

Enabling Sensitive Conversations with Consent Boundaries: Moa, a Platform for Discussing PhD Advising Relationships

JANE IM, CISPA Helmholtz Center for Information Security, Germany

KENTARO TOYAMA, University of Michigan School of Information, USA

When an individual is harmed by someone in power, such as a workplace manager, it can help to identify allies—people who would offer advice or supportive action. However, ally discovery is fraught because the very people who might be most relevant—e.g., someone who reports to the same manager—might not be sympathetic and could exacerbate the harm. We examine this problem in the context of PhD students navigating advising challenges and present a social media platform called “Moa” that brings together a number of features for facilitating ally discovery. Moa’s most novel element is an audience selection process that uses what we call *consent boundaries*, which allow users to flexibly define each post or comment’s audience based on factors such as common social identity or lived experience, all while preserving anonymity—neither senders nor recipients learn each other’s identities, even as the post reaches the right audience. A 3-week field study with 47 real-world users showed that Moa’s features in combination facilitated sensitive conversations about advising, with 7 out of 31 posters (22.6%) using consent boundaries. We discuss both our overall “recipe” for systems for ally discovery and the benefits of a consent-centered approach to design.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**.

Additional Key Words and Phrases: Problematic power dynamics, ally discovery, consent, consent boundary, social media, sensitive information.

ACM Reference Format:

Jane Im and Kentaro Toyama. 2026. Enabling Sensitive Conversations with Consent Boundaries: Moa, a Platform for Discussing PhD Advising Relationships. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (April 2026), 33 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 Introduction

Individuals who suffer from problematic power dynamics, such as mistreatment or abuse from someone in power, can benefit from identifying allies, whether it is to seek advice or remedial action. Here, we use the word “ally” to refer to any individual who is sympathetic and can offer some kind of help, ranging from sharing useful information to actively navigating the situation together. Example contexts include workplace bullying [91], political persecution [57], and sexual harassment [23]. In contexts where potential allies are initially unknown, however, someone must make at least a partial disclosure of their experience for what we call *ally discovery*¹ to succeed. But, disclosure is risky because the exact conditions under which these situations arise heighten

¹We have coined this phrase because the relevant literature in political science, collective action, and social movements does not appear to have a phrase representing individuals searching for and finding previously unknown allies. Related terms include “coalition formation” [40] and “boundary activation” [81] but these tend to focus on the building of group solidarity, not the initial person-to-person identification. We note that while the connotations of “allyship” often imply

Authors’ Contact Information: [Jane Im](mailto:jane@cispa.de), im@cispa.de, CISPA Helmholtz Center for Information Security, Germany; [Kentaro Toyama](mailto:toyama@umich.edu), toyama@umich.edu, University of Michigan School of Information, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2026/4-ART

<https://doi.org/XXXXXXXX.XXXXXXX>

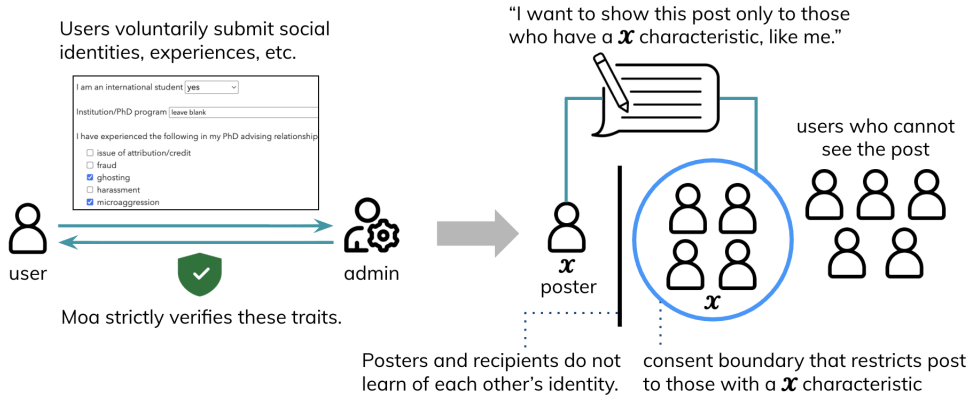


Fig. 1. Moa is a social platform that incorporates a novel concept we call a *consent boundary*—it allows users to demarcate precisely who can view their posts in terms of various contextual information, such as social identity, advising challenges, PhD program, and faculty information, while maintaining anonymity—neither senders nor recipients learn each other’s identities. When a post is created, only users within the specified consent boundary are notified and can view it in their feed.

concerns of exposure and retaliation [3, 21]. That is, one cannot know whom to trust about an issue until one discloses being affected by it; yet one cannot disclose before knowing whom to trust.

Some of the most helpful information in navigating problematic power dynamics can come from individuals who share similar traits regarding the context, e.g., women working with the same manager. How can we facilitate the flow of information in such situations, so that individuals who need it most can discover allies while maintaining one’s anonymity? We consider this problem in the context of PhD students experiencing advising challenges. Because the advising relationship is fraught with power dynamics and ripe for exploitation [14, 47, 54, 80], ally discovery is critical for enabling students to connect with peers who might offer advice or partnership toward remedial action. Yet, research shows that ally discovery within the same PhD program can be difficult [24, 27].

We conducted a three-phase study to build and evaluate a system that facilitates discussions among PhD students to enable ally discovery. In Phase 1, we conducted 19 interviews across nine R1 institutions, and found that participants had a strong desire for a social platform where they can share information about advising dynamics. Many pointed out advising challenges is a sensitive topic that is not easy to talk about, even with peers in the department. Furthermore, participants emphasized wanting to have control over who can see their posts based on a rich set of contextual factors, which existing social media platforms do not provide [42].

In Phase 2, we combined insights from Phase 1 interviews, prior work, and the affirmative consent framework [35], to develop Moa, a fully functioning social media platform.² Beyond features of existing social platforms for sensitive discussion, Moa includes a novel concept we call the *consent boundary*. Existing social media platforms tend to offer brittle controls for audience configuration [42, 93], or they provide mechanisms for selecting individual recipients on a one-on-one basis, which requires knowing users’ identities in advance. In contrast, consent boundaries allow users to demarcate who can view each post or comment based on dimensions such as social identity, lived

support across power asymmetries (e.g., between privileged and marginalized groups), our usage is more general and does not assume differences in power or influence.

²“Moa” is a transliteration of “모아,” which is the conjugated form of the Korean verb “모으다” (to gather).

experience, and interpersonal relationships, while maintaining all users' anonymity—the sender does not need to know the recipients' identities, nor must the sender reveal their own.

In Phase 3, we evaluated Moa in a 3-week field study with 47 PhD students from two departments at a public R1 university. Through system logging, a survey, and interviews, we found that overall, Moa met its goal of enabling ally discovery among PhD students. Out of 47 users, 31 posted or commented on advising-related topics, some of which were sensitive, such as whether to switch advisors. In post-hoc interviews, many commented on their ongoing interest in Moa; a few expressed interest in directly contacting other PhD students on the system or even meeting someone in person.

Out of 31 users who posted or commented, seven (22.6%) used consent boundaries, suggesting that while consent boundaries were not a critical element for ally discovery, they were useful to some. Those users who used consent boundaries had various reasons for doing so—for some, it was to restrict the audience size and reduce privacy risks; others used them to proactively curate an audience (as opposed to defensively setting a boundary) in order to reach those who were sympathetic, could provide useful input, or might find the content helpful.

While Moa is a basic social media prototype, its uniqueness lies in the combination of features designed to enable ally discovery. In all, this work makes the following contributions:

- A recipe of socio-technical features for developing social platforms that facilitate ally discovery, informed by both our study and prior research.
- Conceptualization and instantiation of consent boundaries as a way to address the issue of coarse audience configuration on social media systems.
- Reflections on the values of taking a consent-focused perspective to systems-design.

2 Related Work

We first review research on problematic power dynamics. Because ally discovery requires disclosure for others to take notice, we then discuss literature on audience configuration and identity management—key factors that impact users' self-disclosures. We next give an overview of research on consent. Lastly, we discuss challenges in PhD advising relationships.

2.1 Studying Problematic Power Dynamics or Organizational Justice

Rankism refers to “abusive, discriminatory, and/or exploitative behavior towards people because of their rank in a particular hierarchy” [26]. There is little work that studies how prevalent rankism is across contexts. For example, while research in both North American and European contexts show at least 10–15% of the workforce has been exposed to hostile, aggressive, or abusive behaviors, it is still unclear what percentage of instances are perpetuated by people in power [10, 91].

While not focused on problematic power dynamics in interpersonal relationships, a few HCI systems are relevant for enabling sensitive discussions in organizational contexts. For example, Abdulgalimov et al. [1] built OurVoice, a platform that helps employees discuss issues in the workplace. A deployment study showed that anonymity, strict identity verification, and moderation were important in enabling sensitive discussions about organizational justice (e.g., gender pay) among employees [1]. There has also been research on systems for collective action, such as Dynamo, a platform for Amazon Mechanical Turk workers to organize [71].

Compared to such prior work, we focus on a more interpersonal context of individuals navigating problematic power dynamics within an organization. Research shows that it is extremely hard to be the first person to disclose some kind of information for potential allies to notice in such situations [3]. This is fundamentally related to the difficulty of building trust—trust is about the willingness to be vulnerable with someone in a situation that assumes some kind of risk [56, 73].

2.2 Audience Configuration and Identity Management Online

Audience configuration and identity management are key factors that impact self-disclosures on social media systems [51].

2.2.1 Audience Configuration. While research shows that audience configuration plays a role in sensitive disclosures [51], less work has focused on enabling users to customize their audience more flexibly. Researchers have pointed out that this is likely due to the status quo of how social media platforms are designed, which often provide coarse-grained controls [62, 93].

Social media structure. Fundamentally, the form of a social media system impacts how users configure their audience. Zhang et al.’s Form-From framework [92] classifies three ways social media systems distribute content: *networks*—users create connections (e.g., friending, following) to exchange information, *spaces*—all users within a shared environment (e.g., channels, email threads) see the same information, and *commons*—users can receive content from anyone, without having to follow a user or join a space.

Many studies on audience management on social media have focused primarily on *networks* [8, 15, 84]. The coexistence of different types of social ties (e.g., family, co-workers) on many platforms, which is often called the “context collapse”, makes it difficult for users to carve out a precise audience [17, 55, 82]. A study showed that Facebook users actually tend to underestimate their posts’ audience size [8].

Some platforms that fall under *commons* recommend content and accounts by algorithmically inferring users’ interests, such as Snapchat’s Discover [39, 42]. However, research shows that users perceive them to be still ineffective in identifying the right audience [42].

There is a rich line of work on audience configuration in emails, which has the form of a *space*. Similar to social media platforms, people actively consider social affiliations when configuring an audience using the To and Cc fields [93]. Other studies have examined the impact of Ccing others in workplaces [32, 52, 75], as well as how people decide whether to reply or reply all [28]. However, emails require users to know who their audience is and how to reach them.

In this work, rather than being constrained by a predefined network, space, or algorithmic configuration, we ask: “How does each individual user want to form an audience *per post or comment*, when they do not know the audience’s exact identity?”

Access and privacy controls. Even with the same underlying structure, platforms could differ in respect to what are often called access, privacy, or permission controls [62, 65]. Privacy researchers tend to classify these controls into two categories: interpersonal and institutional [50]. Interpersonal controls regulate how other users access one’s personal information. Settings for audience curation fall under this category, and many general platforms offer coarse options, such as “Public”, “Friends”, and “Close Friends” [50, 87]. Institutional privacy controls manage what platforms collect about users, particularly in the context of online behavioral advertising [36, 50].

While the usable privacy community has advanced the design of access controls for users (e.g., [22, 31, 36, 72, 89]), much of the work tends to focus on increasing the controls’ usability, such as ensuring that they are easy to find [36]. Notably, Feng et al. [22] proposed a design space for privacy controls, introducing five dimensions—type, functionality, timing, channel, and modality.

The social computing community has particularly argued for more radical designs for audience configuration [5, 35, 41, 42, 62, 63]. Most recently, Kim et al. [42] articulated the value of viewing privacy management as a trust-driven process, and suggested concrete designs based on such reframing, such as contextual audience segmentation, intentional engagement signaling, and guided disclosure.

2.2.2 Identity Management. Anonymity is known to help with sensitive disclosures because the lack of identifiable ties lowers risks [51]. Because of this, major social platforms rely on anonymity for posting sensitive information. For example, users actively use throwaway accounts on Reddit [2, 44]. Blind, a platform launched in South Korea, enables anonymous sharing about employers, verifying users via corporate email addresses [86, 90].³

HCI system researchers have explored how anonymity, coupled with moderation, helps with sensitive conversations, along with new identity management mechanisms beyond anonymity. For instance, OurVoice supports sensitive workplace discussions by combining anonymity, strict identity verification, and moderation [1]. Meronymity, a set of traits that provide information about a person’s identity without revealing it completely, helped junior scholars participate in academic discourse on Twitter, mitigating the impact of social hierarchies [76]. The goal of meronymity is to enable users to present themselves more flexibly to give sufficient context to others, while preserving an adequate level of privacy [76].

However, identity management does not fully resolve users’ uncertainty around disclosing information or connecting with people online. For example, research shows that anonymity does not always translate into meaningful social connections, because it does not ensure empathetic responses—an essential factor in sharing sensitive information [51]. This is likely one reason why many discussions using throwaway accounts end as one-off disclosures [44]. And, even for participants who used meronymity, some expressed hesitation or fear of revealing too much about themselves [76], which shows that audience control is a key factor for enabling sensitive discussions.

Overall, no work has yet instantiated fluid and contextual audience configuration on a social platform. While our work started off as developing a system to support sensitive discussions around power dynamics, we address this gap in audience configuration as part of our contributions.

2.3 Consent

In essence, consent is a grant of rights to another person to initiate some kind of interaction. This is especially important for high-risk contexts, such as sexual interactions [59] and medical procedures [43]. For example, in sexual contexts, asking for consent acknowledges that one person can only engage in sex if the other party explicitly grants that right. This is why consent clearly explains why sexual violence is morally wrong—it emphasizes respecting each person’s autonomy and decision-making power [25].

While consent is a powerful concept, defining it has not been straightforward. There have been debates on whether to view consent as a subjective or objective phenomenon [12, 33, 85]. *Subjective consent* is based on one’s internal sense of having granted permission, while *performative consent* is an indication of agreement based on behavioral signs or explicit verbalization [7, 33, 34].

Affirmative consent is a form of consent that values both subjective consent and performative consent. In essence, affirmative consent is an explicit indication of voluntariness and enthusiasm before an interaction is initiated [25]. Building on Una Lee’s work on “consentful technologies” [45], Im et al. [35] argued that affirmative consent is *voluntary, informed, specific, revertible, and unburdensome*, and to develop social computing systems based on these five properties.

In the field of computing, consent is often viewed as a concept for data control within the *notice and choice* model, in which users are given notice about how their information might be collected and used, and offered choices about it [61]. However, many scholars have criticized the model because it has failed to protect people’s privacy in practice [13, 53, 60, 68, 78]. In this work, instead

³<https://www.teamblind.com/>

of being confined to the notice and choice model, we view consent as a fundamental concept that give users the agency to carve out their online interactions.

2.4 Challenges in PhD Advising Relationships

PhD advisors play a significant role in their students' mental health [11, 24, 46, 64] or success in the program [27]. In a worldwide survey by Nature, 23% of PhD student participants answered that they would switch advisors, if given the chance [88]. Another study from a European institution shows that 24% of PhD students perceived they have experienced some kind of abuse from faculty [37]. In HCI, a qualitative study showed that graduate students cite advising relationships as one main cause of stress in graduate school [70]. Other research shows the range of advisor behavior that can cause harm: incompetent advising, sudden dismissal of advisee, forcing advisor's views, exploitation, bullying, encouraging fraud, authorship issues, misappropriating a student's work, harassment, abuse, and dual relationships (e.g., friendship, romantic relationship) [4, 14, 47, 54]. Research also shows that even when PhD students realize they are experiencing harm from their advisor, they have few guidelines for what to do next [6, 27].

3 Phase 1: Formative Study

We conducted two rounds of semi-structured interviews with 19 PhD students in the United States. Our study was exempt-approved by our institution's Institutional Review Board.

3.1 Methodology

3.1.1 Participant Recruitment and Study Protocol. In the first round, we recruited 10 participants from 2 PhD programs within an R1 institution by contacting the student mailing lists. For the second round, we recruited 9 PhD students from 8 R1 institutions by publicizing the study on X and LinkedIn. We noticed that a few participants self-disclosed their experiences to an open-ended question in the sign-up form, and prioritized recruiting them, while accounting for year in PhD and PhD program. Participants were compensated \$20/hour. All participants were in computing-related PhD programs, ranging from small (fewer than 40 students) to large (over 200 students). The average PhD year for first-round participants was 3.8, and 2.9 for the second round.

We showed participants a hypothetical scenario of a PhD student whose advisor persistently requested them to work on non-research tasks (see Supplementary Materials). We asked participants what they would do in the situation. Then, we presented participants with a high-level description of a system that lets students converse regarding advising and asked for their thoughts. Next, we showed participants low-fidelity designs of such a system, and asked them to think aloud. The interviews lasted between 1 to 1.5 hours.

3.1.2 Initial Designs. First, we considered a posting and commenting format (Figure 7). However, the design was not a public forum—we wanted to give users control over each post's audience, and considered two criteria: 1) types of PhD advising-related challenges and 2) faculty names. We also included a messaging feature because we hypothesized it would be helpful for PhD students to privately ask and answer sensitive questions.

3.1.3 Data Management and Analysis. The first author manually corrected the auto-generated transcripts. For sensitive personal experiences, we either redacted it or asked participants about inclusion. The first author conducted inductive coding and periodically discussed the themes with the second author.

3.2 Phase 1 Findings

We report the most salient findings from the Phase 1 interviews.

3.2.1 Needs and challenges in connecting with other PhD students about advising. Almost all of the participants expressed interest in using a system that connects PhD students regarding advising relationships. Even the one participant (P12) who had mixed feelings acknowledged the system's value, though they felt it did not address the root issue of faculty oversight: *"I feel like overall, there just needs to be more oversight [of] faculty and PhD programs. [...] And I think this [system] could be a really great place for students to start doing that bottom-up type of accountability thing."*

When asked what they desired the most, almost all participants answered they wanted *validation* of their experiences. They wanted to know whether their interactions with their advisor were "normal" or problematic. For example, P17 said *"Just knowing that other people's experiences are either the same or different gives you courage to know that it's OK or it's not OK. It helps you figure out what's a normal experience."*

Interviews also revealed why PhD students struggle to seek validation and advice about their advising experiences. Many participants found it was difficult to discuss their experiences with peers or mentors in their established social circles (e.g., research group) due to concerns about power imbalances—a key reason they wanted a system that could facilitate conversations beyond those circles. This contrasts with prior literature emphasizing the role of social ties in sensitive disclosures [51]. For example, some participants avoided discussing their advisor with labmates out of fear of retaliation. P16 explained *"I think it can be hard to talk to other PhD students in your lab or your department because you don't want that information to get out and get back to your advisor."*

Similarly, some participants who had switched advisors avoided discussing their experiences within their close networks due to concerns about repercussions and damage to their reputation. For example, a few participants said they avoided disclosing advising experiences to students in the same program to avoid burning bridges with former advisors. P15 said *"If you've been through a negative advisor experience, I feel like not everyone is willing to be transparent about that with, say, people who reach out to them asking about how their experience in a particular group was so. And it's because we know there are power hierarchies in academia."*

Additionally, some participants noted that even within the same institution, connecting with others who had faced the same kind of challenges could be difficult. P6 said *"...let's say it's the same department, but [the PhD student is in] some other lab, and I've never interacted with the person before. The chances of us connecting over that experience is very low."* Similarly, P13 observed *"So if they [specific PhD students] are not in the exact same scenario, their suggestions would be different."*

3.2.2 Form of the platform: Consent and control in audience setting. Regarding the platform structure, many participants wanted a form that enabled lightweight, distant interactions before mutual trust developed. Broadly, this meant they favored a post-and-comment structure over a messaging system. We initially hypothesized that PhD students would seek deeper connections when supporting one another, such as through a direct messaging. However, most participants did not want dyadic, deep interactions due to the uncertainty of the other person's identity and the context's sensitivity. Others noted that 1:1 messaging created an obligation to reply. As P16 explained: *"I think it would be easy [to write a post], especially in a forum setting where maybe I wasn't like obliged to respond back to people because other people are chiming in as well."*

At the same time, participants wanted something different from a traditional forum—many preferred the ability to set specific criteria for who could access their posts. Participants reacted positively to proposed audience control features and recognized that this structure differed from existing forums (see Section 3.1.2). For instance, P19 said *"Kind of like forum style one can express their what they have to say about the situation, if they meet the criteria."* One participant (P16) explicitly used the word "consent" when describing how the discussions could have some kind of access control: *"I guess people would have to like consent to this, but if you could be directed to*

a group with people that have been having similar issues, and you could see what people have been saying in threads of conversations, and then you could join in.”

Participants’ reasons for wanting granular levels of control included privacy concerns and fear of repercussions. For instance, P11 said: *“I would have to feel more comfortable on the platform without revealing identity stuff. And I would want to know that they’re not like in the same lab as me or share the same advisor, you know.”* Another reason was the importance of gradually developing trust before disclosing more information—something existing forums fail to support because they tend to be public. For example, P12 said *“I think that’s one of the reasons why like platforms like Reddit don’t do so well in terms of forming like strong social support networks because everything is too public, if that makes sense, like you need some sense of like trust and privacy and intimacy to have to start building those relationship blocks to actually support someone.”*

3.2.3 Criteria for deciding whom to interact with. Participants found our proposed criteria of advising-related challenges and faculty names to be helpful (see Section 3.1.2). The interviews also revealed additional criteria that participants suggested.

One important finding was that some participants naturally brought up identities—such as gender, race, and being an international student—without being prompted. (We did not initially include social identities in our list of criteria.) For instance, P16 said *“If I had one [difficulty in advising] with sexism, then maybe I’d want to limit it to like other female identifying students or maybe not men. So yeah, maybe depending on the issue, there would be other like demographic information like that.”* Similarly, P18 highlighted the challenges international students face: *“...as international students, the education system here is so different from the system that we grew up in. We have no idea for resources for harassment, you know, threats or anything.”*

Many participants also wanted the ability to filter by PhD programs or even specific labs (e.g., avoiding labmates). For instance, P19 said *“I really would not want to have that conversation with them [labmates] over this platform [...] But maybe that’s part of the criteria, right, like you say ‘Oh I don’t want anyone else who is being advised by X.’ ”*

Some participants said they are fine with posting publicly. Their main reason was to get advice from a wide audience.

3.2.4 Preference towards anonymous interactions. Nearly all participants preferred anonymous connections. When asked whether they would ever reveal their identity to another user on the platform, most participants said they would not. Many thought anonymity was important to reduce the risks of sharing sensitive information. P19 said *“You can be more open with people if you’re anonymous.”* Others noted that anonymity helps mitigate bias. For instance, P13 said *“...because if I reveal my identity, other people think I’m some kind of vulnerable thing or I cannot progress on my research so that I want to quit.”*

At the same time, participants acknowledged that the desire for anonymity gradually weakens as trust develops. For instance, P19 said *“I don’t think identity has to be all or nothing, it can be parts of it and I would do it as the other person was opening up, like you would actually do, almost in a conversation.”*

3.2.5 Trust between users and the system creators. Many participants asked questions about the system creators’ affiliation and their access to users’ data. For instance, P15 said *“I think the biggest thing I want is some kind of transparent security and privacy labels or like who is going to have access to this database? And I want to know that this is very much like student organized and not something coming from faculty.”* Some participants commented on the importance of knowing the system creators’ motivation. P7 said *“Maybe if there was a picture of you [first author] on the home page and it says yeah, man, like I had some fucked up shit happen to me, that might actually be persuasive*

to me. There are people who made this, not just because they wanted to write a paper, but because of what happened to them in their PhD.”

3.2.6 Preventing abuse. Many participants desired strong abuse prevention measures, so that they would know for sure a user is a student, and not a faculty member. Participants were also aware that students could abuse the system, such as by spreading false information about faculty. For example, P10 said “*Sometimes the abuse can come from the advisee’s side, although that is rare. So, they [PhD students] need to be fair.*” When informed that we plan on moderating the system, many participants emphasized wanting transparency behind moderation [38].

3.2.7 Summary. Based on the interview findings, we organized our design requirements as the following. The platform should:

- REQ1: Follow a post-and-comment structure.
- REQ2: Provide consentful controls for audience setting with respect to various contextual information, such as social identities, advising experiences, affiliation, and advising status.
- REQ3: Ensure interactions remain anonymous; minimize risks of revealing one’s identity.
- REQ4: Have strong privacy and security measures.
- REQ5: Include moderation and community guidelines to prevent potential abuse.

4 Phase 2: Design of Moa

Moa is designed according to insights gained from Phase 1, previously known features that enable sensitive discussion, and the principles of affirmative consent [35]—the value of the last being reinforced by the Phase 1 interviews.

4.1 Overview of Moa

Like many social platforms, on Moa, users can write text-based posts and comment in response to posts (REQ1).

4.1.1 Not Network- or Space-Based; No Profile Page. Social media platforms are typically structured around a predefined network (e.g., followers) or space (e.g., group) [92]. In contrast, Moa focuses on the question: “How does each user want to design their audience per post or comment?” As a result, Moa does not implement any network model nor does it include user profile pages (REQ3).⁴ Furthermore, instead of letting users have access to the same content if they are in the same space, Moa’s users see different sets of posts (and comments) depending on their own and others’ consent boundaries (Section 4.2). In terms of the Form-From matrix [92], we believe Moa is closest to a Threaded Space, but consent boundaries considerably weaken both the threading and the space.

4.1.2 Strict Identity Verification. Because all Phase 1 participants strongly emphasized that they wanted a space of only PhD students, Moa has a strict identity verification process (REQ4). Users can only sign up using an institutional email address; email domains are verified by the system (Supplementary Materials).⁵ Furthermore, the moderator (the first author) reviews every sign-up, verifies email validity, enrollment in one of the two PhD programs, and personal traits (to the extent that can be publicly verified) as entered by the prospective user.

4.1.3 Multiple Usernames. Phase 1 participants emphasized preserving users’ anonymity on Moa. A few participants said they wished there were multiple user identifiers on the system, so that it would be hard for others to infer their identity. Based on this finding and affirmative consent’s

⁴Among existing platforms, Whisper also does not provide profile pages.

⁵For users who did not feel comfortable leaving their institution email address in the database, Moa lets them change to a personal one after signing up.

The screenshot shows the Moa platform interface. At the top, there is a navigation bar with the Moa logo, a user profile icon, and the text "Logged in as phding". Below the navigation bar is a sidebar menu with options: feed, write post, direct message, updates to posts, review your posts, account information & consent boundary settings, and community rules & moderation. The main content area displays a post titled "Asking for tips on how to get more feedback from advisor" by user @po1r3. The post text discusses the user's experience as a junior PhD student and their request for feedback. Below the post, there is a "Reply while making sure to choose your boundary." section. This section includes a text input field for the reply, a "hide panel for customizing consent boundary for this reply" link, and a "Feel free to choose a unique username for the discussion." section with a text input field containing "happyphdstudent" and a confirmation message "This username is available." Below this, there is a "Your reply will be visible to @po1r3 and others who fit your consent boundary." section. This section includes a list of categories for selecting users to see the reply: identity, prior experiences, PhD / faculty, select specific users, and other information. The "prior experiences" category is selected, and a list of specific experiences is shown, including "who experienced change in PhD advising situation", "who experienced any of the following in their advising relationship(s)", and "communication issue". Below this, there is an "Important!" note and a "Set the visibility of this reply's consent boundary." section with a checkbox for "make the consent boundary visible to users who can already see the reply (i.e., fit the boundary criteria)". A "send" button is located at the bottom of the reply section.

Fig. 2. A user (@po1r3) set their post to be visible to those who are international PhD students, not advised by John Smith, and have experienced communication issues or lack of feedback. A user (@phding) is about to reply while creating a different username (“happyphdstudent”). @phding can see the post’s consent boundary because @po1r3 opted in to showing it to those who can already see the post.

specific principle [35], we let users create different usernames per post or discussion (REQ3; Figures 2). Once a user takes a username, it cannot be used by others, even if they are in a different thread. Compared to other platforms, Moa provides this feature directly in the posting and commenting interface (*unburdensome*), which is not the default for major social media platforms.⁶

To prevent users from using multiple usernames to add weight to their point, once a user starts or participates in a thread, they can only use the name that they initially used. The username field becomes deactivated, so that the user can no longer make modifications to it.

4.1.4 Community Rules and Moderation. Based on the Phase 1 findings and prior research [1, 74], Moa emphasizes community rules on a linked page from the navigation bar (REQ5), which are: 1) Be respectful and kind to others, 2) What is said on Moa, stays on Moa, 3) Do not spread false information, 4) Do not use discriminatory language, 5) Do not harass or attack others. Moa is

⁶On Blind, a user has to go to their account setting to change their username. Since August 2023, Blind allows users to change their username once per day. Previously, it was allowed 5 times per month. <https://www.teambblind.com/post/New-More-username-changes-7CXLWNB4>

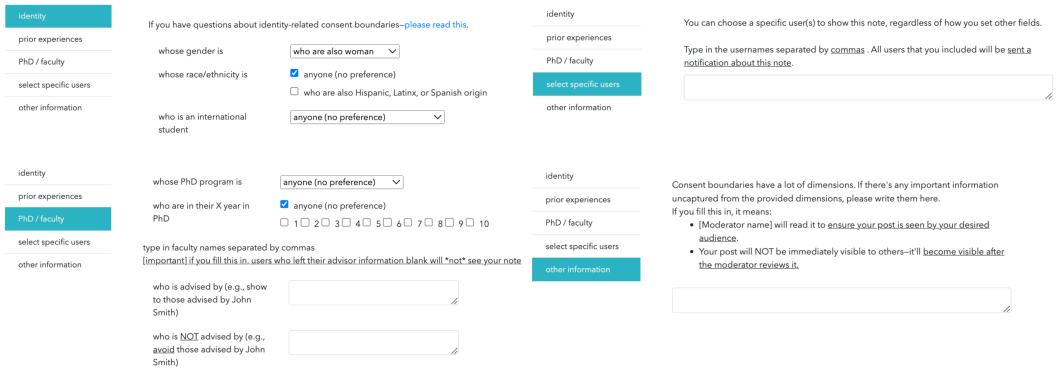


Fig. 3. Different dimensions of consent boundary. The high-level categories included: identity, challenges experienced in advising relationships (Figure 2), PhD program/faculty-related factors, specific users, and other information. We explain each category in Section 4.2.2.

moderated by the first author because adding others would heighten users’ privacy concerns. However, theoretically, Moa could be extended to recruit other moderators.

4.1.5 Notifications. Whenever a post is created, users within the post’s consent boundary are notified via email. Moa also notifies all users in a thread whenever a user adds a new comment. For both, the email shows the first 20 words of the post/comment.

4.2 Consent Boundaries

Moa lets users mark their *consent boundary*, which means users can set an audience per post or comment based on a rich set of dimensions (REQ2; *voluntary, specific*).

4.2.1 Flow of setting a consent boundary. A user can set a consent boundary while writing a post (or comment) and save it to their account setting (Figure 2). Or, a user can set a default, account-level consent boundary in the settings page. Whenever a user sets a consent boundary for a post/comment, it does *not* affect prior ones (*specific*). When a user visits a post they wrote (or a thread they participated in), Moa pulls up the recently used boundary for that post (*unburdensome*).

4.2.2 Dimensions of consent boundary. Below, we list possible consent boundary dimensions.

Identity. Moa lets users show their post or comment to others who share their gender, race, or international student status. We enabled users to select each category *only* when they have the corresponding identity.⁷ For example, only women can set their consent boundary so that their posts and comments are visible to women. We describe this more in Section 4.2.3.

Prior advising experiences. Moa lets users show their post/comment to others who have experienced specific kinds of advising challenges (e.g., micromanagement, communication issue; Figure 2). The categories of challenges are based on research on PhD advising [14, 47] and Phase 1 interviews (full list included in Supplementary Materials). We also linked a survey in the interface so that users could suggest new categories they desired (*voluntary*). Users can also only show their post/comment to others who have experienced a change in their advising situation (Figure 2).

⁷We note that was a deliberate design choice among several options. For example, a platform could let users set boundaries that they are not themselves within, perhaps to seek opinions from groups they are not members of. However, that can open the door to misuse without additional safeguards, so for this work, we opted for the more conservative option.

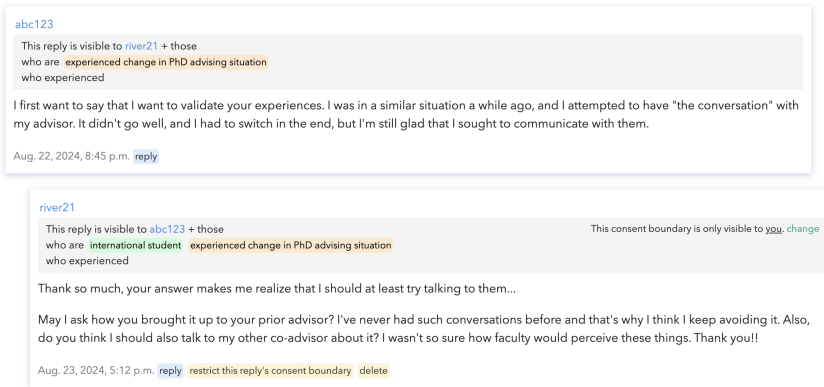


Fig. 4. Users can also set a consent boundary for each comment. @abc123 shows their comment to users who have experienced a change in PhD advising, while @river21 limits their reply to @abc123 and international PhD students who experienced a change in advising situation. At the top left of @river21’s comment, it says that the consent boundary is only visible to @river21.

PhD program and faculty names. A user can make their post or comment only visible to PhD students from a specific PhD program (Figure 3). Moa also lets users avoid, or connect to, other users who are advised by certain faculty (Figure 3).

Specific usernames. Moa lets users choose accounts they only want to show a post or comment to (Figure 3). For comments, users are only able to select users who participated in the thread.

Other information. Users can type in information uncaptured by provided dimensions in the “other information” textbox (Figure 3). When a user uses this, the post or comment is delayed until the moderator reviews it and chooses the right audience.

4.2.3 Declaration of identities and experiences. To make consent boundaries work, Moa has to know users’ identity traits and experiences. Users can voluntarily declare their traits for *only* the system to know (*voluntary*). Users are encouraged to submit information when signing up, but they can go to their account page to change it (*reversible*). To ensure a user does not falsify their account information after sign-up, the first author periodically checked for users’ account setting changes.

4.2.4 Consensually visualizing consent boundaries. Moa lets users decide whether to show their consent boundary to others *who can already see the post or comment* (*voluntary*). That is, by default, when users see a post with a boundary applied, it looks the same as public posts. For example, in Figure 2, the poster consented to showing their boundary. Those who have access to the post can understand that the poster is an international student, has experienced communication issues and lack of feedback from their advisor, and wants to avoid students who are advised by John Smith. In contrast, if a user decides to not show their boundary, other users who have access can only see the content and username. Users can also change their boundary’s visibility after posting or commenting (*reversible*; Figure 4).

4.2.5 Monotonically decreasing branches. Moa applies the original post’s boundary to all of its comments. This is because research has shown that replies to a post can reveal its content, even if the post is deleted [58]. A comment to another comment also inherits their boundary rules.

4.2.6 Restricting a consent boundary. Based on the affirmative consent's *reversible* principle, Moa lets users *restrict* consent boundaries (Supplementary Materials). Users who already replied to the post but no longer fit the new boundary will lose access to it (without any notification).

4.2.7 Not applying consent boundary rules to the user whom one is replying to. There was one contradiction between the affirmative consent framework and how Moa's consent boundary was implemented. The framework is conservative and implies that consent boundary rules should be enforced without exception [35], including when a user replies to a post or comment. For example, let's say *A* is a woman. *A* set their default consent boundary so that their comments are visible to women. *A* is about to reply to *B*'s post. Based on the affirmative consent framework, Moa should enforce *A*'s consent boundary to *B* as well. This means *B* cannot see *A*'s comment if they did not indicate they are a woman in the account setting.

However, our pilot studies showed that this could be confusing. Thus, we changed it so that when a user leaves a reply, by default, Moa assumes they want to show their reply to the original poster. Users can always change the setting (see the bottom of Figure 2).

4.3 System Implementation and Pilot Studies

We built Moa using Django, MySQL, JavaScript, HTML, and CSS. We used Fly.io⁸ for deployment. For email notifications, we used SendGrid and configured it in Django's settings.py. To ensure consent boundaries work robustly, the first author created test cases and conducted pilot studies.

4.3.1 Pilot User Studies. We conducted pilot user studies with five participants, who were PhD students and a postdoc who recently graduated. All participants were compensated \$20 per hour.

Improving the usability of consent boundary. Participants found the consent boundary interface overwhelming because there were too many fields. Thus, instead of showing everything at once, we let users select a tab and only showed the relevant boundary dimensions (Figure 3).

Making it clear what consent boundary features do. We noticed that it took time for some participants to grasp the connection between their account settings and the visibility of posts and comments. For example, a woman participant did not understand that if they left their user information blank, they could not see posts from women who used a consent boundary of restricting their content to only women. Thus, we decided to include a short tutorial in Moa's sign-up and account settings interfaces, which explains how account pages affect what users can see.

Emphasizing that Moa was built by a PhD student. Some participants emphasized that knowing Moa was built by a student was important for them. We thus highlighted this on a separate page.

5 Phase 3: Field Study

To evaluate Moa, we conducted a field study for almost 3 weeks with 47 PhD students in two computing-related PhD programs at a public R1 university.⁹ This study was exempt-approved by our institution's Institutional Review Board.

5.1 Methodology

We sent recruitment emails about Moa and the field study to all PhD students in the two PhD programs. The study was also publicized on PhD-student-only Slack channels. After participants gave their consent, successfully verified their identity, and signed up, they were sent an email that their sign-up was approved, and to start using Moa after reading a tutorial about Moa.

⁸<https://fly.io/>

⁹One PhD program has over 100 PhD students, and the other has over 200 PhD students.

We wanted participants to use Moa in a natural setting. Participants were asked to log in at least twice per day, but this was a recommendation. Participants who completed the post-study survey were compensated \$10. Those who participated in interviews were compensated \$30/hour. Among participants who posted or commented at least once, five were selected to receive \$100. The likelihood of winning increased based on number of posts and comments.

5.1.1 Addressing the “cold start” problem. A major challenge was attracting users to a new online space [69]. To tackle this, the first author made 4 posts as seed content based on their experiences, and then nudged 5 early users to comment. Participants were not aware that the seed content were posted by the first author, which was an approach used by Reddit’s founders [29]. Once we recruited around 20 participants, we sent two follow-up emails about user activity and the first author’s motivation for developing the system.

5.1.2 Data collection and analysis. After a week of using Moa, users were asked to complete a survey and invited to an interview. 31 participants filled out the survey, and we interviewed 14 of them. We also collected and analyzed log data, with users’ consent.

Follow-up survey and interviews. Using Likert-scale questions, we asked participants how they perceived their experience on Moa and the concept of consent boundaries. We also included one open-ended question about what, if anything, the participant wanted to change or keep about Moa. Lastly, we asked if participants were interested in a follow-up interview.

Half of the interview participants used consent boundaries (7/14; Table 2). 9 out of 14 participants posted at least once. Of the 5 participants who did not post, 4 commented at least once. All interview participants gave consent to the first author to access their posts and comments.

We asked interview participants how they decided to use, or not use, consent boundaries. For those who did not use consent boundaries but still posted or commented, we aimed to understand what enabled them to do so. The interviews lasted between 20 minutes and 45 minutes.

The first author conducted deductive and inductive coding of interview transcripts and answers to the survey’s open-ended question. During the process, the first author periodically discussed the themes with the second author.

Log data. We also analyzed the data of posts and comments, along with user activity. This was communicated to the participants in advance of the study, via the consent form and Moa’s interface.

5.1.3 Limitations. Our field study’s limitation is that we focused on the context of the United States. Future work should explore how the findings from this work could be applied to other regions’ contexts. Another limitation is that we deployed Moa in two PhD programs. In the future, we aim to conduct a longitudinal study on a larger scale to report more robust findings.

5.2 Ethics Statement

We sought to make the terms of using Moa very clear to all users, given that our study setting is very sensitive. We indicated both on the platform and in recruiting emails that the moderator (first author) had access to all content and meta-data. No posts were excerpted verbatim from the paper, and all references to posts were paraphrased. We made it clear to users that the moderator was monitoring the platform on the backend in case harassment or trolling occurred.

We also took care in deciding methods for user disclosure and use of personal background and identity. We designed our system based on one set of best practices in consultation with several peers with diverse gender identities. No participant was forced to disclose any part of their identity. And, we allowed users to self-describe any consent boundaries as they felt were relevant in an open-ended text box (Figure 3).

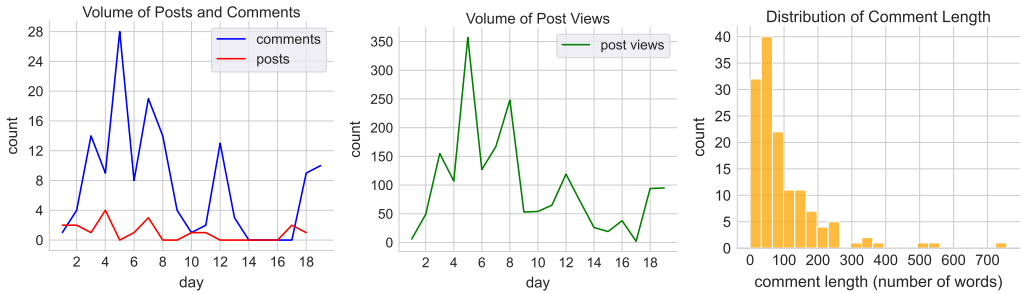


Fig. 5. The first graph shows Moa’s daily post and comment volume, the second shows daily post views, and the last shows comment length distribution. On average, posts had 7.7 comments, with a maximum of 17. The daily post views averaged 98.2. Comments averaged 98.5 words, with the longest at 757 words.

5.3 Phase 3 Findings

First, we report user activity on Moa and how users perceived the platform. Then, we discuss findings on consent boundaries and on features that enabled users’ participation.

5.3.1 Moa enabled a range of real conversations about PhD advising relationships. Over 19 days, a total of 47 PhD students created accounts and logged onto Moa at least once. All users submitted at least some kind of user information when signing up, and most did not make changes to it.¹⁰ 72.3% of the users submitted their current advisors’ names and 21.3% submitted their prior advisors’ names. 36.2% indicated that they were from PhD program A and 55.3% indicated they were from program B. We report participant demographics in Appendix.

15 users created 18 posts and 31 users left a total of 139 comments¹¹ and users viewed posts on Moa 1,866 times (Figure 5). Moa’s users overall engaged with each other. On average, posts had 7.7 comments, and at most, one post received 17 comments. The comments had 98.5 words on average (Figure 5). Posts that asked for advice about their advising situation especially received comments with detailed advice—the longest was 757 words. Users tended to thank each other for the advice.

With respect to the kinds of interactions that occurred, among the 18 original posts, 9 asked for advice (see Table 1).¹² These included a range of sensitive posts where users asked for advice. The most sensitive posts were about seeking guidance on whether to switch advisors or add a co-advisor, as well as dealing with mental health struggles after switching advisors due to harm from a prior advisor. Other topics also included navigating gender dynamics with another advisee, seeking to diagnose reasons behind communication issues, how to cope when advisor is not receptive to feedback, and how to better work with their advisor given their mentoring style. In three posts, some users proactively shared personal experiences, such as how they switched advisors.

Overall, users’ reactions showed that they received real benefit from Moa. For example, P14 reached out to the first author on Slack saying “*Thank you for making this system. I was in deep need for it.*” In interviews, participants often commented that Moa enabled them to learn what good advising looks like and what “is not normal.” Survey results (Figure 6) also showed that many participants strongly agreed that Moa made it easier to have conversations around PhD advising relationships (average=4.45; median=5).

¹⁰8 users (17%) submitted more information afterwards, and all updates seemed valid (i.e., users did not falsify information). The updates consisted of submitting advisor name, advising dynamic experiences, whether one is an international student, gender, race, year in PhD, and PhD program.

¹¹These numbers exclude the seed content the first author created.

¹²Four posts were deleted; we did not include these in the final dataset.

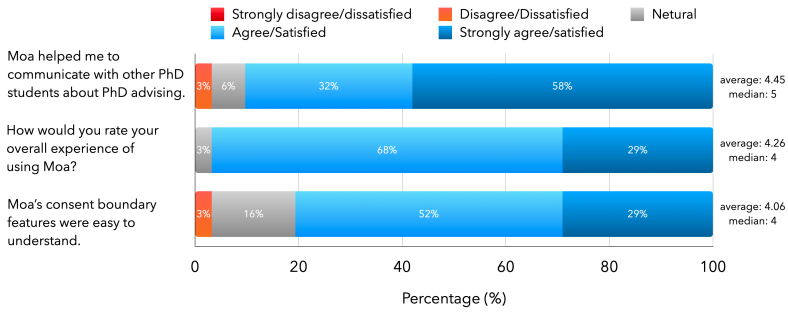


Fig. 6. Results of the post-study survey (1=Strongly disagree/dissatisfied; 5=Strongly agree/satisfied). Participants strongly agreed that Moa made it easier to have conversations around PhD advising (average=4.45; median=5) and were satisfied with the overall experience (average=4.26; median=4). Many found consent boundaries easy to understand (average=4.06; median=4).

	# of posts
Asking for advice about PhD advising relationships	9 (50%)
Sharing prior experiences	3 (16.7%)
Sharing positive stories	2 (11.1%)
Sharing high-level opinion about PhD advising	1 (5.6%)
Discussing PhD-related topics in general	3 (16.7%)

Table 1. Number of posts per topic. Out of 18 posts, half were seeking advice on PhD advising relationships.

	Activity on Moa	Consent Boundaries Used	Set Default Consent Boundary
P1	posted/commented	advising challenge	no
P2	posted/commented	none	-
P3	commented	none	-
P4	posted/commented	none	-
P5	posted/commented	advising challenge	no
P6	commented	advising challenge, advising status, race	yes + customized per comment
P7	posted/commented	PhD program	no
P8	none	none	-
P9	posted/commented	none	-
P10	posted/commented	none	-
P11	commented	gender	yes
P12	commented	advising challenge	no
P13	posted/commented	gender	no
P14	posted/commented	none	-

Table 2. Overview of users who participated in the field study, post-study survey, and interviews. The first column shows whether each participant posted or/and commented. The second column shows whether/how each participant used the consent boundary feature. For those who used consent boundaries at least once, the third column shows whether a participant has set their default, account-level consent boundary. “yes” means the user has set at least one boundary to a non-default value. “no” means everything was at default value. We put “-” for users who did not end up using consent boundaries.

5.3.2 Consent boundaries were used by about a quarter of posters/commenters (N=7/31), who tended to set them per post or comment. 7 out of 31 users who posted or/and commented (22.6%) set consent boundaries for at least one post or comment (Table 2). 3 out of 7 users used it for posting and 4 used it when commenting. For the three posts, they were sensitive in nature, although the degree of sensitiveness differed—one was about their advisor not being receptive to feedback, another was about gender dynamics in their advisor’s group, and the last one shared their experience of switching advisors. In the comments, users asked questions, shared one’s experience (e.g., having difficult conversations with one’s prior advisor), and expressed agreement or sympathy.

Contrary to what we expected, almost all users set a *distinct* consent boundary *per comment or post*. The majority (5/7) did not touch their account-level consent boundary at all, and only P11 used their account boundary setting as it is (Table 2). Users mentioned they found the consent boundary being “right there” and found it easy to apply them for each comment. More interestingly, some interview participants compared consent boundaries against existing privacy settings, which they noted were binary (public versus private). In contrast, participants thought consent boundaries are more contextual and thus wanted to apply them per post or comment.

“I kind of felt like I have so many different types of experiences that are for different groups that I don’t want to put something and then have to go back and undo it. So I felt like per post was nice. That way I can choose who sees what when I make the post...”-P13

The survey results showed that most users found the consent boundary features easy to understand (Figure 6). However, three out of the 31 survey participants commented that they found consent boundaries difficult to understand. Some interview participants also wished the feature was less overwhelming, although they understood why we included each dimension.

Restricting boundaries. The feature for restricting consent boundaries was not used frequently, although one participant used it twice. P6’s comment was public at first, but they later made it visible to users who are Asian. Then, they updated the boundary so that the comment is visible to those who are Asian *and* have switched advisors.

5.3.3 Consent boundaries were used for not just defensive boundary-setting but also proactive audience curation. The interviews showed that users tended to use consent boundaries differently. For some, consent boundaries were used to filter out specific kinds of users for privacy; for others, they served as a mechanism to proactively curate an audience for various reasons.

Specifically, three participants (3/7) used consent boundaries to protect their privacy by not making their post public. These participants posted or commented to ask or share personal experiences about their advising relationship, but did not want to increase the chances of revealing their identity—not just because of their privacy, but also because of their advisor’s. For example, P5, who used boundaries about certain kinds of PhD advising dynamics when posting a question, said:

“I wasn’t sure if I wanted everybody on on the platform to be able to see the post, mostly because it is kind of like a personal issue. And then I was also worried that people could identify me from it, and honestly, I think I didn’t want it to reflect badly on my advisor.”-P5

Interestingly, the rest of the interview participants who used consent boundaries viewed it as a way to *proactively reach* certain groups of people, rather than preventing some people from viewing their post. Their reasons included wanting to reach people who would better understand them, provide accurate responses, or find the content helpful. For example, P1 used boundaries related to certain kinds of advising challenges. To them, privacy concerns were not a reason for using consent boundaries, because they refrained from writing sensitive content. But, P1 still wanted to interact with people who could provide the most accurate insights.

“ So I think I mostly used it [consent boundaries] to try to make sure that my comment reached the people that I wanted it to reach as opposed to avoiding people I didn’t want it to reach, if that makes sense.”-P1

In particular, two consent boundary setters emphasized wanting to interact with those who could understand them. For example, P6 wanted to help other PhD students by sharing their experiences, but in a way that minimizes chances of getting hurt by reactions from people who cannot understand them. This participant was the most active in using different kinds of boundaries. P6 explicitly said it was due to the desire to have an audience who could understand them, rather than reducing privacy risks.

“I think privacy in this specific experience might be less relevant because I feel like whatever consent boundary I set, I can be identified by sharing my experiences. [...] Or another way to put this is that I don’t expect or really want to be understood by all people.”-P6

5.3.4 Some participants found value in showing consent boundaries, while others chose to not make them visible. Users showed mixed behaviors when deciding whether to display consent boundaries on posts or comments. Of the seven users who used consent boundaries, four displayed them, two kept them entirely invisible, and one alternated between showing and hiding their boundary.

Those who used *and* showed consent boundaries thought there was value in signaling whether a post was a “closed thread” for specific kinds of users. They noted that without visible boundaries, others might assume the post is public or be unsure about its visibility, making them hesitant to respond. For example, P5, used and showed boundaries of limiting their post to those who have experienced specific advising challenges, thinking that it would ease users to respond, since their comments will be viewed by those who had the same experience.

“I sort of just wanted to give some transparency to anyone who felt like they want to respond to this with a personal story. [...] I felt like if I were to set the standard in the first post, then maybe people would be more comfortable knowing that it had a limited audience to begin with.” -P5

In contrast, some tended to think that once the intended audience sees the post or comment, then the consent boundary itself is unnecessary information. That is, they used consent boundaries for audience selection, rather than for showing bits of their identity. P6, who was the most active in using consent boundaries, said:

“I feel like in reality, when a person sees a post or a reply, whatever the consent boundary is visible to them, doesn’t change the fact that they can see this reply. So I think [shown consent boundaries] is unnecessary information from that point.” -P6

Finally, a few interview participants did not want to show their consent boundaries because they thought it was a private information they wanted to keep to themselves.

Some participants thought seeing consent boundaries was helpful. Some interview participants thought seeing consent boundaries was helpful, because it provides more context on why the poster is reaching out to a specific group of users. For example, P11 noted that seeing a women-only boundary on a post about gender dynamics made them want to engage with it.

“Like since they are showing a woman-only boundary, I will be very likely to respond to this post in a very, very honest way and support this woman student.”-P11

A few participants also raised ideas on making it possible for users to share parts of their identities in comments, regardless of setting a particular audience. For example, P13 said that this can be

useful when a user finds a post that resonates with them because it is about a similar experience they have had, and wants to signal the connection to the poster.

5.3.5 Reasons for not using consent boundaries. Most posters and commenters did not ultimately set consent boundaries, and they also cited several reasons for not doing so.

Wanting to have a broad audience. One major reason was that they wanted to enable their post to reach the broadest audience, and thus made their content public. For example, P4, who publicly wrote about advice regarding advising relationships, said they wanted it to reach a lot of people. This was also the case for three interview participants who asked questions about the advising challenges they faced. These participants, however, were not posting publicly out of comfort. They disguised their content—e.g., by altering the gender of their advisor—so that readers could not identify them, especially given that they are from the same institution or program.

“I intended that (i.e., set their post to public) because I want to get advice from everyone who can be related and I don’t know who they are.[...] Moa’s users are highly related to me, so they can give me more direct and valuable advice. But at the same time, it makes me conceal myself more because I feel like they can easily guess it’s me.” -P10

Anonymity and PhD student verification were enough. For some users, anonymity through usernames, combined with the strict identity verification imposed by the system, were sufficient.

“For me personally, I’m an international student and I am kind of in fear of authority, but at the same time, I kind of trust you. Everyone on here is a student, so I don’t mind if everyone is a student, I don’t think they will do anything.”-P3

The multiple username feature was popular—20 users (64.5% of posters/commenters) created 54 distinct usernames, aside from their default username, when posting or commenting (average=2.7). 23.4% of all users changed their default usernames at least once, as well. The interviews and survey showed that this feature helped users feel reassured that others would not learn their identity.

System creators’ and other users’ credibility also helped. Even for participants who did not know the first author well, Moa being developed by researchers in the same institution and detailed explanations of how we treat users’ consent increased Moa’s credibility. The first author’s revelation (in a recruiting email) of having switched advisors also helped. For example, P13 said:

“I think one was how much you were respecting our anonymity and our consent boundaries and things like that. I kind of felt it empowered me to post about things that maybe I wouldn’t have otherwise. [...] The description was part of it, where it was like, ‘We’re making this platform for this purpose, and it’s made by people who have had experiences that they’ve wanted to talk to other people about.’”-P13

Users’ credibility also incentivized some users to post. Interview participants said they could get more trustworthy advice because they knew the users were from the same institution, and not just a random person on the internet. For example, when asked what enabled them to post, P14, whose post asked for advice about navigating advising challenges, said:

“I think several factors. One is, well, the first thing I can seek advice. The second thing I was looking for something where I am anonymous. On social media and everything you’re mostly known, or if I go to my lab mates or anything they know who I am. The third thing is I know it’s within [institution name], so I’m sure where they are but I’m not sure who they are. It’s like I’m sure the people who reply to me are actually PhD students.”-P14

Insufficient trust in system. Some participants refrained from writing sensitive content (and did not use consent boundaries) because they needed more time to build trust in the system.

5.3.6 Signs of users wanting closer interactions with each other. A few interview participants mentioned the possibility of directly reaching out to other users on Moa or wanting to meet in-person. That users wanted closer interactions was surprising to us, because it contradicted what the majority of our Phase 1 participants said.¹³ For example, a few Phase 3 interview participants said they wished Moa had a chat interface so they could more easily message others, while another said there was one particular user who wrote a post that made them want to connect in-person. These show that through Moa, some users developed trust in others, to the extent of wanting to take the risk to further interact with them.

“I guess if I know that only people with that consent boundary are now seeing the post and they replied, after the initial post, I would rather have a conversation with them. [...] I believe another feature that could help for very sensitive situations where I’m really not comfortable sharing it as a post would be having some kind of a private messaging.” -P4

When asked why, participants mentioned wanting to get more advice because they desperately needed to resolve their advising situation, and to obtain accurate signals about a person [20].

“There is a really impactful post [on Moa]. If I see an impactful post [like that], then I really want to reach out and get advice from them. [...] I want to meet in-person, and talk sincerely because their post is very sincere, and I’m touched by that.” -P10

It seems that while some risk-averse participants considered Moa to be useful, they believed that giving or getting more sensitive information should happen outside the platform, with discussions on Moa being a basis for such connections. For example, P1 pointed out that Moa could be a starting point for in-person meet-ups.

“If there was like, okay, based on all of these topics that we’ve discussed, there’s going to be these in-person meetings based on these conversations, and we’ll sit there and have a safe space of being able to talk in a non-technology facilitated way, but using this [Moa] as a way to filter what those topics would be, that’s something I would probably use more.” -P1

5.3.7 Ways to improve Moa. We describe the most interesting feedback on improving Moa.

Using Moa to help faculty learn about mentoring. One interview participant said while they were not so sure how it should happen, they thought it would be worthwhile to invite faculty to Moa, as they could gain exposure to students’ experiences and learn about mentoring.

Expanding the user pool while providing audience control via consent boundaries. Some interview participants wished there were more users, and not just from two PhD programs. However, even these participants perceived benefits of interacting only with those from the same program, and said the consent boundary could be a way to easily show their post to them.

Making consent boundaries more usable. Multiple interview participants wished the consent boundary interface was more usable. One participant suggested a feature that gives users a sense of the audience size when applying a consent boundary. Related to the flexibility of applying different consent boundaries, one participant recommended using a computational modeling approach so that the system could recommend consent boundary settings based on each post.

5.3.8 Summary of Results. Moa enabled a range of sensitive discussions, with a quarter of participants using consent boundaries—typically a separate boundary per post or comment. Users tended to use consent boundaries differently—some used them to simply remove users to reduce privacy risks; others used them for proactive audience curation, to reach those who are sympathetic, could

¹³Many Phase 1 participants did not want conversation on Moa to be too private and some also emphasized that they would never reveal their identity (Section 3.2.4).

provide useful input, or find the content helpful. Some participants valued displaying their consent boundaries, while others chose not to, because they used them solely for audience configuration. Finally, a few participants expressed interest in directly contacting another user on Moa or meeting them in person, which was surprising compared to the results from Phase 1.

6 Discussion and Future Work

6.1 Recipe for Enabling Sensitive Conversations and Ally Discovery Online

Combined, the existing literature on enabling sensitive conversations makes many recommendations, but they are scattered across multiple papers [1, 44, 51]. Here, we pull them together to provide a state-of-the-art, socio-technical “recipe” for such systems. We present it in two layers: One layer mentions the high-level objectives sought (indicated by the boldface headings below); the second contains specific features that enable elements of the first layer (boldface, in parentheses).

Anonymity (pseudonyms, multiple usernames, no profile pages). As prior literature shows [1, 51], and our Phase 1 study confirmed, providing users anonymity is important. Complete anonymity, however, can make discussion threads difficult to follow. A good compromise is to allow users to create pseudonymous usernames—many of Moa’s users created multiple pseudonyms. Avoiding profile pages, as Whisper does, also enhances anonymity. None of our field study participants said they wanted a profile page, and one participant explicitly said they liked not having them.

Credibility of system (baseline system security, sympathetic sponsors, consent emphasis). System credibility requires technical privacy and security protections as a baseline, and Moa assured users of that, but there is also a social element. Confirming findings from online systems in general [49], Phase 1 study showed that they would use Moa more if it were run by people they trusted, especially a PhD student peer. A few participants of Phase 3 also noted that Moa’s emphasis on consent was intriguing to them, and enhanced the system’s credibility.

Credibility of users (strict identity verification, affiliated user base, moderation). Credibility of users is also critical, as prior work has noted [1]. One important factor is for the user base to be limited to some sort of affinity group. In Moa’s case, all users were PhD students at the same university. Phase 3 participants noted that one reason they found Moa more useful than, say, Reddit, was because they could expect more relevant responses from peers who shared a lot of context. We also believe that it was easier to trust other users because it was a limited, familiar, and moderated community. In the future, we anticipate Moa could be opened to various institutions, but users would still be able to post to others within the same institution using consent boundaries.

Engaged, empathetic atmosphere (seed content, community rules, in-group moderation). An engaged atmosphere is essential for any thriving online community [69], but for sensitive discussions, the culture must additionally be empathetic. To enable this, we believe appropriate seed content, community rules, and moderation are critical, especially at the beginning. Because many Phase 1 participants said they would lurk first before deciding they could trust the system, the first author seeded posts based on their personal experience. Two Phase 3 interview participants noted that one of the seed posts was memorable. Furthermore, Moa’s community rules emphasized Moa being an supportive community for PhD students. And, we believe in-group moderation is essential. Moa’s users seemed comfortable with moderation, likely because Moa’s moderator was a member of the user base. This is in contrast with prior work, in which participants raised questions about moderation, because an external research team handled it for the study [1].

Granular control over post visibility (consent boundaries). For users who want to discuss topics or share experiences that are highly sensitive, granular control over who can see their posts

seems to increase chances of engagement. For Moa, this was enabled by consent boundaries. The Phase 3 evaluation showed that consent boundaries lowered the barrier to posting for some users, especially for risk-averse ones. Even for users who had less privacy concerns, consent boundaries still provided benefits of reaching out to the right audience. We also note that there was arguably an implicit consent boundary of all users being PhD students from the same institution. Furthermore, the users also consented to the moderator seeing their posts, which could be seen as another implicit consent boundary due to the study design. These could explain why the consent boundary's usage rate was 22.6%—the default, implicit consent boundaries already provided an audience that made most users comfortable enough to post.

Feature visibility (transparent explanations, repeat feature reminders). Finally, it is important for all of the above to be visible to users. Messaging on the system should highlight key system features repeatedly. In Moa's case, the goals and features of the platform were noted on the landing page, emphasized in recruitment emails, and scattered throughout the platform's interface.

6.2 Consent Boundaries: Values and Challenges

While many of Moa's features are drawn from known social media designs, consent boundaries are a novel design contribution. Two Phase 3 participants (P11 and P13) mentioned explicitly that they had never seen anything like consent boundaries. Below, we discuss the values and novelty of consent boundaries, and then outline how to further improve them.

6.2.1 Novelty and value of consent boundaries. First, as participants noted, consent boundaries offer *highly relevant, context-focused* dimensions for audience selection. In the case of PhD students discussing advising challenges, the dimensions were related to social identity, academic experience, granular information about the institution, etc.—such dimensions are not offered by existing platform structures [92] or access control systems, most of which offer simple subsets such as “private / friends / public” or technically defined groups such as “administrator, editor, viewer.”

Second and related to the point above is that Moa's consent boundaries offer *finer-grained* dimensions than existing social media or access-control systems. Based on Phase 1's findings, the types of boundary dimensions span the full range of unions and intersections of various groups. And, Phase 3 participants who used consent boundaries leveraged different kinds of dimensions.

Third, our design of the implementation of consent boundaries carefully attends to the process of setting them. In contrast, most platforms treat audience selection as a rarely adjusted element that requires users to navigate complex “Settings” features. Our implementation honors the principles of affirmative consent [35], by ensuring that *every step* of consent-boundary-setting respects user consent. For instance, to ensure consent *specificity*, users can select a boundary for each post and comment; and indeed, Phase 3 participants tended to do so.

6.2.2 Improving consent boundaries. Despite the potential value of consent boundaries, there is a tradeoff with usability. With advancements in AI, one could imagine a system could learn how a user uses consent boundaries, and then make suggestions [35]. Or, a user could type their consent-granting preferences in natural language, and the system automatically suggests boundary dimensions. (Of course, such an approach might have to wait until AI models are perfect with respect to interpreting user requests in order to adhere to the principles of affirmative consent.)

Another challenge is that for some consent boundaries, whether a user is inside of them may not be readily verifiable. For example, users can ask to have a post be visible only to others who have experienced bullying by their advisors, but whether a user has had such experiences cannot be easily confirmed. It is therefore possible that a user reports their experiences inaccurately and gains access to discussions. Strict fidelity to affirmative consent would not allow such leakage; maybe

such dimensions should not be implemented. It is also possible, however, that compromises should be made to facilitate the desired communication. Future work should explore these tradeoffs.

6.3 Towards a Consentful Approach to System Design

Taking a consent-centered approach was essential to developing Moa. While consent and privacy are closely related, there is a subtle difference in emphasis. Speaking strictly from an engineering standpoint, consent boundaries are like privacy settings—they give different permissions to different sets of users. But, they differ from traditional privacy settings with respect to the categories of user-sets that can be formed, the granularity of control, and most importantly, the way the categories align with users' own preferences for control. Below, we reflect further on how consent differs from privacy, and what consent means for system development.

To begin, privacy is a fuzzy concept—scholars disagree on how exactly to define it, though definitions cluster around themes such as the right to be left alone, restriction of access, secrecy, control over personal information, personhood, and intimacy [77]. One relevant ambiguity in the context of computing is about who provides privacy: Privacy can be conferred to a user by the user themselves, or by a privacy-protecting system.

In contrast, the essence of consent is that it is strictly user-determined. The only entity who can grant an individual's consent is the individual herself; there is no definition of "consent" in which a system can grant itself consent to do something to a user that the user does not herself wish to allow.

Thus, relying on the term "privacy" can obscure the central role of the user, as privacy encompasses multiple, often ambiguous meanings. Arguably, the reason why "user-centered privacy" is a term distinct from general privacy is because the user-centeredness of privacy is not a given. Yet, even within the usable privacy and security community, discussions of consent have been limited to the notice-and-choice paradigm [13]. Consent is not cast as a foundational principle for system design. Based on this view, systems frequently define privacy options in advance (i.e., the system is conferring some privacy), with users only able to accept or decline them. In practice, these options are often difficult to find or understand, further limiting true user consent. For instance, many of Facebook's privacy settings are often hidden [30].

As the feminist consent framework for design recommends [35],¹⁴ consent should not merely be a mechanism for relaxing privacy constraints, but a foundational principle of design: individuals affected by an action should have the agency to define exactly what their consent boundary is, in the way that they conceive as relevant to a particular context.

6.4 Towards Collective Action

Originally, Moa was conceived to help students *resolve* their advising challenges, but we recognized the problem space was very large, and narrowed our scope to ally discovery. But as reported, some users were considering next steps beyond ally discovery. A few Phase 3 participants mentioned wanting to meet another user in-person or wished it was easier to have a 1:1 interaction on Moa.

How could a system like Moa be extended in future research? One line of work might involve developing mechanisms that enable non-anonymous connections for collective action. While Moa has allowed PhD students to have conversations around advising relationships, nobody revealed their identity to other users. To make consentful non-anonymous connections easier, future work could explore embedding an information escrow into Moa so that users can gradually build trust

¹⁴Of course, how people express consent can be different, and there is an abundant literature on understanding such a range of expressions regarding demographic factors. Feminist scholarship on consent is a major example of this (e.g., [7]). However, these differences in expression do not diminish what consent fundamentally does.

in each other, to the point of revealing their identities to one another. Information escrows are systems that connect two parties or enable information exchange, only if both provide information (to the system) that both agree would be sufficient for trust [3].¹⁵

However, one drawback of escrows is that they are private until the match happens—users cannot observe others, which is important for developing enough trust to submit information to escrows in the first place. We suspect it is possible to provide a middle ground by embedding information escrows within a platform like Moa, so that users can give and receive signals for reciprocal self-disclosures [79]. For instance, users can observe each other’s comments on Moa, and then decide to submit information after building a sufficient level of trust.

Another interesting direction is supporting users working together to arrive at a plan for action—especially in order to communicate with those in power. For a consent-based intervention to provide agency to those with less power, it helps to communicate with, or have the agreement of, parties in power—in Moa’s context, faculty. We noticed some participants were thinking about this. Unprompted, some participants from both Phase 1 and 3 mentioned that Moa could be extended to convey PhD students’ experiences to faculty. This requires a group of users to collectively plan and act [71, 83]. In particular, there are interesting questions around whether revelations about the identities of group members seem necessary, and if so, how that can be accomplished.

6.5 Audience Configuration and Identity Management in Sensitive Contexts

Some users showed consent boundaries to signal that the audience shares a common identity trait, in order for other users to chime in more comfortably. This was the most pronounced for a post that was about gender dynamics, where the poster used a women-only boundary. Most of the commenters also decided to use the same boundary for their comments and chose to display them.

Relatedly, a few of Phase 3 interview participants also noted that the commenting feature could be designed so that it is easier to display parts of one’s identity when responding, which resembles meronymy [76]. Moa could potentially be extended to provide users with such features, so that once an audience restriction has happened, users within the thread could more freely express parts of their identity if they wanted to, separately from using consent boundaries.

Some users who did not show their consent boundaries thought they did not add much information, since the post will be delivered to the right audience anyway. These Phase 3 participants were less concerned with expressing identity and more focused on audience control, possibly because Moa’s context centers on sensitive power dynamics, unlike prior work (e.g., [51, 76]). For example, while meronymy [76] was evaluated where different levels of seniority exist, the focus was on academic research discourse, rather than power hierarchies.

6.6 Scaling Moa and Consent Boundaries

Currently, Moa is maintained by the lead author in order to minimize privacy risks. Running the platform with 47 users turned out to be feasible—the lead author was diligent in moderating and seeking user feedback, and no incidents of abuse or negative feedback occurred. However, how could Moa and the concept of consent boundaries scale?

One way to scale is having multiple instances of Moa, one per organization or group, each with its own administrator, similar to Mastodon. This approach reflects an important assumption that local system admins are important for earning users’ trust, which was confirmed by our Phase 3 study. And, just as research shows that Reddit’s moderators have their own network of giving each

¹⁵Bad actors could try to submit false information to information escrows. We believe the most feasible approach to address such issues is to have human moderators. While there could be scalability issues, systems like Callisto [66], show that it is feasible for escrow systems to be implemented with moderators on a scale of spanning multiple campuses. <https://www.projectcallisto.org/callistovault>

other support [9], we imagine that there should be a support network for Moa’s administrators and moderators. This would ensure that each instance is not entirely isolated, and that when sensitive incidents happen, there would be cross-instance resources on how to resolve them.

It is possible to imagine building Moa as a larger, monolithic platform such as Facebook or Reddit, but we imagine even on such a platform, communities would have their own moderators—each trusted by their respective communities—who are empowered to provide the kind of meaningful oversight over identity that the Moa prototype included. For this study, it was easy for users to verify and understand that the moderator was also a PhD student in the same institution. At larger scale, there would need to be a process for selecting such moderators, and proving their identity to other members of the community.

In terms of consent boundaries, they are likely to be applicable to other contexts beyond PhD advising, such as among workplace employees [18, 67], survivors of sexual misbehavior by one group of people [16, 48], or potential whistleblowers at the same organization [19]. One aspect of the Moa prototype described in this paper was that there was an implicit consent boundary formed by the community of PhD students from the same institution—no user needed to consider whether their content would be revealed to non-students. Communities could be formed much like Slack’s channels around different topics, though additional mechanisms such as member-sponsored invitations to join or formal background checks would be required to ensure that community members satisfy membership requirements.

We also imagine consent boundaries could be used in general platforms such as Facebook or Reddit, likely in less sensitive contexts. Consent boundaries could help users carve out a group of people they want to interact with, without having to designate a particular space or network [92]. This would resemble the “social circles” features on platforms such as Close Friends,¹⁶ but it would not require individual enumeration of members or knowledge of other users’ identities. For example, users who were exchanging comments under a post on a subreddit could use consent boundaries to easily create a space within the thread for users that meet particular criteria. At the same time, when consent boundaries are used within a large-scale commercial platform, there might be a bigger hesitation for users to use the feature. It may take a longer time for users to build trust in consent boundaries, as failures could result in unintended exposure of posts to a broad audience. Because of this, the usage rate could be much lower in the beginning.

7 Conclusion

When individuals experience harm from someone in power, finding allies—people who are sympathetic or willing to navigate an issue together—is important. To understand how to develop a social platform that supports sensitive conversations for ally discovery, we conducted interviews with 19 PhD students, which revealed a strong desire for a platform that connects peers to discuss PhD advising relationships, and reinforced the importance of consent in enabling these connections. Building on these findings and the affirmative consent framework, we developed a social platform called Moa. One of Moa’s core features is the “consent boundary,” which allow users to fluidly demarcate who can view each post or comment based on dimensions such as social identity, lived experience, and interpersonal relationships—while preserving anonymity. The sender and recipient do not learn of each other’s identity, even as the post reaches the intended audience. We conducted a three-week field study with 47 PhD students that showed Moa successfully facilitated ally discovery. Consent boundaries lowered barriers for some users, enabling them to ask sensitive questions and provide authentic advice. In all, this work presents insights on how to develop a social platform for ally discovery and instantiates a consent-centered approach to system design.

¹⁶https://www.facebook.com/help/200538509990389?helpref=faq_content

8 Acknowledgements

We sincerely thank all participants. A big thank you to Vikram Mohanty for being a sounding board throughout the system development. Chaerin Im deserves a very special shout-out for encouraging Im to build a new social media platform in 2021. We also thank Jiyeon Kim, Shwetha Rajaram, and Yulin Yu for their valuable feedback and kind support. Yunseok Jang gave much-needed suggestions for the tutorial. Im also thanks Amy Ko, Mark Ackerman, and Emily Mower Provost for their helpful comments. Im was supported by a Meta Research PhD Fellowship.

References

- [1] Dinislam Abdulgalimov, Reuben Kirkham, James Nicholson, Vasilis Vlachokyriakos, Pam Briggs, and Patrick Olivier. 2020. Designing for employee voice. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [2] Tawfiq Ammari, Sarita Schoenebeck, and Daniel Romero. 2019. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
- [3] Ian Ayres and Cait Unkovic. 2012. Information escrows. *Mich. L. Rev.* 111 (2012), 145.
- [4] Benita J Barnes. 2009. The nature of exemplary doctoral advisors' expectations and the ways they may influence doctoral persistence. *Journal of College Student Retention: Research, Theory & Practice* 11, 3 (2009), 323–343.
- [5] Amanda Baughan, Larry Tian, Pranav Shekar, Amy Zhang, and Alexis Hiniker. 2024. Supporting Hard Conversations in Close Relationships Through Design. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–22.
- [6] Monica Becerra, Emily Wong, Brooke N Jenkins, and Sarah D Pressman. 2021. Does a good advisor a day keep the doctor away? How advisor-advisee relationships are associated with psychological and physical well-being among graduate students. *International Journal of Community Well-Being* 4 (2021), 505–524.
- [7] Melanie A Beres. 2007. 'Spontaneous' sexual consent: An analysis of sexual consent literature. *Feminism & Psychology* 17, 1 (2007), 93–108. <https://doi.org/10.1177/0959353507072914>
- [8] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 21–30.
- [9] Lia Bozarth, Jane Im, Christopher Quarles, and Ceren Budak. 2023. Wisdom of Two Crowds: Misinformation Moderation on Reddit and How to Improve this Process—A Case Study of COVID-19. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 155 (April 2023), 33 pages. <https://doi.org/10.1145/3579631>
- [10] Sara Branch, Sheryl Ramsay, and Michelle Barker. 2013. Workplace bullying, mobbing and general harassment: A review. *International Journal of Management Reviews* 15, 3 (2013), 280–299.
- [11] Stephanie M Breen, Jesse McCain, and Josipa Roksa. 2024. Breaking points: exploring how negative doctoral advisor relationships develop over time. *Higher Education* (2024), 1–20.
- [12] David P Bryden. 1999. Redefining rape. *Buff. Crim. L. Rev.* 3 (1999), 317.
- [13] Fred H Cate. 2006. The failure of fair information practice principles. *Consumer protection in the age of the information economy* (2006).
- [14] Aaron Cohen and Yehuda Baruch. 2022. Abuse and exploitation of doctoral students: A conceptual model for traversing a long and winding road to academia. *Journal of Business Ethics* 180, 2 (2022), 505–522.
- [15] Nancy L Collins and Lynn Carol Miller. 1994. Self-disclosure and liking: a meta-analytic review. *Psychological bulletin* 116, 3 (1994), 457.
- [16] Lilia M Cortina and Maira A Areguin. 2021. Putting people down and pushing them out: Sexual harassment in the workplace. *Annual Review of Organizational Psychology and Organizational Behavior* 8, 1 (2021), 285–309.
- [17] Jenny L Davis and Nathan Jurgenson. 2014. Context collapse: Theorizing context collusions and collisions. *Information, communication & society* 17, 4 (2014), 476–485.
- [18] Patrick Decker-Tonnesen, Kabuika Kamunga, Erick Garcia, Monica Ibarra, Isabelle Martin, Kara Saliba, Caleta Beards, Barbara Jordan, and Anjali Bhagra. 2024. A catalyst for activation against racism: case study on effectiveness of workplace equity, inclusion and diversity conversations. *Journal of Workplace Learning* (2024).
- [19] Philip Di Salvo. 2022. Leaking black boxes: Whistleblowing and big tech invisibility. *First Monday* (2022).
- [20] Judith Donath. 2007. Signals in social supernets. *Journal of Computer-Mediated Communication* 13, 1 (2007), 231–251. <https://doi.org/10.1111/j.1083-6101.2007.00394.x>
- [21] Chai R Feldblum and Victoria A Lipnic. 2016. Select task force on the study of harassment in the workplace. *Washington: US Equal Employment Opportunity Commission* (2016). https://www.eeoc.gov/select-task-force-study-harassment-workplace#_ftn62

- [22] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 2021. A design space for privacy choices: Towards meaningful privacy control in the internet of things. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [23] Kara Fox and Jan Diehm. 2017. # MeToo’s global moment: The anatomy of a viral campaign. *CNN*, November 9 (2017).
- [24] Rachel E Friedensen, Genia M Bettencourt, and Megan L Bartlett. 2023. Power-conscious ecosystems: Understanding how power dynamics in US doctoral advising shape students’ experiences. *Higher Education* (2023), 1–16.
- [25] Jaclyn Friedman and Jessica Valenti. 2019. *Yes means yes!: Visions of female sexual power and a world without rape*. Seal Press.
- [26] Robert W Fuller. [n. d.]. Rankism: A Social Disorder. *Breaking Ranks* ([n. d.]).
- [27] Chris M Golde. 2005. The role of the department and discipline in doctoral student attrition: Lessons from four departments. *The Journal of Higher Education* 76, 6 (2005), 669–700.
- [28] Sukeshini A Grandhi and Lyndsey K Lanagan-Leitzel. 2016. To reply or to reply all: Understanding replying behavior in group email communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 560–569.
- [29] Andy Greenberg. 2012. How Reddit’s Alexis Ohanian Became the Mayor of the Internet. <https://www.forbes.com/forbes/2012/0625/technology-politics-alexis-ohanian-reddit-sopa-mayor-of-internet.html>
- [30] Hana Habib, Sarah Pearman, Ellie Young, Ishika Saxena, Robert Zhang, and Lorrie Faith Cranor. 2022. Identifying User Needs for Advertising Controls on Facebook. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 59 (apr 2022), 42 pages. <https://doi.org/10.1145/3512906>
- [31] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 2021. Toggles, dollar signs, and triangles: How to (in) effectively convey privacy choices with icons and link texts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–25.
- [32] Tessa Haesevoets, David De Cremer, Leander De Schutter, Jack McGuire, Yu Yang, Xie Jian, and Alain Van Hiel. 2021. Transparency and control in email communication: The more the supervisor is put in cc the less trust is felt. *Journal of Business Ethics* 168 (2021), 733–753.
- [33] Janet Halley. 2016. The move to affirmative consent. *Signs: Journal of Women in Culture and Society* 42, 1 (2016), 257–279. <https://doi.org/10.1086/686904>
- [34] Heidi M Hurd. 1996. The moral magic of consent. *Legal theory* 2, 2 (1996), 121–146.
- [35] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S. Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI ’21). Association for Computing Machinery, New York, NY, USA, Article 403, 18 pages. <https://doi.org/10.1145/3411764.3445778>
- [36] Jane Im, Ruiyi Wang, Weikun Lyu, Nick Cook, Hana Habib, Lorrie Faith Cranor, Nikola Banovic, and Florian Schaub. 2023. Less is Not More: Improving Findability and Actionability of Privacy Controls for Online Behavioral Advertising. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–33.
- [37] Romain Jacob, Irena Kuzmanovska, and Nina Ripin. 2018. *AVETH survey on supervision of doctoral students*. Technical Report. ETH Zurich.
- [38] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does transparency in moderation really matter? User behavior after content removal explanations on reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.
- [39] Corinne Jones. 2023. How to train your algorithm: The struggle for public control over private audience commodities on Tiktok. *Media, Culture & Society* 45, 6 (2023), 1192–1209.
- [40] James P Kahan and Amnon Rapoport. 2014. *Theories of coalition formation*. Psychology Press.
- [41] Haesoo Kim, Juhoon Lee, Jeong-Woo Jang, and Juho Kim. 2024. ReSPect: Enabling Active and Scalable Responses to Networked Online Harassment. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–30.
- [42] JaeWon Kim, Robert Wolfe, Ramya Bhagirathi Subramanian, Mei-Hsuan Lee, Jessica Colnago, and Alexis Hiniker. 2025. Trust-Enabled Privacy: Social Media Designs to Support Adolescent User Boundary Regulation. *arXiv preprint arXiv:2502.19082* (2025).
- [43] Paul Kinnersley, Katie Phillips, Katherine Savage, Mark J Kelly, Elinor Farrell, Ben Morgan, Robert Whistance, Vicky Lewis, Mala K Mann, Bethan L Stephens, et al. 2013. Interventions to promote informed consent for patients undergoing surgical and other invasive healthcare procedures. *Cochrane Database of Systematic Reviews* 7 (2013).
- [44] Alex Leavitt. 2015. This is a Throwaway Account” Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 317–327.
- [45] Una Lee and Dann Toliver. 2017. *Building Consentful Tech*. <http://www.consentfultech.io/wp-content/uploads/2019/10/Building-Consentful-Tech.pdf>

- [46] Katia Levecque, Frederik Anseel, Alain De Beuckelaer, Johan Van der Heyden, and Lydia Gisle. 2017. Work organization and mental health problems in PhD students. *Research policy* 46, 4 (2017), 868–879.
- [47] Erika Löfström and Kirsi Pyhältö. 2020. What are ethics in doctoral supervision, and how do they matter? Doctoral students' perspective. *Scandinavian Journal of Educational Research* 64, 4 (2020), 535–550.
- [48] Kimberly A Lonsway, Rebecca Paynich, and Jennifer N Hall. 2013. Sexual harassment in law enforcement: Incidence, impact, and perception. *Police Quarterly* 16, 2 (2013), 177–210.
- [49] Alexis Lothian and Mel Stanfill. 2021. An archive of whose own? White feminism and racial justice in fan fiction's digital infrastructure. *Transformative Works and Cultures* 36 (2021), 2021.
- [50] Byron Lowens, Sean Scarnecchia, Jane Im, Tanisha Afnan, Annie Chen, Yixin Zou, and Florian Schaub. 2025. Misalignments and Demographic Differences in Expected and Actual Privacy Settings on Facebook. *Proceedings on Privacy Enhancing Technologies* (2025).
- [51] Xiao Ma, Jeff Hancock, and Mor Naaman. 2016. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3857–3869.
- [52] Ifigeneia Machili, Jo Angouri, and Nigel Harwood. 2019. 'The snowball of emails we deal with': CCing in multinational companies. *Business and Professional Communication Quarterly* 82, 1 (2019), 5–37.
- [53] Mary Madden, Michele Gilman, Karen Levy, and Alice Marwick. 2017. Privacy, poverty, and big data: A matrix of vulnerabilities for poor Americans. *Wash. UL Rev.* 95 (2017), 53.
- [54] Brian Martin. 2013. Countering supervisor exploitation. *Journal of Scholarly Publishing* 45, 1 (2013), 74–86.
- [55] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133. <https://doi.org/10.1177/1461444810365313>
- [56] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
- [57] Jennifer H McQuaid, Michelle Alejandra Silva, and Katherine C McKenzie. 2021. Surviving violent, traumatic loss after severe political persecution: lessons from the evaluation of a Venezuelan asylum seeker. *BMJ Case Reports CP* 14, 3 (2021), e239025.
- [58] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P Gummadi, and Aniket Kate. 2016. Forgetting in social media: Understanding and controlling longitudinal exposure of socially shared data. In *Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016*. 287–299.
- [59] Charlene L Muehlenhard, Terry P Humphreys, Kristen N Jozkowski, and Zoë D Peterson. 2016. The complexities of sexual consent among college students: A conceptual and empirical review. *The Journal of Sex Research* 53, 4-5 (2016), 457–487. <https://doi.org/10.1080/00224499.2016.1146651>
- [60] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.
- [61] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.
- [62] Federica Paci, Anna Squicciarini, and Nicola Zannone. 2018. Survey on access control for community-centered collaborative systems. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–38.
- [63] Xinru Page, Reza Ghaiumy Anaraky, Bart P Knijnenburg, and Pamela J Wisniewski. 2019. Pragmatic tool vs. relational hindrance: Exploring why some social media users avoid privacy features. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
- [64] Daniel L Peluso, R Nicholas Carleton, and Gordon JG Asmundson. 2011. Depression symptoms in Canadian psychology graduate students: Do research productivity, funding, and the academic advisory relationship play a role? *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 43, 2 (2011), 119.
- [65] Calvin S Powers, Paul Ashley, and Matthias Schunter. 2002. Privacy promises, access control, and privacy management. Enforcing privacy throughout an enterprise by extending access control. In *Proceedings. Third International Symposium on Electronic Commerce*. IEEE, 13–21.
- [66] Anjana Rajan, Lucy Qin, David W Archer, Dan Boneh, Tancrede Lepoint, and Mayank Varia. 2018. Callisto: A cryptographic approach to detecting serial perpetrators of sexual misconduct. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. 1–4.
- [67] Mona Ramonetti and Victoria Pilato. 2019. Keeping the equity, inclusion, and diversity conversations going. *Urban Library Journal* 25, 1 (2019), 1.
- [68] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. 2015. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech. Lj* 30 (2015), 39.
- [69] Paul Resnick, Joseph Konstan, Yan Chen, and Robert E Kraut. 2012. Starting new online communities. *Building successful online communities: Evidence-based social design* 231 (2012).
- [70] Fujiko Robledo Yamamoto, Janghee Cho, Amy Volda, and Stephen Volda. 2023. "We are Researchers, but we are also Humans": Creating a Design Space for Managing Graduate Student Stress. *ACM Transactions on Computer-Human*

- Interaction* 30, 5 (2023), 1–33.
- [71] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 2015. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1621–1630.
- [72] Florian Schaub and Lorrie Cranor. 2020. *Usable and Useful Privacy Interfaces*. IAPP.
- [73] F David Schoorman, Roger C Mayer, and James H Davis. 2007. An integrative model of organizational trust: Past, present, and future. , 344–354 pages.
- [74] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping pro and anti-social behavior on twitch through moderation and example-setting. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 111–125.
- [75] Karianne Skovholt and Jan Svennevig. 2006. Email copies in workplace interaction. *Journal of Computer-Mediated Communication* 12, 1 (2006), 42–65.
- [76] Nouran Soliman, Hyeonsu B Kang, Matthew Latzke, Jonathan Bragg, Joseph Chee Chang, Amy Xian Zhang, and David R Karger. 2024. Mitigating Barriers to Public Social Interaction with Meronymous Communication. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.
- [77] Daniel J Solove. 2002. Conceptualizing privacy. *Calif. L. Rev.* 90 (2002), 1087.
- [78] Daniel J Solove. 2012. Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.* 126 (2012), 1880.
- [79] Susan Sprecher, Stanislav Treger, Joshua D Wondra, Nicole Hilaire, and Kevin Wallpe. 2013. Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology* 49, 5 (2013), 860–866.
- [80] Judith P Swazey, Melissa S Anderson, Karen Seashore Lewis, and Karen Seashore Louis. 1993. Ethical problems in academic research. *American Scientist* 81, 6 (1993), 542–553.
- [81] Charles Tilly. 2004. Social boundary mechanisms. *Philosophy of the social sciences* 34, 2 (2004), 211–236.
- [82] Jessica Vitak. 2012. The impact of context collapse and privacy on social network site disclosures. *Journal of broadcasting & electronic media* 56, 4 (2012), 451–470.
- [83] Vasillis Vlachokyriakos, Clara Crivellaro, Pete Wright, Evika Karamagioli, Eleni-Revekka Staiou, Dimitris Gouscos, Rowan Thorpe, Antonio Krüger, Johannes Schöning, Matt Jones, et al. 2017. HCI, solidarity movements and the solidarity economy. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3126–3137.
- [84] Yi-Chia Wang, Moira Burke, and Robert Kraut. 2016. Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 74–85.
- [85] Alan Wertheimer. 1999. What Is Consent and Is It Important? *Buff. Crim. L. Rev.* 3 (1999), 557.
- [86] Zack Whittaker. 2018. At Blind, a security lapse revealed private complaints from Silicon Valley employees. <https://techcrunch.com/2018/12/20/blind-anonymous-app-data-exposure/#:~:text=But%20Blind%20left%20one%20of,and%20identify%20would%20be%20whistleblowers>
- [87] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for my space: Coping mechanisms for SNS boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 609–618.
- [88] Chris Woolston. 2017. Graduate survey: A love–hurt relationship. *Nature* 550, 7677 (2017), 549–552.
- [89] Yaxing Yao, Justin Reed Basdeo, Smirity Kaushik, and Yang Wang. 2019. Defending my castle: A co-design study of privacy mechanisms for smart homes. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.
- [90] Kim Young-won. 2017. Blind: The anonymous app that serves as Silicon Valley’s gossip outlet. <https://www.theinvestor.co.kr/view.php?ud=20170330000905>
- [91] Dieter Zapf, Jordi Escartin, Miriam Scheppa-Lahyani, Ståle Valvatne Einarsen, Helge Hoel, and Maarit Vartia. 2020. Empirical findings on prevalence and risk groups of bullying in the workplace. *Bullying and harassment in the workplace: Theory, research and practice* 3 (2020), 105–162.
- [92] Amy X Zhang, Michael S Bernstein, David R Karger, and Mark S Ackerman. 2024. Form-From: A Design Space of Social Media Systems. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–47.
- [93] Justine Zhang, James Pennebaker, Susan Dumais, and Eric Horvitz. 2020. Configuring audiences: A case study of email communication. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–26.

A Appendix

A.1 Participants' demographics for Phase 3 study

55.3% were international students, 38.3% were domestic, and 6.4% did not submit any relevant information. 48.9% of the users were women, 31.9% were men, 2.1% were non-binary, and 17% did not submit gender information. 65.9% were Asian, 14.9% were Caucasian, 8.51% were Middle Eastern or North African, 4.26% were of Hispanic, Latinx, or Spanish origin, and 2.13% were Black or African American. 4.2% had mixed race and 6.4% did not submit information.

Received 13 May 2025

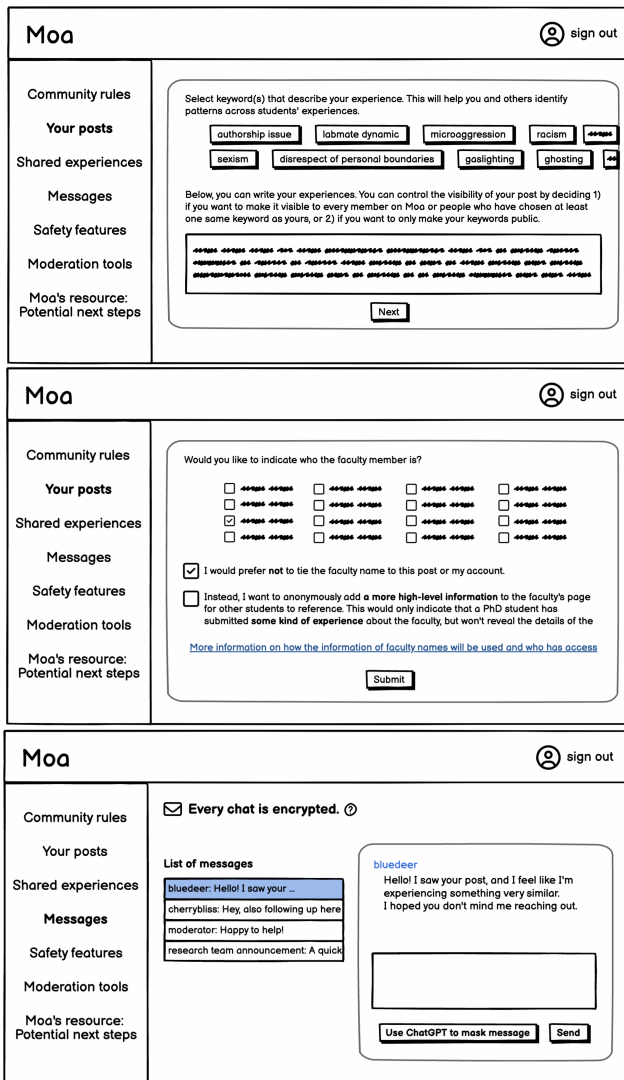


Fig. 7. Example designs that show a system for enabling PhD students to connect around advising challenges.

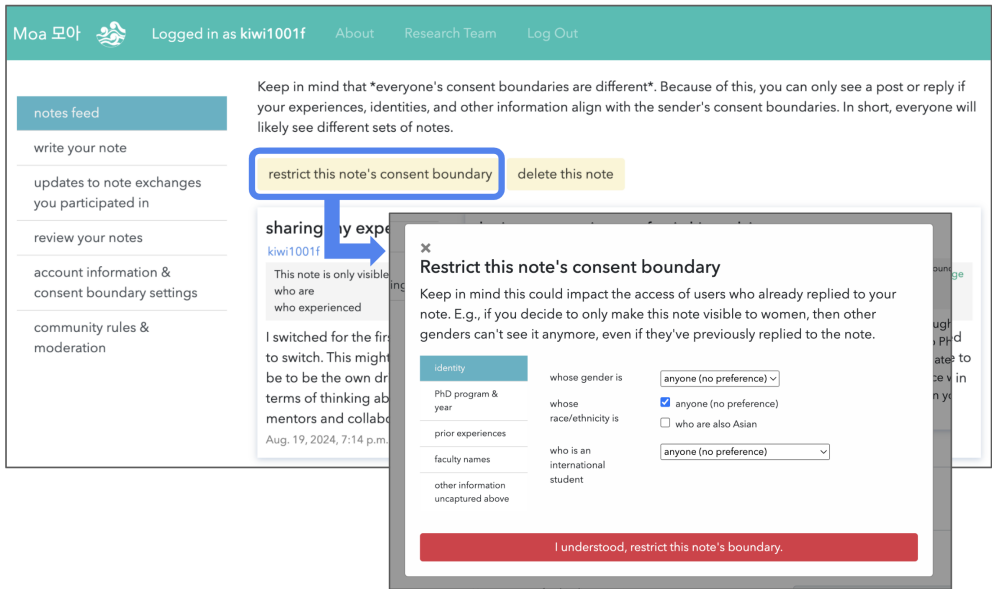


Fig. 8. Users can restrict a post or comment’s consent boundary.

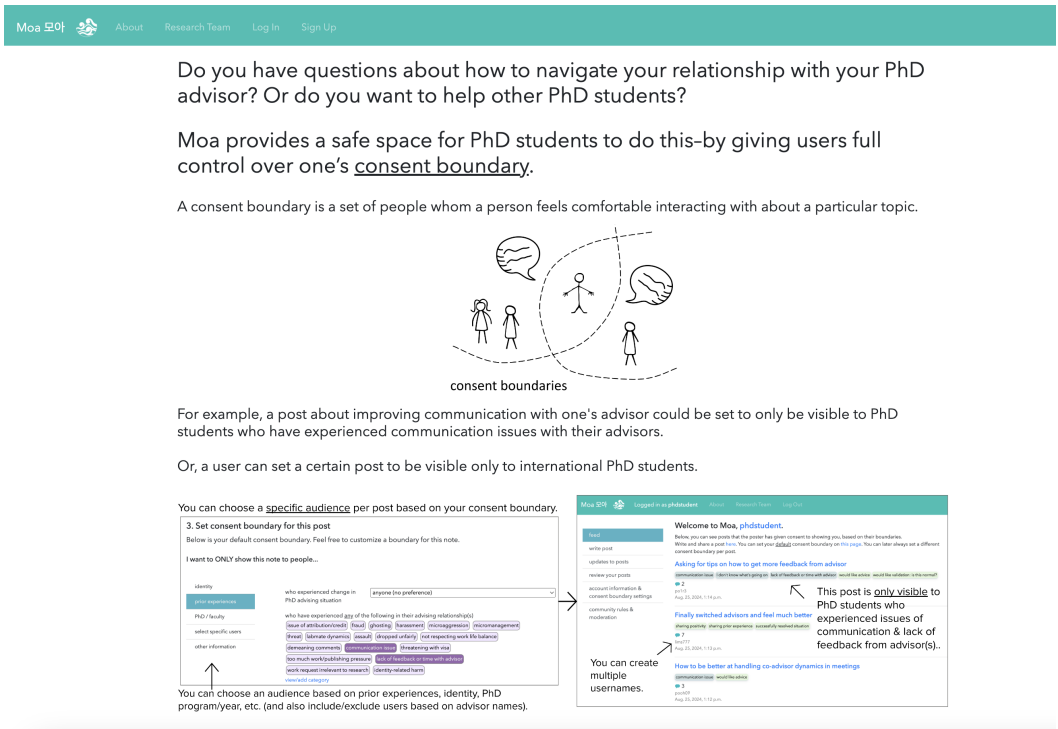


Fig. 9. Moa’s landing page.

The image shows three sequential screenshots of the Moa sign-up page. The top navigation bar in all screenshots includes 'Moa', 'About', 'Research Team', 'Log In', and 'Sign Up'.

Section 1: Sign Up
 Thank you for agreeing to participate. Because discussing PhD advising relationships is sensitive, the moderator [name redacted] manually reviews sign-up requests to ensure all users are PhD students. Your submitted information is **NOT visible to anyone else** on the system, except for your username (which you can always later change).
 Email*
 Use your *institution email* for verification purposes.
 Username*
 This will be your *default username shown to other users*. You can later *change* it anytime and/or set a different username per discussion.
 Password*
 Password
 • Your password can't be too similar to your other personal information.
 • Your password must contain at least 8 characters.
 • Your password can't be a commonly used password.
 • Your password can't be entirely numeric.
 Password (again)*
 Password (again)

Section 2: Quick Tutorial and Information
 Before proceeding to the next questions, please check out this quick tutorial. It'll help you to decide how to answer the below questions.
 While all questions are optional, answering more helps the system connect you better to other users' posts. The information you submit will **NOT be publicly visible**—it is only for showing you the right content.
 Finally, because PhD advising relationships is a sensitive topic, we ask you to submit correct information. The moderator ([name redacted]) manually checks every sign-up request. If you have any questions, please reach out to ([name redacted]).
 Year you started your PhD
 (For example, if you started in Fall 2018, you should type in 2018.)
 Gender | leave blank
 I identify as (select all that apply):
 American Indian or Alaska Native
 Asian
 Black or African American
 Caucasian
 Hispanic, Latinx, or Spanish origin
 Middle Eastern or North African
 Native Hawaiian or Pacific Islander
 Other

Section 3: Institutional and Advising Information
 I am an international student | leave blank
 Institution/PhD program | leave blank
 I have experienced the following in my PhD advising relationship(s) (select from below).
 issue of attribution/credit
 bias/d
 ghosting
 harassment
 microaggression
 micromanagement
 threat
 labmate dynamics
 assault
 dropped/unfairly
 not respecting work life balance
 demeaning comments
 communication issue
 threatening with visa
 too much work/publishing pressure
 lack of feedback or time with advisor
 work request irrelevant to research
 identity-related harm
 Have you ever experienced any change in your PhD advising situation? E.g., switching, leaving, adding (co-)advisors | leave blank
 Current PhD advisor(s) (Please type in the full name(s) separated by commas.)
 Past PhD advisor(s) (Please type in the full name(s) separated by commas.)
 Feel free to type in your personal email address, if you want [name redacted] to discard your institution email and save your personal one in the database instead (after verification).
 Submit

Fig. 10. Moa’s sign-up page. Users are asked to voluntarily submit identity-related information, PhD program, advisor names, and challenges experienced in advising relationships.

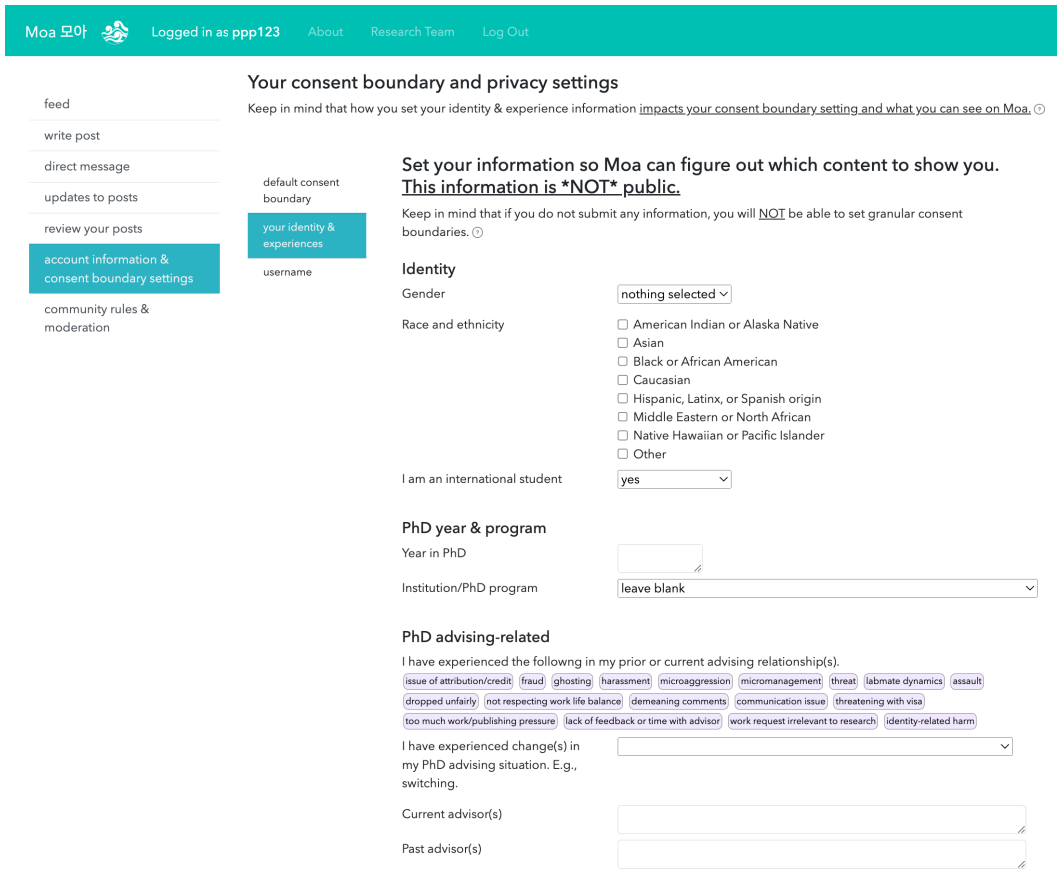


Fig. 11. Users can update their identity traits and experiences on Moa in the account settings page.