

Public Perceptions, Critical Awareness, and Community Discourse on AI Ethics: Evidence from an Online Discussion Forum

Subhasree Sengupta
Clemson University
subhass@g.clemson.edu

Swapnil Srivastava
Clemson University
srivass7@g.clemson.edu

Nathan McNeese
Clemson University
mcneese@clemson.edu

Abstract

As Artificial Intelligence (AI) becomes increasingly ingrained into society, ethical and regularity concerns become critical. Given the vast array of philosophical considerations of AI ethics, there is a pressing need to understand and balance public opinion and expectations of how AI ethics should be defined and implemented, such that it centers the voice of experts and non-experts alike. This investigation explores a subreddit r/AIethics through a multi-methodological, multi-level approach. The analysis yielded six conversational themes, sentiment trends, and emergent roles that elicit narratives associated with expanding implementation, policy, critical literacy, communal preparedness, and increased awareness towards combining technical and social aspects of AI ethics. Such insights can help to distill necessary considerations for the practice of AI ethics beyond scholarly traditions and how informal spaces (such as virtual channels) can and should act as avenues of learning, raising critical consciousness, bolstering connectivity, and enhancing narrative agency on AI ethics.

Keywords: AI Ethics, Informal Learning, Critical Awareness, Public Perceptions, Reddit.

1. Introduction

Artificial Intelligence (AI) is becoming a key player across various social and cultural functions of our day-to-day existence (Makridakis, 2017). Given the increased use and adoption of AI-based artifacts for several critical decision-making considerations (such as being deployed in healthcare, education, and service sectors), AI's ethical impact and footprint have become important (Birhane et al., 2022). As AI is foreseen to grow in its level of sentience and may also be expected to act at par with human counterparts, the ethical aspect accrues further significance (Makridakis, 2017). It becomes imperative to consider how AI systems can be designed with ethical considerations

in mind and how ethical visions will impact the use and deployment of AI across different application domains (Crawford, 2021). However, a challenge with such initiatives is understanding how ethical perspectives shall be defined and how implicit power asymmetries that give rise to ethical concerns should be tackled and holistically dismantled (D'ignazio and Klein, 2020). Prior studies highlight how automated systems may, in some instances, amplify social prejudices and structural inequalities, causing harm and subjugation, further perpetuating structures of social segregation and exclusion (Mohamed et al., 2020). These explorations also indicate the challenges of infusing and incorporating the complex interconnected web of humanistic values and experiential wisdom into designing and implementing AI systems (Birhane, 2021).

Given the lingering issues with understanding and implementing ethical AI systems, it becomes essential to understand the public viewpoints and concerns so that these systems can be designed to augment societal tasks more responsibly and cater to the voices of the public (Crawford, 2021). Such visions also call for the need to initiate literacy visions aimed at educating and empowering the public to tackle the impending influence of AI in various societal functions. Understanding how to increase alertness towards ethical visions and tensions becomes critical (Long and Magerko, 2020). While some explorations have looked into public perceptions of AI, in general (Kelley et al., 2021), there is a scholarly void associated with understanding how people view, frame, and wish to engage in establishing a practice for the AI ethics community (Garrett et al., 2020). It is critical to better gauge how the public creates mental models and an understanding of AI-based systems as the importance of creating frameworks for designing ethical AI systems gains momentum (Sartori and Bocca, 2022). Prior work has indicated how the public creates mental frameworks that capture the visions, perspectives, expectations, and future goals people associate with AI and how

it will impact society (Sartori and Bocca, 2022). Understanding how public perceptions can impact AI ethics as a practice is crucial in creating a climate that increases public readiness to thrive in an AI-driven society (Sartori and Bocca, 2022). While most prior studies have looked into public perceptions through surveys (Kelley et al., 2021), to broaden further the scope and reach of the platform and forum through which the public voices its thoughts and opinions, we investigate how an online discussion forum can provide insights into the public perceptions associated with AI ethics. When considering online discussion channels, Reddit has emerged as a space that affords discussion and deliberation on diverse topics (Anderson, 2015). Research has also indicated how Reddit and its many conversational affordances help bolster disclosure and offer safe spaces for deliberation on sensitive issues (De Choudhury and De, 2014). Studies surmise how such online forums may enthuse and bolster narrative agency, empowering the views and voice of the public (Schoot, 2022). Given this background, we investigate how a subreddit *r/AIethics*¹ fuels conversations that help to discern public perceptions and expectations around the practice of AI ethics, especially concerning how it is perceived and discussed in the forum of interest (*r/AIethics*). Further, we also draw on prior research on how online communities may serve as augmented avenues for informal learning to highlight how peer-led conversations may provide a collaborative arena for enrichment to develop competencies to tackle the issue of AI ethics at the societal level (Long and Magerko, 2020; Sengupta, 2021). The key questions we address in this study are:

- *What content themes can be discerned from r/AIethics?*
- *Based on the emergent network characteristics, what participation trends and roles do community members assume in r/AIethics?*
- *What key sentiment characteristics characterize the conversations on r/AIethics?*

2. Related Work

2.1. The Social Stage of AI - Perceptions, Narratives, and Civic Empowerment

As AI's technical development and advancement booms at breakneck speed, the societal implications of AI use have gained importance, especially considering the sociological impacts and influence of this technical

innovation (Martinho et al., 2021). Despite its computational supremacy and promise of prowess, scholars have indicated how AI may introduce several harms that magnify exclusions, stigma, and prejudice of the historically oppressed and marginalized (D'ignazio and Klein, 2020). Calls for democratizing AI conceptualization and use cases, de-centering and dismantling latent power structures embedded within the echelons of organizations at the helm of the AI revolution have become critical (Crawford, 2021). Increasingly, taking a humanistic turn, critical visions toward AI advocate for the need and vision to repurpose and recalibrate practices, values, and societal functions associated with AI development and deployment (Crawford, 2021).

The need to understand the public sentiment, responses, perceptions, and expectations becomes crucial as we consider the impact of AI and attempt to develop collective awareness of the future footprint of AI (Kelley et al., 2021). A deeper understanding of public perceptions can not only help to grow awareness toward how AI is viewed and capture associated public sentiment, but it can also help to fuel and drive the agenda of literacy initiatives, providing a window into the needs and visions that such causes should support and scaffold (Sartori and Bocca, 2022). Prior work on public perception has captured temporal trends in the public's view of AI, the influence of media portrayals, narratives of apprehension, expectations, and the overall sociocultural climate engulfing the public's association with AI as a sociotechnical artifact (Sartori and Bocca, 2022). Ethical visions of AI thus become the nucleus of the social manifestation of AI's present and future footprint. While the focus of prior research has primarily been on AI, this investigation expands this prior trajectory of research by exploring an online community on AI ethics to capture how, through the medium of an online forum, visions, imaginaries, and expectations associated with AI ethics are constituted and advocated for.

2.2. Online Forums - Knowledge Sharing, Learning, and Community Building

A burgeoning array of scholarship narrates how online forums have quickly acquired widespread popularity and use for a cross-section of purposes, serving as a chief outlet for knowledge sharing, rapport building, and creative expression and may even as platforms for developing shared sensibilities of empowerment, morality, and advocacy (Sengupta and Tacheva, 2022). Further, the various affordances of online discussion forums (such as Reddit) have

¹<https://www.reddit.com/r/AIethics/>

empowered collective sharing and have propelled support to be provided for critical and stigmatized causes, which have further fueled the propensity of open sharing and disclosure (Andalibi et al., 2016). Through conversation, online communities also offer avenues for relationship building, establishing social capital that nurtures a sense of shared vision and purpose, enriching conversations, and shaping the relational fabric of such virtual enclaves (Martey and Stromer-Galley, 2007). Knowledge sharing, personal development, and informal learning are also bolstered by peer-led activities that online forums scaffold. For example, prior work highlights how discussion forums nurture a combination of socialization needs, learning needs, and peer-driven mentoring initiatives, augmenting the resources available for self-enhancement and professional development (Sengupta, 2021).

As online forums are created and maintained by the participation and engagement of their members, it becomes critical to understand how different actions and roles assumed by community members sustain such conversational spaces (Johnson et al., 2020). Roles in online forums capture the functions that community members implicitly or explicitly assume to motivate contribution and set the conversational habits of the forum (Johnson et al., 2020). Roles typically are marked by behavioral characteristics captured through community members' discourse and activity traces (Johnson et al., 2020). Understanding such contribution patterns can help to provide insight into the structural characteristics and how community members assimilate the community's norms (Sengupta and Tacheva, 2022). Such norms may mediate the nature of sociality, relationship maintenance, and the extent to which such online enterprises boost narrative agency development (Andalibi et al., 2016). Hence, through various roles, actions, and conversational routines, each online forum assumes a unique collective identity replete with a distinct communal affinity and task orientation, shaping its overarching vision and purpose (Martey and Stromer-Galley, 2007).

3. Methods

To address the research questions of interest, we employed a combination of content analysis, network analysis, and sentiment analysis to observe critical trends of participation and discourse that emerge in these forums. Given that online forums are flexible and driven by the content produced and consumed by the community participants, applying a combined toolkit of different methodological perspectives helps to elucidate varied dimensions of conversation, engagement, and

community culture that are nurtured and sustained in such forums (Kou et al., 2018).

3.1. Data Collection

The data was collected from the creation date of the subreddit (06/08/2016) to the present day (05/01/2023) using a combination of existing databases on Reddit data (PRAW) and Python API wrappers (PMAW) (Sengupta, 2021). An initial data sanitization was conducted to remove data points that were deleted or removed. Further, to construct a network, all those data points that did not have author information were also pruned out. The final dataset comprised 808 conversational elements (including top-level posts and replies). Further details are provided in table 2.

3.2. Analysis Details

For content analysis, a combined pipeline of quantitative and qualitative techniques was followed (Sengupta, 2021). First, topic modeling (using the non-negative matrix factorization approach) augmented with topic coherence yielded an initial estimate of 8 topics. NMF was used instead of other methods (such as LDA (Latent Dirichlet Allocation)), given that prior research has demonstrated its effectiveness and robustness when considering social media data (Habbat et al., 2021). Topic coherence was employed to compute the best number of topics, comparing across different topic models (Habbat et al., 2021). Table 1 details the initial 8 themes inspected based on the NMF results. Figure 1 visualizes the performance of the topic models. Following a constructivist approach to grounded theory, qualitative thematic coding was conducted to prune, refine, and add contextual relevance to this initial computational categorization (Charmaz, 2006; Habbat et al., 2021). The initial quantitative insights from the topic modeling were used as a guiding mechanism, akin to sensitizing devices as described in (Charmaz, 2006). Two stages of this qualitative refinement were conducted. The first stage delved into identifying the themes based on the topical modeling results. Two coders independently coded an initial random sample of 50 conversations, arriving at an inter-coder reliability of 0.88 (Charmaz, 2006). The random sample was selected such that each topical category was present proportionately in the selected sample. Disagreements were discussed to reach a consensus regarding each category's final descriptions and nuances. The thematic categorization achieved after the 1st round is further detailed in table 1. In the 2nd round of qualitative coding, the initial schema was consolidated, combining topics with conceptual parity and overlap. For example,

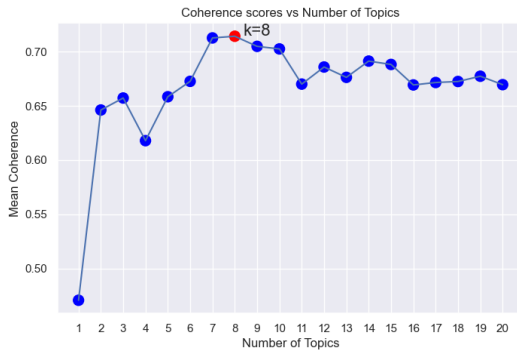


Figure 1. Topic modeling results

the philosophical considerations and critical probing categories were paired since they had similar contextual and conceptual motives. Such refinement also helps to overcome any shortcomings of the topic modeling approach, given that it is driven by semantic properties that are computationally defined (Egger and Yu, 2022). The final refined schema comprised 6 themes (described in subsequent sections).

Sentiment analysis adds nuance to the content categorization by helping to capture the type of sentiment and spectrum of emotions that manifest in the community discourse (Neidhardt et al., 2017). The VADER python library was used to perform sentiment analysis (Hutto and Gilbert, 2014). While content and sentiment analysis help capture the disposition of community members based on their conversational traces, social network analysis (SNA) helps establish the social structure and interaction patterns within each community by representing each community member as a node whose connections to other members, or edges, denote comments issued or received (Rabbany et al., 2014). Theoretically, this approach helps understand how individuals assume positions or roles that impact participation and communication patterns (Rabbany et al., 2014). The social network was compiled as a directed reply network, with nodes representing authors and each edge corresponding to a comment mentioned to a top-level post in a thread (Staudt Willet and Carpenter, 2020). Figure 3 depicts the network visualization, and table 3 shows the key metrics.

4. Results

This section describes the details of the insights captured from the conversational traces of r/AIEthics. First, the emergent content themes are discussed (Figure 2), and subsequently, additional observations based on sentiment (figures 4, 5) and network analysis (table

Table 1. Initial NMF topics and key identifiers

Topical theme	Top 5 keywords associated with each theme
State of AI ethics scholarship	<i>AI, Ethics, Research, Bias, Article</i>
Resource sharing for collective consciousness	<i>Reddit, YouTube, Link, Share, Read</i>
Perspectives on ethical AI development	<i>Goals, Function, Behavior, Utility, Consciousness</i>
Community building for collective action	<i>Thanks, Please, Love, Help, Posting</i>
Influence of ethical AI on society	<i>People, Think, Good, Cars, Ethical</i>
Considerations for ethical practice of AI	<i>Data, Algorithms, Assumptions, Race, Code</i>
Philosophical considerations for outlining ethical AI agenda	<i>Morals, Theories, Actions, Autonomous, Model</i>
Critical probing of ethical considerations	<i>Rights, humans, Intelligence, Laws, Suffering</i>

Table 2. Data statistics for the Forum

Data set descriptor	Value
Number of conversations	808
Earliest creation date	07/01/2016
Latest creation date	04/01/2023
Number of unique threads	219
Number of unique authors	271
Average score	2.364

3, figure 3) highlight nuances about participation and engagement trends that emerge in this subreddit.

4.1. Description of content themes

Critical probing of ethical considerations: These conversations provoked critical thought about how ethical stances and visions should be understood. These conversations ignite and provoke discussion around the agenda and definition of ethics, blending many different rationales from varied schools of thought (Whittlestone et al., 2019). Thus, these conversations aimed at sparking profound thought and deepening critical insight, creating a trajectory for advocating for the need to expand the scope of how ethics should be conceptualized when considering implementation and societal ramifications associated with employing AI. An example quote aligned with these visions stated:

Table 3. Network Statistics for the forum

Network metric	Value
Number of nodes	202
Number of edges	291
Average degree	1.44
Density	0.007
Number of connected components	18
Average clustering coefficient	0.056

“We probably still do many immoral things today and want future AI systems to learn better. There are many standards of morality - take, for example, the Catholic Church, yet none are absolute. So finalizing a set of values needs more thought”. Some of these conversations also revolved around decision-making when ethical dilemmas arise and attempted to advocate for embedding deeper reflection on varied possibilities and experiences an AI agent may have to tackle. An example quote aligned with this theme stated: “If an automated car pilot had to choose between saving the human driver or an animal on the road, what should it do? How should it decide? Is it that an animal’s life will have less value than a human’s?”. Such quotes also highlight latent flaws associated with how ethics is practiced in society, which may impede how the agenda of AI ethics is delineated and envisioned (Whittlestone et al., 2019). A vital dimension of this discussion centered on the issue of *machine consciousness*, the definition, and the associated implications. These themes depict the dilemma around how consciousness between human and AI agents shall be distinguished and how ensuing tensions and comparative foundations shall be navigated to address the issue of embedding *sentient* machines into society (Gamez, 2018). For example, a quote mentioned: “Something has to be conscious before it can suffer or before the appearance of its suffering takes on any moral valence. Can a robot surgeon have the same level of sentience as an automated car pilot?”. This quote indicates that definitions and visions for the ethical design of automated agents may need to account for contextual variances, potentially dismissing a monolithic approach to AI ethics as a panacea (Omrani et al., 2022).

Influence of ethical AI on society: Under this category, the posts elicited concerns and perspectives around how the emergent debates on ethics may not only influence the AI community and perspectives for imbuing AI agents with ethical visions but also may, in turn, impact how human society embraces AI and evolves in parallel. Conversations discussed pertinent considerations regarding how AI training initiatives may also need to incorporate policy-linked education that

provides a basis for critical reasoning to bolster AI ethics initiatives (Garrett et al., 2020). For example, a comment stated: “An AI designer should be aware of how the human mind works and try to reproduce the most important traits with a much cleaner and simpler design.”. Another stream of thought focused on understanding how ethical AI may influence changes in societal order and value systems (Crawford, 2021). For example, a quote stated: “How do you think having an ethical AI agent will change our rights? Will it also impact our way of deciding ethical matters?”. Some posts also discussed how imbuing AI with ethical perspectives can impact social structure and national policy. An example comment stated: “Legislation will be necessary to not only control harmful intent and actions of AI but also to understand how human rights may change with AI systems exhibiting ethical thought inline with human values”.

Perspectives on ethical AI implementation and decision-making: These conversations discussed rationales and perspectives on how AI systems should reason and engage in ethical decision-making. This theme captures public opinion and views on the implementation details and associated policy considerations needed with the issue of AI ethics. Many of the conversations indicated limitations of mathematical thought that constrain the incorporation of ethical perspectives into AI agents. An example comment stated: “One of AI’s limitations is the emotion feedback loop. Without the ability to feel pain, it will be hard to grasp what it feels like to cause pain to others.”. Other pertinent conversational traces also advocated for public participation and contribution to defining the trajectory of the AI ethics community and policy considerations needed to safeguard the safety of vital AI assets to thwart any attempts to impinge models that govern ethical visions of AI. For example, a comment indicated “The value of lives should be decided by society at large, instead of just engineers and researchers”. This conversational trace urges the need for egalitarian perspectives to define the AI ethics agenda (Schiff et al., 2020).

Community building to cultivate narrative agency on AI ethics: These conversations encouraged participation and engagement with the Reddit forum. Akin to prior studies, many comments were geared towards encouraging, validating, and acknowledging other community members for their thoughts and ideas. An example comment stated: “I just logged into Reddit and saw a post from r/AIethics trending, I am happy to be here!”. Other posts also detailed strategies for improving information curation and knowledge management practices (Bettoni et al., 2007).

An example comment stated: “We are creating new flairs for people with different experiences, there will be separate flairs for computer science and philosophy”. Some posts also depicted attempts to create more personalized connections. For example, a quote stated: “Let us continue to discuss this, I have sent you a chat request”. Some conversations under this category revolved around understanding and interpreting community norms. An example post stated: “Why was my post downvoted, you may disagree with my views, but it is still an exciting area to unpack”. All these conversations highlight how community members develop a shared rapport and sense of togetherness when collaborating in virtual spaces (Sengupta, 2021). Further, some posts aimed to encourage participation in initiatives that advocate for AI ethics. An example quote stated: “Definitely consider volunteering for organizations that are trying to build that climate (referring to ethical stances and practices)”. Thus, there was a thrust to invigorate participation and collective purpose toward AI ethics.

Resource sharing to boost collective ethical consciousness: Complementing the visions of community building, this category catered to conversations that share essential materials and resources that can help to elucidate critical details and directions for the AI ethics field. For example, a post stated: “A few good reads on the topic (referring to AI ethics) that highlight how science and philosophy come together”. Some posts also asked about pertinent information, as one comment stated: “I am so interested in algorithmic fairness, anyone can share some essential resources for getting started on this topic?”. Some posts also shared details about other virtual forums or groups (discord channels), which can serve as resources for developing knowledge and insight into ethics. These conversations highlight the knowledge capital created and maintained in these virtual spaces, making them powerful sources of learning and understanding multidimensional perspectives associated with the topic of ethics (Hall and Graham, 2004). This theme also highlights how community members tap into other community members’ experiential wisdom and knowledge, indicating their mental model of the type of wisdom and knowledge held by the community (Sharratt and Usoro, 2003).

4.2. Inferences from sentiment analysis

Sentiment analysis helps capture the community’s overall emotional orientation (Elhadad et al., 2014). Prior studies investigating public sentiment around AI

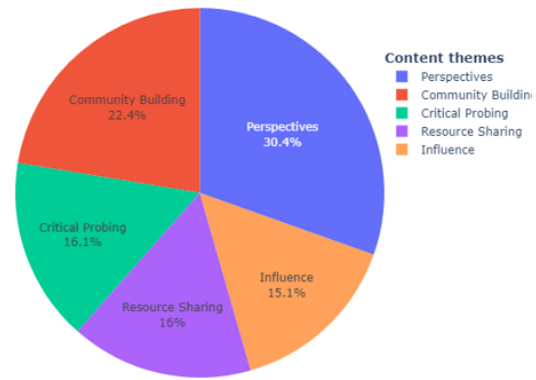


Figure 2. Distribution of content themes

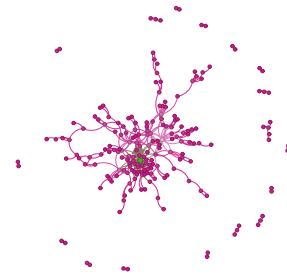


Figure 3. Network representation of the online forum

have also highlighted how differences in sentiment can help provide a basic understanding of public acceptance, adoption, and potential aversion towards AI. Thus, such an understanding of sentiment can serve as a powerful metric to account for, including public demands and visions (Kelley et al., 2021). Overall, most conversations depict neutral sentimental attitudes; however, a strong inclination towards an expression of negative sentiment was also discerned from the analysis. Positive sentiment manifested least in the context of the overall data set. Overall, such emotional valence essentially highlights the underlying tensions, hopes, and visions that characterize how the public (particularly this Reddit community) ascribes viewpoints concerning the issue of AI ethics.

To gain deeper insight into how emotions align with the various topical dimensions, we find that community building is a content theme that captures the most significant proportion of positive sentiment. This can be attributed to positive sentiment associated with acknowledgments, validations, and appeals for camaraderie among community members.

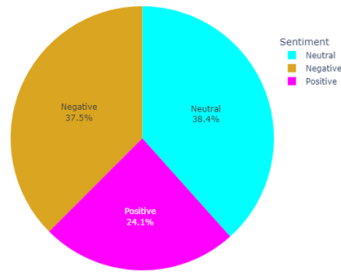


Figure 4. Overall Sentiment distribution in the forum

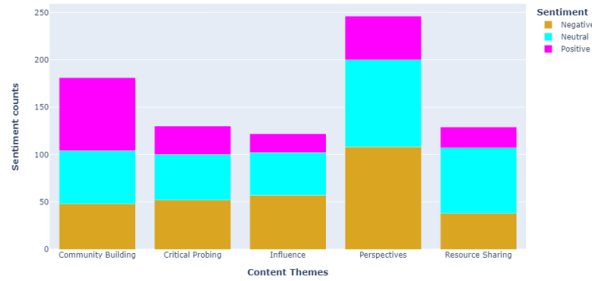


Figure 5. Sentiment across the different topical dimensions

Thus demonstrating how this Reddit community places importance on welcoming conversations and community growth. The prevalence of negative emotions among this same topical theme is associated with posts that question the relevance of content and preservation of community norms. The higher margin of negative sentiment is also associated with perspectives on ethical AI implementation and decision-making concerning the posts that discuss policy considerations and the need for greater incorporation of societal values and perspectives in how ethical stances are defined. These capture public disillusionment and resentment associated with how organizations and authorities of power manage the stipulation and implementation of ethical guidelines associated with AI deployment and use (Crawford, 2021). Thus, a close inspection of sentiment and the content themes highlights not only the emotional responses associated with the topic of AI ethics but also elucidates how community members engrave normative conventions of articulation, expression, and participation as they engage in such virtual forums (Ivaturi and Chua, 2019).

4.3. Inferences from network analysis

To further gauge the activity trends, insights from network analysis reveal structural aspects and mechanisms through which information is shared and curated in these online spaces (Haythornthwaite,

2005). A sparse, fragmented network emerges based on the macro or the group-level metrics (Table 3). This can be inferred given the many connected components and low average clustering coefficient. Complementing these views, based on the network visual (Figure 3), we can confirm that the level of cohesion in the network is low, indicating conversational silos and participation in specific topics by different members of the community (Rabbany et al., 2014). We can see the emergence of a central core of connected actors surrounded by several peripheral participants. Further inspection shows that this central core is associated with conversations under the topical umbrella of *resource sharing* and *perspectives on ethical AI implementation and decision-making*. This core's emergence could also indicate commitment and dedication to the community, indicating an attempt to engender connectivity and collective purpose in the virtual forum (Haythornthwaite, 2005).

Beyond the macro-level (or the collective footprint) of the community, as discussed above, to distill further inferences about individual node-level tendencies, we investigate the top 25 nodes ranked based on degree centrality, betweenness centrality (the propensity of bridging subgraphs) and closeness centrality (a measure of proximity to other nodes in the network) and the type of topics these actors contribute to (Rabbany et al., 2014). As per degree centrality, the top nodes participate in conversations on *resource sharing*, *community building*, and *perspectives on ethical AI implementation and decision-making*. This highlights that the community ushers conversations that enable the community to grow and exchange information on AI ethics to enhance the knowledge pool curated by the community. Interestingly, as per betweenness centrality, the top nodes participate most prominently in *perspectives on AI implementation and decision-making*, *critical probing on ethical considerations*, and *discussions on community building and resource sharing*. Thus, the role of these actors involves infusing critical thoughts and opinions that spark novel strands of discussions among community members (Faraj et al., 2011). The trends associated with the top 25 actors, as per closeness centrality, mirror the trends discussed for the other two centrality measures. This indicates that a set of actors within the community's core is instrumental in propelling conversation and collective enrichment. While some are crucial in binding the community, others guide and lead the community. In aggregate, the network analysis helps to discern how community members assume roles essential for such spaces to thrive and survive.

5. Discussion

While a vast array of scholarship has investigated public perceptions of AI, this study explores a critical extension to this stream of scholarship by particularly looking into perceptions of AI ethics. Further, investigating virtual forums such as Reddit is another novel layer of inspection associated with this study. Triangulating the inferences from the three different methodological tools employed in this inspection, we find several significant findings. Firstly, the five content themes provide a preliminary but pertinent typology of the perceptions discussed in such online spaces. Augmented with network and sentiment analysis, we can understand who drives the topical enrichment of these spaces and the overall emotional contour of the community. The insights from these forums can inform the technical and social aspects of the overarching agenda of AI ethics in various ways.

Themes such as *critical probing of AI* and *influence of ethical AI on society* raise vital questions regarding how AI should be defined and constituted, especially regarding capturing the public's voice and visions about this topic. Such discussions distill the need to arouse public consciousness, awareness, proactivity, and overall competency in ensuring AI is implemented, used, and adopted responsibly and judiciously. Such information can play a role in understanding how to develop training and competencies that can shape how the public develops a form of readiness and preparedness toward the growing use of AI artifacts (Long and Magerko, 2020). Further, concerning the issue of competency development, through this analysis, we elucidate how online discussion forums, through collaborative discussion and information exchange, serve as an informal ground to develop such competencies. For example, the thematic category of critical probing highlights how community members engage in conversations that heighten the level of *critical consciousness*, provoking and expanding the margins of critical insight associated with the issue of AI ethics, sharpening the critical competency of the community (Watts et al., 2011). Such critical aptitude can also awaken and sharpen ethical sensibilities and widen the spectrum of perspectives associated with nurturing awareness of ethical visions (Garrett et al., 2020).

Beyond the topical dimensions of how these online forums can serve as informal pedagogical devices, the roles that actors assume depict how peer-led activities serve as a basis for knowledge advancement and aid in community enrichment and sustenance. Such forums can thus play a pivotal role in dispelling such fears

by sharing resources and perspectives on AI ethics' emergent regulations and practices. For example, the category of *community building* encompasses communicative acts that attempt to appropriate the type of conversations that are discussed through and on such forums, which highlights how community members, through their experience in engaging with community discourse, implicitly assume the role of information regulators (Gorwa, 2019). Also, the theme of resource sharing, which encapsulates the sharing of information and materials on varied topics associated with AI ethics, highlights how these online forums become avenues of informal learning. These conversations enrich and add to the knowledge capital of such collectives (Haythornthwaite, 2005). Such insights can have design and policy implications for how online forums can and should play a part in nurturing pedagogy about the emerging field of AI ethics (Ibáñez and Olmeda, 2022).

Many of these themes, particularly the one on *influence*, indicate the visions and gaps that need to be fulfilled to create a practice of AI Ethics that values scholarly practices and addresses public needs and demands. For example, the views on privacy or implicit training needs for developing ethical alertness of the public indicate how the public envisions involvement in developing an agenda for AI ethics. Such viewpoints can play a crucial role in understanding and addressing the role of non-experts or how non-experts and experts can collaboratively tackle the issue of designing ethical AI systems (Sartori and Bocca, 2022). Further, remarks about instituting training for technical experts and social policy experts indicate that the pressing need to address ethics will call for restructuring the curriculum and operational expertise of both technical and social fields, which in the long term can impact literary initiatives of the AI discipline, in addition, the competency needs of the general public as discussed above (Garrett et al., 2020). This conversational analysis shows how the public enacts agency in outlining the agenda surrounding AI ethics. Thus, such informal virtual discussion channels can enable, embolden, and empower the public to develop a form of narrative agency enacted by sharing their views, opinions, sentiments, and desires for the practice of AI ethics.

6. Conclusion and Future Work

Our investigation provides a multi-method, multi-level analysis of how conversations in online forums can help capture public perceptions, imaginaries, and needs for how the practice of AI ethics should be constituted and aligned with the visions of society at large. Such insights can help provide directions

toward how non-experts should be engaged to tackle the many confounding dilemmas that plague the AI ethics conundrum. Thus, in the aggregate, through this analysis, we provide an initial foray into how online enclaves act as avenues for peer-led knowledge activities to empower and enrich public awareness and bolster participation in the issue of AI ethics. Such insights can impact AI implementation, competencies, and training development on a larger scale, potentially informing AI policy and societal preparedness. Yet, this work is limited as it captures Reddit data from a single Reddit forum. Hence, enriching these initial visions with additional investigations and inquiries into public needs and demands will further add to the relevance of this study. Future work can expand the analysis's scope by providing more robust content analysis trends and network insights. Longitudinal inspection of the data can also be a line of future expansion. Further, additional insights can be captured from a complementary survey and interview analysis to understand perspectives, needs, and ways to more holistically engage and cater to the public as the field and practice of AI ethics are defined and implemented.

References

- Andalibi, N., Haimson, O. L., De Choudhury, M., & Forte, A. (2016). Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. *Proceedings of the 2016 CHI conference on human factors in computing systems*, 3906–3918.
- Anderson, K. E. (2015). Ask me anything: What is reddit? *Library Hi Tech News*, 32(5), 8–11.
- Bettoni, M. C., Andenmatten, S., & Mathieu, R. (2007). Knowledge cooperation in online communities: A duality of participation and cultivation. *Electronic Journal of Knowledge Management*, 5(1), pp1–6.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), 100205.
- Birhane, A., Ruane, E., Laurent, T., S. Brown, M., Flowers, J., Ventresque, A., & L. Dancy, C. (2022). The forgotten margins of ai ethics. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 948–958.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
- Crawford, K. (2021). *The atlas of ai: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- De Choudhury, M., & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. *Proceedings of the international AAAI conference on web and social media*, 8(1), 71–80.
- D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.
- Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, 7, 886498.
- Elhadad, N., Zhang, S., Driscoll, P., & Brody, S. (2014). Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions. *AMIA Annual Symposium Proceedings, 2014*, 516.
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization science*, 22(5), 1224–1239.
- Gamez, D. (2018). *Human and machine consciousness*. Open Book Publishers.
- Garrett, N., Beard, N., & Fiesler, C. (2020). More than” if time allows” the role of ethics in ai education. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 272–278.
- Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2), 1–22.
- Habbat, N., Anoun, H., & Hassouni, L. (2021). Topic modeling and sentiment analysis with lda and nmf on moroccan tweets. *Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications*, 147–161.
- Hall, H., & Graham, D. (2004). Creation and recreation: Motivating collaboration to generate knowledge capital in online communities. *International Journal of Information Management*, 24(3), 235–246.
- Haythornthwaite, C. (2005). Social networks and internet connectivity effects. *Information, Community & Society*, 8(2), 125–147.
- Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, 8(1), 216–225.
- Ibáñez, J. C., & Olmeda, M. V. (2022). Operationalising ai ethics: How are companies bridging the gap between practice and principles? an

- exploratory study. *AI & SOCIETY*, 37(4), 1663–1687.
- Ivaturi, K., & Chua, C. (2019). Framing norms in online communities. *Information & Management*, 56(1), 15–27.
- Johnson, J., Black, R. W., & Hayes, G. R. (2020). Roles in the discussion: An analysis of social support in an online forum for people with dementia. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1–30.
- Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T., & Woodruff, A. (2021). Exciting, useful, worrying, futuristic: Public perception of artificial intelligence in 8 countries. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 627–637.
- Kou, Y., Gray, C. M., Toombs, A. L., & Adams, R. S. (2018). Understanding social roles in an online community of volatile practice: A study of user experience practitioners on reddit. *ACM Transactions on Social Computing*, 1(4), 1–22.
- Long, D., & Magerko, B. (2020). What is ai literacy? competencies and design considerations. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–16.
- Makridakis, S. (2017). The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. *Futures*, 90, 46–60.
- Martey, R. M., & Stromer-Galley, J. (2007). The digital dollhouse: Context and social norms in the sims online. *Games and Culture*, 2(4), 314–334.
- Martinho, A., Poulsen, A., Kroesen, M., & Chorus, C. (2021). Perspectives about artificial moral agents. *AI and Ethics*, 1(4), 477–490.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial ai: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659–684.
- Neidhardt, J., Rümmele, N., & Werthner, H. (2017). Predicting happiness: User interactions and sentiment analysis in an online travel forum. *Information Technology & Tourism*, 17, 101–119.
- Omrani, N., Riviuccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? an assessment of trust in ai-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*, 181, 121763.
- Rabbany, R., Elatia, S., Takaffoli, M., & Zarane, O. R. (2014). Collaborative learning of students in online discussion forums: A social network analysis perspective. In *Educational data mining* (pp. 441–466). Springer.
- Sartori, L., & Bocca, G. (2022). Minding the gap (s): Public perceptions of ai and socio-technical imaginaries. *AI & SOCIETY*, 1–16.
- Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for ai ethics, policy, and governance? a global overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158.
- Schoot, J. M. (2022). *Youth's participation in online support forums on reddit, and the influence it has on the communion and agency identity development processes* (Master's thesis). University of Twente.
- Sengupta, S. (2021). A tale of two virtual communities: A comparative analysis of culture and discourse in two online programming communities.
- Sengupta, S., & Tacheva, J. (2022). Flames of justice in a virtual garden: An analysis of a digital campaign on twitter surrounding the death of an indian celebrity. *Information for a Better World: Shaping the Global Future: 17th International Conference, iConference 2022, Virtual Event, February 28–March 4, 2022, Proceedings, Part II*, 243–258.
- Sharratt, M., & Usoro, A. (2003). Understanding knowledge-sharing in online communities of practice. *Electronic Journal of Knowledge Management*, 1(2), pp18–27.
- Staudt Willet, K. B., & Carpenter, J. P. (2020). Teachers on reddit? exploring contributions and interactions in four teaching-related subreddits. *Journal of Research on Technology in Education*, 52(2), 216–233.
- Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *New directions for child and adolescent development*, 2011(134), 43–57.
- Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in ai ethics: Towards a focus on tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200.