

The Double Burden of Moderators: Safeguarding Communities While Facing Identity-Based Harassment

Nazanin Sabri^{ID} and Kristen Vaccaro^{ID} | University of California, San Diego
Mai ElSherief^{ID} | Northeastern University

Moderators help safeguard online spaces but can experience personal attacks. As platforms introduce new technologies, users more easily infer characteristics like gender and sexuality. This article explores how increased visibility affects moderators, highlighting key challenges and important future research directions.

Introduction

Men who are moderating spaces, because of the tone that they use, they can easily navigate their rooms and people will have this respect for them. [...] But] when women moderate, you can tell that men [...] usually don't know how to submit to the person that is running the room. [...]. You can hear their ideals of what they feel that women should be doing. And that's been a lot of basically the war of the sexes, the battle of the sexes happening live. (Clubhouse Moderator 1)^a

Moderators are often positioned as protectors of online communities, tasked with ensuring respectful

interactions, mitigating conflicts, and upholding community norms. Traditionally, text-based platforms allowed moderators to interact anonymously, often shielding their identities from users. However, with the rise of social audio and virtual reality (VR) platforms, moderators' voices, avatars, and behavioral patterns can provide more visible identity cues, making it easier for users to infer demographic characteristics such as gender, race, disability, or sexual identity. This ease could introduce new challenges to moderation, including harassment or resistance moderators might face tied to assumptions about their identity.

In social audio platforms, such as Clubhouse or X^b Spaces, moderators manage real-time voice discussions, guiding conversations, resolving disputes, and removing harmful or disruptive participants. Similarly, in VR spaces like Horizon Worlds or VRChat, moderators oversee virtual environments, where they manage interactions not only through voice but also through avatars, often

^aThis work discusses harms experienced by moderators. Profanity or potentially triggering words, as shared by our participants, have been replaced with high-level descriptions of the event. Additionally, quotes have been edited to remove hedging language.

^bFormerly Twitter.

facilitating or regulating virtual events. Their tasks are demanding as they must act quickly to address inappropriate behavior or content in an ephemeral, fast-moving setting, requiring a balance of authority, diplomacy, and sometimes rapid decision making to protect the community while ensuring fair treatment for all participants.¹

The quote displayed at the beginning of this work showcases one example of the types of challenges moderators of social audio platforms such as Clubhouse could face because of their gender. This experience was extracted from a set of interviews conducted with volunteer social audio and VR moderators. This dataset, previously published in Sabri et al.,¹ explored moderation practices and challenges. A challenge that emerged through these interviews was moderators discussing instances in which they were directly attacked (i.e., virtually or verbally) by participants of the event. Another moderator, of a social VR space, discussed how “after several years of hosting [this event], I had discovered that I’m genderfluid. And I made a few adjustments ... [to] how I look and the title and everything to better reflect [who I am]” (AltspaceVR Moderator 1). Despite the moderator’s happiness with their updated appearance, their audience did not respond well. In the next event they hosted, a regular member came up to them and said, “What happened to your mustache? You look like [an antigay slur]” (AltspaceVR Moderator 1). The participant expressed how the shock of this comment rendered them incapable of taking action and that other moderators in the space needed to step in to remove the offending user.

Another moderator of social VR shared an experience in which, based on their perceived gender, they were targeted for sexual harassment. They described an event in which they noticed a user moving around the space oddly, in difficult to predict ways. The moderator noted that “because your avatar has a really limited perspective view,” multiple users can work in tandem to provide distraction, obstruction, and interference “almost like a real world Snatch and Grab [...] it’s coordinated [...] they’ll come in and they’ll distract someone” (AltspaceVR Moderator 2). As a result of this distraction, the moderator could not see that another user had approached them from behind. Their team observed this user “[coming] up behind me and started stroking my avatar and grabbing my avatar” (AltspaceVR Moderator 2). While they expressed that “I couldn’t feel it and I couldn’t see it,” they described how their avatar had been sexually harassed and assaulted. Again, it was team members who, after observing this behavior, removed the offending user from the event.

Mentions of these instances came up naturally when moderators discussed negative experiences on these platforms more broadly. Interviewees were not prompted to discuss how aspects of their identity (e.g.,

gender) affected their experiences.^c Nevertheless, these instances showcase that moderators often face targeted attacks and biases based on their perceived identity, affecting both their authority and the overall moderation process. In the remainder of this work, we discuss future areas of work that could look into this issue and investigate strategies that could be used to manage such incidents. This study contributes to the understanding of how identity visibility shapes the dynamics of moderation and highlights the need for further understanding of the issue, as well as the design of interventions to support moderators in these evolving digital environments.

Ephemeral Content Moderation

Ephemeral Social Media Platforms

We consider ephemeral social media interactions to be those that disappear immediately after they are experienced. Some examples of ephemeral interactions include 1. live-streamed content on platforms such as Instagram, YouTube, or Twitch; 2. real-time audio discussions on platforms such as Clubhouse, Discord, or X Spaces; 3. real-time immersive interactions on social VR platforms such as Horizon Worlds and VRChat; and 4. real-time video conversation on platforms like Zoom and Slack Huddles. These platforms allow users to engage with others through voice and gestures in real time.

Social audio platforms such as Clubhouse and X Spaces allow users to create events,^d where they can engage in real-time audio-only conversations with other users. On some platforms, like Clubhouse, the creator of an event is automatically assigned as a moderator and expected to manage the overall progress and health of the conversation. Typical activities a moderator engages in include: setting the agenda for the conversation, managing participation (e.g., granting or taking away speaking privileges), and dealing with conflicts and harms. The event creator can assign others as comoderators. Participants in events might be able to communicate with others through chat messages or speak with others, depending on the settings of features and privileges. It is worth noting that even though this design means most events are often short (i.e., less than two hours), we have observed events that use the help of multiple moderators to carry on for significant amounts of time (e.g., 24 hours).^e

However, not all ephemeral spaces are designed with the same requirement for moderator presence. For

^cThe full list of interview questions is provided in the supplementary material of the published work,¹ available here: <https://dl.acm.org/doi/full/10.1145/3544548.3581329#supplementary-materials>

^dDifferent platforms have different terminology for these spaces, including channels (e.g., Discord), rooms (e.g., Clubhouse), and worlds (e.g., Horizon Worlds).

^eLonger events are achieved through assignment of comoderators, allowing the original event creator to leave the event without ending the event.

example, Horizon Worlds events can take the form of persistent virtual worlds that are accessible to users during all hours of the day and do not require moderators to be present. While these spaces are not consistently moderated, some world creators elect to host events in their spaces. During these events, a team of hosts and moderators is present in the space, managing the event and moderating behavior. Participants can engage in these spaces using verbal contributions, gestures, and movement. Social VR moderators can also choose to remove speaking privileges if they wish to do so. VR spaces can be designed to look like anything (e.g., nature, house, or outer space), but in our prior work, we found that physical obstructions can prevent effective moderation (i.e., by blocking bad behaviors from view).¹

Interactions on these voice and VR platforms are often informal and unscripted.^f Volunteer moderators on the majority of these ephemeral platforms take on the tasks of both content moderation and content creation. This often means that moderators are involved in the content and share their personal opinions and thoughts in the events.¹ The social dynamics of these spaces (e.g., increasing one's number of followers) seem to play a stronger role in motivating users to host or participate in events on some platforms (e.g., Clubhouse) than on others. In cases where higher participation is desired, moderators have been observed to engage in highly contentious topics or encourage conflicts.

Content Moderation

Alongside the prosocial content shared online, some users share harmful, detrimental, or objectionable content or harass others.² A large body of research has investigated the moderation needs and challenges associated with online spaces and offered tools and techniques to address them.^{2,3,4,5} Content moderation, defined as “the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse,”⁶ is conducted by all platforms.⁴ When platforms have limited their moderation, they have faced negative consequences, including removal from app stores, loss of hosting services, restricted payment processing, and advertiser withdrawals, as companies seek to distance themselves from unmoderated harmful content. Platforms often engage in content moderation through a mix of automated detection and flagging of content by users.² Once a potentially harmful piece of content has been identified, a decision is made either automatically or by human moderators, who can inspect the content manually and determine whether it violates the terms of service.⁵

The aforementioned mechanisms of content moderation view it from the “policy and platform” perspective,³ assigning the responsibility of moderation to the platform. Recent work introduced a “communities perspective” as a new view of the task, which focuses on self-organization, through volunteer moderators as a way to address community governance.³ These moderators are members of the community who “are responsible for building communities, often from their inception, and guiding them toward positive cultures of social interaction,”³ often with no monetary gain. Reddit is one of the biggest social media platforms that heavily relies on volunteer moderators for its day-to-day operations. The majority of events on social audio and VR platforms are currently moderated by volunteer moderators.

Users of platforms that incorporate ephemerality engage in real-time interactions through voice and/or VR, and user-generated “content” on such platforms is no longer in the form of static text or images but instead takes the form of real-time conversations or actions. Given these differences, the moderation process on these platforms has also changed. Studies on volunteer moderation in ephemeral spaces have highlighted the need for immediate intervention as moderators must address misconduct in real time to prevent harm and maintain order.¹ This need for immediate intervention arises for two key reasons: 1. because content in ephemeral spaces disappears, volunteer moderators lack evidence of misconduct after it occurs, requiring them to act in real time, and 2. if left unaddressed, harmful behavior can disrupt the remainder of the event.

Identity, Bias, and Psychological Burden in Moderation

Given that human decision making is a major component of the moderation process, it introduces a number of challenges. One of these challenges is that it can lead to considerable psychological burdens for the moderators. Through interviews with Facebook moderators, prior work has shown how repeated exposure to the worst types of conduct made possible through these online spaces can have negative effects on moderators' mental health.⁵

Prior work has also examined how biases in this human judgment could result in unfair decisions with regard to content posted by members of marginalized communities.⁷ For instance, political conservatives, Transgender people, and Black people have been shown to experience content and account removals at higher frequencies compared with others.⁷ While these examinations of how the identity of the user/poster could impact content moderation have been done,⁷ to the best of our knowledge, little attention has been paid to how the identity of the moderator could impact how they are treated and how effectively they are able to conduct their work.

^fIt is worth noting that Elon Musk has hosted interviews with politicians on X spaces before, and these spaces can also be used for work functions.

Since moderators of ephemeral social media interact with other members of the community using voice and their virtual presence, they end up revealing aspects of their identity. Traditionally, in text-based spaces, moderators were able to hide their identities as either they did not need to directly interact with the user or their interactions were done fully through text. An author's gender identity can sometimes be inferred or assumed based on their writing style or word choices,⁸ but these inferences are not always straightforward and can often be minimized through the use of brief or automated responses. However, even in text-based spaces, limited studies have shown examples of how gender could affect how moderators are treated. For instance, a female moderator who worked on the comment section of a website received messages filled with profanity and threats of physical and sexual violence once commenters found out they were a woman.⁵ This study expands on this prior work, exploring how the additional visibility of moderators' identities could impact their work.

Challenges Ahead

Unlike text-based interactions, where identity can often remain less salient, social audio and social VR communication platforms make it harder for users to control which aspects of their identity are disclosed. Voice, avatar choice, and behavioral patterns can be used to make assumptions about different aspects of a user's identity, such as gender, race, disability status, and sexuality. For instance, some social VR platforms allow users to design avatars that use wheelchairs, thus revealing one type of account holder disability. Voice features, such as pitch, can also be used by listeners to make assumptions about speakers' gender, race, or even sexuality.^{9,10}

Self-Presentation and Signaling of Identity

In everyday face-to-face interactions, characteristics such as age, gender, or height are observable, although not always accurately perceived (e.g., estimating someone's age incorrectly). In online spaces, however, users might rely on signals, defined in prior work as "a perceivable action [... that indicates] an otherwise imperceivable quality about the signaler or the signaler's environment,"¹¹ to infer these characteristics. For instance, a user's word choices, voice, or avatar design can signal different aspects of the user's identity. The reliability of signals is proportional to the strength of incentives for honesty, goals behind signaling, and imposed costs. For instance, people might not have much incentive to use deceptive bumper stickers but might instead have an incentive to signal a different gender in online gaming platforms because it might mean they are disturbed less frequently.¹¹ The privacy concerns associated with social video, audio, and VR environments could be an important cost to consider when thinking about

signaling in these online platforms. Sharing one's voice or visual representation online can expose moderators to unwanted scrutiny or identification. This risk is especially pronounced in discussions of stigmatized topics (e.g., mental health) or politically divisive issues, where moderators' identities may make them targets of harassment or discrimination. These concerns could result in additional incentives for unreliable signaling by moderators.

It is worth noting that even in everyday interactions, whether intentionally or habitually, people are putting on performances. Goffman's *The Presentation of Self in Everyday Life*¹⁶ discusses how these performances might be done with the goal of impression management and portraying a specific public image. These concepts of signaling and self-presentation strategies closely relate to topics of identity inference we have discussed, such as avatar selection.

Self-presentation of users in social VR has been explored in the human-computer interaction literature.¹² The moderators, whose experiences were presented in the "Introduction" section, used human-like avatars with features corresponding to their real-world identities. That is not always the case, however. Users have discussed "switching the gender" of their avatar to experience the fun of cross-gender play.¹² The ability to use nonhuman avatars (e.g., animals) is also provided on some social VR platforms (e.g., VRChat), which is being used by some users due to the feelings of freedom and creativity they promote.¹² Prior work has discussed how avatar selection could be platform-specific, reflecting the norms and atmosphere of the platform.¹² While there is some work on the topic of avatar selection, to the best of our knowledge, this work is done from the perspective of everyday users and participants, not moderators. However, moderators face unique challenges related to self-presentation, authority signaling, and identity management. Unlike regular users, moderators must balance approachability with authority, and their avatar choices may influence how they are perceived and treated by participants. Additionally, visibility of identity markers (e.g., gender, race, or cultural cues in avatar design) may shape interactions, potentially exposing moderators to bias, harassment, or challenges to their authority. Future research could explore how moderators navigate these issues and whether specific self-presentation strategies enhance their effectiveness in maintaining order and engagement.

Understanding New Identity-Based Harms

The ability to make assumptions about a moderator's identity can lead to prejudicial treatment and harassment, as demonstrated by the experiences shared by two of the moderators in the "Introduction" section. For instance, we discussed how a female social VR moderator had been sexually assaulted or how a genderfluid

social VR moderator was targeted with hate speech due to their avatar's presentation. While we offered some motivating examples for the potential existence of identity-based discrimination against moderators in social audio and VR platforms, future work should explore this issue in more depth. Future studies could conduct more interviews with moderators to create a taxonomy of the types of harm targeted toward moderators, when these types of harm occur, and how the moderator's conduct could reduce such instances.

Challenges to Authority

Prior work has pointed out how identity factors such as gender identity, race, and ethnicity can result in discriminatory behavior with regard to authority.¹³ This discrimination can occur both on the path to authority and at the point of authority. In other words, minorities are given less access to skill development opportunities (e.g., courses and mentorship) to be able to reach leadership roles and are also supported less when they express interest in applying for promotions (termed "path to authority"). Additionally, research suggests that they are also treated differently when they are given the position of authority (point of authority) and what their experience is like (i.e., how they will be treated and evaluated). Our motivating examples presented in the "Introduction" section reflected the existence of such issues as well. For instance, we discussed how the female Clubhouse moderator had observed that her authority was often undermined by male users simply because of her gender. Prior work has indicated that when leaders are displaying stereotypically masculine behavior, male leaders are evaluated as more effective.¹³ When leaders are displaying stereotypically feminine behaviors, there is mixed evidence on their evaluation, with some suggesting women are favored over male leaders.¹³ Similar differences based on the type of conduct could be present in moderation. Future directions for research include understanding the ways that moderators' authority is challenged, how instances of disregard or dismissal differ based on perceived identity, and how moderators experience and respond to these events.

The backlash effect, defined as experiencing reprisals or backlash for acting in counterstereotypical ways, is another issue discussed in the literature.¹³ For instance, when women assert their agency, they are often perceived as unlikely, which can reduce their chances of being hired for leadership positions.¹³ Backlash effects are observed across political spheres and racial lines.¹³ Future work could investigate whether this effect is experienced by moderators as well.

Challenges to Onboarding and Retention

In the "Challenges to Authority" section, we described how women and racial minorities could face discrimi-

nation on their path to authority. These challenges can include lower chances of being hired for or promoted to positions. Future work could investigate how recruiting decisions of volunteer or commercial content moderators are made and if the challenges suggested by our examples influence the hiring or onboarding process. Additionally, future research could investigate challenges to the retention of hired minorities given the additional chances of harassment they might face in social audio and VR platforms. As one of our motivating examples showed, moderators can be targeted by sexual harassment, hate speech, and other harmful experiences, and when they are targeted, it can interfere with their ability to carry out moderation, as with one individual who shared how they froze and needed to be helped by other moderators. Studies could examine strategies that are implemented or that could be put in place to help relieve some of the stress put on moderators due to attacks or harassment and improve the chances of keeping them on the team for longer.

Anonymization

Anonymization and the ability to make anonymous contributions online could be one of the future directions that are explored with the hope of giving people greater identity-preserving measures.¹⁴ The majority of text-based social media platforms offer this anonymization by allowing users to select usernames that are not associated with their real identity, the ability to have accounts with nonhuman profile pictures, and so on. On social audio or VR platforms, where individuals need to make the majority of their contributions using spoken language, that is no longer the case.

A number of studies have begun looking into voice anonymization techniques for online voice-based discussions.^{14,15} These studies of anonymization have focused on making modifications to the user's audio/voice. It is important to note that the efforts thus far are being explored from the perspective of users and conversation participants, not moderators. Considering the distinct needs and roles of the two groups might reveal the need for different methods of anonymization. For instance, given that moderators on social audio platforms do not only address harms but also facilitate and participate in the conversation, they might want to have more than one persona. They might want to contribute to the conversation using their own voice or an anonymous voice that corresponds with their gender identity. On the other hand, they might prefer to perform moderation duties using an anonymous voice from another gender or with specific features. Thus, one could imagine assigning a moderator two identities, one they could use to contribute to the discussion and another they could use to perform the moderation tasks others might be less receptive to.

It is important to note a considerable concern with such an approach: a concern about the balance between mitigating harm and protecting expression. While allowing moderators to change their voice so it more closely aligns with assumptions about a different gender might reduce harm due to a lack of authority or gender-based harassment, it could also lead to more underrepresentation of minorities as moderators on these platforms.

Teamwork and Human–Artificial Intelligence Cooperation to Address Challenges

Within our results, we showcased that in all three scenarios, delegation and teamwork were at the heart of the solution moderators devised for the challenges they faced. This approach is fundamentally a form of *help seeking*, where moderators reach out to others within their moderation team to address situations they could not resolve alone. Help seeking in this context serves as a strategic response to mitigate the emotional and psychological toll of harassment while ensuring effective community management. Having multiple volunteer moderators within an event who had different genders or sexual identities allowed moderators to respond to the challenge by switching who was responding to participants. Future work can look into the different teamwork strategies employed by moderators to address identity-based harms.

Artificial intelligence (AI)-based automated support could be another potential solution to address these challenges. Future work could explore the development of AI-driven moderation tools that can identify potentially biased or identity-targeted interactions and provide context-sensitive intervention options. For example, automated models could be trained to detect disrespectful tone, discriminatory language, or behaviors specifically aimed at undermining a moderator's authority based on inferred identity markers such as gender, race, or sexuality. These tools could analyze real-time voice data, patterns of interaction, or even avatar-based behaviors to recognize subtle forms of harassment that might otherwise be overlooked. Importantly, AI tools could assist moderators in maintaining control without forcing them to repeatedly engage with hostile or disrespectful individuals, reducing emotional labor and mental strain.

Integrating AI-driven moderation support could serve as a complement to human moderators, empowering them to carry out their roles more effectively while minimizing the risks associated with the visibility of their identities. Future research could examine how such tools can be designed to balance automation with human discretion, ensuring that moderators retain control over final decisions while benefiting from the efficiency and consistency offered by AI systems.

This work discussed the impacts of visible gender, racial, and sexual identity on the effectiveness of volunteer moderators in social audio and social VR platforms. Motivated by interviews with moderators, we hypothesized that visible identity markers can lead to targeted verbal attacks and reduced authority in moderation roles. We then connected these issues to other literature on identity, authority, and aggression. We concluded by discussing future areas of work and potential issues to explore in future work. ■

Acknowledgment

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of California San Diego Institutional Review Board (IRB) under Application No. IRB 804578.

References

1. N. Sabri, B. Chen, A. Teoh, S. P. Dow, K. Vaccaro, and M. Elsherief, "Challenges of moderating social virtual reality," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2023, pp. 1–20.
2. V. U. Gongane, M. V. Munot, and A. D. Anuse, "Detection and moderation of detrimental content on social media platforms: Current status and future directions," *Social Netw. Anal. Mining*, vol. 12, no. 1, 2022, Art. no. 129, doi: [10.1007/s13278-022-00951-3](https://doi.org/10.1007/s13278-022-00951-3).
3. J. Seering, "Reconsidering self-moderation: The role of research in supporting community-based models for online content moderation," *Proc. ACM Human-Comput. Interact.*, vol. 4, no. CSCW2, pp. 1–28, 2020.
4. T. Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT, USA: Yale Univ. Press, 2018.
5. S. T. Roberts, *Behind the Screen*. New Haven, CT, USA: Yale Univ. Press, 2019.
6. J. Grimmelmann, "The virtues of moderation," *Yale J. Law Technol.*, vol. 17, p. 42, Apr. 2015. [Online]. Available: <https://ssrn.com/abstract=2588493>.
7. O. L. Haimson, D. Delmonaco, P. Nie, and A. Wegner, "Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas," *Proc. ACM Human-Comput. Interact.*, vol. 5, no. CSCW2, pp. 1–35, 2021.
8. N. Cheng, R. Chandramouli, and K. P. Subbalakshmi, "Author gender identification from text," *Digital Invest.*, vol. 8, no. 1, pp. 78–88, 2011, doi: [10.1016/j.diin.2011.04.002](https://doi.org/10.1016/j.diin.2011.04.002).
9. S. Sulpizio et al., "The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages," *PLoS One*, vol. 10, no. 7, 2015, Art. no. e0128882, doi: [10.1371/journal.pone.0128882](https://doi.org/10.1371/journal.pone.0128882).
10. J. Kim and C. Davis, "Knowing what to look for: Voice affects face race judgements," *Vis. Cognition*, vol. 18, no. 7, pp. 1017–1033, 2010, doi: [10.1080/13506281003616057](https://doi.org/10.1080/13506281003616057).

11. J. Donath, "Signals in social supernets," *J. Computer-Mediated Commun.*, vol. 13, no. 1, pp. 231–251, Oct. 2007.
12. G. Freeman and D. Maloney, "Body, avatar, and me: The presentation and perception of self in social virtual reality," *Proc. ACM Human-Comput. Interact.*, vol. 4, no. CSCW3, pp. 1–27, 2021.
13. C. Begeny, C. Y. Edwina Wong, T. A. Kirby, and F. Rink, "Gender, race & leadership," in *Oxford Research Encyclopedia of Psychology*, M. Hogg, Ed., London, U.K.: Oxford Univ. Press, 2021, pp. 1–37.
14. W. Kang, M. A. Hughes, and D. Roy, "Anonymization of voices in spaces for civic dialogue: Measuring impact on empathy, trust, and feeling heard," *Proc. ACM Human-Comput. Interact.*, vol. 8, no. CSCW2, pp. 1–22, 2024.
15. T. Simonite, "These deepfake voices can help trans gamers," *WIRED*, Sep. 21, 2021. [Online]. Available: <https://www.wired.com/story/deepfake-voices-help-trans-gamers/>
16. E. Goffman, "The presentation of self in everyday life," in *Social Theory Re-wired*, J. E. Cote and C. D. Levine, Eds., Oxford, U.K.: Routledge, 2023, pp. 450–459.

Nazanin Sabri is a Ph.D. student at the University of California, San Diego, La Jolla, CA 92093 USA. Her research interests include human–computer interactions and

computational social science. Sabri received a Master's of Science in artificial intelligence from the University of Tehran. Contact her at nsabri@ucsd.edu.

Kristen Vaccaro is an assistant professor of computer science and engineering at the University of California, San Diego, La Jolla, CA 92093 USA. Her research interests include how to design the algorithmic experience of machine learning systems, with a particular focus on questions of agency and control and also user understanding, justice, and policy. Vaccaro received a Ph.D. in computer science from the University of Illinois, Urbana-Champaign. Contact her at kvaccaro@ucsd.edu.

Mai ElSherief is an assistant professor of AI and social justice at the Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115 USA. Her research interests include natural language processing and human–computer interaction, specifically, the study of online safety and prosocial behavior, as well as responsible NLP, computational social science, and language and mental health. ElSherief received a Ph.D. in computer science from the University of California, Santa Barbara. Contact her at m.elsherief@northeastern.edu.



IEEE TRANSACTIONS ON BIG DATA

► SUBSCRIBE AND SUBMIT

For more information on paper submission, featured articles, calls for papers, and subscription links visit: www.computer.org/tbd

TBD is financially cosponsored by IEEE Computer Society, IEEE Communications Society, IEEE Computational Intelligence Society, IEEE Sensors Council, IEEE Consumer Electronics Society, IEEE Signal Processing Society, IEEE Systems, Man & Cybernetics Society, IEEE Systems Council, and IEEE Vehicular Technology Society

TBD is technically cosponsored by IEEE Control Systems Society, IEEE Photonics Society, IEEE Engineering in Medicine & Biology Society, IEEE Power & Energy Society, and IEEE Biometrics Council



IEEE
COMPUTER
SOCIETY



IEEE