Towards Youth-Sensitive Hateful Content Reporting: An Inclusive Focus Group Study in Germany

Julian Bäumler
Science and Technology for Peace and Security (PEASEC)
Technical University of Darmstadt
Darmstadt, Germany
baeumler@peasec.tu-darmstadt.de

Helen Bader
Science and Technology for Peace and Security (PEASEC)
Technical University of Darmstadt
Darmstadt, Germany
helen.bader@stud.tu-darmstadt.de

Marc-André Kaufhold
Science and Technology for Peace and Security (PEASEC)
Technical University of Darmstadt
Darmstadt, Germany
kaufhold@peasec.tu-darmstadt.de

Christian Reuter
Science and Technology for Peace and Security (PEASEC)
Technical University of Darmstadt
Darmstadt, Germany
reuter@peasec.tu-darmstadt.de

Abstract

Youth are particularly likely to encounter hateful internet content, which can severely impact their well-being. While most social media provide reporting mechanisms, in several countries, severe hateful content can alternatively be reported to law enforcement or dedicated reporting centers. However, in Germany, many youth never resort to reporting. While research in human-computer interaction has investigated adults' views on platform-based reporting, youth perspectives and platform-independent alternatives have received little attention. By involving a diverse group of 47 German adolescents and young adults in eight focus group interviews, we investigate how youth-sensitive reporting systems for hateful content can be designed. We explore German youth's reporting barriers, finding that on platforms, they feel particularly discouraged by deficient rule enforcement and feedback, while platform-independent alternatives are rather unknown and perceived as time-consuming and disruptive. We further elicit their requirements for platform-independent reporting tools and contribute with heuristics for designing youth-sensitive and inclusive reporting systems.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**.

Keywords

Youth, Adolescents, Young adults, Hateful content, Focus Groups, Reporting, Social Media

ACM Reference Format:
Julian Bäumler, Helen Bader, Marc-André Kaufhold, and Christian Reuter. 2025. Towards Youth-Sensitive Hateful Content Reporting: An Inclusive Focus Group Study in Germany. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3706598.3713542

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713542

1 Introduction

Youth, i.e., adolescents and young adults, are particularly likely to encounter hateful internet content [57, 67]. In Germany, they use social media more frequently than older demographics [74], which increases opportunities to come into contact with such content. A 2024 survey shows that in Germany, 18.5% of students aged seven to 20 experienced cyberbullying [8], while data from 2023 indicates that 89% of 14 to 24-year-olds have already encountered online hate speech [67]. This can severely impact their well-being by causing anxiety, feelings of fear and insecurity, or even sleeping disorders and psychological conditions [7, 35, 57]. However, young Germans are also willing to react. In 2023, 52% of 14 to 24-year-olds who came across such content have already reported it to platforms, compared to only 30% across all age groups [67]. Given their particularly high exposure and willingness to respond, the perspectives of German youth on hateful content reporting merit scientific attention.

Many social media platforms allow their users to flag, i.e., report, hateful content that might violate community guidelines, which is subsequently reviewed and, if a violation is found, sanctioned [21, 88]. In jurisdictions where severe instances of hateful content are subject to criminal law [18], it is alternatively possible to file reports to law enforcement agencies (LEAs), and in a number of countries, there additionally exist non-profit hate speech reporting centers [44, 77]. As victim-centered organizations, these centers typically accept reports via self-administered web portals and act as intermediaries between reporting individuals and platform operators and LEAs, supporting victims in the deletion and, if criminal law is applicable, prosecution of respective content by leveraging established contacts and communication channels with these actors [15, 80]. Of the few youth-centered studies in human–computer interaction (HCI) on content moderation [90, 101, 111] or addressing online hate [4, 47, 50, 100], none have been conducted in the German context. However, Germany constitutes an interesting case, since the 2017 Network Enforcement Act (NetzDG), as the first national law of its kind, mandates large social media platforms to remove clearly illegal hate speech 24 hours after a report [40, 61]. It

shaped the Digital Services Act (DSA) of the European Union (EU), which emphasizes notice-and-takedown obligations and greater moderation transparency [52]. With the designation of some German reporting centers as trusted flagger under the DSA [17], their role in combating hateful content increases. As similar organizations exist across the EU [44, 77], such developments might follow in other countries. Nonetheless, in 2023 around half of Germans under the age of 24 who encountered hateful content stated that they had never resorted to platform-based reporting, and almost none indicated to have filed a report to LEAs [67]. For platform-independent reporting centers, no data is available. Considering this as well as LEAs' and reporting centers' significance in national efforts against hate, the German context seems particularly suited to explore what discourages youth from reporting and how their uptake of reporting options, especially platform-independent ones, can be improved.

Initial work has already explored individual characteristics [75] and contextual factors [25] that influence youth's willingness to report. In addition, it was found that the content moderation approaches of many platforms, which primarily focus on sanctions, do not accommodate all needs of young online harm victims [90, 111]. Yet, whether the design of reporting systems influences this demographic's willingness to report has not been studied. By contrast, adults' perspectives on platform-based reporting, including those affected by platforms' sanctions [73] or using reporting mechanisms [59, 62, 113], have already been investigated. Beyond that, there is an active community in HCI that is developing user-centered technologies to support victims and bystanders of hateful or harassing content, including detection tools [11, 71, 81], documentation tools [39, 98], chatbots [47, 63, 99] and peer support networks [10, 28, 70] for the provision of information and support, and tools to facilitate coping [83, 103]. In summary, the design of reporting systems has received limited attention in HCI thus far. The few available user studies focus on adults' perspectives on platform-based reporting, while, to the best of our knowledge, there is no work on the user-centered design of platform-independent reporting tools for hateful content. Furthermore, we are not aware of any research on youth's perspectives on reporting systems for such content, irrespective of whether they are platform-based or platform-independent. To address these gaps in research, our paper is guided by the following overarching research question: **How can reporting systems for hateful content be designed in a youth-sensitive manner?** More concretely, we focus on the perspective of German adolescents and young adults with diverse identities and seek to answer the following three sub-questions:

- What barriers impede German youth from reporting hateful content? (RQ1)
- Which requirements do German youth have on the design of platform-independent reporting tools? (RQ2)
- What heuristics should guide the design of youth-sensitive and inclusive reporting systems? (RQ3)

Based on our literature review and identified research gaps (Sec. 2), we employ an exploratory qualitative research design comprising eight focus group interviews (FGIs) [64, 68] with a sample of diverse German adolescents and young adults (N=47) and a subsequent structuring qualitative content analysis following Kuckartz [65] (Sec. 3). We not only recruit participants of different age, gender, and formal level of education, but also involve youth who have a migration biography[1], are black, indigenous, people of colour (BIPoC), or identify as lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other (LGBTQIA+), as in Germany these groups are particularly targeted by hateful content [9]. In summary, we provide three contributions to the state of research in HCI on the reporting of online harms and youth's handling of hateful content. In our results (Sec. 4), we contribute with empirical insights on German youth's barriers in relation to the platform-based and -independent reporting of hateful content (C1) and their requirements on the design of platform-independent reporting tools for such content (C2). As these insights are generated independently of concrete technical solutions, they have the potential to inform the development of different types of reporting tools, e.g., apps, plugins, or chatbots. Finally, as a third contribution (Sec. 5), we derive five heuristics for the design of youth-sensitive and inclusive reporting systems on basis of our empirical findings and previous work (C3). These explicitly apply to both platform-independent and platform-based reporting. We then outline our limitations and close the paper with a brief conclusion (Sec. 6).

2 Background and Related Work

As perspectives on hateful content reporting are our focus, we first want to clarify key concepts. In the FGIs, we explored the subject in relation to hate speech and cyberbullying, as many German youth encounter these phenomena [8, 67]. While a universally accepted definition of hate speech remains elusive [79, 93], we follow the view that it attacks or diminishes groups or individuals based on actual or ascribed group affiliations or identities [30, 92]. Cyberbullying, by contrast, can be seen as aggressive behavior that occurs via the internet and is directed at individuals, where perpetrators intentionally and repeatedly exploit their superior position to inflict harm [51, 112]. This can involve hate speech.

2.1 Adolescents and Young Adults Navigating Hateful Content

Youth, i.e., adolescents and young adults[2], are more likely to encounter hateful content than older internet users [9, 57, 67]. For Germany, a 2023 representative survey found that the vast majority (89%) of respondents between 14 and 24 have already personally encountered hateful content, which is significantly higher than the proportion of the overall population (76%) [67]. Exposure to hateful content can severely impact their well-being, e.g., by causing anxiety, feelings of fear and insecurity, or even sleeping disorders and psychological conditions [7, 35, 57, 107]. It has been found that LGBTQIA+ youth are particularly at risk of self-harm due to online harassment [100]. In response to online hate, youth employ various strategies. Passive strategies comprise, e.g., ignoring content or avoiding the platform, while active strategies include blocking,

[1] In Germany, individuals are considered to have a migration biography (Migrationshintergrund) if either they or one of their parents were born without German citizenship (see official definition [78]). We adopted this notion due to our study's scope.
[2] In this study, we differentiate youth into adolescents aged between 14 and 17 and young adults aged between 18 and 29. This follows the adolescent definition of German law and further reflects the age limit up to which the OECD defines youth [29].

educating, or shaming of offenders, as well as counter speech, fact-checking, seeking support, and reporting content [9, 107]. With regard to reporting, representative data for Germany shows that only half (52%) of those between 14 and 24 who encountered hateful content have at least once reported it to a platform, while almost none of them indicated to have ever reported it to LEAs [67].

In light of such findings, the factors influencing youth's willingness to report are of interest. Surveying youth from eight European countries, Naderer et al. [75] discovered that individual factors like higher media literacy or past experience with online harassment correlate with a higher intention to report. Furthermore, DeSmet et al. [25] found that Belgian adolescents' willingness to engage in pro-social cyberbullying bystander behavior, which includes reporting, is particularly dependent on contextual factors like social status and group membership as well as social cues from peers and authority figures. Finally, studies in the United States (US) have established that the sanction-oriented retributive justice approaches of most social media platforms do not sufficiently meet the expectations and needs of young victims of online harms [90, 111]. They show that approaches of restorative justice, which center on victims and prioritize rehabilitation and community involvement [16, 43, 111], show great potential for addressing online harms involving youth.

2.2 Platform-Based and Platform-Independent Reporting of Harmful Content

Large social media platforms usually provide complex content moderation systems that combine automated and human moderation approaches [36, 38, 85]. Often they involve users by providing the option to flag, i.e., report, content that might violate community guidelines, which is subsequently reviewed and, if a violation is found, sanctioned [21, 88]. As a finite number of human moderators cannot match a rapid growth in users and content, many platforms automate moderation tasks, including the review of reported content [21, 89, 95]. Research has revealed systemic deficits in platform-based reporting mechanisms. It was found that some discourage reporting through manipulative dark patterns [108]. Transparency deficits also constitute a major issue. Often there is no communication of the time frame in which reports will be reviewed and on some platforms, users don't receive notifications on reports' outcome [99]. Inconsistencies in sanctioning practice, which exist both within and between platforms [88, 95], can further create an impression of biased rule enforcement [41]. Finally, reporting mechanisms of platforms have been found to offer insufficient information on victims' legal rights, support or counseling services, and possibilities to involve LEAs [3, 99].

Some empirical studies examine adults' perspectives on platform-based reporting. Myers West [73] surveyed users that were sanctioned and discovered a tendency to attribute content removals to human intervention, e.g., reporting, which is reinforced by inadequate communication of underlying reasons and decision-making processes. Other studies focus on those submitting reports. Kou and Gui [62] have shown in context of online gaming that reporting is met with mistrust but nonetheless used, including for non-intended purposes. In a survey experiment with YouTube users, Kim et al. [59] further found that the willingness to report sexist hate speech is increased by the presence of counterspeech, but this effect varies by its articulator's gender and number of upvotes. Finally, Zhang et al. [113] interviewed social media users that recently reported to examine their motivations, mental models, and concerns. They discovered that users report due to an desire to protect oneself and others, social pressure or encouragement by the personal environment, an perceived obligation to maintain a clean online environment, or dissatisfaction with content. They further uncovered that reporting activity is often discouraged by perceived inaction, unclear communication, and the cognitive burden of submitting reports, and thus recommend to improve procedural transparency and feedback.

In a number of countries, hateful content can not only be reported on the respective platform, but also to platform-independent actors. As in some jurisdictions severe instances of hate can be subject to criminal law [18], there is the option of filing reports to LEAs. In some countries, e.g., Germany, LEAs provide online channels for this, which are typically realized as web-forms [80]. Beyond that, non-profit hate speech reporting centers exist in Germany and several other European countries [44, 77]. They are either run by non-governmental organizations (NGOs) or authorities and complement platforms' content moderation ecosystems and LEAs in their efforts to combat online hate [104]. As victim-centered organizations, they typically accept reports via e-mail or self-administered web-portals and act as intermediaries between reporting individuals and other relevant actors [15, 80]. If they assess that content violates a platform's community guidelines, they forward it to the operator for deletion, while content deemed to be criminally relevant is submitted to LEAs [104]. If there is demand, some also provide free counseling [97]. Only few studies have engaged with these centers or LEAs in context of platform-independent reporting. Patz et al. [80] conducted interviews with employees of three centers to analyze countermeasures to hate speech in Germany. While Demus et al. [23] cooperated with one German center to develop a dataset of hateful X/Twitter posts, Bäumler et al. [14] analyzed interviews with staff of German centers and LEAs to create a domain-specific hate speech classification scheme. With regard to reporting hate crime in offline contexts, Gatehouse et al. [34] found that some LGBTQIA+ youth from the United Kingdom are hesitant to approach LEAs, as reporting implies victimhood.

2.3 User-Centered Tools to Support Victims and Bystanders of Hate and Harassment

HCI developed various user-centered tools to support victims and bystanders of hateful and harassing content. Some assist social media users in the identification of such content. Modha et al. [71] developed a browser plug-in that leverages artificial intelligence (AI) to detect and visually highlight textual online aggression on Facebook and X/Twitter. Another browser plug-in likewise detects hateful textual content on Reddit, but then hides it in real-time to reduce exposure [11]. There are also tools to support reporting and documentation. One German reporting center offers a reporting app for mobile Android and iOS devices [45], but its development was not accompanied by user-centered research. However, targets of harassment have been involved in the design of tools for evidence documentation. Informed by a survey and interviews with Bangladeshi victims of gender-based online harassment, Sultana

et al. [98] developed a tool for documenting authentic evidence on gender-based online harassment. Their browser plugin enables the creation of screenshots with metadata and the publication of incidents in a dedicated Facebook group. Goyal et al. [39] have instead analyzed the needs and challenges of female harassment victims on social media by conducting a focus group and interviews with journalists, activists, and employees of NGOs. On that basis, they developed and evaluated a prototype that allows for the documentation of content and the creation of reports that can be downloaded or shared.

Beyond that, research explored how tools can empower victims and bystanders through the provision of informational resources and support. In this regard, a recent strand of user-centered research investigated the use of AI-based chatbots, e.g., for educating adolescents about cyberbullying with bystander role-play scenarios [47], facilitating victims in reporting sexual harassment on platforms [99], and offering mental health support [63]. Technology-enabled peer support networks like Heartmob [10], Squadbox [70], or TrollBusters [28] constitute an alternative, as they connect harassment victims with volunteers that provide emotional or practical support. Finally, there is work on technical interventions to facilitate victims' coping processes. To et al. [103] involved BIPoC in the design of storyboards for tools to support victims before, during, and after racist interactions. Reid et al. [83] instead developed in-game tools with targets of toxicity that provide emotional support, e.g., through a provision of friendly messages or cute animal pictures.

2.4 Research Gap

Our study advances HCI research on the reporting of online harms and youth's handling of hateful content with an investigation into how both platform-based and platform-independent reporting systems can be designed in a youth-sensitive manner. It is situated at the intersection of three research gaps:

First gap: Reporting barriers of youth. While individual [75] and contextual factors [25] that influence youth's willingness to report have been identified and HCI research found that they perceive restorative justice approaches as particularly suited to address online harms [90, 111], it has not been investigated which barriers related to reporting systems and their providers obstruct hateful content reporting by this demographic.

Second gap: User requirements on reporting systems. Previous works have revealed systemic deficits of platform-based reporting mechanisms, including the use of dark patterns [108], insufficient transparency [99], and inadequate victim support and legal information [3, 99]. Other research has empirically examined user perspectives on them, both amongst those affected by sanctions [73] and those reporting [59, 62, 113]. While some of these studies derived implications for their design, we are not aware of any work in HCI that surveys prospective users directly about design requirements.

Third gap: Human perspectives on platform-independent reporting. In HCI, there is user-centered work on the design of tools to support victims and bystanders of hate or harassment with an automated identification of content [11, 71, 81], a provision of information and support through chatbots [47, 63, 99] or peer-support networks [10, 28, 70], a promotion of coping processes [83, 103], and a facilitation of documentation [39, 98]. By contrast, user perspectives on the design of tools to support the platform-independent reporting of hateful content have not been explored.

3 Research Design and Method

By addressing these research gaps, we seek to contribute to the state of research with insights on youth's barriers in relation to reporting hateful content (C1), their requirements on platform-independent reporting tools (C2), and heuristics for the design of youth-sensitive and inclusive reporting systems (C3). In order to answer our research questions, we employed a qualitative research design comprising eight FGIs with German adolescents and young adults and a subsequent qualitative content analysis.

3.1 Data Collection: Focus Group Interviews

3.1.1 Study Procedure. Due to the exploratory character of our research, we decided to conduct FGIs. Designed to facilitate open dialogue and discussion, they constitute an established method in HCI [68, 91, 94]. As they can yield a broad range of participant perspectives and opinions [68], they enable us to gather in-depth data on youth's perspectives in the context of reporting hateful content, including insights into individual needs and rationales. Since interactions within FGIs are particularly well suited for elucidating collective opinions [5, 64], we considered them conducive to the identification of barriers and requirements with relevance across individuals. Moreover, in contrast to individual interviews, youth participate within a more familiar setting, which can empower them to openly express opinions in their own language as they interact with peers [58, 82]. Finally, discussions can raise unanticipated issues and ideas [96], which suits our research objective.

From May to June 2024, we conducted and audio-recorded eight FGIs with German adolescents and young adults (N=47) in German language. Our intended group size was six to eight participants, which is recommended for this age group [20, 27, 64]. This was achieved with one exception (see Tab. 1). Seven FGIs were held on partner organizations' premises and one at our university. They were moderated by two researchers, lasted approximately 90 minutes, adhered to methodological recommendations [1, 20, 64], and followed a structured guideline with standardized language. Fig. 1 shows typical interview situations. As our duration is at the upper end of recommendations for youth [27, 58], we followed advice to adopt a stimulating and engaging interview design [20, 58, 64]. We developed a multi-step interview procedure (see Sec. A.1) with a variety of stimuli (S1-S5; see below and Sec. A.3) to ensure that the participants are introduced to the complex and emotionally demanding topic of our study in an age-appropriate manner, while maintaining their attention and establishing a common baseline of knowledge. We incorporated both direct questions and indirect prompts to encourage an articulation of perceived barriers and requirements. To guarantee age-adequacy, we consulted teachers and youth social workers in advance. In addition, we conducted a pre-test with five young adults that was followed by a discussion of deficits and resulted in a final revision of the procedure.

All FGIs shared the same structure. After an introduction and the collection of demographic data (see Sec. A.2) and declarations of informed consent, we showed a video introducing cyberbullying

Figure 1: Photos of typical interview situations from two FGIs. The participants' faces were blurred to protect their identity. Participants approved a blurred use for publication.

and hate speech (S1) and moderated a discussion about social media preferences (1). Then, we presented response strategies to hateful content (S2) and initiated a discussion about them (2). This was followed by a discussion about reporting experiences and barriers (3). Stimulated by a video on a reporting center (S3) and cards with potential reporting solutions (S4), participants further talked about their technology preferences (4) and design requirements in context of platform-independent reporting (5). Finally, participants familiarized themselves with the web form of the reporting center Hessen gegen Hetze (S4), submitted a dummy report in groups of two[3], and discussed suggestions for improvement (6). Two stages (2, 3) focused on barriers (RQ1), while two (4, 5) focused on requirements (RQ2). Stage six covered both. Thematic time allocation was thus balanced.[4] Insights from all stages contributed to the design heuristics (RQ3).

In FGIs with minors, effective moderation is key to ensuring inclusive participation. To foster engagement, discussions began with an icebreaker question, enabling contributions without extensive reflection [64]. We also encouraged rather quiet participants by inviting them to express opinions through (dis-)agreeing with others. Furthermore, we used non-verbal cues, such as eye contact and nodding, and emphasized the value of dissenting opinions to encourage participation, while respecting preferences to remain silent [22, 64]. All participants contributed, but not to the same extent. In most FGIs, particularly motivated participants with more speaking time emerged, whose opinions were frequently echoed by others. While our measures cannot fully compensate for limitations concerning social desirability and dominating voices [22], participants nonetheless articulated differing views on various issues. We did not intervene during interactions, but occasionally asked follow-up questions or encouraged further contributions before transitioning to the next topic.

3.1.2 Recruitment and Participants. To gather perspectives of German youth with diverse backgrounds, we followed a convenience sampling approach by establishing cooperations with three organizations situated in the city of Darmstadt, which has more than 150,000 inhabitants and is located in a metropolitan area in western Germany. This has implications for our results' validity (see Sec. 5.4). Nonetheless, we considered a metropolitan context particularly conducive to recruiting diverse youth. Given the potentially distressing themes of our research, alternative sampling approaches were not practicable since we could only recruit minors through organizations with whom we were able to establish a relationship of trust. We could only establish such relationships with organizations located in the city of our university by leveraging personal contacts and previous partnerships. Altogether, 47 individuals participated in the study, distributed across eight FGIs (see Tab. 1). We cooperated with an NGO advocating for the rights of LGBTQIA+, an integrated comprehensive school whose students are mixed with regard to their intended school leaving qualifications, and a youth center that is popular among youth with a migration biography. After representatives of these organizations expressed interest, we met them to clarify organizational details. Subsequently, they recruited youth that visit their organization. We involved the integrated comprehensive school to ensure that we cover diverse education levels and learning abilities. The specific classes were selected on the basis of time availability. Three FGIs were conducted with ninth graders (FGIs 2-4) and three with tenth graders (FGIs 5-7). Moreover, we

[3]The participants were provided devices (smartphones, tablets, laptops). They were not confronted with hateful content and disclosed no personal data.
[4]When scoring stage six for both, about 45 minutes were allocated to answering both RQ1 and RQ2 (see Sec. A.1). Thanks to moderation interventions in case of significant delays, we maintained our schedule across all FGIs with only minor deviations.

#	Type of Organization	Age Group	Venue	N
1	LGBTQIA+ organization	Young adults (18-29)	University	6
2	Integrated comprehensive school	Adolescents (14-17)	School	7
3	Integrated comprehensive school	Adolescents (14-17)	School	6
4	Integrated comprehensive school	Adolescents (14-17)	School	4
5	Integrated comprehensive school	Adolescents (16-17)	School	6
6	Integrated comprehensive school	Adolescents (16-17)	School	6
7	Integrated comprehensive school	Adolescents (14-17)	School	6
8	Youth center	Adolescents & young adults (14-21)	Youth center	6

Table 1: Overview of the FGIs with respective organizations, age groups, venues, and number of participants.

involved the NGO and youth center to ensure representation of youth that consider themselves as LGBTQIA+, BIPoC, or having a migration biography. We focused on adolescents between the ages of 14 and 17 (N=40). Since we were unsuccessful in recruiting queer adolescents through our partner NGO and other local organizations, we decided to instead involve young adults (N=7) to account for queer perspectives. Prior exposure to hateful content and reporting systems were no recruitment criteria, as barriers and requirements of those without such experiences were also of interest. During the FGIs, most youth (37) indicated prior use of platform-based reporting, while only one used platform-independent options.

Participant data was gathered with paper-based questionnaires. The average participant age was 17, with an average of 16 when excluding young adults. Regarding gender, 27 participants identified as female, 18 as male, one as non-binary, and one as demi-girl. The majority of participants (N=39) were students enrolled at an integrated comprehensive school, i.e., they have the option to pursue different secondary school leaving qualifications (Abitur, Mittlere Reife, or Hauptschulabschluss), while one participant visited a Realschule (Mittlere Reife as qualification) and one a vocational gymnasium (Abitur as as qualification).[5] One was already employed, while five were university students. Since we also wanted to assess whether we realized a diverse sample, participants had the voluntary option to disclose their affiliation to social groups. We explained potentially ambiguous terms like BIPoC, LGBTQIA+, and migration biography. Fig. 2 summarizes their responses. In the results section of this paper, participants' statements are referenced using individual identifiers (1a-8f). Detailed participant data and identifiers can be found in Sec. A.4.

3.2 Data Analysis: Qualitative Content Analysis

After transcribing and anonymizing the FGIs, we performed a structuring qualitative content analysis following Kuckartz [65]. As a common method in HCI (see, e.g., [54, 60, 110]), it allows a content-based systematization of data with an iteratively developed category system. Thus, it enables us to transparently elaborate participants' perceived reporting barriers and requirements on platform-independent reporting tools, while its mixed approach to category development, i.e., both deductive and inductive, matches the exploratory character of our research. Our analysis was guided by the categories *1 - Barriers* and *2 - Design Requirements*, each with subcategories for further differentiation. Whereas we established these categories deductively on basis of the research objective, subcategory development had both deductive and inductive aspects. We deduced initial subcategories from related work, which were then revised and supplemented by subcategories inductively derived during coding. All categories can be found in the coding scheme (see Sec. A.5).

We used the software MAXQDA 2024 for coding. Participants' statements constituted coding units. During coding, statements that match the definition and coding rules of a (sub-)category were assigned to it. A simultaneous assignment to several categories was possible. Prior to the first coding iteration, we familiarized ourselves with the data to check the adequacy of deductive subcategories, supplement initial inductive subcategories, and develop definitions and coding rules. Then one author coded all empirical material, amended the codebook, and selected examples for each category. After this, two authors coded two FGIs (1, 7) simultaneously. On this basis, we assessed intercoder agreement at the segment level with MAXQDA, which revealed a kappa coefficient following Brennan and Prediger [12] of 0.79.[6] This can be interpreted as a substantial result [66]. Nevertheless, we discussed improvement potentials, on the basis of which we revised and finalized the codebook. Eventually, one author coded all interviews in a final iteration. We then assembled all statements assigned to one subcategory using MAXQDA and analyzed them together.

3.3 Ethical Considerations

As our study involved minors, we were careful to ensure ethical integrity throughout the entire conception and execution of it. We obtained approval from our university's ethics committee (IRB Number EK 35/2024) in advance. During the creation of the interview guide and selection of stimuli for the FGIs, we coordinated with teachers and youth social workers from the involved organizations to include their practical expertise in working with youth. All study material was carefully tailored to the age of the participants, avoiding any potentially distressing content. The videos we used as stimuli were age-appropriate and introduced with a trigger warning [37]. The anonymity of the participants was guaranteed

[5] In the German school system, students can obtain different qualifications depending on the school type and duration of education. Within the 2011 International Standard Classification of Education (ISCED), Abitur is equivalent to ISCED 34 (higher secondary education, 12 or 13 years), while Mittlere Reife (10 years) and Hauptschulabschluss (9 years) are equivalent to ISCED 24 (lower secondary education) [76].

[6] We determined kappa with MAXQDA and followed the recommendations of Kuckartz and Rädiker [66]. We considered codings for segments a match if there was at least 99% overlap between the segments specified by both coders.

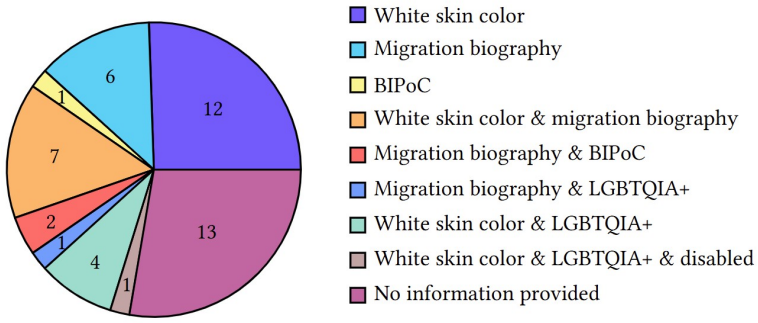

Figure 2: Voluntarily disclosed information about participants' affiliation to social groups.

during the entire study. In simplified language, we obtained explicit informed consent from participants and, if they were minors, their legal guardians in written form. Prior to the FGIs, we reminded participants that they can end participation at any time. Our moderation team was gender-balanced and we aimed to create a safe, comfortable, and respectful interview environment. During the FGIs, we provided participants with refreshments and afterward, they received a €15 book voucher as a token of appreciation and a curated list of counseling services.

4 Results

In this section, we present our qualitative findings from the content analysis of the FGIs. First we describe barriers that impede youth's platform-based and platform-independent reporting of hateful content (Sec. 4.1). Then we provide insights on their requirements on the design of platform-independent reporting tools (Sec. 4.2).

4.1 Youth's Reporting Barriers

Our participants described nine barriers that complicate or even inhibit the reporting of hateful content (see Tab. 2). We will first present those mentioned in context of platform-based reporting (B1-B4) and then proceed with those indicated in context of platform-independent reporting (B5-B9).

4.1.1 Platform-Based Reporting. Throughout the FGIs, participants outlined four barriers that, from their perspective, obstruct the platform-based reporting of hateful content. Based on our data, we cannot assess whether there is a factual basis for them. Some of them relate to reporting mechanisms themselves (B2, B4), while others relate to more abstract perceptions or attitudes in context of the platforms' handling of hateful content (B1, B3).

Across different FGIs, 16 participants indicated that an impression of a **deficient enforcement of community guidelines (B1)** discourages them from reporting on platforms. 4a has summarized her reasoning in this regard as follows: *"If I know that reporting doesn't help me, then I won't report."* Due to this impression, two respondents described platform-based reporting as *"mostly ineffective"* (1e) or *"relatively ineffective"* (1b). Youth considered it a serious issue that some platforms sanction hateful content not solely based on the presence of a rule violation, but also on the number of received reports (1e, 5c, 6c/f, 8b/f). As 1e described, this can result in frustrating reporting experiences: *"But on TikTok, you always get a notification that no violation was found, even when I clearly see a violation. I reported a video multiple times, and in the end, the account went private, but it was always 'no violation' for me."* Such frustration was shared by others (5c, 6c, 8f). To get hateful content against herself removed, one participant recalled that she even reported via different devices, without success: *"I kept trying to report it from my mom's phone, my siblings' phones, and my dad's phone, but it never worked" (8f).*

Inadequate or lack of feedback (B2) by platform operators was seen by twelve participants as an impediment to reporting. Several youth recounted that they had already experienced a complete lack of feedback after submitting reports on hateful content, either on social media (1a/c/e/f, 2e, 6d/f) or online gaming platforms (3b/d), which led them to question the general significance of platform-based reporting. Two participants considered it impractical that even though some platforms provide feedback on the assessment of reports, users have to proactively check for it because there is no explicit notification (1c, 8a). For 1c this is an issue, as he may not recall submitting a report: *"After I've reported something, I honestly forget quite quickly that I've done it. I can't keep track of whether something is happening."* Finally, some deemed the content of follow-up messages unsatisfactory (1f, 4a, 5b). Particularly negative feedback, i.e., that no violation was found, was perceived as too generic, as typically no reasoning is specified (4a, 5b). This left 4a frustrated, as it remained vague when platforms will consider hateful content a violation: *"It just says it's not against the community rules, that doesn't help me much."*

A fundamental barrier of some youth towards platform-based reporting is **distrust in platform operators (B3)**. Eight participants questioned their general commitment to tackle hate. For instance, 7f highlighted the prevalence of hateful and violent content on Instagram, criticizing a lack of response: *"There are so many accounts that just keep posting this stuff, and you don't feel like Instagram is doing anything about it."* Additionally, for three participants the suspected use of AI by platform operators erodes their trust (4a/b/c). They stated that they view assessments of reported content as superficial, assuming that an algorithmic analysis without human involvement is conducted. This can create a feeling of not being taken seriously: *"Why waste time if you know that TikTok won't respond, that they don't care?"* (4a). Finally, one adolescent voiced her suspicion that platform-based reporting merely serves to create

	#	Barrier	∑	Participants
Platf.-based	B1	Deficient enforcement	16	1a/b/e/f, 2d, 4a/b, 5a/b/c, 6c/f, 7f, 8b/c/f
	B2	Inadequate or lack of feedback	12	1a/c/e/f, 2e, 3b/d, 4a, 5b, 6d/f, 8a
	B3	Distrust in platform operators	8	1e, 2d, 3a/b, 4c/d, 7f, 8c
	B4	Standardized incident categories	6	1b/e, 4c, 6a/f, 7f
Platf.-indep.	B5	Unawareness of reporting options	18	1a/b/c/d/e/f, 2b/e, 3a/d/f, 4a, 5a/c, 6a, 7e/f, 8c
	B6	Disruption of social media use	17	1a/b/d/e/f, 2c/e/f, 3c, 4d, 5c, 6a/b/e, 7b, 8b/e
	B7	Time-consuming reporting procedure	13	1a/b/c/d/e, 2c/e/f, 3c, 4a, 6b, 8c/e
	B8	Distrust in law enforcement	12	1a/b, 2f, 3d, 6b/c/f, 7b/d/f, 8c/f
	B9	Complicated generation of URLs	4	1a/c/f, 8e

Table 2: Overview of the reporting barriers described during the FGIs with frequency of mention by participants (∑). They are differentiated by their relation to platform-based or platform-independent reporting.

a sense of security: *"I think social media offers these options so we feel safe here, but usually nothing happens"* (8c).

A barrier that is directly related to the current design of many platform-based reporting mechanisms is the possibility to only report **standardized incident categories (B4)**. Six participants stated that they feel restricted by the often highly standardized reporting interfaces. Lists with pre-determined incident categories, which only allow for the selection of one option, were viewed particularly negative in this context, as youth struggled to accurately categorize individual instances of hateful content. For example, 1e felt that individual content may be covered by several categories simultaneously: *"Sometimes I couldn't quite categorize something in a comment that I just wanted to report. Whether it was hate or harassment or something else ... somehow several things fit or it was something in between."* 6f generally perceived the amount of available categories confusing and would have preferred an option to briefly describe the incident herself, while both 7f, with regard to TikTok, and 1b, with regard to Instagram, recalled situations where they were dissuaded from reporting because the available categories did not cover the concrete nature of the incident.

4.1.2 Platform-Independent Reporting. With regard to platform-independent reporting, youth highlighted five other barriers. These were raised both in light of individual experiences and in relation to the reporting systems presented and discussed as stimuli. As in the previous section, they are perceived barriers. Some relate directly to reporting systems (B6, B7, B9), while others are linked to general awareness (B5) or trust (B8).

Unawareness of platform-independent reporting options (B5) constituted a barrier for 18 participants. Only four acknowledged that they were familiar with concrete opportunities to report hateful content to LEAs or dedicated reporting centers prior to the FGIs (1e/f, 2b, 3d), and only 1f previously filed a report to such organizations. As 1a noted, *"it's an extra effort to find an external reporting channel, especially if you don't know one in advance"*. Given her unawareness, 1d feels discouraged by the effort required to identify a suitable option: *"There's a mental barrier to researching how to report something."* Another participant explains that for her, not only knowledge about a reporting option but also about the actual process would be important: *"If I knew exactly how it works, I would probably use it"* (4a). This procedural unawareness about reporting options is compounded by substantial unawareness about what types of hateful content could be reported. Several participants highlighted a perceived severity threshold when it comes to platform-independent reporting (1b/c, 2b, 3a/d/f, 5a/c, 6a, 7e/f). 1b stated that they would only report *"doxing, death threats, or calls for violence"*. Similarly, others would file reports only *"if it involved serious, ongoing bullying"* (3d), *"in very drastic cases"* (2a), *"in cases of explicit threats"* (7e), or *"in an emergency"* (3f). As 1b points out, this is related to uncertainty about the criminal relevance of hateful content: *"I'm not always sure if something is legally relevant or if reporting a general issue is adequate."*

For 17 participants, the **disruption of social media use (B6)** that is necessary to proceed with platform-independent reporting, which is caused by a missing integration of such options on social media platforms or apps, introduces significant friction. Some youth highlighted that they see social media usage as an important leisure activity, where they want to relax while consuming content and prefer not to be interrupted (1c, 2c, 5c). While 8b views the necessity to interrupt social media usage as merely *"a bit annoying"*, 6e argues that it could discourage from reporting at all: *"If you don't feel like switching websites and stuff, you just let it be."* Particularly when using social media apps on mobile devices, the necessity to switch to reporting forms on the web browser is seen as disruptive (1b, 4d, 5c, 8e/f). 1b summarizes their reservations as follows: *"If there would be a way to access the form directly from the app I'm using, like how you report things on the platform itself, I would use it. But going to another browser, searching, opening, and then copying the link from one to the other is just too much."*

The **time-consuming reporting procedure (B7)** of platform-independent services constituted a significant barrier for 13 youth. A comment by 5c summarizes this general sentiment: *"I wouldn't use anything that takes too much time."* More concretely, several participants emphasized that they felt that platform-independent reporting involved significantly more steps than platform-based reporting (1b/d, 3c). Using the example of TikTok, 3c illustrates this issue: *"When I report on TikTok, I press a button, select the reason, and it's done. But with external reporting, I have to take screenshots and go to another website — it takes too long."* Similarly, 1d argues that the reporting processes presented during the FGIs comprise *"too many steps"* and come with *"quite a big effort"*. Moreover, 2c emphasized that she has parental restrictions on daily phone use and does not want to spend her limited time on lengthy reporting:

"It's a waste of time. I have limited time on my phone, and if I have more time, I use it for myself, not to report."

During the discussions, **distrust in law enforcement (B8)** emerged as a significant barrier among twelve participants. Two LGBTQIA+ youth expressed fundamental distrust in LEAs (1a, 1b). *"I don't have much trust in the authorities"* stated 1b and explained that this implies avoiding any non-essential interaction with them. Similarly, 1a preferred to not involve LEAs unless absolutely necessary and argued that knowing that a report might trigger a criminal investigation without her approval would deter her from reporting. A recurring theme among younger and particularly female youth was that they expect to not be taken seriously by LEAs when filing a report (2f, 3d, 6f, 7b/d/f, 8c/f). 7f argued that youth's concerns about hateful content get often dismissed by adults: *"Often grown-up people don't take us seriously. Even if you would talk seriously about something important, they would still say we're only 16, it's just fun."* 8c described how individual officers did not take her seriously when she approached police about a case of online harassment. Other participants further questioned the ability of LEAs to adequately respond to online hate (1a/b, 6b/c, 8f).

Finally, four participants emphasized that the **complicated generation of URLs (B9)** that link to hateful content could obstruct them from platform-independent reporting. They argued that they are challenged by the requirement to submit URLs to individual content, e.g., comments or posts, or offender profiles, as they do not know how to retrieve them, particularly on mobile apps. 1c stated that he had previously experienced this issue with Instagram: *"Once a link is involved, I would find it burdensome because I don't understand how to generate links on Instagram."* As there exist considerable differences, familiarity with the functionality on individual platforms is only of limited help: *"What I can also imagine to be difficult, is to really find a link ... especially since the problem can arise that you don't know exactly how to get to a link on every platform"* (1f).

4.2 Youth's Requirements on the Design of Platform-Independent Reporting Tools

Based on their previous experiences and their engagement with different reporting processes and solutions during the FGIs, the youth articulated eleven requirements for the design of platform-independent reporting tools (see Tab. 3). While five relate to the reporting process (R1-R5), three focus on ensuring feedback and transparency (R6-R8), and three concern the provision of additional support and information (R9-R11).

4.2.1 Reporting Process. 23 participants emphasized that an ideal reporting tool should **minimize effort through intuitive navigation and a streamlined reporting process (R1)**. Well designed interfaces were described to be *"simple"* (2b, 7b/e), *"uncomplicated"* (6b), and *"clear"* (7f). Youth would like to know quickly which input should be entered in which fields and whether it is mandatory or not (3b, 4a/d, 6a, 8b). For the latter, 8b considered the use of asterisks practical: *"Perhaps a quick explanation above saying that those fields with asterisks are mandatory."* The reporting process should be *"relatively short and concise"* (6a). Such expectations were emphasized by several participants (1a/b, 2e/g, 3c, 5a/c, 6a/b/f, 8f), with 6b stating *"less than a minute"* and 6f *"two, three minutes"* as an optimal time frame. To realize this, it was suggested to minimize steps and limit compulsory input to a minimum (1b, 2e, 5a/c, 6a/b).

Many youth expressed concerns about providing personal data during reporting. Accordingly, 20 participants articulated the requirement that tools should **allow the anonymous submission of reports (R2)**. Youth viewed a mandatory input of personal data critically, describing it as *"superfluous"* (6b) or *"unnecessary"* (6a). Meanwhile, 5d argued that *"it would be good if you could decide whether you want to stay anonymous or be known"*, and 4a stated that *"it should definitely be optional so everyone can decide for themselves"*. Youth gave different rationales for this preference. While 5c argued that he wants to protect himself from retaliation by offenders, 5a primarily wanted to avoid personal responsibility after reporting: *"If it's just someone I know from social media ... I'd prefer to stay anonymous to avoid any responsibility or getting involved."*

For seventeen participants, it was important that user interfaces of reporting tools **enable the display of information in foreign or simple language (R3)**. Youth emphasized the importance of making the platform accessible to non-German speakers, suggesting that at least English should be offered (1b/e, 4a, 5c, 7f). To further enhance accessibility, some also advocated that additional languages commonly spoken by individuals with migration biography living in Germany should be offered (7d/f, 8a/e). Beyond that, 7b thinks that *"any language would be useful"*. In addition, some participants argued that information should be displayed in a way *"which is accessible to as many people as possible"* (1f), or that there should at least be an option to display it in simple language to accommodate users with varying cognitive abilities (1e).

Regarding the interfaces of reporting tools, 13 participants expected them to **provide diverse input options for incident data (R4).** Specifically, there was a preference for a combination of structured and unstructured options. Some youth found structured options, such as checkboxes or drop-down lists, particularly convenient (1e/f, 3b). One particpant (1e) explained that for them, filling out free-text fields comes with higher cognitive load: *"This description of the incident, what exactly I'm reporting, would probably be too much effort for me ... because putting thoughts into words is sometimes difficult."* However, others valued their higher flexibility, as they allow for a detailed description and a reporting of incidents not adequately covered by pre-defined categories (3d, 6a, 7f, 8c). This way, *"everyone can simply express their feelings, follow their own thoughts, instead of just checking boxes based on what others think is serious or not" (8c).* As 1a summarizes, both approaches could complement each other: *"It would be good to have some guidance on how to categorize it thematically ... but also the opportunity to write something yourself. In other words, a mixture of both."*

To encourage reporting, ten respondents suggested **leveraging gamification elements (R5).** In this regard, some envisioned using some kind of reward system with incentives for reporting (1a/b/e, 3f, 4c). For instance, 1e argued that *"it would be funny if you could collect points ... and after a certain number, you could get a voucher or something else"* and 3f remarked that rewards *"would motivate me to rather report something instead of ignoring it"*. Youth highlighted that such rewards could be quite simple, e.g., a picture of *"a cute mouse that looks happy"* (1b), *"something to laugh about"* (3f), or *"a sticker"* (1e). However, it was also stressed that gamification elements should

#	Requirements	∑	Participants
	Reporting Process		
R1	Minimize effort through intuitive navigation and a streamlined reporting process	23	1a/b/d, 2b/e/f/g, 3b/c, 4a/d, 5a/c, 6a/b/e/f, 7b/d/e/f, 8b/f
R2	Allow the anonymous submission of reports	20	1a/c/e, 2b/e, 3b/d, 4a, 5a/c/d, 6a/b/d, 7b/d/f, 8b/e/f
R3	Enable the display of information in foreign or simple language	17	1b/e/f, 2e, 3a/b/d, 4a, 5c, 7b/d/f, 8a/b/c/e/f
R4	Provide diverse input options for incident data	13	1a/e/f, 2a, 3a/b/d, 4d, 6a/f, 7f, 8b/c
R5	Leverage gamification to encourage reporting	10	1a/b/d/e, 3d/f, 4c, 6a/b/f
	Feedback & Transparency		
R6	Provide feedback on successful submission, assessment outcomes, and initiated actions	25	1c/d/e, 2e, 3c/d/f, 4a/d, 5a/b/c/e, 6a/c, 7b/c/e/f, 8a/b/c/d/f
R7	Allow for a customization of feedback	10	1a/c, 2d/e, 5c, 8a/b/c/e/f
R8	Disclose information on legal ramifications and data transmission	8	1a/f, 4d, 5c, 6a, 7e/f, 8e
	Additional Support & Information		
R9	Facilitate contact with psycho-social counseling as well as emergency services	20	1a/b/d/e/f, 2d/e, 3d, 4a/d, 5c/d, 6a/f, 7c/d/e/f, 8e/f
R10	Assist evidence documentation through a provision of detailed instructions	5	1d/e/f, 3b, 8e
R11	Provide information resources on hateful content and the applicability of criminal law	4	1d, 4d, 5c, 7d

Table 3: Overview of the design requirements for platform-independent reporting tools articulated during the FGIs with frequency of mention by participants (∑).

be designed in a way that does not encourage the transmission of unsubstantiated reports (1a, 1d, 3d, 6a).

4.2.2 Feedback and Transparency. Among youth, there were also requirements relating to the post-submission stage. 25 respondents expected reporting tools to **provide feedback, particularly on successful submission, assessment outcomes, and initiated actions (R6)**. Some argued that knowing reports' outcome reinforces the feeling that reporting can have meaningful impact, thus enhancing their motivation (1e, 2e, 3d). In this regard, 1e explained: *"What would motivate me more is if I could see that it really makes a difference. If I feel like nothing is happening, then I won't report."* Youth not only wanted to receive a notification about successfully submitting a report (2e, 3c/d, 4d, 5b/c, 6a, 7b, 8a/f), which should acknowledge their effort (1c, 6c, 7f, 8e), but also about the assessment outcome, including the reasoning behind decisions (3c, 4a, 5a/c/e, 6c, 7c/e, 8d), and about the actions taken in response (1e, 2e, 3c/d/e, 4a/d, 5b/c, 6a, 7b/e, 8b). 3e argues that the latter should include information on imposed sanctions: *"I'd like to know if they were suspended or something like that"* (3e). 3c summarized his expectations as follows: *"The best would be if they sent a message to inform whether the report was accepted, whether they are working on it, or even if they rejected it. I want to know if they said no, it's too little, or yes"* (3c).

Ten participants principally welcomed transparency, but suggested that reporting tools should **allow for a customization of feedback (R7)**. 1c argued that he would not like to receive feedback e-mails for every reported incident, and therefore viewed customizable notification settings, e.g., before report submission, as the *"ideal"* solution. 1a and 8e shared this view. Others had similar opinions with regard to tool-internal notifications (8a/c/f). 8c shared his thoughts on this: *"It would also be nice to have an option to adjust notifications, or even turn them off entirely. It can be really annoying when you're constantly being notified, especially when you don't want to be reminded of what you have reported."*

Since platform-independent reporting options are either provided by LEAs or there is at least a possibility of their involvement, eight participants expected that reporting tools **disclose information on legal ramifications and the transmission of data (R8)**. Youth emphasized that prior to submission, they expect to receive unambiguous and transparent information about the organizations that may receive case-related data, as well as the purposes for which it will be used (1a/f, 4d, 5c, 6a, 7e). For instance, 1a stated that she would be hesitant if LEAs could be involved: *"If I knew that it could lead to a criminal complaint, I would probably refrain from reporting ... therefore I believe that it has to be clearly indicated whether a criminal complaint may be filed or to whom the report may be forwarded."* This opinion was shared by 1f, who thinks this information should be provided proactively during reporting. In addition, two participants voiced an expectation to not only receive information about possible consequences for the perpetrator, but also about potential legal ramifications and obligations for themselves (7f, 8e).

4.2.3 Additional Support and Information. During the FGIs it became apparent that age-appropriate support can be essential to accommodate the needs of young victims of hateful content. This is particularly demonstrated by the requirement that reporting tools should **facilitate direct contact with psycho-social counseling as well as emergency services (R9)**, which was articulated by 20 participants. Several youth emphasized the need to make reporters aware of specialized psycho-social counseling services (1a/b/d/e/f, 2d/e, 3d, 4a/d, 5c/d, 6a/f, 7c/d/e, 8e/f). Beyond that, some argued that they should further be provided contact information of organizations that provide support in emergencies, e.g., the police or crisis

hotlines (1a/b, 6a/f, 7d/f, 8e). These services should be available for the user's geographic location (7f) and ideally be available around the clock (1f). As regards implementation, twelve youth recommended that the interface should explicitly highlight support and emergency services during the reporting process (1a/b/d/e, 2d/e, 3d, 4d, 6a/f, 7d, 8e), e.g., automatically when submitting a report (1e, 6a, 8e) or after pressing a prominently marked button (3d). In terms of content, services' scope and contact information, e.g., website URLs, e-mail addresses, or phone numbers, should be provided (1d, 6a/f). 6f envisioned that ideally, users could *"just tap on a number and connect"* with them. There were controversial opinions as to whether chatbots could be used to provide or mediate support. While 8e saw benefits in their provision of *"quick responses"*, 4a found that they tend to *"say the same thing over and over, which isn't very helpful"*. Likewise, 7d considered such a solution inadequate, because *"it is not a human, it does not know what feelings you have"*.

As the provision of URLs or screenshots as evidence can be challenging, five participants would like reporting tools to **assist evidence documentation through detailed instructions (R10)**. Some youth described difficulties in accessing or generating URLs that link directly to individual hateful content or perpetrators' profiles, particularly on mobile social media apps (1e/f, 8e). This can not only discourage reporting but also hinder the assessment of submitted reports if data is incomplete or incorrect. To address this issue, participants suggested that reporting tools might display distinct icons next to input fields that can be used to access additional instructions (1e, 3b, 8e). These could be tagged with a question mark (1e, 8e) and link to a *"page, where it is explained again for each platform how to copy or generate a link"* (1e). In addition, 1d and 1e identified demand for similar assistance regarding the creation and submission of screenshots. 1e could imagine that *"there is a question mark on the side of the screenshot upload and if you click on it an explanation appears. ... You then get pictures, step-by-step explanations, or videos"*.

Beyond that, four participants indicated that they would like reporting tools to **provide information resources on hateful content and the applicability of criminal law (R11)**. To support their decision-making in relation to reporting, some youth considered it helpful to be offered information that comprehensibly explains what qualifies as hate speech and other forms of harassment (1d, 7d). Furthermore, some requested information explaining under what circumstances content could be criminally relevant (4d, 5c, 7d). 1d felt that such information should not be displayed upfront in the user interface, but rather in *"a separate view with further links, with information and articles that you can look at yourself"*. Regarding the presentation of information, 4d articulated some preferences: *"It depends how it is structured. If a video is really long and it keeps going on about other things, then I'd rather look at a text. But if a text is five pages long, I'd rather watch a five-minute video."* Finally, 5c and 7d had the idea that AI, e.g., implemented within a chatbot, might be leveraged to give a preliminary legal assessment of content and answer follow-up questions.

5 Discussion

Youth are particularly at risk of encountering hateful online content [9, 57, 67], yet research in HCI on reporting systems has largely overlooked this demographic. This focus group study with German adolescents and young adults explored **how reporting systems for hateful content can be designed in a youth-sensitive manner**. In this section, we first discuss German youth's reporting barriers (Sec. 5.1) and requirements for platform-independent reporting tools (Sec. 5.2). Thereby, we highlight similarities and differences to other contexts, provided that corresponding findings are available. Then, we synthesize our insights with those from related work to derive design heuristics for youth-sensitive and inclusive reporting systems (Sec. 5.3). Finally we outline this study's limitations and opportunities for future work (Sec. 5.4).

5.1 Youth's Reporting Barriers

In response to RQ1, we contribute with empirical insights on nine barriers discouraging German youth from reporting hateful content (C1), thus extending prior work that instead focused on individual or contextual factors influencing youth's willingness to report [25, 75]. Whereas Zhang et al. [113] have established that perceptions of insufficient community guideline enforcement as well as inadequate transparency on report handling can dissuade adults living in the global north from platform-based reporting, our findings suggest that this also applies to German youth. Furthermore, we not only found that they distrust platforms' content moderation, which is in line with findings on US youth in context of harassment [90], but also that this can constitute a reporting barrier. Another finding that was not reported in work with adults and may thus be youth-specific is that the necessity to classify incidents into standardized categories can discourage reporting.

Beyond that, our study is the first to provide insights into youth's barriers to platform-independent reporting of hateful content. We found that German youth are discouraged from using such offers due to the necessity to disrupt their social media activities, time-consuming reporting procedures, and difficulties in generating URLs for evidence documentation. Bangladeshi women voiced similarly discouraging experiences when reporting online harassment to LEAs [98], and some North American and European adults also perceive reporting on platforms as burdensome [113]. While these barriers are thus not youth-specific, they could be particularly deterring for them. As was shown for young adolescents from various countries [19], and pointed out by one participant, youth may have parental-enforced time limits for internet and device usage. This may further reduce their readiness to engage with a lengthy, disruptive, and complicated reporting process. In addition, procedural unawareness about suitable reporting options and substantial unawareness about 'reportable' content, as well as distrust in LEAs constitute profound barriers for our participants. The first finding is consistent with previous research showing inadequate awareness of legal rights and actions in response to harassment among women from Bangladesh [98] and English-language TikTok users [3]. Distrust was further identified by Gatehouse et al. [34] as a barrier to hate crime reporting among queer youth in the United Kingdom. Across various countries, this also applies to adults [106]. We found that such distrust additionally deters some German youth from reporting to civil society based centers. As such organizations collaborate less intensively with LEAs in other countries, this could be Germany-specific.

5.2 Youth's Requirements on Platform-Independent Reporting Tools

As a second empirical contribution, we identified eleven requirements of German youth on platform-independent reporting tools (C2), thus addressing RQ2. Whereas the HCI community has developed a variety of user-centered tools to support victims and bystanders of online hate and harassment (see Sec. 2.3), our work is the first to empirically examine what features prospective users expect from platform-independent reporting tools. Our findings demonstrate a need to design reporting tools that minimize required workload, allow for maintaining anonymity, and accommodate youth's differences in language competences, cognitive abilities, and input preferences. This mirrors expectations of adults from other countries in different reporting contexts. In a study on supporting female journalists and activists in the documentation of online harassment, the significance of reducing workload was also emphasized [39], and reporters on platforms highlighted the importance of protecting their own identity from other users [113]. In this regard, our results reveal an interesting nuance. Some German youth emphasized that maintaining anonymity is additionally important in relation to the recipients of reports. In both previous studies, adults further advocated an optional free text field to contextualize reports [39, 113]. Going beyond that, youth in our study specifically requested that the often mandatory categorization of content into pre-defined categories should be optional.

Furthermore, youth argued that tools should provide customizable feedback on both assessment processes and outcomes, be transparent with regard to potential legal ramifications and data transmission, and educate users about hateful content and criminal liability. Studies with individuals from different countries show that transparency about platforms' handling of reports is an expectation across ages [4, 62, 113]. What stands out among our young respondents, however, is their request for an option to customize feedback on reports, both in terms of its overall provision and extent. Meanwhile, demands to be informed of reporting's legal ramifications and contents' criminal relevance within tools may be explained by the German context, in which non-governmental reporting centers collaborate with LEAs and hateful content can be subject to various criminal norms [14]. However, relevance for countries with similar regulatory frameworks is conceivable.

Youth further suggested that tools should help in approaching additional psycho-social counseling and emergency services. This corroborates previous findings from the US that the support needs of young victims of online harms extend well beyond retribution and entail the provision of emotional support and validation [4, 90, 111]. What was viewed favorably among several youth and has not yet been established in research with adults is the use of gamification elements, e.g., a provision of simple non-monetary rewards, within reporting tools to encourage reporting and improve one's mood. In particular, it was suggested to show eye bleach pictures, i.e., images of cute animals or memes, following the submission of reports, which Reid et al. [83] have already found to be effective in providing emotional support in the context of online gaming, particularly for women. However, in such serious contexts, gamification should be approached with caution, which was also emphasized by some participants. Finally, some youth articulated a need for assistance in evidence documentation. Similar needs among adult women from both the global south and north that are targeted by online harassment have informed the user-centered design of evidence documentation and sharing tools [39, 98]. This design knowledge should also be considered when developing tools for youth-sensitive reporting.

5.3 Heuristics for Designing Youth-Sensitive and Inclusive Reporting Systems

As a third contribution and in response to RQ3, we derive five design heuristics (D1-D5) for youth-sensitive and inclusive reporting systems (C3). Design heuristics represent simple and practical design-oriented rules that, unlike requirements that specify system goals and functionalities in a specific setting, are generalizable to a class of technologies [26, 86]. As they typically reflect accumulated design knowledge [26, 56, 86], we chose to derive them by consolidating our findings with insights from related work. They are meant to guide researchers and designers in developing and improving reporting systems. Thus, we also discuss initial recommendations in light of currently available solutions.

Ensure simplicity by minimizing mandatory user input, featuring an intuitive interface design, and clearly communicating all workflow steps (D1). With regard to available platform-independent reporting tools, youth considered the necessity to interrupt social media use (B6) and the time required for reporting (B7) as discouraging and therefore formulated the expectation that technical solutions should minimize effort through intuitive navigation and a streamlined process (R1). Similar requirements have already been articulated with regard to in-game support tools for targets of toxicity [83] and documentation tools for online sexual harassment [98]. In this respect, some platform-independent offers provide commendable features. To explain required input without overloading the user interface, individual German reporting centers display additional information after clicking easily comprehensible icons, e.g., question marks [13, 48]. Another links to a guide for creating court-proof screenshots [46]. Innovative solutions could further allow for platform-independent reporting without interrupting social media use. This may include plugins that, like the tool by Sultana et al. [98], semi-automate screenshot and URL capture. Alternatively, platform functionalities, such as tagging or direct messaging, may be leveraged to forward content directly. Since most social media platforms place the primary responsibility for safe online interactions on users [109], and some adults experience their reporting mechanisms as burdensome [113], simplicity should also have significance in this context. While mandatory user input in most of these mechanisms is relatively small when compared to platform-independent offers, our findings still support suggestions from previous work that incident categories should be better explained, delimited, and provided with examples [99, 113].

Accommodate the needs of diverse users by being adaptable to different cognitive abilities, language competences, as well as input and feedback preferences (D2). Our results indicate that some youth struggle with standardized incident categories when reporting hate on platforms, as classifying concrete incidents into abstract categories can be cognitively challenging (B4). Research with disabled content creators further revealed that

these categories often do not account for ableist hate [49]. With regard to platform-independent reporting tools, youth therefore want the option to choose between structured and unstructured input options (R4). For available options in Germany, this often suggests supplementing free text fields with structured alternatives, e.g., drop-down lists with hate types. On some platforms, submitting a qualitative description as part of reports is possible, which can be of value not only for the submitter but also the recipient, e.g., during assessment [21]. As argued in previous work, this allows for better contextualization [99, 113]. However, since many platforms' moderation systems are tailored towards the, often (semi-)automated, processing of large volumes of pre-structured user reports [21, 89, 95], an entirely optional categorization would likely require substantial adjustments of current processes. In addition, several youth argued that reporting tools should enable a presentation of information in foreign or simple language (R3), to allow people with different language proficiencies and cognitive abilities to submit reports. As language barriers deter hate crime reporting by minorities [106], this is particularly relevant. For available reporting mechanisms, we thus recommend an expansion of language options. With the exception of one German reporting center [48], we are not aware of any offers that provide information in simple language. Regarding foreign language support, large platforms mostly allow for a selection from numerous options, while many platform-independent options in Germany are only available in German and sometimes English [13, 33, 46, 48]. Particularly for small organizations, it could be difficult to reconcile language inclusivity with limited resources. For them, it may be sensible to focus on the most commonly spoken languages within their country. Finally, some youth suggested that users should have the opportunity to adapt feedback settings (R7), i.e., what type and amount of notifications one would like to receive. We will reflect on this below. Altogether, our findings on the importance of the adaptability of reporting systems are consistent with the broader observation from content moderation research that a one-size-fits-all approach does not accommodate the needs of young and socially marginalized individuals [49, 69, 72, 90, 100].

Be transparent about data processing and transmission, report assessment and its outcomes, initiated measures, and potential ramifications (D3). The two barriers that were most frequently cited by youth with regard to platform-based reporting of hateful content were a perceived deficiency in the enforcement of community guidelines (B1) and a lack of or insufficient feedback on submitted reports (B2). This corresponds to the observation that many content moderation systems exhibit transparency deficits regarding the time-frame of assessing reports and the communication of outcomes [99, 113]. Our results further show that feedback on platform-independent reporting is important to many youth (R6), which complements similar findings for platform-based reporting [4]. Specifically, there is a demand for transparency on successful submission, assessment outcome, decision rationale, and subsequent measures. While some platform-based and platform-independent systems disclose the first two information types, communicating the latter two rarely occurs and could be challenging to organizations with large reporting volumes or limited personnel. A central interface with a package-tracking-like visualization of reports, as suggested by Zhang et al. [113], appears well suited for communicating such information and could also allow the specification of individual feedback settings (R7). However, this is only feasible for systems which assign reports to an individual account. Many platform-independent offers do not feature this and instead utilize e-mail notifications [13, 33, 46, 48]. Here, checkboxes could query preferences before report submission. Our participants additionally demanded transparency on data disclosure to third parties and any ramifications for themselves (R8). Within the EU, regulations stipulate this. However, since privacy policies are often not read or understood [53], such information should also be provided in an easily comprehensible format during the reporting process.

Support users by facilitating contacts to third-party support services and providing information on legal rights (D4). Youth frequently cited unawareness of platform-independent reporting services and their scope, particularly regarding the 'reportability' of less severe incidents, as a barrier (B5). Similar substantial or procedural unawareness has already been discovered with regard to the reporting of hate crimes [106] or gender-based harassment [98]. To counteract this and thus support decision-making about reporting or seeking support, youth would like to be provided information on the applicability of criminal law to hateful content (R11) and contact details of psycho-social counseling or emergency services (R9). Research on platform-based reporting shows that there is likewise room for improvement in this regard [3, 99]. Given that victims of online harms often seek context-sensitive responses beyond sanctioning perpetrators [102, 111], facilitating contact to alternative support offers could be particularly valuable. Individual German reporting centers provide information on the applicability of criminal norms and victims' rights on their websites and refer to LEAs as emergency contacts in their reporting forms [46, 48]. One even refers to specialized counseling services [46]. Interactive solutions may constitute an alternative. Tan et al. [99] proposed leveraging chatbots to provide feedback and support while reporting sexual harassment. While some youth also saw potential, others were critical whether this could accommodate their needs. Instead, contact with peer support networks like HeartMob [10] could be facilitated. However, youth may not have the necessary competences to provide mental health support to peers [50]. The suitability of such solutions should thus be investigated before adoption. Since victims' needs vary significantly, providing tailored and easily accessible support must generally be reconciled with ensuring the simplicity and usability of the reporting process.

Provide an option to anonymously submit reports and protect the identity and personal data of the reporting individual towards third parties (D5). Our FGIs revealed that some youth are critical of LEAs (B8) or platform operators (B3), not least because they feel that they are not taken seriously. In light of this and fears of repercussions, almost half of the respondents advocated for optionally anonymous platform-independent reporting (R2). Evidence from other contexts supports the importance of anonymity. There are various privacy and security concerns in context of platform-based reporting [113], and it was found that retaliation fears and distrust towards LEAs dissuades members of marginalized communities from reporting hate crimes [106]. With regard to sexual abuse, anonymity in reporting allows vulnerable populations to express themselves more openly, reducing fears of

embarrassment, criticism, or retaliation [2]. However, the significance of anonymity varies among individuals. While many youth value the protection of their identity from third parties, this is less common in relation to reports' recipients. Also, providing an e-mail address is sometimes regarded less critical than real names and address details. There can be a trade-off with transparency considerations. Complete anonymity renders the active provision of feedback impossible. Systems should thus permit a spectrum of anonymity instead of a binary choice. Currently, this is only partially possible. While within the EU reporting on some platforms requires an account (e.g., Bluesky), others alternatively allow it under the provision of a name and address/e-mail contact (e.g., Instagram) or only an e-mail contact (e.g., Reddit). On TikTok, it is instead possible without providing any personal data. Meanwhile, several German reporting centers offer an anonymous option and communicate the absence of feedback in this case [13, 33, 48].

5.4 Limitations and Future Work

Our study is subject to several limitations. *First*, we could only ensure external study quality [65] through a justified participant selection. While we were striving to achieve a diverse sample, our participants nonetheless all lived in a metropolitan area in western Germany. Even though we discussed our results in comparison to studies with other demographics, we cannot assess the validity of our findings for youth from more rural settings, other parts of Germany, or other countries. For instance, there is considerable political attention on hate speech in Germany [6], and support for right-wing parties in urban and western German regions is lower than in others [24]. Both might influence perspectives on countering hateful content. In addition, we could not systematically involve certain societal groups, such as people with disabilities and neurodiverse individuals. Future research could evaluate the generalizability of our findings by examining perspectives on reporting of youth from other countries and communities. *Second*, though we adhered to best practices for FGIs with minors [1, 20, 22, 27, 58], we had to develop our procedure in an ad hoc manner due to a lack of guidance for design-oriented FGI research with this demographic. While some works applied them in such contexts [4, 19, 42], and Schafer et al. [87] explore how participatory design (PD) methods could be harnessed to study misinformation, disinformation, or online hate, we see potential in methodological contributions that synthesize best practices for ethically sound FGI-research with minors in PD. *Third*, our study should only be seen as a first step towards the design of youth-sensitive and platform-independent reporting tools for hateful content. Our heuristics partially suggest contradictory design choices, and reporting tools additionally need to satisfy expectations of reports' recipients, on which there is no research yet. The Value Sensitive Design (VSD) framework might be well suited to guide future design efforts that accommodate these circumstances. It posits that design choices shape possibilities for different stakeholders and therefore promote or subvert human values [32]. By adopting such a perspective, values that should inform reporting tool design can be elicited. As design trade-offs are likely, VSD's proactive engagement with value conflicts seems particularly promising. Value conflicts arise when equally important values suggest incompatible design choices [84, 105]. For instance, the tension between receiving feedback on the processing of reports and ensuring anonymity may be explored as a conflict between the values of transparency and privacy. As constraints on the design space, such conflicts could be negotiated among stakeholders to identify a range of viable solutions [31].

6 Conclusion

In this work, we employed a qualitative research design to investigate how reporting systems for hateful content can be designed in a youth-sensitive manner. In light of a research gap regarding youth perspectives on reporting and platform-independent reporting tools, we conducted eight FGIs with a diverse group of German youth (N=47), followed by a qualitative content analysis. We found that nine barriers complicate or inhibit youth's reporting of hateful content. On platforms, they feel particularly discouraged by deficient rule enforcement and feedback, while platform-independent alternatives are rather unknown and perceived as time-consuming and disruptive. Moreover, we elucidated eleven requirements on the design of platform-independent reporting tools. While some of them relate to the reporting process, others are focused on ensuring feedback and transparency or providing additional support. Finally, based on our findings and previous work, we derived five design heuristics for youth-sensitive and inclusive reporting systems that are centered around simplicity, adaptability, transparency, support, and anonymity.

Acknowledgments

This research has been co-funded by the German Federal Ministry of Education and Research (BMBF) in the project CYLENCE (13N16636) [55], as well as by the BMBF and the Hessian Ministry of Higher Education, Research, Science and the Arts (HMWK) within their joint support of the National Research Center for Applied Cybersecurity ATHENE. We would like to thank all interviewees and partner organizations for their participation.

References

[1] Kristin Adler, Sanna Salanterä, and Maya Zumstein-Shaha. 2019. Focus Group Interviews in Child, Youth, and Parent Research: An Integrative Literature Review. *International Journal of Qualitative Methods* 18 (Jan. 2019), 160940691988727. doi:10.1177/1609406919887274

[2] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3906–3918. doi:10.1145/2858036.2858096

[3] Atieh Armin, Joseph J Trybala, Jordyn Young, and Afsaneh Razi. 2024. Support in Short Form: Investigating TikTok Comments on Videos with #Harassment. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 305, 8 pages. doi:10.1145/3613905.3650849

[4] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3895–3905. doi:10.1145/2858036.2858548

[5] Anna Bagnoli and Andrew Clark. 2010. Focus groups with young people: a participatory approach to research planning. *Journal of Youth Studies* 13, 1 (Feb. 2010), 101–119. doi:10.1080/13676260903173504

[6] Rafael Bauschke and Sebastian Jäckle. 2023. Hate speech on social media against German mayors: Extent of the phenomenon, reactions, and implications. *Policy & Internet* 15, 2 (June 2023), 223–242. doi:10.1002/poi3.335

[7] Natalie Beisch and Wolfgang Koch. 2023. ARD/ZDF-Onlinestudie: Weitergehende Normalisierung der Internetnutzung nach Wegfall

aller Corona-Schutzmaßnahmen. *Media Perspektiven* 23 (2023), 1–9. https://www.ard-zdf-onlinestudie.de/files/2023/MP_23_2023_Onlinestudie_2023_Fortschreibung.pdf

[8] Franz Beitzinger and Uwe Leest. 2024. *Cyberlife V: Spannungsfeld zwischen Faszination und Gefahr – Cybermobbing bei Schülerinnen und Schülern.* Research Report. Bündnis gegen Cybermobbing e.V., Karlsruhe. https://buendnis-gegen-cybermobbing.de/wp-content/uploads/2024/10/Cyberlife_Studie_2024_Endversion.pdf

[9] Lukas Bernhard and Lutz Ickstadt. 2024. *Lauter Hass - leiser Rückzug: Wie Hass im Netz den demokratischen Diskurs bedroht: Ergebnisse einer repräsentativen Befragung.* Research Report. Kompetenznetzwerk Hass im Netz, Berlin. https://kompetenznetzwerk-hass-im-netz.de/wp-content/uploads/2024/02/Studie_Lauter-Hass-leiser-Rueckzug.pdf

[10] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 24 (Dec. 2017), 19 pages. doi:10.1145/3134659

[11] Jack Bowker and Jacques Ophoff. 2022. Reducing Exposure to Hateful Speech Online. In *Intelligent Computing. SAI 2022. Lecture Notes in Networks and Systems*, Kohei Arai (Ed.). Vol. 508. Springer International Publishing, Cham, 630–645. doi:10.1007/978-3-031-10467-1_38

[12] Robert L. Brennan and Dale J. Prediger. 1981. Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement* 41, 3 (Oct. 1981), 687–699. doi:10.1177/001316448104100307

[13] Jugendstiftung BW. 2024. Report Hate Speech! https://meldestelle-respect.de/en/

[14] Julian Bäumler, Marc-André Kaufhold, Georg Voronin, and Christian Reuter. 2024. Towards an Online Hate Speech Classification Scheme for German Law Enforcement and Reporting Centers: Insights from Research and Practice. In *Mensch und Computer 2024 – Workshopband*. Gesellschaft für Informatik, Karlsruhe, Germany, 11 pages. doi:10.18420/muc2024-mci-ws13-124

[15] Julian Bäumler, Thea Riebe, Marc-André Kaufhold, and Christian Reuter. 2025. Harnessing Inter-Organizational Collaboration and Automation to Combat Online Hate Speech: A Qualitative Study with German Reporting Centers. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW093 (2025), 31 pages. doi:10.1145/3710991

[16] Jie Cai, Aashka Patel, Azadeh Naderi, and Donghee Yvette Wohn. 2024. Content Moderation Justice and Fairness on Social Media: Comparisons Across Different Contexts and Platforms. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 84, 9 pages. doi:10.1145/3613905.3650882

[17] Chen-Jung Chan. 2024. Normative Regulierung für algorithmische Inhaltsmoderation auf Internet-Plattformen. In *Künstliche Intelligenz, Ethik und Recht*, Matthias Knauff, Chien-Liang Lee, Yuh-May Lin, and Meinhard Schröder (Eds.). Nomos Verlagsgesellschaft, Baden-Baden, 31–44. doi:10.5771/9783748916499-31

[18] Naganna Chetty and Sreejith Alathur. 2018. Hate speech review in the context of online social networks. *Aggression and Violent Behavior* 40 (May 2018), 108–118. doi:10.1016/j.avb.2018.05.003

[19] Ananta Chowdhury and Andrea Bunt. 2023. Co-Designing with Early Adolescents: Understanding Perceptions of and Design Considerations for Tech-Based Mediation Strategies that Promote Technology Disengagement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 198, 16 pages. doi:10.1145/3544548.3581134

[20] Lauren Clark. 2009. Focus Group Research With Children and Youth. *Journal for Specialists in Pediatric Nursing* 14, 2 (April 2009), 152–154. doi:10.1111/j.1744-6155.2009.00187.x

[21] Kate Crawford and Tarleton Gillespie. 2016. What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18, 3 (March 2016), 410–428. doi:10.1177/1461444814543163

[22] Alison Moriarty Daley. 2013. Adolescent-Friendly Remedies for the Challenges of Focus Group Research. *Western Journal of Nursing Research* 35, 8 (Sept. 2013), 1043–1059. doi:10.1177/0193945913483881

[23] Christoph Demus, Jonas Pitz, Mina Schütz, Nadine Probol, Melanie Siegel, and Dirk Labudde. 2022. A Comprehensive Dataset for German Offensive Language and Conversation Analysis. In *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*. Association for Computational Linguistics, Seattle, Washington (Hybrid), 143–153. doi:10.18653/v1/2022.woah-1.14

[24] Larissa Deppisch, Torsten Osigus, and Andreas Klärner. 2022. How Rural is Rural Populism? On the Spatial Understanding of Rurality for Analyses of Right-wing Populist Election Success in Germany*. *Rural Sociology* 87, S1 (July 2022), 692–714. doi:10.1111/ruso.12397

[25] Ann DeSmet, Charlene Veldeman, Karolien Poels, Sara Bastiaensens, Katrien Van Cleemput, Heidi Vandebosch, and Ilse De Bourdeaudhuij. 2014. Determinants of Self-Reported Bystander Behavior in Cyberbullying Incidents Amongst Adolescents. *Cyberpsychology, Behavior, and Social Networking* 17, 4 (April 2014), 207–215. doi:10.1089/cyber.2013.0027

[26] Alan Dix, Janet Finlay, Gregory D. Abowd, and Russel Beale. 2004. *Human-Computer Interaction* (3rd edition ed.). Pearson Prentice Hall, Harlow.

[27] Christine Efken. 2002. Keeping the focus in teen focus groups. *Young Consumers* 3, 4 (Sept. 2002), 21–28. doi:10.1108/17473610210813583

[28] Michelle Ferrier and Nisha Garud-Patkar. 2018. TrollBusters: Fighting Online Harassment of Women Journalists. In *Mediating Misogyny*, Jacqueline Ryan Vickery and Tracy Everbach (Eds.). Springer International Publishing, Cham, 311–332. doi:10.1007/978-3-319-72917-6_16

[29] Organisation for Economic Co-operation and Development. 2024. Youth. https://www.oecd.org/en/topics/policy-issues/youth.html

[30] Paula Fortuna and Sérgio Nunes. 2018. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.* 51, 4, Article 85 (July 2018), 30 pages. doi:10.1145/3232676

[31] Batya Friedman, David G. Hendry, and Alan Borning. 2017. A Survey of Value Sensitive Design Methods. *Foundations and Trends® in Human–Computer Interaction* 11, 2 (2017), 63–125. doi:10.1561/1100000015

[32] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. In *Early engagement and new technologies: Opening up the laboratory. Philosophy of Engineering and Technology*, Neelke Doorn, Daan Schuurbiers, Ibo Van De Poel, and Michael E. Gorman (Eds.). Vol. 16. Springer Netherlands, Dordrecht, 55–95. doi:10.1007/978-94-007-7844-3_4

[33] Landesanstalt für Medien NRW. 2024. Beschwerde einreichen. https://www.medienanstalt-nrw.de/zum-nachlesen/recht-und-aufsicht/beschwerde.html

[34] Cally Gatehouse, Matthew Wood, Jo Briggs, James Pickles, and Shaun Lawson. 2018. Troubling Vulnerability: Designing with LGBT Young People's Ambivalence Towards Hate Crime Reporting. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173683

[35] Daniel Geschke, Anja Klaßen, Matthias Quent, and Christoph Richter. 2019. *#Hass im Netz: Der schleichende Angriff auf unsere Demokratie. Eine bundesweite repräsentative Untersuchung.* Research Report. Institut für Demokratie und Zivilgesellschaft (IDZ), Jena. 1–158 pages. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf

[36] Tarleton Gillespie. 2019. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale University Press, New Haven. doi:10.12987/9780300235029

[37] Brian Goredema-Braid. 2010. Ethical Research with Young People. *Research Ethics* 6, 2 (June 2010), 48–52. doi:10.1177/174701611000600204

[38] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (Jan. 2020), 205395171989794. doi:10.1177/2053951719897945

[39] Nitesh Goyal, Leslie Park, and Lucy Vasserman. 2022. "You have to prove the threat is real": Understanding the needs of Female Journalists and Activists to Document and Report Online Harassment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 242, 17 pages. doi:10.1145/3491102.3517517

[40] Rachel Griffin. 2022. New school speech regulation as a regulatory strategy against hate speech on social media: The case of Germany's NetzDG. *Telecommunications Policy* 46, 9 (Oct. 2022), 102411. doi:10.1016/j.telpol.2022.102411

[41] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 466 (Oct. 2021), 35 pages. doi:10.1145/3479610

[42] Katrin Hartwig, Tom Biselli, Franziska Schneider, and Christian Reuter. 2024. From Adolescents' Eyes: Assessing an Indicator-Based Intervention to Combat Misinformation on TikTok. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 905, 20 pages. doi:10.1145/3613904.3642264

[43] Amy A. Hasinoff and Nathan Schneider. 2022. From Scalability to Subsidiarity in Addressing Online Harm. *Social Media + Society* 8, 3 (July 2022), 205630512211260. doi:10.1177/20563051221126041

[44] International Network Against Cyber Hate. 2024. Member details. https://www.inach.net/member-details/

[45] HateAid. 2020. App gegen Hass – Mach mit und werde MeldeHeld*in. https://hateaid.org/meldehelden-app/

[46] HateAid. 2024. The HateAid reporting form. The best way to contact us. https://hateaid.org/en/reporting-form/

[47] Michael A. Hedderich, Natalie N. Bazarova, Wenting Zou, Ryun Shim, Xinda Ma, and Qian Yang. 2024. A Piece of Theatre: Investigating How Teachers Design LLM Chatbots to Assist Adolescent Cyberbullying Education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA,

Article 668, 17 pages. doi:10.1145/3613904.3642379

[48] Hessisches Ministerium des Innern, für Sicherheit und Heimatschutz. 2024. Meldeformular. Hate Speech & Extremismus melden. https://hessengegenhetze.de/hate-speech-und-extremismus-melden

[49] Sharon Heung, Lucy Jiang, Shiri Azenkot, and Aditya Vashistha. 2024. "Vulnerable, Victimized, and Objectified": Understanding Ableist Hate and Harassment Experienced by Disabled Content Creators on Social Media. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 744, 19 pages. doi:10.1145/3613904.3641949

[50] Jina Huh-Yoo, Afsaneh Razi, Diep N. Nguyen, Sampada Regmi, and Pamela J. Wisniewski. 2023. "Help Me:" Examining Youth's Private Pleas for Support and the Responses Received from Peers via Instagram Direct Messages. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 336, 14 pages. doi:10.1145/3544548.3581233

[51] Netta Iivari, Leena Ventä-Olkkonen, Sumita Sharma, Tonja Molin-Juustila, and Essi Kinnunen. 2021. CHI Against Bullying: Taking Stock of the Past and Envisioning the Future. 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 357, 17 pages. doi:10.1145/3411764.3445282

[52] Sylvia Jaki and Stefan Steiger (Eds.). 2023. *Digitale Hate Speech: Interdisziplinäre Perspektiven auf Erkennung, Beschreibung und Regulation*. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-662-65964-9

[53] Yousra Javed and Ayesha Sajid. 2024. A Systematic Review of Privacy Policy Literature. *ACM Comput. Surv.* 57, 2, Article 45 (Nov. 2024), 43 pages. doi:10.1145/3698393

[54] Patrick Jost and Monica Divitini. 2020. Game elicitation: exploring assistance in delayed-effect supply chain decision making. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 40, 10 pages. doi:10.1145/3419249.3420154

[55] Marc-André Kaufhold, Markus Bayer, Julian Bäumler, Christian Reuter, Stefan Stieglitz, Ali Sercan Basyurt, Milad Mirabaie, Christoph Fuchß, and Kaan Eyilmez. 2023. CYLENCE: Strategies and Tools for Cross-Media Reporting, Detection, and Treatment of Cyberbullying and Hatespeech in Law Enforcement Agencies. In *Mensch und Computer 2023 - Workshopband*. Gesellschaft für Informatik e.V., Rapperswil, Switzerland, 8 pages. doi:10.18420/muc2023-mci-ws01-211

[56] Marc-André Kaufhold, Thea Riebe, Markus Bayer, and Christian Reuter. 2024. 'We Do Not Have the Capacity to Monitor All Media': A Design Case Study on Cyber Situational Awareness in Computer Emergency Response Teams. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 580, 16 pages. doi:10.1145/3613904.3642368

[57] Teo Keipi, Matti Näsi, Atte Oksanen, and Pekka Räsänen. 2016. *Online Hate and Harmful Content: Cross-national perspectives* (1 ed.). Routledge, London. doi:10.4324/9781315628370

[58] Christine Kennedy, Susan Kools, and Richard A. Krueger. 2001. Methodological Considerations in Children's Focus Groups. *Nursing Research* 50, 3 (2001), 184–187.

[59] Jae Yeon Kim, Jaeung Sim, and Daegon Cho. 2023. Identity and Status: When Counterspeech Increases Hate Speech Reporting and Why. *Information Systems Frontiers* 25, 5 (Oct. 2023), 1683–1694. doi:10.1007/s10796-021-10229-2

[60] Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 429, 16 pages. doi:10.1145/3544548.3580707

[61] Torben Klausa. 2023. Graduating from 'new-school' – Germany's procedural approach to regulating online discourse. *Information, Communication & Society* 26, 1 (Jan. 2023), 54–69. doi:10.1080/1369118X.2021.2020321

[62] Yubo Kou and Xinning Gui. 2021. Flag and Flaggability in Automated Moderation: The Case of Reporting Toxic Behavior in an Online Game Community. 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 437, 12 pages. doi:10.1145/3411764.3445279

[63] Theodora Koulouri, Robert D. Macredie, and David Olakitan. 2022. Chatbots to Support Young Adults' Mental Health: An Exploratory Study of Acceptability. *ACM Trans. Interact. Intell. Syst.* 12, 2, Article 11 (July 2022), 39 pages. doi:10.1145/3485874

[64] Richard A. Krueger and Mary Anne Casey. 2014. *Focus Groups: A Practical Guide for Applied Research* (5 ed.). SAGE Publications, Thousand Oaks, CA, USA.

[65] Udo Kuckartz. 2014. *Qualitative text analysis: A guide to methods, practice and using software*. SAGE Publications Ltd, London.

[66] Udo Kuckartz and Stefan Rädiker. 2019. *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Springer International Publishing, Cham. doi:10.1007/978-3-030-15671-8

[67] Landesanstalt für Medien NRW. 2023. *Hate Speech forsa-Studie 2023. Zentrale Untersuchungsergebnisse*. Technical Report. Landesanstalt für Medien NRW. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2023_Praesentation.pdf

[68] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. Interviews and focus groups. In *Research Methods in Human Computer Interaction*, Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser (Eds.). Morgan Kaufmann, Cambridge, 187–228. doi:10.1016/B978-0-12-805390-4.00008-X

[69] Yao Lyu, Jie Cai, Anisa Callis, Kelley Cotter, and John M. Carroll. 2024. "I Got Flagged for Supposed Bullying, Even Though It Was in Response to Someone Harassing Me About My Disability.": A Study of Blind TikTokers' Content Moderation Experiences. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 741, 15 pages. doi:10.1145/3613904.3642148

[70] Kaitlin Mahar, Amy X. Zhang, and David Karger. 2018. Squadbox: A Tool to Combat Email Harassment Using Friendsourced Moderation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174160

[71] Sandip Modha, Prasenjit Majumder, Thomas Mandl, and Chintak Mandalia. 2020. Detecting and visualizing hate speech in social media: A cyber Watchdog for surveillance. *Expert Systems with Applications* 161 (Dec. 2020), 113725. doi:10.1016/j.eswa.2020.113725

[72] Tyler Musgrave, Alia Cummings, and Sarita Schoenebeck. 2022. Experiences of Harm, Healing, and Joy among Black Women and Femmes on Social Media. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 240, 17 pages. doi:10.1145/3491102.3517608

[73] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. doi:10.1177/1461444818773059

[74] Thorsten Müller. 2024. Ergebnisse der ARD/ZDF-Medienstudie. Zahl der Social-Media-Nutzenden steigt auf 60 Prozent. *Media Perspektiven* 2024, 28 (2024), 1–8. https://www.ard-media.de/fileadmin/user_upload/media-perspektiven/pdf/2024/MP_28_2024_ARD_ZDF-Medienstudie_2024._Zahl_der_Social-Media-Nutzenden_steigt_auf_60_Prozent.pdf

[75] Brigitte Naderer, Ruth Wendt, Marko Bachl, and Diana Rieger. 2023. Understanding the role of participatory-moral abilities, motivation, and behavior in European adolescents' responses to online hate. *New Media & Society* (Oct. 2023), 14614448231203617. doi:10.1177/14614448231203617

[76] Federal Ministry of Education and Research. 2024. ISCED 2011 – International Standard Classification of Education. https://www.datenportal.bmbf.de/portal/en/G294.html

[77] Council of Europe. 2024. Reporting to National Bodies. https://www.coe.int/en/web/no-hate-campaign/reporting-to-national-bodies

[78] Federal Statistical Office of Germany. 2024. Personen mit Migrationshintergrund. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/_inhalt.html

[79] Charis Papaevangelou. 2023. The role of citizens in platform governance: A case study on public consultations regarding online content regulation in the European Union. *Global Media and China* 8, 1 (March 2023), 39–56. doi:10.1177/20594364221150142

[80] Janine Patz, Matthias Quent, and Axel Salheiser. 2021. *#Kein Netz für Hass - Staatliche Maßnahmen gegen Hate Speech im Internet. Die Bundesländer im Vergleich*. Research Report. Institut für Demokratie und Zivilgesellschaft (IDZ), Jena, Germany. https://www.amadeu-antonio-stiftung.de/wp-content/uploads/2021/03/Studie_Kein_Netz_fÃijr_Hass_BundeslÃďndervergleich_Hate_Speech_MaÃ§nahmen___Campact-_Institut_fÃijr_Demokratie_und_Zivilgesellschaft.pdf

[81] Tejas Pradhan, Ganesh Bhutkar, and Aditya Pangaonkar. 2022. Prototype Design of a Multi-modal AI-Based Web Application for Hateful Content Detection in Social Media Posts. In *Sense, Feel, Design. INTERACT 2021. Lecture Notes in Computer Science*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Marta Larusdottir, Lucio Davide Spano, José Campos, Morten Hertzum, Tilo Mentler, José Abdelnour Nocera, Lara Piccolo, Stefan Sauer, and Gerrit Van Der Veer (Eds.). Vol. 13198. Springer International Publishing, Cham, 404–411. doi:10.1007/978-3-030-98388-8_36

[82] Samantha Punch. 2002. Interviewing strategies with young people: the 'secret box', stimulus material and task-based activities. *Children & Society* 16, 1 (Jan. 2002), 45–56. doi:10.1002/chi.685

[83] Elizabeth Reid, Regan L. Mandryk, Nicole A. Beres, Madison Klarkowski, and Julian Frommel. 2022. Feeling Good and In Control: In-game Tools to Support Targets of Toxicity. *Proc. ACM Hum.-Comput. Interact.* 6, CHI PLAY, Article 235 (Oct. 2022), 27 pages. doi:10.1145/3549498

[84] Thea Riebe, Julian Bäumler, Marc-André Kaufhold, and Christian Reuter. 2023. Values and Value Conflicts in the Context of OSINT Technologies for Cybersecurity Incident Response: A Value Sensitive Design Perspective. *Computer Supported Cooperative Work (CSCW)* 33 (April 2023), 205–251. doi:10.1007/s10606-022-09453-4

[85] Sarah T. Roberts. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media.* Yale University Press, New Haven.

[86] Corina Sas, Steve Whittaker, Steven Dow, Jodi Forlizzi, and John Zimmerman. 2014. Generating implications for design through design research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1971–1980. doi:10.1145/2556288.2557357

[87] Joseph S. Schafer, Kate Starbird, and Daniela K. Rosner. 2023. Participatory Design and Power in Misinformation, Disinformation, and Online Hate Research. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1724–1739. doi:10.1145/3563657.3596119

[88] Brennan Schaffner, Arjun Nitin Bhagoji, Siyuan Cheng, Jacqueline Mei, Jay L Shen, Grace Wang, Marshini Chetty, Nick Feamster, Genevieve Lakier, and Chenhao Tan. 2024. "Community Guidelines Make this the Best Party on the Internet": An In-Depth Study of Online Platforms' Content Moderation Policies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 486, 16 pages. doi:10.1145/3613904.3642333

[89] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R. Brubaker. 2021. A Framework of Severity for Harmful Content Online. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 368 (Oct. 2021), 33 pages. doi:10.1145/3479512

[90] Sarita Schoenebeck, Carol F. Scott, Emma Grace Hurley, Tammy Chang, and Ellen Selkie. 2021. Youth Trust in Social Media Companies and Expectations of Justice: Accountability and Repair After Online Harassment. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 2 (April 2021), 18 pages. doi:10.1145/3449076

[91] Andrew Sears, Julie A. Jacko, and Julie A. Jacko (Eds.). 2007. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition.* CRC Press, Boca Raton. doi:10.1201/9781410615862

[92] Andrew Sellars. 2016. *Defining Hate Speech.* Research Publication 2016-20. Berkman Klein Center for Internet and Society, Cambridge, MA, USA. https://www.ssrn.com/abstract=2882244

[93] Alexandra A. Siegel. 2020. Online Hate Speech. In *Social Media and Democracy: The State of the Field, Prospects for Reform*, Nathaniel Persily and Joshua A. Tucker (Eds.). Cambridge University Press, Cambridge, MA, USA, 56–88.

[94] Jesper Simonsen and Toni Robertson (Eds.). 2012. *Routledge International Handbook of Participatory Design* (1 ed.). Routledge, London. doi:10.4324/9780203108543

[95] Mohit Singhal, Chen Ling, Pujan Paudel, Poojitha Thota, Nihal Kumarswamy, Gianluca Stringhini, and Shirin Nilizadeh. 2023. SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE, Delft, Netherlands, 868–895. doi:10.1109/EuroSP57164.2023.00056

[96] Emily Skop. 2006. The methodological potential of focus groups in population geography. *Population, Space and Place* 12, 2 (March 2006), 113–124. doi:10.1002/psp.402

[97] Jens Struck, Daniel Wagner, Thomas Görgen, Samuel Tomczyk, Antonia Mischler, Pia Angelika Müller, and Stefan Harrendorf. 2022. Menschenverachtende Online-Kommunikation – Phänomene und Gegenstrategien. In *Radikalisierungsnarrative online*, Sybille Reinke De Buitrago (Ed.). Springer Fachmedien Wiesbaden, Wiesbaden, 171–195. doi:10.1007/978-3-658-37043-5_8

[98] Sharifa Sultana, Mitrasree Deb, Ananya Bhattacharjee, Shaid Hasan, S.M.Raihanul Alam, Trishna Chakraborty, Prianka Roy, Samira Fairuz Ahmed, Aparna Moitra, M Ashraful Amin, A.K.M. Najmul Islam, and Syed Ishtiaque Ahmed. 2021. 'Unmochon': A Tool to Combat Online Sexual Harassment over Facebook Messenger. 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 707, 18 pages. doi:10.1145/3411764.3445154

[99] Yuying Tan, Eva Nave, Heidi Vandebosch, Sara Pabian, and Karolien Poels. 2023. *Methods for reporting online sexual harassment.* Research Report. NETHATE. doi:10.31235/osf.io/p92t44

[100] Tangila Islam Tanni, Mamtaj Akter, Joshua Anderson, Mary Jean Amon, and Pamela J. Wisniewski. 2024. Examining the Unique Online Risk Experiences and Mental Health Outcomes of LGBTQ+ versus Heterosexual Youth. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 867, 21 pages. doi:10.1145/3613904.3642509

[101] Katie Salen Tekinbaş, Krithika Jagannath, Ulrik Lyngs, and Petr Slovák. 2021. Designing for Youth-Centered Moderation and Community Governance in Minecraft. *ACM Trans. Comput.-Hum. Interact.* 28, 4, Article 24 (July 2021), 41 pages. doi:10.1145/3450290

[102] Kurt Thomas, Patrick Gage Kelley, Sunny Consolvo, Patrawat Samermit, and Elie Bursztein. 2022. "It's common and a part of being a content creator": Understanding How Creators Experience and Cope with Hate and Harassment Online. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 121, 15 pages. doi:10.1145/3491102.3501879

[103] Alexandra To, Hillary Carey, Geoff Kaufman, and Jessica Hammer. 2021. Reducing Uncertainty and Offering Comfort: Designing Technology for Coping with Interpersonal Racism. 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 398, 17 pages. doi:10.1145/3411764.3445590

[104] UNESCO. 2021. *Assessing internet development in Germany: using UNESCO's Internet Universality ROAM-X Indicators.* Research Report. United Nations Educational, Scientific and Cultural Organization, Paris, France. https://unesdoc.unesco.org/ark:/48223/pf0000378902

[105] Ibo Van De Poel and Lambèr Royakkers. 2011. *Ethics, Technology, and Engineering: An Introduction.* Wiley-Blackwell, Malden.

[106] Matteo Vergani and Carolina Navarro. 2023. Hate Crime Reporting: The Relationship Between Types of Barriers and Perceived Severity. *European Journal on Criminal Policy and Research* 29, 1 (March 2023), 111–126. doi:10.1007/s10610-021-09488-1

[107] Joyce Vissenberg, Leen d'Haenens, and Sonia Livingstone. 2022. Digital Literacy and Online Resilience as Facilitators of Young People's Well-Being?: A Systematic Review. *European Psychologist* 27, 2 (April 2022), 76–85. doi:10.1027/1016-9040/a000478

[108] Ben Wagner, Krisztina Rozgonyi, Marie-Therese Sekwenz, Jennifer Cobbe, and Jatinder Singh. 2020. Regulating transparency? Facebook, Twitter and the German Network Enforcement Act. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 261–271. doi:10.1145/3351095.3372856

[109] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Franziska Roesner, and Kurt Thomas. 2023. "There's so much responsibility on users right now:" Expert Advice for Staying Safer From Hate and Harassment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 190, 17 pages. doi:10.1145/3544548.3581229

[110] Irmtraud Wolfbauer, Mia Magdalena Bangerl, Katharina Maitz, and Viktoria Pammer-Schindler. 2023. Rebo at Work: Reflecting on Working, Learning, and Learning Goals with the Reflection Guidance Chatbot for Apprentices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 244, 7 pages. doi:10.1145/3544549.3585827

[111] Sijia Xiao, Coye Cheshire, and Niloufar Salehi. 2022. Sensemaking, Support, Safety, Retribution, Transformation: A Restorative Justice Approach to Understanding Adolescents' Needs for Addressing Online Harm. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 146, 15 pages. doi:10.1145/3491102.3517614

[112] Wenjie Yin and Arkaitz Zubiaga. 2021. Towards generalisable hate speech detection: a review on obstacles and solutions. *PeerJ Computer Science* 7 (June 2021), e598. doi:10.7717/peerj-cs.598

[113] Alice Qian Zhang, Kaitlin Montague, and Shagun Jhaver. 2024. Cleaning Up the Streets: Understanding Motivations, Mental Models, and Concerns of Users Flagging Social Media Posts. doi:10.48550/ARXIV.2309.06688 Version Number: 3.

A Appendix

The appendix contains information supplementary to this work. First, an English translation of the German-language FGI guide (Sec. A.1) and preliminary questionnaire (Sec. A.2), as well detailed information on the stimuli (Sec. A.3) are provided. This is followed by the demographic details of the participants (Sec. A.4) and an English translation of the coding scheme (Sec. A.5).

A.1 Interview Guide

Below, an English translation of the German-language guide for the FGIs is provided. Stimuli are referenced with abbreviations (S1-S5). To ensure conciseness, the original guide has been slightly shortened. Some statements and activities of the interviewers are

summarized (signified by the use of *italics*). Substantive questions, explanations, and instructions are not summarized.

Introduction (5 minutes)

- *Introduction of the interviewers; explanation of the purpose and procedure of the FGI; reminder of the sensitivity of the topic and potential triggers as well as the option to leave at any time; repetition of the data protection information; obtaining consent for audio recording.*
- *Collection of demographic participant data via questionnaire.*
- *Clarification of open questions and start of the audio recording.*

Background on hate speech & cyberbullying (10 minutes)

- Introduction of S1 and trigger warning: First of all, we would like to show you a video. In it, a person will describe how they have been personally affected by online hate, so we would like to encourage you to protect yourself if this topic could cause negative feelings or refresh previous, difficult experiences. You are welcome to leave the room now or at any time during the video if you feel uncomfortable.
- *Presentation of S1 on monitor/projector.*
- *Age-appropriate explanation of hate speech and cyberbullying and clarification of any comprehension issues.*
- Q1: Could you please share with us which social media platforms or messenger services you use most often and for which purposes?

Strategies against hateful content (10 minutes)

- Introduction of S2: On this poster you can see different strategies for responding to hateful content. In addition to the strategies shown, there are of course other approaches. Take your time to read the poster.
- Q2: How do you react when you see hateful content directed at other people on social media? Please tell us which three strategies you use most frequently.
- Q3: Which of the strategies do you think are particularly promising and why?

Experiences and barriers in context of reporting (15 minutes)

- Q4: One option is reporting hateful content directly on a social media. Who has already done this?
- Q5: Where there any challenges or difficulties when you reported on social media platforms? If you have never reported to them, would you tell us why?
- Q6: Another option is reporting hateful content to platform-independent reporting centers or the police. Who has already done this?
- Q7: Where there any challenges or difficulties when you reported to platform-independent reporting centers or the police? If you have never reported to them, would you tell us why?

Preferences for technical reporting solutions (10 minutes)

- Introduction of S3 and trigger warning: One example of a platform-independent reporting center that also works together with the police is called Respect. We will show you a video clip that explains how their reporting process works. At one point, a racist hate comment can be seen for a moment. You are welcome to leave the room now or at any time during the video if you feel uncomfortable.
- Introduction of S4: If you want to report hateful content to such centers, different technical solutions are imaginable. We have placed cards on the table showing such solutions. Take your time to look at them and feel free to ask if you need us to explain something.
- *Clarification of any comprehension issues.*
- Q8: Which two technical solutions for reporting hateful content to platform-independent reporting centers or the police would you prefer and why?

Requirements for reporting technologies (15 minutes)

- Q9: Which features should a perfect tool for reporting hateful content to reporting centers or the police have?
- Q10: What would be most important to you before and after submitting a report?
- Q11: In addition to assisting with the reporting itself, what other features would you like to see in such a tool?

Dummy reporting (20 minutes)

- Introduction of S5: Together we will now take a closer look at the reporting tool of a German reporting center. Two of you will use one of the devices we have prepared to report a post to Hessen gegen Hetze. The dummy report is coordinated with the center. We have already opened the post you are supposed to report on the devices. It is no real hateful content. The reporting center's website is also already open in your browser. When filling out the reporting form, please skip the section concerning your personal data. Please pay particular attention to any difficulties or challenges that may arise during the reporting process. You will have about seven minutes to complete the report. You can always ask questions.
- *Clarification of any comprehension issues.*
- Q12: What difficulties did you encounter during dummy reporting?
- Q13: What did you like and dislike about the tool?
- Q14: Imagine you are developers of a new reporting tool. What would you do differently?

Conclusion (5 minutes)

- *Stop of the audio recording.*
- *Provision of information about counseling services.*
- *Expression of gratitude for participation and answering of questions about the study.*

A.2 Questionnaire to Capture Demographic Information

Below, an English translation of the German-language paper-based questionnaire is provided. It was completed by the participants prior to the FGIs. Answering question five was optional.

1. Please specify your assigned participant number (e.g., 1a, 3c).

- _____

2. Please specify your type of educational institution (e.g., gymnasium, university) or employment status if you already finished education (e.g., employed, unemployed).

- _____

3. Please specify your age.

- 14-15 years
- 16-17 years
- 18-19 years
- 20-21 years
- 22-23 years
- 24-25 years
- 26-27 years
- 28-29 years

4. Please specify your gender. You can use the free text field if your gender is not listed.

- Male
- Female
- Diverse
- _____

5. Please indicate whether you consider yourself a member of one or several of these social groups. You do not have to share this information if you do not wish to. You can use the free text field if you want to add groups.

- Black, indigenous, people of colour (BIPoC)
- People with white skin color
- Migration biography
- Lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other (LGBTQIA+)
- People with disabilities
- _____

A.3 Information on FGI Stimuli

Tab. 4 provides additional information on and sources for the stimuli (S1-S5) used during the FGIs. Fig. 3 depicts the poster and the cards used as S2 and S4 as well as the reporting form used as S5.

A.4 Detailed Information on FGI Participants

In Tab. 5, the demographic details of the participants that were collected using the preliminary questionnaire as well our participant identifiers are listed. FGI 1 was conducted in cooperation with a LGBTQIA+ advocacy organization, FGIs 2-7 in cooperation with an integrated comprehensive school, and FGI 8 in cooperation with a youth center.

A.5 Coding Scheme

Coding scheme created and applied during the structuring content analysis with main categories (**bold**) and subcategories (*italics*). The full codebook is available in the supplementary material.

1 - Barriers

1-I Barriers for platform-based reporting

- *1-I-a Deficient enforcement*
- *1-I-b Inadequate feedback*
- *1-I-c Distrust in platform operators*
- *1-I-d Standardized incident categories*

1-II Barriers for platform-independent reporting

- *1-II-a Unawareness of reporting options*
- *1-II-b Disruption of social media use*
- *1-II-c Time-consuming reporting procedure*
- *1-II-d Distrust in law enforcement*
- *1-II-e Complicated generation of URLs*

2 - Design Requirements

- *2-a Minimized effort*
- *2-b Anonymous submission option*
- *2-c Foreign & simple language*
- *2-d Diverse input options*
- *2-e Gamification elements*
- *2-f Feedback on report*
- *2-g Customizable feedback*
- *2-h Transparency about consequences*
- *2-i Facilitating additional support*
- *2-j Evidence documentation instructions*
- *2-k Background information on hateful content*

No.	Type	Description & Additional Information
S1	Video	Excerpt from the German-language video "ganz konkret: Hate-Speech und Cybermobbing | Zeit für Politik" by the Bavarian State Agency for Civic Education (Bayerische Landeszentrale für politische Bildungsarbeit). Created for use in school from year eighth onward. A youth describes how he has been personally affected by cyberbullying and how he coped with it. Then an interview with a public prosecutor is shown, who characterizes and differentiates cyberbullying and hate speech. A trigger warning was given before the video was shown. Timeframe: 0:00 - 3:20 Link: https://www.youtube.com/watch?v=FljdLX8gk6A
S2	Poster	Printed poster in DIN A1 size giving an overview about potential strategies to respond to hateful content, created by the authors specifically for the FGIs. Outlined strategies include ignoring, counterspeech, shaming, correcting, blocking, reporting on platforms, reporting to platform-independent reporting centers, reporting to law enforcement, initiating a private conversation with the perpetrator, approaching specialized counseling services, consulting friends or family members, and other. Fig. 3 depicts an English translation of the German poster.
S3	Video	Excerpt from the German-language video "ganz konkret: Hate-Speech und Cybermobbing | Zeit für Politik" by the Bavarian State Agency for Civic Education (Bayerische Landeszentrale für politische Bildungsarbeit). Created for use in school from year eighth onward. The reporting process of the German platform-independent reporting center REspect! is presented. A trigger warning was given before the video was shown. Timeframe: 4:04 - 4:37 Link: https://www.youtube.com/watch?v=FljdLX8gk6A
S4	Cards	Seven paper cards each showing one potential solution to report hateful content to platform-independent reporting centers. They contain the name of the solution, pictograms as illustrations, and sometimes additional explanations. The solutions comprise an app, a web form, a chatbot, a browser plugin, direct messages, tagging of reporting centers, or usage of other platform features (e.g., groups).
S5	Web-form for reporting	Web-based reporting form of the German reporting center Hessen gegen Hetze in German language. In the form, it can be specified whether an incident was offline or online and what exactly happened. Then a link to the content and screenshots can be provided. After a specification whether oneself is affected and wants to file a criminal complaint, personal data (name, address, e-mail, phone number) can be entered optionally. The form ends with an option to consent to the forwarding of data to LEAs and to the data protection declaration. In groups of two, the participants used prepared devices to file a dummy report on non-hateful content. When filling out the reporting form, participants were requested to skip the section concerning their personal data. Fig. 3 provides screenshots of the interface. Link: https://hessengegenhetze.de/hate-speech-und-extremismus-melden

Table 4: Detailed information on the stimuli used during the FGIs.

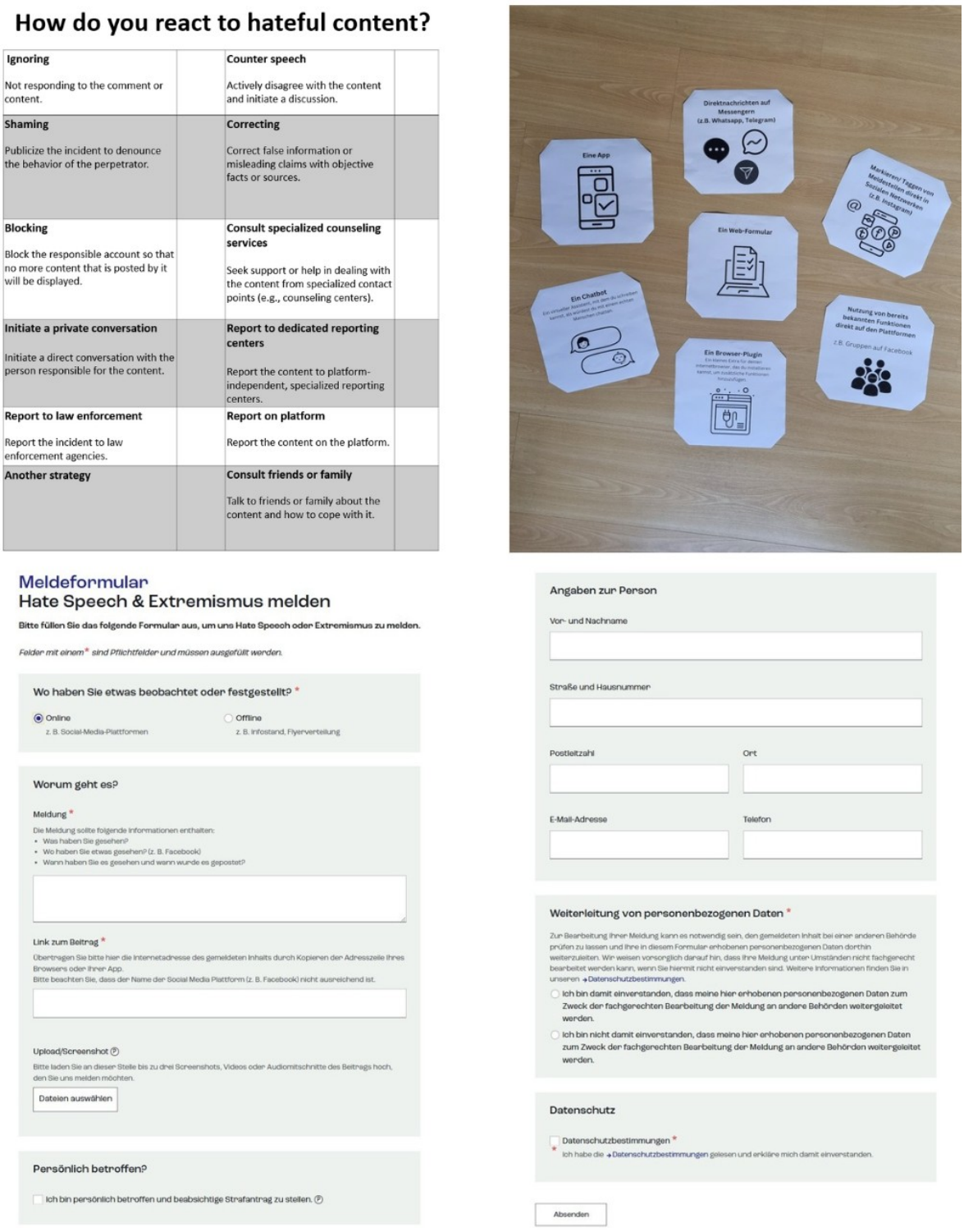

Figure 3: On top are images of an English translation of the German-language poster that was used as S2 (left) and the German-language cards used as S4 (right). At the bottom are screenshots of the web-based reporting form of the German platform-independent reporting center Hessen gegen Hetze that was used for dummy reporting (S5). On the actual form, all details and input boxes are displayed at one page.

No.	Age	Gender	Educational Institution	Group affiliation(s)
1a	20-21	female	University	Migration biography; LGBTQIA+
1b	22-23	non-binary	University	Migration biography; LGBTQIA+
1c	22-23	male	University	White skin color; LGBTQIA+
1d	26-27	female	University	White skin color; LGBTQIA+
1e	18-19	demi-girl	Vocational gymnasium	White skin color; LGBTQIA+
1f	28-29	female	Employed	White skin color; LGBTQIA+; Disabled
2a	14-15	male	Integrated comprehensive school	White skin color; Migration biography
2b	16-17	male	Integrated comprehensive school	Migration biography
2c	14-15	female	Integrated comprehensive school	White skin color
2d	14-15	female	Integrated comprehensive school	White skin color
2e	14-15	female	Integrated comprehensive school	White skin color; Migration biography
2f	14-15	male	Integrated comprehensive school	White skin color; Migration biography
2g	14-15	male	Integrated comprehensive school	Not specified
3a	14-15	male	Integrated comprehensive school	Not specified
3b	14-15	male	Integrated comprehensive school	Not specified
3c	14-15	male	Integrated comprehensive school	White skin color
3d	16-17	female	Integrated comprehensive school	White skin color
3e	14-15	female	Integrated comprehensive school	Not specified
3f	14-15	male	Integrated comprehensive school	Not specified
4a	14-15	female	Integrated comprehensive school	Migration biography
4b	14-15	female	Integrated comprehensive school	Migration biography
4c	16-17	male	Integrated comprehensive school	White skin color
4d	14-15	female	Integrated comprehensive school	White skin color
5a	16-17	female	Integrated comprehensive school	Not specified
5b	16-17	female	Integrated comprehensive school	Not specified
5c	16-17	female	Integrated comprehensive school	Migration biography
5d	16-17	male	Integrated comprehensive school	White skin color
5e	16-17	male	Integrated comprehensive school	White skin color
5f	16-17	female	Integrated comprehensive school	White skin color
6a	16-17	male	Integrated comprehensive school	Not specified
6b	16-17	male	Integrated comprehensive school	Not specified
6c	16-17	male	Integrated comprehensive school	Not specified
6d	16-17	male	Integrated comprehensive school	Not specified
6e	16-17	female	Integrated comprehensive school	Not specified
6f	16-17	female	Integrated comprehensive school	Not specified
7a	14-15	female	Integrated comprehensive school	White skin color; Migration biography
7b	16-17	female	Integrated comprehensive school	White skin color
7c	16-17	female	Integrated comprehensive school	White skin color
7d	16-17	male	Integrated comprehensive school	White skin color
7e	16-17	male	Integrated comprehensive school	Migration biography
7f	16-17	female	Integrated comprehensive school	White skin color; Migration biography
8a	14-15	female	Integrated comprehensive school	Migration biography
8b	16-17	female	Integrated comprehensive school	Migration biography; BIPoC
8c	16-17	female	Integrated comprehensive school	BIPoC
8d	20-21	female	University	White skin color; Migration biography
8e	16-17	female	Integrated comprehensive school	White skin color; Migration biography
8f	14-15	female	Realschule	Migration biography; BIPoC

Table 5: Identifiers and demographic details of the participants. Displayed are age (in cohorts), gender identity, attended educational institution / employment status, and membership of specific social groups. Abbreviations: LGBTQIA+ = lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other; BIPoC = black, indigenous, people of colour.