

Flattening Fantasies: Analyzing Stereotypes in Pornographic Discourse*

Sarmad Chandio[†], Osama Khalid[†], Sharaf Zia, Ethan Kutlu

University of Iowa

{sarmad-chandio, osama-khalid, sharaf-zia, ethan-kutlu}@uiowa.edu

Abstract

This study presents a large-scale quantitative analysis of 5.4 million video titles from xHamster, the world’s fourth most visited pornography platform, to examine linguistic patterns used to describe different dimensions of identities (e.g., race and gender). Specifically, we investigate how nationalities, ethnic groups, and gender identities are portrayed in pornographic discourse, focusing on stereotypes and deviations from implicit norms of pornography. The results show the dominance of American pornography over other groups, which acts as the de facto norm for the adult industry. We identify persistent ethnic stereotypes reflecting colonial power dynamics, religious fetishization, and the sexualization of cultural taboos. The analysis reveals gendered representational differences across cultures, with a prevalent Madonna-Whore dichotomy in the portrayal of women and stereotypical depictions of men in physically dominant roles.

1 Introduction

Pornography and its consumption patterns have been subjects of academic research for over half a century (Laura 1980; Marcus 1966). Research on this topic has explored not only pornography’s direct relationship to sexuality, but also its associations with violence, sexual assault, and relationship dynamics (Resch and Alderson 2014). With the proliferation of internet access, the use of pornography-related language is becoming more normalized and visible in mainstream discourse. Terms and concepts such as ‘MILF’ (Mother I’d Like to Fuck), ‘thirst’ (excessive sexual desire), and ‘NSFW’ (Not Safe For Work), once confined to pornographic contexts, have gradually permeated the mainstream.

The digital era has ushered in unprecedented access to pornographic content, humorously encapsulated by the mu-

*This paper contains analyses of pornographic language and discusses the perpetuation of ethnic, racial, and gender stereotypes and cultural taboos within that discourse. Some readers may find the language, themes, and discussions of stereotypes disturbing or offensive.

[†]These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

sical Avenue Q in the song “The Internet is for Porn” (Lopez, Marx, and Whitty 2003). This increased availability has sparked new discussions about the content and impact of pornography, bringing to the forefront concerns about representation and the perpetuation of harmful stereotypes. Psychological research has investigated these stereotypes through both laboratory and online studies, where participants report either their own experiences or their perceptions of others’ beliefs regarding different forms of stereotyping (Greenwald and Banaji 1995). However, to date, no quantitative work has been done, at scale, on the language used in pornography to examine the nature of these stereotypes presented, and their prevalence across geographical and cultural spaces. This study seeks to address this gap. We argue that pornography, while ostensibly about deriving pleasure through observing sexual behavior, often provides a private space to explore taboos and fantasies where sexual behavior intersects with social representations of various ethnic, racial, and gender groups. In particular, it can reflect how marginalized and stereotyped groups including women, racialized communities, and sexual minorities, are depicted by dominant groups and portrayed in the content that is uploaded by content creators. As such, it provides a space for researchers to examine how individuals engage with stereotypes through language that is used to describe pornographic content.

Earlier work has already identified how pornography can reduce women’s roles to simplistic caricatures, erasing complexity and reinforcing damaging stereotypes, reducing them to one-dimensional objects of desire. This reductionist approach is considered a product of the ‘male gaze’. Film theorist Laura Mulvey introduces this concept to describe the depiction of women in visual arts and literature from a masculine, heterosexual perspective (1975). This framing is not limited to what men do but reflects a broader societal structure where patriarchal norms dictate how women are viewed and valued. Similar reductionist approaches are evident across history. For instance, the imperial gaze (Fischer 2002) privileged white women primarily for their reproductive roles while simultaneously hypersexualizing women of color, reducing diverse communities of color to essen-

tialized, monolithic categories, erasing multilingual, multicultural identities and reinforcing otherness (Fanon 2016). These dynamics have been widely documented across social science disciplines (Flores and Rosa 2015).

In pornography, this gaze manifests in the portrayal of women primarily as vessels for cis-men’s pleasure, often devoid of agency or emotional depth (Dines 2010). The male gaze in pornography not only objectifies women but also reinforces harmful gender stereotypes, presenting women as constantly available and willing participants in male sexual fantasies (Dworkin 1981).

This reduction can extend beyond sexual objectification to the imposition of rigid, binary tropes that reflect the Madonna–Whore Dichotomy, a psychological and cultural framework that categorizes women as either pure and submissive or sexually deviant. In pornographic discourse, this is mirrored through recurring titles and categories such as ‘teen’, ‘innocent’, or ‘schoolgirl’, which emphasize youth, naivety, and chastity, contrasted with categories like ‘slut’, ‘nasty’, or ‘gangbang’ which signal promiscuity and deviance. These roles carry distinct assumptions about women’s sexual value and morality, and can reinforce harmful binaries that shape broader societal attitudes toward femininity and desire.

This flattening of gendered experiences also intersects with racial and ethnic categories (Cervulle and Rees-Roberts 2009; Nagrath 2003). Rooted in colonial and post-colonial power dynamics, the imperial gaze reduces complex cultural experiences into simplistic, often fetishized categories like ‘Asian’ or ‘Brown’. For instance, titles like “*amateur submissive Asian Wife [...]*”¹ portray Asian women as submissive yet hypersexualized, embodying a fantasy that combines docility with exotic allure. These are not isolated cases but part of a broader pattern repeatedly observed across our dataset.

In pornography, Black men are frequently portrayed through a racialized lens that emphasizes hypermasculinity and hypersexuality. These representations tend to exaggerate physical dominance, aggression, and sexual prowess. While these traits are superficially positive, however they reduce Black male identities to caricatures defined solely by physicality and virility. For example, titles like “*Big black man pounds [...]*” exemplify this pattern, where Black men are repeatedly framed as overpowering, sexually aggressive figures. This persistent framing draws on long-standing racial stereotypes and contributes to a dehumanizing narrative that equates Black male worth with sexual performance and physical domination (Williams 2004).

Such stereotypes, while considered racially insensitive in mainstream discourse, are often amplified in pornography, and can potentially translate into real-world violence and hate crimes (Guardian 2021; Campbell, Hudson, and Ratliff 2023). Prior research has shown that the language and tropes found in pornography closely mirror those used in misogynistic online communities, such as involuntary celibate or ‘incel’ groups (Tranchese and Sugiura 2021).

The potential for pornographic content to shape societal

attitudes and reinforce harmful stereotypes should be understood within the broader media ecosystem, underscoring the importance of systematically studying how different racial and gender categories are represented within pornographic content. Given its widespread and often private consumption, the question becomes: what is the incidence of pornographic stereotypes (i.e. can text data substantiate the presence of stereotypes posited in research noted above), and what is their prevalence (i.e. how widespread is the usage of stereotypical language and terms with demographic and ethnic groups)?

While pornography is primarily consumed as visual media, the experience is mediated by textual and linguistic elements such as search terms, tags, descriptions, titles, etc. These components serve as entry points and framing devices for consumers and, as prior research suggests, may contribute to the perpetuation of stereotypes and the simplification of complex sexual dynamics and identities (Ramírez-Rubio et al. 2022). Rather than treating individual titles as isolated acts of representation, our approach identifies recurring linguistic patterns that position social groups, contributing to broader discourses that normalize and reinforce stereotypical portrayals. Importantly, this research can provide a venue for scholars to investigate what type of processes can be implemented to reduce the perpetuation of these stereotypes and biases in online spaces.

We investigate how gendered and ethnic stereotypes manifest in text by analyzing pornographic video titles. Specifically, we ask the following questions:

- **RQ 1:** What emergent clusters of ethnic representation arise from patterns in the language used in pornographic video titles?

We propose a title-level representation approach to quantitatively analyzing pornographic content. This method allows for a systematic comparison of linguistic patterns in adult content grouped by video-level nationality tags that include labels, such as ‘India’, ‘German’, or ‘American’ that are used by the platform to organize content, into clusters that broadly coalesce across ethnic categories.

- **RQ 2:** How are ethnic stereotypes and cultural taboos represented in language?

We investigate the ways in which linguistic choices in pornographic video titles reflect and engage with ethnic stereotypes and cultural taboos. Our analysis highlights recurring patterns of ethnic exoticization, where identities are represented through simplified, sexualized, and often racialized language. These representations vary across cultural contexts, revealing how particular words, phrases, and themes are used to signify otherness, taboo, or fantasy. This approach allows us to examine how stereotypes are encoded and circulated through the language of pornographic metadata, without presupposing intent or effect.

- **RQ 3:** How are gender roles represented in pornographic titles across different cultural contexts?

We examine how gendered language is used in pornographic titles across different cultural contexts, focus-

¹Please note that all the titles used in examples are from our dataset.

ing on the terms and descriptors associated with masculinity and femininity. By analyzing the semantic relationships between gendered terms within each nationality group, we observe cross-cultural patterns and divergences in how male and female gender roles are constructed. This includes variations in sexual agency, physical descriptors, and relational framing, offering insight into how gendered identities are linguistically encoded within pornographic discourse.

2 Related Work

The study of pornography is an active, multidisciplinary field of research (Haywood 2024; Nocella 2024; Franco and Webber 2024; Cole 2014; Virdis 2015; Moorman 2024; Thorneycroft and Smilges 2023). Much of the research on pornography within Computer Science has focused on network analysis and content distribution. Works such as those by Morichetta et al. (2021) and Ahmed, Shafiq, and Liu (2016) examine HTTP logs to explore patterns in porn traffic across networks. Unlike these approaches, our work aims to examine linguistic patterns in published content, specifically how language is used within pornography to socially construct identity categories and associated stereotypes.

2.1 Pornography and Language

Mulvey’s foundational work (1975) on the ‘male gaze’ in cinema has been instrumental in explaining how visual media, including pornography, often reduces women to objects of male pleasure. Building on this concept, Pihlaja’s study (2016) analyzed around 22,000 comments from the 100 most popular videos on Pornhub. This research revealed that while comments were generally not overtly offensive, they often contained negative descriptions of women, reflecting the objectifying nature of the male gaze in pornographic discourse.

Mach’s work (2022), similar to ours, explored the language of pornography by analyzing a corpus of 17,000 Pornhub video titles, focusing on gender representation and the roles assigned to men and women. Specifically, they examined how agency is constructed through verb-object relationships (primarily using the verb ‘fuck’) and how roles are linguistically encoded using nouns and adjectives associated with each gender. Their method relied on part-of-speech tagging. While useful for quantifying grammatical features (e.g. nouns, verbs etc.), this approach has limited capacity for extracting semantic meaning. More germane to our work, is prior research that has looked at emergent patterns of language and meaning from text on the internet. Caliskan and colleagues (2017) found through a large-scale language corpora that the internet holds the imprints of our historic biases such as associations between gender and professional roles. Crucially, they showed that these biases are not inert; they are replicated and even amplified in algorithmic systems. Subsequent research has shown that algorithmic systems such as large language models (e.g., ChatGPT) can exhibit gender bias in tasks like machine translation, particularly in languages with gender-neutral pronouns (Ghosh and Caliskan 2023). These findings highlight the broader concern that language used in digital environments does not

merely reflect social stereotypes but can also amplify them at scale (Noble 2018; Habib et al. 2024).

2.2 Pornography and Tags

While our study focuses on the language of video titles, we use tags, specifically nationality tags, as a way to organize and group the content for analysis. Tags function as metadata that assign categorical identity to videos (e.g., ‘Indian’, ‘German’), enabling us to compare linguistic patterns across these tagged groups.

Wong et al.’s study (2020) that analyzed video tags for 3 million xHamster videos. By representing tags as a network of relationships, they observed a hierarchy in which some tags are more frequent and dominant while others are niche. However, this study did not examine how tags relate to questions of representation, gender, or race. Similarly, others quantitatively analyzed tags on XNXX and xHamster to explore how nationalities were represented through co-occurring tags (Mazières et al. 2014). We extend this work by using nationality tags as an organizational framework to examine the linguistic construction of stereotypes in video titles.

3 The Pornographic Ecosystem.

3.1 How Pornographic Content Is Produced, Distributed, and Categorized Online

The pornography industry is a multifaceted and complex ecosystem with various players, each contributing to the production, distribution, and consumption of adult content.

Producers in the porn industry include both individual content creators and large production studios like Faphouse and BangBros. These entities have the most autonomy during production, deciding on the contents’ themes, performers, and overall direction.

Consumers are the people who view pornography. Their preferences and viewing habits significantly influence the types of content produced and made available on various platforms.

Platforms such as Pornhub, xHamster, and XNXX dominate the distribution of pornographic content. These websites serve as the primary venues where consumers access and view videos.

Uploaders are a significant part of the ecosystem, with most videos on these platforms being uploaded by pseudonymous users rather than the platforms or producers. These uploaders often re-upload existing videos with new titles, tags, and categories, reflecting their own interpretations and market demands.

The anatomy of a video: In addition to video content, a pornographic video includes associated metadata such as the date of upload, view counts, duration, and comments, along with its title, tags, and categories. The title is a free-text field that reflects the uploader’s description of the video. Tags are specific keywords or phrases that describe details or aspects of the content, such as themes like ‘threesomes’, performers, or scenario-based content like ‘plumber’. Categories are broader classifications that group content into general themes or genres, such as ‘Amateur’, ‘Professional’,

or ‘Popular’ as well as nationalities ‘Chinese’, ‘Danish’ and racial identities ‘Latino’. These fields help organize the content library on the websites, making it easier for users to navigate and find content related to their specific interests.

4 Dataset Description

We utilize publicly available data from xHamster, the fourth most visited pornographic website globally in 2024 (Semrush 2024). Our dataset, collected in July 2024, comprises metadata for 5,412,529 videos, including titles, tags, descriptions, upload dates, engagement metrics, and duration. Figure 2(a) summarizes the rate of video uploads over time. The data shows a steady, linear upload growth until 2019, followed by a sharp spike in 2020 and 2021, likely driven by increased online activity during the COVID-19 pandemic (Zattoni et al. 2021). This surge peaks in 2021 with 709,395 uploads before dropping in 2022 and 2023 as user behavior returns to pre-pandemic trends. The 2024 data, representing uploads for six months (from January to July).

Out of the 5.4 million videos, 3,547,673 were not tagged for any nationality², and 1,841,860 had exactly one nationality tag. The remaining videos had two or more nationalities tagged. The most common nationality was tagged American with 450,865 videos, followed by German and Japanese (215,989 and 146,914 videos, respectively) shown in Figure 2(b). We note nationalities follow a long-tail distribution. In our analysis, we focused on nationalities that had at least 1000 videos. The resulting set of 71 nationalities³ represents 99.5% of the tagged videos.

Figure 1 shows the upload duration of videos from various groups (discussed in section Section 4). Across all groups, we observed that video durations were predominantly within the 5-20 minute range.

This observation is consistent with data from other major pornographic platforms; for instance, PornHub reports an average time spent on the website is 9 minutes and 20 seconds (Pornhub 2024). This finding is also consistent with previous works that found the average session to be 15 minutes or fewer (Morichetta, Trevisan, and Vassio 2019).

While all groups show some degree of right-skewed distribution (indicating the presence of longer videos), the extent of this long tail varies notably between groups. The most prominent long tail is observed in the Asia Pacific group. An extended tail is also visible in the Western European group. This extended right tail indicates a higher proportion of lengthier videos than other groups. The remaining groups show a relatively compact distribution adhering to shorter, more standardized videos.

5 Empirical Findings

5.1 Representing Nationalities and Titles

We define videos based on the nationalities they are tagged for by the uploaders. We group videos into nationality-groups based on these tags. For our analysis, we study

²Nationalities are explicitly tagged in the xHamster data structure.

³In our analysis, we treat the Unmarked videos as a separate nationality

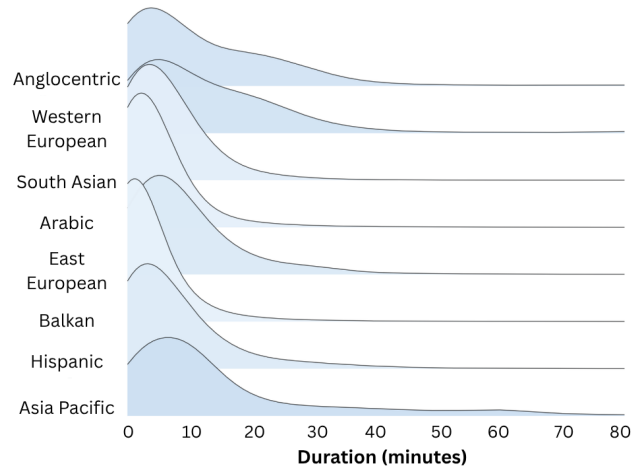


Figure 1: The distribution of video duration for each ethnic cluster. All clusters had a mean video length < 20 minutes.

the language of video titles corresponding to different nationality-groups.

Each video in the pornographic dataset has a title and one or more tagged category⁴. Of the 981 unique categories, 182 specifically correspond to nationalities, which we use to define our nationality-groups. A video tagged for nationality makes the performer’s perceived national identity a central focus, explicitly associating the content with a specific ethnic or cultural background.

Our analysis represents each nationality-specific group using the language in video titles. The free-text of user-generated titles often contains nuanced or descriptive language offering insights into how users and content creators frame their content. We quantify the association between a group c and a word w in video titles by representing all video titles that belong to a group as a single pseudo-document and calculating the TF-IDF (Robertson 2004):

$$\text{TF-IDF}(c, w) = \text{TF}(w|c) \cdot \text{IDF}(w)$$

Where $\text{TF}(w|c)$ is the frequency of the word w in all the videos of c and $\text{IDF}(w)$ is the inverse-document frequency of w in all the groups. We represent each group c as a vector of TF-IDF values of the group and all the words in the vocabulary of all the preprocessed titles⁵.

To illustrate how this method captures group-specific linguistic patterns, consider the following example: if the word ‘maid’ appears frequently in titles tagged with ‘French’ but rarely in titles from other nationality groups, it will have a high TF-IDF score for the ‘French’ group. This would indicate that ‘maid’ is particularly characteristic of how ‘french’ content is linguistically framed. In contrast, a word like ‘sex’ may appear frequently across all groups and thus have a high term frequency but a low TF-IDF score. As such it lacks

⁴The tags/categories are predefined by xHamster and uploaders select from these options to categorize the uploaded video.

⁵The titles were preprocessed using standard NLP operations i.e. lowercasing, removing English stopwords, and lemmatization.

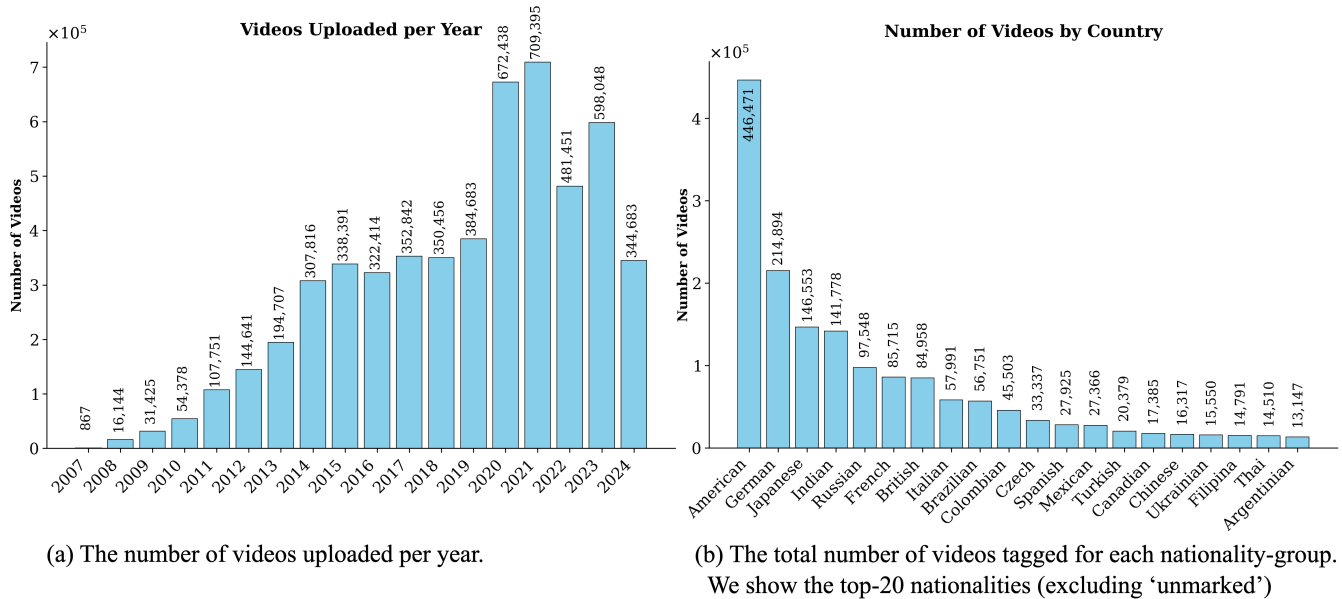


Figure 2: Breakdown of the trends observed in the dataset.

discriminatory power and does not help distinguish the linguistic framing of any particular group. TF-IDF helps isolate language that is specific and meaningful to a group’s representation, rather than just commonly used across all content.

Importantly, this representation accounts for the varying number of videos within each group by inherently adjusting for document frequency across the entire corpus.

Not every video is tagged with a nationality, and the absence of such tags suggests that the performer’s national origin is not emphasized or is treated as default and unmarked. In our analysis, we consider the videos tagged for nationalities and those that are not tagged, allowing us to explore how the presence and absence of explicit nationality may influence representations in the dataset. To minimize tag-spamming⁶, we limit our dataset to only include videos with either one or no nationality tags.

5.2 Clustering of Nationalities

We represent a nationality-group as a vector of the TF-IDF values of all the words across all nationalities. We use Ward’s minimum variance method to cluster (Ward Jr 1963) the TF-IDF vectors.

Figure 4 shows the hierarchical clusters of the TF-IDF vectors. These clusters indicate emergent ethnic categories of shared cultures and geography. Expectedly, nationalities in geographical proximity in the real world (e.g., Pakistani-Indian and German-Swiss) are clustered closer to each other based on their title descriptions. In some cases, however, nationalities cluster not due to geographical proximity but due to similarities in the way they are described in pornographic discourse. For example, American and Australian titles cluster closely, suggesting a convergence in the sexu-

alized language used to represent these groups. We refer to such groupings as “culturally similar” in the sense that the titles evoke overlapping stereotypes, vernaculars, or themes, as identified through unsupervised clustering.

We can broadly categorize the clusters into 8 groups based on either language or region: Anglosphere (American, Australian, Canadian, Nigerian), Western European (German, Swiss, French), South Asian (Indian, Bangladeshi, Pakistani), Middle-East and North Africa — MENA (Moroccan, Saudi, Egyptian), Eastern European (Russian, Ukrainian), Balkan (Serbian, Croatian, Turkish), Hispanic (Spanish, Mexican, Argentine), and Asia Pacific (Japanese, Chinese, Malaysian, Indonesian). To explore how language is used to describe ethnic categories in pornographic content, we focus on five nationalities as exemplars: French (Western European), Chinese (Asia Pacific), Saudi (MENA), Mexican (Hispanic), and Indian (South Asian). These exemplars were selected to represent a wide range of cultural, geographic, and linguistic clusters, while also reflecting nationalities with substantial representation in the dataset and distinctive stereotype patterns. Together, they allow for a comparative analysis across dominant cultural tropes in different regions.

5.3 Cultural and Gendered Differences in Porn

Cultural differences. Cultures and cultural identities are complex constructs that can not be modeled simplistically. We use nationality tags as proxies of cultural representations. Even though this model is not a perfect reflection of culture, it provides us with a useful lens to study representations in pornography.

We investigate whether the representation of a particular nationality-group c significantly differs from the general pornographic representation. Our null hypothesis posits no

⁶tag-spamming: excessive or irrelevant tagging.

similarity = 0.93), indicating the centered-ness of the category in pornographic discourse (Figure 4). The high similarity of ‘American’ can not be attributed to its volume alone. We find ‘German’ — the 3rd largest group by volume (4% of total videos) — only had a 0.53 cosine similarity with the random universe, whereas ‘Canadian’ the 16th most populous (0.3% of total videos) category had a much higher similarity (0.80).

The globalization of pornographic language. The general pattern reflects a gradient of ‘Westernization’ in pornographic discourse, with ‘Unmarked’ and ‘American’ as the defacto norms. This pattern, however, is not uniform across all nationality-tagged groups. Some non-Western nationalities showed high similarity to the random category, notably ‘Colombian’(0.72) and ‘Nigerian’ (0.71), suggesting the adoption of Western pornographic language norms. For these two groups, among the top-20 words with the highest TF-IDF scores in these groups, most were sexually explicit generic terms like ‘hot’, ‘sex’, and ‘fuck’, with only ‘nigeria’ and ‘latino’ hinting at local content.

On the other hand, Asian and Middle-Eastern nationalities predominantly use localized language in titles. This reinterpretation and reimagining of pornographic discourse into local vernaculars manifest in the use of region-specific terms, ‘fuck’ is replaced by ‘chadai’ in the ‘Indian’ context, ‘sex’ with ‘entot’ in the ‘Indonesian’, ‘blowjob’ with ‘sakso’ in ‘Turkish’ discourses.

We hypothesize that deviations from the (Western) norm can primarily be attributed to two factors, First, using local languages in titles, indicating that video uploaders are likely native speakers or fluent in these languages. Second, the development of culture-specific pornographic genres and terminology. One example is the relatively high use of ‘hentai’ with ‘Japanese’.

5.4 Ethnic Stereotypes

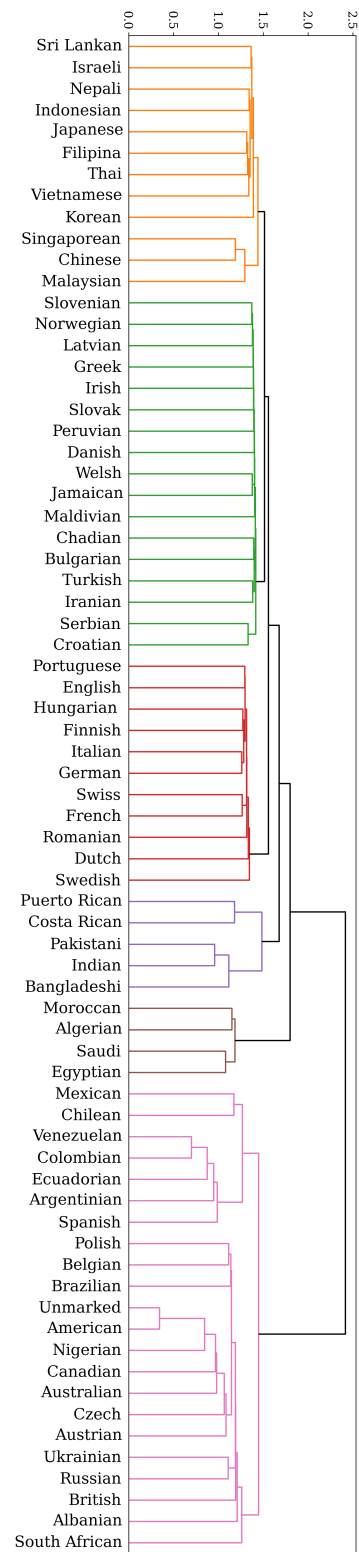
Ethnic stereotyping in pornographic titles can be divided into two categories: outgroup stereotyping and ingroup stereotyping. Outgroup stereotyping manifests as ‘Otherization’ and involves ethnic fetishization, while ingroup stereotypes engage with culturally specific taboos and their subversion.

Outgroup

Colonial Otherization. Orientalism, according to scholars (Said 1979), is fundamentally about the power to define and represent outgroups, racial and ethnic categories distinct from one’s own. The language of pornography also represents an orientalist gaze that reinforces the relationship between the colonizer ‘Self’ and the colonized ‘Other’. This is apparent in the language used in the pornography of the former colonizer (‘French’) and the formerly colonized (‘Arab’). A prominent example is the use of ‘beurette’ in both the Saudi and French sets.

This French slang describes female immigrants from the Arab Maghreb. It appears in video titles like: “*danseuse marocaine arab beurette fesse ass*” (trans: Moroccan dancer arab beurette spans ass)

Some scholars argue that Maghrebi women have been fetishized since the start of the French colonial rule in North



(a)

Figure 4: Dendrogram of title-level vector similarity. Lower branch connections indicate greater similarity. Cutting the tree at different levels reveals various clustering arrangements

Africa (Hafsa 2017). French colonizers in Algeria categorized prostitutes by their ethnic and physical characteristics. Such characterization continues to this day and is represented in and reinforced by the language of pornography we look at for French porn. Such stereotypical representation continues to dehumanize women from minority groups to satisfy a dominant ingroup's sexual fetishization of a minority outgroup (Taraud 2011).

Religion. Religious symbols and identities are similarly appropriated and sexualized in porn, particularly within Middle Eastern contexts. In the 'Saudi' set, terms like '*hijab*' and '*muslim*' were in the top-20 words when ranked by TF-IDF scores. They were used in titles like "*hijab arab lady* [...]" and "*shameless muslim hijabi*⁷ [...]" suggesting a fetishization of religious taboos and the intersection of cultural and religious identity within sexual content. The rise of 'porno ethnique' (ethnic porn) in cinema, and the use of hijab as a form of striptease, can be thought of as a by-product of colonial repression and ridicule of symbols deemed incompatible with Western ideas of progress (Mack 2017).

Such religious fetishization indiscriminately orientalistizes and codes performers from the Middle-East as Muslim. For example, titles like "*muslim pornstar mia khalifa sucks* [...]" code Mia Khalifa as Muslim, based on her brownness, disregarding her actual Christian identity for the sake of the video (Bella 2018).

The ethnic exoticization observed, especially in the context of former colonial relationships, highlights how historical power dynamics continue to shape sexual representation and fantasy. The fetishization of cultural and religious symbols, such as the hijab, further underscores the complex interplay between sexuality, culture, and power. The focus on the hijab not only exoticizes the Other but also mirrors the objectification found in the sexualized exposure of women's nudity. While one renders power through concealment and symbolic cultural markers, the other does so through exposure and bodily availability. Both function as mechanisms of objectification that reduce women's bodies to signifiers of erotic difference. These findings raise important questions about the role of pornography in perpetuating harmful stereotypes and colonial tropes. They also highlight the need for more critical engagement with pornographic content, as a reflection of broader societal attitudes towards gender, race, and sexuality.

Ingroup

Cultural specific taboos. Previous studies have argued that pornography is a domain where otherwise taboo sexual fantasies are realized, and it is used as a tool to release pent-up sexual repression (Ellis 2006; Mondin 2017). We find sexualization within kinship relations and domestic settings as a manifestation of taboo fantasies. As an example, within the 'Indian' category, terms like '*bhabhi*' (sister-in-law) or '*kaamwali*' (maid) frequently appear in titles, as in "*bhabhi cheating* [...]" and "*Indian Kaamwali Maid Fucked by* [...]". This sexualization of familial and household roles reflects a distinct form of exoticization that plays on cultural taboos.

⁷'Hijab' is Arabic for veil. 'Hijabi' is a person who wears hijab.

However, this exoticization is not only specific to the Indian context. We also find this phenomenon of sexualizing visible in other cultures. Pornographic categories tagged '*mom*' in 'American' and '*maid*' in 'Japanese' or 'French' cultures (Figure 5) represent similar taboo fantasies and serve similar purposes, invoking an immediate caretaker fantasy for men fantasizing about sex with women who are in close physical space to them.

We find the pervasiveness of patterns of ethnic stereotyping and exoticization within pornographic discourse, as evidenced by the use of racialized language in the set of the top-20 TF-IDF words for nationalities like 'French' and 'Saudi'. This exoticization is not only limited to outgrouping or racialization, it also presents itself as an artifact of cultural repression and fantasies as illustrated by kinship and caretaking themes in various communities, including but not limited to 'Indian', 'Japanese', and 'American'.

While the findings above do not fully address the extent to which pornography reinforces and potentially amplifies societal biases related to race, ethnicity, and culture, the observed incidence and prevalence of stereotype-rich language underscores the need to critically examine how ethnic and cultural identities are represented in adult content.

Gendered differences. Pornography, as a reflection of societal attitudes and desires, often presents exaggerated and stereotypical representations of gender. While sexualization is prevalent across all genders in pornographic content, the nature and extent of this sexualization can vary significantly between masculine and feminine portrayals. We explore and quantify these differences, shedding light on the ways in which gender is constructed and presented in adult content across different cultural contexts. Due to the limitations of the data collection, we used man/woman dichotomy to represent gender differences and, therefore, were unable to implement non-binary identities.

Method. We use Word2Vec (Mikolov et al. 2013) to create word-embeddings of the pornographic video titles from each nationality-group in our dataset. These embeddings capture the semantic relationships between words and concepts specific to each cultural context. Titles serve as concise summaries of content, capturing key themes, performers, and actions in a highly condensed format.

These embeddings allow us to quantify semantic relationships by measuring vector similarities between pornographic concepts and gendered reference vectors. For instance, if we find that the word '*maid*' is in closer proximity to the feminine vector compared to the masculine vector, this would indicate that the term is more semantically aligned with femininity in the context of pornographic discourse.

We use previous methods from studies to create representative gender vectors by averaging the embedding vectors of gender-specific words and pronouns (Bolukbasi et al. 2016). For the masculine dimension, we average the embeddings of the masculine words, while for the feminine dimension, we use the embeddings of the feminine words⁸. Given a word

⁸masc. words = ['man', 'he', 'his', 'him', 'lad', 'boy', 'men', 'guy', 'male']

fem. words = ['woman', 'she', 'her', 'female', 'lady', 'girl', 'lass']

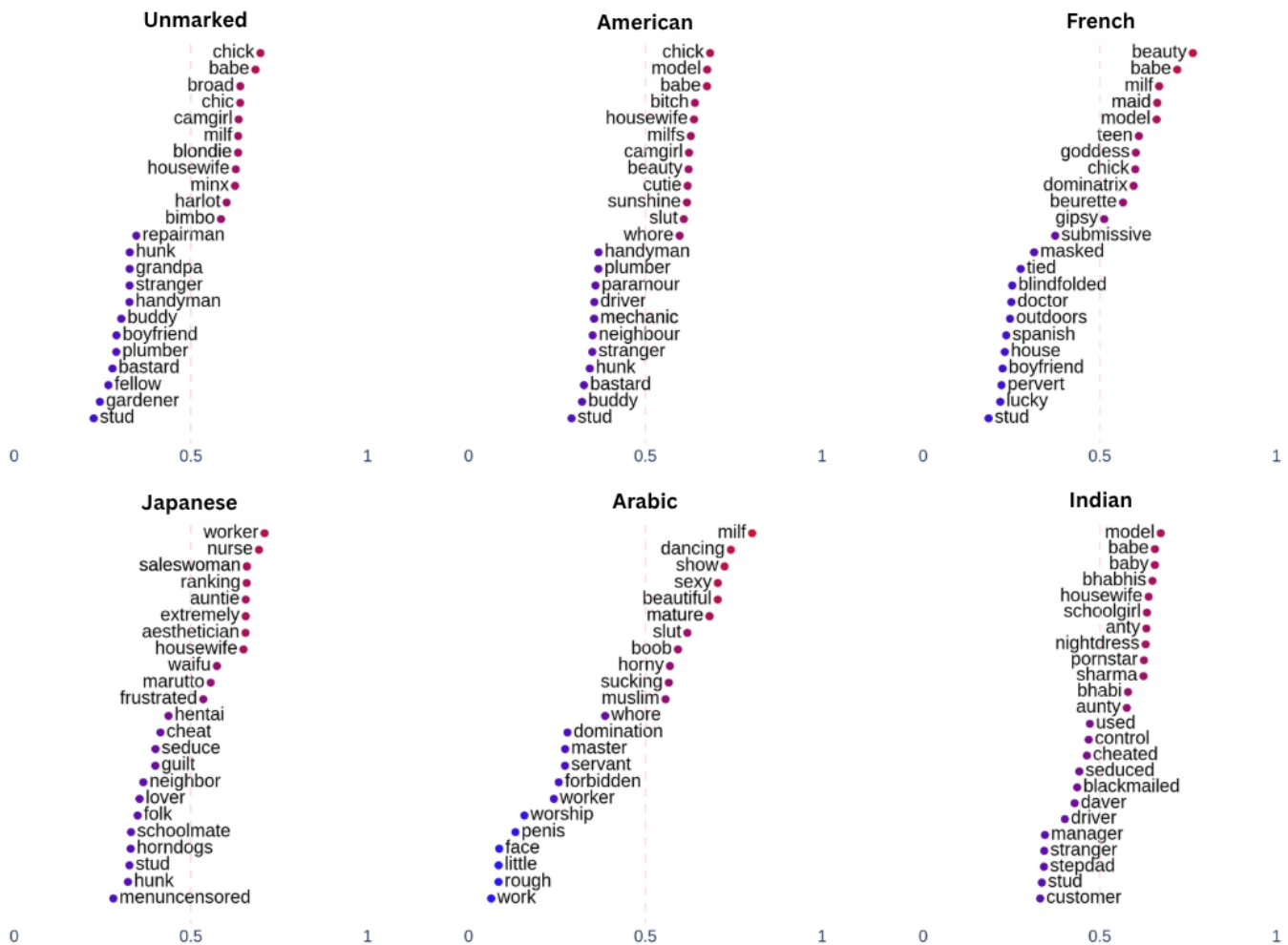


Figure 5: The gender association ratio $G(w)$ of a sample of the words (discussed in Section 5.3) of the 6 exemplars (we use Arabic nationality-group due to the small size of its exemplar). Words above 0.5 share a greater context to the feminine vector, and words below 0.5 share a greater context to the masculine vector. For instance, in the Unmarked category (top-left), the word ‘stud’ ($G(w) = 0.28$) shares a greater context with the masculine vector than the feminine vector and vice versa for the word ‘chick’ ($G(w) = 0.70$).

w in our embedding-space, we represent the relative gender association $G(w)$ as :

$$G(w) = \frac{\cos(\vec{w}, \vec{m})}{\cos(\vec{w}, \vec{f}) + \cos(\vec{w}, \vec{m})}$$

where $\cos(\vec{w}, \vec{m})$ and $\cos(\vec{w}, \vec{f})$ represent the cosine distances of the embedding of w from the average masculine and feminine vectors, respectively.

Evaluation. The $G(w)$ ratio ranges from 0 to 1, where values closer to 1 indicate stronger feminine associations, values closer to 0 suggest stronger masculine associations, and 0.5 represents equal association with both genders.

Unmarked and American portrayals In the ‘Unmarked’ category, feminine descriptors predominantly emphasize physical attributes. Conversely, masculine descriptors often relate to occupations traditionally associated with manual labor or technical skills (e.g., “*Blonde MILF wants a plumber*

to fuck her”). The unmarked baseline establishes a general trend of objectification for women and occupational identity for men in pornographic content.

The representation of men in ‘American’ follows the same pattern as the ‘Unmarked’ category and is closely associated with labor-intensive professions. In contrast, women follow an appearance-heavy slant where words such as ‘*beauty*’, ‘*blondie*’, ‘*model*’, and ‘*chick*’ generally refer to their appearance or hypersexualized character and are used in titles like “*Beautiful slattern gives ... her [· · ·]*”.

Western European depictions. Within the Western European group, we find associations between French women and the word ‘*maid*’. While the fetishization of the maid reflects a power dynamic emphasizing a master-servant relationship (Apter 2018), in the ‘French’ context, as noted earlier, it also demonstrates a racialization of the power dynamic as the ‘maid’ trope is being used in conjunction with

racialized groups (“*French Arab Muslim maid*”).

On the other hand men are linked with terms such as ‘*pervert*’, ‘*tied*’, ‘*masked*’, and ‘*submissive*’, that suggest a prevalence of BDSM themes and a subversion of traditional masculine dominance in French pornography, as hinted by titles like (“*french dominatrix fuck submissive husband [···]*”; “*tied hand stepbrother [···]*”)

Asian portrayals. In the ‘Japanese’ set, we observe a distinct characterization of male roles that focuses on themes of moral transgression and power imbalances, and is indicated by association of words such as ‘*cheat*’, ‘*seduce*’, ‘*neighbor*’, and ‘*guilt*’ leaning towards the masculine ($G(w) < 0.5$).

These themes manifest in titles such as “[···] *cheating husband*” and “[···] *teacher seduces schoolgirl*”, highlighting a prevalent narrative of men in positions of trust or authority violating social norms or ethical boundaries. The “*teacher seduces schoolgirl*” trope, in particular, underscores a problematic power dynamic where men in authoritative roles exploit their position to initiate sexual encounters with women subordinates or dependents. This representation contrasts with, and complements, the portrayal of women in the Japanese set, who are often depicted in service-oriented or submissive roles like ‘*nurse*’ and ‘*saleswoman*’ in titles like (“*Door-to-door saleswoman offers more than just goods*”; “*beautiful japanese nurse [···]*”).

Madonna-Whore Dichotomy. The Madonna-Whore Dichotomy (MWD) is a psychological framework that categorizes women into two mutually exclusive groups: the virtuous, pure ‘*Madonnas*’ and their sexually promiscuous counterparts, ‘*whores*’. This dichotomy reflects a deeply problematic view of female sexuality and personhood (Bareket et al. 2018). Finding evidence of MWD in pornographic content is significant because it suggests that these materials can not only reflect but potentially reinforce harmful stereotypes about women. This can contribute to the objectification of women, difficulties in forming healthy relationships, and the perpetuation of sexist attitudes in society at large (Kahalon et al. 2019). The presence of MWD in pornography across different cultures indicates that this is a widespread issue and not a localized issue for particular sub-groups.

Universally, we find sexist language targeted against women across almost every group. This includes derogatory terms like ‘*sharmota*’ (Arabic for whore) in the Saudi set and ‘*prostituta*’ (Spanish for prostitute) in the Mexican set. We also observe sex acts indicating dominance over women, such as ‘*facialisee*’ (French for the sex-act: facialized) in the French set.

Interestingly, kinship terms are also sexualized, like ‘*Mer-tua*’ (Indonesian for Mother-in-Law) or ‘*Bhabi*’ (Hindi for sister-in-law). This dichotomy of women being sexualized and dehumanized as ‘*bad*’, ‘*promiscuous*’, and ‘*seductive whores*’, as well as the use of kinship terms to indicate them as either ‘*pure*’ and ‘*chaste*’, signals the presence of a Madonna-Whore Dichotomy in pornographic language (Bareket et al. 2018).

6 Discussion

In this study, we present a quantitative analysis of 5.4 million video titles from xHamster to examine linguistic patterns used to describe different dimensions of identities (e.g., race and gender). Specifically, we investigate how nationalities, ethnic groups, and gender identities are portrayed in pornographic discourse, focusing on stereotypes, their nature and prevalence across pornographic titles. We summarize our findings below for their cross-cultural patterns and colonial ideologies in pornography.

Our analysis reveals consistent patterns in the ethnic stereotyping and over-sexualization of women. Women are frequently portrayed as objects of male desire, with their value tied to their ethnic and racial background for outgroup members, and their status within domestic frameworks tied to their value as sexual objects for ingroup members pursuing sexual taboos. In both cases, however (ingroup and outgroup), a woman’s appearance and her ability to please men is placed at the core of pornographic content with words like ‘*beauty*’, ‘*housewife*’, and ‘*model*’ etc. used to describe women. In contrast, men are generally depicted as dominant, masculine figures who engage in physically active, male dominated fields — ‘*plumbers*’, ‘*construction workers*’, etc. In both cases, we note small divergences from the norm where usual masculine and feminine stereotypes are subverted.

The prevalence of stereotyping and exoticizing across different nationalities also suggests that flattening the diverse sexual behaviors, appetites, and needs of women, particularly those from racial and ethnic minorities, is a widespread issue in pornographic discourse, and not simply limited to any single cultural context. It highlights the need for more critical engagement with how women are represented within pornography, and how ethnicity and culture factor into these representations, and what, if any, are the downstream effects on consumers who continue to watch this content.

As noted above, while the hypersexualization of women was observed in all 8 subgroups, the nature and degree of hypersexualization differed based on which ethnic group was under consideration. Our analysis reveals that historical power dynamics, especially sustained through former colonial relationships, continue to shape sexual representation and fantasy creation in pornographic content. These patterns suggest that pornographic discourse often flattens complex identities into simplistic, stereotypical portrayals that can reinforce harmful biases. While the dataset is extensive, it centers on a single platform, examines only title-level metadata, and focuses on nationality and gender without addressing intersecting identity dimensions such as sexuality, body type, or age. Future research would also benefit from studying the provenance and permeation of pornographic language in the mainstream and gauging its impact on mainstream culture.

6.1 Limitations and Future Work

Our analysis focuses solely on nationality as a category, an approach that, though informative, focuses only on a narrow aspect of individual identity in an area (pornography) where a multitude of identities co-exist and interact simul-

taneously to generate and guide sexual behavior. Future research should examine how nationality intersects with other facets of identity, such as sexual orientation, age, and body type in pornographic content. This would provide a more comprehensive picture of how the influence of various identifiers informs characterization and especially stereotyping within porn.

Additionally, our data comes from a single, albeit major, adult website. A broader analysis across multiple platforms could reveal whether these patterns are consistent across the industry or specific to certain contexts. Longitudinal studies could also shed light on how these representations have evolved over time, revealing what interpretations and characterizations have endured and which ones have been retired.

Further, as stated above, while our work shows that problematic language exists in porn that reinforces biases present in the real world, the extent to which such language permeates into mainstream discourse remains to be seen. Specifically, a language diffusion analysis can be done to understand pathways of language adaption from porn into mainstream culture. Reddit communities, some of which are centered around the consumption of porn (eg. r/watchitfortheplot), are ideal for such analyses as they represent an opportunity to see how the language of porn has informed mainstream discourse around sex, especially as adult content.

Finally, the results of this study underscore the need for more critical engagement with pornographic content, not just as a matter of personal consumption, but as a barometer of broader cultural biases and power dynamics. As pornography continues to be a significant part of many people's media consumption, understanding and addressing when and how representation impacts consumers will become increasingly crucial for fostering more equitable and respectful societal attitudes and healthy sexual behaviors.

References

2019. hiQ Labs, Inc. v. LinkedIn Corp. Opinion by Judge Berzon; Concurrence by Judge Wallace.
- Ahmed, F.; Shafiq, M. Z.; and Liu, A. X. 2016. The internet is for porn: Measurement and analysis of online adult traffic. In *2016 IEEE 36th international conference on distributed computing systems (ICDCS)*, 88–97. IEEE.
- Apter, E. 2018. 8 Master Narratives/Servant Texts: Representing the Maid from Flaubert to Freud.
- Aylo. 2024. Our Brands. Accessed: 2024-10-09.
- Bareket, O.; Kahalon, R.; Shnabel, N.; and Glick, P. 2018. The Madonna-Whore Dichotomy: Men who perceive women's nurturance and sexuality as mutually exclusive endorse patriarchy and show lower relationship satisfaction. *Sex Roles*, 79(9): 519–532.
- Bella, T. 2018. You Don't Know Mia Khalifa. *Playboy*. Accessed 29 September 2024.
- Bolukbasi, T.; Chang, K.-W.; Zou, J. Y.; Saligrama, V.; and Kalai, A. T. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
- Caliskan, A.; Bryson, J. J.; and Narayanan, A. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334): 183–186.
- Campbell, J. T.; Hudson, S.-k. T. J.; and Ratliff, K. A. 2023. The influence of perceiver and target race in hostile and benevolent sexist attitudes. *Sex Roles*, 89(11): 644–657.
- Cervulle, M.; and Rees-Roberts, N. 2009. Queering the Orientalist porn package: Arab men in French gay pornography. *New Cinemas: Journal of Contemporary Film*, 6(3): 197–208.
- Cole, K. L. 2014. Pornography, censorship, and public sex: exploring feminist and queer perspectives of (public) pornography through the case of Pornotopia. *Porn Studies*, 1(3): 227–241.
- Dines, G. 2010. *Pornland: How porn has hijacked our sexuality*. Beacon Press.
- Dworkin, A. 1981. *Pornography: Men Possessing Women. Women's*.
- Ellis, J. 2006. On pornography. *Pornography: Film and Culture*, 25–47.
- Fanon, F. 2016. Black skin, white masks. In *Social theory re-wired*, 394–401. Routledge.
- Fischer, K. 2002. *The imperial gaze: Native American, African American, and colonial women in European eyes*. Oxford: Blackwell Publishers.
- Flores, N.; and Rosa, J. 2015. Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard educational review*, 85(2): 149–171.
- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.
- Franco, R.; and Webber, V. 2024. 'This is fucking nuts': the role of payment intermediaries in structuring precarity and dependencies in platformized sex work. *Porn Studies*, 1–18.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Ghosh, S.; and Caliskan, A. 2023. Chatgpt perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across bengali and five other low-resource languages. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 901–912.
- Greenwald, A. G.; and Banaji, M. R. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1): 4.
- Guardian, T. 2021. Man accused of killing eight people in Atlanta spa shootings indicted on murder charges.
- Habib, H.; Stoldt, R.; High, A.; Ekdale, B.; Peterson, A.; Biddle, K.; Ssozi, J.; and Nithyanand, R. 2024. Algorithmic amplification of biases on Google Search. *arXiv preprint arXiv:2401.09044*.
- Hafsa. 2017. The Beurette – A Symbol of The Sexualization of Maghrebi Women in French Cinema.
- Haywood, C. 2024. Becoming pornographic, being ethnographic: exploring the heteroerotic, hierarchies of desire and affective atmospheres in swinger clubs. *Porn Studies*, 1–15.

- Jacquemien, D. 2021. Operators of the notorious porn site xHamster unmasked. *Bluewin*. Accessed: 2025-09-15.
- Kahalon, R.; Bareket, O.; Vial, A. C.; Sassenhagen, N.; Becker, J. C.; and Shnabel, N. 2019. The Madonna-whore dichotomy is associated with patriarchy endorsement: Evidence from Israel, the United States, and Germany. *Psychology of Women Quarterly*, 43(3): 348–367.
- Laura, L. 1980. Take Back the Night: Women on Pornography. *New York, William Morrow*.
- Lopez, R.; Marx, J.; and Whitty, J. 2003. Avenue Q. Live performance. Musical, premiered in 2003. Winner of Best Musical, Book, and Score at the 2004 Tony Awards.
- Mach, T. 2022. Smitten by the Written: Corpus Analysis of Agency and Roles in Pornographic Video Titles.
- Mack, M. A. 2017. *Sexagon: Muslims, France, and the sexualization of national culture*. Fordham Univ Press.
- Marcus, S. 1966. *The Other Victorians: A Study of Sexuality and Pornography in Mid-nineteenth-century England*. A Bantam book. Basic Books. ISBN 9780465053865.
- Mazières, A.; Trachman, M.; Cointet, J.-P.; Coulmont, B.; and Prieur, C. 2014. Deep tags: toward a quantitative analysis of online pornography. *Porn Studies*, 1(1-2): 80–95.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Mondin, A. 2017. ‘Tumblr mostly, great empowering images:’blogging, reblogging and scrolling feminist, queer and BDSM desires. *Journal of Gender Studies*, 26(3): 282–292.
- Moorman, J. 2024. Bi for pay? On notions of authenticity and women’s ‘compulsory bisexuality’ in US adult video. *Porn Studies*, 1–19.
- Morichetta, A.; Trevisan, M.; and Vassio, L. 2019. Characterizing web pornography consumption from passive measurements. In *International Conference on Passive and Active Network Measurement*, 304–316. Springer.
- Morichetta, A.; Trevisan, M.; Vassio, L.; and Krickl, J. 2021. Understanding web pornography usage from traffic analysis. *Computer Networks*, 189: 107909.
- Mulvey, L. 1975. Visual pleasure and narrative cinema. *Media and cultural studies: Keywords*, 342–352.
- Nagrath, S. 2003. (En) countering orientalism in high fashion: A review of India fashion week 2002. *Fashion Theory*, 7(3-4): 361–376.
- Noble, S. U. 2018. Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press.
- Nocella, R. R. 2024. The stigma around porn work: inhibiting the enforcement of health and safety regulations. *Porn Studies*, 1–20.
- Pihlaja, S. 2016. Expressing pleasure and avoiding engagement in online adult video comment sections. *Journal of Language and Sexuality*, 5(1): 94–112.
- Pornhub. 2024. Long vs. Short: How Your Porn Watching Time Has Changed. Accessed: 2024-10-09.
- Ramírez-Rubio, A.; Villanueva-Moya, L.; Sánchez-Hernández, M. D.; Herrera, M. C.; and Expósito, F. 2022. Social perceptions of women in pornography: Attitudes and gender stereotypes. *Psicologia sociale*, 17(2): 277–307.
- Resch, M. N.; and Alderson, K. G. 2014. Female partners of men who use pornography: Are honesty and mutual use associated with relationship satisfaction? *Journal of sex & marital therapy*, 40(5): 410–424.
- Robertson, S. 2004. Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of documentation*, 60(5): 503–520.
- Said, E. 1979. *Orientalism*. Near eastern studies : Cultural history. Knopf Doubleday Publishing Group. ISBN 9780394740676.
- Sandvig, C.; Hamilton, K.; Karahalios, K.; and Langbort, C. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, 22(2014): 4349–4357.
- Semrush. 2024. Trending Websites - Global (Adult). Accessed: 2024-10-09.
- Taraud, C. 2011. La prostitution féminine juive dans l’Algérie coloniale. *Archives Juives*, 44(2): 77–85.
- Thorneycroft, R.; and Smilges, J. L. 2023. Choosing disability: bug chasing and gift giving in gay pornography. *Porn Studies*, 10(1): 99–115.
- Tranchese, A.; and Sugiura, L. 2021. “I don’t hate all women, just those stuck-up bitches”: How incels and mainstream pornography speak the same extreme language of misogyny. *Violence against women*, 27(14): 2709–2734.
- Viridis, D. F. 2015. Eroticizing female and male bodies: a linguistic investigation of a pornographic novel from the Victorian magazine *The Pearl*. *Porn Studies*, 2(1): 19–34.
- Ward Jr, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301): 236–244.
- Williams, L. 2004. Skin flicks on the racial border: Pornography, exploitation and interracial lust. *FGS–Freiburger GeschlechterStudien*, 10(2): 9–10.
- Wong, C.; Song, Y.-D.; and Mahanti, A. 2020. YouTube of porn: longitudinal measurement, analysis, and characterization of a large porn streaming service. *Social Network Analysis and Mining*, 10: 1–19.
- Zattoni, F.; Gül, M.; Soligo, M.; Morlacco, A.; Motterle, G.; Collavino, J.; Barneschi, A. C.; Moschini, M.; and Moro, F. D. 2021. The impact of COVID-19 pandemic on pornography habits: a global analysis of Google Trends. *International journal of impotence research*, 33(8): 824–831.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disre-

spect to societies or cultures? Yes, and the research focuses on analyzing publicly available data from xHamster and aims to critique the representation of stereotypes in adult content, which aligns with broader ethical concerns regarding media literacy and bias critique. The paper describes how stereotypes are perpetuated rather than reinforcing them.

- (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the abstract and introduction clearly state that the paper addresses the prevalence of ethnic, racial, and gender stereotypes in pornographic discourse.
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we use TF-IDF and Ward's method to analyze linguistic patterns, which is appropriate for identifying and quantifying stereotypical language across large datasets. The methods are clearly explained in the methodology section.
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, the paper discusses how the dataset is based solely on title-level metadata and nationality tags, which are limited in capturing the full complexity of ethnic and gender identities. We mention the potential for overrepresentation of certain nationalities and the limitations of using only xHamster data.
- (e) Did you describe the limitations of your work? Yes, the limitations are discussed in the "Limitations and Future Work" section, where the authors mention the narrow focus on nationality, and the use of a single platform.
- (f) Did you discuss any potential negative societal impacts of your work? Yes, the paper discusses how its findings could reinforce harmful stereotypes if misinterpreted, and emphasizes the need for critical engagement with pornographic content to avoid legitimizing stereotypes.
- (g) Did you discuss any potential misuse of your work? Yes, the potential for reinforcing stereotypes through the paper's analysis of pornographic language is acknowledged. We have tried our best to avoid amplifying harmful representations.
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, the paper adheres to ethical scraping practices and ensures the data does not include personal or sensitive information. It also provides a link to the notebook used to process the data for transparency.
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...

- (a) Did you clearly state the assumptions underlying all theoretical results? Yes, the assumptions underlying

the analysis are clearly described in the methodology, particularly regarding the use of TF-IDF and the grouping of videos by nationality tags.

- (b) Have you provided justifications for all theoretical results? Yes, the results are grounded in theoretical frameworks drawn from existing literature on stereotypes, Orientalism, film theory and the male gaze.
- (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, we reference previous works that explore similar themes and contrast our findings with other theoretical perspectives.
- (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, alternative explanations, such as cultural differences in media consumption and platform-specific language use, are considered in the paper.
- (e) Did you address potential biases or limitations in your theoretical framework? Yes, we acknowledge the potential for biases in our interpretation of the data, especially considering the limitations of analyzing a single platform.
- (f) Have you related your theoretical results to the existing literature in social science? Yes, our work builds on existing frameworks in gender studies, and includes discussion on the male gaze, Orientalism, and gender representation in media. We relate our findings to broader discussions of power dynamics and stereotype perpetuation.
- (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, our work calls for more critical engagement with pornography and its societal implications, we argue for further research into media literacy and the effects of pornographic content on public perceptions of gender and race.

3. Additionally, if you are including theoretical proofs...

- (a) Did you state the full set of assumptions of all theoretical results? N/A
- (b) Did you include complete proofs of all theoretical results? N/A

4. Additionally, if you ran machine learning experiments...

- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we provide a link to the notebook used for data processing and analysis, which is available at an OSF repository.
- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? We do not train a machine learning model
- (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? N/A

- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? *N/A*
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? *N/A*
 - (f) Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? *N/A*
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity...**
- (a) If your work uses existing assets, did you cite the creators? *Yes, the paper cites relevant datasets and research, including previous studies on pornographic content and language.*
 - (b) Did you mention the license of the assets? *N/A*
 - (c) Did you include any new assets in the supplemental material or as a URL? *Yes, the dataset and associated analysis notebook are made available via an OSF repository link for reproducibility.*
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? *Yes, the data used in the study is publicly accessible and no personally identifiable information is included in the analysis.*
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? *Yes, the data does not contain personally identifiable information and that the analysis focused on public metadata. We do acknowledge that because of the sensitive nature of the topic, some of the content might be offensive or disturbing to some readers.*
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? *N/A*
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Geburu et al. (2021))? *N/A*
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity...**
- (a) Did you include the full text of instructions given to participants and screenshots? *NA, this work does not involve human subjects or crowdsourcing.*
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? *NA, this work does not involve human subjects.*
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? *NA, this work does not involve human subjects.*
 - (d) Did you discuss how data is stored, shared, and deidentified? *NA, this work does not involve human subjects.*

A Appendix

A.1 Data Collection Ethics

This research involves the systematic scraping of data from the adult website, xHamster, conducted strictly within the platform’s rate limits to minimize any impact on its infrastructure. Our crawls were consistent with typical auditing studies (Sandvig et al. 2014) and are legally permissible ((hiq 2019)). We adhered to ethical web scraping practices to ensure that our data collection did not interfere with user experience or the platform’s operations.

The analysis of pornographic content, particularly in its portrayal of different nationalities and ethnicities, brings forth significant ethical concerns. By categorizing and analyzing content based on ethnicity, there is a risk of reinforcing harmful stereotypes, precisely the biases we seek to critique. Our goal is to expose how pornographic discourse perpetuates racial, ethnic, and gender biases, but we approach this with caution to avoid inadvertently legitimizing or amplifying these stereotypes.

The data collected from xHamster is publicly accessible and does not contain sensitive personal information such as individual identities, private sexual preferences, or personally identifiable data. Our collection focused solely on title-level and keyword-level metadata related to content portrayal. We took care to handle the data responsibly and followed best practices for privacy protection and ethical research, ensuring that no personal or confidential information was divulged.

A.2 Data Availability

The notebook used to process the data can be found at an OSF repository: https://osf.io/kbzx6/?view_only=43b9d8aee7174c02a9cfe80e7f498222. Due to the large size of the dataset, we can make it available through direct communications with the authors.