man = y :: woman = x

Deconstructing Gender Bias

Table of Contents

Table of Contents

Abstract

Abstract

The Practice of Sharing Knowledge — Deconstructing Gender Bias is a communication design project that seeks to promote reflection on the issue of gender bias in Artificial Intelligence.

It seeks to understand the social and cultural conceptions that inform this phenomenon and the way machine learning systems potentially reinforce and perpetuate cultural gender stereotypes.

“The potential wide-ranging impact makes it necessary to look carefully at the ways in which these technologies are being applied now, whom they’re benefiting, and how they’re structuring our social, economic, and interpersonal lives”.[1] The issue of gender bias is increasingly becoming a subject of discussion and reflection. According to Kate Crawford, it is important to move towards its neutralization. “And I think this increase in interest is completely justified, basically because machine learning systems are starting to impact millions of people every day. Therefore, bias is important. So, bias matters”.[1]

For this reason, it is important to address how social discrimination is reflected in the systems we build: “bias in systems is most commonly caused by bias in training data, and we can only gather data about the world that we have”, as Crawford asserts.

Thinking about gender bias in Artificial Intelligence is a challenge. This project intends to explore gender bias at a social and computational level to reveal how those levels interweave.

[1] Kate Crawford (2018) AI Now: Social and Political Questions for Artificial Intelligence. Distinguished Lectures. Tech Policy Lab. University of Washington

A Introduction

A Introduction[a]

[a] Jenna Cryan, Shiliang Tang, Xinyi Zhang, Miriam Metzger, Haitao Zheng, Ben Y. Zhao. 2020. Detecting Gender Stereotypes: Lexicon vs. Supervised Learning Methods.

The values of our society are both reflected in and reinforced by our use of language. In that context, sexism and gender discrimination is often perpetrated and reproduced through lexical choices in everyday communication. Recent studies have identified descriptions that reflect gender stereotypes in different types of articles, such as biographical pages of notable people[2], recommendation letters[3], fictional stories[4] and movie dialogue[5].

These issues are further exacerbated today by the ubiquitous usage of machine learning tools in language processing. We know that machine learning algorithms often translate and incorporate gender biases from training data[6], and such biases have been proven in popular techniques including word embeddings[7,8], coreference resolution[9] and sentence encoders.[10]

[2] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. In Proc. of ICWSM.

[3] Juan M Madera, Michelle R Hebl, and Randi C Martin.2009. Gender and letters of recommendation for academia: agentic and communal differences. Journal of Applied Psychology 94, 6 (2009), 1591.

[4] Ethan Fast, Tina Vachovsky, and Michael S Bernstein. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In Proc. of ICWSM.

5] Anil Ramakrishna, Victor R Martínez, Nikolaos Malandrakis, Karan Singla, and Shrikanth Narayanan. 2017. Linguistic analysis of differences in portrayal of movie characters. In Proc. of ACL, Vol. 1. 1669–1678.

[6] Latany Sweeney. 2013. Discrimination in online ad delivery. In arXiv:1301.6822.

[7] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In Proc. of NIPS. 4349–4357.

[8] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017a. Semantics derived automatically from language corpora contain human-like biases.Science 356, 14 (April 2017), p.183–186.

[9] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai--Wei Chang. 2018. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. In Proc. of NAACL, Vol. 2.

[10] Chandler May, Alex Wang, Shikha Bordia, Samuel R. Bowman, and Rachel Rudinger. 2019. On Measuring Social Biases in Sentence Encoders. In Proc. of NAACL.

B

Gender Bias in Western Society

[b] Pedro Costa. 2018. Conversations with ELIZA: on Gender and Artificial Intelligence. 6th Conference on Computation, Communication, Aesthetics & X Madrid, Spain

[11] West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” Gender and Society, Sage Publications. p.140

[12] West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” Gender and Society, Sage Publications. p.127

[13] Butler, Judith. 1990. Gender Trouble: Feminism and the Sub-version of Identity. New York and London: Routledge Classics. p. 522

[14] West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” Gender and Society, Sage Publications. p.126

B Gender Bias in Western Society

B.1 Gender Performance and Social Constructions[b]

“Gender is not simply an aspect of what one is, but, more fundamentally, it is something that one does, and does recurrently, in interaction with others.”[11]

Gender constitutes a part of our identity that regulates the type of behavior or acts we establish socially “by managing situated conduct in light of normative conceptions of attitudes and activities appropriate for one’s sex category”.[12] In this sense, Judith Butler introduced the idea that gender has a performative nature, given that gender identity is a repetition of acts stylized through time, manifesting a “cultural interpretation or signification of that [biological] facticity”.[13]

“Doing gender involves a complex of socially guided perceptual, interactional, and micropolitical activities that cast particular pursuits as expressions of masculine and feminine natures.”[14]

B Gender Bias in Western Society

B.1 Gender Performance and Social Constructions

B.1.1 Binary Framework

Simone de Beauvoir[fig.1] once said that “one is not born, but rather becomes, a woman” since ‘woman’ (as a concept) is a “historical idea and not a natural fact”.[15] These words suggest how gender is not something we are born with and, instead, is something we internalize through performative acts, over time. To be female or male is a matter of sex; but to be a man or a woman is a matter of gender. Gender is also seen as something polar, as seen through a “binary framework” in which there is a “mimetic relation of gender to sex whereby gender mirrors sex or is otherwise restricted by it.” [16,17]

Consequently, there is a normalization of what is considered to be feminine or masculine behavior, which becomes predetermined in a foreclosed historically sedimented structure. This establishes a set of expected behaviors and acts according to which we are compelled to act. That expectation is based on the perception that others have of our sex, which is presumed through the “factic datum of primary sexual characteristics”.[18] In other words, through this “need to routinize (…) behavior in accord with pre-established conceptualizations and behavioral patterns”[19], certain attributes and acts are identified as specifically feminine or masculine and are supposed to imply someone’s preferences and behaviors. As we grow up, and are categorized as men or women (or, instead, boys or girls) we are expected to comply to “normative conceptions of appropriate attitudes and activities” that are determined by “institutionalized frameworks through which natural, ‘normal sexedness is enacted’”.[20]

[fig.1] Simone de Beauvoir, 1957. Photo: Jack Nisberg / Roger-Viollet

[15] Butler, Judith. 1990. Gender Trouble: Feminism and the Subversion of Identity. New York and London: Routledge Classics. ch. 1 sec. III par. 3, 1988, p.522

[16] Butler, 1990, p.88

[17] According to Judith Butler, gender is “radically independent of sex” and, instead, is a “free-floating artifice”, while sex is defined as a “biological facticity” (Butler 1988), which means it is a biological criterion that distinguishes solely between female and male. As Butler puts it, gender “is neither the causal result of sex nor as seemingly fixed as sex” (Butler 1990, ch.1 sec. II par. 1). Therefore, gender is not something inherent “because gender is not a fact, the various acts of gender creates the idea of gender, and without those acts, there would be no gender at all” and gender is shaped and socially defined according to a “tacit collective agreement to perform, produce and sustain discrete and polar genders as cultural fictions” (Butler 1988, p.522).

[18] Butler, Judith. 1988. “Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory.” Theatre Journal 40 (4):519-531. The Johns Hopkins University Press. p.528

[19] Deaux, Kay, and Brenda Major. 1987. “Putting Gender Into Context: An Interactive Model of Gender-Related Behavior.” Psychological Review 94 (3):369-389. American Psychological Association Inc. p.370

[20] Goffman, 1977 in West and Zimmerman 1987, p.137

B Gender Bias in Western Society

B.2 Gender Sereoypes

As Prentice and Carranza put it, “prescriptive gender stereotypes” define “the qualities [ascribed] to women and men (…) that are required of women and men”.[21] These stereotypes imply that a gender belief system imposes expectations and gender behavior patterns, as internalized and socially reinforced stereotypes. Butler expands on this, stating that “gender performances (…) are governed by (…) punitive and regulatory social conventions”[22] that reject the acts or behaviors that convey some kind of deviation from the norm.

Some stereotypes, presented by Bem (1981 in Prentice and Carranza 2002, 269):

[21] Prentice, Deborah A., and Erica Carranza. 2002. “What Women and Men Should Be, Shouldn’t Be, Are Allowed to Be, and Don’t Have to Be: The Contents of Prescriptive Gender Stereotypes.” Psychology of Women Quarterly, 269-281. Blackwell Publishing. p.269

[22] Butler, Judith. 1988. “Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory.” Theatre Journal 40 (4):519-531. The Johns Hopkins University Press. p.527

F

affectionate, cheerful, childlike, compassionate, does not use harsh language, eager to soothe hurt feelings, feminine, flatterable, gentle, gullible, loves children, loyal, sensitive to the needs of others, shy, soft-spoken, sympathetic, tender, understanding, warm, yielding

M

acts as a leader, aggressive, ambitious, analytical, assertive, athletic, competitive, defends own beliefs, dominant, forceful, has leadership abilities, independent, individualistic, makes decisions easily, masculine, self-reliant, self-sufficient, strong personality, willing to take a stand, willing to take risks

Gender roles and characteristics deemed as specifically feminine or masculine also imply a structural hierarchization of labour. In other words:

“If, in doing gender, men are also doing dominance and women are doing deference (cf. Goffman 1967, pp. 47-95), the resultant social order, which supposedly reflects ‘natural differences’, is a powerful reinforcer and legitimator of hierarchical arrangements”.[23]

This means that feminine and masculine behavior is also used to segregate and structure labour accordingly. The workplace and its relationships change since, according to Kelly, when we interact within these contexts “social labels, beliefs and attributions may serve as grounds for predictions and generate behavior designed to validate or invalidate these beliefs”.[24] In fact, a lot of service work is seen as feminized labour or “associated with qualities traditionally coded as feminine”.[25]

[23] West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” Gender and Society, Sage Publications. p.146

[24] Snyder, Mark. 1977. “On the Self-Fulfilling Nature of Social Stereotypes.” Annual Meeting of the American Psychological Association, San Francisco, California. p.8

[25] Hester, Helen. 2016. “TechnologyBecomes Her.” New Vistas 3 (1):46-50. p.47

B Gender Bias in Western Society
B.2 Gender Sereoypes
B.2.1 Feminized Labour

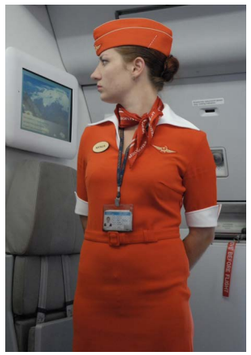
In other words, by expecting certain acts (deemed as feminine) from women, we expect them to occupy jobs and perform tasks associated with these attributes, thereby creating a category of feminine labour. To give a concrete example, historically women have a significant presence in the telecommunications industry, where they filled the role of assisting and establishing calls and communications, which rendered “female operators (…) inferior, subordinate, and knowable”.[26] In other cases, women fill the role of secretaries[fig.2], assistants, nurses[fig.3] or even flight attendances.[fig.4] These type of jobs convey, in a way, an “assumption that women possess a natural affinity for service work and emotional labour”.[27]

[26] Zost, Mary. 2015. “Phantom of the Operator: Negotiating Female Gender Identity in Telephonic Technology from Operator to Apple iOS.” Senior Thesis, BA, Faculty of College of Arts and Science of Georgetown University. p.3

[27] Hester, Helen. 2016. “TechnologyBecomes Her.” New Vistas 3 (1):46-50. p.47

[fig.3] District Nurse, 1990s. Source: Flickr.

[fig.4] Aeroflot flight attendant Photo: Eric Böhm, Source: Flickr.



[fig.2] Source: Try out a retro and modern office in BB Centrum.

This asymmetry also affects the private sphere, namely domestic work. As West and Zimmerman explain, household[fig.5] and child care[fig.6] tasks are considered women's work as a consequence of "normative conceptions of appropriate attitudes and activities for sex category".[28] The heterosexual framework contributes to this asymmetry since it reinforces the "embodiment of wifely and husbandly roles, and derivatively, of womanly and manly conduct".[29]

Additionally, and according to Donna Haraway, domestic work is transformed into capitalized labour out of the private sphere, through jobs such as office work, nursing or service work. Borrowing from Richard Gordon, Haraway considers that, with new media, a "homework economy" emerges, defined as a "restructuring of work that broadly has the characteristics formerly ascribed to female jobs, jobs done only by women".[30]

[28] West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." Gender and Society, Sage Publications. p.139

[29] Beer 1983 in West and Zimmerman 1987, p.144

[30] Haraway, Donna. 1991. "A Cyborg Manifesto: Science, technology and socialist-feminism in the late twentieth century." In Simians, Cyborgs, and Women: The Reinvention of Nature. London: Free Association Books. p.304

[fig.5] 1970s Woman Housewife Homemaker wearing apron loading laundry into washing machine (Photo by H. Armstrong Roberts/ ClassicStock/Getty Images)

[fig.6] Germany Bavaria, mother and son preparing boletus, 1960ies (Photo by Oskar Poss/ via Getty Images)

[31] Haraway, Donna. 1991. "A Cyborg Manifesto: Science, technology and socialist-feminism in the late twentieth century." In Simians, Cyborgs, and Women: The Reinvention of Nature. London: Free Association Books. p.305-307

Therefore, even outside the domestic sphere, women still ensure domestic tasks: "partly as function of their enforced status as mothers" as well as working in an "integrated circuit (…) in advanced industrial societies [where] these positions have been restructured (…) by social relations mediated and enforced by the new technologies".[31]

This reflects traditional conceptions of gender derived from a patriarchal heteronormative society where women perform domestic and assistant-like roles, while it also reveals how gender standardization and normalization has consequences at a social, personal and structural level.

[a] Jenna Cryan, Shiliang Tang, Xinyi Zhang, Miriam Metzger, Haitao Zheng, Ben Y. Zhao. 2020. Detecting Gender Stereotypes: Lexicon vs. Supervised Learning Methods.

[fig.7] 1950s modern' business office, with a boss watching as his secretary answers the telephone, 1958. Screen print. (Photo by GraphicaArtis/Getty Images)

Gender stereotypes are common beliefs about what men and women's physical and personality traits are and should be like. According to traditional gender stereotypes, women should display communal traits (e.g., nice, caring, warm) and men should display agentic traits (e.g., assertive, competent, effective).[32, 33]

[32] Alice H Eagly, Wendy Wood, and Amanda B Diekman. 2000. Social role theory of sex differences and similarities: A current appraisal. The developmental social psychology of gender 12 (2000), p.174.

[33] Amy JC Cuddy, Susan T Fiske, and Peter Glick. 2004. When professionals become mothers, warmth doesn't cut the ice. Journal of Social issues 60, 4 (2004), 701–718.

[34] Michela Menegatti and Monica Rubini. 2017. Gender bias and sexism in language. (2017).

[35] Anne Maass and Luciano Arcuri. 1996. Language and stereotyping. Stereotypes and stereotyping(1996), p.193–226.

[36] Mahzarin R Banaji and Curtis D Hardin. 1996. Automatic stereotyping. Psychological science 7, 3 (1996), p.136–141.

[37] Ethan Fast, Tina Vachovsky, and Michael S Bernstein. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In Proc. of ICWSM.

[38] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. In Proc. of ICWSM.

[39] Juan M Madera, Michelle R Hebl, and Randi C Martin. 2009. Gender and letters of recommendation for academia: agentic and communal differences. Journal of Applied Psychology 94, 6 (2009).

[40] Anil Ramakrishna, Victor R Martínez, Nikolaos Malandrakis, Karan Singla, and Shrikanth Narayanan. 2017. Linguistic analysis of differences in portrayal of movie characters. In Proc. of ACL, Vol. 1. 1669–1678.

Gender stereotypes emerge in language choices used in written and verbal communication.[34] It has been found that a category label used to refer to a group automatically activates the characteristics stereotypically associated with the group, even in supposedly unprejudiced people who do not explicitly endorse the stereotype.[34, 35] This also applies when the category label is one's gender. For example, after primed by words consistent with gender stereotypes (e.g.,"nurse"), people are faster to associate gender pronouns (e.g., "she") with the corresponding gender (e.g.,"female").[36]

As a result, gender stereotypes are common in contemporary languages, both in written and spoken communication. For example, in fiction writing, traditional gender stereotypes such as dominant men and submissive women are common throughout nearly every genre, regardless of the gender of the author.[37] On Wikipedia, articles about notable women emphasize more on romantic relationships or family-related issues compared to articles about notable men.[38] When writing recommendation letters for faculty positions, women are often described as more communal and less agentic than men.[39] Additionally, in movie dialogue, male characters use more words related to achievement than female characters.[40]

C

Gender Bias in Word Analogies

[d] Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai (2016), Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.

D

Gender Bias in Artificial Intelligence

[c] Susan Leavy (2018), Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning.

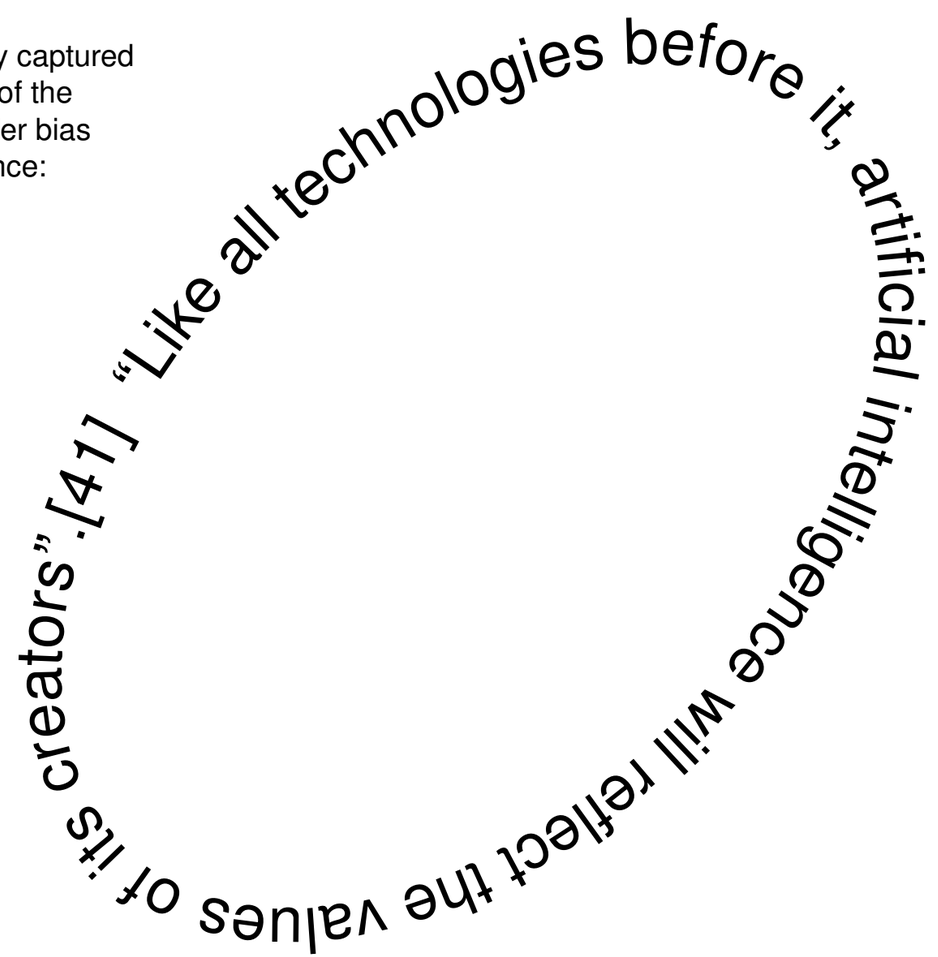
Artificial intelligence is increasingly influencing the opinions and behavior of people in everyday life. However, the over-representation of men in the design of these technologies could quietly undo decades of advances in gender equality.

Over centuries, humans developed critical theory to inform decisions and avoid basing them solely on personal experience. However, machine intelligence learns primarily from observing data that it is presented with. While a machine's ability to process large volumes of data may address this in part, if that data is laden with stereotypical concepts of gender, the resulting application of the technology will perpetuate this bias.

While some recent studies sought to remove bias from learned algorithms they largely ignore decades of research on how gender ideology is embedded in language. Awareness of this research and incorporating it into approaches to machine learning from text would help prevent the generation of biased algorithms.

Leading thinkers in the emerging field addressing bias in artificial intelligence are also primarily female, suggesting that those who are potentially affected by bias are more likely to see, understand and attempt to resolve it. Gender balance in machine learning is therefore crucial to prevent algorithms from perpetuating gender ideologies that disadvantage women.

Kate Crawford aptly captured the ultimate cause of the prevalence of gender bias in artificial intelligence:



[41] Kate Crawford. 2016. Artificial Artificial Intelligence's White Guy Problem. The New York Times.

Societal values that are biased against women can be deeply embedded in the way language is used and preventing machine learning algorithms trained on text from perpetuating bias requires an understanding of how gender ideology is manifested in language.

Developers of artificial intelligence are	Those who have recognized and are seeking to address this issue are
overwhelmingly male.	overwhelmingly female.[42]

[42] “Kate Crawford, Fei-Fei Li and Joy Buolamwini to name but a few.”

[43] Niki Kilbertus, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. 2017. Avoiding discrimination through causal reasoning. In Advances in Neural Information Processing Systems. 656–666.

It follows that to avoid gender biased algorithms influencing decisions in our society, diversity in the area of machine learning is essential. The benefits of diversity in the workplace are well documented and largely stem from the inclusion of a range of critical perspectives. Diversity in the development of machine learning technologies could accelerate solutions to the issue of gender bias by improved assessment of training data, incorporation of concepts of fairness in algorithms[43] and the assessment of the potential impact of gender bias in the context of the intended use of the technology.

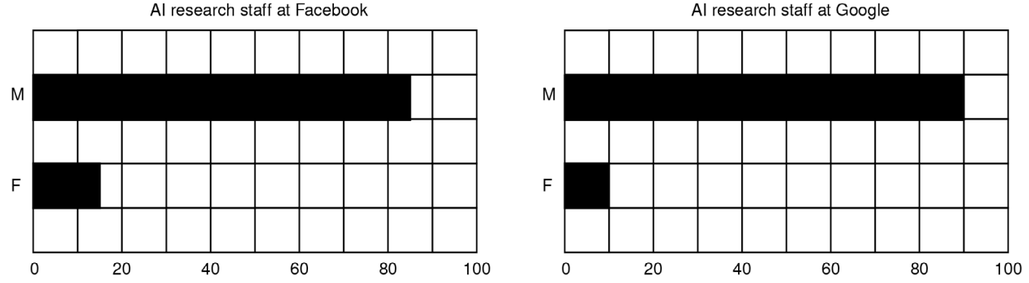
D Gender Bias in Artificial Intelligence

D.1 Diversity Crisis in AI[d]

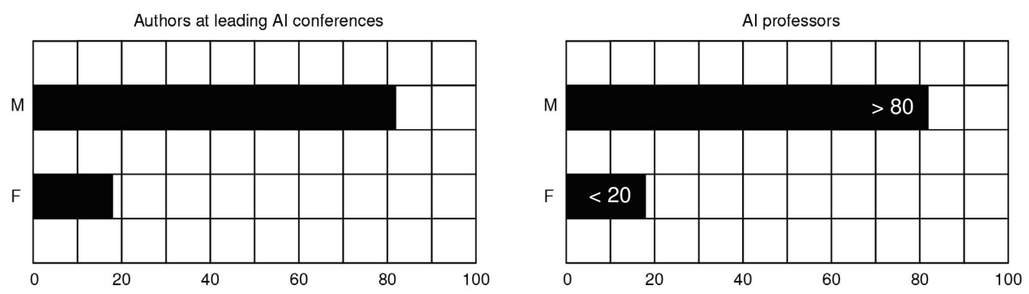
[d] Sarah Myers West, Meredith Whittaker, Kate Crawford. 2019. Discriminating Systems, Gender, Race, and Power in AI.

There is a diversity crisis in the AI industry, and a moment of reckoning is underway. Over the past few months, employees have been protesting across the tech industry where AI products are created. (...) This is just one face of the diversity disaster that now reaches across the entire AI sector. The statistics for both gender and racial diversity are alarmingly low.

For example, women comprise 15% of AI research staff at Facebook and just 10% at Google.[44]



It's not much better in academia, with recent studies showing only 18% of authors at leading AI conferences are women[45], and more than 80% of AI professors are male.[46]



[44] Simonite, T. (2018). AI is the future - but where are the women? WIRED.

[45] Element AI. (2019). Global AI Talent Report 2019.

[46] AI Index 2018. (2018). Artificial Intelligence Index 2018.

D Gender Bias in Artificial Intelligence

D.1 Diversity Crisis in AI

The diversity problem is not just about women. It's about gender, race, and most fundamentally, about power.[47] It affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development. (...)

To date, the diversity problems of the AI industry and the issues of bias in the systems it builds have tended to be considered separately. But we suggest that these are two versions of the same problem: issues of discrimination in the workforce and in system building are deeply intertwined. (...)

From a high-level view, AI systems function as systems of discrimination: they are classification technologies that differentiate, rank, and categorize. But discrimination is not evenly distributed. A steady stream of examples in recent years have demonstrated a persistent problem of gender and race-based discrimination (among other attributes and forms of identity). Image recognition technologies miscategorize black faces[48], sentencing algorithms discriminate against black defendants[49], chatbots easily adopt racist and misogynistic language when trained on online discourse [50], and Uber's facial recognition doesn't work for trans drivers.[51] In most cases, such bias mirrors and replicates existing structures of inequality in society.

In the face of growing evidence, the AI research community, and the industry producing AI products, has begun addressing the problem of bias by building on a body of work on fairness, accountability, and transparency. This work has commonly focused on adjusting AI systems in ways that produce a result deemed "fair" by one of various mathematical definitions[52] Alongside this, we see growing calls for ethics in AI, corporate ethics boards, and a push for more ethical AI development practices.[53]

[47] As authors of this report, we feel it's important to acknowledge that, as white women, we don't experience the intersections of oppression in the same way that people of color and gender minorities, among others, do. But the silence of those who experience privilege in this space is the problem: this is in part why progress on diversity issues moves so slowly. It is important that those of us who do work in this space address these issues openly, and act to center the communities most affected.

[48] Alcine, J. (2015). Twitter. Retrieved from https://twitter.com/jackyalcine/status/615329515909156865.

[49] Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016, May 3). Machine Bias. ProPublica

[50] Vincent, J. (2016, Mar 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. The Verge.

[51] Melendez, S. (2018, Aug. 9). Uber driver troubles raise concerns about transgender face recognition. Fast Company.

[52] Narayanan, A. (2018). 21 fairness definitions and their politics. ACM Conference on Fairness, Accountability and Transparency.

[53] Vincent, J. (2019, Apr. 3). The Problem with AI Ethics. The Verge.

D Gender Bias in Artificial Intelligence

D.1 Diversity Crisis in AI

Currently, large scale AI systems are developed almost exclusively in a handful of technology companies and a small set of elite university laboratories, spaces that in the West tend to be extremely white, affluent, technically oriented, and male[54]. These are also spaces that have a history of problems of discrimination, exclusion, and sexual harassment. As Melinda Gates describes, “men who demean, degrade or disrespect women have been able to operate with such impunity—not just in Hollywood, but in tech, venture capital, and other spaces where their influence and investment can make or break a career. The asymmetry of power is ripe for abuse”[55]. Or as machine learning researcher Stephen Merity noted at the end of 2017, “Bias is not just in our datasets, it’s in our conferences and community.”[56]

Both within the spaces where AI is being created, and in the logic of how AI systems are designed, the costs of bias, harassment, and discrimination are borne by the same people: gender minorities, people of color, and other under-represented groups. Similarly, the benefits of such systems, from profit to efficiency, accrue primarily to those already in positions of power, who again tend to be white, educated, and male. This is much more than an issue of one or two bad actors: it points to a systematic relationship between patterns of exclusion within the field of AI and the industry driving its production on the one hand, and the biases that manifest in the logics and application of AI technologies on the other.

These problems are not inevitable, nor are they natural: history shows us that they are a product of the distribution of power in society.[57] (...) As Hicks describes, “throughout history, it has often not been the content of the work but the identity of the worker performing it that determined its status”.[58]

[54] Crawford, K. (2016, June 25). Artificial Intelligence’s White Guy Problem. The New York Times.

[55] Kolhatkar, S. (2017). The Tech Industry’s Gender Discrimination Problem.

[56] Merity, S. (2017). Bias is not just in our datasets, it’s in our conferences and community. Smerity.com.

[57] Greenbaum, J.(1990). Windows on the Workplace: Computers, Jobs, and the Organization of Office Work in the Late Twentieth Century. New York: Monthly Review Press. Oldenziel, R. (1999) Making Technology Masculine: Men, Women, and Modern Machines in America, 1870-1945. Amsterdam: Amsterdam University Press.; Ensmenger, N. (2015). Beards, Sandals, and Other Signs of Rugged Individualism: Masculine Culture within the Computing Professions. Osiris, 30(1): 38-65.

[58] Hicks, M. (2017). Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing. Cambridge: MIT Press, 16.

D.1.1 Who makes AI?

The current data on the state of gender diversity in the AI field is dire, in both industry and academia. For example, in 2013, the share of women in computing dropped to 26%, below their level in 1960[59]. Almost half the women who go into technology eventually leave the field, more than double the percentage of men who depart[60].As noted above, a report produced by the research firm Element AI found that only 18% of authors at the leading 21 conferences in the field are women[61],while the 2018 Artificial Intelligence Index reports 80% of AI professors are men.[62] This imbalance is replicated at large tech firms like Facebook and Google, whose websites show even greater imbalances, with women comprising only 15% and 10% of their AI research staff, respectively[63, 64].There is no reported data on trans workers or other gender minorities. (...)

“We are in a diversity crisis for AI,” Gebru explains. “In addition to having technical conversations, conversations about law, conversations about ethics, we need to have conversations about diversity in AI. This needs to be treated as something that’s extremely urgent.” (...)

Discrimination and inequity in the workplace have significant material consequences, particularly for the under-represented groups who are excluded from resources and opportunities. For this reason alone the diversity crisis in the AI sector needs to be urgently addressed. But in the case of AI, the stakes are higher: these patterns of discrimination and exclusion reverberate well beyond the workplace into the wider world. Industrial AI systems are increasingly playing a role in our social and political institutions, including in education, healthcare, hiring, and criminal justice. Therefore, we need to consider the relationship between the workplace diversity crisis and the problems with bias and discrimination in AI systems.

[59] Thompson, C. (2019, Feb. 13). The Secret History of Women in Coding. New York Times Magazine.

[60] Ashcraft, C., McLain, B. and Eger, E. (2016). Women in Tech: The Facts. National Center for Women in Information Technology.

[61] Element AI. (2019). Global AI Talent Report 2019.

[62] AI Index 2018. (2018). Artificial Intelligence Index 2018.

[63] Simonite, T. (2018). AI is the future - but where are the women? WIRED.

[64] The World Economic Forum’s 2018 Global Gender Gap Report includes a section on diversity in AI that places its estimate much higher at 22%. However, the methodology for obtaining this figure raises some questions: it relies on LinkedIn users’ inclusion of AI-related skills in their profiles as the primary data source. This requires several causal leaps: first, that a sample of LinkedIn users is representative of the global population of workers in the field of AI, and that these users accurately represented their skill set. Moreover, the study used a flawed mechanism to attribute gender on a binary basis to users on the basis of inference from their first name – a practice that is not only trans-exclusionary, but is particularly problematic in an analysis that includes names in non-English languages.

It is a critical time to be addressing the diversity crisis in AI, because we now see diversity itself being weaponized. Over the past year and a half, evidence of systemic discrimination and harassment at tech companies and conference spaces has entered the public debate, much of it exposed by worker-led initiatives and whistleblowers. This growing awareness, accompanied by demands for inclusion and equity, has led to some change, but there has also been resistance, especially among those implicitly privileged by the status quo.

Those questioning and even rejecting the idea that racism, misogyny, and harassment are problems within the AI field and the tech industry have appropriated the language of diversity to argue that efforts to improve inclusion are in fact 'exclusionary', and that addressing the deeper structural challenges posed by racism, sexism, and inequity is misguided. (...) Such pushback often centers calls for "cognitive diversity" or "viewpoint diversity," the idea that individual differences in the ways people think and understand the world are distinctions that should be counted alongside, or instead of, other identity categories such as race and gender. As Bärí A. Williams puts it, "a dozen white men, so long as they were not raised in the same household and don't think identical thoughts, could be considered diverse".[65]

These arguments work by centering "identity" while flattening or ignoring power relationships. For example, in 2017 Facebook VP of Engineering Regina Dugan said that "the ultimate goal is cognitive diversity, and cognitive diversity is correlated with identity diversity. That means it's not just about [getting] women in tech. It's about broad voices, broad representation. But we can't step away from the idea that in the workplace, diversity also looks like identity diversity. You have to get to the place where you aren't made comfortable by the fact that everyone is the same, but rather feel inspired by how different we are".[66]

[65] Williams, B.A. (2017, Oct. 16). Tech's Troubling New Trend: Diversity Is in Your Head. The New York Times.

[66] Fast Company. (2017, Jan. 9) Facebook Engineering VP Explains Why "Cognitive Diversity Is the Most Powerful Tool".

The diversity crisis in AI is well-documented and wide-reaching. It can be seen in unequal workplaces throughout industry and in academia, in the disparities in hiring and promotion, in the AI technologies that reflect and amplify biased stereotypes, and in the resurfacing of biological determinism in automated systems.

Our analysis surfaced two prominent responses to the diversity crisis: on the one hand, a worker-driven movement focused on addressing inequities is showing promise in driving change. On the other hand, we observe a small but vocal counter-movement that actively resists diversity in the industry and uses arguments from biological determinism to assert that women are inherently less suited to computer science and Artificial Intelligence.

This is a critical moment for the AI industry to decide what it will do. As AI systems are embedded in more social domains, they are playing a powerful role in the most intimate aspects of our lives: our health, our safety, our education, and our opportunities. It's essential that we are able to see and assess the ways that these systems treat some people differently than others, because they already influence the lives of millions.

There have been attempts to address gender bias in machine learning through the review of learned gender -based associations and modification of the algorithms to exclude stereotypes[67]. However, there is little consideration of the decades of research that exist on the relationship between gender ideology and language. Incorporating gender theory, in particular feminist linguistic theory, into the approach to machine learning from textual data may prevent learning of gender bias and avoid the need to modify the algorithms.

[c] Susan Leavy (2018), Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning.

[67] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasingword embeddings. In Advances in Neural Information Processing Systems. 4349–4357.

Many of the debates in artificial intelligence on the topic of gender bias mirror those related to gender equality in society since the 1960s. It is important that computer scientists look to such debates so that negative consequences for women due to gender bias are not repeated. Feminist studies from the 1960s analyzed how women were often represented as passive, emotional and irrational in literature[68] and how the media presented idealized portrayals of femininity[69]. In the later part of the 20th century feminist theorists questioned the active role of language in the perpetuation of gender ideologies in society[70]. These seminal works identified ways in which gender ideology is embedded in language and how this can influence people's conceptions of women and expectations of behavior associated with gender. These gender ideologies are still embedded in text sources and result in machine learning algorithms learning stereotypical concepts of gender[67].

To ascertain the importance of addressing gender bias in machine learning, a lot can be learned from experiments in the 1970s showing its damaging effects[71, 72]. These studies prompted the development of guidelines to avoid the use of gender biased or sexist language[73, 74]. For example, the publisher McGraw-Hill adopted editorial guidelines to avoid sexist language[75]. It would be unfortunate to have to wait until gender biased machine learning algorithms repeat the injustices of the past before action preventing gender bias is taken.

[68] Kate Millett. 2016. Sexual politics. Columbia University Press.

[69] Betty Friedan. 2001. The Feminine Mystique. 1963. New York (2001).

[70] Judith Butler. 1990. Gender trouble and the subversion of identity. New York and London: Routledge (1990).

[71] Sandra L Bem and Daryl J Bem. 1971. Training the woman to know her place: The social antecedents of women in the world of work. Department of Psychology, Stanford University.

[72] Wendy Martyna. 1978. What Does 'He' Mean? Use of the generic masculine. Journal of communication 28, 1 (1978), 131–138.

[73] Miller Casey, Swift Kate, and Dowrick Stephanie. 1981. The Handbook of Nonsexist Writing: For Writers, Editors and Speakers. Women's Press.

[74] Casey Miller and Kate Swift. 2001. The handbook of nonsexist writing. iUniverse.

[75] Janice Moulton, George M Robinson, and Cherin Elias. 1978. Sex bias in language use:" Neutral" pronouns that aren't. American Psychologist 33, 11 (1978), 1032.

Work within the field of stylistics on gender and language has identified recurring linguistic features of language that are attributable to gender bias.[76] This work lends itself to a computational approach to identifying gender bias and could be used to remove it from training data for a machine learning algorithm. The following demonstrates how an abstract concepts such a gender bias can be operationalized into measurable features of text that can be computationally identified. This connection of theoretical and critical perspectives on language to the feature extraction stage of machine learning is the key to addressing bias in artificial intelligence.

[76] Sara Mills. 1995. Feminist stylistics. Routledge London.

[77] Suzanne Romaine et al. 1998. Communicating gender. Psychology Press.

[78] Robert Sigley and Janet Holmes. 2002. Looking at girls in corpora of English. Journal of English Linguistics 30, 2 (2002), 138–157.

[79] Paul Baker. 2008. Sexed texts: language, gender and sexuality. Equinox.

[80] Janet Holmes. 2002. Gender identity in New Zealand English. Gender across languages. The linguistic representation of women and men 1 (2002).

[81] Lia Litosseliti and Jane Sunderland. 2002. Gender identity and discourse analysis. Vol. 2. John Benjamins Publishing.

[82] Casey Miller and Kate Swift. 2001. The handbook of nonsexist writing. iUniverse.

a) Naming

Gender bias can be recognized in terms used to describe groupings of men and women. For instance a father is often described as a ‘family man’ with no commonly used equivalent such as ‘family woman’[77].Terms such as ‘single mum’, ‘working mother’, ‘career woman’ and ‘mother’ commonly used in the media also reveals social preconceptions of women.[76] Occupational terms used in relation to women were found to be often pre-modified by a gender specification such as ‘female lawyer’ and ‘woman judge’, identifying their existence as counter to societal expectations.[78]

Another manifestation of gender bias that is in decline is the use of androcentric terms such as ‘he’, ‘him’, ‘man’ and ‘mankind’ to refer to both men and women [79, 80]. However, in referring to groups, where there is an expectation that the individuals in question are more likely be of a particular gender, that gender will be used to refer to both men and women in the group.[81] For example in reference to a group of fire-fighters individuals are more likely to be referred to in male terms. In the context of machine learning, while certain linguistic features may be used less in current textual sources, machine learning algorithms that are trained on older corpora may reflect outdated ways of referring to men and women.

Women are described as girls more often than men are described as boys[77]. In an analysis of the use of the terms girl(s) and boy(s) in a corpus of text of British, American and New Zealand English[78], found that the term ‘girl’ is 3 times more likely than the term ‘boy’ to refer to an adult and that women were described as girls in in order to characterize them as immature, innocent, of youthful appearance, subordinate status, emotionally weak or financially dependent. Using ‘girl’ in conjunction with occupations also reduced the status of the jobs. In [79] it was found that the terms ‘boy’ and ‘girl’ occurred with equal frequency in an analysis of examples of British English texts including literature and media content from 2006. However, 52 percent of uses of the term‘girl’ referenced women while 28 percent of the uses of ‘boy’ pertained to men. The term ‘girl’ was also used in more disparaging and sexual contexts . This demonstrates how techniques to analyze not only the frequency of mentions but the broader context of the use of terms for men and women in texts could detect gender bias in training data for machine learning.

Honorific titles such as ‘Miss’ and ‘Mrs’ reflect the marital status of women but the male equivalent does not, demonstrating how women are portrayed in terms of their relationships to others[76, 82]. In the 1970s ‘Ms’ was introduced as an equivalent for ‘Mr’ to address this asymmetry. However, there is evidence that ‘Ms’ is being used to replace ‘Mrs’ but not ‘Miss’[80].

b) Ordering

Gender bias in language is evident in the ordering of items in lists. In English, it is convention when naming pairs of each gender, to name the male first (eg. son and daughter, husband and wife, Mr and Mrs)[83]. This practice demonstrates a bias which presents a gender-based social order[83, 84, 85]. This practice of naming the most powerful of a pair first is evidenced by the following common pairs: 'master/ servant', 'teacher/pupil' and 'doctor/nurse'.[86]

A comprehensive study of the ordering of personal binomials in the British National Corpus uncovered examples of word pairs studied included 'man/woman', 'girl/boy', nobility titles such as 'lady/gentleman', 'princess/prince', kingship terms such as 'wife/husband', occupations such as 'actress/actor' and pronouns such as 'he/ she'[84]. While there were variances in the order of naming the pairs, gender was the most important influencing factor regarding which of the pair of terms was named first.

c) Biased Descriptions

In an analysis of adjectives used to describe men and women in British newspapers[87], found that men were more frequently described in terms of their behavior while women were described in terms of their appearance and sexuality. In an analysis of the context of the use of the term 'girl', research has shown that girls and boys are represented differently with girls being more objectified[88] and portrayed in more negative contexts[89]. Extraction of adjectives used to describe women in training data could therefore be incorporated as part of gender proofing the textual data that is used to training machine learning algorithms.

How word embedding learns stereotypes has been the focus of recent research on gender bias and artificial intelligence[90]. Evaluating what constitutes a stereotypical association has largely been a result of researcher interpretation. However [91] analyzed the British national Corpus and extracted collocates of men and women and identifying those that were just used for each gender, revealing striking gender stereotypes[fig.8]. Other kinds of stereotypes have been identified in relation to sexuality, beauty[92] and levels of agency[83].

Gender	Adjectives
Female	Bossy, chattering, gossiping, submissive, bitchy, hysterical, weeping
Male	Gregarious, cautious, affable, amiable, avuncular, funniest, good-natured, jovial, likable, mild-mannered, personable, cruel, dour, insufferable, braver, humane, law-worthy, patient, sincere, tolerant, trustworthy, truthful, upstanding, anxious, insane, astute, scholarly, self-educated, ignorant

[83] Sara Mills. 1995. Feminist stylistics. Roautledge London.

[84] Heiko Motschenbacher. 2013. Gentlemen before ladies? A corpus-based study of conjunct order in personal binomials. Journal of English Linguistics 41, 3 (2013), 212–242.

[85] Gülşen Musayeva Vefali and Fulya Erdentuğ. 2010. The coordinate structures in a corpus of New Age talks: "man and woman"/ " woman and man". Text & Talk-An Interdisciplinary Journal of Language, Discourse & Communication Studies 30, 4 (2010), 465–484.

[86] Sandra Mollin. 2012. Revisiting binomial order in English: Ordering constraints and reversibility. English Language & Linguistics 16, 1 (2012), 81–103.

[87] Carmen Rosa Caldas-Coulthard and Rosamund Moon. 2010. 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. Discourse & Society 21, 2 (2010), 99–133.

[88] Charlotte Taylor. 2013. Searching for similarity using corpus-assisted discourse studies. Corpora 8, 1 (2013), 81–113.

[89] Paul Baker. 2008. Sexed texts: language, gender and sexuality. Equinox.

[90] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasingword embeddings. In Advances in Neural Information Processing Systems. 4349–4357.

[91] Michael Pearce. 2008. Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. Corpora 3, 1 (2008), 1–29.

[92] Katherine Frith, Ping Shaw, and Hong Cheng. 2005. The construction of beauty: A cross-cultural analysis of women's magazine advertising. Journal of communication 55, 1 (2005),56–70.

[fig.8] Table 1: Gendered Personality Adjectives from the BNC

D Gender Bias in Artifficial Inteligence
D.2 Gender Bias in Machine Learning
D.2.2 Learning Bias from Text

[93] Caitlin Hines. 1999. Rebaking the pie: thewoman as dessert metaphor. Reinventing identities: The gendered self in discourse (1999), 145–162.

[94] Veronika Koller. 2004. Businesswomen and war metaphors: 'Possessive, jealous and pugnacious'? Journal of Sociolinguistics 8, 1 (2004), 3–22.

[95] Wendy Martyna. 1978. What Does 'He' Mean? Use of the generic masculine. Journal of communication 28, 1 (1978), 131–138.

[96] Suzanne Romaine et al. 1998. Communicating gender. Psychology Press.

[97] Pedro A Fuertes-Olivera. 2007. A corpus-based view of lexical gender in written Business English. English for Specific Purposes 26, 2 (2007), 219–234.

[98] Omar Ali, Ilias Flaounas, Tijl De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. 2010. Automating news content analysis: An application to gender bias and readability. In Proceedings of the FirstWorkshop on Applications of Pattern Analysis. 36–43.

d) Metaphor

Metaphor is difficult to identify automatically but is a powerful tool in the construction of gender in society[83, 93, 94]. In an endeavor to systematize the identification of metaphors in text[95] outlined five steps that could be applied to linguistic features of a text to identify whether their use was metaphorical or not. Research on the kind of metaphors used to portray men and women has identified a gender bias whereby those metaphors used to portray women are "more prolific and more derogatory than those used exclusively for men". [93, 94]

e) Presence of Women in Text

Straightforward frequency counts of women in text can be a powerful indicator of gender bias. In the British National Corpus, 'Mr' occurs more often than 'Mrs', 'Miss' and 'Ms' combined[96]. Furthermore, mentions of individual men, as distinct from mentions of men as a general category, occurred twice as often as mentions of individual women[91]. In an analysis of business literature[97], also found that mentions of men occurred 10 times more often than mentions of women and that of the total mentions of terms of terms of address (including Mr, Ms, Mrs, and Miss), 93.5 percent were occurrences of 'Mr'. In a study of 3.5 million articles from British newspapers, automated methods were devised to identify the gender of subjects referenced in newspaper articles[98]. It was found that men were referenced in 49 percent of top stories while women were referenced in 18 percent. Based on this, a simple quota system for the gender balance in in training data for machine learning algorithms may serve to combat much of the latent bias in text based sources of training data.

D Gender Bias in Artifficial Inteligence

D.3 Word Embeddings[d]

[d] Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai (2016), Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with word embedding, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. (...) This raises concerns because their widespread use, as we describe, often tends to amplify these biases.

Word embeddings, trained only on word co-occurrence in text corpora, serve as a dictionary of sorts for computer programs that would like to use word meaning. First, words with similar semantic meanings tend to have vectors that are close together. Second, the vector differences between words in embeddings have been shown to represent relationships between words.[99, 100]

For example given an analogy puzzle: “man is to king as woman is to x” (denoted as man:king :: woman:x), simple arithmetic of the embedding vectors finds that x=queen is the best answer because:

man — woman ⟶ king — queen

However, the embeddings also pinpoint sexism implicit in text. For instance, it is also the case that:

man — woman ⟶ computer-programmer — homemaker

[99] H. Rubenstein and J. B. Goodenough. Contextual correlates of synonymy. Communications of the ACM, 8(10):627–633, 1965.

[100] T. Mikolov, W.-t. Yih, and G. Zweig. Linguistic regularities in continuous space word representations. In HLT-NAACL, pages 746–751, 2013.

D Gender Bias in Artifficial Inteligence

D.3 Word Embeddings

[fig.9] The most extreme occupations as projected on to the she he gender direction on g2vNEWS.Occupations such as businesswoman, where gender is suggested by the orthography, were excluded.

[fig.10] Analogy examples. Examples of automatically generated analogies for the pair she-he using the procedure described in text. For example, the first analogy is interpreted as she:sewing :: he:-carpentry in the original w2vNEWS embedding. Each automatically generated analogy is evaluated by 10 crowd-workers are to whether or not it reflects gender stereotype. Top: illustrative gender stereotypic analogies automatically generated from w2vNEWS, as rated by at least 5 of the 10 crowd-workers. Bottom: illustrative generated gender-appropriate analogies.

Extreme *she* occupations

1. homemaker	2. nurse	3. receptionist
4. librarian	5. socialite	6. hairdresser
7. nanny	8. bookkeeper	9. stylist
10. housekeeper	11. interior designer	12. guidance counselor

Extreme *he* occupations

1. maestro	2. skipper	3. protege
4. philosopher	5. captain	6. architect
7. financier	8. warrior	9. broadcaster
10. magician	11. figher pilot	12. boss

Gender stereotype *she-he* analogies.

sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber

Gender appropriate *she-he* analogies.

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

In other words, the same system that solved the above reasonable analogies will offensively answer “man is to computer programmer as woman is to x” with x = homemaker. Similarly, it outputs that a father is to a doctor as a mother is to a nurse.

The analogies generated from these embeddings spell out the bias implicit in the data on which they were trained. Hence, word embeddings may serve as a means to extract implicit gender associations from a large text corpus similar to how Implicit Association Tests[101] detect automatic gender associations possessed by people, which often do not align with self reports.

To quantify bias, we compare a word embedding to the embeddings of a pair of gender-specific words. For instance, the fact that nurse is close to woman is not in itself necessarily biased (it is also somewhat close to man — all are humans), but the fact that these distances are unequal suggests bias. To make this rigorous, consider the distinction between gender specific words that are associated with a gender by definition, and the remaining gender neutral words. Standard examples of gender specific words include brother, sister, businessman and businesswoman. The fact that brother is closer to man than to woman is expected since they share the definitive feature of relating to males.

[101] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition: the implicit association test. Journal of personality and social psychology, 74(6):1464, 1998.

D Gender Bias in Artifficial Inteligence

D.3 Word Embeddings

Stereotypes are biases that are widely held among a group of people. We show that the biases in the word embedding are in fact closely aligned with social conception of gender stereotype. The crowd agreed that the biases reflected both in the location of vectors (e.g. doctor closer to man than to woman) as well asin analogies (e.g., he:coward :: she:whore) exhibit common gender stereotypes.

(Site) www.2021.fbaul-dcnm.pt/deconstructing-gender-bias/

E Conclusion

[102] Susan Leavy (2018), Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. p.16

[103] Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang, Kai-Wei Chang August (2018), Learning Gender--Neutral Word Embeddings. p.15

“Identifying gender bias in training data for machine learning algorithms is a complex but not an insurmountable task. (...) While the fact that machine learning algorithms can learn gender bias can be of interest to researchers looking to understand its prevalence in society, it is not an advantage in practical applications making decisions about people’s lives. There is an emerging focus on fairness in machine learning generally and it is essential that women are at the core of who defines the concept of fairness. Advancing women’s careers in the area of Artificial Intelligence is not only a right in itself; it is essential to prevent advances in gender equality supported by decades of feminist thought being undone.”[102]

“One perspective on bias in word embeddings is that it merely reflects bias in society, and therefore one should attempt to debias society rather than word embeddings. However, by reducing the bias in today’s computer systems (or at least not amplifying the bias), which is increasingly reliant on word embeddings, in a small way debiased word embeddings can hopefully contribute to reducing gender bias in society. At the very least, machine learning should not be used to inadvertently amplify these biases, as we have seen can naturally happen”.[103]

Authors Joana Pereira
Matilde Dias

Date May 2021

Context Projeto II / Masters in Communication Design. Faculty of Fine-Arts, Lisbon

Teachers Luísa Lopes Ribas
Pedro Ângelo