

Safety by Design: Closing the Regulation-Design Gap

From broad legal concepts in the Digital Services Act to feasible design affordances

Raquel De Haro i Pérez

Center for Law & Economics, ETH Zurich & Centre for Protecting Women Online, The Open University,
rdeharo@ethz.ch

Mariana Valente

Universität St. Gallen & InternetLab, mariana.valente@unisg.ch

Olga Jurasz

Centre for Protecting Women Online, The Open University, olga.jurasz@open.ac.uk

The latest decade has been marked by a surge in digital regulation, with countries around the world aiming to regulate the digital economy. Regulation frequently comes with technical and design requirements for intermediary services, particularly online platforms and social media, creating a translation gap between regulatory expectations and feasible implementation in code, interface design, and product architecture. The European Union and the United Kingdom have developed instruments like the Digital Services Act (DSA) and the Online Safety Act (OSA) to make internet services safer. These laws often include broad, undetermined duties on platforms and make references to design and interface architecture, yet they fail to specify how such requirements should be implemented in practice. Using Ofcom's guidelines on creating a safer life online for women and girls under the OSA and selected DSA provisions, this workshop examines how abstract obligations can be translated into concrete design affordances. We aim to explore technical feasibility, trade-offs, and unintended consequences, particularly in the context of online violence against women and girls, and to generate design-informed, participatory insights that can guide regulators and practitioners in implementing safety by design.

CCS CONCEPTS • Human-centered computing~Interaction design~Interaction design theory, concepts and paradigms • Security and privacy~Human and societal aspects of security and privacy~Social aspects of security and privacy • Social and professional topics~Computing / technology policy • Applied computing~Law, social and behavioral sciences~Law

Additional Keywords and Phrases: TFGBV, Safety by design, law, DSA

ACM Reference Format

1 INTRODUCTION

The latest decade has been marked by a surge in digital regulation, with countries around the world aiming to regulate the digital economy. Regulation frequently comes with technical and design requirements for intermediary services, particularly online platforms and social media. The result is a translation gap between regulatory expectations and feasible implementation in code, interface design, and product architecture.

Over the past decade, the European Union has positioned itself as a global leader in digital regulation, with instruments like the General Data Protection Regulation (GDPR), the AI Act or the Digital Services Act (DSA). The United Kingdom has followed a similar path, with rules like the Online Safety Act (OSA), the UK's equivalent to the DSA. The overarching goal of both the DSA and the OSA is to make internet services safer. These initiatives have developed in parallel with an accelerating global AI race and have therefore often been framed as hindering innovation. The more pressing question is not the false choice between regulation and innovation [1], but rather how to translate broad, vague regulatory objectives into technically feasible design choices.

At the same time, online violence against women and girls (OVAWG) has gained public and political attention recently, especially after X's AI chatbot Grok was used to generate 3 million images in 11 days [2]. Despite warnings from academia and civil about the lack of 'safety by design' measures (particularly for women and girls) in the deployment of AI systems and in platforms' affordances since at least 2017, the regulatory landscape remains focused on criminalizing *ad-hoc* behaviours that do not protect against the yet-to-exist forms of harms. At the same time, it imposes broad, undetermined duties on platforms, often leaving detailed instructions to voluntary guidelines. This is the case of Ofcom's guidance under the OSA on creating a safer life online for women and girls [3]. In this workshop, we use these guidelines to examine, and together with the assistants, brainstorm, how broadly drafted obligations under the DSA might be translated into concrete design affordances.

2 OFCOM'S GUIDELINES UNDER THE ONLINE SAFETY ACT (OSA) AND THE DIGITAL SERVICES ACT (DSA)

The Guidelines include 3 chapters with 3 specific action measures that service providers should adopt, respectively. For the purposes of this workshop, we will focus on Action 6: 'Reduce the circulation of content depicting, promoting or encouraging online gender-based harms', Action 7: 'Give users better control over their experiences', and Action 8: 'Reporting systems'.

Action 6 suggests persuasive measures like introducing friction by design into the user journey to give people time to think about what they are posting, especially when dealing with gender-based harms. Action 7 wants to add more granularity over the user experience and suggests measures like blocking and muting multiple accounts simultaneously. Action 8 focuses, among others, in the complaint systems, stating that they should be "*easy to find, easy to access, and easy to use and allow users to add supporting information to their complaints*" (5.26.b).

In turn, the DSA makes explicit and implicit references to 'design' throughout the text. These mentions are frequently framed in abstract terms. For example, Article 16 on notice and action mechanisms claims that "*those mechanisms shall be easy to access and user-friendly (...)*", in line with Ofcom's Action 8. Article 25's title expressly refers to "online interface design and organization" and prohibits providers of online platforms to "*design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of the service (...)*". Article 25.3 even claims that the Commission may issue guidelines on how this applies. Data protection scholars analyse this provision from the lens of 'dark patterns' [4], recognizing that Article 25.3 points (a) to (c) mention three types of 'dark patterns', namely giving more prominence to certain choices when asking the recipient for a decision; repeatedly request the recipient to make a choice where that choice has already been made; and making the procedure for terminating a service more difficult than subscribing to it.

Article 31 of the DSA also covers 'compliance by design' when regulating how online platforms allow consumers to conclude distance contracts with traders. Lastly, Articles 34 and 35 refer to the design of very large online platforms and very large online search engines when regulating risk assessments and measures for the mitigation of those risks. For

example, Article 35.1(a) suggests “adapting the design, features or functioning of their services, including their online interfaces” as a measure to tailor systemic risks identified in their assessments. Gender-based violence is recognized as a systemic risk in Article 34.1(d) of the DSA.

These provisions implicate design and interface architecture, yet they fail to specify how such requirements should be implemented in practice. In particular, Article 16 on notice and action mechanisms is the basis for other rights recognized in the law to be guaranteed, which may lead to a ‘domino effect of non-compliance’ [5]. If the reporting mechanisms are not user-friendly, users do not report content or abuse the reporting system. If users are not notified about the outcome, they do not challenge it and, ultimately, do not use out-of-court dispute settlement bodies to appeal a decision, which are an innovative resource introduced by the DSA.

Hence, using the Guidelines’ suggested action points we want to move from the DSA’s abstract obligations to concrete interface and architectural choices and bring to the debate how principles like ‘safety by design’ can and must be implemented in practice.

3 WORKSHOP CONTRIBUTION

This workshop addresses a translation gap: laws include broad, undetermined principles but dialogue with engineers and designers is limited. We want to create a space to brainstorm, explore, discuss and debate with designers, HCI researchers, engineers, and others about the unintended consequences of regulatory expectations, the technical feasibility (or desirability) of design choices, the burdens these choices place on platforms, and, in particular, how all of this is deployed in a context of increasing hostility against women and girls online.

Our goal is then threefold:

1. Translate specific Ofcom actions and DSA provisions into concrete design affordances.
2. Critically evaluate what is technically feasible and identify the trade-offs (e.g., privacy vs. safety).
3. Assess whether voluntary guidelines should become binding and explore alternative ways to improve online safety by design.

4 INTENDED OUTCOME

Importantly, the outcome of this workshop is not intended to be a legal blueprint, but rather a set of design-informed, participatory insights that can guide regulators (such as the European Commission when developing further guidance under the DSA, or the Ofcom on the feasibility of their existing Guidelines), and practitioners tasked with implementing safety by design practices.

REFERENCES

- [1] Anu Bradford. 2024. *The False Choice Between Digital Regulation and Innovation*. Northwestern University Law Review 119, 2 (2024), 377–453. Retrieved from <https://scholarlycommons.law.northwestern.edu/nulr/vol119/iss2/3/>
- [2] Robert Booth. 2026. Grok AI generated millions of sexualised images in a month, research says. *The Guardian* (22 January 2026). Retrieved from <https://www.theguardian.com/technology/2026/jan/22/grok-ai-generated-millions-sexualised-images-in-month-research-says>
- [3] Ofcom. 2025. *Guidance: A Safer Life Online for Women and Girls*. Ofcom, London, UK; 25 November 2025. Retrieved from <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-on-draft-guidance-a-safer-life-online-for-women-and-girls/statement-docs/guidance-a-safer-life-online-for-women-and-girls.pdf>
- [4] Cristiana Santos, Nataliia Bielova, Sanju Ahuja, Christine Utz, Colin Gray, and Gilles Mertens. 2024. Which Online Platforms and Dark Patterns Should Be Regulated under Article 25 of the DSA? (July 22, 2024), 39 pages. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4899559
- [5] Júlia Tar. 2026. Digital rulebook wins flagged by EU Commission as pressure rises. *MLex* (11 February 2026). Retrieved from <https://www.mlex.com/mlex/data-privacy-security/articles/2440023/-digital-rulebook-wins-flagged-by-eu-commission-as-pressure-rises>