Kirsten Salpukas

Dec 4, 2022

GSWS 1200 - Trans Studies

Final Project Draft

Intro

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.

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. 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.

Discussion of Gender (Gender by Jules Gill-Peterson)

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). This is important to understand as much of NLP seeks to categorize text using the idea of sex or gender.

Why is there a problem to begin with? ("Academia")

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 more academic sources such as the Wall Street Journal and Wikipedia. Having NLP focus on academic 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, the major source of data comes from a traditional academic site. 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 sites 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 cannot infer information that it does not have.

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 ("Where Are All the Trans Women in Byzantium")

While there are many articles and research done to examine the gender bias within NLP there is almost none being done to examine the gender binary. 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.

Similarly, 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 ("The Transgender Turn: Eleanor Rykener Speaks Back")

A large part of the issue with the way that gender is currently treated within NLP is the idea of cisnormativity. 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.

The (mis)treatment of gender identity ("If Sex is Not a Biological Phenomenon")

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.

Why does it matter?
- More real-world examples:
  - 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 not recognizing neopronouns

Broader Scale
- Minorities & tech interaction
  - Where else do we see issues?

Solution?

Conclusion

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