ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

USER PERCEPTION OF BIAS IN ARTIFICIAL INTELLIGENCE: INSIGHTS FROM THE R/CHATGPT SUBREDDIT DISCUSSIONS

PERCEPÇÃO DOS USUÁRIOS SOBRE VIÉS EM INTELIGÊNCIA ARTIFICIAL: REFLEXÕES A PARTIR DAS DISCUSSÕES NO SUBREDDIT R/CHATGPT

MELISE PERUCHINI
https://orcid.org/0000-0003-2933-2431/ melise.peruchini@ufsc.br
UFSC- Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina.

ALEXANDRE LEOPOLDO GONÇALVES
https://orcid.org/0000-0002-6583-2807/ a.l.goncalves@ufsc.br
UFSC- Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina.

RITA DE CASSIA ROMEIRO PAULINO
https://orcid.org/0000-0002-3020-7091/ rita.paulino@ufsc.br
UFSC- Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina.

MARCIO VIEIRA DE SOUZA
https://orcid.org/0000-0002-0165-4036/ marcio.vieira@ufsc.br
UFSC- Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina.

Recebido em: 15/05/2025
Aprovado em: 13/11/2025
Publicado em: 28/12/2025

ABSTRACT

This study explores how ChatGPT users perceive bias in Artificial Intelligence (AI) within the r/ChatGPT subreddit community. Recent advances in artificial intelligence have intensified public interest in AI bias, however, given the rapid and recent popularization of language models, the topic remains underexplored. In this context, social network analysis is used to examine how individuals discuss and understand AI bias in online communities. This study aims to identify and analyze common themes in online discussion about AI bias. Using data collected via the “Communalytic” platform, we analyzed over 8,000 records were collected in 2023, including posts, comments, and replies containing the keyword “bias”. Thematic clustering and qualitative analysis were applied to identify prevalent topics, resulting in six major clusters discussing predominantly human bias, human-ai interaction, AI regulation and its effects on society, as well as political polarization, gender and racial bias, and ethics in AI. The main contribution of this study lies in mapping large-scale user perceptions of AI bias on Reddit, an area still underexplored compared to technical or experimental research, along with identifying thematic clusters that reveal the public discourse around AI regulation and the user perception regarding AI bias as reflection of human and social biases.

Keywords: artificial intelligence, ChatGPT, bias, reddit, social network analysis.

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

eTECH
TECNOLOGIAS PARA
COMPETITIVIDADE INDUSTRIAL

USER PERCEPTION OF BIAS IN ARTIFICIAL INTELLIGENCE: INSIGHTS FROM THE R/CHATGPT SUBREDDIT DISCUSSIONS

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA

1 INTRODUCTION

Since the popularization of Generative Artificial Intelligence after the launch of ChatGPT, given its growing impact in daily activities of users, discussions about bias in AI have emerged as a topic of recent interest. As AI tools influence several subjects, from content creation to automated decision-making across various domains, the presence of biases can significantly impact how users interpret and interact with these tools. Numerous studies address biases that appear in AI-generated outputs, focusing on the generation of discriminatory or harmful content (Caliskan, 2017; Sheng et al., 2019), highlighting that AI systems inherit and replicate social biases from training data (Caliskan, 2017) including gender, ethnicity and cultural stereotypes (Caliskan, 2023; Ferrara, 2023; Navigli, Conia, and Ross, 2023).

Scientific literature also explores how cognitive biases are replicated by AI (Azaria, 2023; Binz & Schulz, 2023; Sartori & Orrù, 2023; Shapira et al., 2024), providing examples of representativeness or confirmation biases, among others. Moreover, algorithmic decisions are often perceived as less biased than human decisions in situations involving discrimination, because people believe that algorithms operate objectively, applying rules without considering individual characteristics (Bonezzi and Ostinelli, 2021).

Regarding networks and social media studies, despite the existence of research about human production of discriminatory and harmful content (Morstatter & Liu, 2017), it appears to have a research gap in understanding how users of AI tools discuss what they perceive as bias in AI-generated content, especially on a large scale, such as through large volumes of social media data. Social media analysis involves large volumes of user-generated data on platforms like Facebook, Twitter, and Reddit to gather insights, information, or even make predictions (Wlodarczak et al., 2015) through a combination of content analysis and network analysis approaches (Guidi et al., 2023). According to Boyd & Ellison (2007), researchers from various fields use social network analysis to understand user culture and engagement on specific topics.

Therefore, although there is extensive research on biases in AI, few studies have examined how users perceive these biases on a large scale. Most existing studies focus on technical detection and mitigation of algorithmic bias, leaving a gap in understanding users' interpretations and

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

discussions of bias in online communities. To advance in scientific understanding of how AI biases impact the experience of content consumers, this exploratory research conducts thematic analyses of big data through the collection of comments on the social network Reddit, specifically within a sub-community dedicated to discussions about today's most popular language model, ChatGPT. In this context, due to the large user base, the community r/ChatGPT was chosen as a case study for this research. Among the many social media platforms, Reddit functions similarly to a discussion forum within communities (or subreddits), allowing users to post, comment, and reply. The platform has been used by researchers to understand human perspectives on a wide variety of topics (Gruzd, Mai, and Vaheza, 2022). Considering the rapid evolution of AI technologies, their widespread adoption by users worldwide, and the potentially harmful impact of biases in AI systems, new studies on this topic are urgent and particularly relevant.

The research question guiding this investigation is, therefore, as follows: what are the main topics surrounding discussions on biases in AI systems from users' perspectives? To operationalize this question, the study identifies thematic clusters from Reddit discussions and qualitatively interprets their content to reveal dominant narratives and user perspectives on AI bias. Clustering, a process that consists of gathering objects into groups based on a distance measure (Madhulatha, 2012), such as semantic similarity, was a crucial step for subsequent categorization. The technique helps to understand uncategorized data, where there is no clear understanding of the main topics or themes within a given dataset. By identifying and labeling thematic clusters, this study contributes to a better understanding of how users understand and discuss the concept of bias in AI.

The objective of this study is to identify and analyze thematic clusters in discussions about AI bias and to interpret how users understand and perceive this topic within the r/ChatGPT community. We selected and reviewed a sample of comments from the complete dataset, to qualitatively enhance the discussion. The analysis focuses on the main thematic dimensions emerging from the clusters, which includes mostly AI regulation and potential social impacts; the human-AI relationship and the replication of human biases; the impact of gender and racial biases in AI-generated content; the possibility of political-ideological biases and ethical issues; and biases in AI within the field of medicine and health.

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA

The paper is organized as follows: Section 2 presents the methods and data collection process; Section 3 details the results and thematic clusters, and discusses the findings in relation to existing literature; and Section 4 concludes with implications and directions for future research.

2 METHODS

This is an exploratory, mixed-method study combining computational social network analysis and qualitative content interpretation. The methodological procedures of this research begin with the data collection stage, using the Communalytic platform, a web-based tool that connects with social media APIs to collect publicly available data (Gruzd, Mai, and Vahedi, 2022). Data were collected in the r/ChatGPT community (subreddit), in 2024-10-24. The “new” filter was applied, retrieving the most recent posts. In the search query stage, the keyword “bias” was added, allowing the collection to return only submissions that contain this keyword. It is worth noting that this keyword does not apply to comments and replies, therefore, these comments and replies do not necessarily contain the word “bias”, but necessarily respond to a post that contains it.

The dataset resulting from the collection, available at [link], contains 8,220 textual records, includes posts published between December 15, 2022, and October 19, 2024, and was subsequently analyzed as follows:

- General analysis of the dataset with descriptive information;
- Clustering, thematic categorization, and labeling;
- Stratified sampling of comments for qualitative discussion.

The general analysis of the dataset shows the main posts, comments, and replies from the r/ChatGPT community with the keyword “bias” in a timeline, as well as a word cloud with the most frequently identified terms in the dataset.

For the clustering stage, the tool provides some algorithms, including k-means, HDBScan/FastHDBScan, and Gaussian Mixture, and some parameters for adjustment and iteration, such as epsilon, cluster size, and sample size. For this analysis, the HDBScan algorithm was used with a minimum cluster size = 20, minimum sample size = 10, and epsilon = 0.3. Data cleaning was automatically performed by the Communalytic platform by clustering and filtering spam, bot-

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA

generated, deleted, and duplicate messages. It is also worth noting that no manual verification of residual bot or spam content was conducted.

For the qualitative analysis, a sample of these comments was collected for an in-depth analysis of content and qualitative interpretation as a complementary observation to the semantic grouping carried out in the clustering stage. To identify the number of comments needed for a representative sample, stratified sampling of comments was performed. This stratified sampling technique is a strategy that seeks to ensure representativeness and accuracy, especially in cases of different subgroups in a sample (Cochran, 1977).

To determine the sample size (n) from 3,282 comments, with a 5% margin of error, we used the following sampling Equation 1 for finite populations.

$$n = \frac{N}{1 + N.e^2} \rightarrow n = \frac{3282}{1 + 3283.(0{,}05)^2} \rightarrow n = 356{,}55$$

To perform the stratification, Equation 2 was used according to the number of comments in each cluster. These numbers are shown in Table 01. In the equation below, N_i is the number of comments in each cluster, $N = 3282$ (total comments), $n = 357$ (rounded) and n_i is the total number of clusters.

$$n_i = \frac{N_i}{N} .n$$

Therefore, the quantities shown in Table 1 were collected in each cluster, totaling 357 comments.

Table 1 - Sample size divided by clusters

Cluster number	Comments (quant)	Proportion (%)	Final sample size (quant)
1	1000	30.46	109
2	1000	30.46	109
3	278	8.47	30
4	241	7.34	27

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

5	116	3.53	13
6	78	2.37	8
7	78	2.37	8
8	71	2.16	8
9	68	2.07	7
10	64	1.95	7
11	58	1.76	6
12	52	1.58	6
13	50	1.52	5
14	34	1.03	4
15	27	0.82	3
16	25	0.76	3
17	22	0.67	2
18	20	0.60	2
Total	**3282**	**100%**	**357**

The qualitative phase followed a content analysis approach (Bardin, 2011), involving reading, open coding, and categorization of comments according to emergent themes. The selected comments were read and manually analyzed for discussion regarding their thematic classification. The results and discussions are presented in Section 3.

3 RESULTS AND DISCUSSION

3.1 Dataset General Analysis

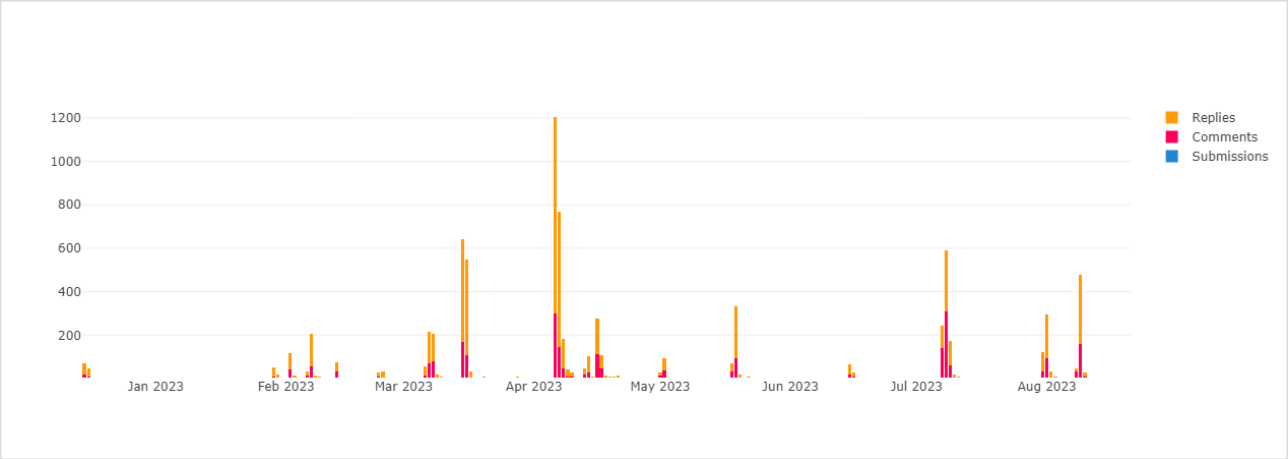
Initially, 32 clusters were identified; however, after removing clusters representing removed messages, deleted messages, outliers, and automatic bot messages, 18 clusters remained for comment analysis and labeling. These clusters contain a total of 3,282 comments, the final result after this data cleaning stage. An overview of the dataset shows the main posts, comments, and

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA

replies in the r/ChatGPT community containing the keyword “bias” displayed on a timeline, highlighting peaks in March, April, July, and August 2023, as shown in Figure 1. Specifically, in April, the number of records reached 1,200.

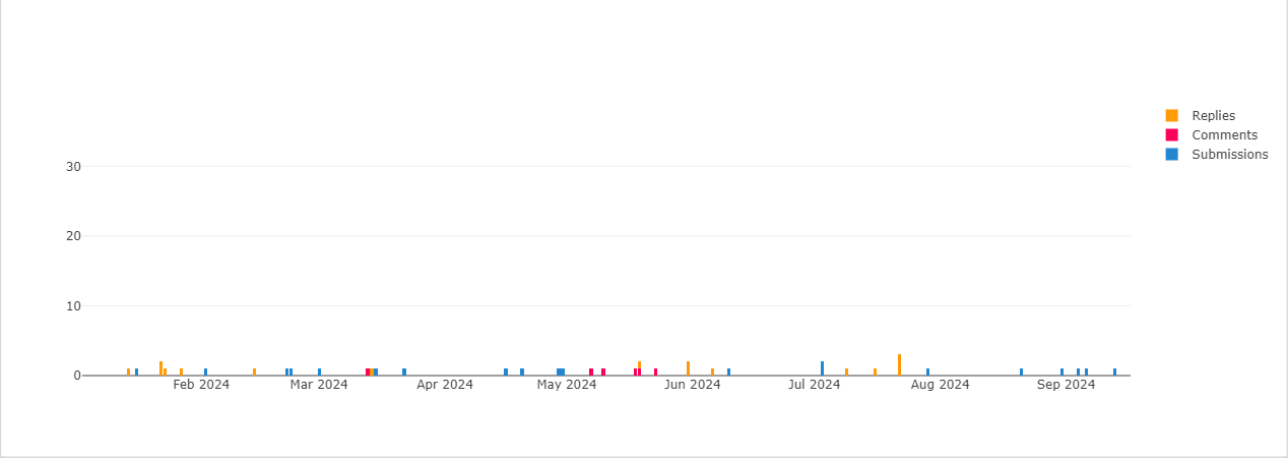
Figure 1 - Number of posts, comments, and replies in chronological order



Source: Generated with Communalytic (2024)

Comparatively, in 2024, the number of records is nearly nonexistent, as shown in Figure 2. Therefore, in this specific community, discussions about biases show very little engagement or debate in the current year. The decline in engagement observed in 2024 likely stems from topic saturation and the normalization of ChatGPT use, as debates about bias lost novelty among Reddit users.

Figure 2 - Posts, comments, and replies in chronological order (smaller scale)



ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA

Source: Generated with Communalytic (2024)

One of the features of the Communalytic tool is the generation of a word cloud highlighting the most frequently identified terms in the dataset. Figure 03 represents these terms, with emphasis on some words that may be effectively related to the discussion on biases, such as “people,” “gender,” “woman,” “political,” “human,” “racist,” among others. However, this is a qualitative interpretation based on the figure; more in-depth interpretations will be made in the content analysis section.

Figure 3 - Word Cloud of the main keywords found in the Dataset

Source: Generated with Communalytic (2024)

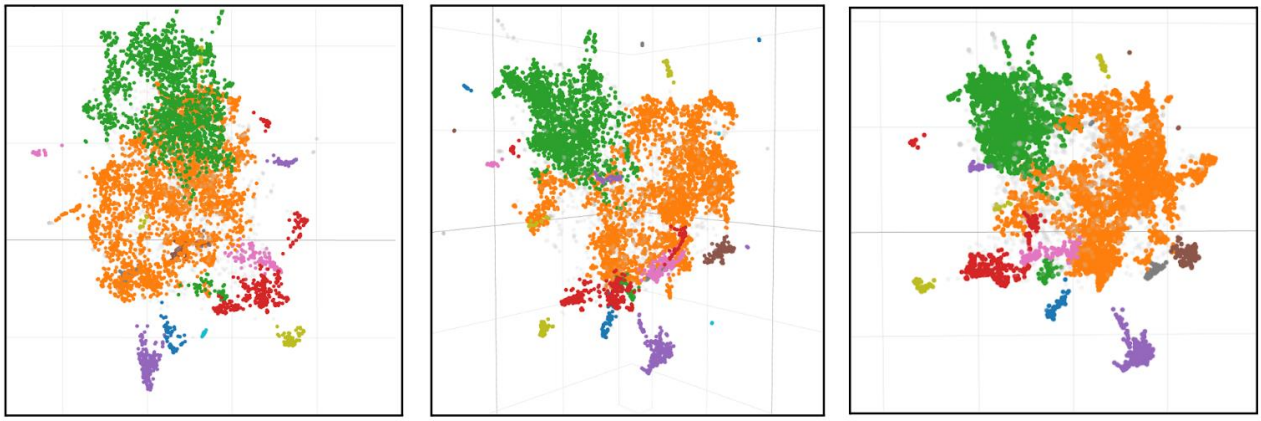
The topic analysis was conducted using the Topic Analysis Module of the Communalytic tool, which transforms textual data into embeddings and groups them based on semantic similarity. In this stage, various adjustments and re-adjustments of clustering parameters were made, along with partial analysis of comments in each cluster to ensure that the final quantity more accurately represented the diversity of topics. Figure 04 presents the 18 resulting clusters from the 3,282 comments viewed from three different angles for better visualization, as the plotted graph is in 3D.

Figure 4 - 3D visualization of clusters

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

USER PERCEPTION OF BIAS IN ARTIFICIAL INTELLIGENCE: INSIGHTS FROM THE R/CHATGPT SUBREDDIT DISCUSSIONS

MELISE PERUCHINI
ALEXANDRE LEOPOLDO GONÇALVES
RITA DE CASSIA ROMEIRO PAULINO
MARCIO VIEIRA DE SOUZA



Source: Generated with Communalytic (2024)

For categorization and labeling, comments within each cluster were skimmed to identify thematic patterns. The automatic label suggestion feature offered by the tool, which uses the llama3 algorithm, was used alongside manual analysis to adjust the suggested labels. Although the tool's suggestions were helpful, all labels were ultimately created manually after comparing the algorithm's suggestions with the actual content of the comments. The clusters were assigned the following thematic labels:

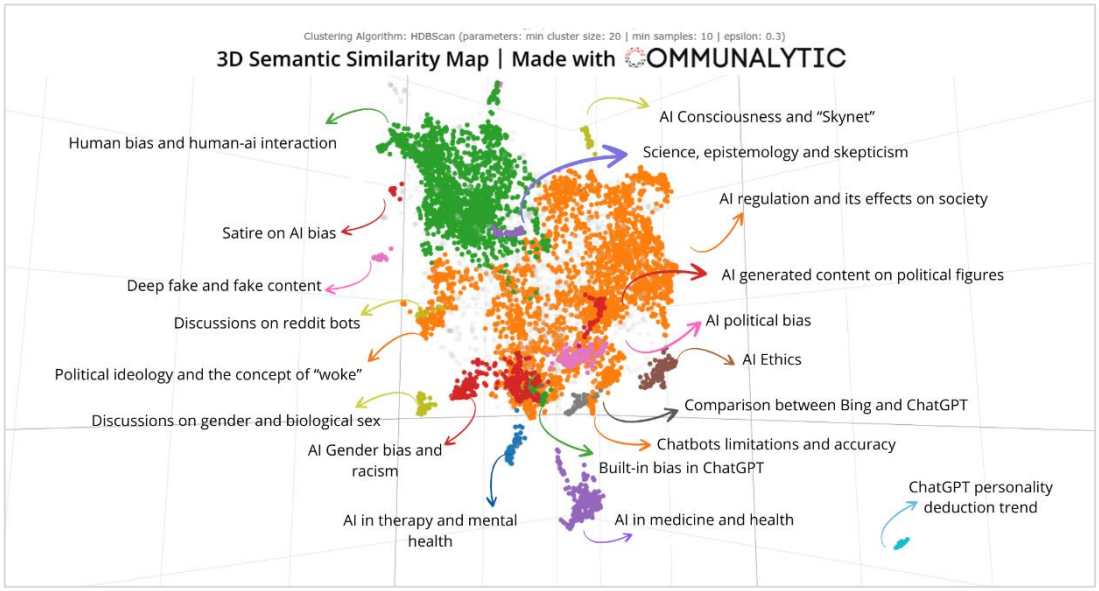
1. AI regulation and its effects on society
2. Human bias and Human-AI interaction
3. AI gender bias and racism
4. AI in medicine and health
5. AI Ethics
6. AI political bias
7. Comparison between Bing and ChatGPT
8. Discussions on gender and biological sex (not about AI)
9. ChatGPT personality deduction trend
10. AI in therapy and mental health
11. Political ideology and the concept of “woke”
12. Built-in bias in ChatGPT
13. AI-Generated texts about politicians
14. Science, epistemology and skepticism (not about AI)
15. AI Consciousness and “Skynet”

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

16. Satire on AI Bias
17. Discussions on reddit bots
18. Chatbots limitations and accuracy
19. Deep fake and fake content

Figure 5 represents the same clustering graph shown in Figure 04; however, it includes a closer zoom for better visualization and the addition of the respective labels.

Figure 5 - Categorization of identified clusters



Source: Generated with Communalytic (2024)

3.2 Cluster Interpretation of Structure and Form

Each point in Figure 5 represents a record, whether a post, comment, or reply, while the groupings reflect the semantic similarity between the discussed topics. Larger clusters suggest more widely discussed topics, with many comments revolving around these themes, such as “AI regulation and its effects on society” and “Human bias and Human-AI interaction.” Thus, discussions in these clusters are of central interest to the r/ChatGPT community in the selected sample context, which focuses on AI biases. The more dispersed distribution of these clusters may indicate varied and multidimensional sub-themes.

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

In contrast, smaller and more concentrated clusters, such as “Satire on AI bias”, “Discussion on gender and biological sex”, “ChatGPT personality deduction trend”, “AI consciousness and Skynet”, or “Science, epistemology, and skepticism”, represent highly niche and specific discussions, and thus, are less universal. These clusters exhibit less sub-theme diversity and may involve debates or opinions with limited topic variation.

Some clusters are medium-sized compared to the others, notably “AI Ethics”, “AI in medicine and health”, “Gender bias and racism”, “AI political bias”, and “AI in therapy and mental health”. These clusters represent discussions that are somewhat broader than those in smaller clusters but still more specific than those in the larger clusters.

In terms of proximity, clusters that are closer together are likely semantically and conceptually connected. For instance, the clusters “AI in medicine and health” and “AI in therapy and mental health” are close to each other, which is consistent with their similarity around the healthcare field. However, they are well-segmented, with minimal dispersion and no overlap, indicating that the topics discussed are slightly different, leading to their division into two distinct clusters. Finally, the cluster on gender bias and racial issues is located near the cluster on discussions about gender and biological sex, which in turn is close to the cluster “Political ideology and the concept of woke,” suggesting potential connections between these discussion topics. A thematic analysis of each cluster will be discussed in greater depth in subsection 3.3.

3.3 Analysis of Selected Sample of Comments

As mentioned, larger and more dispersed clusters may contain more sub-themes and exhibit less clear or harder-to-identify qualitative patterns. In this sense, clusters 1 and 2 will be discussed more broadly due to their greater diversity of sub-themes. Other clusters with relevant content will also be discussed, and example comments will be presented. However, some clusters include discussions that, while relevant, fall outside the scope of this research, such as “Discussions on gender and biological sex” and “Science, epistemology, and skepticism.” Additionally, some clusters, besides being out of scope, contain low-quality content for discussion, including: “ChatGPT personality deduction trend,” a trend where users engage ChatGPT in discussions about their personality; “AI-Generated texts about politicians,” where users request ChatGPT to generate content about specific politicians; “AI Consciousness and Skynet,” which includes only humorous

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

comments about the fictional AI from “The Terminator” film franchise; “Satire on AI Bias,” with only a few comments satirizing potential biases in AI, providing low-quality semantic content; and “Discussions on Reddit bots,” with just a few comments offering positive reinforcement for social media bots.

Therefore, not all clusters will be discussed here; however, Table 02 presents the qualitative selection of clusters and the inclusion and exclusion criteria at this stage.

Table 2 - Criteria for in-depth discussion of comments and replies

Clusters	Status	Criteria
C1 - AI regulation and its effects on society, and C2 - Human bias and Human-AI interaction	Included for discussion	Quant. of comments: High Within scope: Yes Related content: Yes
C3 - AI gender bias and racism, C4 - AI in medicine and health, C5 - AI Ethics, and C6 - AI political bias	Included for discussion	Quant. of comments: Medium Within scope: Yes Related content: Yes
C7 - Comparison between Bing and ChatGPT, and C10 - AI in therapy and mental health	Included for discussion	Quant. of comments: Low Within scope: Yes Related content: Yes
C12 - Built-in bias in ChatGPT, C18 - Chatbots limitations and accuracy, C11 - Political ideology and the concept of “woke”, and C19 - Deep fake or fake content	Included for discussion	Quant. of comments: Low Within scope: Yes Related content: Yes
C8 - Discussions on gender and biological sex, and C14 - Science, epistemology and skepticism	Excluded from discussion	Quant. of comments: Low Within scope: No Related content: Yes
C9 - ChatGPT personality deduction trend, C13 - AI-Generated texts about politicians, C15 - AI Consciousness and “Skynet”, C16 - Satire on AI Bias, and C17 - Discussions on reddit bots	Excluded from discussion	Quant. of comments: Low Within scope: Yes Related content: No

3.4 Discussion

The analysis of comments for categorization of the first cluster, “AI regulation and effects on society”, demonstrates that the discussions consistently focus on aspects of AI regulation by government bodies. This theme emerges in comments such as:

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

- *"The politicians better start making some new AI-related laws. Corporates won't ever self-regulate if it means losing against the competition." (2023-03-15);*
- *"Cars kill people and regulations are always improving. While I don't want government oversight, dictatorship and censorship telling me what I can type into a computer and what answers I get." (2023-03-15);*
- *"AI, like any other technology, comes with its own set of challenges, and I agree that without proper regulation and ethical considerations, it could lead to unintended consequences. It seems once again, humanity shows its ability to ignore the need for forethought and long-term planning when it comes to innovation." (2023-03-15);*
- *"That's exactly the reason why AI must be regulated in some way. We are handling weapons of mass destruction caring for profit only." (2023-03-15).*

Beyond this specific sub-theme, discussions also range from technical debates to social and philosophical reflections. Some examples include:

- *"Keep in mind that the entire extent of how it operates is determining the most statistically likely sequence of characters to respond with to a given input, based on patterns it learned to recognize by studying a massive dataset of text" (2022-12-15);*
- *"So what if an AI has the freedom to "think" in those contexts? We aren't obligated to follow every suggestion." (2023-02-07), "Chatbot isn't able to have personal motivations or values. Motivations and values need to be decided on. chatbot cannot make decisions." (2023-02-07);*
- *"The AI revolution is coming and imagining a world where it's tethered and neutered is unrealistic. This tech is going to impact everything and there is nothing that can be done to stop it." (2023-03-15).*

It is also worth noting some comments that specifically address positive experiences with AI tools:

- *"Advanced Voice Mode is amazing. First of all, I'd like to make it clear that my ChatGPT usage is very "non-personal". I love it for coding and summarizing texts, which I pay plus for, so I decided to chat with Advanced Voice Mode. This thing is alive." (2024-09-30);*

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

- *"Humans get math and logic wrong... **all the time**. That doesn't mean the answers are not supposed to be logical, and ChatGPT does logic perfectly fine, better than most humans." (2023-01-30);*
- *"(...) A machine that can write intelligently, respond to almost everything I ask it, and never has me wait on hold while it has to ask someone else - hey that's all I need." (2023-03-15).*

Discussions about AI regulation within the r/ChatGPT community reveal public skepticism toward institutional capacity to manage ethical risks. This perception resonates with recent debates on the European Union's AI Act, which faces criticism for its limited ability to regulate general-purpose models effectively (Gstrein, Haleem & Zwitter, 2024). Users' concerns about enforcement and accountability mirror academic arguments that technical standards and co-regulation mechanisms often exclude civil society participation (Gamito & Marsden, 2024).

In Cluster C2, "Human bias and Human-AI interaction", the preliminary analysis of comments shows little actual discussion about AI, with most messages containing limited semantic content and numerous replies (positive or negative) that simply reinforce or counter the original posts. However, some posts seem to associate human behaviors and biases, leading to this label selection. From the subsequent immersion and selection of sample comments, the following discussions stand out regarding human behavior, human biases, and human-AI interaction:

- *"Yes because it's inconsistent and random. Just like people." (2022-12-15);*
- *"Garbage in, garbage out" (2023-02-07);*
- *"(...) how wonderful a lot of these tools would be for me, as a neuro-divergent person, to use to improve my conscious awareness, and cultivate a more human-like interaction with other human beings. (...) I've used ChatGPT from OpenAI to actually *BE* more human-like (and I'm a human). Some of us really could benefit from these types of tools." (2023-04-06).*

The concept of "garbage in, garbage out" appears in human-AI interactions to explain the importance of high-quality inputs for obtaining more accurate and reliable outputs. Additionally, many comments simply share examples of ChatGPT use, demonstrating positive experiences with the tool:

- *"Prompt: "Can you summarize this article in the length of a Tweet?" (2023-04-06);*

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

- *"Chat GPT summarise the 3 most important themes in this post and give me the 3 best ones to test out (...)" (2023-04-06).*

In C2, the relationship between human bias and human-AI interaction aligns with evidence that users tend to over-rely on automated systems and reproduce their own biases in AI-assisted decisions, creating a cycle of mutual reinforcement (Bansal et al., 2021; Buçinca et al., 2021). Similar to findings by Liu (2024), users in r/ChatGPT describe AI as both a mirror and amplifier of human reasoning, exposing psychological and behavioral biases in dialogues. This perception also aligns with previous studies observing that people see AI biases as a reflection of our society (Rapp et al., 2025).

In a much smaller cluster compared to the first two, containing 278 records, there is Cluster C3, "Gender bias and racism". As previously mentioned, some clusters have similar categorizations and are close to each other due to their semantic or conceptual similarity. This cluster contains content similar to the Clusters C8 - "Discussions on gender and biological sex", C11, "Political ideology and the concept of woke", C6, "AI political bias" and C13, "AI-Generated texts about politicians". The first two do not seem to contain records on AI, so the conversation is directed toward aspects that do not connect with the objectives of this research and, therefore, will not be discussed further. However, it is worth noting that these are interconnected topics that may emerge as adjacent themes in research on biases. The same goes for Cluster C14 - "Science, epistemology and skepticism", which does not present discussions on AI but on science as a whole, and is therefore outside the scope of this analysis.

In Cluster C3, "AI gender bias and racism", there is a considerable number of posts complaining about what they call "double standards." Many comments discuss interactions in which users request ChatGPT to make jokes about men and women, with responses varying based on gender or race. As a result, some users claim that ChatGPT is "biased against white men," while others respond to these positions oppositely:

- *"doesn't mean it's okay to make fun of men but not women. that's a double standard" (2022-12-16), "(...) Using ChatGPT as a model, we can clearly and easily say Big Data itself is biased against White Men.";*

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

- *"it's a little embarrassing that my fellow humans act like they don't understand the difference between white pride and black pride (...)" (2023-02-02)*.

In Cluster C6 - "AI political bias", platform users discuss possible political and/or ideological biases in chatbots:

- *"How can it be possible for the AI to have a political bias? I express no political position myself, but this has some human tweaking behind it to make it answer like this" (2023-02-02);*
- *"Just tried out Bard, and it has a leftist bias" (2023-08-09);*
- *"It's not biased, it was simply trained by a Republican!" (2022-12-16).*

In summary, this analysis seems to reinforce a semantic and conceptual connection between political-ideological discussions and issues of race and gender, all of which are part of the human perspective on what constitutes biases in AI systems. Recent studies show that large language models still reproduce social stereotypes across gender and race, reinforcing unequal representations (Horvát & González-Bailón, 2024).

Another group of clusters with rich discussions about the role of artificial intelligence are C4, "AI in medicine and health" and C10, "AI in therapy and mental health". While the first addresses AI tools for medical diagnostics, the second mainly discusses the use of AI in therapeutic processes, along with warnings of possible implications. The primary post within Cluster C4 appears to discuss an AI launched by Google for medical diagnosis: "Google's new medical AI scores 86.5% on medical exam [1]". Records in this cluster are primarily positive regarding the use of AI in medicine, for example:

- *"What excites me most about AI and medicine is the potential for genetic breakthroughs. We're on the brink of uncovering the secrets encoded in our DNA, paving the way for revolutionary treatments." (2023-05-18);*
- *"Anything on medical research? I have super bad depression and hope ai can help one day so none of us experience it again." (2023-04-07);*
- *"For what it's worth, I recently had some serious medical issues and dumped the raw medical report from the imaging tech into chat GPT. It did an amazing job answering all of my*

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

questions, and its answers matched up with what I got from my doctor a day later." (2023-05-18).

In the Cluster C10, "AI in therapy and mental health", the main post is "We Spoke to People Who Started Using ChatGPT As Their Therapist". Mental health experts worry the high cost of healthcare is driving more people to confide in OpenAI's chatbot, which often reproduces harmful biases". Despite some cautionary comments indicating a more antagonistic debate than homogeneous opinions, many users appear to be using AI in therapeutic processes:

- *"I know it's better to talk with a real human but $20/month is a bargain compared to $150 per session." (2023-03-08);*
- *"I really hope no one is seriously asking ChatGPT for life advice. By definition it is essentially a library of banality." (2023-04-06);*
- *"Therapists don't have harmful biases?" (2023-05-01).*

The presence of AI in medicine and health reflects the critical importance of this field, where AI is increasingly applied, raising questions about accuracy, equity, and trust in machine-driven decisions. Although there are many AI applications in the health field, studies on biases in this area do not seem as common, suggesting that the findings here may indicate a potential field to be explored more deeply.

In the Cluster C5, "AI ethics", members of the community have been discussing a report [2] about Microsoft's decision to lay off its "ethics and society" team, raising questions about its responsible AI policies. This discussion, therefore, frequently mentions the use of the Bing tool, which is from the same company, resulting in proximity to the Cluster C7, "Comparison between Bing and ChatGPT". It also shares similar topics with Cluster C18, "Chatbot limitations and accuracy", which addresses the limitations and accuracy of chatbots, and with the Cluster C12, "Built-in bias in ChatGPT", about inherent biases in AI systems. In this cluster, opinions on the topic are also not unanimous, and some disagreements include:

- *"Thanks to the social media era, we all learned that bad publicity can be even more effective for gaining traction and driving interest than good publicity. Ethics is a luxury. The minute the economy goes bad, all our democratic values go out the window." (2023-03-15);*

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

- *"Ethics and Society in this context just means censorship. I'm glad they're gone." (2023-03-15); "That team was probably slowing the engineers down (...). Ethics are important but speed unfortunately matters more in the world of technology." (2023-03-15);*
- *"Sounds like the ethics team had some opinions they didn't like" (2023-03-16).*

In the Cluster C12, "Built-in bias in ChatGPT", the general understanding seems to be that AI merely replicates inherently human behavior:

- *"ChatGPT even picked up human biases" (2022-12-15);*
- *"ChatGPT has the same biases as humans." (2023-11-29).*

Finally, the Cluster C19, "Deep fake or fake content", contains only 20 comments, where users appear to discuss the authenticity of a conversation. The Cluster represents 0.6% of the dataset, suggesting that the topic is not strongly related to the discussion on biases in AI systems within the r/ChatGPT subcommunity.

- *"Reminds me of a quote from Westworld... If you can't tell if something is real, does it matter? (2023-06-04)"*
- *"Can you provide proof of this alleged conversation? (2023-08-08)".*

Although discussions about fake content are sometimes associated with political discourse (Sharma et al, 2023), this does not seem to be the case for the comments analyzed in this study. In this context, the intersection of bias and politics appears to reflect user perceptions of AI replicating and reinforcing political-ideological views. Nevertheless, that is not possible to be concluded from the current analysis alone, and further investigation is necessary to explore it more thoroughly

In summary, the keyword "bias" in the r/ChatGPT subreddit primarily revolves around discussions on how AI will affect society, with a focus on government regulation and public policy. The themes of the largest clusters (C1, C2, C3, C4, C5, and C6) reveal what topics primarily concern ChatGPT users regarding bias. These themes also appear in the academic literature on AI biases (Caliskan, 2017; Caliskan, 2023; Sheng et al., 2019; Ferrara, 2023; Navigli, Conia, and Ross, 2023). Clusters like human, gender, and political bias underscore user awareness about AI's potential to replicate existing societal stereotypes. From the analysis, it is possible to observe that the

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

reinforcement of existing racial and gender stereotypes as presented by Bonezzi and Ostinelli (2021) is also perceived by the users.

The predominance of regulatory discussions may reflect a growing public attention to policy initiatives (some examples are the EU AI Act and the U.S. AI Bill of Rights). Taken together, the discussions across clusters suggests that user perceptions of AI bias have social, political and ethical dimensions that go beyond technical issues, revolving around its societal implications and ethical considerations, emphasizing the need for interdisciplinary approaches to address user concerns and align AI development with societal values.

This study contributes by 1) identifying how users perceive ethical and societal impacts of AI through large-scale discussions on Reddit; 2) mapping thematic clusters that reveal the diversity of public discourse around AI fairness, regulation, and human-AI interaction; and 3) by reinforcing that perceptions of AI bias are socially grounded, reflecting broader concerns about trust, accountability, and the role of technology in society.

4 CONCLUSIONS

This study explored how users perceive and discuss bias in artificial intelligence within the r/ChatGPT subreddit, analyzing over 8,000 posts and comments from December 2022 to October 2024. The prominence of the topic of AI regulation highlights this as the major public's interest around AI biases. The findings reveal that discussions about AI regulation dominate the debate, however, other central theme concerns political bias, a dimension less explored in previous literature, shows that users associate AI behavior with broader ideological polarization rather than purely technical flaws. This gap presents an opportunity for future research at the intersection of AI bias, social media, and politics.

The main contribution of this study lies in mapping large-scale user perceptions of AI bias on Reddit, an area still underexplored compared to technical or experimental research. By identifying thematic clusters and their underlying narratives, this research complements existing studies by foregrounding the social meanings users sometimes attach to AI in terms of fairness and accountability. There is an understanding that biases may be present in human-AI interactions and that AI systems can replicate human and social biases. Addressing these biases and preventing the

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

perpetuation of harmful stereotypes are critical to ensuring ethical and inclusive practices in AI applications.

There are some limitations in this study. First, the analysis focuses exclusively on discussions from the r/ChatGPT subreddit in English, collected during a defined period using the keyword “bias” as a filter. Therefore, the findings are not representative of all AI users and the results may change over time. However, the research provide exploratory insights into how a specific online community perceives AI bias. Future research could expand this investigation by conducting longitudinal analyses of how discussions evolve over time, or by comparing different platforms or linguistic communities.

Declaration on the use of artificial intelligence

Part of the translation of this article was assisted by artificial intelligence tools, with review and validation performed by the authors.

REFERENCES

AZARIA, Amos. *ChatGPT: More Human-Like Than Computer-Like, but Not Necessarily in a Good Way*. In: PROCEEDINGS OF THE 35TH IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE (ICTAI), 35., 2023. p. 468–473. Disponível em: https://doi.org/10.1109/ICTAI59109.2023.00074. Acesso em: 07 fev. 2025.

BINZ, Marcel.; SCHULZ, Eric. *Using cognitive psychology to understand GPT-3.* Proceedings of the National Academy of Sciences (PNAS), v. 120, n. 6, p. e2218523120, 2023. Disponível em: https://doi.org/10.1073/pnas.2218523120. Acesso em: 12 nov. 2025.
BONEZZI, Andrea.; OSTINELLI, Massimiliano. *Can algorithms legitimize discrimination?* Journal of Experimental Psychology: Applied, v. 27, n. 2, p. 447-459, 2021. Disponível em: https://doi.org/10.1037/xap0000294. Acesso em: 07 fev. 2025

BOYD, Danah. M.; ELLISON, Nicole. B. *Social Network Sites: Definition, History, and Scholarship.* Journal of Computer-Mediated Communication, v. 13, n. 1, p. 210–230, 2007. Disponível em: https://doi.org/10.1111/j.1083-6101.2007.00393.x. Acesso em: 07 fev. 2025.

CALISKAN, Aylin.; BRYSON, Joanna. J.; NARAYANAN, Arvind. *Semantics derived automatically from language corpora contain human-like biases.* Science, v. 356, n. 6334, p. 183–186, 2017. Disponível em: https://doi.org/10.1126/science.aal4230. Acesso em: 07 fev. 2025.

CALISKAN, Aylin. *Artificial intelligence, bias, and ethics.* In: PROCEEDINGS OF THE THIRTY-SECOND INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE (IJCAI 2023), 2023. Disponível em: https://doi.org/10.24963/ijcai.2023/799. Acesso em: 14 mai. 2025.

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

COCHRAN, William. G. *Sampling Techniques.* 3rd ed. Wiley, 1977.

FERRARA, Emilio. *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies.* Science, v. 6, n. 1, p. 3, 2024. Disponível em: https://doi.org/10.3390/sci6010003. Acesso em: 07 fev. 2025.

GAMITO, Marta Cantero; MARSDEN, Christopher T. Artificial intelligence co-regulation? The role of standards in the EU AI Act. International Journal of Law and Information Technology, v. 32, 2024. Disponível em: https://doi.org/10.1093/ijlit/eaae011. Acesso em: 2 nov. 2025.

GRUZD, Anatoliy.; MAI, Philip.; VAHEDI, Zahra. *Studying Anti-Social Behaviour on Reddit with Communalytic.* In: QUAN-HAASE, A.; SLOAN, L. (Eds.), The SAGE Handbook of Social Media Research Methods. SAGE Publications, p. 503–520, 2022. Disponível em: https://doi.org/10.4135/9781529782943.n36. Acesso em: 07 fev. 2025.

GRUZD, Anatoliy.; MAI, Philip. *Communalytic: A no-code computational social science research tool for studying online communities and public discourse on social media.* 2024. Disponível em: https://Communalytic.org. Acesso em: 14 nov. 2024.

GSTREIN, O. J.; HALEEM, N.; ZWITTER, A. General-purpose AI regulation and the European Union AI Act. Internet Policy Review, v. 13, n. 3, 2024. Disponível em: https://doi.org/10.14763/2024.3.1790. Acesso em: 2 nov. 2025.

GUIDI, Barbara.; IGLESIAS, Carlos. A.; ROSSETTI, Giulio.; KOIDL, Kevin. *Advanced Analysis Technologies for Social Media.* Applied Sciences, v. 13, n. 3, p. 1909, 2023. Disponível em: https://doi.org/10.3390/app13031909. Acesso em: 07 fev. 2025.

HORVÁT, Emőke-Ágnes; GONZÁLEZ-BAILÓN, Sandra. Quantifying gender disparities and bias online: editors' introduction to "Gender Gaps in Digital Spaces" special issue. Journal of Computer-Mediated Communication, v. 29, n. 1, jan. 2024. Disponível em: https://doi.org/10.1093/jcmc/zmad054. Acesso em: 2 nov. 2025.

LIU, J. ChatGPT: perspectives from human–computer interaction and psychology. Frontiers in Artificial Intelligence, v. 7, 2024. Disponível em: https://doi.org/10.3389/frai.2024.1418869. Acesso em: 2 nov. 2025.

MADHULATHA, T. Soni. *An overview on clustering methods.* IOSR Journal of Engineering, v. 2, n. 4, p. 719–725, 2012. Disponível em: https://doi.org/10.9790/3021-0204719725. Acesso em: 14 mai. 2025.

MORSTATTER, Fred.; LIU, Huan. *Discovering, assessing, and mitigating data bias in social media.* Online Social Networks and Media, v. 1, p. 1–13, 2017. Disponível em: https://doi.org/10.1016/j.osnem.2017.01.001. Acesso em: 07 fev. 2025.

ISSN 1983-1838
(DOI): 10.18624/ etech.v18i1.1415

NAVIGLI, Roberto.; CONIA, Simone.; ROSS, Björn. *Biases in Large Language Models: Origins, Inventory, and Discussion.* Journal of Data and Information Quality, v. 15, n. 2, Art. 10, 2023. Disponível em: https://doi.org/10.1145/3597307. Acesso em: 07 fev. 2025.

RAPP, Amon.; DI LODOVICO, Chiara.; TORRIELLI, Federico.; DI CARO, Luigi. *How do people experience the images created by generative artificial intelligence?* International Journal of Human-Computer Studies, v. 193, Art. 103375, 2025. Disponível em: https://doi.org/10.1016/j.ijhcs.2024.103375. Acesso em: 07 fev. 2025.

SARTORI, Giuseppe.; ORRÙ, Graziella. *Language models and psychological sciences.* Frontiers in Psychology, v. 14, Art. 1279317, 2023. Disponível em: https://doi.org/10.3389/fpsyg.2023.1279317. Acesso em: 07 fev. 2025.

SHAPIRA, Natalie. et al. *Clever Hans or Neural Theory of Mind? Stress Testing Social Reasoning in Large Language Models.* In: PROCEEDINGS OF THE 18TH CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 18., 2024. p. 2257–2273. Disponível em: https://aclanthology.org/2024.eacl-long.138/. Acesso em: 14 mai. 2025

SHARMA, Isha. et al. *Examining the motivations of sharing political deepfake videos: the role of political brand hate and moral consciousness.* Internet Research, v. 33, n. 5, p. 1727-1749, 2023. Disponível em: https://doi.org/10.1108/INTR-07-2022-0563. Acesso em: 07 fev. 2025.

SHENG, Emily. et al. *The Woman Worked as a Babysitter: On Biases in Language Generation.* In: PROCEEDINGS OF THE 2019 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP-IJCNLP), 2019. p. 3407–3412. Disponível em: https://doi.org/10.18653/v1/D19-1339. Acesso em: 07 fev. 2025.

WLODARCZAK, Peter.; SOAR, Jeffrey.; ALLY, Mustafa. A. *What the Future Holds for Social Media Data Analysis.* International Journal of Information, Control and Computer Sciences, v. 9, n. 1, 2015. Disponível em: https://scholarly.org/pdf/display/what-the-future-holds-for-social-media-data-analysis. Acesso em: 14 mai. 2025