Research Article

Julian Bäumler*, Georg Voronin and Marc-André Kaufhold

Cyber hate awareness: information types and technologies relevant to the law enforcement and reporting center domain

https://doi.org/10.1515/icom-2024-0062
Received November 22, 2024; accepted February 14, 2025;
published online March 7, 2025

Abstract: In Germany, both law enforcement agencies (LEAs) and dedicated reporting centers (RCs) engage in various activities to counter illegal online hate speech (HS). Due to the high volume of such content and against the background of limited resources, their personnel can be confronted with the issue of information overload. To mitigate this issue, information filtering, classification, prioritization, and visualization technologies offer great potential. However, a nuanced understanding of situational awareness is required to inform the domain-sensitive implementation of supportive technology and adequate decision-making. Although previous research has explored the concept of situational awareness in policing, it has not been studied in relation to online HS. Based on a qualitative research design employing a thematic analysis of qualitative expert interviews with practitioners from German LEAs and RCs ($N = 29$), we will contribute to the state of research in human-computer interaction with a systematization of 23 information types of relevance for situational awareness of online HS in the law enforcement and RC domain. On that basis, we identify victim, perpetrator, context, evidence, legal, and threat awareness as domain-specific situational awareness sub-types and formulate ten implications for designing reporting, open-source intelligence, classification, and visual analytics tools.

***Corresponding author: Julian Bäumler**, Science and Technology for Peace and Security (PEASEC), Technical University of Darmstadt, Darmstadt, Germany, E-mail: baeumler@peasec.tu-darmstadt.de.
https://orcid.org/0000-0002-4535-8036
Georg Voronin, Business Information Systems and Digital Transformation (SAP-Endowed), University of Potsdam, Potsdam, Germany,
E-mail: georg.voronin@uni-potsdam.de.
https://orcid.org/0009-0002-4871-6297
Marc-André Kaufhold, Knowledge Engineering (KE), Technical University of Darmstadt, Darmstadt, Germany,
E-mail: kaufhold@peasec.tu-darmstadt.de.
https://orcid.org/0000-0002-0387-9597

Keywords: hate speech; situational awareness; law enforcement; reporting centers; usable safety and security

1 Introduction

Over three-quarters of German citizens have already been exposed to hate speech (HS) online.[1] Countering its dissemination has come to the attention of German policymakers, not least after the murder of Walther Lübcke, the governmental district president of Kassel in Hesse, who was a target of HS due to his refugee-friendly stance and was shot in 2019 by a right-wing extremist who also published such content online.[2] German law enforcement agencies (LEAs) at both the federal and state levels prosecute illegal HS content. However, as only a few conduct proactive monitoring, their activities partly depend on the volume of reporting by civil society actors, primarily dedicated HS reporting centers (RCs).[3] Due to the high volume of HS content and against the background of limited resources or inadequate technology support, both LEAs' and RCs' personnel may be overwhelmed with managing excessive amounts of potentially relevant information. Such an information overload may either occur during the processing of reported content[4] or during social media monitoring.[5] Not least in light of political intentions to oblige social media operators to report content potentially criminally relevant to LEAs,[6,7] and given the designation of selected RCs as trusted flaggers with privileged reporting channels to platforms,[8] a further increase in information volume is to be expected.

Information overload can significantly impair situational awareness in LEAs and RCs. Situational awareness encompasses the perception of environmental elements within a particular temporal and spatial context, the comprehension of their significance, and the projection of their status into the future.[9] Social media can be leveraged to establish or enhance situational awareness if their content includes real-time descriptions, they have a large and active user base, and their data is easily accessible.[10] Concerning online HS, cross-case situational awareness has significance

for case prioritization, resource allocation, trend recognition, and ultimately counter-strategy adaptation. Since HS dissemination can also be linked to physical hate crimes,[11] it may improve risk assessment in this regard and inform the implementation of preventive measures.[12] Research in human–computer interaction (HCI) demonstrated the potential of user-centered technologies for information filtering, classification, and prioritization in mitigating information overload.[5,13] It also adapted situational awareness to the cybersecurity domain.[14,15] However, as regards technologies for HS response, there has been little linkage to the concept.[10] Moreover, technologies for documenting and reporting hateful content,[16,17] artificial intelligence (AI) models for its detection[18–21] or subtype-classification,[22–24] and dashboards for its visual analysis[12,25] have mainly been researched without any involvement of LEA and RC staff.

Therefore, precursory conceptual and empirical work is required to ensure technologies' applicability.[26–28] Whereas initial user-centered research has involved staff from such organizations in the development of annotated datasets for HS detection,[29] classification schemes,[30] and user interfaces for respective classification systems,[27] there remains a research gap regarding an empirically-grounded systematization of information types that are relevant for situational awareness in this domain. This can serve as a conceptual basis for a corresponding situational awareness model. Morevoer, it can also ensure that future HS response technologies gather, prioritize, analyze, and visualize information types that are actually relevant for the work of LEAs and RCs. With this work, we thus seek to address the following research questions:

- **RQ1:** What information types are relevant for acquiring situational awareness of online HS in the German law enforcement and RC domain?
- **RQ2:** What sub-types of situational awareness can be deduced from these information types?

Based on a qualitative research design employing a thematic analysis[31] of qualitative expert interviews[32,33] with practitioners from German LEAs and RCs ($N = 29$), we provide three contributions (C1–3) to the state of research in HCI. We systematize 23 information types of relevance for situational awareness of online HS in German LEAs and RCs (C1), identify six high-level and domain-specific sub-types of situational awareness (C2), and derive ten implications for designing assistive technologies (C3). After defining online HS, introducing the concept of situational awareness, and reviewing research on technologies for HS response (Section 2), we outline our approach for data collection and analysis (Section 3). Then, we present our findings (Section 4). On this basis, we derive design implications and discuss our limitations (Section 5). We close with a brief conclusion (Section 6).

2 Background and related work

In this section, we elaborate an HS definition suitable for the law enforcement and RC domain (Section 2.1), introduce the fundamentals of situational awareness (Section 2.2), and review research on technologies for HS response (Section 2.3).

2.1 Defining online hate speech

HS is a contested concept,[34] which is rarely systematically conceptualized despite extensive research on its causes and harms as well as adequate responses.[35] In academia, HS is often defined in relation to the respective research motivation.[35] For computer science, this means that research datasets are often underpinned by differing conceptualizations of it.[36] This creates challenges for the evaluation and generalizability of detection models trained on them.[37,38] In law, definitions and thus the sanctionability of different types of expressions as prohibited forms of HS vary by national jurisdiction.[39] Finally, online platforms also follow varying definitions that inform their content moderation practice.[40]

Appropriately defining HS can be challenging. *First*, the term is often used as an umbrella term for numerous negative or aggressive statements, including harsh criticism and simple insults.[41] *Second*, it is often insufficiently distinguished from broader concepts, e.g., abusive or offensive language, as well as specific forms of group-related hostility, e.g., racism.[42] *Third*, even though there is some agreement that it targets groups or individuals based on their group membership or identity,[35,37] there is disagreement about whether HS can only target pre-determined 'protected', i.e., minority, groups.[41] The group-relatedness of HS becomes apparent when it is situated within the taxonomy of harmful online content of Banko et al. (see Figure 1).[36] With the category Hate and Harassment they cover content that torments, humiliates, undermines, or frightens victims. While they recognize that content types are not mutually exclusive, they differentiate Doxing, Insult, Sexual Aggression, Threat of Violence, Identity Attack, and Identity Misrepresentation. Since HS can involve not only overt group-related attacks but also derogatory stereotypes,[20] it covers both of the latter two content types. *Fourth*, HS definitions often focus on one of three aspects: The intention of the person expressing it (1), the content of the speech and its form (2), or the effect on the target (3).[41,42]

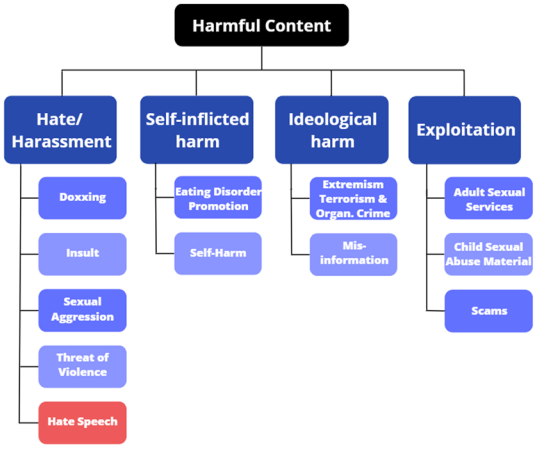

Figure 1: HS within the taxonomy of harmful online content by Banko et al.[36] Adaptation of the taxonomy by the authors.

Developing a novel HS definition goes beyond the scope of our work. However, based on the narrative review of conceptual literature above and given our domain of interest, we regard the following aspects as critical for the selection of an adequate definition:

1. HS should be defined with reference to its content. This seems suitable for LEAs and RCs since they also identify it on this basis. Articulators' intentions and effects on victims may be difficult to ascertain.[42]
2. It should also cover subtle forms of HS that work, e.g., with negative stereotypes, metaphors, irony, or humor,[43,44] as these can also justify or reinforce discrimination against groups.[20]
3. It should not narrow down targeted groups, as new targets can emerge over time.[20] This also seems important regarding the work of LEAs and RCs.
4. Since HS is not defined in German criminal law,[45] LEAs and RCs may also be concerned with cases below the threshold of criminal liability.

The HS definition by Fortuna and Nunes [20, p.5] corresponds to these considerations but needs to be adapted to suit the law enforcement and RC domain. *First*, it should be narrowed down to online content. *Second*, it should be defined irrespective of data modality, as not only textual but also visual, audio, and audio-visual material can constitute HS. *Third*, reference to exemplary targeted group characteristics should be omitted to ensure brevity. Accordingly, we adopt the following definition in this work:

> Online HS is internet content, irrespective of data modality, that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics. It can also be expressed in subtle or humorous ways.

2.2 Situational awareness

In order to prevent and respond to HS effectively, LEAs and RCs need to establish situational awareness and undertake informed decisions. Stanton et al. distinguish three major conceptions of situational awareness in research.[46] These focus on the interaction between individuals and the environment,[47] individuals' cognitive sub-systems,[48] or individuals' perception and understanding of the environment with progressive degrees of insight.[49] As we want to investigate practice-relevant information types within the environment of LEA and RC staff engaging with HS, the latter conception with its three awareness levels is particularly suitable as a conceptual framework. Specifically, Endsley [9, p.36] differentiates "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". In an effort to adopt the concept for cyberspace, the term cyber situational awareness was established. It refers to the state of knowledge of actors that enables them to perceive the relevant elements in the cyber environment (e.g., current situation, impact of the attack, adversarial behavior, as well as quality and trustworthiness of available information) within a particular volume of time and space, to comprehend their meaning, and to project their status in the near future.[14,15,50]

Although it is conceptualized as a subset of situational awareness, it cannot be considered in isolation because events in cyberspace often impact the physical world, e.g., financially, socially, or politically.[51–53] Its growing importance within public administration is reflected in numerous national cybersecurity strategies.[14] Especially professionals in formalized security organizations like computer emergency response teams (CERTs), whose work is particularly challenging due to the necessity of coordination and information exchange, benefit from continuously improving their situational awareness.[54] The term cyber situational awareness is often related to knowledge about occurrences in one's own network, but CERTs have to look way further "to gain a common operational picture of the threat environment in which the constituency is operating" [55, p.17]. Due to the growing number of attacks and security breaches, the volume and variety of potentially relevant data and data sources are increasing, and thus, maintaining adequate cyber situational awareness is increasingly dependent on the properties and capacities of the tools used.[56,57]

However, due to its focus on cybersecurity threats, the notion of cyber situational awareness does not account for the specifics of managing cases of hateful content, such as in the HS response activities of LEAs and RCs. Still, some works

examine situational awareness of LEAs, including a mixed-method study for identifying themes of police-specific situational awareness[58] and an interview study on its role in everyday and high-risk tactical police interventions.[59] Moreover, to enhance situational awareness, researchers propose an ecosystem for the combined use of augmented reality and AI in patrolling and tactical scenarios,[60] an interactive dashboard for fire and police departments,[13] and a

Figure 2: HS classification scheme developed in Bäumler et al.[30] The grey labels characterize the classification tasks. PMK = politically motivated crime.

mobile visual analytics approach for criminal, traffic, and civil incidents.[61] Yet, none of these conceptual, empirical, and technical studies focus on the specifics of situational awareness in HS response. Finally, the HS classification scheme for German LEAs and RCs of Bäumler et al. that differentiates HS by targeted group, the conveyance of an immediate threat, and criminal relevance (cf. Figure 2),[30] is of interest to our work but does not adopt a situational awareness perspective.

2.3 Hate speech response technologies

Computer science researches various technologies that may be leveraged to assist LEAs and RCs in online HS response. Concerning HS gathering, there are three major research directions. First, there is user-centered research in HCI on tools that support victims or witnesses in *documenting and reporting* hateful content. Whereas LEAs and RCs usually offer e-mail contacts and self-administered web portals for reporting,[3,26] one center provides a mobile app whose development, however, was not accompanied by research.[62] Yet, there is user-centered research on tools to support targets of gender-based harassment in evidence documentation and report generation.[16,17]

Second, *detection* algorithms can identify HS content as part of proactive monitoring. There exists extensive research in AI, machine learning (ML), and natural language processing (NLP) on its binary detection.[63–65] While most approaches focus on textual HS,[18,20] there is some research on its detection in visual[21,66] or audiovisual data.[19] While so far most HS detection models are implemented as black boxes, i.e., without disclosing the reasoning behind algorithmic decisions,[67] some works investigate how interface design can improve algorithmic transparency and explainability for content moderators,[67,68] LEA staff,[27] or internet users.[69]

Third, beyond content detection and reporting tools, *open-source intelligence (OSINT)* can be leveraged for gathering information on HS. OSINT refers to "the collection, analysis, and use of data from open sources for intelligence purposes" [70, p.677]. Among other use cases, OSINT methodologies are crucial in monitoring alternative social media and understanding the spread of hateful content across platforms and channels.[71] In context of HS, integrating them with advanced computational techniques, e.g., NLP, ML, or graph neural networks, can enable the recognition of perpetrator, target, and topic relationships,[72] the identification of malicious users,[73] and the tracking of hate-related keywords.[74]

Concerning analysis, there is research on *sub-type classification* and visual analytics. For content moderators, LEAs, and RCs, the type, target, severity, or legality of HS may be relevant. Thus, granular HS classification attracts increasing attention. AI and NLP research often addresses this as a multi-class problem.[65] However, this assumes mutually exclusive labels. Individual content may include varying hostility patterns and target different group affiliations simultaneously.[22,75] Moreover, in some jurisdictions, several criminal norms may apply to HS content simultaneously.[30] Multi-label HS classification addresses this issue but still receives limited attention.[63] Models are available for hate types,[23,24,76] targets,[22,76,77] and hate patterns,[78,79] while for criminal norms, only one training dataset exists.[29]

Beyond that, there is research on the *visual analysis* of online HS data. Some works investigate the visualization of temporal and geographic HS diffusion with maps.[80,81] Others propose dashboards to examine its dissemination with social network analysis, e.g., by visualizing trends, illustrating associations with real-world events, or highlighting perpetrator relationships.[12,25]

Most of these works do not involve users. At the same time, those with an empirical grounding are mostly unrelated to the law enforcement or RC domain, instead targeting internet users or content moderators. Thus, a research gap exists regarding the identification and systematization of information types relevant to this domain that can inform the design of HS response technologies.

3 Methods

To investigate relevant information types for acquiring situational awareness of online HS, we adopted a qualitative research design encompassing 29 expert interviews[32,33] with LEA and RC staff (Section 3.1) and a subsequent thematic analysis of the empirical data (Section 3.2). To create the interview guide, we engaged with the interview design of studies on technology support in other security-critical contexts.[82–84] It included open questions on organizational structures and services, reporting and monitoring HS, analyzing and handling it, collaborative practices, and challenges. As we conducted the interviews as part of a research project to develop technologies and strategies against HS,[28] not all interview content is relevant to this study. The guide varied to some extent between LEAs and RCs to accommodate their different functions.

3.1 Data collection: expert interviews

Following a targeted convenience sampling approach, we contacted German LEAs and RCs accepting reports on online HS from Germany. Twelve individuals from eleven LEAs

and 17 individuals from eleven RCs agreed to participate, of which 14 were male and 15 were female. We use identifiers to refer to the interviews throughout this paper (L1–12; R1–17). From LEAs, we interviewed officials from five state (L1–5) and three federal (L6–8) police authorities, as well as public prosecutors from four attorney general's offices (L9–12). From RCs, we interviewed staff of three centers focusing on online HS (R1–7), three centers accepting reports on different types of illegal content, including criminally relevant HS (R8–11), and centers that handle reports of antisemitic (R12–13), antiziganist (R14), or queer-hostile incidents (R15) and digital violence (R16–17). We involved organizations with varying scope, as they all engage with at least some types of online HS. Table 1 provides detailed information on the interviewees and organizations. After obtaining their informed consent, we held all interviews digitally between March 2023 and August 2024. On average, they lasted 60 min. We audio-recorded and held them, with one English exception (R7), in German. Participation was not compensated.

3.2 Data analysis: thematic analysis

After transcribing and anonymizing the interviews, we performed a thematic analysis following Braun and Clarke[31] with the software MAXQDA 24 to systematize the empirical data. Given the flexibility of the approach and the exploratory character of our research, we considered it well-suited for elaborating information types relevant to acquiring situational awareness of HS. As there is no previous work on situational awareness in this domain, the themes guiding our analysis were identified inductively during the exploration and coding of the empirical material without a pre-defined coding frame. For inductive theme discovery, we followed Mayring's processual approach for category formation.[85] Based on our research question, we defined our selection criterion for relevant content and categories' level of abstraction. Then, one author analyzed the material in an initial iteration, coding all text that matches the selection criterion. He formulated initial thematic categories for concrete relevant information types

Table 1: Organizations' type and geographic area of responsibility, as well as the interviewees' organizational role and gender.

No.	Organization	Area	Interviewee's role	Gender
L1	Criminal police	State	Chief commissioner	Male
L2	Police	State	Police director	Female
L3	Criminal police	State	Chief commissioner	Male
L4	Police	State	Commissioner	Male
L5	Criminal police	State	Chief commissioner	Female
L6	Police	National	Commissioner	Female
L7[a]	Criminal police	National	Chief commissioner	Female
L8[a]	Criminal police	National	Criminal director	Male
L9	Public prosecutor's office	State	Public prosecutor	Female
L10	Public prosecutor's office	State	Senior public prosecutor	Male
L11	Public prosecutor's office	State	Senior public prosecutor	Female
L12	Public prosecutor's office	State	Public prosecutor	Female
R1[a]	HS RC	National	Head of team	Male
R2[a]	HS RC	National	Case manager	Female
R3[a]	HS RC	National	Case manager	Male
R4[a]	HS RC	National	Case manager	Female
R5[a]	HS RC	National	Case manager	Female
R6	HS RC	National	Head of team	Male
R7	HS RC	International	Operational manager	Male
R8	Illegal content RC	National	Desk officer	Male
R9	Illegal content RC	State	Desk officer	Female
R10[a]	Illegal content RC	National	Desk officer	Male
R11[a]	Illegal content RC	National	Legal counsel	Male
R12	Antisemitism RC	National	Scientific officer	Female
R13	Antisemitism RC	State	Case manager	Male
R14	Antiziganism RC	National	Scientific officer	Male
R15	Queer-hate RC	State	Head of team	Female
R16[a]	Digital violence RC	National	Head of counseling	Female
R17[a]	Digital violence RC	National	Legal counsel	Female

Subsequent interviews marked with [a] were conducted with employees of the same organization. RC = reporting center; HS = hate speech.

Table 2: Overview of the information types described by interviewees with frequency of mention by individual LEAs and RCs, differentiated by awareness sub-type.

	Type	LEAs	RCs	Interviews
Victim	Hate type	9	6	L1–6, L8, L10, L12, R1, R3, R4, R7, R12, R14–16
	Targeted group	4	6	L2, L6, L7, L9, R1–5, R6, R9, R12–14
	Victim's identity & contact	9	5	L1–7, L11, L12, R1, R3, R13–17
	Victim's location	5	6	L1, L2, L7, L11, L12, R1, R12–16
Perpetrator	Perpetrator's profiles	10	2	L2–12, R2, R4, R5, R9
	Perpetrator's activities	9	1	L2–4, L6–12, R5
	Perpetrator's IP-address	5	–	L2, L5, L8–10
	Perpetrator's identity	11	6	L1–12, R8–10, R14, R15, R17
	Perpetrator's location	8	3	L1–5, L8, L9, L12, R2, R8, R9
Context	Platform	10	7	L1–6, L8–19, L12, R1, R3, R4, R6, R7, R9–12, R16
	Event relation	6	2	L2–4, L10–12, R1–4, R17
	Discourse relation	9	8	L1, L3–11, R1–6, R8, R11–14, R17
	Extremism relation	5	4	L2, L4, L7, L9, L11, R1–6, R8, R16
Evidence	Direct link	8	7	L3–9, L11, L12, R1–6, R8–11, R15, R16
	(Audio-)visual evidence	10	8	L2–12, R1–6, R8–10, R13, R15–17
	Reporter's identity & contact	7	10	L1, L5–7, L10–12, R1–14, R16
	Incident time	4	4	L6, L9–11, R1–5, R8, R12, R14
	Documentation time	3	2	L7, L9, L10, R1, R5, R8
	Content availability	5	3	L5, L7–10, L12, R1, R2, R9, R16
Legal	Criminal relevance	11	11	L1–12, R1–17
	Criminal norm	11	9	L1–12, R13–6, R8–12, R14–17
	PMK type	5	1	L1–4, L7, L8, R6
Thr.	Physical threat	5	3	L1, L2, L5, L7, L8, L11, R1–5, R8, R16

Thr. = threat; PMK = politically motivated crime.

during this process. If the selected text matched an already established category, it was assigned to that category. Otherwise, a new category was created. After coding a quarter of the interviews, the authors reviewed the categories and decided that the level of abstraction and selection criterion were adequate. Then, the rest of the material was coded. This resulted in a coding scheme that the authors reviewed a second time. During this step, the categories and their definitions were refined. In addition, meta-themes that reflect not concrete information, e.g., perpetrators' identities, but high-level awareness sub-types, e.g., perpetrator awareness, were inductively formed through thematic clustering. Two researchers then used the revised coding scheme to code all material in a second iteration. They discussed difficulties and borderline cases to enhance consistency and coded six interviews parallely to assess coding quality (~20 %; L2, L4, L8, L10–12). MAXQDA was used to check intercoder agreement. The resulting kappa coefficient of 0.78 following Brennan and Prediger[86] indicates substantial agreement.[87,1]

The final coding scheme consisted of six meta-themes and 23 categories and can be found in the Appendix, whereas Table 2 provides an overview of the categories and their occurrence in the individual interviews. In the subsequent section, we structure our results by the identified meta-themes and use core statements for illustration.

4 Results

In the interviews, 23 information types of relevance for acquiring situational awareness of online HS were raised. We organized them into six awareness sub-types, inductively generated by thematic clustering (cf. Table 2). They represent abstract ideal types, each covering one specific dimension of situational awareness of online HS. In the following, we will refer to the overall situational awareness of online HS among LEA and RC staff, which encompasses these sub-types, as cyber hate awareness (CHA). The information types allocated to the sub-types are primarily relevant for establishing situational awareness on the respective dimension. However, we must emphasize that they can secondarily inform other dimensions. Consequently, there

1 We followed Kuckartz and Rädiker[87] to assess intercoder agreement at the segment level and considered codings a match if there was at least 95 % overlap.

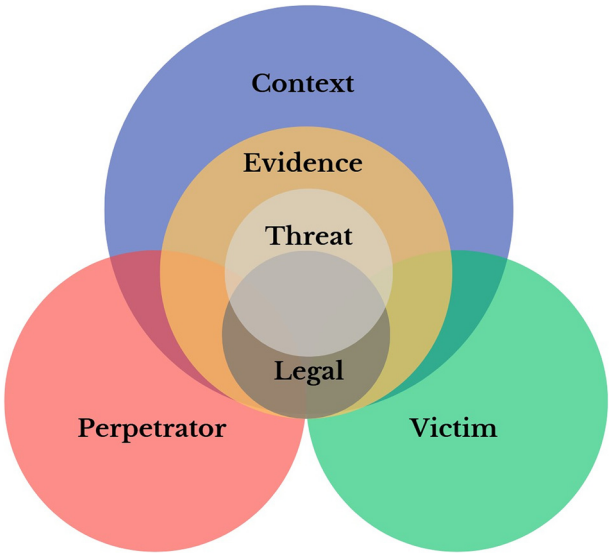

Figure 3: Illustration of the six identified CHA sub-types and their intersections. Intersections indicate that information types primarily assigned to one sub-type are secondarily relevant to others. Lower-level sub-types typically inform higher-level ones.

are intersections (cf. Figure 3). In the following subsections, we first introduce the respective CHA sub-type and outline its relationship to the others. We then describe the information types, with a focus on their significance for LEAs' and RCs' work, relations with information from other sub-types, and relevance for the projection level of CHA. We use direct quotes to illustrate key points.

4.1 Victim awareness

Victim awareness refers to the perception of information relating to those targeted by HS and the comprehension of its significance in the situational context. High awareness may further permit projections of future developments regarding hate targets. Victim-related information can further be significant for establishing context or evidence awareness and may be of value in assessing the legal and threat dimension of hate. Four information types can primarily be assigned to this sub-type.

As HS attacks or denigrates targets based on their group membership or identity, information on **hate types**, i.e., which abstract group affiliations are targeted by it, has significance for several LEAs and RCs. The same applies to the specific **targeted groups**. These differ from hate types, as they do not refer to the kind of group affiliation, e.g., religion, but to the specific group attacked, e.g., Muslims. Individual HS may be attributed to several hate types and target several groups at once, as many victims are attacked in connection to more than one group-related attribute:

> That is often the case, especially when it concerns women, that there is not only one angle. So it is often the case that one says: Based on skin color and from a gender perspective, women are an affected group (R1).

Particularly, when slang or veiled group-related slurs are used, it may be necessary to consider the individual context when establishing which hate types and targeted groups are present. Both information types have minor relevance for prosecution but are an invaluable component of situational pictures on the dissemination of hateful online content within specific jurisdictions or online spaces. Their cross-case analysis allows to comprehend the HS exposure of different social groups over time and an anticipation of upcoming trends. In addition, some RCs leverage the information to recommend tailored, e.g., group or hate type specific, external counseling services to victims.

Information on victims' specific **identity**, **contact** details, and **location** is, by contrast, of greater significance for the day-to-day operation of LEAs and RCs, especially for prosecution. However, this information is only available for content that is targeted at concrete persons. It is particularly important for instances of complaint offenses (Antragsdelikte), such as defamation or insults (cf. Section 4.5). Here, a criminal complaint must be obtained from affected individuals, which is only possible if their identity and contact details are known:

> In these cases, it is simply not possible for the victims to remain anonymous. This means that we actually need the contact details (L12).

In the case of ex-officio offenses (Offizialdelikte), such as incitement, this information may be helpful but is not essential. Accordingly, most RCs allow anonymous reporting. Knowledge of victims' location, e.g., their place of residence, is furthermore relevant in three regards. First, many organizations use it to assign a case to a regionally responsible police station or public prosecutor's office if the perpetrator's place of residence is still unknown (cf. Section 4.2). Second, in some RCs, victims are referred to local counseling services if they know their place of residence. Third, it is relevant for all cross-case situational pictures on the dissemination of online HS that encompass a spatial dimension, e.g., on targeted groups or criminal offenses within a certain jurisdiction. For instance, hate crime hotspots may be identified on that basis:

> Of course, it's nice for us to be able to recognize areas of concentration. Is it a regional hotspot and what is the reason for this? Perhaps it's due to a perpetrator who is a persistent offender (L2).

4.2 Perpetrator awareness

Perpetrator awareness refers to the perception of information related to perpetrators of HS and the comprehension of its significance in the situational context. Information on this can also be useful for understanding the broader context of hateful content and may constitute evidence. Moreover, it often contributes to comprehending the legal and threat dimension of hate. Five information types can be primarily assigned to this sub-type.

Considering a perpetrator's online **activities** can be essential to comprehend the scope and dynamics of HS. Information on a perpetrator's online behavior, such as the frequency and nature of their posts, affiliations to groups, and interactions with others, provides valuable insights for LEAs. Such information helps to identify patterns, assess the threat level, and determine the appropriate legal response. Repeated HS dissemination amplifies severity and may influence the outcome of criminal proceedings:

> Repetition of the actions is an important criterion. Individual incidents can also be serious, but repeated behavior strengthens the case as a hate crime (L6).

Alongside online activities, perpetrator's social media **profile(s)** and **IP-Address** are information types that help LEAs and RCs determine the **identity** of anonymous perpetrators. For LEAs, this information is crucial for initiating criminal proceedings, assessing potential threats, and developing prevention programs. However, uncovering perpetrator identities in the digital realm can be challenging. Perpetrators frequently use pseudonyms, faked e-mail addresses, virtual private networks (VPNs), and disposable accounts, rendering identification complex:

> The challenge lies in determining who is behind the account. That is our everyday work.… Many use VPN tunnels and fake emails that they use for 10 minutes to register an account and then delete (L3).

LEAs employ various methods to overcome identification challenges, including OSINT investigations and filing data requests to platform operators. Some also rely on cross-referencing information from multiple cases to detect patterns and connections. As part of this, they prioritize data that may uniquely identify an account, such as account IDs, account URLs, IP addresses, and timestamps, as those usually remain constant. Especially IP addresses are often a key piece of data for identification, but obtaining and utilizing them can be challenging due to legal and procedural constraints.

Finally, information on the perpetrator's **location**, i.e., their place of residence, is of importance for both criminal proceedings and operational purposes in LEAs. As hate speech often transcends regional boundaries, knowing where a perpetrator resides allows assigning HS cases to LEAs with appropriate regional competencies:

> Of course, if you have a perpetrator, or if it comes through a reporting office and you can identify a perpetrator, then it goes to the station where the perpetrator is based (L1).

When the perpetrator's location is unknown, LEAs may resort on the victim's location to determine the responsible authority. In cases where a physical threat is likely, the location is also relevant for initiating emergency measures, e.g., sending police officers to the perpetrator's address.

4.3 Context awareness

Context awareness refers to the perception of information related to the social contexts in which HS is disseminated, the comprehension of its situational significance, and the anticipation of developments in this regard. It intersects with victim and perpetrator awareness as information associated with it may contribute to those sub-types and vice versa. Contextual information can be evidence and crucial for determining content's legality and threat potential. It encompasses four primary information types.

Information regarding the **platform** or website on which HS content was published is of significance to LEAs and RCs. First, for many RCs and police agencies, knowledge of this is a prerequisite for notifying operators about such content and thus initiating its deletion. For LEAs, this is also important if user data has to be requested from operators to identify anonymous perpetrators. Second, the information is also relevant when generating situational pictures. In light of CHA levels, this relates not only to understanding the current situation by pinpointing temporal trends, but also projecting them into the future. Changes may indicate that hate dissemination is shifting to particular platforms:

> Things are often promoted via well-known platforms, but then the users migrate to others and that's where all the criminal activity actually happens. So it's always important to recognize these trends and developments early on (L8).

Such insights allow LEAs and RCs the adaption of monitoring activities and the demand-oriented design of prevention and awareness-raising measures.

Several LEAs and RCs emphasize that knowledge on the relation of HS content to specific **events**, such as criminal

incidents or political events, is highly relevant. This can not always be objectively determined. Often an interpretation within the conversational context is required if references are only implied or emerge from other content. Some LEAs and RCs conduct HS monitoring in context of specific events to understand the dynamics of HS dissemination in their vicinity:

> We also do some monitoring and check whether you can really pin it down to one trigger, one date, one event and what effect it has. For example, state elections. (L2)

Some also record event-relations of processed cases and present corresponding data to decision-makers. Since knowledge on respective trends may allow a projection of incident numbers, event attribution also informs the projection level of CHA. This can influence resource planning, case prioritization, and the anticipation of potential threats (cf. Section 4.6) necessitating immediate action:

> I think an acute situation like this has the most potential for something to happen. In other words, that some kind of action will follow (R4).

For many LEAs and RCs, the relation of HS to broader social or political **discourses**, e.g., on the war in Ukraine or the Covid-19 pandemic, has comparable relevance. Such links are not always apparent in the content itself, but often stem from its conversational context and may thus require interpretation. Links to discourses are particularly relevant for situational reports on case volumes, as it allows conclusions about the evolution of topics on which HS is disseminated. Such insights are not only valuable for understanding the current situation, but also its projection. In some instances, discourse-related trend data not only serves as a basis for decision-making within the organization itself, but is also provided to other administrative or political decision-makers. It can inform the initiation of measures, such as increased protection of facilities, necessitate upgrading technologies such as AI detection algorithms, affect the allocation of resources, and trigger the initiation of targeted monitoring:

> We always have to be up to date with current events. There's a weekly report, a quick monitoring on current cases. For example, the Ukraine war, the coronavirus pandemic (R2).

Finally, to some LEAs and RCs the extent to which HS content or its creators are connected to anti-constitutional activities, i.e., whether there is an **extremism relation**, is important. Often, this cannot be determined from the content itself and requires interpretation. Respective information may influence case prioritization and the forwarding of content and profiles to domestic intelligence services, which focus on monitoring and countering extremist efforts. In context of CHA, it is also used to anticipate future extremist activities:

> You have to look at this separately from criminal law, because that is often not even about specific offenses, but about activities that can then lead to them. A storming of the parliament or whatever (R1).

4.4 Evidence awareness

Evidence awareness refers to the perception of information that is relevant as evidence during the investigation and prosecution of potentially illegal HS and the comprehension of its significance in the context of the current situation. It intersects with victim, perpetrator, and context awareness, as information pieces from those often constitute evidence, and contributes to legal and threat awareness. Six information types can be primarily assigned to it.

A **direct link** to specific HS content is required by most RCs and LEAs. This type of information primarily refers to the content's URL. For certain content types, e.g., longer texts or video and audio files, additional information may also be valuable:

> If a video has thousands of comments and we get the URL, it takes us a long time to find the comment. If the video is two hours long and there is a potentially relevant minute, then a time code helps us in addition to the URL (R6)

Direct links are particularly important during evidence documentation, for RCs before a case is forwarded to other organizations and for LEAs in the context of investigations. If they are unavailable, a time-consuming search may be necessary, or it may be impossible to proceed. Furthermore, direct links are also relevant for requesting content deletion from platform operators. Therefore, some RCs require the indication of a URL in their reporting portals.

As HS content may be deleted or hidden before or during its processing by LEAs or RCs, **(audio-)visual evidence** is required for a successful assessment and subsequent investigations and criminal proceedings. In case of textual or visual content, this usually entails screenshots. In case of audio-visual content, it can also be recorded and saved as a file. Three different types of (audio-)visual evidence are considered particularly relevant. First, evidence documenting the specific HS content. Second, evidence documenting the immediate conversational context, e.g., an initial post or thread. Third, evidence documenting the perpetrator's profile. As regards screenshot quality, it is emphasized that

they should ideally include the URL of the depicted content or profile and a time stamp:

> We always need the criminally relevant comment or post. Then the initial post, the context in which it was posted, and then a screenshot of the user's profile. The screenshots should all contain the URL and preferably the time (L9).

Visual evidence should be documented as early as possible. RCs often offer uploading options for screenshots on their reporting portals. In addition, RCs and LEAs conduct separate evidence preservation for both reported and self-identified content, if possible.

Information on the **reporter's identity and contact details** is only available if HS content has been externally reported. Since many RCs also allow anonymous reporting, they and subsequently LEAs often do not have this information. The data can be identical to the identity and contact details of the victim (cf. Section 4.1). During case processing it is helpful but not essential to have a contact option, e.g., by e-mail. Especially if reported HS has already been deleted or is not publicly accessible, evidence and contextual information may be requested:

> We have cases were things are reported from closed user groups …and if we say there is not enough for us to evaluate, we ask if there is anything that can be sent afterwards (R1).

If criminal proceedings are initiated, the reporter might also appear as a witness. Furthermore, RCs can only provide feedback if contact information has been provided.

The **incident time**, i.e., the date and time HS content was published, has significance for LEAs and RCs as it represents the time of offense for criminally investigated cases. In some RCs, it further influences the case processing order. Across cases, the information is relevant for all analyzes, reports, and situational pictures on the dissemination of online HS that cover temporal trends or specific time frames.

The **documentation time** of HS content, i.e., the concrete date and time evidence was preserved, is instead only relevant for individual investigations and criminal proceedings. It is important to enhance the admissibility of (audio-)visual evidence in court. Ideally, it is incorporated directly into the evidence:

> So I would want a piece of evidence. With a good time stamp. Because we have people claiming that their account was hacked. That they weren't responsible (L7).

The **availability** of HS content, i.e., whether it was and still is publicly accessible, is relevant in two regards. First, it is important information during evidence documentation. Publicly accessible content can be documented by RCs or LEAs themselves. In the case of unavailable content, it is necessary to contact the reporting party and request evidence. On the other hand, the availability and therefore potential reach of content can be relevant in criminal proceedings, e.g., when determining the sanction.

4.5 Legal awareness

Legal awareness refers to the perception of information concerning the legality of HS, the comprehension of its case-specific and cross-case significance, and the projection of the dissemination of illegal HS into the future. It is obtained by interpreting information from the other CHA sub-types and comprises three main information types.

For almost all RCs and LEAs, the general **criminal relevance** of online HS constitutes important information. An often preliminary assessment of criminal relevance determines how cases are handled. Most RCs forward potentially criminally relevant cases to responsible LEAs or support victims in filing a criminal complaint. In non-criminal cases, victims are nonetheless often supported, either in content deletion, by providing counseling, or by referral to external counseling services. Public prosecutor's offices initiate an investigation if there is an initial suspicion of a criminal offense, in the course of which they and supporting police authorities gather evidence, clarify the perpetrator's identity, and specify criminal norms. If the suspicion is confirmed, they will file charges. Otherwise, the investigation will be discontinued. Determining whether content is criminally relevant requires an interpretation of various information from other CHA sub-types. Accordingly, extensive efforts are usually required to generate this information:

> In the best case, I have all this information.… Otherwise, the identity of the suspect must be established, then the statement and the context must be evaluated (L10).

Whereas RCs' determination of criminal relevance is more preliminary, LEAs undertake a more extensive, rigorous, and evidence-bound assessment, as criminal proceedings may initiated on this basis:

> This can sometimes be very difficult, extensive. A statement that the perpetrator made in one minute may require a lengthy examination by the public prosecutor, who may also have to consult case law (L10).

Beyond that, data on criminally relevant HS can also be used to deduce and project trends. This can, e.g., inform resource planning within the organization.

When examining criminal relevance, most LEAs and RCs check the applicability of individual **criminal norms**.

Several German criminal norms are potentially applicable to online HS.[30] While it may be possible to assess candidates for some offenses solely based on content, e.g., public incitement to commit offenses (§ 111 StGB), it is particularly important to consider contextual factors for others. For instance, when assessing cases of defamation (§ 187 StGB) it is crucial to consider the (un-)truthfulness of person-related statements. Generally, several criminal norms can apply to individual HS simultaneously. Information on potentially applicable criminal norms is particularly important for RCs, as different types of offenses are forwarded to separate authorities. The differentiation is also important for LEAs, since in the case of complaint offenses, the victim of HS must file a criminal complaint for further prosecution, which is not required in the case of ex-officio offenses. In the former case, the authority must have knowledge of the victim's identity and contact details (cf. Section 4.1).

The applicable criminal norms are naturally also relevant for subsequent criminal proceedings. Beyond that, aggregated data can also provide a valuable basis for comprehending and projecting trends:

> We are also heavily indicator-based. To manage our resources and identify deficiencies, but also to provide our partners with key figures regarding offenses, which offenses are in focus, and the success of our measures (L8).

Another information that should be considered separately is the attribution of criminally relevant HS to a **politically motivated crime (PMK) type**. While this information has only relevance for one RC, interviewees from several LEAs stated that they evaluate HS according to discernible political motivations, primarily to provide data for crime statistics. There is a uniform federal framework with the following mutually exclusive PMK-types: PMK – right, PMK – left, PMK – foreign ideology, PMK – religious ideology, PMK – other/not assignable. Within CHA, aggregated data on this is essential for understanding and projecting trends regarding the political motivations of perpetrators. It can inform future crime prevention efforts and internal training:

> If there has been an extreme increase, for example in the area of PMK – right, you can say: Okay, we need to do more in prevention, perhaps make the departments more aware, and also look into why there has been this increase (L3).

4.6 Threat awareness

Threat awareness refers to the perception, comprehension, and projection that an immediate **physical threat** to individuals, groups, or institutions is conveyed by HS content. We consider it separately because, although it involves the interpretation of information from other sub-types, it does not match any of them. In addition, LEAs and RCs often take exceptional measures in case of a corresponding characterization of content. Several of them review reported or identified HS early to determine whether a threat is conveyed:

> The first thing you look at is whether measures need to be taken immediately. Does it involve any threats, any specific dangers for someone? And that is of course always the first thing that is checked (L1).

If a potential threat can be deduced, e.g., in case of direct or veiled calls for violence, these organizations immediately inform the responsible sections within the state criminal police office (LKA) or federal criminal police office (BKA) so that those can initiate security measures. This may involve sending police units to threatened actors or perpetrators:

> With experience, you know what tends to indicate lethal or serious violence. We immediately have in mind that there is a risk aspect.… And we approach the authorities in the respective state quite quickly. If it's acute, they can visit the house. Or talk with the threatening person (L7).

This awareness sub-type is heavily dependent on interpretation. Perpetrators' intentions must be comprehended and available information must be evaluated to project the likelihood of a threat.

5 Discussion and implications

To investigate information types relevant for acquiring situational awareness of online HS in the German law enforcement and RC domain, we employed a qualitative research design encompassing semi-structured expert interviews. By conducting a thematic analysis of the interview data, we identified, as our first contribution, 23 information types of practical relevance for these organizations (C1). They corroborate previous findings on domain-relevant hate speech differentiations with regard to targeted groups, immediate threats, and criminal relevance,[29,30] but also encompass additional aspects. To systematize the types in light of the situational awareness framework, we inductively clustered them by thematic dimensions. The resulting six high-level and domain-specific situational awareness sub-types victim, perpetrator, context, evidence, legal, and threat awareness are our second contribution (C2).

Both findings constitute initial steps towards a domain-sensitive situational awareness framework. Following the situational awareness concept of Endsley[9] and extensions for the cybersecurity domain,[14,15] we view cyber hate

awareness (CHA) as a state of knowledge of actors that enables them to perceive information pieces related to online HS within a specific volume of time and space (1), to comprehend their meaning and significance (2), and to project their status to the near future (3). As online HS can have consequences beyond the internet, e.g., psychological effects on victims[88] or physical hate crimes,[11] and information from non-digital sources can be important for its handling, we do not consider CHA to be strictly limited to cyberspace.

HS response technologies can enhance CHA in LEAs and RCs, especially in the face of information overload due to high case volumes. Based on our findings and in consideration of the state of research, we derive ten design implications (D1–10) for reporting, OSINT, AI-classification, and visual analytics tools as a third contribution (C3). Then, we outline the limitations of this work and suggest directions for future research.

5.1 Implications for reporting tools

The majority of HS cases that LEAs and RCs handle are based on reports. To ensure that transmitted data can contribute meaningfully to CHA, in particular evidence awareness, both the scope and quality of information must satisfy the recipients' requirements. If this is not the case, a time-consuming inquiry for additional information or manual search and documentation effort may be necessary.[26] This has implications for reporting tools.

Reporting tools should account for both reporter and recipient perspectives (D1): Both LEAs and RCs offer public channels for reporting HS, mostly e-mail contacts and web portals.[3,26] However, their appeal and usability remain uncertain as there is a lack of user-centered research on their design. Previous HCI research involved targets of gender-based harassment in developing documentation and report creation tools.[16,17] However, none of these tools allow the submission of reports to LEAs, RCs, or similar organizations, and the requirement elicitation was limited to the reporting side. Researchers designing such tools should thus engage with both reporters and recipients to ensure good usability, user experience, completeness and actionability of transmitted data, and data privacy and security.

Reporting tools should facilitate evidence documentation (D2): The fact that LEAs and RCs frequently receive insufficient URLs and (audio-)visual evidence suggests that securing and submitting such information can be challenging for the reporting party. This could be addressed by providing easy-to-understand instructions within reporting tools that explain step-by-step how to create sufficient screenshots, i.e., with timestamp and URL, and generate URLs pointing directly to HS content. Technical support for evidence documentation would be even more convenient, e.g., by enabling reporting tools to capture sufficient (audio-)visual evidence and automatically retrieve correct URLs. Previous research on online harassment documentation tools offers suggestions to this end.[16,17]

5.2 Implications for OSINT tools

If reported information is not sufficient for handling cases or initiating criminal proceedings, or if LEAs proactively examine cases, OSINT investigations can be utilized to gather additional information alongside traditional investigative approaches and data requests at platforms. We can formulate two implications for OSINT tools that assist with this.

OSINT tools should allow evidence documentation (D3): Some interviewees, especially from police agencies, use OSINT to collect perpetrator-related information. Our findings suggest that OSINT tools for detecting and characterizing hate networks,[72] tracking hate-related keywords,[74] or identifying malicious users[73] presented in recent research may be beneficial for enhancing perpetrator, victim, and context awareness in LEAs and RCs. However, what these tools have in common is that they have no documentation functionality for acquired information. Yet, evidence is essential in our domain, especially in preparation for criminal proceedings. Thus, OSINT tools should provide documentation functionalities, such as generating (audio-)visual evidence or saving URLs and metadata.

OSINT tools should follow privacy by design principles (D4): With OSINT tools, LEAs and RCs can effectively collect relevant information and thus strengthen their CHA. However, authorities are subject to particular restrictions regarding the collection and processing of personal data, especially from uninvolved third parties. State-of-the-art tools for hateful content have no safeguards in this regard.[72–74] Riebe et al. further show that privacy issues are rarely considered for similar tools in cybersecurity contexts.[89] Privacy by design principles should thus inform the development of novel solutions to ensure their applicability and the legal admissibility of gathered evidence. Research on OSINT in other law enforcement contexts can provide guidance.[70,90]

5.3 Implications for AI-classification tools

In LEAs and RCs, the assessment of HS with regard to its relevance under criminal law, present hate types, targeted groups, extremist relevance, and immediate threats is essential for the initiation of responses. Especially if large

data volumes induce information overload,[4,5] there is potential to use AI to evaluate corresponding content according to these aspects. In light of our findings, there are several implications for designing such tools.

Classification tools should be capable of multi-label classification (D5): Some of the most important assessment tasks in the investigated domain can entail the simultaneous assignment of cases to several subcategories, e.g., extremism types or criminal norms. However, multi-label HS classification has generally received little scientific attention.[63] Available models for hate types or targeted groups are limited to a few classes.[22–24,76,77] For classification by extremism type or criminal norm, no models are available, and end user-centered tools are limited to binary classification.[12,67–69,91] Thus, designers of classification tools should put a strong emphasis on implementing models for practice-relevant multi-label tasks.

Classification tools should only perform a pre-assessment (D6): The assessment of HS is often complex,[92] highly context-dependent,[27] and influences far-reaching decisions, e.g., on criminal proceedings or security measures. Erroneous decisions may thus have serious consequences. Therefore, within classification tools, AI should only perform a pre-assessment to support human decision-making, e.g., as part of a preliminary case prioritization or allocation. To avoid algorithmic over-reliance, user interfaces should highlight the preliminary nature of the assessment, communicate limitations and error rates, and encourage users to question and alter the assessment. As part of this, explainable artificial intelligence (XAI) approaches may render decisions more comprehensible. Benefits of XAI in HS classification have been researched,[93] but only for binary tasks.

Classification tools should be adaptable to evolving requirements and contexts (D7): Within the law enforcement and RC domain, assessment criteria for HS can evolve over time. For instance, the assessment of content's criminal relevance may be affected by the advancement of the legislative framework and case law, or novel hate types or targets may emerge due to a shift in discourse. In addition, some LEAs and RCs conduct monitoring in context of specific events or discourses, where generic HS classification tools may only be of limited use. Designers should, therefore, implement steps to ensure the long-term applicability and short-term adaptability of detection tools. One way to increase model flexibility might be to tailor pre-trained large language models (LLMs) to the application domain.[94] These could then be fine-tuned for specific tasks or contexts, i.e., by using few-shot learning approaches that require only a small amount of training data.[95] Data augmentation, i.e., the generation of artificial training data,[96] might be particularly helpful when fine-tuning such models if only limited data for novel hate types, events, or discourses is available.

5.4 Implications for visual analytics tools

The goal of visual analytics research is to turn information overload, which might be induced by the large volume of reported incidents or irrelevant public data considered for OSINT, into an opportunity. Decision-makers, such as in LEAs or RCs, should be enabled to examine large-scale, multi-dimensional, multi-source, and time-varying information to achieve a situational overview and make effective decisions in time-critical situations.[97] From this perspective, our research suggests three design implications concerning data collection, data modeling, and analysis of hate speech incidents.

Visual analytics tools should facilitate the modular integration of reporting and OSINT data sources (D8): While reporting tools are the backbone for data collection in LEAs and RCs, our interviews indicated that OSINT tools help to enrich reported data or allow for the proactive identification of additional cases. Both channels rely on diverse information sources, such as apps, e-mail, social media, and web portals. Combined with the issue of regularly changing application programming interfaces (APIs),[98,99] it seems important to provide opportunities for the modular integration of data sources. While results from reporting and OSINT might be shown in separate feeds with individual cases, the interface should facilitate relationship awareness[100] if, e.g., OSINT data is required for the handling of a reported case or a cross-case analysis is conducted to identify connections and patterns.

Visual analytics tools should enable the analysis of context, perpetrator, and victim information (D9): The execution of visual analytics tasks requires the definition of a proper data model.[97] By conducting our interview study, we established a nuanced understanding of context, perpetrator, and victim information, which is required to develop such a data model in future work (see, e.g., a data model for cyber situational awareness[101]). This is also a prerequisite for the display, aggregation, visualization, and exploration (e.g., zooming, filtering, and details on demand[56]) of CHA data on the interface level. Based on aggregated temporal data on incidents, affected social groups, and used platforms, visualizations should enable the analysis of specific timeframes (see, e.g., [102]), temporal trends, and future projections as a foundation for mitigation and response strategies. Furthermore, a map view might be a promising complement to highlight the locations of victims, perpetrators, responsible authorities, or counseling services (see, e.g., [13,103]).

Visual analytics tools should integrate XAI into their processing pipeline (D10): Our results suggest that AI should be used to enrich factual information with predictive information to allow for a rapid prioritization of incoming reports, especially if an immediate physical threat risk was predicted. While our interviewees acknowledged the importance of human decision-making, the integration of visual analytics and XAI[104] seems promising to facilitate the understanding of algorithmic classifications. This allows for a more effective model reconfiguration, e.g., by leveraging data augmentation[96] and few-shot learning,[95] if the performance is insufficient or legal conditions change. Another potential of visual analytics lies in the visualization of AI-detected discourses, topics,[105] and events.[106] This could help in the detection of temporal trends for projections and reactive measures.

5.5 Limitations and future work

Our research is subject to some limitations. *First*, the findings of qualitative research are influenced by subjectivity. Our interviewees' responses are shaped by their individual perspectives and experiences. Moreover, our interviewing style can influence response behavior, and our data analysis depends on our interpretations.[107] Therefore, a standardized quantitative survey on relevant information types with staff at LEAs and RCs could be implemented in future research. Triangulation with our data could decrease the impact of subjectivity, while in-depth insights into rationales and practices are still feasible.[108] *Second*, our sample exhibits shortcomings. On the part of the LEAs, we only interviewed mid-level or senior staff, often with a coordinating function. This was not intentional, but resulted from contacted organizations' suggestions of interview partners. Given the lack of representation of case managers or staff with other non-coordinating roles, certain information types of relevance to their work may have been missed. Future research could uncover potential biases by obtaining a more heterogeneous sample. *Third*, while we also cannot claim validity for all German LEAs and RCs, the generalizability of our findings outside the German context is particularly difficult to assess. Since we only interviewed one RC with an international scope and no non-German LEAs, some information types and CHA sub-types might be specific to the national context, especially those centered around criminal investigations. Cross-country research could thus examine which information types are of international relevance. *Fourth*, we specifically focused on the law enforcement and RC domain to approach situational awareness of online HS. We did not consider the perspectives of other stakeholders engaging with online hate, such as content moderators, intelligence agencies, and civil society monitoring projects. While there is user-centered work on combating HS with these actors, situational awareness perspectives are missing. Therefore, we see potential in following the example of cyber situational awareness research,[14,15] pursuing a sound conceptualization of CHA in future work.

6 Conclusions

In this work, we employed a qualitative research design encompassing semi-structured expert interviews ($N = 29$) with staff from eleven LEAs and eleven RCs and a thematic analysis to investigate information types relevant for acquiring situational awareness of online HS in the German law enforcement and RC domain. We identified 23 information types of practical relevance and clustered them thematically into six high-level situational awareness sub-types. Victim, perpetrator, context, evidence, legal, and threat awareness constitute key elements of domain-specific cyber hate awareness (CHA) and partly inform each other. Finally, we deduced ten design implications by discussing our findings in light of the state of research. Thus, we contribute to HCI research with an empirical foundation for the user-centered design of HS response technologies and an initial step towards a situational awareness framework for online HS.

Acknowledgments: We thank all interviewees for their participation in our study and Julen Grether for his editorial support.
Research ethics: The ethics committee of the Technical University of Darmstadt considered the study to be exempt from review.
Informed consent: Informed consent was obtained from all individuals included in this study, or their legal guardians or wards.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. Julian Bäumler and Georg Voronin collected and analyzed the empirical data. Julian Bäumler coordinated the publication project, conceptualized the research design, and wrote the manuscript, supported by Georg Voronin in the Related Work and Results Section and Marc-André Kaufhold in the Related Work and Discussion Section. Marc-André Kaufhold led the funding acquisition for the projects leading to this publication.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The authors state no conflict of interest.

Research funding: This work has been co-funded by the German Federal Ministry for Education and Research (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 and by the BMBF in the project CYLENCE (13N16636).
Data availability: Not applicable.

Appendix A: Coding scheme

Coding scheme created and applied during the thematic analysis with meta themes (**bold**) and categories (*italics*).

V – Victim Awareness
- *V-1 Hate Type*
- *V-2 Targeted Group*
- *V-3 Victim's Identity & Contact*
- *V-4 Victim's Location*

P – Perpetrator Awareness
- *P-1 Perpetrator's Profiles*
- *P-2 Perpetrator's Activities*
- *P-3 Perpetrator's IP-Address*
- *P-4 Perpetrator's Identity*
- *P-5 Perpetrator's Location*

C – Context Awareness
- *C-1 Platform*
- *C-2 Event Relation*
- *C-3 Discourse Relation*
- *C-4 Extremism Relation*

E – Evidence Awareness
- *E-1 Direct Link*
- *E-2 (Audio-)visual Evidence*
- *E-3 Reporter's Identity & Contact*
- *E-4 Incident Time*
- *E-5 Documentation Time*
- *E-6 Content Availability*

L – Legal Awareness
- *L-1 Criminal Relevance*
- *L-2 Criminal Norm*
- *L-3 PMK Type*

T – Threat Awareness
- *T-1 Physical Threat*

References

1. Landesanstalt für Medien NRW. *Hate Speech forsa-Studie 2023. Zentrale Untersuchungsergebnisse*; Landesanstalt für Medien NRW: Düsseldorf, 2023. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2023_Praesentation.pdf (accessed 2025-03-03).
2. Bauschke, R.; Jäckle, S. Hate Speech on Social Media Against German Mayors: Extent of the Phenomenon, Reactions, and Implications. *Policy Internet* **2023**, *15* (2), 223—242.
3. Patz, J.; Quent, M.; Salheiser, A. *#Kein Netz für Hass - Staatliche Maßnahmen gegen Hate Speech im Internet. Die Bundesländer im Vergleich*; Institut für Demokratie und Zivilgesellschaft (IDZ): Jena, Germany, 2021. https://www.amadeu-antonio-stiftung.de/wp-content/uploads/2021/03/Studie_Kein_Netz_für_Hass_Bundesländervergleich_Hate_Speech_Maßnahmen___Campact-_Institut_für_Demokratie_und_Zivilgesellschaft.pdf (accessed 2025-03-03).
4. Link, D.; Hellingrath, B.; Ling, J. A Human-is-the-Loop Approach for Semi-Automated Content Moderation. In *Proceedings of the ISCRAM 2016 Conference*; ISCRAM: Rio de Janeiro, Brazil, 2016; pp 1—13.
5. Plotnick, L.; Hiltz, S. R. Barriers to Use of Social Media by Emergency Managers. *J. Homel. Secur. Emerg. Manag.* **2016**, *13* (2), 247—277.
6. Simon, E. Das Gesetz zur Bekämpfung von Rechtsextremismus und Hasskriminalität. *Jurist. Rundsch.* **2020**, *2020* (11), 599—607.
7. Krempl, S. *Kampf gegen Hass und Rechts: BKA hat erst 1950 Meldungen bearbeitet*; heise online: Hannover, 2022. https://www.heise.de/news/Kampf-gegen-Hass-und-Rechts-BKA-hat-erst-1950-Meldungen-bearbeitet-7238085.html (accessed 2025-03-03).
8. Chan, C.-J. Normative Regulierung für algorithmische Inhaltsmoderation auf Internet-Plattformen. In *Künstliche Intelligenz, Ethik und Recht*; Knauff, M., Lee, C.-L., Lin, Y.-M., Schröder, M., Eds.; Nomos Verlagsgesellschaft mbH & Co. KG: Baden-Baden, 2024; pp 31—44.
9. Endsley, M. R. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **1995**, *37* (1), 32—64.
10. Lamsal, R.; Harwood, A.; Read, M. R. Socially Enhanced Situation Awareness from Microblogs Using Artificial Intelligence: A Survey. *ACM Comput. Surv.* **2023**, *55* (4), 1—38.
11. Müller, K.; Schwarz, C. Fanning the Flames of Hate: Social Media and Hate Crime. *J. Eur. Econ. Assoc.* **2021**, *19* (4), 2131—2167.
12. Pereira-Kohatsu, J. C.; Quijano-Sánchez, L.; Liberatore, F.; Camacho-Collados, M. Detecting and Monitoring Hate Speech in Twitter. *Sensors* **2019**, *19* (21), 1—37.
13. Kaufhold, M.-A.; Rupp, N.; Reuter, C.; Habdank, M. Mitigating Information Overload in Social Media During Conflicts and Crises: Design and Evaluation of a Cross-Platform Alerting System. *Behav. Inf. Technol.* **2020**, *39* (3), 319—342.
14. Franke, U.; Brynielsson, J. Cyber Situational Awareness — A Systematic Review of the Literature. *Comput. Secur.* **2014**, *46*, 18—31.
15. Husák, M.; Jirsík, T.; Yang, S. J. SoK: Contemporary Issues and Challenges to Enable Cyber Situational Awareness for Network Security. In *Proceedings of the 15th International Conference on*

Availability, Reliability and Security; ACM: Virtual Event, Ireland, 2020; pp 1—10.
16. Goyal, N.; Park, L.; Vasserman, L. "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*; ACM: New Orleans, LA, USA, 2022; pp 1—17.
17. Sultana, S.; Deb, M.; Bhattacharjee, A.; Hasan, S.; Alam, S.; Chakraborty, T.; Roy, P.; Ahmed, S. F.; Moitra, A.; Amin, M. A.; Islam, A. N.; Ahmed, S. I. Unmochon': A Tool to Combat Online Sexual Harassment over Facebook Messenger. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*; ACM: Yokohama, Japan, 2021; pp 1—18.
18. Ayo, F. E.; Folorunso, O.; Ibharalu, F. T.; Osinuga, I. A. Machine Learning Techniques for Hate Speech Classification of Twitter Data: State-of-the-Art, Future Challenges and Research Directions. *Comput. Sci. Rev.* **2020**, *38*, 1—34.
19. Boishakhi, F. T.; Shill, P. C.; Alam, M. G. R. Multi-Modal Hate Speech Detection Using Machine Learning. In *2021 IEEE International Conference on Big Data (Big Data)*; IEEE: Orlando, FL, USA, 2021; pp 4496—4499.
20. Fortuna, P.; Nunes, S. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.* **2019**, *51* (4), 1—30.
21. Sai, S.; Srivastava, N. D.; Sharma, Y. Explorative Application of Fusion Techniques for Multimodal Hate Speech Detection. *SN Comput. Sci.* **2022**, *3* (2), 122.
22. Liu, H.; Burnap, P.; Alorainy, W.; Williams, M. L. Fuzzy Multi-Task Learning for Hate Speech Type Identification. In *The Web Conference 2019*; ACM: San Francisco, CA, USA, 2019; pp 3006—3012.
23. Mazari, A. C.; Boudoukhani, N.; Djeffal, A. BERT-Based Ensemble Learning for Multi-Aspect Hate Speech Detection. *Clust. Comput.* **2024**, *27* (1), 325—339.
24. Mishra, S.; Prasad, S.; Mishra, S. Exploring Multi-Task Multi-Lingual Learning of Transformer Models for Hate Speech and Offensive Speech Identification in Social Media. *SN Comput. Sci.* **2021**, *2* (*2*), 1—19.
25. Paschalides, D.; Stephanidis, D.; Andreou, A.; Orphanou, K.; Pallis, G.; Dikaiakos, M. D.; Markatos, E. MANDOLA: A Big-Data Processing and Visualization Platform for Monitoring and Detecting Online Hate Speech. *ACM Trans. Internet Technol.* **2020**, *20* (2), 1—21.
26. Bäumler, J.; Riebe, T.; Kaufhold, M.-A.; Reuter, C. Harnessing Inter-Organizational Collaboration and Automation to Combat Online Hate Speech: A Qualitative Study with German Reporting Centers. *Proc. ACM Hum. Comput. Interact.* **2025**, *9* (2), 1—31.
27. Hildebrandt, J. R.; Ziefle, M.; Calero Valdez, A. Entscheidungsautonomie und KI — Methodische Hinweise zur Untersuchung von KI-Nutzung in Sicherheitsbehörden. In *Mensch und Computer 2022 — Workshopband*; Gesellschaft für Informatik e.V.: Darmstadt, Germany, 2022.
28. Kaufhold, M.-A.; Bayer, M.; Bäumler, J.; Reuter, C.; Stieglitz, S.; Basyurt, A. S.; Mirbabaie, M.; Fuchss, C.; Eyilmez, K. 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, 2023.
29. Demus, C.; Pitz, J.; Schütz, M.; Probol, N.; Siegel, M.; Labudde, D. 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), 2022; pp 143—153.
30. Bäumler, J.; Kaufhold, M.-A.; Voronin, G.; Reuter, C. 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 e.V.: Karlsruhe, Germany, 2024; pp 1—11.
31. Braun, V.; Clarke, V. Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* **2006**, *3* (2), 77—101.
32. Blandford, A.; Furniss, D.; Makri, S. *Qualitative HCI Research: Going Behind the Scenes*; Morgan & Claypool Publishers: Kentfield, CA, 2016.
33. Wood, L. E. Semi-Structured Interviewing for User-Centered Design. *Interactions* **1997**, *4* (2), 48—61.
34. Siegel, A. A. *Social Media and Democracy: The State of the Field, Prospects for Reform*; Persily, N.; Tucker, J. A., Eds.; Cambridge University Press: Cambridge, MA, USA, 2020; pp. 56—88.
35. Sellars, A. *Defining Hate Speech*; Berkman Klein Center for Internet and Society: Cambridge, MA, USA, 2016. https://www.ssrn.com/abstract=2882244 (accessed 2025-03-03).
36. Banko, M.; MacKeen, B.; Ray, L. A Unified Taxonomy of Harmful Content. In *Proceedings of the Fourth Workshop on Online Abuse and Harms*; Association for Computational Linguistics, 2020; pp 125—137.
37. MacAvaney, S.; Yao, H.-R.; Yang, E.; Russell, K.; Goharian, N.; Frieder, O. Hate Speech Detection: Challenges and Solutions. *PLoS One* **2019**, *14* (8), 1—16.
38. Yin, W.; Zubiaga, A. Towards Generalisable Hate Speech Detection: A Review on Obstacles and Solutions. *PeerJ Comput. Sci.* **2021**, *7*, 1—38.
39. Chetty, N.; Alathur, S. Hate Speech Review in the Context of Online Social Networks. *Aggress. Violent Behav.* **2018**, *40*, 108—118.
40. Gagliardone, I.; Gal, D.; Alves, T.; Martinez, G. *Countering Online Hate Speech*; UNESCO Publishing: Paris, 2015.
41. Paasch-Colberg, S.; Strippel, C.; Trebbe, J.; Emmer, M. From Insult to Hate Speech: Mapping Offensive Language in German User Comments on Immigration. *Media Commun.* **2021**, *9* (1), 171—180.
42. Poletto, F.; Basile, V.; Sanguinetti, M.; Bosco, C.; Patti, V. Resources and Benchmark Corpora for Hate Speech Detection: A Systematic Review. *Lang. Resour. Eval.* **2021**, *55* (2), 477—523.
43. Nielsen, L. B. Subtle, Pervasive, Harmful: Racist and Sexist Remarks in Public as Hate Speech. *J. Soc. Issues* **2002**, *58* (2), 265—280.
44. Vidgen, B.; Harris, A.; Nguyen, D.; Tromble, R.; Hale, S.; Margetts, H. Challenges and Frontiers in Abusive Content Detection. In *Proceedings of the Third Workshop on Abusive Language Online*; Association for Computational Linguistics: Florence, Italy, 2019; pp 80—93.
45. Lehmann, J. Hate Speech: Rechtsansprüche und Rechtsprechung. In *Recht & Netz*; Albers, M., Katsivelas, I., Eds.; Nomos Verlagsgesellschaft mbH & Co. KG: Baden-Baden, 2018; pp 89—126.
46. Stanton, N.; Chambers, P.; Piggott, J. Situational Awareness and Safety. *Saf. Sci.* **2001**, *39* (3), 189—204.

47. Smith, K.; Hancock, P. A. Situation Awareness is Adaptive, Externally Directed Consciousness. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **1995**, *37* (1), 137—148.
48. Bedny, G.; Meister, D. Theory of Activity and Situation Awareness. *Int. J. Cognit. Ergon.* **1999**, *3* (1), 63—72.
49. Endsley, M. R. Design and Evaluation for Situation Awareness Enhancement. *Proc. Hum. Factors Soc. Annu. Meet.* **1988**, *32* (2), 97—101.
50. Ask, T. F.; Knox, B. J.; Lugo, R. G.; Helgetun, I.; Sütterlin, S. Neurophysiological and Emotional Influences on Team Communication and Metacognitive Cyber Situational Awareness during a Cyber Engineering Exercise. *Front. Hum. Neurosci.* **2023**, *16*, 1—24.
51. Cashell, B.; Jackson, W. D.; Jickling, M.; Webel, B. The Economic Impact of Cyber-Attacks. In *CRS Report for Congress RL32331*; Congressional Research Service: Washington, DC, 2004; pp 1—41. https://archive.nyu.edu/bitstream/2451/14999/2/Infosec_ISR_Congress.pdf (accessed 2025-03-03).
52. Maness, R. C.; Valeriano, B. The Impact of Cyber Conflict on International Interactions. *Armed Forces Soc.* **2016**, *42* (2), 301—323.
53. Bada, M.; Creese, S.; Goldsmith, M.; Mitchell, C.; Phillips, E. *Computer Security Incident Response Teams (CSIRTs): An Overview*; The Global Cyber Security Capacity Center: Oxford, 2014. https://ssrn.com/abstract=3659974 (accessed 2025-03-03).
54. Gutzwiller, R.; Dykstra, J.; Payne, B. Gaps and Opportunities in Situational Awareness for Cybersecurity. *Digit. Threats: Res. Pract.* **2020**, *1* (3), 1—6.
55. Ruefle, R.; Dorofee, A.; Mundie, D.; Householder, A. D.; Murray, M.; Perl, S. J. Computer Security Incident Response Team Development and Evolution. *IEEE Secur. Priv.* **2014**, *12* (5), 16—26.
56. Jiang, L.; Jayatilaka, A.; Nasim, M.; Grobler, M.; Zahedi, M.; Babar, M. A. Systematic Literature Review on Cyber Situational Awareness Visualizations. *IEEE Access* **2022**, *10*, 57525—57554.
57. Drodt, M.; Pagel, L.; Biedorf, T. Einbindung Datenschutz und Betriebsrat beim Aufbau eines SIEM. In *Cybersecurity Best Practices*; Bartsch, M., Frey, S., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, 2018; pp 271—284.
58. Huhta, J.-M.; Di Nota, P. M.; Hietanen, T.; Ropo, E. Deriving Expert Knowledge of Situational Awareness in Policing: A Mixed-Methods Study. *J. Police Crim. Psychol.* **2023**, *38* (3), 539—554.
59. Hansson, J.; Borglund, E. A. M. Situation Awareness in Tactical Police Interventions. *J. Police Crim. Psychol.* **2024**, *39* (3), 527—538.
60. Apostolakis, K. C.; Dimitriou, N.; Margetis, G.; Ntoa, S.; Tzovaras, D.; Stephanidis, C. DARLENE — Improving Situational Awareness of European Law Enforcement Agents Through a Combination of Augmented Reality and Artificial Intelligence Solutions. *Open Res. Eur.* **2022**, *1*, 1—26.
61. Razip, A. M. M.; Malik, A.; Afzal, S.; Potrawski, M.; Maciejewski, R.; Yun, J.; Elmqvist, N.; Ebert, D. S. A Mobile Visual Analytics Approach for Law Enforcement Situation Awareness. In *2014 IEEE Pacific Visualization Symposium*; IEEE: Yokohama, 2014; pp 169—176.
62. Hate Aid. *App gegen Hass — Mach mit und werde MeldeHeld*in*; Hate Aid: Berlin, 2020. https://hateaid.org/meldehelden-app/ (accessed 2025-03-03).
63. Alkomah, F.; Ma, X. A Literature Review of Textual Hate Speech Detection Methods and Datasets. *Information* **2022**, *13* (6), 1—22.
64. Jahan, M. S.; Oussalah, M. A Systematic Review of Hate Speech Automatic Detection Using Natural Language Processing. *Neurocomputing* **2023**, *546*, 1—30.
65. Van Aken, B.; Risch, J.; Krestel, R.; Löser, A. Challenges for Toxic Comment Classification: An In-Depth Error Analysis. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*; Association for Computational Linguistics: Brussels, Belgium, 2018; pp 33—42.
66. Kiela, D.; Firooz, H.; Mohan, A.; Goswami, V.; Singh, A.; Ringshia, P.; Testuggine, D. The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 2611—2624.
67. Meske, C.; Bunde, E. Design Principles for User Interfaces in AI-Based Decision Support Systems: The Case of Explainable Hate Speech Detection. *Inf. Syst. Front.* **2023**, *25*, 743—773.
68. Bunde, E. AI-Assisted and Explainable Hate Speech Detection for Social Media Moderators — A Design Science Approach. In *Proceedings of the 54th Hawaii International Conference on System Sciences*; Association for Information Systems: Kauai, Hawaii, USA, 2021; pp 1264—1273.
69. Sontheimer, L.; Schäfer, J.; Mandl, T. Enabling Informational Autonomy Through Explanation of Content Moderation: UI Design for Hate Speech Detection. In *Mensch und Computer 2022 — Workshopband*; Gesellschaft für Informatik e.V.: Darmstadt, Germany, 2022.
70. Koops, B.-J.; Hoepman, J.-H.; Leenes, R. Open-Source Intelligence and Privacy by Design. *Comput. Law Secur. Rep.* **2013**, *29* (6), 676—688.
71. Walther, S.; McCoy, A. US Extremism on Telegram: Fueling Disinformation, Conspiracy Theories, and Accelerationism. *Perspect. Terrorism* **2021**, *15* (2), 100—124.
72. Zapata Rozo, A.; Campo-Archbold, A.; Díaz-López, D.; Gray, I.; Pastor-Galindo, J.; Nespoli, P.; Gómez Mármol, F.; McCoy, D. Cyber Democracy in the Digital Age: Characterizing Hate Networks in the 2022 US Midterm Elections. *Inf. Fusion* **2024**, *110*, 1—16.
73. Khan, Z.; Khan, Z.; Lee, B.-G.; Kim, H. K.; Jeon, M. Graph Neural Networks Based Framework to Analyze Social Media Platforms for Malicious User Detection. *Appl. Soft Comput.* **2024**, *155*, 1—14.
74. Azumah, S. W.; Adewopo, V.; Elsayed, Z.; Elsayed, N.; Ozer, M. A Secure Open-Source Intelligence Framework for Cyberbullying Investigation. In *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*; IEEE: Houston, TX, USA, 2024; pp 1—8.
75. Salminen, J.; Almerekhi, H.; Milenković, M.; Jung, S.-G.; An, J.; Kwak, H.; Jansen, B. Anatomy of Online Hate: Developing a Taxonomy and Machine Learning Models for Identifying and Classifying Hate in Online News Media. *Proc. Int. AAAI Conf. Web Soc. Media* **2018**, *12* (1), 1—10.
76. Mollas, I.; Chrysopoulou, Z.; Karlos, S.; Tsoumakas, G. ETHOS: A Multi-Label Hate Speech Detection Dataset. *Complex Intell. Syst.* **2022**, *8* (6), 4663—4678.
77. Ousidhoum, N.; Lin, Z.; Zhang, H.; Song, Y.; Yeung, D.-Y. Multilingual and Multi-Aspect Hate Speech Analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*; Association for

Computational Linguistics: Hong Kong, China, 2019; pp 4674—4683.
78. Abburi, H.; Parikh, P.; Chhaya, N.; Varma, V. Fine-Grained Multi-Label Sexism Classification Using a Semi-Supervised Multi-Level Neural Approach. *Data Sci. Eng.* **2021**, *6* (4), 359—379.
79. Parikh, P.; Abburi, H.; Badjatiya, P.; Krishnan, R.; Chhaya, N.; Gupta, M.; Varma, V. Multi-Label Categorization of Accounts of Sexism Using a Neural Framework. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*; Association for Computational Linguistics: Hong Kong, China, 2019; pp 1642—1652.
80. Bosco, C.; Patti, V.; Bogetti, M.; Conoscenti, M.; Ruffo, G.; Schifanella, R.; Stranisci, M. Tools and Resources for Detecting Hate and Prejudice Against Immigrants in Social Media. In *2017 Annual Convention of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour*; AISB: Bath, United Kingdom, 2017; pp 79—84.
81. Capozzi, A. T.; Lai, M.; Basile, V.; Poletto, F.; Sanguinetti, M.; Bosco, C.; Patti, V.; Ruffo, G.; Musto, C.; Polignano, M.; Semeraro, G.; Stranisci, M. Computational Linguistics Against Hate: Hate Speech Detection and Visualization on Social Media in the "Contro L'Odio" Project. In *Proceedings of the Sixth Italian Conference on Computational Linguistics*; CEUR Workshop Proceedings: Bari, Italy, 2019; pp 1—6.
82. Ludwig, T.; Reuter, C.; Siebigteroth, T.; Pipek, V. CrowdMonitor: Mobile Crowd Sensing for Assessing Physical and Digital Activities of Citizens during Emergencies. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*; ACM: Seoul, Republic of Korea, 2015; pp 4083—4092.
83. Reuter, C.; Ludwig, T.; Pipek, V. Ad Hoc Participation in Situation Assessment: Supporting Mobile Collaboration in Emergencies. *ACM Trans. Comput. Hum. Interact.* **2014**, *21* (5), 1—26.
84. Riebe, T.; Kaufhold, M.-A.; Reuter, C. The Impact of Organizational Structure and Technology Use on Collaborative Practices in Computer Emergency Response Teams: An Empirical Study. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5* (CSCW2), 1—30.
85. Mayring, P. Qualitative Content Analysis. *Forum Qual. Sozialforsch./Forum Qual. Soc. Res.* **2000**, *1* (2), *1*—10.
86. Brennan, R. L.; Prediger, D. J. Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educ. Psychol. Meas.* **1981**, *41* (3), 687—699.
87. Kuckartz, U.; Rädiker, S. *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*; Springer International Publishing: Cham, 2019.
88. Geschke, D.; Klaßen, A.; Quent, M.; Richter, C. *#Hass im netz: Der schleichende angriff auf unsere demokratie. eine bundesweite repräsentative untersuchung*; Institut für Demokratie und Zivilgesellschaft (IDZ): Jena, 2019. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf (accessed 2025-03-03).
89. Riebe, T.; Bäumler, J.; Kaufhold, M.-A.; Reuter, C. Values and Value Conflicts in the Context of OSINT Technologies for Cybersecurity Incident Response: A Value Sensitive Design Perspective. *Comput. Support. Coop. Work* **2023**, *33*, 205—251.
90. Cuijpers, C. Legal Aspects of Open Source Intelligence — Results of the VIRTUOSO Project. *Comput. Law Secur. Rep.* **2013**, *29* (6), 642—653.
91. Vrysis, L.; Vryzas, N.; Kotsakis, R.; Saridou, T.; Matsiola, M.; Veglis, A.; Arcila-Calderón, C.; Dimoulas, C. A Web Interface for Analyzing Hate Speech. *Future Internet* **2021**, *13* (3), 1—18.
92. Zufall, F.; Hamacher, M.; Kloppenborg, K.; Zesch, T. A. Legal Approach to Hate Speech — Operationalizing the EU's Legal Framework Against the Expression of Hatred as an NLP Task. In *Proceedings of the Natural Legal Language Processing Workshop 2022*; Association for Computational Linguistics: Abu Dhabi, United Arab Emirates (Hybrid), 2022; pp 53—64.
93. Mandl, T. KI-Verfahren für die Hate Speech Erkennung: Die Gestaltung von Ressourcen für das maschinelle Lernen und ihre Zuverlässigkeit. In *Digitale Hate Speech*; Jaki, S., Steiger, S., Eds.; Springer: Berlin, Heidelberg, 2023; pp 111—130.
94. Caselli, T.; Basile, V.; Mitrović, J.; Granitzer, M. HateBERT: Retraining BERT for Abusive Language Detection in English. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*; Association for Computational Linguistics: Online, 2021; pp 17—25.
95. Wang, Y.; Yao, Q.; Kwok, J. T.; Ni, L. M. Generalizing from a Few Examples: A Survey on Few-Shot Learning. *ACM Comput. Surv.* **2021**, *53* (3), 1—34.
96. Bayer, M.; Kaufhold, M.-A.; Reuter, C. A Survey on Data Augmentation for Text Classification. *ACM Comput. Surv.* **2023**, *55* (7), 1—39.
97. Keim, D. A.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H. Visual Analytics: Scope and Challenges. In *Visual Data Mining*; Simoff, S. J., Böhlen, M. H., Mazeika, A., Eds.; Springer: Berlin, Heidelberg, 2008; pp 76—90.
98. Kaufhold, M.-A.; Reuter, C.; Ludwig, T. Big Data and Multi-platform Social Media Services in Disaster Management. In *International Handbook of Disaster Research*; Singh, A., Ed.; Springer Nature: Singapore, 2023; pp 573—593.
99. Bruns, A. After the 'APIcalypse': Social Media Platforms and their Fight against Critical Scholarly Research. In *Disinformation and Data Lockdown on Social Platforms*; Walker, S., Mercea, D., Bastos, M., Eds.; Routhledge: New York, 2021; pp 14—36.
100. Kaufhold, M.-A.; Riebe, T.; Bayer, M.; Reuter, C. '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*; ACM: Honolulu, HI, USA, 2024; pp 1—16.
101. Komárková, J.; Husák, M.; Laštovička, M.; Tovarňák, D. CRUSOE: Data Model for Cyber Situational Awareness. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*; ACM: Hamburg, Germany, 2018; pp 1—10.
102. Kaufhold, M.-A.; Bayer, M.; Reuter, C. Rapid Relevance Classification of Social Media Posts in Disasters and Emergencies: A System and Evaluation Featuring Active, Incremental and Online Learning. *Inf. Process. Manag.* **2020**, *57* (1), 1—32.
103. Ley, B.; Ludwig, T.; Pipek, V.; Randall, D.; Reuter, C.; Wiedenhoefer, T. Information and Expertise Sharing in Inter-Organizational Crisis Management. *Comput. Support. Coop. Work* **2014**, *23*, 347—387.
104. Alicioglu, G.; Sun, B. A Survey of Visual Analytics for Explainable Artificial Intelligence Methods. *Comput. Graph.* **2022**, *102*, 502—520.
105. Haupt, M. R.; Chiu, M.; Chang, J.; Li, Z.; Cuomo, R.; Mackey, T. K. Detecting Nuance in Conspiracy Discourse: Advancing Methods in Infodemiology and Communication Science with Machine

Learning and Qualitative Content Coding. *PLoS One* **2023**, *18* (12), 1—29.
106. Imran, M.; Castillo, C.; Diaz, F.; Vieweg, S. Processing Social Media Messages in Mass Emergency: A Survey. *ACM Comput. Surv.* **2015**, *47* (4), 1—38.
107. Creswell, J. W.; Poth, C. N. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 4th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2018.
108. Flick, U. *An Introduction to Qualitative Research*, 6th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2018.