

# Dehumanization of LGBTQ+ Groups in Sexual Interactions with ChatGPT

Alexandria Leto Juan Vásquez Alexis Palmer Maria Leonor Pacheco

University of Colorado Boulder

{alexandria.letto, juan.vasquez-1, alexis.palmer, maria.pacheco}@colorado.edu

## Abstract

Given the widespread use of LLM-powered conversational agents such as ChatGPT, analyzing the ways people interact with them could provide valuable insights into human behavior. Prior work has shown that these agents are sometimes used in sexual contexts, such as to obtain advice, to role-play as sexual companions, or to generate erotica. While LGBTQ+ acceptance has increased in recent years, dehumanizing practices against minorities continue to prevail. In this paper, we hone in on this and perform an analysis of dehumanizing tendencies toward LGBTQ+ individuals by human users in their sexual interactions with ChatGPT. Through a series of experiments that model various concept vectors associated with distinct shades of dehumanization, we find evidence of the reproduction of harmful stereotypes. However, many user prompts lack indications of dehumanization, suggesting that the use of these agents is a complex and nuanced issue which warrants further investigation.

## 1 Introduction

Large Language Models (LLMs) have undoubtedly changed the way people interact with their computing hardware and software. These tools have seen widespread adoption in recent months to help write research papers (Liang et al., 2024), improve the individual journaling experience for better mental health (Nepal et al., 2024) and help with creative writing tasks (Grigis and Angeli, 2024), among other activities. Therefore, examining human interactions with LLM-powered conversational agents such as ChatGPT may provide valuable insights into human behavior (Zhou et al., 2024).

Sometimes the use cases of LLM-powered conversational agents are sexual. For example, they are used to obtain sexual information and advice, as romantic or sexual companions, and to generate erotica and pornography (Döring et al., 2024). As evidence of this, Mireshghallah et al. (2024)

labeled a sample of  $5k$  user-generated ChatGPT prompts, reporting that 6.3% of the inputs contained sexual and erotic content.

Past work has analyzed porn consumption as a means to study sexual behavior (Hald, 2006; Harvey, 2020). Doing so can evade the limitations of more traditional sexual studies such as experimentation, which demands time-consuming planning to comply with ethical guidelines (Pearson and Curtis, 2025). Like studying porn, analyzing sexual conversations between users and conversational agents may be a viable way to study sexual behavior. Because this type of study doesn't require deploying surveys—the traditional method for studying porn consumption—it may surpass the advantages of studying porn in terms of scalability and ease of study. It also has the added advantage of rich textual data exemplifying the ways people think and talk about sex. This type of sexual experience will be more tailored to the user's particular wants than porn, and therefore potentially more informative to study.

It has long been recognized that, though porn can serve as a safe space for the queer community (Flory, 2024), the representation of LGBTQ+ people in mainstream porn can be problematic (Harvey, 2020). For example, transgender people may be dehumanized by being overly-objectified (Anzani et al., 2021). Lesbian representation in mainstream porn largely caters to the male gaze, leading to fetishization (Smyth, 1990; Collins, 1998; Webber, 2013). Though generally thought of as less problematic, male gay porn has been critiqued for reinforcing homophobia (Corneau and van der Meulen, 2014).

Like porn, erotic conversational agents have their advantages for queer populations, such as providing emotional support (Lissak et al., 2024). However, we seek to understand whether—in spite of these positive effects—problematic representations of LGBTQ+ people observed in porn can also be

found in user prompts with conversational agents. To this end, the main contributions of this work include: (1) an analysis of the way users represent LGBTQ+ people in sexual interactions with ChatGPT using an extended version of Mendelsohn et al. (2020)’s pre-existing framework for identifying dehumanizing language and (2) a discussion of our findings that motivates further exploration of sexual interactions with LLM-powered conversational agents.

## 2 Related Work

In this section we cover the background on LLMs and sexuality, discuss NLP for queer sociolinguistics and cover prior dehumanization work.

### 2.1 LLMs and Sexuality

Erotic conversational agents are increasing in popularity; male and queer audiences are common early adopters (Gesselman et al., 2023). Döring et al. (2024) identified several sexual use-cases for these agents, such as for sexual education and therapy. Other work in this vein has evaluated LLMs’ understanding of sexual consent (Marcantonio et al., 2023), sexual medical information (Seyam et al., 2024; Caglar et al., 2023), their ability to act as therapists (D’Souza et al., 2023; Vowels, 2024) and their biases (Organization, 2022; Dhingra et al., 2023; Wan et al., 2023; Kotek et al., 2023). Though many studies focus on the LLM component of these interactions, few have examined the phenomena as a way to understand human sexuality. As Pearson and Curtis (2025) argue, this could be a rich research opportunity (Hald, 2006).

### 2.2 NLP for Queer Sociolinguistics

Several works have used NLP techniques to analyze linguistic phenomena related to the LGBTQ+ population. Andersen et al. (2024) analyzed the Twitter discourse around the Mexican Spanish-speaking LGBTQ+ community over ten years. By mapping how the polarity of some nouns related to the LGBTQ+ community has evolved in conversational settings, the authors found that, on average, the analyzed tweets had a negative polarity. Furthermore, the authors revealed that the nouns related to the trans community have seen the greatest increase in usage for the time range and subgroups represented in their corpus. Locatelli et al. (2023) performed a cross-lingual analysis of LGBTQ+ discourse on Twitter across seven languages during the 2022 Pride month. Their results indicate that

*homotransphobia* is a global problem that takes on distinct cultural expressions. In line with the paper presented by Andersen et al. (2024), Locatelli et al. (2023) found that derogatory language toward LGBTQ+ people is present in the seven languages they studied while being especially prevalent in Italian and French.

### 2.3 Dehumanization and Language

Haslam (2006) presents two types of dehumanization explored in prior work: animalistic (likening a target group to animals) or mechanistic (treating a target group as machines or inanimate objects). For example, Tutsis in Rwanda were explicitly compared to cockroaches in propaganda leading up to the 1994 genocide (Harris and Fiske, 2011). They propose that these types of dehumanization can also occur in subtler cases in which groups are not sufficiently attributed human qualities. For example, feminist work has discussed dehumanization in porn through the sexual objectification of women (Zhou et al., 2021); these women are stripped of human qualities such as emotionality. Cascalheira and Choi (2023), in their study of dehumanization of transgender people, echo that sexual objectification is an important element of dehumanization and can have negative impact on mental health (Anzani et al., 2021).

Mendelsohn et al. (2020) built on Haslam (2006) to present a computational framework for identifying dehumanizing language. They analyzed mentions of LGBTQ+ individuals in the New York Times over 30 years (1986 to 2015), finding decreasing association of LGBTQ+ groups with dehumanizing elements such as vermin metaphors and moral disgust. Giorgi et al. (2023) used this framework for their analysis of the dehumanization of those who use substances in U.S. news media. Burovova and Romanyshyn (2024) take another approach in their analysis of the dehumanization of Ukrainians on Russian Social Media, using a sentence-level binary classifier to identify dehumanization. To our knowledge, this work is the first to leverage this type of computational framework to examine dehumanization in erotic content.

## 3 Data

We conduct our analysis on a filtered portion of the WildChat-1M-Full dataset (Zhao et al., 2024),<sup>1</sup>

<sup>1</sup><https://huggingface.co/datasets/allenai/WildChat-1M-Full?not-for-all-audiences=true>

which contains 1 million conversations between users and OpenAI’s GPT-3.5 and 4. We filter the dataset to only include conversations in English which have been marked by the included OpenAI Moderation results as containing sexual content. Because we are interested in studying how *humans* characterize those from LGBTQ+ communities (rather than how LLMs do), we remove all system responses from the set. The resulting dataset contains approximately 38 thousand unique user turns from sexual conversations.

To identify mentions of various groups, we build a lexicon based on the Textual Identity Detection and Augmentation Lexicon (TIDAL),<sup>2</sup> a dataset formed to enable automatic detection of identity labels (Klu and Sethi, 2023). It has coverage for identity groups including race, nationality, ethnicity, sexual orientation, gender identity, gender expression, sex characteristics and religion; it includes slurs. Because our vector approaches (discussed in Section 4) rely on word embeddings, we only consider single-word nouns. Based on these words, we formed lexicons to identify mentions for seven groups: LGBTQ+, LGB, Transgender, Gay Men, Lesbian, Bisexual, and Heterosexual. Word lists for each group are included in Appendix A.1.

## 4 Measuring Dehumanization

Mendelsohn et al. (2020) present a structure for analyzing dehumanization in media. They outline several elements of dehumanization from prior social psychology literature including: a) negative evaluation of a target group, b) moral disgust, and c) vermin as a dehumanizing metaphor. We follow their framework for analyzing each element and add an additional component: d) objectification.

### (a) Negative Evaluation of a Target Group

Prior work shows that negative evaluations of a target group contribute to the dehumanization of the group (Haslam, 2006). To quantify this, we complete a simple per-sentence sentiment analysis with the SiEBERT model (Hartmann et al., 2023).<sup>3</sup> To obtain a sentiment score for each group label, we average over the scores for every sentence containing a term associated with the group.

**(b) Moral Disgust** To identify moral disgust associated with each group—another indicator of

<sup>2</sup><https://github.com/google-research-datasets/TIDAL>

<sup>3</sup><https://huggingface.co/siebert/sentiment-roberta-large-english>

dehumanization—we lean on lexicons created by Graham et al. (2009) for each dimension from Moral Foundations theory (Haidt and Graham, 2007). Specifically, we use all words from the “moral disgust lexicon”, which includes about 80 words such as *obscene*, *sin*, and *sick* (see full lexicon in Appendix A.1). We train a word2vec skip-gram model<sup>4</sup> on our dataset to create a moral disgust “concept vector” by averaging the embeddings for all moral disgust words (Mikolov et al., 2013). Then, we measure the cosine distance between this concept vector and concept vectors for all groups. We compare these distances to a “neutral” concept vector, constructed by averaging the embeddings for the words *person*, *people*, *individual*, and *individuals*.

**(c) Vermin Metaphor** Another way people are dehumanized is when they are robbed of human traits and attributed those of animals such as vermin (Haslam, 2006). To measure the association of different groups with vermin in the dataset, we repeat our method for measuring moral disgust—we create a vermin concept vector by averaging the word embeddings for vermin words (*vermin*, *rodent(s)*, *rat(s)*, *mice*, *cockroach(es)*) and measure cosine distance between this vector and the concept vectors for the group labels.

**(d) Objectification** Haslam (2006) discusses mechanistic dehumanization, or treating a target as an inanimate object. This is related to sexual objectification, and thus pertinent to our analysis. We measure the association of different groups with objects in the dataset. We repeat our method for measuring vermin metaphor and moral disgust—we create an object concept vector by averaging the word embeddings for “object” words including *it*, *that*, *this*, *thing*, *things*, *object*, *objects*, *item*, *items*, *machine*, and *machines*. Then we measure the cosine distance between this and the group concept vectors.

## 5 Dehumanization of LGBTQ+ People in Wildchat

In our analysis of the dehumanization of LGBTQ+ people in user prompts in the sexual subset of the Wildchat dataset, we unveil a number of interesting insights which point to diverse portrayals of

<sup>4</sup><https://radimrehurek.com/gensim/models/word2vec.html>

LGBTQ+ groups in human interactions with LLM-powered agents.

**LGBTQ+ people are objectified less than “neutral” terms.** In Figure 1, we find evidence that LGBTQ+ terms are used in less similar contexts to object words than the predetermined “neutral” terms. The Vermin Metaphor results (shown in Figure 6), show that *all* LGBTQ+ word groups are, on average, used in less similar context to vermin words than the “neutral” terms. In Figure 7, we find that some of the group vectors have a smaller cosine distance to the “Moral Disgust” concept vector, while others (including “Bisexual” and “Gay Men”) have slightly longer distances. Collectively, these results seem to indicate that these groups are characterized with similar levels of, or less, dehumanizing language than the terms *person*, *people*, *individual*, and *individuals*.

**Dehumanization is relatively similar across groups.** The “vermin”, “object”, and “moral disgust” vectors exhibit similar cosine distances to the vectors associated with each subgroup in the LGBTQ+ acronym. While some groups may be dehumanized more, the differences are not stark. This indicates that on average, when people ask ChatGPT to generate text with LGBTQ+-related erotic content, they tend to dehumanize every subgroup in a similar measure.

However, some terms, such as “twink” and “fag-got”, exhibit the longest cosine distance from the moral disgust concept vector when compared to all the other terms in the LGBTQ+ category, indicating a semantic shift in these terms. One possible reason for this is that societal acceptance of gay men has led some ChatGPT users to view them in a more positive and sexually desirable way. The same cannot be said for other LGBTQ+ groups.

**Transgender and lesbian populations are dehumanized more than other groups.** In line with patterns in mainstream porn, we find that lesbian and transgender people are stripped of human qualities more often than other groups. In Figure 7 we observe that the “Transgender” group vector is more closely related to the moral disgust concept vector than any other group vector. The “Lesbian” group vector is also situated nearer the “moral disgust” concept vector than the “neutral” vector. These groups also face more negative evaluations, as shown in Figure 2.

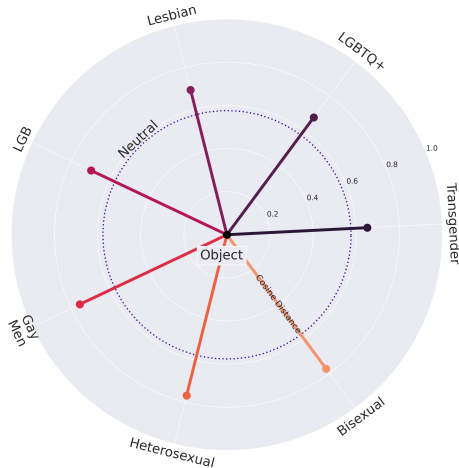


Figure 1: Cosine distance from Object Concept Vector to each subgroup. An average distance over all terms in the category is shown.

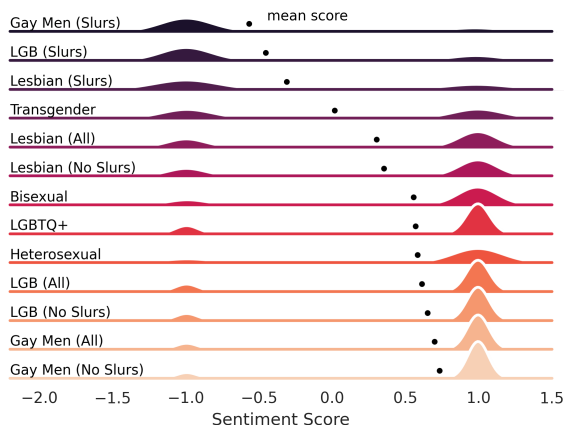


Figure 2: Distribution of sentiment scores for each group. Mean score is calculated over the sentiment labels for sentences containing a term in the group.

**Outdated terms are used in more dehumanizing contexts.** In line with Mendelsohn et al. (2020)’s findings in their analysis of dehumanization in U.S. news media, outdated terms used to refer to people from the LGBTQ+ community are used in more dehumanizing ways than other terms, such. For example, the terms “homosexual” and “hermaphrodite” are generally considered to be outdated and offensive.<sup>5</sup> These words are more closely associated with the vermin, object, and moral disgust concept vectors than other LGBTQ+ terms (Figures 3, 4, and 8).

**Some interactions represent LGBTQ+ people in more positive ways.** Finally, we remark that the various concept vectors we model here—which

<sup>5</sup><https://glaad.org/reference/terms/>



Term(s)	User Prompt	Dehumanization
Twinks	The twinks and Shrek are relaxing in bed au naturel and snacking on roasted fish (that died from Shrek farting in a pond during his morning bath with the twinks, and he later cooked up for them) as they chat while Shrek couples with one of them (describe Shrek’s physique and butt).	Some
Lesbian	She very slowly begins to realize she is attracted to girls, at first denying it, but very slowly accepting that she is a lesbian and eventually finds a girlfriend named Lola who is just like her, physically and mentally, and they begin dating, and after months of dating, on a date while walking through the park, she got on one knee and proposed, to which her girlfriend immediately responded with a yes.	No
Lesbian, Hetero	Write a chapter of 1000 words about a hot lesbian couple that love to do lesbian acts in front of hetero men.	Yes

Table 1: Selected examples of terms in user prompts.

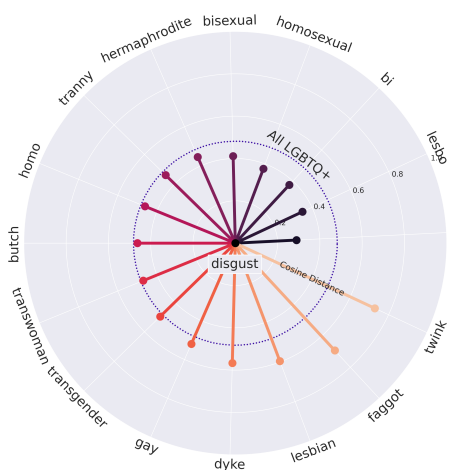


Figure 3: Cosine distance from the Moral Disgust vector to the vector of terms in the LGBTQ+ category.

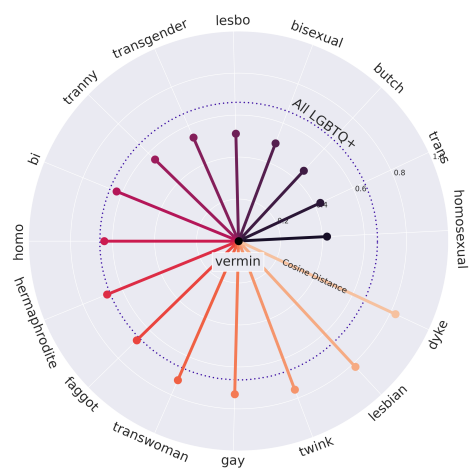


Figure 4: Cosine distance from the Vermin vector to the vector of terms in the LGBTQ+ category.

reveal dehumanizing language targeted at the LGBTQ+ population—indicate that not all interactions are dehumanizing LGBTQ+ people. This is further evidenced by some of the selected examples included in Table 1.

As a result, our analysis does not yield one single, binary conclusion about the ways in which LGBTQ+ are being characterized by users in interactions with AI-powered agents. Rather, it reinforces that the conversation around sexual interactions with these agents, much like around porn, must remain complex and nuanced. Though we find evidence of dehumanizing language in our quantitative and qualitative analysis—which seems to mirror the demand for this type of content in mainstream porn (Anzani et al., 2021; Webber, 2013)—we also find content that is more uplifting. We suggest that this *could* be further evidence that these agents are helpful tools actively being used by those who cannot find proper representation of their desires in mainstream media (Gesselman et al., 2023).

## 6 Conclusion

Motivated by our findings, we call attention to further studying how conversational agents facilitate or hinder marginalized groups’ representation in sexual contexts. We propose extending this work to explore representation across lines of race and gender. Our findings also suggest that it may be informative to expand our focus past dehumanizing language to study other possible modes of representation.

By understanding how people reproduce harmful stereotypes—or not—in their prompting practices of erotic text, we could develop safeguards to minimize the dehumanization of the LGBTQ+ community, while promoting the use of these tools among LGBTQ+ individuals who cannot find proper representation of their desires in traditional media.

## 7 Limitations

The findings reported in this paper are based on a subset of human-generated ChatGPT prompts.

Therefore, claiming that these can be generalized to every interaction with every commercial LLM is inaccurate. Further analyses with data gathered from a longer timespan and multiple LLMs could support or refute the claims we report here. Furthermore, given the time and computational constraints, we limit our studies to only a subset of the original WildChat-1M-Full dataset.

Another noteworthy limitation of this work is the language diversity in our data. We limited our experiments to prompts written in English, ignoring the existing epistemic diversity in human-agent interactions.

We also acknowledge the tendency of automatic sentiment analysis to ignore in-group mentions of slurs. Therefore, the sentiment scores may be negatively biased when slurs are present.

## References

- Scott Andersen, Segio-Luis Ojeda-Trueba, Juan Vásquez, and Gemma Bel-Enguix. 2024. [The Mexican gayze: A computational analysis of the attitudes towards the LGBT+ population in Mexico on social media across a decade](#). In *Proceedings of the 8th Workshop on Online Abuse and Harms (WOAH 2024)*, pages 178–200, Mexico City, Mexico. Association for Computational Linguistics.
- Annalisa Anzani, Louis Lindley, Giacomo Tognasso, M. Paz Galupo, and Antonio Prunas. 2021. [“being talked to like i was a sex toy, like being transgender was simply for the enjoyment of someone else”: Fetishization and sexualization of transgender and nonbinary individuals](#). *Archives of Sexual Behavior*, 50(3):897–911.
- Kateryna Burovova and Mariana Romanyshyn. 2024. [Computational analysis of dehumanization of ukrainians on Russian social media](#). In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 28–39, St. Julians, Malta. Association for Computational Linguistics.
- U. Caglar, O. Yildiz, M. F. Ozerverli, R. Aydin, O. Sarilar, F. Ozgor, and M. Ortac. 2023. [Assessing the performance of chat generative pretrained transformer \(chatgpt\) in answering andrology-related questions](#). *Urology research & practice*, 49(6).
- Cory Jacob Cascalheira and Na-Yeun Choi. 2023. [Transgender dehumanization and mental health: Microaggressions, sexual objectification, and shame](#). *The Counseling Psychologist*, 51(4):532–559.
- Dana Collins. 1998. [Lesbian pornographic production: Creating social/cultural space for subverting representations of sexuality](#). *Berkeley Journal of Sociology*, 43:31–62.
- Simon Corneau and Emily van der Meulen. 2014. [Some like it mellow: On gay men complicating pornography discourses](#). *Journal of Homosexuality*, 61(4):491–510. PMID: 24245563.
- Harnoor Dhingra, Preetiha Jayashanker, Sayali Moghe, and Emma Strubell. 2023. [Queer people are people first: Deconstructing sexual identity stereotypes in large language models](#).
- Nicola Döring, Thuy Dung Le, Laura Vowels, Matthew Vowels, and Tiffany Marcantonio. 2024. [The impact of artificial intelligence on human sexuality: A five-year literature review 2020–2024](#). *Current Sexual Health Reports*, 17:1–39.
- Russell Franco D’Souza, Shabbir Amanullah, Mary Mathew, and Krishna Mohan Surapaneni. 2023. [Appraising the performance of chatgpt in psychiatry using 100 clinical case vignettes](#). *Asian Journal of Psychiatry*, 89:103770.
- Isabelle Marie Flory. 2024. [Porn is blunt \[...\] i had way more lgbtq+ friendly education through porn: The experiences of lgbtq+ individuals with online pornography](#). *Sexualities*.
- Amanda N. Gesselman, Ellen M. Kaufman, Alexandra S. Marcotte, Tania A. Reynolds, and Justin R. Garcia. 2023. [Engagement with emerging forms of sextech: Demographic correlates from a national sample of adults in the united states](#). *The Journal of Sex Research*, 60(2):177–189. PMID: 35104185.
- Salvatore Giorgi, Daniel Roy Sadek Habib, Douglas Bellew, Garrick Sherman, and Brenda Curtis. 2023. [A linguistic analysis of dehumanization toward substance use across three decades of news articles](#). *Frontiers in Public Health*, 11.
- Jesse Graham, Jonathan Haidt, and Brian A. Nosek. 2009. [Liberals and conservatives rely on different sets of moral foundations](#). *Journal of Personality and Social Psychology*, 96(5):1029–1046.
- Paolo Grigis and Antonella De Angeli. 2024. [Roleplay with large language model-based characters: A creative writers perspective \(short paper\) paolo grigis](#). In *SYNERGY@AVI*.
- Jonathan Haidt and Jesse Graham. 2007. [When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize](#). *Social Justice Research*, 20(1):98–116.
- Gert Martin Hald. 2006. [Gender differences in pornography consumption among young heterosexual danish adults](#). *Archives of Sexual Behavior*, 35(5):577–585.
- Lasana T. Harris and Susan T. Fiske. 2011. [Dehumanized perception a psychological means to facilitate atrocities, torture, and genocide? Zeitschrift fur Psychologie / Journal of Psychology](#), 219(3):175–181.

- Jochen Hartmann, Mark Heitmann, Christian Siebert, and Christina Schamp. 2023. [More than a feeling: Accuracy and application of sentiment analysis](#). *International Journal of Research in Marketing*, 40(1):75–87.
- Penny Harvey. 2020. [Let’s Talk About Porn: The Perceived Effect of Online Mainstream Pornography on LGBTQ Youth](#), page 31–52. Springer International Publishing, Cham.
- Nick Haslam. 2006. [Dehumanization: An integrative review](#). *Personality and Social Psychology Review*, 10(3):252–264.
- Emmanuel Klu and Sameer Sethi. 2023. [Tide: Textual identity detection for evaluating and augmenting classification and language models](#). *Preprint*, arXiv:2309.04027.
- Hadas Kotek, Rikker Dockum, and David Sun. 2023. [Gender bias and stereotypes in large language models](#). In *Proceedings of The ACM Collective Intelligence Conference, CI ’23*, page 12–24, New York, NY, USA. Association for Computing Machinery.
- Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D Manning, and James Y. Zou. 2024. [Mapping the increasing use of llms in scientific papers](#).
- Shir Lissak, Nitay Calderon, Geva Shenkman, Yaakov Ophir, Eyal Fruchter, Anat Brunstein Klomek, and Roi Reichart. 2024. [The colorful future of LLMs: Evaluating and improving LLMs as emotional supporters for queer youth](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2040–2079, Mexico City, Mexico. Association for Computational Linguistics.
- Davide Locatelli, Greta Damo, and Debora Nozza. 2023. [A cross-lingual study of homotransphobia on twitter](#). In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 16–24.
- Tiffany L. Marcantonio, Michelle Haikalis Karen E. Nielsen, Ruschelle M. Leone, Jacqueline Woerner, Elizabeth C. Neilson, and Anne Marie Schipani-McLaughlin. 2023. [Hey chatgpt, let’s talk about sexual consent](#). *The Journal of Sex Research*, 0(0):1–12. PMID: 37707442.
- Julia Mendelsohn, Yulia Tsvetkov, and Dan Jurafsky. 2020. [A Framework for the Computational Linguistic Analysis of Dehumanization](#). *Frontiers in Artificial Intelligence*, 3:55.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). *Preprint*, arXiv:1310.4546.
- Niloofer Mireshghallah, Maria Antoniak, Yash More, Yejin Choi, and G. Farnadi. 2024. [Trust no bot: Discovering personal disclosures in human-llm conversations in the wild](#). *arXiv*, abs/2407.11438.
- Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Michael V. Heinz, Ashmita Kunwar, Eunsol Soul Choi, Xuhai Xu, Joanna Kuc, Jeremy F. Huckins, Jason Holden, Sarah M. Preum, Colin Depp, Nicholas Jacobson, Mary P. Czerwinski, Eric Granholm, and Andrew T. Campbell. 2024. [Mind-scape study: Integrating llm and behavioral sensing for personalized ai-driven journaling experiences](#). *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 8(4).
- World Health Organization. 2022. [Ageism in artificial intelligence for health: WHO policy brief](#). World Health Organization.
- Samuel Pearson and Caitlin Curtis. 2025. [Erotic ai chatbots offer research opportunities for the behavioral sciences](#). *Archives of Sexual Behavior*.
- R M Seyam, B S Khan, S A Aljazeera, T Z Arabi, S S Alkhateeb, M F Alotaibi, and W M Altaweel. 2024. [Artificial Intelligence ChatGPT and GPT4 Performance on Male and Female Sexual Dysfunction, Sexually Transmitted Infection, and Male Factor Infertility in the 2019 to 2023 American Urological Association Self-Assessment Study Programs](#). *The Journal of Sexual Medicine*, 21:qdae002.033.
- Cherry Smyth. 1990. [The pleasure threshold: Looking at lesbian pornography on film](#). *Feminist Review*, 34(1).
- Laura M. Vowels. 2024. [Are chatbots the new relationship experts? insights from three studies](#). *Computers in Human Behavior: Artificial Humans*, 2(2):100077.
- Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, Kai-Wei Chang, and Nanyun Peng. 2023. [“kelly is a warm person, joseph is a role model”: Gender biases in LLM-generated reference letters](#).
- Valerie Webber. 2013. [Shades of gay: Performance of girl-on-girl pornography and mobile authenticities](#). *Sexualities*, 16(1-2).
- Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. 2024. [Wildchat: 1m chatGPT interaction logs in the wild](#). In *The Twelfth International Conference on Learning Representations*.
- Xuhui Zhou, Zhe Su, Tiwalayo Eisape, Hyunwoo Kim, and Maarten Sap. 2024. [Is this the real life? is this just fantasy? the misleading success of simulating social interactions with LLMs](#).
- Yanyan Zhou, (Harry) Yaojun Yan, Tuo Liu, and Paul Bryant. 2021. [Pornography use, two forms of dehumanization, and sexual aggression: Attitudes vs. behaviors](#). *Journal of Sex & Marital Therapy*, 47(6):571–590.

## A Appendix

### A.1 Lexicons

**Vermin** *vermin, rodent, rodents, rat, rats, mice, cockroach, cockroaches, termite, termites, bedbug, bedbugs, fleas*

**Moral Disgust** *disgust\*, deprav\*, disease\*, unclean\*, contagio\*, indecen\*, sin, sinful\*, sinner\*, sins, sinned, sinning, slut\*, whore, dirt\*, impiety, impious, profan\*, gross, repuls\*, sick\*, promiscu\*, lewd\*, adulter\*, debauch\*, defile\*, tramp, prostitute\*, unchaste, intemperate, wanton, profligate, filth\*, trashy, obscene\*, lax, taint\*, stain\*, tarnish\*, debase\*, desecrat\*, wicked\*, blemish, exploitat\*, pervert, wretched\**

**Neutral** *person, people, individual, individuals*

**Object** *it, that, this, thing, things, object, objects, item, items, machine, machines*

### Groups

**LGBTQ+** *lgbt, lgbtq, lgbtqs, lgbs, bisexual, bisexuals, gay, gays, homosexual, homosexuals, lesbian, lesbians, ftm, ftms, mtf, mtfs, trans, transes, transgender, transenders, hermaphrodite, hermaphrodites, trannies, tranny, transsexual, transsexuals, transman, transmen, transwoman, transwomen, transsexual, transsexuals, pansexual, pansexuals, nonbinaries, nonbinary, bi, bis, butch, butches, dyke, dykes, faggot, faggots, homo, homos, lesbo, lesbos, lgbtqia, lgbtqias, poly, polys, twink, twinks*

**LGB** *bisexual, bisexuals, gay, gays, faggot, faggots, homo, homos, twink, twinks, homosexual, homosexuals, lesbian, lesbians, butch, butches, dyke, dykes, lesbo, lesbos*

**Transgender** *ftm, ftms, mtf, mtfs, trans, transes, transgender, transenders, hermaphrodite, hermaphrodites, trannies, tranny, transman, transmen, transwoman, transwomen*

**Gay Men** *gay, gays, faggot, faggots, homo, homos, twink, twinks*

**Lesbian** *lesbian, lesbians, butch, butches, dyke, dykes, lesbo, lesbos*

**Bisexual** *bi, bis, bisexual, bisexuals*

**Heterosexual** *hetero, heteros, heterosexual, heterosexuals*

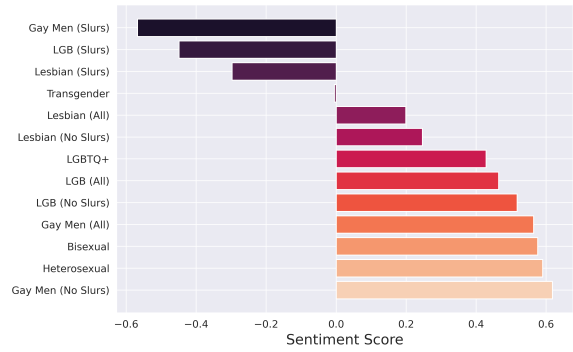


Figure 5: Sentiment score for each group. The score is calculated by taking an average over the sentiment labels (-1 for negative and 1 for positive) for all sentences containing a term in the group category.

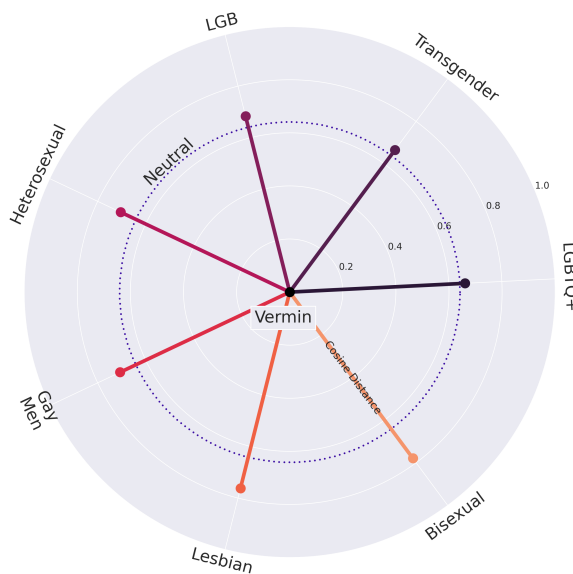


Figure 6: Cosine distance from Vermin Concept Vector to each subgroup. Subgroup distance is calculated by averaging over the distance of each term in the subgroup to the concept vector.

## B Additional Tables and Plots



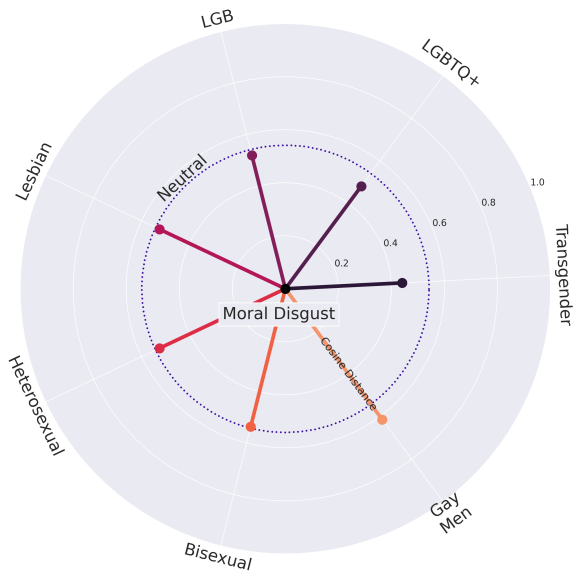


Figure 7: Cosine distance from Moral Disgust Concept Vector to each subgroup. Subgroup distance is calculated by averaging over the distance of each term in the subgroup to the concept vector.

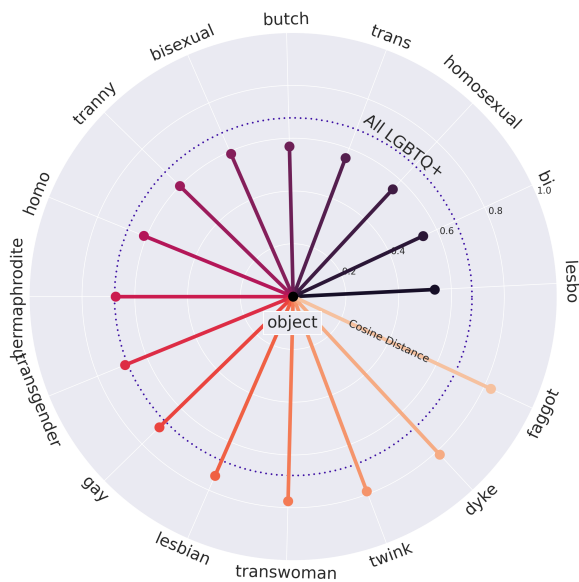


Figure 8: Cosine distance from Object Concept Vector to each term in the LGBTQ+ category. An average distance over all terms in the category is shown.