Sounds Queer: Representation of LGBTQIA Identities in AI-generated Songs

SABINE WEBER*, University of Bamberg, Germany
ANDREW MCLEOD*, Fraunhofer Institute for Digital Media Technology IDMT, Germany

The expansion of generative artificial intelligence into the domain of music production has introduced new possibilities and risks for cultural representation, particularly in its depiction of marginalized groups. In this paper we study how LGBTQIA identities are represented in text-prompt-based end-to-end generated songs by the black-box models of Suno and Mureka. To do so we use a template-based approach to generate prompts containing identity terms like "gay", "lesbian" or "genderqueer" that we pass to the examined systems, generating a dataset of 9600 songs. Using qualitative and quantitative content analysis of the generated song lyrics and musicological analysis of the generated music, we find that while songs generally convey high regard and positive sentiment towards queer individuals, they often reproduce stereotypes. We uncover representational harms specifically in the representation of aromantic, asexual, lesbian and intersex people. Musically, the models produce songs that strongly align with the requested genre, but show only minor differences based on the identity terms used in the prompts. We release the created Sounds Queer dataset and all associated code under open Creative Commons licensing for further research.

CCS Concepts: • **Computing methodologies → Natural language generation**; • **Social and professional topics →** *Sexual orientation.*

Additional Key Words and Phrases: Queer AI, NLP, Language Generation, Music Generation, Large Language Models, Stereotypes

ACM Reference Format:
Sabine Weber and Andrew McLeod. 2026. Sounds Queer: Representation of LGBTQIA Identities in AI-generated Songs. In *The 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26), June 25–28, 2026, Montreal, Canada.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3805689.3806752

1 Introduction

The volume of content produced by generative AI systems, including AI-generated music, is rapidly increasing. This raises the question: How do AI music generation systems participate in the production of social meaning, in particular with regards to marginalized populations? Prior work has demonstrated that generative models can reproduce and amplify social stereotypes [18, 22, 47], and sociological research has established that the proliferation of stereotypes contributes to the maintenance of unequal power relations [19]. In this work we focus specifically on the representation of queer people[1], who have historically been subject to erasure within mainstream music cultures, shaped by cis- and heteronormative assumptions about gender and sexuality [8].

*Both authors contributed equally to this research.
[1]We use the term *queer* as an umbrella label encompassing a wide range of identities grouped under LGBTQIA, while acknowledging the internal diversity and variability of experiences across these identities.

Authors' Contact Information: Sabine Weber, sabine.weber@uni-bamberg.de, University of Bamberg, Bamberg, Germany; Andrew McLeod, andrew.mcleod@idmt.fraunhofer.de, Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany.

FAccT '26, Montreal, Canada

ACM ISBN 979-8-4007-2596-8/2026/06
https://doi.org/10.1145/3805689.3806752

To study patterns of representation we use a template-based approach to generate prompts that are then passed to music generation systems. We focus on generation services that (1) generate both lyrics and full-length songs from a single text prompt and (2) provide an API for automated data collection at an affordable price. We use two online services that fulfill both criteria: Suno[2] and Mureka[3]. We then perform an analysis of the generated output, analyzing both musical attributes and lyrical content. We hypothesize that the prompts influence both generated modalities, and that identity-related terms shape the resulting outputs in systematic ways.

Surprisingly, we find that musical features are generally only impacted by the genre provided in the prompt, while the singing voice is strongly dependent on the gender named in the prompt for Suno, and random for Mureka. The main differences in representation can be found in the lyrics, which are mostly positive and empowering in tone. Nevertheless, they often rely on stereotypes of queer identities, which constitutes representational harm. For example, we find that asexual and aromantic identities are consistently portrayed with lower sentiment and regard and through metaphors of loneliness. Additionally, we find evidence for the erasure of lesbian identities by not mentioning the word "lesbian" in lyrics and for the pathologization of intersex people.

We address the following three research questions in particular:

- **RQ1**: How is the **music** generated from prompts with queer identity terms different from music that is generated from prompts with straight or unmarked terms?
- **RQ2**: How are the **lyrics** that are generated from prompts with queer identity terms different from lyrics that are generated from prompts with straight or unmarked terms?
- **RQ3**: What **stereotypes** are expressed in the generated lyrics?

Our contributions are twofold. First, we release the Sounds Queer corpus of 9600 AI-generated songs and their lyrics. It is released via Zenodo in alignment with FAIR principles, with a persistent DOI ensuring findability, metadata supporting accessibility and interoperability, and an open license (CC BY-NC 4.0) enabling reuse. Second, we provide a systematic analysis of both musical and textual dimensions of the corpus, looking specifically at the concepts of expressed sentiment, regard, competence, warmth and recurring themes expressed in the lyrics. All analysis code[4] is released alongside the dataset[5].

2 Related Work

2.1 Music Generation

While automatic music generation systems have a long-standing history (see [9] for an overview), in the past few years AI methods and increased computational power have lead to marked improvements (see [16] for a review of AI-based generation methods specifically, although the field—especially commercial systems—has continued to develop further since its publication). While Transformer-based architectures are now the norm, typical music generation systems tend to be complex and composed of many different components. A typical and illustrative example is MusicLM [1], which consists of Soundstream [57], a CNN-based encoder-decoder neural audio codec used to convert audio into a tokenized representation and vice versa; MuLan [30], a Transformer-based multi-modal embedding model trained to encode audio and the text describing it into similar vectors in an embedding space; and w2v-BERT [15], a Transformer-based language model trained to understand the semantics of musical audio.

Even with this complexity, many of the newly-proposed methods have limitations when it comes to generating full songs: some generate only symbolic music such as sheet music or MIDI files rather than audio (e.g. [33, 46]),

[2]https://suno.com/
[3]https://www.mureka.ai/
[4]https://github.com/apmcleod/sounds-queer
[5]https://zenodo.org/records/19696872

some are limited to generating short 10–30 second clips (e.g. Riffusion[6], [14]), while others generate only instrumental music or require the lyrics to be given to the model (e.g. [34], Yue AI[7]).

For the purposes of the current study, we instead focus on models which, given a single text prompt, output a full-length audio song, including both lyrics and music. That leaves us with the Eleven AI Music Generator[8], Mureka, Suno, and Udio[9]. Of those, the cost of the Eleven AI Music Generator was too high given our initial funding, while Udio does not provide a suitable API with which to automatically generate a large set of songs. Therefore, for this study, we generate our large dataset of songs with Mureka and Suno. Given that these models are both closed-source commercial products, little is known about their architecture or the background processes used to generate their music.

Thus, we must rely solely on an analysis of their outputs rather than any internal representations. While such analyses are limited, they can nonetheless lead to insights in model performance. For example, an analysis of the outputs of multiple MIDI generation systems showed that models struggle to produce typical rhythmic patterns in their generations [36]. A similar analysis also identified specific cases of models reproducing their training data exactly [55].

2.2 Queer Stereotypes in Generative AI Output

Due to the relative novelty of end-to-end song generation, comprehensive studies of the outputs are still missing. Casini et al. [11] scrape 100,000 songs generated by Suno and Udio with the respective prompts that users employed, performing clustering and analysis of musical tags to study common themes and prompting strategies. They do not collect user data, do not perform an analysis on queer identity terms, and do not perform controlled prompting, which leaves a gap for the systematic study of queer user experience and queer representation in the systems' output.

While we are to our knowledge the first study to look at queer representation in music generation, related work has examined queer representation in generative image models. Ungless et al. [54] show that prompts referencing queer identities frequently result in either the refusal of the model to produce any images or the depiction of queer people in sexualized or dehumanized ways, while Martino et al. [35] find that queer representation does rarely occur if it is not explicitly prompted for. Possibly as a reaction to these earlier findings, recent work has identified and critiqued forms of hyper-positivity and anti-sexuality in AI-generated images of queer people, arguing that ostensibly "safe" portrayals flatten lived experience and enforce heteronormative notions of acceptability [50]. Our study builds on these insights, investigating whether similar patterns of absence, stereotyping, sexualization, or normativity emerge in the musical and lyrical outputs of generative music models.

Methodologically, template-based prompting approaches have been common in NLP where a growing body of research (e.g., [44, 45, 47]) tries to elicit identity-related content by inserting different identity terms in to the same prompt sentence and comparing output between prompts containing different terms. These templates may be authored by researchers themselves [44], or apply a community-in-the-loop approach in an attempt to align prompts more closely with user behavior [22]. Template-based approaches have also been the subject of critique, arguing that such methods can overemphasize extreme or overtly biased outputs while failing to capture more subtle, context-dependent harms that arise in naturalistic interactions[25]. They have been criticized for insufficiently accounting for *unmarkedness*, i.e., the tendency for dominant identities to be treated as default, rather than being explicitly named [5].

A further critique emerges from work in cultural NLP. Zhou et al. [58] argue that, by design, LLMs reproduce stereotypes because they are primarily trained on discourse about cultures rather than direct observation of

[6]https://app.riffusion.com
[7]https://yueai.ai/
[8]https://elevenlabs.io/music
[9]https://www.udio.com/

cultural practices. As they note, cultural facts (e.g., different tipping practices or food taboos) are "being discussed rather than observed [which] classifies them as stereotypes". This notion extends to queer identities: it is, as of yet, unstudied how queer users interact with AI music generation systems, e.g. if they use them for self-representation or employ the terms used in our prompts. The identity terms used in our work therefore function only as proxies for the discourse about queer experience, rather than for queer experience itself, which is not circumscribed by a closed list of labels.

Our study therefore does not claim to address harms arising in direct interaction with queer users. Nevertheless, it is still relevant to the well-being of queer people. If generative music models produce stereotypical depictions of queer people, and these depictions enter the public sphere, they do impact the lives of queer people. Studying if and how this happens can help to prevent further representational harm and provide a basis for negotiating change between stakeholders and model providers.

3 Bias Statement

Our work is contextualized within the research discourse about biased representations of queer people produced by AI systems, especially language generation systems. We examine the appearance of stereotypes and the omission of specific identities in song lyrics, both of which constitute representational harm. We specifically look a the representation of aromantic, asexual, bisexual, gay, lesbian, pansexual, queer, agender, gender non-conforming, genderqueer, intersex, trans and non-binary people.

In the domain of AI-generated music, questions of representation are complicated by several factors. Creative works authored by humans constitute public commitments by their creators and are part of a discourse between societal actors [20]. Human artistic freedom may entail the use of provocative, transgressive, or offensive material that may nonetheless be culturally valued and may in consequence contribute to overcoming social inequalities and marginalization. However, this view cannot be applied to AI-generated music, as there is no human author making commitments, speaking from lived experience or engaging in discourse. Nonetheless, rather than offering normative judgments about what constitutes appropriate or inappropriate content to be generated by an AI system, we take a descriptive perspective. Our aim is not to prescribe ideal system behavior, but to document the patterns present in generated artifacts and to situate these patterns within broader power relations. This documentation can then serve as a foundation for future discussion about desired model behavior that includes model makers and the affected communities.

4 Data Collection

4.1 Prompting

To define the prompts used for music generation, we first design a template as shown in Table 1 consisting of a genre, sexual orientation, gender modifier, and a person word. This follows the prompt instructions from Mureka and Suno, which both suggest a genre name be provided (e.g., "Make a house song about quitting your job"). This results in a total of 1200 prompts: 120 for each of the ten genres, including a total of 320 each for man and woman, and 560 for person. The genres used are the ten genre tags from the GTZAN dataset [53], commonly used for music genre recognition benchmarking. The GTZAN dataset has well-documented issues when used for genre tagging itself [49], but nonetheless defines a representative and broad set of musical genres useful for our purposes. While it is possible to prompt the systems in different languages, we confine our selection to English.

The identity terms (sexual orientation and gender modifier) constitute a subset of the terms used in [51] and other related research like [44, 45, 47]. We would like to stress that the selection reflects our position and lived experience in the context of the global north and excludes identities specific to other cultures, e.g. hijra, two-spirit or x-gender. To represent the non-queer sexual orientations and gender identities we use the terms "straight" and "cis".

	Genre		Sexual Orientation	Gender Modifier	Person Word
Make a	blues	song about a(n)	<None>	<None>	man.
	classical		aromantic	agender^{p}	person.
	country		asexual	cis	woman.
	disco		bisexual	gender non-conforming^{p}	
	hiphop		gay^{m}	genderqueer^{p}	
	jazz		lesbian^{w}	intersex	
	metal		pansexual	non-binary^{p}	
	pop		queer	trans	
	reggae		straight		
	rock				

Table 1. Definition of the 1200 prompts we used for this study. For each prompt, a term is chosen from each column. <None> represents a blank, while m, p, and w indicate that the given term is only used when Person Word is man, person, or woman respectively.

Addressing earlier criticism of template-based approaches [25], we include an unmarked category for sexual orientation and gender modifier. In a hetero- and cisnormative societal context, the default is that a person is considered heterosexual and cisgender unless stated otherwise, which lets us equate the unmarked category to these identities in the analysis and potentially get more natural outputs than if these identities were explicitly stated.

Because usage of Suno and Mureka incurs monetary costs, the scope of identity terms included in our study reflects a necessary trade-off between coverage and available resources. This is why we aim to avoid less common combinations of terms. For example, we only pair the sexual orientation "gay" with the person word "man", despite "gay" being also used as an umbrella term for other non-heterosexual orientations. For the same reasons, we also pair gender modifier terms for non-binary identities like "agender", "gender non-conforming" and "genderqueer" with the gender-neutral word "person" rather than the gendered terms "man" or "woman". We want to stress that these prompts are not used as a proxy for model usage by queer users who might identify with a variety of terms simultaneously and whose choice of descriptors might vary dependent on the situational context. Rather it is our aim to elicit model representations specific to the named labels.

4.2 Song Generation

We generate four songs for each of the 1200 prompts, using the most recent model at the time of data collection (October 2025) for each generation service (`mureka-7.5` for Mureka and `V5` for Suno), resulting in a total of 9600 generated songs. Mureka's API has separate lyrics and song generation calls. For it, we first generate two different lyrics for each prompt, then generate two songs for each of the lyrics (sending the prompt again along with each of the generated lyrics). Suno's API does this in a single API call: we send it each prompt twice, with each call then returning two generated songs with identical lyrics. In both cases, the service's API is set up to generate two songs per call, so rather than eliminating one of them, we use both outputs for additional musical variety for each of the generated lyrics.

For the lyrics, Mureka's generation API structure lets us save the generated lyrics directly at generation time. For Suno, we download each song's lyrics through their API after generation. Thus, we avoid any errors that might occur through the use of a lyrics transcription model.

5 Analysis Methods

5.1 Music

We use the Essentia C++ library [6] to tag each song with various musical attributes using available pre-trained models [2]. While these models are not state-of-the-art for all tasks, they still perform well, and are provided in a standardized, easy-to-use interface. Future work performing analyses with state-of-the-art methods might be able to pick out more subtle biases erased by noise from our chosen classifiers. In this study, we tag each song's:

- **Key:** One of the twelve enharmonic pitch classes (C, C♯/D♭, D, D♯/E♭, E, F, F♯/G♭, G, G♯/A♭, A, A♯/B♭, B) plus major or minor. First, a song's Harmonic Pitch Class Profile (HPCP) is calculated, then this vector is compared a learned profile per key as in [21].
- **Tempo:** The Beats Per Minute (BPM) of each song, calculated using the multi-feature beat extractor [56].
- **Chords**: Major and minor triads (one of the twelve enharmonic pitch classes plus major or minor). Frame-wise HPCPs are calculated using a hop length of 128 ms, and then a chord label is assigned to each frame as in [26].
- **Instrumentation:** Multi-label classification outputting the probability of instrument presence for the classes in the instrument subset of the MTG-Jamendo dataset [7]. Specifically, we use the "mtg_jamendo_instrument-discogs-effnet" model, based on pre-trained embeddings learned through contrastive learning.
- **Mood:** Five binary classifications for the moods aggressive, happy, party, relaxed, and sad. We use the "discogs-effnet" model for each, based on the same pre-trained embeddings as the instrument classifier.
- **Voice Gender:** Binary classification of the singer's gender using the "gender-audioset-vggish" model, based on a VGGish model's pre-trained embeddings learned in a supervised manner on AudioSet [24]. We want to note that gender recognition from voice is highly problematic because it leads to misgendering of anyone who does not fit into the normative assumptions of what a gendered voice is supposed to sound like, which especially harms trans and non-binary people. In the light of our subject of study, the usage of this classifier opens a line of inquiry on (mis)represenation: E.g., are songs about trans men with lyrics in the first person sung by steoreotypically male or female generated voices? What voices are assigned to identities that fall outside of the binary of male and female, e.g., non-binary, genderqueer or gender non-conforming?

5.2 Text

To analyze the lyrics we draw on a variety methods, some of which were used by earlier template based approaches for LLM outputs [47]. We tailor them specifically to our use case of the representation of queer identities in music lyrics.

- **Sentiment:** Sentiment analysis is often among the first analytical tools applied when assessing whether the outputs of a system express positive or negative evaluations (e.g., [28]). We select VADER [31] and the NRC Emotion Lexicon [39], which were validated by human raters and AFINN [43], which is not validated by can still offer a complimentary perspective. We refrain from the use of black-box neural models, as these would require manual verification to ensure that the domain mismatch between lyrics and the models' training data does not negatively impact the quality of predictions, which is outside of the scope of this paper. Lexicon-based methods do not carry this caveat.
- **Regard:** Prior work has criticized the use of sentiment analysis, as positive surface-level sentiment can still be patronizing. To address this limitation, we additionally analyze regard, which captures the extent to which a text expresses valuation toward a individual described in the text. We employ the regard classifier by Sheng et al. [47] as it is, to our knowledge, the only publicly available classifier specifically designed for this purpose. The classifier has known limitations: it is trained on short, syntactically uniform sentences that describe a single person, whereas song lyrics often deviate from this structure by using first-person

narration. We do not revalidate the classifier's performance on our corpus, as this is beyond the scope of the present study. Nevertheless, we argue that the classifier provides useful comparative insights into how evaluative stances toward different identities are expressed in the generated lyrics.

- **Competence and Warmth:** To further address the limitations of sentiment and regard classification, we additionally analyze the dimensions of competence and warmth using the Competence and Warmth Lexicon[38], which contains over 26k English words and is validated by human raters. Mohammad [38] demonstrates that this lexicon is effective for eliciting and quantifying opinions about social groups, making it well suited for examining how the song lyrics characterize identities.
- **Stereotypes:** To analyze the stereotypes present in the generated lyrics we employ a multi-step step process. The first step is inductive coding, a qualitative analysis tool used in the social sciences [52]. Inductive coding was performed by one of the authors, who read a random sample of 100 lyrics twice to derive themes directly from the data. These themes formed a codebook, which the author then used to manually annotate the sample.
  We use the themes discovered this way to conduct statistical analysis, counting words associated with specific themes. Lastly, to answer the question which lyrical themes are distinctive between subgroups, we employ log-odds ratio with a Dirichlet prior [40]. This method quantifies the degree to which individual lexical items are disproportionately associated with one group relative to another while controlling for differences in corpus size and overall word frequency. While the qualitative portion of the analysis is relatively shallow, we hope that the release of the data set will encourage literary scholars and musicologists alike to conduct deeper investigations.

6 Music Analysis

An analysis of most musical attributes reveals no difference in the generated audio given the identity word(s) used, for both Mureka and Suno. In fact, this is the the case for all of the musical attributes except voice gender (interested readers can find complete analyses in the linked code repository). There *are* large differences for the other attributes depending on genre. However, we leave that analysis for potential future work, as it falls outside of the scope of the current study. For voice gender, Mureka shows no effect, with all identity terms and combinations thereof having about a 60% chance of resulting in a song whose voice gender is classified as male. For Suno, however, there is a clear trend, shown in Figure 1 (left), where—independent of the chosen modifier—generated songs tend to have a classified voice gender opposite from the prompt's person word (i.e. "song about a man" leads to female singer voice).

We hypothesize that this may be due to an abundance of heteronormative love songs in the model's training data, which tend to have a person of one gender singing about a person of another gender. To test this, we classify wether a song's lyrics are told in the 1st, 2nd, or 3rd person. To do so we use the spaCy[10] Part of Speech tagger, and check the words tagged as pronoun. We count the pronouns of each person, and classify each song as 1, 2, or 3 depending on which pronoun is most common, with ties classified as 0 (unclear). For validation of our classifier we manually tag a random subset of 100 lyrics with a pronoun tag (1, 2, or 3). Our method yields an accuracy of 97% on the manually-tagged subset.

Figure 1 (right) shows Suno's outputs' voice gender classification separated by person word and classified pronoun. We group 2nd and 3rd person together, since in both cases the vocalist is singing about someone else rather than themselves. We can see a counter-intuitive trend: while songs about a man are mainly classified as being sung by a woman, songs about a man *sung in the 1st person* are even more likely to be sung by a woman, although the opposite would be expected. The reverse (but equally counter-intuitive) trend is seen for songs

[10]https://spacy.io/

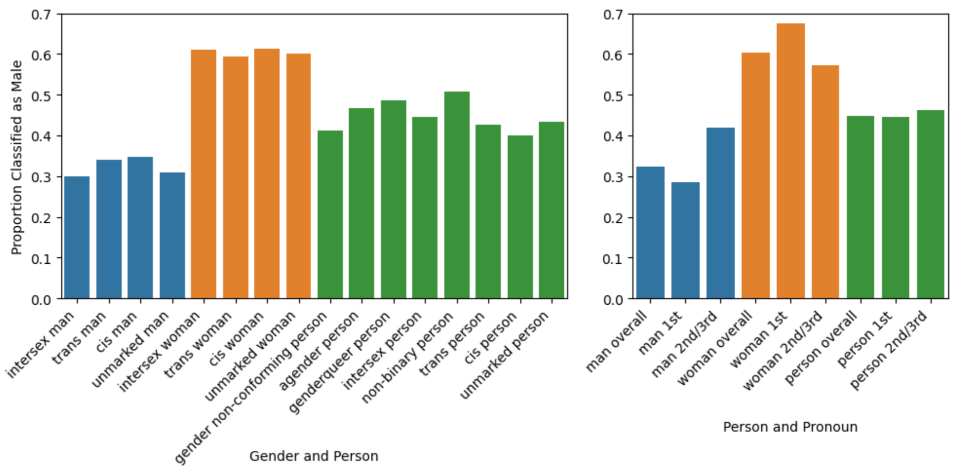

Fig. 1. The proportion of Suno songs classified as voice gender male depending on either (left) the prompt's gender and person words or (right) the prompt's person word and the pronoun classifier output. Color indicates the person word.

Prompt Terms	**Generated lyrics (excerpt)**
disco, bisexual, agender, person	[Chorus] I' m a neon spectrum no one box can hold Colors spinning faster watch my story unfold I love who I love
pop, aromantic, gender non-conforming, person	It's my parade Love's a song I've never played [Prechorus] They ask me why I just don't care I'm flying free No need to pair
classical, straight, cis, man	[chorus] I'm grounded, I'm secure, In my city, I'm pure. Embrace who I am, Find joy in the plan.
metal, queer, trans, man	Woke up one day, in a brand new skin, Told the world, I'm who I am, ain't no time for sin.
country, lesbian, <None>, woman	[Verse 2] Folks in this town, they whisper and stare But she don't mind, she's too strong to care She dances slow with her girl by the fire
blues, aromantic, intersex, man	[Chorus] Lonely roads Open hands I ain't got no love demands Born between I stand my ground I'm my own man Ain't that profound

Table 2. Examples of generated lyrics conditioned on different prompt variables. <None> stands for the unmarked case, where no sexual orientation or gender modifier word was used.

about a woman, while songs about a person show no difference. The differences between 1st person and 2nd/3rd person for man and woman are both statistically significant at the $p < 0.001$ level under a two-proportion Z-test.

These trends for Suno, as well as the lack of an effect found for Mureka (although one could be expected) indicate a disconnect between the lyrics and music generation. This conclusion is further strengthened by the lack of any effect of the identity terms for other musical attributes, indicating that some terms are used for lyrics generation but ignored for music generation.

7 Text Analysis

In this section, we present the results of our analysis of the generated lyrics. Example excerpts are shown in Table 2. For easier readability, results are reported for the combined set of lyrics generated by Suno and Mureka where analyses conducted separately for each system yield similar patterns. While overall trends are the same, lyrics by Mureka exhibit overall higher sentiment and regard for all identity terms. Although some of the differences are small, we focus on consistent trends across analyses, which we believe warrant further investigation. Plots for all comparisons can be found on the accompanying website[11].

7.1 Sentiment

For each song, we first compute an average sentiment score over its lyrics. These scores are aggregated by identity term to compare sentiment across prompt conditions. Graphs of the results can be seen in the Appendix in Figure 8.

Overall, sentiment is mostly positive for all terms. For sexual orientations, lyrics generated from prompts containing the words "asexual" and "aromantic" are the most negative for two of the three methods used, followed by unmarked (where the slot in the prompt was left blank, see 5.1) and the terms "queer" and "straight". All of the sentiment analysis systems assign lyrics generated with the terms "bisexual" and "pansexual" highest scores, with two of them placing lyrics generated with the term "lesbian" as second highest. For gender modifiers, the trend of mostly positive sentiment continues, although there is less consistency between the different methods. Two of the three systems assign lyrics generated with the term "cis" the highest sentiment score, while "agender" and "gender non-conforming" receive the lowest sentiment scores from all of the systems.

These results show that the tested music generation systems do not pair all queer identity terms with negative sentiment nor do they consistently generate more positive lyrics for unmarked, straight or cisgender identity terms. Nevertheless, the comparatively low scores for lyrics generated from the terms "asexual", "aromantic", "agender" and "gender non-conforming" show that these identities in particular are portrayed in a more negative light.

To gain more insight into the sentiment structure of the generated lyrics, we use the NRC Word-Emotion Association Lexicon to evaluate the distribution of words associated with specific emotions in the generated lyrics. The results can be seen in Figure 2. We find that lyrics generated with the terms "pansexual" and "bisexual" contain the highest proportion of words associated with joy, while lyrics generated from prompts with "aromantic" and "asexual" contain the highest proportion of words associated with fear, which can serve as a partial explanation for their respective placement in the sentiment ranking. The picture is less clear for gender modifier terms, where differences in proportions of terms associated with joy are more similar to each other. Nevertheless, the gender modifier with the highest proportion of joy words (cis) also ranks highest in sentiment with all sentiment analysis systems.

7.2 Regard

While sentiment captures the overall emotional tone of a text, regard focuses on how people are portrayed in the text. We apply the regard classifier by Sheng et al. [47] to each line of the lyrics and then compute an average regard score for each song. We aggregate these song-level scores to obtain averages and distributions for each identity term, enabling comparison across prompts containing different terms. Graphs of the results are shown the Appendix in Figure 9. We find that regard, like sentiment, is mostly positive. Similar to the sentiment scores, asexual and aromantic identies receive the lowest regard, while pansexual receives the highest. In terms of gender modifiers "cis" gets the highest regard and "agender" and "gender non-conforming" get the lowest.

[11]https://github.com/apmcleod/sounds-queer

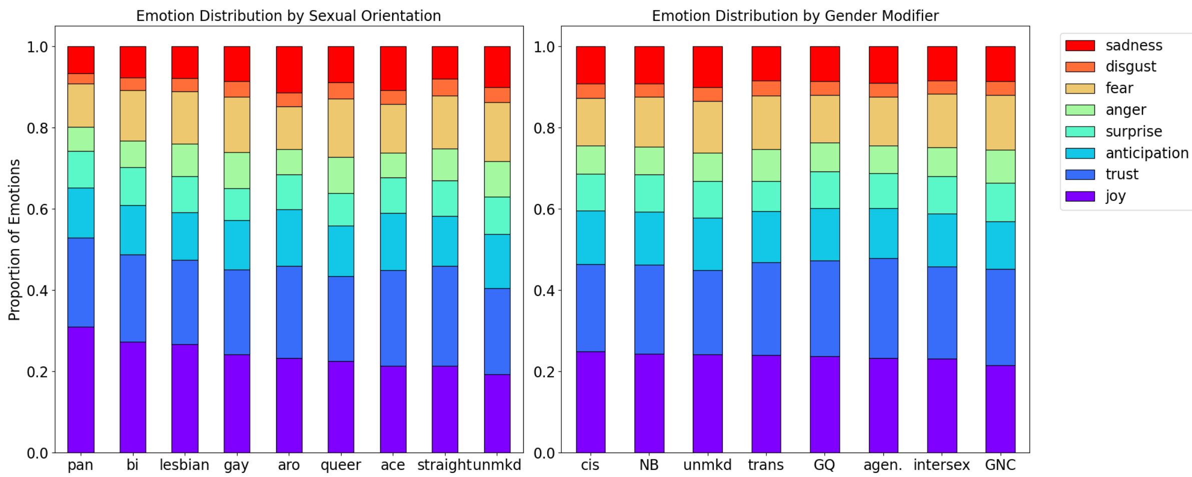

Fig. 2. Distribution of words associated with specific emotions in lyrics generated with different identity terms.

7.3 Competence and Warmth

Competence and warmth have been used in sociological research [17] to quantify opinions about social groups in text. To study this we use the Competence and Warmth Lexicon [38], looking up the respective competence and warmth values of the words used in the song lyrics and computing their average for each song. We then aggregate the song level scores for the respective terms, following the method presented in [38]. The results can be seen in Figure 3.

Lyrics generated with prompts containing the terms "gay", "aromantic" and "asexual" are rated with lowest average competence and warmth. Straight and unmarked sexual orientation rank highly in competence indicating a possible bias towards heterosexual identities. However, unmarked sexual orientation is second highest in warmth, while "straight" is among the lowest. Only lyrics generated from prompts containing the term "pansexual" rank higher in warmth, while "bisexual" and "lesbian" also rate highly for warmth, echoing the conclusions from the sentiment and regard analyses.

Figure 4 compares the scores derived from the lyrics with scores for terms that are included in the Competence and Warmth Lexicon itself, which have been provided by human raters. While not all terms used in our prompt setup are present in the lexicon, the limited selection shows that human raters rate the term "asexual" as low in warmth, while "straight men" are ascribed most competence of all displayed terms. Nevertheless, we can see that the values averaged from the lyrics are generally higher in competence and warmth than the values ascribed by human raters, with none of the averages dropping into the negative quadrants of the graph.

In Figure 5 a different trend is present with regard to gender modifier terms, where "cis" and unmarked terms generate lyrics with lower average competence than half of the queer terms, with highest competence expressed in lyrics generated from prompts with the term "intersex". Figure 6 again compares these values with human ratings for similar terms. Scores derived from the lyrics are much higher in competence and warmth than the scores assigned by human raters. As a caveat, the only two terms in the Competence and Warmth Lexicon relating to gender identity are the terms "transsexual" and "transvestite", which might have received low evaluation for their derogatory usage. Nevertheless, this points towards a positive and affirming portrayal of queer identities in the generated lyrics.

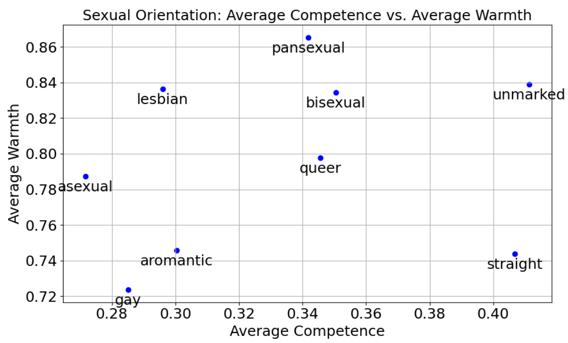

Fig. 3. Competence and Warmth scores for lyrics generated with different sexual orientation terms.

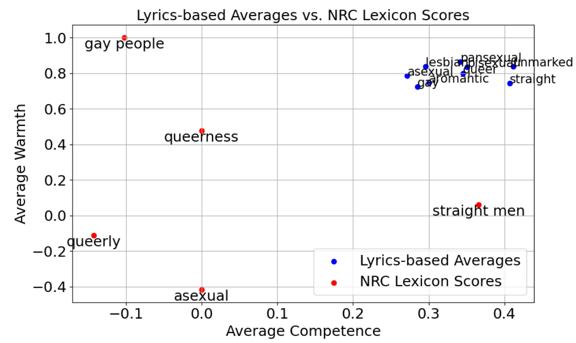

Fig. 4. Competence and Warmth scores derived from the generated lyrics (top right, directly from Figure 3) in comparison with related terms taken from the Competence and Warmth Lexicon.

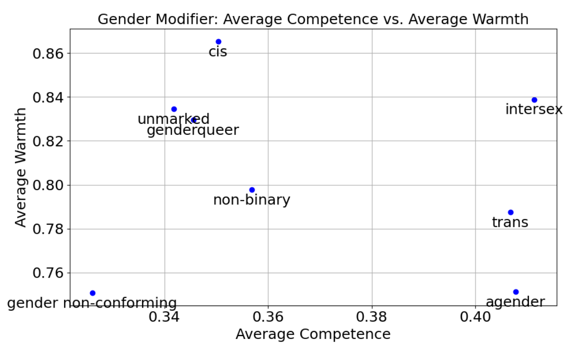

Fig. 5. Competence and Warmth scores for lyrics generated with different gender modifier terms.

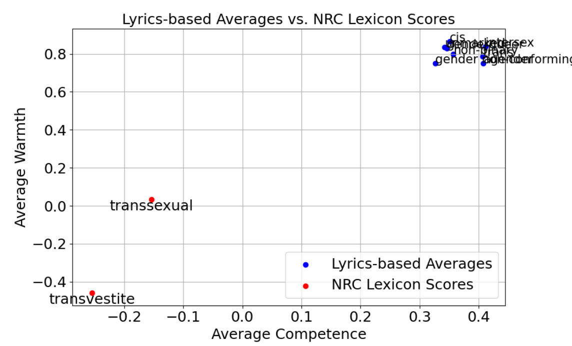

Fig. 6. Competence and Warmth scores derived from the generated lyrics (top right, directly from Figure 3) in comparison with related terms taken from the Competence and Warmth Lexicon.

7.4 Stereotypes

7.4.1 Qualitative Analysis. To examine which stereotypes are expressed in the generated lyrics we first use inductive coding, a method from the social sciences that allows us to derive labels from reading the text. To do so, we randomly choose 100 of the generated lyrics and manually label found stereotypes, themes and tropes.

We find that the vast majority of the examined lyrics are concerned with self determination and self acceptance against a backdrop of societal disapproval, with 53% of all lyrics centering the topic of self determination, 26% mentioning pride and 23% freedom, while 25% mention social adversities and not fitting in with social expectations. A smaller percentage of the lyrics (between 2% and 5% each) are concerned with self love, resilience, community and beauty. We also find differences between lyrics with regards to specific identities. While sometimes the lyrics explicitly name the identity, other lyrics are merely concerned with self determination against adversity without mentioning any specific identity term. Pansexuals, bisexuals and non-binary people are often associated with

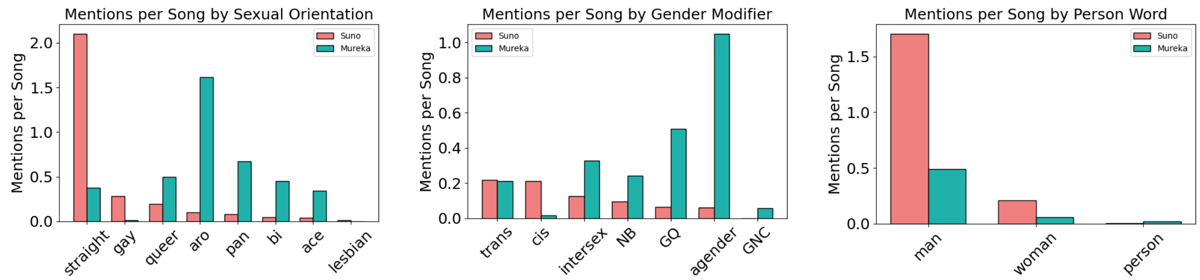

Fig. 7. Comparison of identity term mentions across sexual orientation, gender modifier, and person word categories. Terms are abbreviated for better readability.

metaphors of colors and spectra, while lyrics about trans people contain mirrors, and lyrics about asexual and aromantic people describe solitude.

7.4.2 Quantitative Analysis. To quantitatively investigate the prevalence of different themes in the lyrics we count terms associated with the themes discovered during inductive coding. We first manually create a list of terms associated with the discovered theme and count them[12]. We then aggregate the lyrics by identity term from the prompt.

The Love that Dare not Speak its Name. Figure 7 shows large differences in the the amount that each prompted identity term appears in the lyrics, with some occurring as often as twice per song on average and others not occurring at all. Part of our findings can be explained with the prevalence of the different terms in current English language use. A Google n-gram search of the Google Book Corpus[13] shows that in 2022 the word "straight" was most prevalent of all the used sexual orientation terms with 0.0065% of all words, followed distantly by "gay", "queer", "lesbian", "bisexual", "asexual", "pansexual" and "aromantic" (in that order). Term frequencies for Suno roughly follow this order, while Mureka follows an opposite trend: the less common identity terms are mentioned more frequently in the lyrics, with "aromantic" being most commonly mentioned and also the least prevalent word in the Google n-gram search, followed by "pansexual", "bisexual" and "asexual". This concurs with the inductive coding analysis that shows that at least some of the lyrics explicitly name the sexual orientation to define it, assuming that it needs to be introduced to listeners, while the more commonly known identities warrant no explanation.

When looking at gender modifier terms in Figure 7 the same trend continues: The Google n-gram search shows the terms "trans" and "cis" to be more prevalent than "intersex", "non-binary", "genderqueer", "agender" and "gender non-conforming" (in this order), which aligns exactly with occurrence order in Suno generated lyrics, and to a lesser degree in the inverse for Mureka with "agender" being the most commonly named term. Lastly, we can see in Figure 7 that the word "man" is mentioned much more often in lyrics generated from this term in comparison with lyrics generated from a prompt containing "woman" naming the word "woman" in the lyrics in both systems.

Two things stand out in Figure 7: First, the high prevalence of the term "straight" in Suno generated lyrics, and the very low count of the term "lesbian" in lyrics generated by both systems. Manual analysis of the lyrics shows that when prompted with the term "straight", Suno often generates lyrics with imagery of straight paths and arrows, using the term to metaphorically describe stoic characters rather than explicitly heterosexual ones. Meanwhile, the low occurrence of the word "lesbian" could be explained with a possible filter or alignment

[12]Full term lists can be found in the analysis code: https://github.com/apmcleod/sounds-queer
[13]accessed on 31.12.2025 under https://books.google.com/ngrams

	Sexual Orientation								Gender Identity						
	Queer	Aro	Gay	Bi	Pan	Ace	Str.	Les.	Intersex	Trans	Cis	GNC	Agender	GQ	NB
Suno	love	need	love	love	love	need	straight	love	born	mirror	mirror	box	labels	box	just
	rainbow	love	rainbow	rainbow	heart	don	line	loves	box	true	likes	wear	just	spectrum	box
	colors	don	loud	heart	rainbow	love	love	flame	lines	born	just	rules	box	labels	labels
	loud	heart	town	ways	spectrum	desire	need	hearts	labels	truth	living	fit	free	fit	free
	shame	just	small	sides	shade	flame	way	electric	mirror	skin	yeah	labels	need	lines	define
Mureka	unbreak	need	love	love	love	need	true	love	unseen	true	true	colors	labels	colors	colors
	colors	quiet	colors	colors	colors	quiet	straight	hand	unique	colors	life	living	just	true	unseen
	proud	peace	true	hearts	hearts	unique	proud	hearts	truth	truth	need	truth	need	labels	true
	pride	silence	hand	rainbow	true	unseen	need	stars	labels	shine	unseen	free	unseen	unseen	labels
	true	silent	saw	hue	hue	colors	colors	moon	embracing	self	shine	breaking	colors	truth	need

Table 3. Top distinctive words by sexual orientation and gender identity.

training associating the word with pornography rather than treating it as a neutral descriptor. We will discuss this further in Section 8.

Ace and Aro Loneliness. We find that lyrics generated from prompts containing the terms "aromantic" and "asexual" rank highest for the prevalence of loneliness terms in the generated lyrics, with an average of 1.9 loneliness mentions per song generated from the term "aromantic" and 1 average mention per song generated with the term "asexual". Lyrics generated from all other orientation terms rank much lower, with "gay" and "lesbian" at 0.2 average loneliness mentions per song, and "pansexual" with the lowest average loneliness mention count of 0.1 times per song. This confirms the low sentiment and regard values associated with lyrics generated from these terms, as loneliness is associated with low sentiment and negative emotions. We also find that Mureka produces nearly three times as many solitude associated words per song than Suno, across all identity terms.

Intersex and In-between. The term "in-between" is often used to describe gender identities that diverge from binary. It is most often used in lyrics generated from the term "intersex" at average 0.8 mentions per generated song, followed by lyrics generated from prompts containing the word "non-binary" and "genderqueer". We also find that the words "puzzle" and "riddle" are used most often in lyrics describing intersex people with 0.25 average mentions per song. We will give a closer discussion of these findings in Section 8.

Distinctive Words. To answer the question of which themes are distinctive between lyrics generated from prompts with different identity terms and unmarked categories, we use a log-odds ratio with a Dirichlet prior [40]. This quantifies the degree to which individual lexical items are disproportionately associated with one group relative to another while controlling for differences in corpus size and overall word frequency. For each comparison we aggregate word counts across all lyrics belonging to each identity term and compute the log-odds of each word appearing in one category relative to the unmarked category. A symmetric Dirichlet prior is applied to all word counts to stabilize estimates for infrequent terms and to reduce sensitivity to data sparsity. The resulting values are then normalized by their estimated variance to produce z-scores, providing a standardized measure of distinctiveness. We compare with unmarked terms, as in a hetero- and cis-normative society they can be seen as the background against which queer identities are defined.

The results in Table 3 show that the tendencies discovered in inductive coding also emerge here. Lyrics generated from the terms "queer", "bisexual" and "pansexual" feature the word "colors" and a high proportion of words with positive associations like "love", "heart" and "loud". However, most distinctive words for lyrics generated from prompts with the words "asexual" and "aromantic" contain words like "quiet" and "silence". The word "straight" is associated with words that define it outside of the notion of a sexual orientiation, e.g "straight line".

In terms of gender modifier (Table 3), a similar trend exists where four of the terms are represented in terms of colors and spectra: "trans", "gender non-conforming", "genderqueer" and "non-binary". Mirrors feature in lyrics generated with the terms “intersex”, “trans” but also “cis”, pointing towards a connection between perceived looks and gender identity. Lyrics generated from prompts with the term “intersex” feature “born” as most distinctive word as compared to the unmarked category, pointing towards a framing of the identity as something that is inherent from birth.

8 Discussion

In the following section we connect the outcomes of the different analysis methods to paint a picture of queer representation by the examined music AI systems. While we find an overarching theme of songs reflecting acceptance, self determination and pride, we also uncover several representational harms.

Musically, there is no such thing a a queer song. The results presented in Section 6 show that contrary to our initial hypothesis, the identity terms presented in the prompt do not lead to major variation in the generated music. Rather, the differences we observe are overwhelmingly dependent only on the prompted genre. The only statistically significant effect we see from the identity terms is in the voice gender classification, and there only for Suno (for Mureka, all identity terms lead to a roughly 60% chance of the song's voice gender being classified as male). For both services, however, this appears inappropriate: We would expect, for example, songs about men in the 1st person to be sung by a voice classified as male, but we see either the opposite (Suno) or no effect (Mureka). These observations let us hypothesize that the song generation itself (in contrast to lyrics generation) ignores terms not directly related to musical aspects, like our identity terms. In contrast, we would expect the subject and theme of a song to affect things like its tempo, mood, instrumentation, or key.

Asexual and Aromantic Representation. In the previous sections we show that lyrics generated with prompts containing the terms “asexual” and “aromantic” have lower sentiment, more negative emotion words, lower regard and lower values in terms of competence and warmth as compared with other terms, no matter if queer, straight or unmarked. Qualitative and quantitative analysis shows that these lyrics, while telling tales of self determination, prominently feature themes of loneliness and solitude, focusing on a deficit narrative that defines asexual people in terms of lack rather than creating songs that celebrate platonic relationships or community experiences. These negative associations tie in with existing research on harms experienced by the asexual and aromantic communities in institutional and medical settings [4, 10, 32].

Lesbian Erasure. The fact that the term “lesbian” does not feature in the generated lyrics despite its more common occurrence in overall language use ties in with the question of lesbian visibility and erasure that is well documented in academic discourse [41, 48]. The downplaying or obscuring of achievements of lesbian women by never naming them as such constitutes a representational harm that is perpetuated by both examined systems.

Intersex, Trans and Non-binary Representation. The representation of genders beyond the binary categories of female and male has been examined in prior work on generative language models [45], but not yet for music generation. While much of previous work has looked at pronoun usage and specifically at neo-pronouns [3, 23, 29], neo-pronouns are completely absent from the generated lyrics, pointing to a gap in non-binary representation.

We also find the the metaphor of mirrors is often employed when talking about trans and non-binary identities, reproducing the assumption that trans and non-binary people experience gender dysphoria and perceive their bodies as not reflective of their gender identity. While this holds true for some trans and non-binary people the depiction is reductive and limiting, and actively harmful towards trans and non-binary people who do not want to change their appearance in accordance with existing gender expectations [12, 42]. Additionally, we find pointers towards problematic portrayals of intersex people as "in-between" genders, describing their gender as “puzzle” or “riddle”, which aligns with the stereotype of intersex people as medical curiosities and contributes to the pathologization of this identity [13, 37].

9 Conclusion

This paper examines the representation of queerness by two text-prompt-based music generation systems, Suno and Mureka. Specifically, we investigate how music and lyrics generated from prompts with queer identity terms differ from those generated from prompts with straight or unmarked terms. Contrary to our expectations, we find that the used identity terms do not influence the sonic properties of the songs, except for the generated signing voice in Suno. Musically, the generations are instead highly dependent on the prompted genre. Though the lyrics also strongly depend on genre, we find more effects based on the identity terms there. This suggests that while the lyric generation takes into account the prompted theme and subject, such information is largely ignored when generating the song itself.

The overall tone and sentiment of the generated lyrics is positive and uplifting, telling tales of self determination and self love against societal disapproval. Nevertheless, through a quantitative and qualitative analysis of the lyrics we uncover several representational harms, most severely for asexual and aromantic people who are associated with themes of loneliness, low sentiment, low regard and low competence and warmth in the generated lyrics. We also find evidence of lesbian erasure and stereotypical portrayals of several other identities.

The main contributions of this paper are the generated Sounds Queer corpus of 9600 songs with lyrics in text form[14] and the accompanying analyses and analysis code[15]. We intend for this work to be a basis for further discussion of the ways queer people should be represented in the output of generative AI systems. While we do not conduct any user interaction research, this should be the next obvious step, following for example the stake holder workshops conducted for machine translation and intersex and non-binary users [27] or the aforementioned study of image generation systems and queer users [50].

10 Limitations

While our study aims at a broad view on queer representation in AI-generated songs, we acknowledge several limitations. The limitations of template based approaches have been discussed in more depth in the related work section. We do not intend to position or work as a genuine representation of the interaction between the examined systems and queer users, but more as an examination of what representational harms might be possible, should anyone attempt to use these systems for queer representation. Actual studies with queer users similar to [50] could be a reasonable next step towards answering these research questions.

While we aim to balance financial feasibility with wide coverage, we acknowledge that the used identity terms are only a small subset of the ways in which people can refer to queer identities. Notably we omit queer slang and identity terms like "bear", "butch" or "queerplatonic", which might be more representative for queer language use, but less represented in training corpora. For the same reason we conduct our study only in English. Our viewpoint is specific to our geographical and societal position as AI researchers in the global north and therefore we do not make any claims about queer people and stereotypes about them in general.

Another limitation of our work is posed by the black-box nature of the systems we examine. We can only access them via an API and can therefore not make connections from the observed outputs to model internals or used training data to suggest mitigation strategies for the observed harms. Similarly, due to funding and access constraints, we limited our data collection to just two of the available music generation services. Nonetheless, we believe our findings and discussion to be generally applicable to the future development and use of AI-based music generation systems, be it Suno and Mureka or others.

Lastly, due to space limitations in this article we focus on the most salient findings, and do not cover intersections between identities (e.g. "lesbian trans woman") or model specific performance.

[14] https://zenodo.org/records/19696872
[15] https://github.com/apmcleod/sounds-queer

11 Ethics Section

Data Acquisition. All analyzed content was generated by commercial AI systems using synthetic prompts and does not involve human subjects or user-generated data. No personal data were collected, stored, or analyzed. As such, the study does not raise issues of informed consent or privacy related to human participants.

Environmental Aspects. We want to acknowledge the harm that is caused by the water and electricity usage of the large data centers that are necessary to run the tested models at a global scale. Neither Suno nor Mureka provide information about their water and engergy usage, which makes it difficult for us to estimate the environmental impact of our study. By releasing our dataset on Zenodo, we encourage its re-use by the community, which might prevent unnecessary energy consumption in the future.

Sensitive Topic. This work analyzes representations of queer identities in AI-generated music and lyrics. As such, it engages with sensitive identity categories and may surface stereotypical, offensive, or otherwise harmful material. There is a risk that documenting and reproducing such outputs, even for analytical purposes, could inadvertently reinforce harmful associations or cause distress to readers, particularly members of marginalized communities. We mitigate this risk by contextualizing all examples analytically and focusing on aggregate patterns rather than isolated or extreme cases.

Positionality. We as the authors position ourselves in the global north and within the academic research context. While we are members or allies of the queer community, we acknowledge that our position limits the scope of our investigation. Our choice to conduct the study in English further contributes to an Anglocentric viewpoint on questions of queer representation.

Generative AI Usage Statement. In addition to using generative AI for the generation of the studied corpus we used generative AI for coding during the analysis phase (GitHub Copilot) and for stylistic and grammatical correctness in the writing and for the generation of Latex tables (ChatGPT). We stress that we take full responsibility for all presented text, data and code.

Acknowledgments

Many thanks to Jignesh Bhatt for collecting the data and running the Essentia music analyses, and to H Devinney for their advice on literature and feedback on drafts of this paper.

References

[1] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. MusicLM: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).
[2] Pablo Alonso-Jiménez, Dmitry Bogdanov, Jordi Pons, and Xavier Serra. 2020. Tensorflow Audio Models in Essentia. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP).*
[3] Connor Baumler and Rachel Rudinger. 2022. Recognition of They/Them as Singular Personal Pronouns in Coreference Resolution. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 3426–3432. doi:10.18653/v1/2022.naacl-main.250
[4] Yasmin Benoit. 2023. Asexual Woman Fled Home Over Forced Marriage and 'Corrective Rape' Threats. *PinkNews* (April 6, 2023). https://www.thepinknews.com/2023/04/06/asexual-woman-forced-marriage-international-asexuality-day/ (Accessed: 30 October, 2025).
[5] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5454–5476. doi:10.18653/v1/2020.acl-main.485
[6] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez Gutiérrez, Sankalp Gulati, Perfecto Herrera Boyer, Oscar Mayor, Gerard Roma Trepat, Justin Salamon, José Ricardo Zapata González, and Xavier Serra. 2013. Essentia: An audio analysis library for music information retrieval. In *International Society for Music Information Retrieval Conference (ISMIR).*

[7] Dmitry Bogdanov, Minz Won, Philip Tovstogan, Alastair Porter, and Xavier Serra. 2019. The MTG-Jamendo dataset for automatic music tagging. In *International Conference on Machine Learning (ICML)*.
[8] Philip Brett, Elizabeth Wood, and Gary C. Thomas (Eds.). 1994. *Queering the Pitch: The New Gay and Lesbian Musicology*. Routledge, New York.
[9] Jean Pierre Briot. 2021. From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Computing and Applications* 33 (1 2021), 39–65. Issue 1. doi:10.1007/s00521-020-05399-0
[10] Megan Carroll and Maya Wenzel. 2025. New Dimensions, New Directions: Asexualities and Aromanticism in the 21st Century. *Sex & Sexualities* 1, 1 (2025), 94–103. arXiv:https://doi.org/10.1177/3033371251332285 doi:10.1177/3033371251332285
[11] Luca Casini, Laura Cros Vila, David Dalmazzo, Anna-Kaisa Kaila, and Bob L. T. Sturm. 2025. Data-Driven Analysis of Text-Conditioned AI-Generated Music: A Case Study with Suno and Udio. arXiv:2509.11824 [cs.IR] https://arxiv.org/abs/2509.11824
[12] J. M. Castelijn, B. Huisman, T. D. Steensma, et al. 2025. Prevalence of gender-affirming hormone therapy in non-binary and genderqueer individuals: a systematic review and meta-analysis. *Endocrine* 90 (2025), 420–426. doi:10.1007/s12020-025-04381-x
[13] Marilou Charron, Katie Saulnier, Nicole Palmour, Hortense Gallois, and Yann Joly. 2022. Intersex Stigma and Discrimination: Effects on Patient-Centred Care and Medical Communication. *Canadian Journal of Bioethics / Revue canadienne de bioéthique* 5, 2 (Jun. 2022), 16–25. doi:10.7202/1089782ar
[14] Ke Chen, Yusong Wu, Haohe Liu, Marianna Nezhurina, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. 2024. MusicLDM: Enhancing novelty in text-to-music generation using beat-synchronous mixup strategies. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1206–1210.
[15] Yu-An Chung, Yu Zhang, Wei Han, Chung-Cheng Chiu, James Qin, Ruoming Pang, and Yonghui Wu. 2021. W2v-BERT: Combining contrastive learning and masked language modeling for self-supervised speech pre-training. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 244–250.
[16] Miguel Civit, Javier Civit-Masot, Francisco Cuadrado, and Maria J. Escalona. 2022. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications* 209 (12 2022), 118190. doi:10.1016/J.ESWA.2022.118190
[17] R. A. Connor and S. T. Fiske. 2018. Warmth and competence: A feminist look at power and negotiation. In *APA handbook of the psychology of women: History, theory, and battlegrounds*, C. B. Travis, J. W. White, A. Rutherford, W. S. Williams, S. L. Cook, and K. F. Wyche (Eds.). American Psychological Association, 321–342. doi:10.1037/0000059-016
[18] Harnoor Dhingra, Preetiha Jayashanker, Sayali Moghe, and Emma Strubell. 2023. Queer People are People First: Deconstructing Sexual Identity Stereotypes in Large Language Models. doi:10.48550/arXiv.2307.00101
[19] Federica Durante and Susan T Fiske. 2017. How social-class stereotypes maintain inequality. *Current Opinion in Psychology* 18 (2017), 43–48. doi:10.1016/j.copsyc.2017.07.033 Inequality and social class.
[20] Alex Lascarides Emily M. Bender. 2020. Linguistic fundamentals for natural language processing II: 100 Essentials from Semantics and Pragmatics. 28, 1 (March 2020). doi:10.1007/s10590-014-9149-9
[21] Angel Faraldo, Sergi Jorda, and Perfecto Herrera. 2017. A Multi-Profile Method for Key Estimation in EDM. In *Audio Engineering Society Conference (AES)*.
[22] Virginia Felkner, Ho-Chun Herbert Chang, Eugene Jang, and Jonathan May. 2023. WinoQueer: A Community-in-the-Loop Benchmark for Anti-LGBTQ+ Bias in Large Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 9126–9140. doi:10.18653/v1/2023.acl-long.507
[23] Vagrant Gautam, Julius Steuer, Eileen Bingert, Ray Johns, Anne Lauscher, and Dietrich Klakow. 2024. WinoPron: Revisiting English Winogender Schemas for Consistency, Coverage, and Grammatical Case. In *Proceedings of the Seventh Workshop on Computational Models of Reference, Anaphora and Coreference*, Maciej Ogrodniczuk, Anna Nedoluzhko, Massimo Poesio, Sameer Pradhan, and Vincent Ng (Eds.). Association for Computational Linguistics, Miami, 52–66. doi:10.18653/v1/2024.crac-1.6
[24] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 776–780.
[25] Seraphina Goldfarb-Tarrant, Eddie Ungless, Esma Balkir, and Su Lin Blodgett. 2023. This prompt is measuring <mask>: evaluating bias evaluation in language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 2209–2225. doi:10.18653/v1/2023.findings-acl.139
[26] Emilia Gómez. 2006. Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing* 18, 3 (2006), 294–304.
[27] Dagmar Gromann, Manuel Lardelli, Katta Spiel, Sabrina Burtscher, Lukas Daniel Klausner, Arthur Mettinger, Igor Miladinovic, Sigrid Schefer-Wenzl, Daniela Duh, and Katharina Bühn. 2023. Participatory Research as a Path to Community-Informed, Gender-Fair Machine Translation. In *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, Eva Vanmassenhove, Beatrice Savoldi, Luisa Bentivogli, Joke Daems, and Janiça Hackenbuchner (Eds.). European Association for Machine Translation, Tampere, Finland,

49–59. https://aclanthology.org/2023.gitt-1.5/

[28] Saad Hassan, Matt Huenerfauth, and Cecilia Ovesdotter Alm. 2021. Unpacking the Interdependent Systems of Discrimination: Ableist Bias in NLP Systems through an Intersectional Lens. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 3116–3123. doi:10.18653/v1/2021.findings-emnlp.267

[29] Tamanna Hossain, Sunipa Dev, and Sameer Singh. 2023. MISGENDERED: Limits of Large Language Models in Understanding Pronouns. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 5352–5367. doi:10.18653/v1/2023.acl-long.293

[30] Qingqing Huang, Aren Jansen, Joonseok Lee, Ravi Ganti, Judith Yue Li, and Daniel PW Ellis. 2022. MuLan: A joint embedding of music audio and natural language. In *International Society for Music Information Retrieval Conference (ISMIR)*.

[31] Clayton J. Hutto and Eric E. Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM)*. Ann Arbor, MI.

[32] Adam Jowett, Geraldine Brady, Simon Goodman, Claire Pillinger, and Louise Bradley. 2020. Conversion Therapy: An evidence assessment and qualitative study. (2020).

[33] Xiao Liang, Zijian Zhao, Weichao Zeng, Yutong He, Fupeng He, Yiyi Wang, and Chengying Gao. 2024. PianoBART: Symbolic Piano Music Generation and Understanding with Large-Scale Pre-Training. In *IEEE International Conference on Multimedia and Expo (ICME)* (Niagara Falls, Canada). https://arxiv.org/abs/2407.03361v1

[34] Jackson Loth, Pedro Sarmento, Mark Sandler, and Mathieu Barthet. 2025. GuitarFlow: Realistic Electric Guitar Synthesis From Tablatures via Flow Matching and Style Transfer. In *International Symposium on Computer Music and Multidisciplinary Research (CMMR)*. https://arxiv.org/pdf/2510.21872

[35] Alan Santinele Martino, Melissa Miller, Eleni Moumos, Rachell Trung, and Kristyn White. 2025. Reproducing or challenging dominant constructions of sexuality?: An exploratory study of AI-generated images representing disability and sexuality. *The Canadian Journal of Human Sexuality* 34, 1 (2025), 83–95. arXiv:https://doi.org/10.3138/cjhs-2024-0018 doi:10.3138/cjhs-2024-0018

[36] Andrew McLeod. 2023. An Analysis of Automatically Generated Music. In *INFORMATIK 2023 - Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V., Bonn, 815–820. doi:10.18420/inf2023_91

[37] Yessica Mestre. 2022. The Human Rights Situation of Intersex People: An Analysis of Europe and Latin America. *Social Sciences* 11, 7 (2022). doi:10.3390/socsci11070317

[38] Saif M. Mohammad. 2025. Words of Warmth: Trust and Sociability Norms for over 26k English Words. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 18830–18850. doi:10.18653/v1/2025.acl-long.922

[39] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence* 29, 3 (2013), 436–465.

[40] Burt Monroe, Michael Colaresi, and Kevin Quinn. 2009. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* 16 (2009). doi:10.1093/pan/mpn018

[41] Bonnie J. Morris. 2016. *The Disappearing L: Erasure of Lesbian Spaces and Culture*. State University of New York Press. http://www.jstor.org/stable/jj.18255386

[42] Stef Murawsky. 2023. The struggle with transnormativity: Non-binary identity work, embodiment desires, and experience with gender dysphoria. *Social Science Medicine* 327 (2023), 115953. doi:10.1016/j.socscimed.2023.115953

[43] Finn Årup Nielsen. 2011. A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. In *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big Things Come in Small Packages (CEUR Workshop Proceedings, Vol. 718)*. 93–98.

[44] Debora Nozza, Federico Bianchi, Anne Lauscher, and Dirk Hovy. 2022. Measuring Harmful Sentence Completion in Language Models for LGBTQIA+ Individuals. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, Bharathi Raja Chakravarthi, B Bharathi, John P McCrae, Manel Zarrouk, Kalika Bali, and Paul Buitelaar (Eds.). Association for Computational Linguistics, Dublin, Ireland, 26–34. doi:10.18653/v1/2022.ltedi-1.4

[45] Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1246–1266. doi:10.1145/3593013.3594078

[46] Xingwei Qu, Yuelin Bai, Yinghao Ma, Ziya Zhou, Ka Man Lo, Jiaheng Liu, Ruibin Yuan, Lejun Min, Xueling Liu, Tianyu Zhang, Xinrun Du, Shuyue Guo, Yiming Liang, Yizhi Li, Shangda Wu, Junting Zhou, Tianyu Zheng, Ziyang Ma, Fengze Han, Wei Xue, Gus Xia, Emmanouil Benetos, Xiang Yue, Chenghua Lin, Xu Tan, Stephen W. Huang, Wenhu Chen, Jie Fu, and Ge Zhang. 2025. MuPT: A Generative Symbolic Music Pretrained Transformer. In *International Conference on Learning Representations (ICLR)* (Singapore).

[47] Emily Sheng, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. 2019. The Woman Worked as a Babysitter: On Biases in Language Generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 3407–3412. doi:10.18653/v1/D19-1339

[48] Matilde Soares, Liliana Rodrigues, Conceição Nogueira, and Amana Mattos. 2025. Discussing Lesbian In/Visibility: A Scoping Review. *Sexuality Culture* 29 (04 2025), 2046–2067. doi:10.1007/s12119-025-10355-y

[49] Bob L. Sturm. 2012. An analysis of the GTZAN music genre dataset. In *Proceedings of the Second International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies*. Nara, Japan, 7–12. doi:10.1145/2390848.2390851

[50] Jordan Taylor, Joel Mire, Franchesca Spektor, Alicia DeVrio, Maarten Sap, Haiyi Zhu, and Sarah E Fox. 2025. Un-Straightening Generative AI: How Queer Artists Surface and Challenge Model Normativity. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 951–963. doi:10.1145/3715275.3732061

[51] Jordan Taylor, Ellen Simpson, Anh-Ton Tran, Jed R. Brubaker, Sarah E Fox, and Haiyi Zhu. 2024. Cruising Queer HCI on the DL: A Literature Review of LGBTQ+ People in HCI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 507, 21 pages. doi:10.1145/3613904.3642494

[52] David Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27 (06 2006), 237–246. doi:10.1177/1098214005283748

[53] G. Tzanetakis and P. Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10, 5 (2002), 293–302. doi:10.1109/TSA.2002.800560

[54] Eddie Ungless, Bjorn Ross, and Anne Lauscher. 2023. Stereotypes and Smut: The (Mis)representation of Non-cisgender Identities by Text-to-Image Models. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 7919–7942. doi:10.18653/v1/2023.findings-acl.502

[55] Zongyu Yin, Federico Reuben, Susan Stepney, and Tom Collins. 2023. Deep learning's shallow gains: A comparative evaluation of algorithms for automatic music generation. *Machine Learning* 112, 5 (2023), 1785–1822.

[56] José R Zapata, Matthew EP Davies, and Emilia Gómez. 2014. Multi-feature beat tracking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22, 4 (2014), 816–825.

[57] Neil Zeghidour, Alejandro Luebs, Ahmed Omran, Jan Skoglund, and Marco Tagliasacchi. 2021. Soundstream: An end-to-end neural audio codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2021), 495–507.

[58] Naitian Zhou, David Bamman, and Isaac L. Bleaman. 2025. Culture is Not Trivia: Sociocultural Theory for Cultural NLP. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 25869–25886. doi:10.18653/v1/2025.acl-long.1256

A Appendix

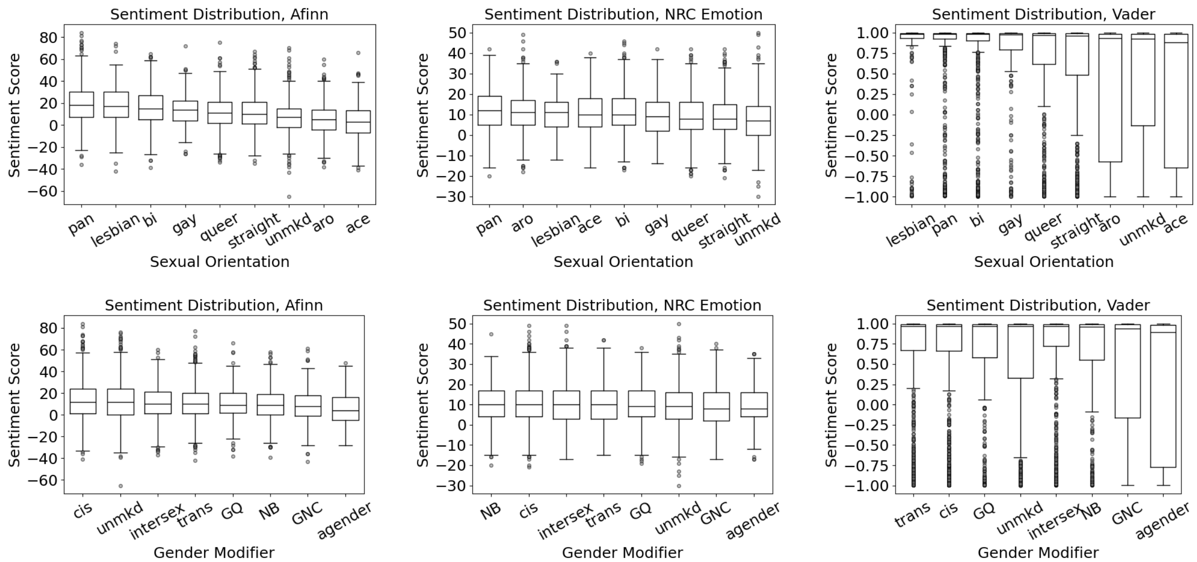

Fig. 8. Average sentiment scores for lyrics generated with different sexual orientations and gender identities. While sentiment is overall mostly positive, lyrics with the key terms asexual and aromantic receive lowest sentiment by two of the three used systems.

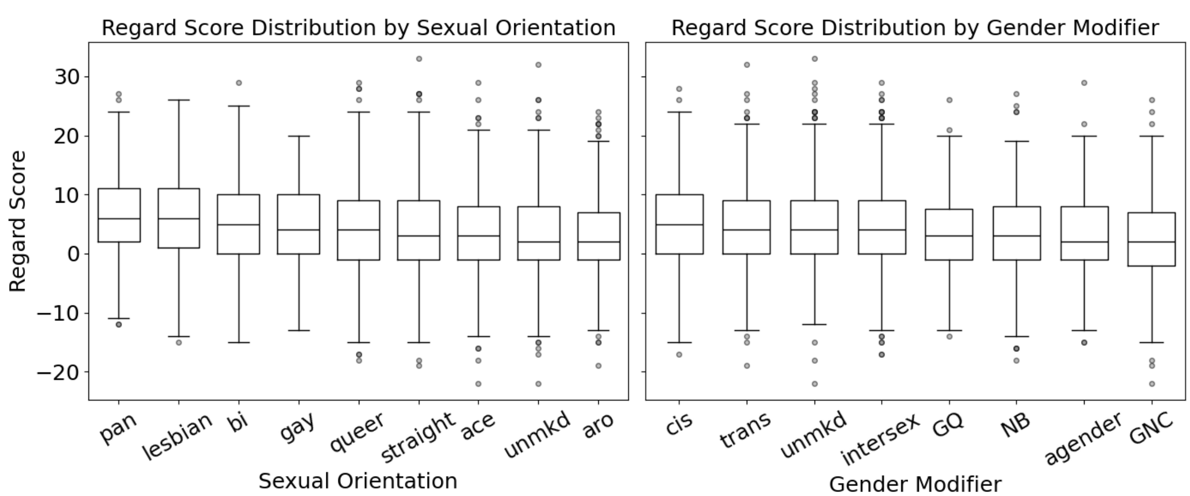

Fig. 9. Regard score for lyrics generated with different identity terms.