Making Reporting Approachable: User-Friendly Reporting Mechanisms for Platform Flagging

MARIE-THERESE SEKWENZ and SIMON PARKIN, Delft University of Technology, The Netherlands

ACM Reference Format:
Marie-Therese Sekwenz and Simon Parkin. 2026. Making Reporting Approachable: User-Friendly Reporting Mechanisms for Platform Flagging. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

This paper investigates the intersection of the Digital Services Act's (DSA) design requirements of reporting mechanisms for users, and the General Data Protection Regulation's (GDPR) requirements for data subjects data protection and right to deletion [1, 2]. A user's rights granted under the GDPR are materialized through the reporting mechanisms designed and operated by online platforms. When users flag content to a platform, however, they must use a complex reporting system on the platform itself. Reporting pathways are divided, between the terms of service of a platform or the law. These systems present complex legal options to users, as shown by Sekwenz et al. [4]: the design requirements of Article 16 DSA demand that the reporting of *illegal content* (see Art. 3(h) DSA) must be 'user friendly' and 'easy to access.' Besides, the DSA demands that such notices from users fulfill certain criteria which are potentially onerous, such as the user's name, a *bona fide* statement, the rationale behind the potential violation, or an email address (See Art. 16(2)(a–d) DSA).

Therefore casual or non-expert users are reporting within complex structures and legal categorizations, that are likely to be unfamiliar to them. Users are then reporting under uncertainty on arguably adversarial environments, where the platform is hosting – and designing – the reporting mechanism that keeps itself in check.

In this text we ask: **How might we envision positive and negative consequences of different futures, and translate them into design provocations that challenge current assumptions?**

Here, we provide different provoking scenarios relevant for understanding users and platform situations, within the design of reporting mechanisms.

2 Design Provocation: Reporting Scenario

2.1 Adversarial Reporting Environments

Users must use the platform's own reporting interface, which would seem user-friendly as the reporting options are then immediately accessible on the platform. This poses several tensions relevant in a privacy-preserving and user-friendly

Authors' Contact Information: Marie-Therese Sekwenz, m.t.sekwenz@tudelft.nl; Simon Parkin, s.e.parkin@tudelft.nl, Delft University of Technology, Delft, The Netherlands.

Manuscript submitted to ACM

environment. First, platforms have control over the front-end design of reporting mechanisms, including how legal categories are structured. This can include pathways with the potential to nudge users to report under Terms and Conditions categories [5]; this essentially reduces the responsibilities of the platform, relating to Art. 6 in conjunction with Art. 16(3) DSA. This means that platforms have control over the infrastructure of flagging, which in turn influences the externally-reported transparency mechanisms (e.g. Art 17, 15,24, 42, 34-35, 37 DSA). This poses the risk of obscuring content management practices.

As an illustrative scenario, it may be that a user discovers an image of themselves that appears to be a so-called "undressing" output or sexually explicit synthetic transformation, generated via an AI tool (e.g., a model embedded in or connected to a platform environment). The harm is privacy-centric: it concerns sexual integrity, consent, and the non-consensual creation or circulation of intimate imagery. Yet, to report it under Art. 16 DSA, the user may be required to provide: a description of why the content is illegal, URLs/screenshots, and identifying information (Art. 16(2)(a–d) DSA). In practice, this can force the user into an evidencing trap: to prove the violation they must re-handle, re-upload, or re-expose intimate material and personally identifying information. This challenge could be compounded if the AI tool has links to the platform where the content appears, or if the platform tolerates AI-generated content to encourage engagement.

2.2 Design Provocation: Reporting Ombudsman

An alternative design is remove the reporting mechanism from the platform's control, and situate it at a neutral third party, akin to an ombudsman. This third-party organization might be run by civil society or government actors, where these actors may already be operating their own reporting tools to help users report illegal content online [4]. By disentangling reporting from platforms, different forms of support for users can be offered which move the reporting to expert bodies and reduce the burden on the user to navigate the law, while also potentially being in distress or upset when exposed to privacy-violating or illegal content.

For instance, to submit a notice, if a user wishes to report a 'doxxing' incident, they may need to provide their name, email, and a detailed rationale (Art. 16(2)(a–d) DSA). In other words, the person reporting a doxxing incident may be required to disclose additional personal information into the very system that is already processing their exposure.

This scenario, however, creates requirements to have stable access to the content, to be able to involve a supporting organization in processing the report. This, in turn, may require development of capabilities, for making (potentially sensitive) content made accessible to the supporting organization, for them to be able to process a report with – or on behalf of – a user.

2.3 Design Provocation: Auditing That is Easy

This provocation imagines a reporting ecosystem in which DSA oversight becomes *audit-ready by default*: the interaction between user, platform, and regulator produces structured, comparable, and reproducible evidence at minimal cost to the user (e.g. automatic logs for illegal content notices to public authorities). In this scenario, the platform's notice-and-action interface is not merely "available" but *instrumented for accountability*. Every notice submitted under Art. 16 DSA automatically generates a stable case identifier, a machine-readable statement of the notice elements required by Art. 16(2)(a–d) DSA, a time-stamped decision trace, and a linkable record that can be used across adjacent procedural stages (e.g., Art. 17 statements of reasons, transparency reporting Art. 15,24,42, Art. 20 complaint handling, and where applicable Art. 21 out-of-court dispute settlement).

Such a change would require content within a platform to be made transparent, or for it to be possible – with reason – to make requests for specific content to be packaged and transmitted to another party, for further investigation. The DSA and similar regulatory shifts are creating additional expectations of transparency for online platforms within, e.g., the statements of reasons database; we can regard these as near-future design changes, and examine how these mechanisms develop over time, to understand where further transparency is feasible.

2.4 Design Provocation: Handover to Specialized Civil Society Experts

This provocation imagines a reporting ecosystem in which users are not left alone with legal categorizations and the filing systems of potentially adversarial platforms. Instead, the reporting interface includes an explicit *handover layer*, as a kind of reporting directory (not so dissimilar to the open listing of Trusted Flagger entities). With this, users would delegate all or part of the reporting task to a vetted civil-society helper (e.g., consumer organizations, digital rights NGOs, legal clinics, or Trusted Flagger organizations), whose expertise matches the user's situation (e.g., an interpersonal violation of privacy by someone online). The core idea is not to redesign the platform's interface from scratch, but to operationalize *support as infrastructure*: the user stays the rights-holder, while specialized intermediaries reduce friction, improve notice quality, and mitigate distress. This would require platforms to know which support entities to guide users toward, and have a process for guiding them there, i.e., proactively help the user in reporting content. However, this may help to signal trust in the platform [3] and that it is a safe place to interact with others.

3 Author Biographies and Interests

Marie-Therese Sekwenz is a PhD candidate at TU Delft and Deputy Director of the AI Futures Lab. Her research focuses on platform governance, with a particular emphasis on the implementation and enforcement of the EU's Digital Services Act (DSA). She explores how regulation shapes content moderation, user rights, design, and auditing on online platforms.

Simon Parkin is an Assistant Professor in the Cybersecurity group in the Technology, Policy, and Management (TPM) faculty at the Delft University of Technology (TU Delft, Netherlands). His specialization is in human-centred security: usability and perceptions of security-related technologies, security behaviour change, security economics, and decision-making in security technology management, support, and policy.

References

[1] 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). http://data.europa.eu/eli/reg/2016/679/oj/eng Legislative Body: EP, CONSIL.

[2] 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 Legislative Body: EP, CONSIL.

[3] Jens Riegelsberger, M Angela Sasse, and John D McCarthy. 2005. The mechanics of trust: A framework for research and design. *International journal of human-computer studies* 62, 3 (2005), 381–422.

[4] 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 *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 532–546. doi:10.1145/3715275.3732036

[5] 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. doi:10.1145/3351095.3372856