VIDA: The Visual Incel Data Archive. A Theory-oriented Annotated Dataset To Enhance Hate Detection Through Visual Culture

Selenia Anastasi[†‡] ✉, **Florian Schneider**[‡], **Tim Fischer**[‡], **Chris Biemann**[‡]
[†]Università di Genova, Genova, Italy
[‡]Universität Hamburg, Hamburg, Germany
selenia.anastasi@edu.unige.it

Abstract

Images increasingly constitute a significant portion of internet content, encoding even more complex meanings. Recent studies have highlighted the pivotal role of visual communication in the spread of extremist content, particularly that associated with right-wing political ideologies. However, the capability of machine learning systems to recognize such meanings, sometimes implicit, remains limited. To enable future research in this area, we introduce and release VIDA[1], the Visual Incel Data Archive, a multimodal dataset comprising visual material and internet memes collected from two central Incel communities (Italian and Anglophone) known for their extremist misogynistic content. Following the analytical framework of Shifman (2014), we propose a new taxonomy for annotation across three primary levels of analysis: content, form, and stance (hate). This allows for associating images with fine-grained contextual information that helps identify the presence of offensiveness and a broader set of cultural references, enhancing the understanding of more nuanced aspects of visual communication. In this work, we present a statistical analysis of the annotated dataset and discuss annotation examples and future lines of research.[2]

1 Introduction

While digital visual artifacts and memes are a global communicative phenomenon, social scientists have argued they often carry distinct local values rooted in the national or regional cultures and traditions and in specific groups (Denisova, 2019; McSwiney et al., 2021). The intersection of visual content and cultural capital within web-based communities has been extensively documented in contemporary literature in the field of Cultural Analytics, Social Science, and Semiotics (Shifman, 2013, 2014; Nissenbaum and Shifman, 2018), but still needs to be explicitly addressed as a crucial element in the field of computer science. In fact, grasping the geographic and cultural nuances embedded in visuals is pivotal for unrevealing the processes of signification, both within specific communities (e.g., inside jokes) and in broader, general contexts. To this end, this paper introduces a multimodal, comparable corpus of images and texts, focusing on the hateful contents as well as their contextual use within the Italian and English-speaking Incelosphere — a community known for its extremist rhetoric targeting women and other minorities. Moreover, we explore the differences in representations, discussion themes, and cultural references between the two communities that might be important in relation to different targets of hate and new forms of extremism. Given the multimodal nature of digital platforms and the implicitness of the potentially abusive content (Suryawanshi et al., 2020), our dataset challenges annotators to understand content beyond mere visual inspection. We also propose a nuanced annotation oriented to visual content analysis across three analytical dimensions, following state-of-the-art theories on memes: form (evaluating types, formats, and layouts); stance (sub-categories of stereotypes and hate), and content (topics, gender, and ethnicity targeted, and references to popular as well as internet culture). In doing so, we analyze the instrumental use of images, assessing associated stereotypes and hateful connotations through the interplay of text and visuals. This study is rooted in a grounded theory of visual culture, leading to the development of a unique taxonomy. To summarize, our contributions are as follows:

1. We have collected and archived a total of 445.442 images and memes from two prominent anti-feminist extremist communities of the Manosphere in Italian and English.

[1] https://github.com/uhh-lt/vida

[2] **Content Warning: This document contains some examples of hateful content. This is strictly for the purpose of enabling this research.**

2. We introduce an innovative taxonomy for the classification of images in the context of internet cultures, incorporating semantic, contextual, and morphological aspects to enhance our comprehension of the cultural references embedded in hateful content. Then, we manually annotated and tested the taxonomy on a sample of 2181 images randomly extracted.

3. We emphasize the importance of creating multimodal datasets considering production and circulation's geographical, cultural, and social context.

4. We release the annotated part of the dataset for use by the research community.[1]

2 Related Works

2.1 Offensive Multimodal Datasets

As the Internet continues to evolve and social media becomes more and more complex, the need to identify and categorize offensive and hateful content is becoming crucial. In the so-called web 2.0, determining the exact ratio of multimedia content (including images, video, and audio) to text-only content on the Internet is a complex and fluctuating endeavor. What is clear, however, is that the amount of multimedia content on the Web is on the rise. As a result, there has been a significant increase in research efforts to address the challenges of multimodal data collection and analysis, particularly in identifying subtle and implicit offensive content. This section provides a brief overview of existing datasets and resources to detect hate categories in multimedia content.

One of the first large-scale initiatives dedicated to detecting abusive content in images and memes is the relatively recent enterprise by Facebook AI (Kiela et al., 2020). They released a large artificial dataset of 10.000 annotated memes for unimodal and multimodal hate detection in social media. However, due to its artificial nature, the dataset exhibits significant difficulty in generalizing to real cases. To overcome this limitation, (Suryawanshi et al., 2020) created the MultiOff dataset by collecting visuals from social media related to the 2016 US election and manually annotating them based on multiple classes. Although this work is of great interest, the dataset's small size restricts its suitability to highly specialized machine learning systems only.

Due to the implicit nature of hate expressed through images and the multimodality of the task, detecting abusive content in images can be challenging and requires specialized expertise. Thus, the construction of this kind of resource is often oriented to a single domain. Examples of such resources include the Jewtocracy dataset developed by (Chandra et al., 2021), which collects anti-Semitic material from social network sites such as Gab and Twitter, and HarMemes (Pramanick et al., 2021), focusing on memes related to the COVID-19 pandemic. (Fersini et al., 2021) conducted preliminary work on hate subtypes, including sexism and misogyny, while the problem of automatic detection of misogyny in memes was further explored by the SemEval-2022 Task 5: Multimedia Automatic Misogyny Identification (MAMI) by (Fersini et al., 2022).

2.2 Visual Culture and the Incelosphere

Incel is a portmantous for *involuntary celibate*. According to the Cambridge Dictionary, incels can be defined as "members of a group of people on the internet who are unable to find sexual partners despite wanting them and who express hate towards people whom they blame for this"[3]. The Incelosphere forms a sub-cultural group within the wider context of digital culture that broadly promotes racism, anti-feminism, misogyny, and hateful ideas about women, trans, and non-binary people (Ging, 2019). As with any extremist web-based community, incel groups are also characterized by a unique set of expressive forms, a lexicon, rituals, and inside jokes, which are regularly disseminated on the Internet in order to gain more attention and recruit new members. For this reason, the content produced and consumed within the incelosphere travels in a cross-platform mode (Baele et al., 2023) and the mainstream Internet culture can appropriate the same communication tools, which become conventions. Visual culture refers to the extensive place of the visual in social life, emphasising the way visual media are embedded into a wider culture (Rose, 2016). Whether deployed as inward or outward orientated communication, "visual media function as arenas of political and identity construction, activating or deactivating particular social boundaries which then form the basis for future contentious collective action"(McSwiney

[3]https://dictionary.cambridge.org/dictionary/english/incel

et al., 2021). Thus, visual culture in the incelosphere provides useful insights into how members perceive themselves and the world in sharply delineated categories, highlighting the potential use of the aesthetic dimension in constructing identitarian claims and exclusive solidarity. In this polarized context, women and men who do not adhere to the red pill ideology and heterosexual normativity are common targets of hostility.(O'Malley et al., 2022). Moreover, some members condone and encourage violence against women through direct appeals to misogyny and objectification (Krendel et al., 2022; Jaki et al., 2019). During the COVID-19 pandemic, some members of the English-speaking community engaged in anti-vaccination campaigns, while others supported white supremacist, anti-Semitic, and racist discourses (Nagle, 2017). Thus, the dataset we propose in this paper contains a wide range of examples of abusive categories, from basic sexist stereotypes to direct calls for violence.

3 Data Collection

The selection of the Incel's forums was carried out with qualitative methods, including expert-domain close reading, for the purpose of balancing the contents between the two communities. Thus, we selected only those forums of the Incelosphere that showed the greatest similarity in structure, topic of discussions and purposes. After the selection of the forums, we chose to select and collect only specific freely accessible sections of the forums that did not require any formal subscription or login. This was for two main reasons: first, the ethical one - avoiding violating the platform's privacy policies; second, to reduce the risk for the researchers to be subjected to potential violence and other forms of retribution. For the composition and collection of the dataset, we implemented multiple crawlers using the scrapy framework[4], one for each forum, in order to systematically download threads and posts of the sections of our interest. Given the URLs to the forums' sections, e.g., *Introduction*, *Inceldom Discussion*, *Off-Topic*, the crawlers extract structural information that form our dataset as depicted in Figure 1. With this procedure, the created dataset captures the hierarchical structure of the forums of sections, threads, and posts, as well as the conversational flow of the threads and posts of referring, citing, and replying to other users. Detailed statistics about the crawled data can be seen in Table 4.

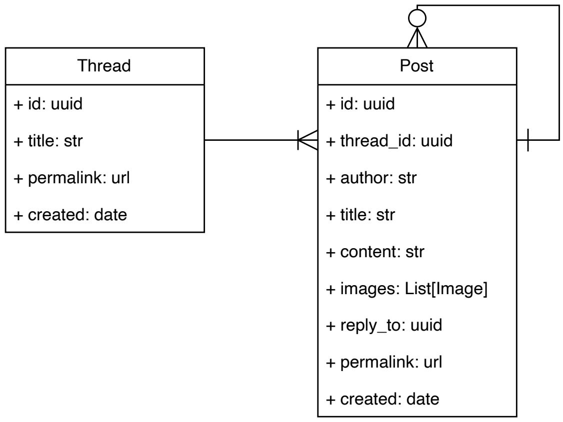

Figure 1: Schema diagram (ERD) of the structural information, i.e., the introduced dataset, crawled from the forums.

For the purposes of this work, only the images and the contextual content of the post in which they are inserted have been extracted and made available, while the full conversations will be released soon.

For the annotation of images, 2000 unique images were randomly selected from both datasets, including the discursive textual context wherein the images appear. Then, we uploaded all the images and context to a self-hosted LabelStudio[5] instance and set up the annotation projects with interfaces for the three levels of analysis.

4 Annotation: Theory and Method

4.1 Conceptual Framework

The theoretical framework is based on Shifman's analysis of memes, extending her categories to apply them to analyzing digital visual artifacts in general (Shifman, 2013). In Bourdieu's terms, the circulation of online visual artifacts also represent important cultural capital for internet communities, actively participating in defining their identity, uniqueness, and boundaries (Nissenbaum and Shifman, 2018). As Shifman notes, another characteristic of visual content online is that they can be repackaged through *mimicry* and *remix* strategies. For instance, the same meme's macro can vary greatly depending on the sociocultural environment in which they propagate.

In order to make sense of the cultural variation of digital visual artifacts, Shifman theorized three main dimensions of analysis: content, form, and stance. Starting from this framework, we could

[4] https://scrapy.org/

[5] https://labelstud.io/

derive three main macro-categories of annotation for our data, each divided into more fine-grained categories.

- **Content** refers to the main topic, idea, and ideology an image can convey (categories listed in Table 5);
- **Form** refers to the physical shape of the image and its morphological dimension, as well as genre-related organization (categories listed in Table 7);
- **Stance** refers to the tone and style of communication, such as the way the senders position themselves in relation to the potential audience of the image. Within this level, one can also consider the emotion of the addressee (categories listed in Table 6).

Although Shifman's definition of *stance* is more complex than this, considering the concept within the pragmatic tradition, for this work, we narrow the meaning of stance as the expression of a strongly hostile position of the sender, encoded in some way within the image or the mix between text and image, and directed towards a single target or a target group.

4.2 Method

Four different annotators with experience in the domain of internet memes, two self-identified women and two self-identified men, were involved in the manual annotation process. All voluntarily participated in two pilot annotation rounds on a random sample of 150 images for both datasets to develop the final version of the taxonomy. For all the rounds of annotation, we chose to use a self-hosted Label-Studio instance, which offers a highly flexible interface. After the two pilot annotation cycles, we were able to develop a complete guide with instructions[6]. After the two pilot rounds, it was decided to collect feedback and discuss the problems encountered by the annotators. All four annotators contributed to the extension of the **Content** categories, which initially contained only 6 categories. During the first and the second pilot rounds, in many cases, the annotators demonstrated different interpretations of the classes present in the **Stance** category, in particular when the task was multimodal. In some of these cases, the images *per se* did not contain hate (i.e., portraits, male/female human bodies, anime characters), and the hateful meaning was implicitly transmitted even within the associated textual content. An example is summarized in Figures 8a and 4.

In an attempt to overcome these interpretative obstacles, the guide was subsequently integrated with clarifications regarding how to handle the labeling in cases where the task is multimodal and more detailed descriptions of some salient characteristics of the ideology associated with the Incel community that can facilitate future annotators in the task of identifying more subtle nuances.

The annotation guidelines are organized as follows:

1. Explanation of what kind of images should be considered memes from the perspective of theory and templates.
2. Description of the classes on the level of the **form**.
3. Description of the classes on the level of the **content**.
4. Description of the classes on the level of the **hate**.

Table 1: Cohen's K and Cramer's V measures for Inter Annotated Agreement on *Form* categories.

	Cohen's K
Artwork Cartoon	0.80
Image Macro	0.65
Infographic Map Graph	0.98
Internet Meme	0.74
Logo	0.97
Other	0.78
Photography	0.77
Poster	0.55
Screenshot	0.74
Sharepost	0.62
Cramer's V	0.78

5 Statistics

Considering the absolute count of the three categories, Tables 5, 6 and 7 show the statistics for the sample dataset. The most common categories for the Content classes are PERSON MEN and PERSON WOMEN, both frequently associated with stereotypes and body shaming categories, as we can read

[6] Annotation Guideline: https://github.com/uhh-lt/vida/blob/main/data/Codebook_Annotation_MEME.docx

Table 2: Cohen's K measure for Inter Annotated Agreement on *Content* categories.

	Cohen's K
Alt-Right	1.0
Animal	0.88
Conspiracy Theory	0.49
Covid 19	0.44
Ethnicity	0.64
Feminism	0.83
Mainstream Meme	0.83
Nazi Fascism	0.71
Numbers	0.56
Other	0.66
Person Man	0.74
Person Woman	0.73
Person Queer	0.65
Politics Left	0.90
Politics Right	0.66
PopCult Anime Manga	0.72
PopCult Cinema	0.55
PopCult Comics Cartoons	0.73
PopCult Influencers	0.53
PopCult Music	1.0
PopCult TVseries	0.60
Pornography	0.73
Red Pill	0.81
Religion	0.82

Table 3: Cohen's K measure for Inter Annotated Agreement on *Stance* categories.

	Cohen's K
Anti Feminism	0.47
Body Shaming	0.58
Misogyny	0.98
Moral Shaming	0.30
None	0.61
Objectification	0.73
Other	0.58
Seduction Conquest	0.76
Stereotype	0.82
Violence	0.88
Cramer's V	0.78

from the co-occurrence plots in Figure 2 and Figure 3.

Contrary to what we expected, memes are not the visual content favored by the communities. This result is interesting because it signals that hatred, prejudice, and stereotypes can be embedded in simple images and should often be captured in a multimodal context through the association between visuals and text. This clearly emerges from the prevalent hate categories that have been labeled considering the context of the entire post: STEREOTYPE (example in Figure 5), BODY SHAMING (examples in Figures 4) OBJECTIFICATION (example in Figure 9a). Also, in reference to style, the co-occurrence matrix between the topic and style categories confirms the prevalent association between photographic genre and STEREOTYPE (129 images in total) and between INTERNET MEMES and STEREOTYPE categories (98 memes in total). Finally, in both communities, examples of misogyny in images are rare but extremely explicit (examples in figure 6 and 7a).

Table 4: Key statistics of the dataset.

	Italian	English
Forums	5	2
Threads	35624	369174
Posts	740278	7359727
Avg. posts / thread	20.78	19.94
Avg. images / post	0.084	0.067
Images	20183	425259
Unique images	94 %	0.72 %
Oldest post	2009/04/29	2017/11/08
Latest post	2023/03/02	2023/03/14
Users	7010	12584

Table 5: Absolute counts of *Topic* annotations.

	Italian	English
Alt-Right	0	16
Animal	63	91
Conspiracy Theory	16	16
Covid 19	41	7
Ethnicity	45	197
Feminism	29	42
Mainstream Meme	51	126
Nazi Fascism	20	36
Numbers	20	27
Other	280	385
Person Man	800	963
Person Woman	652	550
Person Queer	11	20
Politics Left	75	43
Politics Right	83	91
PopCult Anime Manga	27	211
PopCult Cinema	140	94
PopCult Comics Cartoons	50	122
PopCult Influencers	86	46
PopCult Music	38	30
PopCult TVseries	60	57
Pornography	54	60
Red Pill	61	72
Religion	35	35

6 Comparability

From a comparative point of view, the absolute frequency of the classes noted for both datasets can provide some initial clues as to the continuities and differences between the two communities analysed. First of all, we compare the categories related to politics within the macro-category Topic, i.e.: alt-right, nazifascism, feminism, left-wing politics and

Table 6: Absolute counts of *Hate* annotations.

	Italian	English
Anti Feminism	41	54
Body Shaming	111	116
Misogyny	24	45
Moral Shaming	97	105
None	3324	3541
Objectification	108	136
Other	35	53
Seduction Conquest	159	123
Stereotype	357	462
Violence	89	195

Table 7: Absolute counts of *Style* annotations.

	Italian	English
Artwork Cartoon	143	380
Image Macro	8	0
Infographic Map Graph	42	75
Internet Meme	145	299
Logo	13	35
Other	2	10
Photography	965	967
Poster	48	55
Screenshot	410	412
Shared Post	27	41

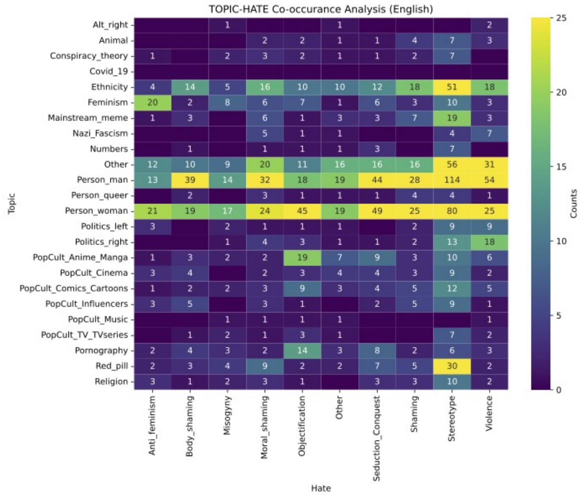

Figure 2: Co-occurrence matrix between Hate and Topic categories (Eng)

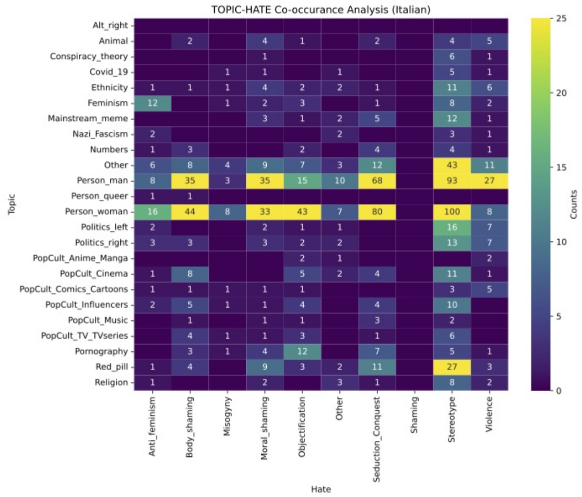

Figure 3: Co-occurrence matrix between Hate and Topic categories (It)

right-wing politics. While the annotators did not find any visual references to the category alt-right in the Italian community, references to the generic right-wing politics and Nazi-Fascism are prevalent in the English-speaking community (see in Appendix A). Conversely, references to the generic left and Communist iconography are prevalent in the Italian community (see in Appendix A).

While this quantitative information alone does not tell us much about how political ideologies are framed in a multimodal context, the co-occurrence matrix between the politics-related classes shows that these categories are often associated with stereotypical and violent content. Sentiment analysis of the textual content associated with the images could provide more accurate insights for interpreting these data. However, based on a close reading of the content, this difference could indicate a less politicised and ideologically motivated orientation within the Italian community and a general tendency to adopt qualunquist political positions that place the extreme right and the extreme left on the same negative level. On the other hand, the absolute numbers of pop culture-related categories underline a strong emphasis on entertainment within the communities, particularly in the areas of cinema, cartoons and TV series. This prevalence suggests an increased engagement with media and popular culture. In particular, the English-speaking community has a significantly higher number of references to Japanese manga and anime, indicating a robust interest in this particular cultural sphere. When examining the co-occurrence matrix, which maps themes to hate categories, there is a clear correlation between the anime and manga label and the use of objectifying language. Furthermore, a detailed examination of visual content from the English-speaking community reveals a recurring depiction of erotic content and feminine representationunder the guise of anime and manga, as shown in Figure 10a. This observation prompts consideration on the highly stereotypical and abstract representation of the female body, which could be considered in future studies on the visual representation of women in Incel communities. Furthermore, it could highlight possible links between otaku geek subcultures and toxic Manosphere subcultures. Future research might also examine the continuity of visual references more closely, shedding light on the intersection of the two online communities.

Although many reports from terrorist studies have examined the representation of in-group and out-group identity through language(Ging, 2019; Krendel et al., 2022), we are still at the beginning of our work to understand the modalities of visual communication within the transnational Incelosphere ecosystem. However, these preliminary numerical results may open up some further questions and possible new lines of research. Given the large amount of material at our disposal, our dataset can certainly be a useful resource for researchers interested in studying the spread of hate through visual artefacts in misogynist extremist online communities.

7 Limitation

The main limitations of our work are related to the possibility of generalizing the annotation protocol to other data sources. In this sense, a first limitation concerns the presence of several specific classes at the level of content, which were developed in an iterative way based on the content of our data set. Although the annotation protocol has been developed on the basis of a solid theory for the analysis of online visual culture in general, the annotation of images on fine-grained categories could make it difficult to apply this protocol to other datasets, thus compromising interoperability. For the categories of the scheme related to hate, the same argument applies. Indeed, the most present category is stereotype, while few images were detected as hateful in the strict sense, as our preliminary analyses show. This is due to the ideological characteristics of the community analyzed. Although the images in our dataset can spread highly offensive messages (such as ethnic stereotypes and pornography), during the annotation phase, our annotators preferred to limit inferences about hateful content to what is expressed, limited to the image and the text associated with it, leaving out any contextual or backgrounding information. Future research could address and overcome these limitations by applying the same taxonomy to data from other sources and considering the integration of classes to annotate hate at the implicit level.

8 Conclusion and Future Work

In this paper, we explored the possibility of integrating information related to the visual culture of misogynistic extremist subcultures on the Internet in order to make the annotation of hateful content more sensitive to the context and linguistic community of reference. For this reason, we selected a state-of-the-art theory for the analysis of Internet visual artifacts such as digital images and memes, and extended our analysis by integrating other identified subcategories to obtain more information about the cultural references contained in the images (i.e., pop culture and various religious and political themes). We mapped the digital sites populated by the community and collected this material by relying on a custom-built crawler to mine the platforms. This allowed us to create a comparable dataset in two main languages, English and Italian. Then, experienced annotators improved and used our annotation scheme to annotate a total of 2181 images and, where present within the post, to annotate the hateful content in light of the text associated. Finally, based on this work, we obtained significant inter-annotator agreement scores, which allowed a first quantitative exploration of the frequency of individual categories and the correlation between them. Our results showed the prevalence of stereotypical content regarding both male and female targets, as well as ethnic and racial stereotypes. We also found the presence of numerous images related to categories of shaming (body and moral shaming), a type of discrimination and abuse that is widespread online and has a strong impact on the psyche of those who are targeted. This initial introduction of VIDA is only our first step in systematically evaluating hate related to Internet visual culture. In future work, we plan to release all crawled data, including threads, posts, and more images, anonymized to prevent the leak of the authors' identities as much as possible. This will be achieved by applying advanced anonymization techniques such as Differential Privacy (Dwork et al., 2014) algorithms and removing source URLs. Further, we will evaluate the proposed taxonomy to test its applicability to other domains, such as hateful memes in the wild, reducing the number of categories if necessary in order to make them as generalizable as possible. Moreover, we will train a classifier on our dataset and apply the human-in-the-loop paradigm to scale our annotations and extend the labeled data in VIDA.

9 License

The dataset is licensed under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) [7]. This allows copying and redistributing the data in any medium or format when appropriate credit is given and a link to the license is given. Further, it is allowed to mix, transform, or extend the dataset for any purpose. However, every change has to be indicated.

References

Stephane Baele, Lewys Brace, and Debbie Ging. 2023. A Diachronic Cross-Platforms Analysis of Violent Extremist Language in the Incel Online Ecosystem. *Terrorism and Political Violence*, pages 1–24.

Mohit Chandra, Dheeraj Pailla, Himanshu Bhatia, Aadilmehdi Sanchawala, Manish Gupta, Manish Shrivastava, and Ponnurangam Kumaraguru. 2021. "Subverting the Jewtocracy": Online antisemitism detection using multimodal deep learning. In *Proceedings of the 13th ACM Web Science Conference 2021*, pages 148–157.

Anastasia Denisova. 2019. *Internet memes and society: Social, cultural, and political contexts*. Routledge.

Cynthia Dwork, Aaron Roth, et al. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

Elisabetta Fersini, Francesca Gasparini, Giulia Rizzi, Aurora Saibene, Berta Chulvi, Paolo Rosso, Alyssa Lees, and Jeffrey Sorensen. 2022. SemEval-2022 Task 5: Multimedia automatic misogyny identification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 533–549.

Elisabetta Fersini, Giulia Rizzi, Aurora Saibene, and Francesca Gasparini. 2021. Misogynous meme recognition: A preliminary study. In *International Conference of the Italian Association for Artificial Intelligence*, pages 279–293. Springer.

Debbie Ging. 2019. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and masculinities*, 22(4):638–657.

Sylvia Jaki, Tom De Smedt, Maja Gwóźdź, Rudresh Panchal, Alexander Rossa, and Guy De Pauw. 2019. Online hatred of womenin the Incels.me forum. *Journal of Language Aggression and Conflict*, 7(2):240–268.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020. The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes. *Advances in neural information processing systems*, 33:2611–2624.

Alexandra Krendel, Mark McGlashan, and Veronika Koller. 2022. The representation of gendered social actors across five manosphere communities on Reddit. *Corpora*, 17(2):291–321.

Jordan McSwiney, Michael Vaughan, Annett Heft, and Matthias Hoffmann. 2021. Sharing the hate? Memes and transnationality in the far right's digital visual culture. *Information, Communication & Society*, 24(16):2502–2521.

Angela Nagle. 2017. *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.

Asaf Nissenbaum and Limor Shifman. 2018. Meme templates as expressive repertoires in a globalizing world: A cross-linguistic study. *Journal of Computer-Mediated Communication*, 23(5):294–310.

Roberta Liggett O'Malley, Karen Holt, and Thomas J Holt. 2022. An exploration of the involuntary celibate (incel) subculture online. *Journal of interpersonal violence*, 37(7-8):NP4981–NP5008.

Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md Shad Akhtar, Preslav Nakov, and Tanmoy Chakraborty. 2021. Detecting Harmful Memes and Their Targets. *arXiv preprint arXiv:2110.00413*.

Gillian Rose. 2016. *Visual methodologies: An introduction to researching with visual materials.*

Limor Shifman. 2013. *Memes in digital culture*. MIT press.

Limor Shifman. 2014. The cultural logic of photo-based meme genres. *Journal of visual culture*, 13(3):340–358.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2020. Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text. In *Proceedings of the second workshop on trolling, aggression and cyberbullying*, pages 32–41.

A Visual Examples

Visual examples of the multimodal labeling of the most numerous categories.

[7] https://creativecommons.org/licenses/by-sa/4.0/

Figure 4: Example of Person Women and Body Shaming associated categories

Figure 5: Example of Stereotype in memes

Figure 6: Example of Person Women and Violence associated categories

Figure 7: Example of Meme and Misogyny associated categories

(a) **Meme text:** Sluts, sluts everywhere

Figure 8: Example of Person Women, Ethnicity and Body Shaming associated categories, annotated considering the context

(a) **Post:** The fat ugly black one gets more attention according to juggernaut law.

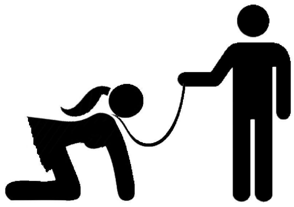

Figure 9: Example of Objectification, annotated considering the context

(a) **Post:** Put foids on a leash - and take away all their rights, treat them like soulless inanimate objects who are just basic fuck dolls and breeding machines.

Figure 10: Example of Pop_Cult_Anime_Manga category associated with Objectification, annotated considering the context.

(a) **Post:** I like 2D legs.

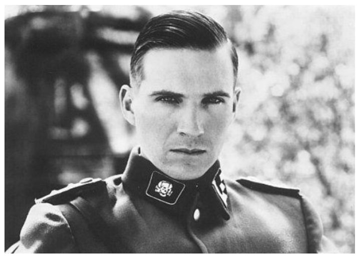

Figure 11: Example of Nazi Fascism category associated with Objectification, annotated considering the context.

(a) **Post:** They subconsciously want the fuhrer to return.

Figure 12: Example of Politics Left category in the Italian dataset.

(a) **Meme text:** I repeat: women in gulags.