"It is unfair, and it would be unwise to expect the user to know the law!" – Evaluating reporting mechanisms under the Digital Services Act

Marie-Therese Sekwenz
Delft University of Technology
Technology, Policy and Management, Multi-Actor Systems, Organization & Governance
Delft, Netherlands
m.t.sekwenz@tudelft.nl

Ben Wagner
IT:U Interdisciplinary Transformation University Austria
Human Rights and Technology Research Group
Linz, Austria
Delft University of Technology
Technology, Policy and Management, Multi-Actor Systems, Organization & Governance
Delft, Netherlands
Inholland
Media, Technology and Society Research Group, Applied Research x Creativity Research Center
The Hague, Netherlands
B.Wagner@tudelft.nl

Simon Parkin
Delft University of Technology
Technology, Policy and Management, Multi-Actor Systems, Organization & Governance
Delft, Netherlands
S.E.Parkin@tudelft.nl

Abstract

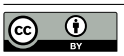
Platforms have a problem with harmful or illegal content online. Flagging, which is an empowering tool for users to report violating content. A new European Union law, the Digital Services Act (DSA), seeks to harmonize the regulation of the flagging process. This paper examines how these flagging mechanisms support user action through semi-structured interviews (N=12) with regulatory authorities and professional reporting experts, using a walkthrough approach (with case studies based on flagging systems on Facebook and TikTok). We found tensions between the empowerment of users with additional reporting options and how it burdens users within service interfaces and processes; users need to understand the law, participate in a legal process, and differentiate between legal options and terms of service. Design choices, like the length of necessary reporting steps, also impacted expectations on the transparency of the reporting process. We close with design insights on support for users and stakeholders in the reporting process.

CCS Concepts

• **Social and professional topics** → **Governmental regulations**.

Keywords

Digital Services Act, Flagging, Notice Action Mechanism, Content Moderation



FAccT '25, Athens, Greece

ACM ISBN 979-8-4007-1482-5/25/06
https://doi.org/10.1145/3715275.3732036

ACM Reference Format:
Marie-Therese Sekwenz, Ben Wagner, and Simon Parkin. 2025. "It is unfair, and it would be unwise to expect the user to know the law!" – Evaluating reporting mechanisms under the Digital Services Act. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3715275.3732036

1 Introduction

Offensive content regularly appears on social media platforms (SMPs), undermining the experience and safety of users. Removing harmful/illegal content is a central task of content moderation on SMPs [26]. As content moderation does not catch all violations, users can report through flagging items or content, also referred to as a notice action, or reporting mechanism (RM) [34]. Content reporting helps platforms catch violations missed by automation and reflects user experiences. For example, in February 2025, a technical failure on Instagram led to a spike in reports after users were exposed to graphic content, prompting a public apology from Meta [4].

Regulatory efforts around the world are moving to reconcile activity on these platforms with legal oversight, as some content goes beyond breaking a platform's Terms of Service (ToS) and into being recognizably against the law. In the European Union, the most important regulation in this domain is the Digital Services Act [15]. It is regulating the process of content moderation, albeit not setting material rules for how the content itself has to be moderated [53]. This raises questions around how the additional user rights afforded through the DSA are empowered, and users appropriately protected from harmful, illegal content while navigating the law amid existing platform-specific ToS.

This study seeks to situate the challenges faced by users and various stakeholders within the framework of the DSA reporting mechanisms (RMs), drawing on interviews with legal and regulatory experts and a heuristic walk-through of reporting interface designs, based on real-world implementations on SMPs.

Our overarching research question therefore focuses on the necessary intersection between legal and design aspects of RMs in scope of the DSA: How do usability and compliance interact within the DSA's requirements of illegal content flagging mechanisms under Article 16 (which mandates that online platforms must implement user-friendly and easy to access systems for reporting illegal content)? Our research questions explore the experiences and expectations of experts regarding what constitutes usability that is compliant with the DSA: RQ1: What are perceived to be different users' DSA-flagging behaviors on social media platforms impacting usability?; RQ2: How do the experts interpret the differences in usability of DSA-reporting on use case examples of prompting heuristic walkthroughs?, and; RQ3: How do the experts explain effects of DSA-reporting in the content moderation process through the design of DSA compliant reporting interfaces?

The goal of the DSA is to "set out harmonized rules for a safe, predictable, and trusted online environment" (Article 1 DSA), and it applies to platforms that offer their 'services' to users in the EU. Similar to Section 230 of the U.S. Communications Decency Act [57], which limits liability for problematic speech on SMPs, the DSA imposes no obligation to 'generally monitor' content (Article 8 DSA). DSA flagging covers illegal content reporting mechanisms as outlined in Article 16(1) DSA. These must be 'easy to access' and 'user friendly' as key obligations under EU law (Recital 50, Recital 52, Article 16 (1) DSA).

This represents a new convergence of legal-interface design. The remit to bring legally-recognized reporting structures direct to users, and at the same time ensure easy-to-use processes, brings potential for a range of tensions in reporting interfaces and associated processes. For instance, users often do not understand the categories or processes involved in reporting content and follow a "hedonic mindset" [43, p. 9]. Flagging is a multi-actor process and "a complex interplay between users and platforms, humans and algorithms, [implementing] social norms and regulatory structures of social media." [10, p. 412].

We argue that the way SMP interfaces are designed and the affordances these interfaces provide are increasingly driven by legal frameworks. As such, we have developed a methodology (Section 4) that takes these legal frameworks into account, providing a systematic perspective on how both law [7] and technology [61] shape human behavior. To this end, we conduct a study with regulators and other expert users rather than the average users of the RMs; these participants have knowledge of the hard bounds of legal requirements. This is reflected in our adaptations to the heuristic walkthrough approach.

Our findings (Section 5) contribute to the developing discourse around *legal design*, and methods to examine social, legal and technological dimensions in tandem. Our expert participants noted that the DSA provides new possibilities for empowering users and flaggers in the reporting process, which on the other hand, also burdens users due to the need for legal understanding of filing illegal content reports, compounded by a need to distinguish illegal content from content against an SMP's ToS.

Our results also illustrate ambiguities in the role of TFs and how non-expert users can reasonably engage in flagging processes. We found tensions between DSA obligations of creating 'easy to access' and 'user friendly' RMs and differences in design choices on SMPs, which can further complicate the reporting process. Besides, interface designs may impact the reporting process, user experience, and what can be expected of the user. Additionally, experts also highlighted that SMP moderation activity remains opaque even to them, highlighting the burden on non-expert users.

Despite the promise of the Digital Services Act to create safer online spaces through user-friendly reporting mechanisms, the complexity of legal requirements and opaque platform practices could undo efforts to encourage user reporting, leaving open questions as to whether reporting systems can effectively empower non-expert users. In discussing our findings (Section 6), we contribute recommendations for practitioners including how to embody legally-recognized user rights within (usable) interface designs, feedback loops for users reporting content and increased process information, and better cooperation between actors in the complex content moderation ecosystem.

2 Background

Where there has been prior research on 'flagging' of content, this has for example been in a USA context [10, 51] where a flag is a report of content, as a mechanism of governance. Increasingly under emerging legislation around the world, flagging can encompass the reporting of content under either Terms of Service (ToS) or as illegal content (which is then subject to legal process 'outside of' the platform on which it is reported). Addressing legal/regulatory compliance and design aspects in tandem, specifically for (user) flagging, is a notable research gap (where prior work has examined platform-driven reporting structures [44, 62]), and is especially relevant for Europe's Digital Services Act (DSA) [15] which creates new rules for reporting systems in Article 16 for online platforms operating in the European Union according to Article 2. The DSA imposes stringent obligations on SMPs for example on the design of illegal content RMs (Article 16(2) DSA). To ensure compliance, national Digital Service Coordinators (DSCs) and, ultimately, the European Commission oversee enforcement, with fines of up to 6% of annual global turnover for violations (Articles 49–52, 56, and 72 DSA).

Other jurisdictions like the UK [46], India's [42], or California's [5] echo the DSA's aim of regulating SMPs. In the EU Germany [17], France [19], and Austria [47], had laws prior to the DSA.

Unlike platform ToS, which platforms set unilaterally, the DSA establishes 'must-moderate' criteria which can overlap with and complement a platform's ToS regarding problematic content such as deepfakes [18], antisemitic material [37], or harm against animals [41]. Therefore, RMs are legal instruments of the "notice and takedown policy" [29, p2].

Few prior EU regulations demand similarly transparent and 'user-friendly' mechanisms to report content [12–14]. The DSA specifically requires that Reporting Mechanisms (RMs) be "clearly identifiable, located close to the information in question" and at least as

accessible as mechanisms for reporting ToS violations (Recital 50, Article 14 DSA). Users must be able to flag multiple pieces of content in a single notice (Recital 53 DSA), and those reporting must be informed and offered means of redress (Recital 58 and Article 16(5) DSA). Additionally, platforms must process notifications promptly, objectively, and transparently, disclosing any automation involved (Recital 61 and Article 16(6) DSA).

The DSA also introduces Trusted Flaggers (TFs), who receive priority treatment for their notices. Accredited by DSCs (see Section 5.2 Recital 61, Article 22 DSA), TFs are expected to demonstrate expertise and unbiased judgment, further professionalizing the moderation process. These Trusted Flaggers would be active alongside Professional Reporters (representing users), and normal users. Article 16(1) DSA regulates 'illegal' content reporting. Because 'illegal' content reporting requires understanding of the legal domain (difference of ToS and law, legal jargon, etc.) and is a "barrier for [normal] user engagement"[51, p. 5] DSA-RMs are an expert system.

3 Related Work

In studying flagging and reporting behaviors in online gaming environments, Kou and Gui [34] note that flagging content is not simply a problem-solving exercise, as content can differ in both 'flaggability' and toxicity. Jhaver et al. study end-user perceptions toward social media content moderation tools [30], finding that participants were reluctant to set content filters, and expected platforms to moderate offensive content on their behalf.

Wei et al leveraged insight from experts to evaluate online safety advice for users in the context of online hate [60], finding that muting content is regarded as a 'best practice' but that it should also be reported to SMPs and civil society organizations (mirroring the social role of flagging mentioned by our participants, Section 5.6.5).

Related research [31] surveying hundreds of US users found that participants managed only their own feeds using content-hiding features, as a means to still allow the free speech of others. Zhang et al. [62] raise questions about the transparency of the reporting process and how content moderation effort is distributed between users and an SMP, where our work identifies many moving parts between stakeholders (all while the SMPs themselves remain as something of a mystery to other stakeholders). Where Gomez et al. [20] raise concerns about the arbitrary nature of content moderation decisions (especially where AI is involved) and the impact on human rights, our findings (Section 5.4) indicate that reporting entities acting on behalf of users equally have little insight into decisions around content removal (thereby compounding harms to affected users). Similarly, Naab et al. [44] explore reporting as a visible bystander intervention, prompting a need for users and platform moderators to have a dialogue.

Previous work on design and compliance with the NetzDG for Reporting Mechanisms (RMs) focused on how Facebook and Twitter nudge users to use Terms of Service (ToS) rather than legal reporting [58], by assessing necessary clicks to report under the law or ToS. Our findings signal that general users are not equipped with the necessary legal knowledge to navigate reporting mechanisms effectively, shifting the burden on users to get educated in how to use the RMs (potentially while upset by content), where a knowledge asymmetry has been noted in prior work on how platform consent functions [36].

The growing body of work around deceptive (or 'dark') patterns in website interfaces and processes is also relevant to our study (e.g., [38, 39], see Section 5.6.3). Gray et al. [24] highlight the need for having legal knowledge in the design domain, seeing it not solely as a limitation but also as empowerment, and actively fostering the integration of legal and UX design [21]. Our findings explored options for integrating flagging into SMP interfaces, but also highlight a platform's power over that design as the 'host', and the interplay between different user groups and their particular needs (e.g. users and Trusted Flaggers). We also characterize challenges in how the navigation of legal-process interfaces relies on prior understanding of the law, in order to be able to competently exercise one's choices.

4 Methodology

Here we describe our approach to answering our Research Questions, by way of an expert interview study. The need to study legal aspects of RM is evident in complying to Article 16 DSA, which regulates the design of RMs.

It has been noted that "legal and procedural complexities often overshadow usability concerns, resulting in reporting interfaces that are challenging for average users" [51, p. 5]. Hence, we interviewed practitioners with knowledge of the legal complexities of the DSA. Here, we bring together the perspectives of Professional Reporters who address legal aspects of user needs daily, experts specializing in challenges related to DSA-reporting, and regulators with expertise in enforcement and evidence within the DSA-regime. Based on their knowledge of the DSA, and prior experience interacting with users in actively reporting content, this serves to anticipate challenges for users who are not familiar with the legally-recognized reporting process, nor how it sits alongside other reporting paths such as ToS.

4.1 Study design

The interview sessions were semi-structured and are based on main Recitals and Articles governing reporting mechanisms in the DSA, which were hinging on the overarching research questions. We did link the relevant provisions in the interview protocol to each question in the Appendix A.1. Each interview opened with questions to understand the work of the participants (with interview questions included in the Appendix A.1.1) , their experience, and their connection to users; this includes whether there is a particular group or kind of user that they represent. The interview would then move into understanding reporting/flagging under the DSA, and expectations on different user groups, relating this to the existing connection the participant has with users who actively report content.

The goal was to discuss how "professionalized" reporters would navigate the interfaces, and how this relates to their expectations for the users they represent in their work, who through the DSA would have similar expectations imposed upon them.

After opening questions, the interview would move to a 'legal design' heuristic walkthrough exercise, as detailed in Section 4.1.1. After the heuristic walkthrough and discussion of live reporting

options, participants were prompted to consider how interfaces should or could be re-engineered, to better encourage and best support reporting by users. The walkthrough mapping would then serve as discussion points for legal requirements and design options.

The interview protocol (as in the Appendix A.1) differed slightly for Regulators ('Regs'), and for User Representatives ('UReps'). The interview protocol for Regs was focusing on oversight activities of the DSA, expected platform compliance, and expectations of TFs and transparency through it. P-UReps on the other hand were providing practical details on flagging, operational challenges, user concerns and expectations.

4.1.1 Legal design heuristic walkthrough. Each interview includes a walkthrough exercise. Our walkthrough approach is based on a heuristic walkthrough [50], which in brief, first examines a user task, then moves to open examination of the interface. Here, the user task is reporting a piece of content on a social media platform. We explored usage through structured examples (such as reporting in different countries, and ToS vs. illegal content). We have adapted the heuristic walkthrough method as, to our knowledge, no visualization method already exists to represent legal concepts in a task flow across interfaces, for use in interface assessment research.

The core adaptation is that we represent legal requirements, representing individual interface screens as steps in a process, alongside salient legal-process information appearing in each step. This is demonstrated in showcasing details of information users have to provide as indicated in Figure 1. Similar to Gray et al.'s examination of dark patterns [24], we include information pertinent to the 'task flow' of a law-related interface. In this case the structure of the steps in the interface process, and specifically the information that a user would have to understand on each screen, and conversely, prompts for the user to provide information. These represent the complexity and amount of text that must be parsed or provided per reporting step. We provide a comparison for how illegal reporting is implemented on Facebook's RM in all different EU member states (constituting in four design versions: F (France), DE (Germany), AT (Austria), and all other member states and a comparison for hiding content on Facebook.

The adapted heuristic walkthrough used a Miro board which was populated with high-level visualizations of wireframes (from the 5.21.2024) of the reporting interfaces and processes for both Facebook and TikTok, as case study examples of design implementations in the scope of the DSA. We focused on elements such as process length or showing all text per process step, or visualizing UI solutions per member state, as illustrated in Figures 4-15 in the repository.[1] This is in contrast to, for example, Gray et al.'s use of wireframes to categorize interface elements according to interface design patterns [23].

The visualizations of reporting details are illustrated in the Figure 1 (and can be found in the repository). We used colors (e.g., blue for ToS and red for the law) annotation labels (describing main RM categories). The visualizations were explained with an overview by the interviewer, in each session. Participants were then invited to think aloud [2, 45] and provide immediate thoughts about the wireframes. Afterwards participants were able to discuss freely aspects relevant to them without any order imposed on them akin to the second, open phase of a heuristic walkthrough. The interviewer or the participants controlled the Miro board. If asked to the interviewer steered participants but shared the screen via Teams. Participants were provided a link to freely explore the Miro board themselves.

[1] https://doi.org/10.4121/16efb52d-b94d-4180-a5f6-89d93028a8c2

4.2 Recruitment

Int	ID	Group	Role	Yrs	Dist.
1	P-Reg1	Reg.	Data analyst	4	A
1	P-Reg2	Reg.	Legal officer	3.5	A
2	P-URep1	URep.	Policy advisor	3	B
3	P-URep2	URep.	Lawyer	25	A
3	P-URep3	URep.	Lawyer	5	A
4	P-URep4	URep.	Legal advisor	3	A
5	P-URep5	URep.	Project mgr., reporting DB	12	C
6	P-Reg3	Reg.	Legal officer	12	A
6	P-Reg4	Reg.	Legal officer	12	A
7	P-URep6	URep.	Dep. Dir., fact-checking org.	4	B
8	P-URep7	URep.	Exec. dir., NGO	7	C
9	P-Reg5	Reg.	Dir. of regulatory ops	10	A
10	P-Auth1	PubAuth	Legal officer	10	D
11	P-URep8	URep.	Senior policy analyst	5	B
12	P-URep9	URep.	Lawyer	1	A

Table 1: Interview Participants: Group, Role, Years of Experience, and Distance to Users. Distance Codes: A = Direct interaction, B = Policy representation, C = Builds tools for users, D = Regulator/public authority.

Similar to Wei et al.'s study of online expectations to reduce online hate content [60], we interviewed subject matter experts. Elias and Bezerianos echo the value of experts who "provide insights into usability challenges that may be inaccessible to regular users" [11, p. 4].

Expert participants will have an awareness of the immutable legal imperatives/constructs which an interface must include (much like how in dark patterns research, a consent process must allow for both 'Accept' and 'Decline', for instance [38], and cannot be removed even were it deemed to improve usability). Further, we build on the reasoning of Habib & Cranor [27] when assessing the usability of privacy choice mechanisms, that recruiting experts can inform an estimate of the burden on users to use a policy/legal interface, including the difficulty for a user to enact specific choices and the alignment of user-facing interface designs with regulatory and legal-expertise needs. This is especially the case with the DSA, as experts have knowledge of later 'offline' steps outside the interface.

Prospective participants were approached through professional contacts, either directly or via an existing contact using snowball sampling to enrich the selection. Interviews were conducted online via Microsoft Teams, and lasted approximately one hour each. Interviews were audio- and video-recorded to inform the analysis of the legal-heuristic walkthrough. Reporting expertise covers illegal content, disinformation, hate speech, human rights aspects,

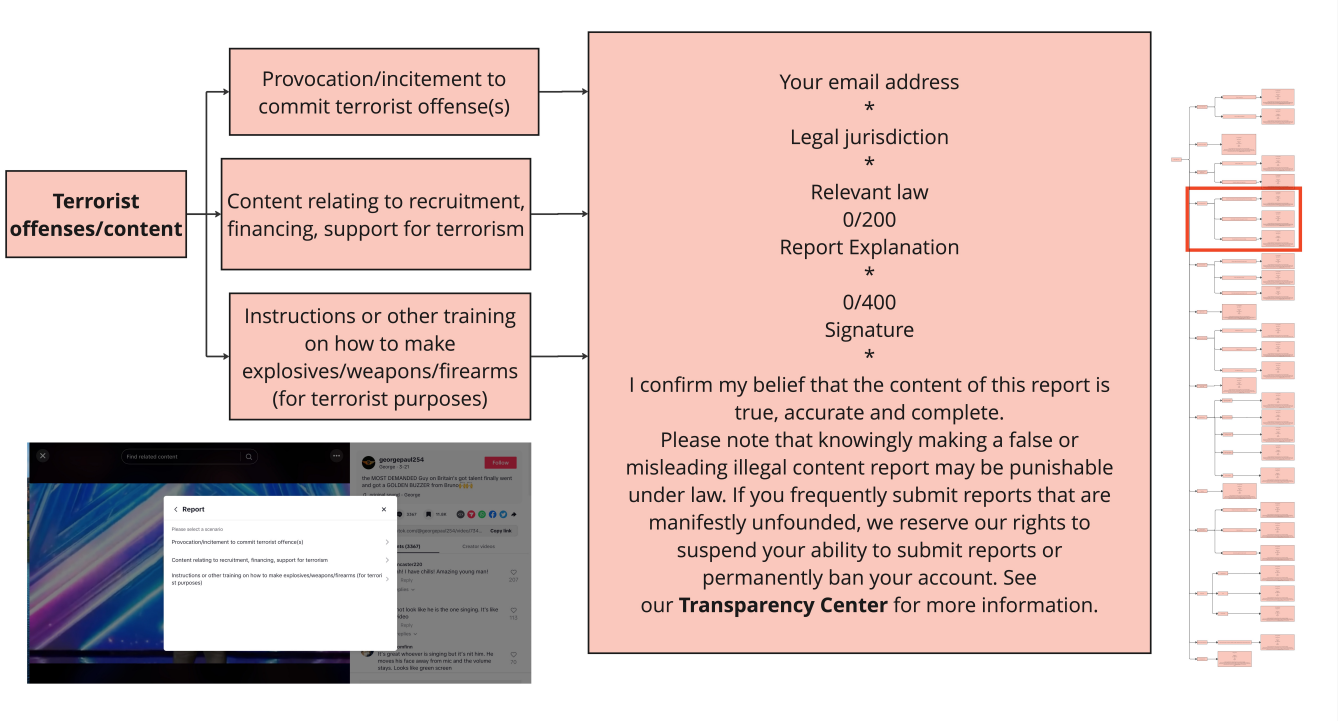


Figure 1: Selection of Wireframe Element for 'Terrorist offense/content' under Illegal Content Reporting on TikTok. Including screenshots of the original interface design, a wireframe element laying out the entire text in that reporting step in a bird's eye perspective, and an overview of where this reporting step is located in the entirety of illegal content reporting options on TikTok.

political advertising and copyright issues. Each Reg. and P-Auth. interview (see Table 1) relates to one country within the EU; to preserve anonymity we do not declare which country. The 'UReps' organizations are all active within an EU country, the UK or the US. Participants were not remunerated for their participation.

4.3 Research ethics

The study was reviewed and approved by our institution's Human Research Ethics Committee (HREC). We gathered consent from participants after they reviewed the informed consent form. Participants could halt the interview at any time, with discussions up to that point being discarded. An overview of the question set was shared with each participants requesting it ahead of the interview. Transcripts were produced from interview audio, then the original recordings deleted and transcripts used for subsequent analysis. There were three interviews (1, 3, and 6) where two participants from the same organization participated. This was suggested and arranged by the contacted participant, and was not initiated by the researchers.

4.4 Data analysis

All interviews were conducted in English. There was one interviewer for all of the interviews.

We applied Thematic Analysis (TA) [6] to analyze the interviews, specifically a codebook-based approach [3] (codebook in Appendix 3). We employed TA to characterize interactions between stakeholders in the complex content reporting and moderation ecosystem, and because identifiable elements of the DSA and online reporting environment are in a state of transition, with stakeholders shifting their views and expectations around them. The lead author, who conducted the interviews, was one coder (Coder1). A second coder, Coder2, is a co-author. Coder1 has a background mainly in law and design, computer science, and journalism; Coder2 a computer science and (security and privacy) usability background.

All transcripts were coded separately by both coders, in small batches of transcripts; codebook meetings occurred after every 2-3 transcripts. After each such batch, both coders would compare and discuss their codebooks. Codebooks were combined by Coder1 after each round of discussion, with both coders then using the reconciled codebook in the next round. Although there was not great consensus between coders, this is acceptable for this kind of TA [3]; crucially, neither coder disagreed with the coding of the other. For instance, Coder1 was looking for explicit mention of "user friendly" discussion (relative to the DSA), and Coder2 was interpreting discussion of users' primary tasks (browsing social media) against reporting – as a secondary task – as a critical consideration for user-friendliness.

There were also many codes which recorded one stakeholder's role in a process or situation, and codes mirroring the reciprocal process from another stakeholder's position (e.g., "uncertainty in

reporting" and "legal uncertainty"). Both coders chose to carry this approach through the whole coding process – to the extent that a passage of text can have multiple codes and classifications – to facilitate coding of a complex, multi-stakeholder set of processes.

4.5 Limitations

We did not talk to users directly. Since the DSA regulates how flagging for "illegal" content has to be organized, focusing on experts that can explain in detail their – and users' – challenges focuses more clearly on our research questions. As a study design decision, we chose instead to use NGOs and flaggers as a proxy for the motivated user to understand their challenges with legal aspects of flagging, similar to [59] and [27]. The most meaningful study of users would be to meet them at the exact moment they are offended by a piece of content and motivated to report it, following the process until after a moderation decision was made, potentially leading to the case reaching the internal complaint-handling system of the platform (Article 20 DSA), or an out-of-court dispute settlement body (Article 21 DSA). Given the positive response we had from participants, we envisage this as future work with these stakeholders, when users liaise with them for guidance.

Our wire-frame visualizations were designed to highlight particular legal design challenges, representing a potential bias in the issues that we aimed to discuss, but nonetheless enriched the discussion with participants; many participants noted the wireframe visualizations as being eye-opening or very useful in their own right as a knowledge artifact – future work will engage more closely with regulators and reporting entities, to explore how these wireframe visualizations can be used further as conversation devices for policymakers.

5 Findings

We structure our findings according to seven themes that emerged from the analysis, as below, with each theme discussed in a subsection. We note which part of the interview protocol they relate to (as in the 'Semi-structured Interview Guidelines' in the Appendix A.1): Content Reporting – A Multi-Actor Process [1.1]; Professional reporting [1.2]; Content moderation and moderation decisions [1.2], Differentiating illegal content from ToS [1.3]; Differences in SMP (via the heuristic walkthrough in SMPs) [1.4]; Design of RMs [1.5]; and; The process of change with DSA [1.5]. Each subsection is opened with an overview summary.

5.1 Content reporting – a multi-actor process

DSA flagging as a form of user empowerment is dependent on a user's motivation to report harmful content while engaging within a legal process, and a need for legal knowledge to navigate Reporting Mechanisms (RMs). Participants did not expect that users should know the law. Regulators and user representatives need to navigate how best to interact with users, to ensure that their legal experience can reach users, for them to be able to report content effectively.

5.1.1 Connection to users. The 12 participants work on digital rights in member state and global contexts (P-URep1, P-URep8), directly enforce user rights through flagging (P-URep2, P-URep3, P-URep4, P-URep9), make data on flagging available to researchers (P-URep5), make users/journalists "aware about platform functionalities" (P-URep7), ensure accurate information environments online for users (P-URep6), or are public authorities enabling flagging for EU member state authorities (P-Auth1), Regulators from three EU member states (P-Reg1-5).

On Social Media Platforms (SMPs), users are seen to focus on "just scroll, [...] rather than [...] report stuff" (P-URep7). Reporting is described as a secondary user task on a platform, where the form of content can influence users' ability and motivation to report it, as "the casual user is gonna be like, 'I don't have time for that' unless they're already like, really motivated" (P-URep5).

Regarding motivations, participants represent users in a range of contexts, such as "consumer protection", users affected or witnessing "digital violence" (P-URep9), like "dehumanizing insults", "defamation", "threats", or "unconsensual or deepfake images", to help users to "file complaints" (P-URep9). Yet other participants support users with in-house legal expertise (P-URep2, P-URep3, P-URep4, P-URep9). The consultation of legal experts is also brought up in the design of reporting mechanisms themselves, where "pages are advising the user even to first consult a lawyer before they fill out this form" (P-URep9).

5.1.2 Legal representation for users. RMs for illegal content, require some degree of legal understanding in their design. According to P-URep9, users must provide an "explanation to why they think the content is illegal" as stated in Article 16(2)(c) DSA. Such statements must be given in *bona fide* in accuracy and completeness as lined out in literal d of the same provision.

P-URep4, on the other hand, does not see Social Media Platforms (SMPs) as having high legal expectations from users in the reporting process. Expectations on users to know the law are evident in design decisions as on TikTok, "stating the jurisdiction" (P-URep9, P-URep8), which is seen as a legal version of asking "[in] what country do you have your permanent residency"? P-Reg1 notes: "[If] users [...] were not able to report the content or [...] it was not handled [...] according to the DSA, [...] users have the right to complain about alleged violations" to regulators. Where Digital Service Coordinators (DSCs) step in to provide help to users in enforcing their rights "in their struggle to find it, or in their despair because they don't get an answer" DSCs provide help to users and are even "contacted by lawyers because the user didn't get...very far with this complaint." (P-Reg1). In many cases however information to "judge the case" is not provided by users (P-Reg1).

5.2 Content reporting

Professional reporting emphasizes the importance of supporting users in the DSA flagging process, such as reporting illegal content through Trusted Flaggers (TFs) acting on behalf of users and utilizing dedicated reporting tools. Empowering and educating users are considered crucial for fostering their active engagement in the reporting process.

5.2.1 Upholding representation of users. TFs are prioritized in the content moderation process (P-Reg5, P-Auth1), though how exactly prioritization is facilitated and against whom "is up to the platforms" (P-URep5). We note here that TFs report according to Article 22 of the DSA, representing a specific expertise and competence in

the DSA flagging process. However, queuing of flags is opaque according to P-Reg2, "if a post gets 1000 notices of being illegal or not compliant with the [Terms of Service] that must get more priority than a post with like one notice, right?".

The TF role is still new and during our interviews only a few were accredited by DSCs. Several participants of this study were in the process of being accredited. Some SMPs have existing 'pre-DSA' flagging programs for similar fast-lane treatment (P-URep7). TFs have the "big problem [of] resources" (P-Auth1), echoed by P-URep7, since TFs are not allocated any specific funding with their accreditation (P-Reg3, P-Reg4, P-Auth1). Resources spent can be in conflict of outcomes – "maybe at the end of the day [...] you've removed, [...] 10 videos with 100 views each and you know, 10s of thousands of Euros and of people's time has been spent doing this" (P-URep7). This raises questions about how much end-users can rely on TFs to fill the 'gap' in their ability.

5.2.2 Reporting that supports the user. To file a report, users also must provide contact information (name and email address) according to Article 16(2) c DSA, "but I could imagine that that will prevent maybe some people who want to stay anonymous from reporting" (P-URep8). Requesting contact information may curb user reporting (P-URep3), where "anonymity" in the process is an important feature (P-URep5), yet it may be that " the lawyer of the perpetrator says [...] You have to come to court" (P-URep3). P-URep7 on the other hand stresses the need of "educating and helping users behave better" in the reporting process and to prevent problematic content in the first place.

5.2.3 The need for building tools. Many User Representatives have built their own tools for users to flag content, to reduce the burden (P-URep2, P-URep3, P-URep4, P-URep5, P-Auth1). Intended users vary though, spanning platform users (P-URep2, P-URep3, P-URep4), researchers and journalists, or "small and medium enterprises" (P-URep5), to authorities and regulators (P-Auth1). These aim to help to, for example, "send notifications or removal orders to hosting service providers" (P-Auth1). The need to build these user support tools often comes from the need for "evidence" (P-URep2, P-URep3), addressing "unawareness" about dealing with illegal content (P-Auth1), the necessity to create a "public record of content removal requests" (P-URep5), or because "users needed a way to help report harmful or illegal content" (P-URep4). Such tools may be intended to collect less information than SMPs' existing reporting processes (P-URep2, P-URep4), or "[help the user] report things [...] even if they aren't sure of the legal definitions" (P-URep9). To expect users to be familiar with the law is seen as "wildly ambitious" (P-URep5).

5.3 Differentiating illegal content from Terms of Service (ToS)

Users cannot be expected to have legal expertise to use DSA Reporting Mechanisms (RMs), and may struggle to differentiate illegal content from Terms of Service (ToS) violations. Perpetuating in differences in member state law that create complexity and define the scope together with European Union law of "illegal content" under Art 3(h) DSA. There can be some content where the (il)legality is not clear-cut. If a user successfully reports content as "illegal" the platform's liability changes since the user's notice creates "actual knowledge" according to Articles 6 and 16(3) DSA removing the platform's liability exception according to Article 8 DSA. Platforms are perceived to have a preference of ToS over legal flagging, according to P-Reg1, P-URep8, P-URep9, where the latter comments that "the platforms [...] more often rely on their Terms of Service because they're private entities, at the end of the day it's their terms of service " (P-URep9). Professionals are expected to use illegal content RMs, while regular users rely on ToS without understanding the differences.

As expressed by P-Reg3: "it's [...] so tough also for us lawyers to decide on this if it's against any criminal law", where there is 'gray content' as a tension in content moderation (P-URep4), as it is "always so hard to know the fine line between illegal and legal" (P-URep2). Differentiating illegal content from ToS violations, P-URep8 expects RMs to use "language that does not require from me to be a lawyer in order to understand what is not allowed." P-URep1 regards a website's DSA reporting path as intended for TFs, while users might use reporting under both – the ToS and the law – to flag as described in Section 2.

Reporting against ToS was seen as "more straight forward" (P-URep8) and "easier" (P-Reg1) than illegal content reporting. The "underlining of legal consequences" (P-URep8), and mention of "punishable under law" (P-URep9), during reporting could dissuade user reporting and lead to under-reporting of content as illegal.

Put together with user expectations and the design of the reporting process (Section 5.1) and user motivations, there is a paradox that by legally recognizing a user right to report content, the process has potentially become too complex and daunting for users who are *unexpectedly* offended and unfamiliar with how to report offensive content.

5.4 Content moderation and moderation decision

Non-transparent moderation complicates understanding of Reporting Mechanisms (RMs) for different users and regulators alike. However, benchmarking across platforms can enhance comparability and DSA compliance.

Regarding certainty and visibility in content moderation, the "decision lies with the companies" (P-Auth1), where there is uncertainty about the involvement of humans and AI in reviewing flags (P-URep1, P-URep2, P-URep8, P-URep9). This is characterized as "you report in hope rather than expectation" by P-URep7.

How SMPs deal with moderation is opaque: "What policy kicked in? How is this being moderated? If it's being even moderated?" (P-URep8). SMPs have different options like "Disable access to content or remove content entirely" (P-Reg1), as well as "geo-blocking" (P-Auth1), "demotion" or "labels" (P-URep6). Content moderation action is not guaranteed, where perhaps "the flag won't be acted upon...for various reasons, e.g., depending on the [...] culture. What's acceptable in Brazil may not be acceptable in Pakistan" (P-Reg5), constituting a further aspect of uncertainty in content moderation.

5.5 Design of reporting mechanisms (walkthrough exercise)

Through the walkthrough exercise, our participants noted that there are numerous overlaps between DSA and Terms of Service (ToS) categories, complicating illegal content reporting, even for content where the legality is unclear.

5.5.1 ToS/law reporting path. Terms of Service flagging can be "more detailed" (P-URep8) and create an overlap with the legal reporting categories. The difficulty for users' understanding between contractual and legal violations becomes evident in the numbers of the DSA transparency reports and the statement of reason database (P-URep5, P-URep8), that show lower numbers of illegal content report flags compared to ToS equivalents. The online statement of reason database captures all content moderation decisions under Article 17 the DSA, but "is not really useful at the moment" (P-Reg3) [8]. Kaushal et al. highlight a similar concern, that despite promising accountability, the implementation of DSA transparency mechanisms risks being " a facade of compliance without substantive insights" [33, p. 8], further stressing that defensibility for legal compliance is often prioritized over user-friendly design – we find similar concerns here among stakeholders tasked with supporting reporting.

It was regarded that users have to "probably click through several pathways" P-URep8 in order to report illegal content compared to ToS violations (P-URep9) (See also Figure 2). This represents a variation on the 'excessive number of legals' concern noted by Rossi et al. [48], where immediately a user would need to decide between a platform-internal report and going down a 'legals' route and entering into a legal process. Further, where Rossi et al. note 'difficult comparability' in assessing legal information, we find for flagging under the DSA, participants note this between legal categories which overlap with each other, as well as legal categories overlapping with ToS categories.

5.5.2 Differences in member state design. Notably, before encountering the design of Facebook's illegal content RM in the EU member states during the walkthrough, no participant was aware of different UIs for different national contexts (like France, Italy or Germany) in implementing the DSA, including regulator participants (see Fig. 16. in the repository *Selection of Wirframe Element for Differences in Illegal Content Reporting in France, Austria, and Germany on Facebook*).

5.6 Design of reporting mechanism

This section addresses the design of reporting mechanisms regarding their legal demands and effects on usability.

5.6.1 The act of reporting: motivation, emotion, and duty. Users might be in upsetting or "shocking" situations like "Oh my God, there is such thing [as this] on internet (P-URep4). So, you have to do something about it." E.g., for cases of "terrorist" or "far-right" content according to P-URep4. P-Reg1 also mentions the influence of encountering problematic content like "seeing hate speech" online, and the potentially deterring influence it might have on a user's flagging behavior. The personal involvement of the user was identified as a core motivation for reporting, as "It really depends if you're personally affected", exemplified in cases of "cybermobbing" (P-URep2). P-URep2 made a general observation, that the reporting of content on SMPs by users and reporting entities is "outsourcing the problem" of finding problematic content. Furthermore, navigation on SMPs should not be an essential step in reporting according to P-URep5, as the user "shouldn't have to go to someplace else in the website and [do] work".

5.6.2 Easy to access interface design. The "three dots" as a metaphor for the UI opening up the reporting button were mentioned regularly by participants, in relation to a need for interfaces to be easy to access (P-Reg1, P-URep1, P-URep9), though the subsequent reporting steps are seen as more "complicated" (P-URep1). Reporting access should be "next to every unit of content" that invites the user "do you want to flag this?" (P-URep5). Also, the explanations of ToS categories, and finding "information easily on the platform, preferably at one place", describe qualities of better reporting mechanisms in the opinion of P-URep8.

5.6.3 Reporting labyrinth – a form of dark patterns? There is an immediate tension in having the reporting mechanism on – and under the control of – a platform. Users were regarded as likely to find the notion of reporting features useful, but not easy to use upon entering the reporting process itself.

P-Auth1 stresses the need to report "swiftly and [in only a] few steps", yet the design may falter if a user has to "click through many things until you actually arrive to what you're looking for" (P-URep8). Such reporting mazes are described by P-URep8 as "dark patterns" regulated in Article 25 DSA, and mentioned in the context of, e.g., privacy policies, in [22, 25]. Dark patterns were also stressed in the context of RMs under the German NetzDG, as influencing other forms of reporting such as numbers in transparency reports [58]. Crucially, the length of the reporting process might dissuade users from reporting, akin to "I guess I'm not going to flag it after all, if I have to click through, you know, one to seven different screens. [...] I just wanted to, you know, put an asterisk on this horrible video" (P-URep5).

5.6.4 Support of different user needs. Participants suggested that platforms direct users to specific representative entities, akin to "Hey, did you know that there are these collective redress organizations that can take your case and here is the contact" (P-URep8). Users often seek support by user representatives because of the lack of response of online SMPs (P-URep4), or a need for the engagement of "real persons", "If it's a really frightening situation" (P-URep2).

5.6.5 Hiding content as the user-friendly default. The option to "mute" content or accounts according to "personal taste" (P-Auth1) is seen as "another option" according to several participants (P-URep6, P-Reg95, P-Auth1), e.g., P-URep6 explains "I don't wanna deal with it, so I just made you invisible" and P-URep7 qualifies hiding as "perfectly reasonable behavior". However, this kind of personal content management "does not help at all against this kind of content being spread online [...] and everyone else still can see it" (P-URep8), where "closing your eyes... it's not the solution" (P-URep6). This is akin to how flagging has been regarded in the literature up to now (e.g., [30]), driven in part by respect for freedom of expression; this then raises further tensions for the future, as to whether a user will be considering whether they are

singularly offended, or reporting and expressing offense 'on behalf of' a community of like-minded users.

5.7 Stakeholder cooperation

The DSA harmonizes content moderation across the EU. Communication is important for enforcement. Clarity is needed on what defines an 'easy to access' and 'user-friendly' Reporting Mechanism (RM). Tensions may arise with other user rights, such as GDPR. In order to make compliance work, "cooperation" and "coordination" are key (P-URep4, P-Reg5, P-Auth1). Regulators have to work together (e.g. in the board Article 61 DSA) and also with Trusted Flaggers (TFs) (P-Reg5, P-Auth1), otherwise there is a risk of fragmented handling (P-Reg5). For example, there may be rare "geo-blocking" in contrast to "removal orders" enforced in "the whole of Europe" (P-Auth1), and global tensions for example with the US 1st Amendment [1]; the latter is already notable in the different ways flagging is discussed here compared to prior literature. The need for cooperation extends to reporting interfaces, for instance in how to design reporting mechanisms under the Terms of Service (ToS), which according to P-URep8 is also not "subject to any reporting obligations".

6 Discussion

Here we reflect on our findings, and the implications for users and the support provided by other stakeholders. Our findings emphasize the urgent need for platforms to balance compliance with the DSA and intuitive design. We found that it is necessary to define channels for support and feedback that connect users with Professional Reporting entities, in order to reduce reporting complexity and burden (where even our practitioner and regulator participants remarked on several pitfalls). These needs are especially critical given how reporting mechanisms are situated on the very same platforms they oversee, where complex reporting processes compete with Terms of Service and content-hiding alternatives as a way for the non-expert user to act on offensive material.

6.1 Users and the law: empowerment and burden

There is the immediate contradiction that by recognizing content reporting as a user right under Article 16 DSA, signaling offense has potentially become a more daunting and unclear process (generally and from the user perspective) (Section 5.6). For instance, does the user need to go to court at some point in the future? This is compounded by users potentially being emotionally-charged at the time they experience offensive content on a platform. This tension is stressed by the current platform shifts, exemplified by Meta, demanding more user interaction in content moderation through less institutionalized content moderation, and more reliance on 'community note' systems [28, 32, 49, 55, 56]. This could lead to two very different digital realities in the US and Europe, for instance leaving content online in the US while having to remove it in the member states.

Some of our participants – both user representatives and regulators – believed that platform users should liaise with user representative or NGO organizations, to make a legal report of offensive content (Sections 5.1 and 5.2). Yet, how users reporting would become aware of these organizations is not clear, as social media platforms do not signal or redirect to specialized organizations (e.g., for copyright or personal harassment). Additionally, such organizations could face resourcing demands and financial shortcomings that have no provision within the DSA, where financial incentives from SMPs and dependencies with these can impact the reporting of content [53]; given that reporting may involve making decisions on divisive topics (e.g. transgender content [54]), power dynamics play a role [16]. It remains to be seen whether easy-to-use reporting interfaces will compete with existing user options to dismiss content from their feeds (Section 5.3). In a sense, the DSA appears to assume a direct correlation of the legally-determined "severity" of offensive content with the motivation and energy it would activate in the offended user.

The DSA is in its early stages, but how to keep users motivated once they report content is unclear, especially if they are expected to make a distinction between offensive and illegal content. One focal point is feedback – the UK Action Fraud service allowed reporting of crimes, but focused on passing information to police forces rather than victim support [9]. Making a comparison to phishing reporting and differentiating between a malicious or benign email [35], users who only receive feedback for reporting a malicious email may begin to doubt whether they are spotting bad content properly and reduce their efforts. We recommend that future effort examines how to support a user who is brought into the (legal) reporting process, and avoid erosion of their motivation by burdensome or uncertain processes.

One direction for future engagement with users would be to fabricate reporting processes within an experimental setting, based on real-world implementations, and lead users through the process (similar to how we distilled real-world implementations here). This could identify 'pain-points' where users encounter problems (given the expectation for legal knowledge and comprehension), where support could be targeted (either in terms of simplifying the process in a way that removes an expectation of legal knowledge, or where expert flagger assistance can best be positioned to compensate for the difficulties users have (to reduce the need for Capability [40] to complete the reporting task). Our approach of using a wireframe of the reporting process, and a legal design walkthrough, can be repurposed as a user journey, to track where any pain-points occur within the reporting process (for instance, identifying issues with comprehension or unclear choices). Given the current positioning of reporting infrastructure on the platforms themselves, there are also issues of Motivation relative to the perceived effort of engaging with the reporting process; specifically, a user may weigh up their sense of offense with a piece of content against their perception of the immediately available (and usable) reporting mechanisms.

A user may engage directly and fully with the legal reporting process, or *satisfice* [52] and accept the first available option that seems satisfactory for getting the content either out of their feed or into a system where it will be scrutinized by the platform (rather than tirelessly seeking out the optimal solution, in this case, reporting within the legal framework of the DSA); they might do this by choosing to 'hide' content from their feed, or reporting it against a platform's own Terms of Service. The aforementioned 'pain-points'

are then not only about usability issues with the interface and decisions, but also the 'usability' of the legal process embedded within the interface and its impact on user motivation [40].

Given the current positioning of reporting infrastructure on the platforms themselves, a potential future user study would be useful to understand how users would weigh up content and their perception of available and usable reporting mechanisms (notably, to hide content, report against Terms of Service, or engage directly with legal reporting).

There are user concerns the DSA will not solve – malicious use of flagging features becomes an increased cost for all stakeholders. Flagging of divisive content (e.g., views on immigration) seem to have been put to one side, deferring instead to the character of the law and there being offensive content that is somewhat "universally" agreed as being illegal. Crawford & Gillespie [10] note how a flag arbitrates "the negotiation around contentious public issues" (Section 5.7). However, the findings of Eyert et al. [16] raise the possibility that societal influence and power dynamics may be what ultimately resolves any such deadlock.

6.2 User action in the presence of platform opacity

Even with the DSA in place, the content moderation processes within Social Media Platforms (SMPs) remain opaque to external actors (Section 5.4). At present, the SMPs alone have the capacity to see what is happening within their own online environments. This leaves questions as to what "good enough" effort by SMP looks like, as the DSA arguably results in monitoring for signs of non-compliance, as opposed to monitoring for proof of utmost compliance (i.e., removal of everything that should be removed) (Section 5.5).

We can articulate our findings around DSA reporting mechanisms, in terms of behavior adoption factors via the COM-B behavior adoption model [40]. To consider reporting entities and Trusted Flaggers (TFs) as (professional) users themselves, our participants perceived a lack of transparency despite the additional controls (impacting capability) (Section 5.2), and expected the same among general users. Participants' legal expertise helped them to understand the scope of specific categories, but did not remove the need to navigate 'gray areas' of content, or decide if content was clearly legal or illegal. Legal aspects were seen to dampen user motivation despite being positioned to empower users. The opportunity to report was seen as being hindered by the placement of reporting mechanisms away from content, but also in part because more convenient options to 'hide' content supplanted reporting features in platform interfaces.

As TF roles expand, another recommendation would be to examine the issue where some participants thought users would consult Professional Reporters (PRs), or that users and flaggers would act independent of one another, or that PR reports would be triangulated with or would verify user reports, once seen together by platform moderators. As the DSA is being established in practice, there is a need to explore how communication and visibility of flagging pathways can be best signaled to users, or otherwise how the interplay between users and TFs can be coordinated – this would offset the perception of a few of our participants, that content moderation is being outsourced to users. Building on the proposal for user studies in Section 6.1, a user study could include a range of conditions, each with different signposting for reporting assistance within a fabricated interface. These could include a user being relied on to interpret the process themselves – as we found is presently the case, unless they are already aware of flagger/reporter entities they can consult – or designed to include a range of signposting options, acting to cue the user to contact and involve a flagger entity, delegate the process to a flagger, etc.

One future direction is to synthesize reporting processes based on real-world implementations, and lead users through the process in a user study. Our mapping approach can be used to track and identify 'pain-points' where a user encounters problems (especially given the expectation for legal knowledge and comprehension). This would be where support could be targeted, either in terms of simplifying the process more toward a layperson, or where expert the assistance of trusted flaggers or other reporting entities can best intervene to compensate for the difficulties users have.

7 Conclusion

In conclusion the DSA is promising EU-harmonization of the regulation of reporting mechanisms (RMs) of Social Media Platforms (SMPs) while also creating new user rights and PR-roles like Trusted Flaggers. Our findings show the entwined and complex relationship between the different actors of the reporting process. RM are a tool for the user but not always designed in a way that supports the user who therefore, has to seek help at user representative organizations to guide them through the legal complexity that filing a report under the DSA can bring. How to better design RMs that adhere to imprecise DSA obligations, like being 'user friendly', while also following user needs is even for experts in the field a challenging task. Further work is needed in the field of legal-design addressing RMs and their regulation. Issues like transparency, UI complexity, and user information, can foster alignment of SMP with the DSA, aiming for a safer online environment for a stakeholder-multitude.

Acknowledgments

This research was supported by the AI Futures Lab on Rights and Justice at TU Delft. We thank all participants for sharing their expertise and perspectives. We are also grateful for the feedback provided by colleagues throughout the development of this study.

References

[1] National Archives. 1791. The Constitution of the United States. , Amendment I pages. www.archives.gov/founding-docs/bill-of-rights

[2] Kelly A. Aschbrenner, Emily R. Haines, Gina R. Kruse, Ayotola O. Olugbenga, Annette N. Thomas, Tanveer Khan, Stephanie Martinez, Karen M. Emmons, and Stephen J. Bartels. 2024. Applying cognitive walkthrough methodology to improve the usability of an equity-focused implementation strategy. *Implementation Science Communications* 5, 1 (Sept. 2024), 95. https://doi.org/10.1186/s43058-024-00630-8

[3] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.

[4] Dylan Butts. 2025. Meta says it fixed 'error' after Instagram users report a flood of graphic and violent content. https://www.cnbc.com/2025/02/27/meta-apologizes-after-instagram-users-see-graphic-and-violent-content.html Section: Technology.

[5] California State Assembly. 2022. Assembly Bill No. 587. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB587

[6] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298.

[7] Julie E. Cohen. 2019. The Regulatory State in the Information Age. In *Between Truth and Power: The Legal Constructions of Informational Capitalism*, Julie E. Cohen (Ed.). Oxford University Press, Oxford, UK, 0. https://doi.org/10.1093/oso/9780190246693.003.0007

[8] European Commission. 2023. Statements of Reasons - DSA Transparency Database. https://transparency.dsa.ec.europa.eu/statement

[9] Sara Giro Correia. 2019. Responding to victimisation in a digital world: a case study of fraud and computer misuse reported in Wales. *Crime Science* 8, 1 (2019), 1–12.

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

[11] Micheline Elias and Anastasia Bezerianos. 2012. Annotating BI visualization dashboards: needs & challenges. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1641–1650. https://doi.org/10.1145/2207676.2208288

[12] European Parliament and Council. 2016. General Data Protection Regulation. http://data.europa.eu/eli/reg/2016/679/oj/eng Legislative Body: EP, CONSIL.

[13] European Parliament and Council. 2018. Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) (codified version) (Text with EEA relevance)Text with EEA relevance. http://data.europa.eu/eli/dir/2010/13/2018-12-18/eng Legislative Body: OP_DATPRO.

[14] European Parliament and Council. 2021. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (Text with EEA relevance). http://data.europa.eu/eli/reg/2021/784/oj/eng Legislative Body: EP, CONSIL.

[15] European Parliament and Council. 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). http://data.europa.eu/eli/reg/2022/2065/oj/eng Legislative Body: EP, CONSIL.

[16] Florian Eyert and Paola Lopez. 2023. Rethinking transparency as a communicative constellation. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, Chicago, IL, USA, 444–454.

[17] Federal Ministry of Justice, Germany. 2017. German Network Enforcement Act (NetzDG). https://www.bundesjustizamt.de/SharedDocs/Downloads/DE/NetzDG/Leitlinien_Geldbussen_en.pdf?__blob=publicationFile&v=3

[18] Pat Foran CTV News Toronto Consumer Alert Videojournalist Follow |Contact. 2024. Ontario man loses $12K to deepfake scam involving Prime Minister Justin Trudeau. https://toronto.ctvnews.ca/ontario-man-loses-12k-to-deepfake-scam-involving-prime-minister-justin-trudeau-1.6823827 Section: Toronto.

[19] French Republic. 2004. Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique.

[20] Juan Felipe Gomez, Caio Machado, Lucas Monteiro Paes, and Flavio Calmon. 2024. Algorithmic Arbitrariness in Content Moderation. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, Rio de Janeiro, Brazil, 2234–2253.

[21] Colin M. Gray, Ritika Gairola, Nayah Boucaud, Maliha Hashmi, Shruthi Sai Chivukula, Ambika R Menon, and Ja-Nae Duane. 2024. Legal Trouble?: UX Practitioners' Engagement with Law and Regulation. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference (DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 106–110. https://doi.org/10.1145/3656156.3663698

[22] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 534:1–534:14. https://doi.org/10.1145/3173574.3174108 event-place: Montreal QC, Canada.

[23] Colin M. Gray, Thomas Mildner, and Ritika Gairola. 2025. Getting Trapped in Amazon's "Iliad Flow": A Foundation for the Temporal Analysis of Dark Patterns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 225, 10 pages. https://doi.org/10.1145/3706598.3713828

[24] Colin M Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery, Yokohama, Japan, 1–18.

[25] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner. 2024. An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3613904.3642436

[26] James Grimmelmann. 2017. The Virtues of Moderation. https://doi.org/10.31228/osf.io/qwxf5

[27] Hana Habib and Lorrie Faith Cranor. 2022. Evaluating the Usability of Privacy Choice Mechanisms. In *Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, USA, 273–289. https://www.usenix.org/conference/soups2022/presentation/habib

[28] Pieter Haek. 2025. Pressure mounts on EU to use legal weapons against Musk's interference. https://www.politico.eu/article/elon-musk-interference-alice-weidel-x-livestream-afd-germany-eu-digital-services-act/

[29] Martin Husovec. 2018. The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which Is Superior? And Why? *JIPITEC* 9, 1 (Jan. 2018), 2–20. https://www.jipitec.eu/issues/jipitec-9-1-2018/4683

[30] Shagun Jhaver, Alice Qian Zhang, Quan Ze Chen, Nikhila Natarajan, Ruotong Wang, and Amy X Zhang. 2023. Personalizing content moderation on social media: User perspectives on moderation choices, interface design, and labor. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33.

[31] Shagun Jhaver and Amy X Zhang. 2023. Do Users Want Platform Moderation or Individual Control? Examining the Role of Third-Person Effects and Free Speech Support in Shaping Moderation Preferences. *New Media & Society* 0 (2023), 1–21. https://doi.org/10.1177/14614448231217993 Published online ahead of print.

[32] Joel Kaplan. 2025. More Speech and Fewer Mistakes. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

[33] Rishabh Kaushal, Jacob Van De Kerkhof, Catalina Goanta, Gerasimos Spanakis, and Adriana Iamnitchi. 2024. Automated Transparency: A Legal and Empirical Analysis of the Digital Services Act Transparency Database. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, Rio de Janeiro, Brazil, 1121–1132.

[34] Yubo Kou and Xinning Gui. 2021. Flag and flaggability in automated moderation: The case of reporting toxic behavior in an online game community. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery, Yokohama, Japan, 1–12.

[35] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, IEEE, San Francisco, CA, USA, 842–859.

[36] Juniper L Lovato, Antoine Allard, Randall Harp, Jeremiah Onaolapo, and Laurent Hébert-Dufresne. 2022. Limits of individual consent and models of distributed consent in online social networks. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, Seoul, Republic of Korea, 2251–2262.

[37] Sapna Maheshwari. 2023. TikTok Pushes Back Against Claims It Fuels Antisemitism. *The New York Times*, Business section. https://www.nytimes.com/2023/11/02/business/tiktok-antisemitism-claims-israel-palestinians.html Accessed July 5, 2024.

[38] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–32.

[39] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery, Yokohama, Japan, 1–18.

[40] Susan Michie, Maartje M Van Stralen, and Robert West. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science* 6 (2011), 1–12.

[41] Dan Milmo. 2023. Social Media Platforms to Face £18m Fines for Publishing Animal Cruelty Content. *The Guardian*, Technology section. https://www.theguardian.com/technology/2023/sep/07/social-media-platforms-to-face-18m-fines-for-publishing-animal-cruelty-content Accessed July 5, 2024.

[42] Ministry of Electronics and Information Technology, Government of India. 2023. Intermediary Guidelines and Digital Media Ethics Code India. https://www.meity.gov.in/writereaddata/files/Information%20Technology%20%28Intermediary%20Guidelines%20and%20Digital%20Media%20Ethics%20Code%29%20Rules%2C%202021%20%28updated%2006.04.2023%29-.pdf

[43] Patricia L. Moravec, Antino Kim, and Alan R. Dennis. 2020. Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media. *Info. Sys. Research* 31, 3 (Aug. 2020), 987–1006. https://doi.org/10.1287/isre.2020.0927

[44] Teresa K Naab, Anja Kalch, and Tino GK Meitz. 2018. Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society* 20, 2 (2018), 777–795.

[45] Janni Nielsen, Torkil Clemmensen, and Carsten Yssing. 2002. Getting access to what goes on in people's heads? reflections on the think-aloud technique. In *Proceedings of the second Nordic conference on Human-computer interaction*

(NordiCHI '02). Association for Computing Machinery, New York, NY, USA, 101–110. https://doi.org/10.1145/572020.572033
[46] House of Commons. 2023. Online Safety Act. https://www.legislation.gov.uk/ukpga/2023/50 Publisher: Statute Law Database.
[47] Republic of Austria. 2023. § 1 Kommunikationsplattformen-Gesetz (KoPl-G), BGBl. I Nr. 151/2020, version of 31 August 2023. https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20011415&FassungVom=2023-08-31&Paragraf=1 Consolidated version of § 1 KoPl-G as of 31 August 2023.
[48] Arianna Rossi, Rossana Ducato, Helena Haapio, and Stefania Passera. 2019. When Design Met Law: Design Patterns for Information Transparency. *Droit de la consommation* 1 (2019), 12–29. Issue No. 1, Université Catholique de Louvain, Centre de Droit de la Consommation.
[49] Adam Satariano and Jeanna Smialek. 2025. E.U. Expands Investigation of Musk's X Over Illicit Content. https://www.nytimes.com/2025/01/17/business/eu-investigation-x-musk.html
[50] Andrew Sears. 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human–Computer Interaction* 9, 3 (Sept. 1997), 213–234. https://doi.org/10.1207/s15327590ijhc0903_2
[51] Yunhee Shim and Shagun Jhaver. 2024. Incorporating Procedural Fairness in Flag Submissions on Social Media Platforms. https://doi.org/10.48550/arXiv.2409.08498 arXiv:2409.08498.
[52] Herbert A Simon. 1956. Rational choice and the structure of the environment. *Psychological review* 63, 2 (1956), 129.
[53] Petros Terzis, Michael Veale, and Noëlle Gaumann. 2024. Law and the Emerging Political Economy of Algorithmic Audits. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, Rio de Janeiro, Brazil, 1255–1267.
[54] Hibby Thach, Samuel Mayworm, Michaelanne Thomas, and Oliver L. Haimson. 2024. Trans-centered moderation: Trans technology creators and centering transness in platform and community governance: Trans-centered moderation. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Rio de Janeiro Brazil, 326–336. https://doi.org/10.1145/3630106.3658909
[55] The Guardian Editorial Board. 2025. The Guardian view on content moderation: Meta did far too little. Now it wants to do even less. *The Guardian*, Opinion section. https://www.theguardian.com/commentisfree/2025/jan/08/the-guardian-view-on-content-moderation-meta-did-far-too-little-now-it-wants-to-do-even-less Accessed January 2025.
[56] Stuart A. Thompson and Kate Conger. 2025. Meet the Next Fact-Checker, Debunker and Moderator: You. https://www.nytimes.com/2025/01/07/technology/meta-facebook-content-moderation.html
[57] United States Congress. 1996. Communications Decency Act. https://www.justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996
[58] Ben Wagner, Krisztina Rozgonyi, Marie-Therese Sekwenz, Jennifer Cobbe, and Jatinder Singh. 2020. Regulating transparency? Facebook, Twitter and the German Network Enforcement Act. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 261–271. https://doi.org/10.1145/3351095.3372856
[59] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Tara Matthews, Sarah Meiklejohn, Franziska Roesner, Renee Shelby, Kurt Thomas, and Rebecca Umbach. 2024. Understanding Help-Seeking and Help-Giving on Social Media for Image-Based Sexual Abuse. *arXiv preprint arXiv:2406.12161* 0 (2024), 18 pages. arXiv:2406.12161 https://arxiv.org/abs/2406.12161 Preprint, accessed July 5, 2024.
[60] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Franziska Roesner, and Kurt Thomas. 2023. "There's so much responsibility on users right now:" Expert Advice for Staying Safer From Hate and Harassment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Hamburg, Germany, 1–17.
[61] Karen Yeung, Martin Lodge, Karen Yeung, and Martin Lodge (Eds.). 2019. *Algorithmic Regulation*. Oxford University Press, Oxford, New York.
[62] Alice Qian Zhang, Kaitlin Montague, and Shagun Jhaver. 2023. Cleaning Up the Streets: Understanding Motivations, Mental Models, and Concerns of Users Flagging Social Media Posts. *arXiv preprint arXiv:2309.06688* 0 (2023), 1–18. arXiv:2309.06688 https://arxiv.org/abs/2309.06688 Preprint, accessed July 5, 2024.

A Appendix

A.1 Semi-Structured Interview Guidelines

A.1.1 Introduction:

(1) Introduce the purpose of the interview and the focus on flagging mechanisms.
(2) Introduce the flagging mechanisms [*Recital 50, Article 16 DSA*] (Terms of Service (ToS) [*Article 14 DSA*] and legal reasons [*Article 3(h) DSA*]) used by platforms like Facebook and TikTok etc. [We asked this question to regulators].
(3) Can you explain your position and your work briefly?
(4) How does your work relate to users? And to which users does it relate?
(5) Which user groups or types do you represent? [Not asked for regulators].

A.1.2 RQ1: How would user circumstances/situations align with the content-flagging process?

(1) The DSA talks about "notice action mechanisms" [*Article 16 DSA*], the process of reporting or flagging (we will use these terms interchanging) describes the process where a UI is utilized by a user to inform a platform about perceived problematic content. How does the flagging process start, and how does it end? When is reporting 'over' or closed? [Not asked for regulators].
(2) What is a 'normal' flagging process for a user?
(3) Describe situations and circumstances of users that motivates / incentivizes users to report? Can you describe situations that would motivate a user to report? (Don't walk on the "bad street" of platform content. You might be associated with problematic/illegal content).
(4) Under which situations do users, in your opinion, flag content?
(5) How do these situations/ circumstances differentiate from/ between user groups normal users, flagging entities, Trusted Flaggers [*Article 22 DSA*])? [We asked this question to regulators].
(6) How many users have to flag a piece of content before consensus is reached to enforce platforms rules (Terms of Service and legal reasons [*Articles 14, 3(h) and 16 DSA*])? When is user representation solidified?
(7) How do these situations/ circumstances differentiate from your own experience when flagging content? [Not asked for regulators].
(8) Are you aware of other alternatives for users than flagging to deal with problematic/upsetting content?
(9) How do professional flagging entities become aware of problematic content [*Articles 9-10 DSA, Article 22 DSA*]?
(10) Are platforms actively approaching flagging entities?
(11) Do users actively approach flagging entities?
(12) What do you expect from Trusted Flaggers [*Recital 61, Article 22 DSA*]?
(13) What do you expect flagging entities and Trusted Flaggers not to catch [*Article 22 DSA*]? [Not asked for regulators].
(14) Introduce the flagging mechanisms (Terms of Service and legal reasons) used by platforms like Facebook and TikTok etc. [Not asked for regulators] [*Article 16(1) DSA*].

(15) How does flagging under the Terms of Service and the law differentiate [*Articles 16(2) and (3) DSA*]?
(16) How does the flagging process start, and how does it end [*Articles 16, 20, 21 DSA*]? [We asked this question to regulators].

A.1.3 RQ2: What are the expectations on users to flag content?

(1) What are the necessities for a user to flag content [*Recitals 52-53, 58 DSA*]? What has to be clear to the user [The second part of the question was not asked for regulators]?
(2) Do you expect that flagging for users can have negative consequences [*Recital 58 DSA*]?
(3) Are there differences in how to flag under the Terms of Service and the law [*Articles 3(h) and 14 DSA*]?
(4) Are users aware of these differences?
(5) Are users able [for regulators "equipped"] to make informed flagging decisions under either alternative (Terms of Service and the law [*Articles 3(h) and 14 DSA]*)?
(6) How are professional flagging entities navigating flagging under the Terms of Service and the law?
(7) Are there different expectations for Trusted Flaggers [*Article 22 DSA*] under the DSA?

A.1.4 Prompts for user flagging challenges (following heuristic walkthrough).

(1) Online platforms have 'notice action mechanisms' or flagging mechanisms that provide options to report content through the Terms of Service and legal reasons [*Recital 50 DSA*]. How does flagging under the Terms of Service and the law differentiate through those mechanisms? And are they differently designed? [*Recital 53 DSA*]
(2) Can you comment on the high level view of flagging mechanisms illustrated for Facebook and TikTok? How do they differentiate?
(3) What are the essential requirements for a user to flag content? What information must be clear to the user before flagging content? [*Recital 53 and Article 16 DSA*].
(4) Are items flagged as 'illegal content' treated differently than Terms and Conditions flags [*Recital 61, Articles 3(h) and 14 DSA*]?
(5) Can you comment on the design of Terms of Service and illegal content flagging of the two platforms?
(6) Are there potential negative consequences for users who flag content? If so, what are they?
(7) Can you articulate your thoughts on the different designs for flagging illegal content for the four Member States for Facebook (France, Austria, Germany, and other) and related regulations in those countries [*Recitals 52-53, Article 3(h) DSA*]?
(8) On point five (referring to the Miro board) you see hiding content or don't show me this again, can you comment on this alternative to flagging under the Terms and Conditions for the user and flagging because of a legal reason [*Recitals 50, 52-53, Article 3(h), 14, 16 DSA*]?

A.1.5 RQ3: How can content-flagging interfaces be designed to be usable and satisfy legal requirements?

(1) The DSA demands "notice action mechanisms" to be "easy to access" and "user friendly" [*Recitals 50, 52-53, 61, and Article 16 DSA*]. In your opinion, what restrictions/expectations do you identify to be compliant with the law?
(2) What requirements are necessary from a user's point of view, in your opinion, to design flagging interfaces in a DSA-compliant way [*Recitals 50, 52-53, 61, and Article 16 DSA*]?
(3) What tensions don you see in user flagging that users have to overcome to flag content and how should flagging be structured in your opinion according to user needs? [Not asked for regulators]
(4) How can flagging mechanisms support different user groups and overcome stakeholder tension (normal users, flagging entities, Trusted Flaggers [*Article 22 DSA*])?
(5) How, in your opinion, do flagging mechanisms have to address flagging under the Terms of Service and the law [*Recitals 50, 52-53, 61, and Article 16 DSA*]?

A.2 Platform Descriptors

Table 2: Wireframe Platform Descriptor Summary

Platform	ToS Cat.	Legal Cat.	ToS Depth	Legal Depth
TikTok	15	14	2/4	2/3
Facebook	14	4	2/4	6/23

A.3 Codebook and Thematic Findings in Themes and Subthemes

Table 3: Codebook: Code Families and Associated Codes

Code Family	Associated Codes
Participants	Expert's position, years of experience, other regulation, GDPR, CDP, NetzDG, Loi nº 2004-575
Stakeholders	User representation, Own tool, NGO representation through the "Right to lodge a complaint," Concerns about user privacy, Whistleblower protection for users, User as a witness, Unreportable content, Insufficient user reporting bootstrapping, User burden, Dark pattern, Offended user, Missing in- formation, Revictimization through reporting, Stakeholder concern about intentional online platform non-compliance
DSA	User right, Information about user rights, Cooperation, Platform obligation, Legal uncertainty
Content	Awareness of problematic content, Awareness of user content screenshots, Comments, Public authority information, Illegal content, Different levels of ease to identify 'manifestly illegal content,' Overlaps in reporting, Fragmentation of definition of illegal content, Terms of service, Contractual freedom
Reporting	Reporting process, Complexity of the reporting process, User follow-up, Information about the reporting process, Information about content moderation decision, Uncertainty in reporting, Uncertainty about the handling of flags, Different treatment of illegal content or ToS flags, Uncertainty about public authorities in the process, User motivation, Social role of reporting, User perception about the impact of own reporting, Professional reporting, Prioritization of trusted flaggers, Accreditation of trusted flaggers, Trust systems in user reporting, Enforcing cross-member state content reporting, Malicious reporting
Transparency	Transparency, Transparency report, Statement of Reason Database, Transparency of the reporting process
Design	Design of reporting mechanism, Symmetry/Relation be- tween ToS and Law reporting, Artificial Intelligence (AI) tools for reporting, Labeling, Design of file uploading options, Single click solution, Progress bar, User friendly
Content Moderation	Moderation decision, Rapid response system, Recommendation system, Unbalanced decision consensus
Hiding Content	Hide/don't show me this again, Information about the effect of hiding

Table 4: Thematic Findings in Themes and Subthemes

No.	Theme	Subtheme	Summary	Related Codes
1	Content reporting – a multi-actor process	Connection to users	Reporting is dependent on the user's motivation. Participants represent users in different positions against various threats of online content by offering legal expertise.	User motivation, user representation, NGO representation through the "right to lodge a complaint", insufficient user reporting bootstrapping, offended user, user burden, professional reporting, design of reporting mechanism, user friendly
2	Content reporting – a multi-actor process	Legal representation for users	DSA reporting empowers the user. Users are not expected to have legal knowledge. Legal understanding is needed in DSA reporting. Regulatory bodies offer enforcement and reporting support to users.	Insufficient user reporting bootstrapping, user burden, illegal content, user right, user follow-up, user representation, NGO representation to lodge a complaint, complexity in the reporting process
3	Content reporting	Upholding representation of users	The new DSA Trusted Flaggers specialize in educating and empowering user representation.	User representation, illegal content, professional representation, NGO, reporting process, complexity of the reporting process
4	Content reporting	Reporting that supports the user	User representatives emphasize creating tools that support user reporting and representation.	Information about the reporting process, user right, information about user rights, own tool, user representation, NGO representation
5	Design of RMs	The need for building tools	User representatives emphasize creating tools that support user reporting and representation.	Own tool, user representation, NGO representation
6	Design of RMs	Accessibility of reporting mechanisms	Reporting mechanisms need to be easy to find and accessible to all users.	User accessibility, simple UI design, user-friendly reporting mechanisms
7	Content moderation	Moderation decision	Users often feel left in the dark about the results of their reporting. Transparency is critical. Reports are sometimes handled inconsistently, leading to perceptions of bias.	User follow-up, transparency, clarity in reporting processes, fairness, report consistency, bias mitigation.
8	Walkthrough exercise	ToS/law reporting path	Platforms choose different layouts and paths for how users can report. These can be complex for users to navigate.	Overlaps in reporting, illegal content, terms of service, user feedback, iterative improvements, user-centric design, symmetry/relation between ToS and Law reporting.
9	Walkthrough exercise	Differences in member state design	Reporting mechanisms show different designs for member states. This makes reporting across national jurisdictions fragmented and complicated.	Insufficient user reporting bootstrapping, User burden, Stakeholder concern about intentional online platform non-compliance
10	Design of reporting mechanism	The act of reporting: motivation, emotion, and duty	Users face shocking content, which influences their motivation and emotional burden while reporting.	User burden, user motivation, uncertainty in reporting
11	Design of reporting mechanism	Easy to access interface design	Reporting mechanisms need to be easy to find and accessible to all users.	User accessibility, simple UI design, user-friendly reporting mechanisms
12	Design of reporting mechanism	Reporting labyrinth – a form of dark patterns?	Reporting processes should empower users without overly burdening them.	User empowerment, reporting burden, process simplicity
13	Design of reporting mechanism	Support of different user needs	External actors, such as NGOs, play a crucial role in supporting effective reporting.	NGO oversight, reporting audits, external expertise
14	Design of reporting mechanism	Hiding content as the user-friendly default	Hiding content can be seen as user control that burdens the user less	Hide/don't show me this again, Information about the effect of hiding
15	Stakeholder cooperation	Collaboration between users	Collaboration between users, platforms, and regulators is key to effective enforcement.	Collaboration, enforcement, stakeholder involvement