

Kirsten Salpukas

Dec 16, 2022

GSWS 1200 - Trans Studies

Final Project

Transgender Exclusion within Language Processing

NLP: “Natural Language Processing”; or “Natural Lawful Prejudice”?

Significant technological advances have been made over the last couple of decades with seemingly neverending new devices and automation and with that has come a new age of how we process information. Long gone are the times when translations had to be done by hand or customer service questions did not first go through an automated chatbot and in its place has emerged the field of Natural Language Processing. Because Natural Language Processing has become such a large and all-encompassing field that has influenced so much of people’s everyday life it has forced people to start looking more closely into how it operates, and, in particular, how it treats different groups of people.

One field that has started to have an influx of attention and research associated with it is that of bias, especially gender bias. While this is an extremely important and relevant subject to be researching, many of the texts surrounding it seems to fall flat when it comes to anyone that does not fall under the umbrella of cis-gender. . When looking at gender bias it is found that “the majority of the articles do not make their theorization of gender explicit, even if they clearly define “bias” (Devinney). Therefore, it is immensely necessary to be looking at how to examine how gender is treated within NLP while not only considering people who are not cis-gendered men and women. Transgender identities are threatened with erasure by NLP and failing to instigate some form of change would mean letting NLP continue to function as a way to maintain the gender binary and uphold incorrect and outdated ideas of gender identity.

Discussion of Gender

To understand why there needs to be such a discussion surrounding the topic of gender we must first understand gender. Gender is typically understood as a separate entity from biology however, looking at how gender is approached today unveils that this thinking is derived from medicine rather than feminist theory. Gill-Peterson in her work “Gender” argues that gender is “an incredibly recent invention” as the medicine connected to it is linked to the mid-twentieth century and anything before that can be attributed to “the category ‘sex’” (Gill-Peterson, 97). However, looking at gender as separate from sex “introduced a cultural concept to critique sex stereotypes, sexed hierarchies, and power imbalances between women and men” which was widely utilized within feminism, likely contributing to the misconception of gender being a byproduct of feminist theory (Gill-Peterson, 98). The concept of gender is important to understand as much of NLP seeks to categorize text using preconceived notions surrounding sex or gender in order to complete tasks such as recognition of names and faces, or translation. This is an issue because, as explained by Gill-Peterson, if gender is a recent invention then it should not be used as a baseline identifier across all data points as it is not an objective measurement. Therefore, using it as such creates a system that is not applicable to all data yet is being used on all data.

What Is NLP?

Natural Language Processing (NLP) refers to a subset of Artificial Intelligence (AI) that is interested in having computers be able to understand and interact with text and speech. However, because the human language is comprised of so much vagueness such as metaphors, acronyms, sarcasm, etc, NLP must use techniques such as tagging parts of speech, recognition of speech, and recognition of different entities. Therefore, the primary goal is to build systems that

can respond with their own text or speech in a similar way that a human would. As a result, this subject is a combination of many different fields such as linguistics, machine learning, and deep learning. In regards to programming, NLP uses mostly the Python programming language as it has a lot of built-in libraries to approach specific NLP actions such as the Natural Language Toolkit, or NLTK, which is an open-source assemblage of tools for NLP. This tool uses libraries that contain the utilization of tasks such as segmentation, tokenization, parsing, lemmatization, etc, which helps to speed run the NLP process.

NLP is the primary force behind systems that are in charge of translating text between languages, reading and compiling text, and understanding human speech. Most people use some form of NLP in their everyday lives without thinking about it with tools such as search results, Email filters, autocorrect, and smart assistants such as Amazon's Alexa and Apple's Siri. However, besides just its everyday use, NLP has also become an extremely important role in how companies approach business ventures such as simplifying business processes and increasing production and profit.

Why Is There a Problem to Begin With?

As with most problems concerning Artificial Intelligence, the bulk of the troubles lies within the data, and in this case, with a lack of data. Within the field of AI, there is an emphasis on creating something for a specific purpose, and therefore, when a model is being built the focus is on how to accomplish that one functionality rather than the humans who will use it. This then leaves a rather gaping human-sized hole in that approach as humans are incredibly complex and such models would benefit from input from the people who will actually use the product.

Within NLP the information that it receives to teach it about the humans it will interact with typically comes from academic sources such as academic journals, books, and mainstream

newspapers such as the Wall Street Journal and The New York Times. Having NLP focus on these types of texts can get tricky when it comes to trans work as academia “is a site of both possibility and constraint for trans people” (Goldberg, 1). On one hand, there is a great deal of opportunity for the production of knowledge as, “across areas of study, the development of scholarly communities within academia has increased knowledge about trans lives and led to trans experiences being addressed in affirming, life-giving ways (Goldberg, 1). This can be extremely useful within NLP because, as mentioned above, a major source of data comes from a traditional academic source. However, a lot of information surrounding trans-ness and trans experiences is not found in these ‘academic’ settings but rather through more user-based platforms (i.e. Reddit, Twitter, and Blogs). Even looking into more traditional user-based text such as journals proves to be a problem as “few journals focus on trans experiences” (Golberg, 1). Within the aforementioned sources typically used for training, if the term “transgender” is mentioned, it’s likely in a medical circumstance and not conducive to teaching any person, let alone a computer, how to interact with transgender individuals. People may be able to read these more scholarly trans texts and infer what that could mean as far as lived experiences for trans people but a computer is only as good as the data it is given. It does not have the human experience, emotions, or empathy to be able to draw further conclusions about how these sources influence transgender people’s everyday life. The only way for it to gain that insight is to feed it sources that talk about real-life experiences.

Another issue surrounding NLP is the subject of the authors and researchers themselves. A 2019 report found that “only about 29% of first authors are female and only about 25% of last authors are female” stating, “notably, this percentage has not improved since the mid-2000s”. This means that a majority of the people in the field are cis-gendered males. Golberg comments

on the ways in which authorship matters within trans studies saying “one potent possibility of Trans Studies is the way the field shifts the narrative of mostly cis authors writing about trans lives to trans people creating knowledge, histories, and culture about their own communities” (Goldberg, 2). Having trans people be scribes and archivists means there is a much greater chance for academia to have work that encompasses trans experiences. This relates to NLP in the sense that those coding and researching are the equivalent of authors and archivists and having no trans individuals within that space likely means excluding the trans voice in the design process for something they will still be using.

The Strict Gender Binary within NLP

While there has been an increasing number of studies and research being done to examine the gender bias within NLP, in contrast, that energy is not being carried over into investigating the gender binary. A bias refers to when something is repeatedly wrong in the same way, so in this case, a gender bias is when something is incorrect based on the perceived gender of the subject. A real-world example would be assuming that girls prefer the color pink and boys prefer the color blue. Assuming this as a fact would mean that you are ignoring all possibilities of girls preferring blue and boys preferring pink, therefore limiting your success rate with that data. It is very well-documented that within NLP there seems to be a strong gender bias in which models tend to lean into certain gender roles. One such example of this bias is when translating the sentence “He is a nurse. She is a doctor.” into Hungarian and back again you get “She is a nurse. He is a doctor” instead (Douglas). While this is certainly an issue it is interesting to see that within these papers surrounding gender there seems to be no mention of how they are actually looking at gender. A survey of “nearly 200 articles concerning gender bias in NLP” found that “the majority of the articles do not make their theorization of gender explicit, even if they clearly

define “bias.”. Going further than that, when looking into the actual models used within these papers “almost none use a model of gender that is intersectional or inclusive of nonbinary genders; and many conflate sex characteristics, social gender, and linguistic gender in ways that disregard the existence and experience of trans, nonbinary, and intersex people” (Devinney). That is to say that the existence of trans individuals within these papers is effectively erased.

This idea of trans erasure is unfortunately not new, Betancourt in their work, “Where Are All the Trans Women in Byzantium” explores this same phenomenon as “the image of women assigned male at birth is all but absent in Byzantine sources” (Betancourt, 297). While this work is about trans erasure it is not focused on the ‘whys’ but rather on “where might we find the traces of these lives by shifting our methodological perspectives” (Betancourt, 297). Similarly, we have already discussed above why there seems to be this problem within NLP so now we must apply our own shifting of perspectives to figure out how to find our own traces. Betancourt discusses how his goal is to “accept that transgender persons are not a modern phenomenon but have existed in the shadows throughout history” and have that influence our reading of Byzantine sources (Betancourt, 299). He is asking us to read texts as they were meant to be read at the time of being written rather than immediately applying our modern knowledge and understanding. The actual language itself must be examined separately from the modern conclusions gathered because it is telling us more about the original culture that surrounded it. He explains that “here my aim is to understand the various glimmers and discussions of women who had been assigned male at birth as medieval antecedents to modern trans lives. Articulating these medieval women as trans allows us to better understand how Greek writers understood, construed, and imagined gender” (Betancourt, 298).

Using this same method with NLP we can look at results that strictly place people into the binaries of men and women and understand that it is only able to do that because it is reading language as a tool to be used to display gender. We can understand gender, not as this rigid binary that is decided for you, but rather as something that is shown through the performativity of your gender. Language is such a key component of this performance and any interaction of NLP and gender reaffirms this connection and creates a space without the binary.

Reaffirming the Existence of Cisnormativity

A large part of the issue with the way that gender is currently treated within NLP is the idea of cis-normativity. Cis-normativity refers to the concept that everyone is cisgender and anything that falls outside of the idea of cisgender is not significant. Even though we as scholars accept the concept that not everyone falls under this cisnormative assumption, there is a large population that adheres to this idea.

Bychowski explains in their work, “The Transgender Turn: Eleanor Rykener Speaks Back” that “most scholarship assumes the cisgender status of any character or historical figure who is presented to readers” (Bychowski, 95). Therefore, assuming that everyone is cisgender, anyone who is not cisgendered becomes “coded by cisgender norms” (Bychowski, 95). NLP does not escape this assumption as there is a “prevalence of cisnormative assumptions” which are said to “constrain thinking about gender” (Devinney). This is especially worrying as this statement comes from a study done on articles about gender bias within NLP so if articles focused on gender still fall under the trap of cisnormativity it would stand to be even worse within the general field.

By assuming that all work is cis-gendered work, it is ignoring the possibility of anyone being outside of that ideology, and therefore, anyone belonging to an identity outside of

cis-gender. It also relates to how the depiction of gender identity on paper impacts the ability to move through life as a trans individual. If your gender identity on paper does not match that of your gender identity in real life then you will face heavy scrutiny and prejudice because of that label.

The (Mis)Treatment of Gender Identity

We have reached a point in the age of technology where many processes are now automated using Artificial Intelligence such as NLP. Therefore, everyone's lives are impacted by the results that it produces, some more so than others. Some of the harms associated with the way that NLP approached gender include "resume scanning systems throwing out resumes from non-binary persons for not having a recognizable name, undercounting instances of discrimination against non-binary persons, and incorrectly processing gender in medical documents, leading to incorrect care" (Dev, 6). While this is certainly a modern situation it is not a solely modern problem as we have seen the way that gender is viewed by those in charge of procedures create issues. The work "If Sex is Not a Biological Phenomenon" discusses gender changes on birth certificates and how people's "ability to move through the world" as their preferred gender "would inevitably be compromised" without the swap in gender (Currah, 31). All of these problems inform us of how important it is for gender identity to be correctly recognized.

Without that recognition, we are left with gender identities that do not match with those that are displayed, and therefore individuals' gender identities are invalidated. This is especially concerning as most sex changes relied on the fact that "individuals would have to prove not only that their genitals had been reconstructed (phalloplasty for men or vaginoplasty for women) but also that they had been sterilized through surgery (a hysterectomy for those assigned as female at

birth, an orchiectomy for those assigned male at birth)” (33). Maintaining the idea that your gender identity is directly related to your genitals is especially problematic because it is not necessary to require any genital surgery in order to validate your gender identity. It should not be required that your genitals are a subject of investigation because that is no one’s business but your own. Genitals do not define gender identity and anything saying otherwise is problematic.

Computer Code Is Binary, Gender Is Not

The question when investigating this topic is why does it matter? Why do we care about how gender is treated with Natural Language Processing? This is an important topic because it impacts everyone. It is such a widely spread concept that it is nearly impossible to escape and it has large impacts.

Models that are starting to be trained have issues such as mistagging “they” as non-person entities and not being able to distinguish when “they” is being used as a singular pronoun. Also any neopronouns such as “ze” and “zir” are not recognized by modern NLP models (Dev).

It is not just gender identity that is being affected. While this is an extremely important topic surrounding gender identity there is far more that goes into it. Any system created using NLP is subject to prejudice connected to those that create it. If the creators are all cis-gendered white men then they will likely not consider how it affects non-cis-gendered non-white men. Therefore all minorities are at risk of being overlooked within NLP models,

We can no longer approach the building of Artificial Intelligence systems as exclusively for data extraction or syntactical analysis when our data relates to people because people are not computer code and do not always behave in an “expected” way. Doing so will only further the already rampant social inequalities faced by society at large. The issue is there is not normally

any financial or social incentive for those creating these systems to create them fairly. The people creating these systems are often the ones benefiting from them and as such, there must be pressure coming from outside sources in order to enact any change.

Therefore it is so important to be having this discussion because understanding how NLPs behave in regard to transgender and non-binary individuals creates a new discourse on the interaction of technology and gender identity. This intersection is so lacking within the current conversation as most studies just ignore the possibility of anyone existing outside of the binary, further feeding into this narrative that anything outside of cis-gender is not worth including.

Works Cited

Goldberg, A. E., & Beemyn, G. (2021). Academia. In *The SAGE encyclopedia of trans studies*. essay, SAGE Reference.

Currah, Paisley. "1 "If Sex Is Not a Biologic Phenomenon". *Sex Is as Sex Does: Governing Transgender Identity*, New York, USA: New York University Press, 2022, pp. 31-53.

Bychowski, M. W. (2021). The Transgender Turn: Eleanor Rykener Speaks Back. In G. LaFleur, M. Raskolnikov, & A. Kłosowska (Eds.), *Trans historical: Gender plurality before the modern* (pp. 95–113). essay, Cornell University Press.

Betancourt, R. (2021). Where Are All the Trans Women in Byzantium? In G. LaFleur, M. Raskolnikov, & A. Kłosowska (Eds.), *Trans Historical: Gender Plurality before the Modern* (pp. 297–321). essay, Cornell University Press.

Gill-Peterson, J. (2021). Gender. *Keywords for Gender and Sexuality Studies*, 96–101. <https://doi.org/10.18574/nyu/9781479808168.003.0034>

Douglas, Laura. 2017. AI is not Just Learning our Biases; It Is Amplifying Them. <https://bit.ly/2zRvGhH>.

Dev, S., Monajatipoor, M., Ovalle, A., Subramonian, A., Phillips, J., & Chang, K.-W. (2021). Harms of gender exclusivity and challenges in non-binary representation in Language Technologies. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/2021.emnlp-main.150>

Devinney H., Björklund, J., & Björklund, H. (2022). Theories of “gender” in NLP Bias Research. *2022 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3531146.3534627>