComputerScience.org
16. Havens, Lucy, Melissa Terras, Benjamin Bach, and Beatrice Alex. "Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research." *arXiv preprint*

arXiv:2011.05911 (2020). [2011.05911] Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research (arxiv.org)

17. "Gender Equality and AI Principles." Artificial Intelligence and Gender Equality: Key findings of UNESCO's Global Dialogue. August 2020. 9-11 https://en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf
18. "Recommendations For Integrating Ge into AI Principles." Artificial Intelligence and Gender Equality: Key findings of UNESCO's Global Dialogue. August 2020. 18. https://en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf
19. Havens, Lucy, Melissa Terras, Benjamin Bach, and Beatrice Alex. "Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research." [2011.05911] Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research (arxiv.org)
20. Angela Chen. "AI picks up racial and gender biases when learning from what humans write." AI picks up racial and gender biases when learning from what humans write - The Verge
21. Stefanie Ullmann and Danielle Saunders. "Google Translate is sexist. What it needs is a little gender-sensitivity training." Scroll.in. April 5, 2021. Google Translate is sexist and it needs a little gender-sensitivity training (scroll.in)
22. Savoldi, Beatrice, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. "Gender bias in machine translation." arXiv preprint arXiv:2104.06001 (2021). https://arxiv.org/abs/2104.06001
23. Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. "Evaluating gender bias in machine translation." arXiv preprint arXiv:1906.00591 (2019). https://arxiv.org/abs/1906.00591
24. Geva, Mor, Yoav Goldberg, and Jonathan Berant. "Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets." arXiv preprint arXiv:1908.07898 (2019) https://aclanthology.org/D19-1107.pdf
25. Julien Lauret. "Amazon's sexist AI recruiting tool: how did it go so wrong?" Becoming Human. August 16, 2019. Amazon's sexist AI recruiting tool: how did it go so wrong? | by Julien Lauret | Becoming Human: Artificial Intelligence Magazine
26. Carmen Niethammer. "AI Bias Could Put Women's Lives at Risk - A Challenge for Regulators." Forbes. March 2, 2020. AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators (forbes.com)
27. Tabassum, Naznin, and Bhabani Shankar Nayak. "Gender Stereotypes and Their Impact on Women's Career Progressions from a Managerial Perspective." IIM Kozhikode Society & Management Review 10, no. 2 (July 2021): 192–208. Gender Stereotypes and Their Impact on Women's Career Progressions from a Managerial Perspective - Naznin Tabassum, Bhabani Shankar Nayak, 2021 (sagepub.com)
28. Jeff Pitman. "Google Translate: One billion installs, one billion stories." Google. April 28, 2021. Google Translate: One billion installs, one billion stories (blog.google)
29. Ioana Latu and Marianne S. Mast. "The Effects of Stereotypes of Women's Performance in Male-Dominated Hierarchies: Stereotype Threat Activation and Reduction Through Role Models." REF (unil.ch)
30. Toni Schmader and Michael Johns. "Converging Evidence That Stereotype Threat Reduces Working Memory Capacity." 10.1037/0022-3514.85.3.440. (PDF) Converging Evidence That Stereotype Threat Reduces Working Memory Capacity (researchgate.net)