The Sanitised Platform

Rachel Griffin*

Abstract:

Feminist legal scholar Vicki Schultz argues that US law on sexual harassment has created a "sanitised workplace", by encouraging employers to suppress any kind of sexual behaviour, while ignoring broader issues around gender equality. This paper employs Schultz's concept of sanitisation as a frame to critique current trends in European social media regulation, focusing on the 2019 Copyright Directive, 2021 Terrorist Content Regulation and the Digital Services Act proposed in 2020. EU law incentivises the deletion of various broadly-defined types of illegal content, which is also likely to suppress large amounts of legal and harmless content. Evidence of how social media platforms moderate content suggests that this over-enforcement will disproportionately suppress marginalised users and non-mainstream viewpoints, while increasing the influence of platforms' commercial goals on online communications. Yet at the same time, by focusing primarily on content (i.e. individual posts and uploads) over broader contextual and design factors, European regulation fails to effectively address many social harms associated with major social media platforms. Schultz's approach not only draws our attention to these failings, but provides theoretical insights as to how private ordering heightens these problems, enforces dominant discourse norms and subordinates online communication to commercial priorities.

A. Introduction

In a widely-cited 2003 article, revisited and updated in 2010, feminist legal scholar Vicki Schultz argues that US law on sexual harassment has created a "sanitised workplace", by encouraging employers to suppress any kind of sexual behaviour, while ignoring broader issues around gender equality[1]. This paper employs Schultz's concept of sanitisation as a frame to critique current trends in European social media regulation. It argues that European law is both under- and overinclusive in ways that parallel Schultz's arguments about the sanitised workplace. It incentivises platforms to frequently suppress harmless or valuable behaviour, while ignoring many individual behaviours and – more importantly – systemic problems that do cause harm. Schultz's approach not only draws our attention to these failings, but provides theoretical insights as to how private ordering heightens these problems, enforces dominant discourse norms and subordinates online communication to commercial priorities.

Schultz forcefully criticises the "sexual model" of sexual harassment prevalent in American jurisprudence on Title VII, the 1964 Civil Rights Act provision which banned sex discrimination in the workplace and was later interpreted (influenced by the campaigns of feminist legal scholars) as making employers liable for failing to prevent workplace sexual harassment. As Schultz's review of the case law shows, a focus on unwanted sexual conduct as the key criterion for unlawful sex discrimination came to eclipse other types of behaviour or features of the work environment which could reasonably be called discriminatory. Schultz argues that the sexual model is both over- and

* PhD candidate at the Law School of Sciences Po, Paris. I would like to thank Séverine Dusollier, Teodora Groza and an anonymous reviewer for helpful comments on earlier drafts. Contact: rachel.griffin@sciencespo.fr

[1] Vicki Schultz, 'The Sanitized Workplace' (2003) 112 *Yale Law Journal*, 2061-2194; Vicki Schultz, 'The Sanitized Workplace Revisited' in Martha Albertson Fineman, Jack E. Jackson and Adam P. Romero (eds), *Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations* (Routledge 2010).

underinclusive, and as a result signally fails to address the real causes and impacts of discrimination in the workplace, while causing significant collateral damage.

Schultz considers the sexual model underinclusive in two ways. First, it excludes important forms of sexist misconduct which are not obviously sexual in nature. Cases based on non-sexualised sexist behaviour have generally been less likely to succeed; claimants have been incentivised to frame hostile behaviour as sexualised to strengthen their claims, even where such interpretations are strained. Second, in focusing on individual sexual misconduct, the sexual model excludes consideration of broader, structural causes and manifestations of gendered discrimination. At the same time, Schultz argues that it is overinclusive, as the threat of liability for sexual misconduct incentivises workplaces to suppress and punish forms of sexualised behaviour which are not harmful. In practice, this typically disproportionately impacts employees from marginalised groups, and ultimately serves managerialist ideology and corporate interests.

In the context of social media governance, some parallels are already evident. Scholars, journalists and activists have long criticised large platforms' content moderation practices for simultaneous under- and overinclusivity, noting that illegal and dangerous content proliferates while legal and harmless content is frequently censored[2]. Moreover, current approaches to social media regulation and to workplace sexual harassment law share some structural features. Both primarily aim to regulate the behaviour of individuals (users/employees), although this may be difficult without also considering how it is influenced by the broader environment. Both utilise liability incentives to delegate the enforcement of legal norms to private actors (platforms/employers), who exercise a degree of direct control over the individuals in question. This paper contends that Schultz's theory of the sanitised workplace provides a useful lens to understand the flaws of current EU regulatory strategies. Her feminist approach to legal scholarship not only shows that the law is not achieving its purported goals, but focuses attention on why it has been interpreted in this way and whose interests it serves, as well as problematising the supposedly clear categories of behaviour it aims to regulate.

The paper proceeds as follows. Section B introduces recent trends in EU regulation of social media. Section C details the parallels between Schultz's arguments for the simultaneous under- and overinclusivity of US sex discrimination jurisprudence and the platform regulation context. Section D considers the relevance of Schultz's underlying theoretical insights. It argues that her feminist and sociolegal approach can sharpen critiques of social media law, by highlighting the ambiguity of the categories used to define "illegal content", and how in practice the enforcement of these rules is subordinated to commercial priorities. Section E concludes by advocating a more structural approach to social media regulation, focusing on platform design and business models over suppressing individual pieces of content.

B. Developments in EU social media regulation

Regulating "big tech" has become a major focus for European policymakers, culminating in the proposals released in late 2020 for the twin Digital Services and Digital Markets Acts, a flagship

[2] Jillian C. York and Corynne McSherry, 'Strong Identity, Strong Borders' (Electronic Frontier Foundation, 29 April 2019) <https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways> accessed 17 November 2021; Caroline Are, 'How Instagram's algorithm is censoring women and vulnerable users but helping online abusers' (2020) 20(5) *Feminist Media Studies* 741 <https://doi.org/10.1080/14680777.2020.1783805> accessed 17 November 2021; Ángel Díaz and Laura Hecht-Fellela, *Double Standards in Social Media Content Moderation* (Brennan Center for Justice, 2021) <https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation> accessed 17 November 2021.

initiative of the Von der Leyen Commission[3]. Social media content has been a prominent strand in these policy debates, spurred by intense media coverage of online extremism and hate speech, the potential influence of "fake news" on elections, and the "infodemic" of health misinformation during the Covid-19 pandemic[4].

Historically, EU regulation of social media content has been relatively light-touch, governed mostly by the "safe harbour" conditional immunity provisions in the 2000 E-Commerce Directive[5]. However, academics agree that we are currently seeing significant and far-reaching changes in the regulatory landscape[6]. Two overarching trends can be identified. First, platforms are subject to increasingly wide-ranging and stringent obligations to rapidly remove illegal content, as detailed in section B(I). Second, they are increasingly expected to undertake extensive private, semi-voluntary content regulation, including in relation to legal content. As section B(II) outlines, this is encouraged both through informal pressure from policymakers, and by legal provisions mandating the establishment of industry best practices, codes of conduct etc.

I. Obligations to remove illegal content

Under Article 14 of the E-Commerce Directive, which remains in force and will be replicated largely unchanged by the Digital Services Act[7], hosting services (which include social media) are immune from liability for making available illegal content posted by users, as long as they are not aware of the illegal content or remove it expeditiously on becoming aware of it. In practice, this has created a notice and takedown regime in which aggrieved parties can contact platforms to inform them about illegal content, with the result that the platform must remove it to avoid liability[8]. However, this general immunity is now complicated by three developments.

First, Article 14 precludes civil or criminal liability for user-generated content, but not injunctions. Since the E-Commerce Directive's introduction, injunctive relief has in particular played a key role in copyright enforcement[9]. More recently, the ECJ has accepted the use of injunctions to impose stringent moderation obligations on social media platforms. In its controversial *Glawischnig-Piesczek*

---

[3] European Commission, 'The Digital Services Act Package' (European Commission, 2020) <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> accessed 18 November 2021.

[4] Kirsten Gollatz and Leontine Jenner, *Hate Speech und Fake News – Zwei verwobene und politisierte Konzepte* (Humboldt Institut für Internet und Gesellschaft, 2018) <https://www.hiig.de/hate-speech-fake-news-two-concepts-got-intertwined-politicised/> accessed 17 November 2021; evelyn douek, 'The Year That Changed the Internet' (*The Atlantic,* 28 December 2021) <https://www.theatlantic.com/ideas/archive/2020/12/how-2020-forced-facebook-and-twitter-step/617493/> accessed 17 November 2021; Věra Jourová, 'Speech of Vice President Věra Jourová on countering disinformation amid COVID-19 "From pandemic to infodemic"' (European Commission, 11 October 2021) <https://ec.europa.eu/commission/presscorner/detail/en/speech_20_1000> accessed 17 November 2021.

[5] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L.178 ('E-Commerce Directive').

[6] Aleksandra Kuczerawy, 'General Monitoring Obligations: A New Cornerstone of Internet Regulation in the EU?' in CiTiP (ed), *Rethinking IT and IP Law: Celebrating 30 years CiTiP* (Intersentia 2019); Giancarlo Frosio and Martin Husovec, 'Accountability and Responsibility of Online Intermediaries' in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020).

[7] Article 5, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC' <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services> accessed 18 November 2021 ('Digital Services Act').

[8] Some member states have formalised this system with explicit provisions on the content and format of notices: see Aleksandra Kuczerawy, 'From "Notice and Takedown" to "Notice and Stay Down": Risks and Safeguards for Freedom of Expression' in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020).

[9] Christina Angelopoulos, 'Harmonizing Intermediary Copyright Liability in the EU: A Summary' in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020).

[2019] decision, the ECJ upheld an Austrian court's imposition of an injunction requiring Facebook not only to delete posts which had been held to defame the claimant, but to find and delete, on an ongoing basis, all identical or equivalent content[10]. This marks a significant shift from its earlier rulings in *Scarlet v SABAM* [2011] and *SABAM v Netlog* [2012] that injunctions could not require an internet service provider to actively check all user uploads for copyright-infringing content[11].

In *Glawischnig-Piesczek*, both the judgment and the Advocate General's opinion attached significant weight to the supposed availability of technological tools which could automatically detect content equivalent to that deemed illegal[12]. Experts consider this confidence in automated moderation tools unwarranted. They remain highly unreliable[13], and their use poses severe risks to users' freedom of expression and privacy rights[14]. Nonetheless, given the political pressure on platforms to take action on harmful content and the at-least-apparent promise that AI technologies can enable more comprehensive enforcement, the use of injunctions to impose such sweeping moderation obligations may become more common.

Second, the EU has introduced different liability regimes in some areas, specifically for terrorist content (under the 2021 Terrorist Content Regulation[15]) and copyright infringement (under the 2019 Copyright Directive[16]). The Terrorist Content Regulation requires platforms to remove terrorist content (which is broadly and vaguely defined, such that it could frequently include journalistic content[17]) within one hour after receiving a removal order from law enforcement[18]. They may also be required by competent national authorities to take further proactive measures to find and remove terrorist content[19]. Article 17 of the Copyright Directive, on the other hand, creates a new liability regime in which platforms are treated as primarily liable for copyright infringement unless they make best efforts to obtain a license from the rightsholder and, in the absence of a license, make best efforts to remove copyright works which have been notified to them by rightsholders and prevent all future uploads[20]. The latter obligation is widely acknowledged by academic experts[21], and by the Advocate General in his recent opinion in Poland's judicial review

---

[10] Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* [2019] (ECJ, 3 October 2019).

[11] Case C-70/10 *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM)* [2011] ECR I-11959; Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* [2012] (ECJ, 16 February 2012).

[12] *Glawischnig-Piesczek* [2019] (n 10).

[13] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance' (2020) 7(1) *Big Data & Society* <https://doi.org/10.1177/2053951719897945> accessed 17 November 2021.

[14] Daphne Keller, 'Facebook Filters, Fundamental Rights, and the CJEU's *Glawischnig-Piesczek* Ruling' (2020) 69(6) *GRUR International* 616 <https://doi.org/10.1093/grurint/ikaa047> accessed 17 November 2021. Keller has further argued that intermediary liability litigation structurally fails to account for users' rights and interests, whether those whose content is removed or the far greater number of users who might have been interested in having access to such content. In *Glawischnig-Piesczek*, as in most intermediary liability cases, users were not represented before the court.

[15] Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online [2021] OJ L.172 ('Terrorist Content Regulation').

[16] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L.130 ('Copyright Directive').

[17] Joris Van Hoboken, *The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications* (Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2019) <https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf> accessed 17 November 2021.

[18] Article 3, Terrorist Content Regulation (n 15).

[19] Article 5, Terrorist Content Regulation (n 15).

[20] Article 17, Copyright Directive (n 16).

[21] Giancarlo Frosio and Sunimal Mendis, 'Monitoring and Filtering: European Reform or Global Trend?' in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020); Maria Lillà Montagnani, 'A New Liability Regime for Illegal Content in the Digital Single Market Strategy' in Giancarlo Frosio (ed), *The Oxford*

case against Article 17[22], to require automated filtering of all user uploads in order to identify and block the notified copyright works.

Both pieces of legislation were highly controversial, due in large part to the perceived risks of "overblocking" of legal content[23]. Kuczerawy[24] and Frosio and Mendis[25] have suggested that, in combination with the *Glawischnig-Piesczek* ruling, these laws could mark the abandonment of the foundational principle in Article 15 E-Commerce Directive, as interpreted by the ECJ in the *SABAM* cases, that platforms cannot be under a general obligation to monitor all content for illegality. The principle has effectively been reinterpreted, such that an impermissible general monitoring obligation will not be taken to exist as long as platforms are only required to search for certain specific content, even if all content on the platform must be monitored for that purpose[26].

Finally, at the same time, some member states have introduced national measures requiring deletion of illegal content within short time limits, such as the German NetzDG[27], Austrian *Kommunikationsplattformen-Gesetz*[28], and French *loi Avia* (although most provisions of the latter were struck down by the Constitutional Council in June 2021[29]). While these laws can be regarded as simply specifying in more detail what constitutes "expeditious" removal under Article 14 E-Commerce Directive, their compatibility with the Directive is questionable, given that its aim was to create harmonised EU-wide standards and that it calls for platforms to be regulated only in the EU member state where they are headquartered[30].

II. Informal pressure and private ordering

A second feature of the developing regulatory landscape is the active encouragement of private ordering, through the encouragement of self-regulation and the creation of legal duties outside the intermediary liability framework[31]. Article 5 of the Terrorist Content Regulation requires platforms designated by regulators as exposed to terrorist content to take "specific measures" to address it.

Handbook of Online Intermediary Liability (Oxford University Press, 2020); Martin Senftleben, 'Institutionalized Algorithmic Enforcement—The Pros and Cons of the EU Approach to UGC Platform Liability' (2020) 14(2) *FIU Law Review* 299 <https://dx.doi.org/10.25148/lawrev.14.2.11> accessed 17 November 2021.

[22] Case C-401/19 *Poland v Parliament and Council*, Opinion of AG Øe.

[23] James Vincent, 'Europe's controversial overhaul of online copyright receives final approval' (*The Verge*, 26 March 2019) <https://www.theverge.com/2019/3/26/18280726/europe-copyright-directive> accessed 17 November 2021; Mathieu Pollet, 'EU adopts law giving tech giants one hour to remove terrorist content' (*Euractiv*, 28 April 2021) <https://www.euractiv.com/section/cybersecurity/news/eu-adopts-law-giving-tech-giants-one-hour-to-remove-terrorist-content/> accessed 17 November 2021.

[24] Kuczerawy, 'General Monitoring Obligations' (n 6).

[25] Frosio and Mendis (n 21).

[26] Bernd Justin Jütta and Giulia Priora, 'On the necessity of filtering online content and its limitations: AG Saugmandsgaard Øe outlines the borders of Article 17 CDSM Directive' (*Kluwer Copyright Blog,* 20 July 2021). <http://copyrightblog.kluweriplaw.com/2021/07/20/on-the-necessity-of-filtering-online-content-and-its-limitations-ag-saugmandsgaard-oe-outlines-the-borders-of-article-17-cdsm-directive/> accessed 17 November 2021.

[27] Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [2017] BGBl. I S. 3352 ('NetzDG').

[28] Bundesgesetz über Maßnahmen zum Schutz der Nutzer auf Kommunikationsplattformen (Kommunikationsplattformen-Gesetz – KoPl-G) [2020] BGBl. I Nr. 151/2020 ('Kommunikationsplattformen-Gesetz').

[29] Décision n° 2020-801 DC du 18 juin 2020, *Loi visant à lutter contre les contenus haineux sur internet* [2020]. Loi visant à lutter contre les contenus haineux sur internet

[30] Marc Liesching, *Stellungnahme zum Entwurf eines Gesetzes zur Änderung des Netzwerkdurchsetzungsgesetzes* (Deutscher Bundestag Ausschuss für Recht und Verbraucherschutz, 2020) <https://www.bundestag.de/resource/blob/700788/83b06f596a5e729ef69348849777b045/liesching-data.pdf> accessed 11 October 2021; Robert Gorwa, 'Elections, institutions, and the regulatory politics of platform governance: The case of the German NetzDG' (2021) 45(6) *Telecommunications Policy* <https://doi.org/10.1016/j.telpol.2021.102145> accessed 17 November 2021.

[31] Montagnani (n 21); Frosio and Husovec (n 6).

These measures remain largely at their own discretion, though one example specified in the provision is introducing new restrictions in their contractual community standards – a notable step towards privatised enforcement[32]. Article 17(10) Copyright Directive and Section 5 of the proposed Digital Services Act both mandate the Commission to work with businesses to develop industry codes and best practices[33]. Such co-regulatory measures have already significantly affected how platforms moderate both legal and illegal content, encouraging them to go beyond notice and takedown regimes and introduce more proactive content removal measures, including increasing use of automated moderation[34].

European policymakers have also placed informal pressure on platforms to introduce new content governance measures, often with the threat that harder regulation will otherwise follow[35]. In response to rising public and political concerns about racist hate speech and disinformation following the 2015 "refugee crisis", the 2016 Brexit referendum and the 2016 US election, leading European policymakers initially showed a clear preference for encouraging industry self-regulation[36]. The Commission negotiated a Code of Conduct on Hate Speech and Code of Practice on Disinformation with leading platforms in 2016 and 2018 respectively[37]. Informal pressure from European policymakers was also instrumental in leading major platforms to set up the GIFCT, an industry body which coordinates the removal of terrorist content across all participating platforms, using a hash database to flag any future uploads which are identical to previously removed content[38].

C. Under- and overinclusive regulation

It is widely recognised that content moderation is inevitably both under- and overinclusive, in the sense that all available methods of identifying banned content involve significant rates of both false negatives and false positives[39]. Land suggests that this is an inherent structural feature of online content moderation: given the scale at which platforms operate and the increasing use of automation, enforcement tends to consider only the content of posts and to ignore contextual factors which would enable a more nuanced consideration of their meaning and whether they are harmful[40]. The inevitability of errors must be taken into account when imposing new moderation obligations on platforms; inadequate safeguards against overblocking were a key point of criticism of both the Terrorist Content Regulation and the Copyright Directive.

---

[32] Van Hoboken (n 17).
[33] Article 17(10) Copyright Directive (n 16); Section 5 Digital Services Act (n 7).
[34] Hannah Bloch-Wehba, 'Automation in Moderation (2020) 53 *Cornell International Law Journal* 41.
[35] Paddy Leerssen, 'Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU' 6(2) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 99 <https://www.jipitec.eu/issues/jipitec-6-2-2015/4271> accessed 17 November 2021; Molly K. Land, 'Against Privatized Censorship: Proposals for Responsible Delegation' 60 *Virginia Law Review* 363.
[36] Gorwa (n 30).
[37] European Commission, *The EU Code of conduct on countering illegal hate speech online* (European Commission, 2016) <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> accessed 18 November 2021; European Commission, *Code of Practice on Disinformation* (European Commission, 2018) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> accessed 18 November 2021.
[38] evelyn douek, *The Rise of Content Cartels* (Knight First Amendment Institute, 2020) <https://knightcolumbia.org/content/the-rise-of-content-cartels> accessed 11 October 2021; Bloch-Wehba, 'Automation in Moderation' (n 34).
[39] evelyn douek, 'Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability' 121(3) *Columbia Law Review* 759. <https://www.columbialawreview.org/content/governing-online-speech-from-posts-as-trumpsto-proportionality-and-probability/> accessed 17 November 2021.
[40] Land (n 35).

However, EU platform regulations do not only create incentives for under- and overinclusive enforcement at the level of individual pieces of content which might be incorrectly left up or deleted. As this section will show, they are also under- and overinclusive in terms of the types of content, behaviour and circumstances which are deemed problematic and targeted for intervention in the first place.

I. Underinclusivity

Schultz argues that the sexual model of workplace sex discrimination both ignores and distracts from other important aspects of discrimination: it diverts employers', employees' and the courts' attention from sexist conduct which is not sexual in nature and from structural discrimination which cannot be reduced to individual misconduct. In platform regulation, it is important to question whether liability for certain types of illegal content distracts attention from other issues. Liability risks evidently influence how platforms allocate resources to moderation and other "trust and safety" programmes: this is illustrated by the major platforms' immediate investment of significant additional resources and moderation staff in Germany following the introduction of NetzDG[41]. However, as recent leaks from within Facebook revealed, even the biggest and wealthiest tech companies make very limited resources available for trust and safety projects[42]. Any deployment of resources and personnel to areas which do not generate revenue is unlikely to be approved by private corporations unless there is another clear financial justification, such as regulatory compliance. Thus, it can be assumed that any regulation requiring platforms to invest resources in one aspect of content governance risks reducing the resources available to investigate and address other social issues.

Like the narrow definition of sex discrimination which Schultz criticises, the tendency in European regulation to single out illegal content for deletion risks diverting attention from other types of harmful behaviour. Taking hate speech as an example, Ben-David and Matamoros-Fernández have documented how hate can systematically be spread on social media through content which does not itself fall under hate speech bans[43]. For example, users can post something just within the law which encourages hate speech in the comments, or like and comment on posts containing hate speech to increase their visibility to other users. Focusing only on the legality of content (posts, comments etc.) also ignores other types of abusive behaviour, such as coordinated malicious reporting of other users for legal or policy violations[44]. This may even be actively facilitated by rules requiring expeditious removal of illegal content, since incentivising quick responses may increase the likelihood of spurious complaints being upheld. There is anecdotal evidence of coordinated

---

[41] Philip Oltermann, 'Tough new German law puts tech firms and free speech in spotlight' (*The Guardian,* 5 January 2018) <https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firmsand-free-speech-in-spotlight> accessed 17 November 2021; Amélie Heldt, 'Reading between the lines and the numbers: an analysis of the first NetzDG reports' (2019) 8(2) *Internet Policy Review* 336 <https://doi.org/10.14763/2019.2.1398> accessed 17 November 2021.

[42] Jeff Horwitz, 'The Facebook Whistleblower, Frances Haugen, Says She Wants to Fix the Company, Not Harm It' (*Wall Street Journal*, 3 October 2021) <https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-to-fix-the-company-not-harm-it-11633304122> accessed October 11 2021.

[43] Anat Ben-David and Ariadna Matamoros Fernández, 'Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain' (2016) 10 *International Journal of Communication* 1167 <https://ijoc.org/index.php/ijoc/article/view/3697/1585> accessed 17 November 2021.

[44] Kate Crawford and Tarleton Gillespie, 'What is a flag for? Social media reporting tools and the vocabulary of complaint' (2016) 18(3) *new media & society* 410 <https://doi.org/10.1177/1461444814543163> accessed 17 November 2021; Stefanie Duguay, Jean Burgess and Nicolas Suzor, 'Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine' (2019) 26(2) *Convergence* 237 <https://doi.org/10.1177/1354856518781530> accessed 17 November 2021; Ari Ezra Waldman, 'Disorderly Content' (2021) 97 *Washington Law Review* (forthcoming) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906001> accessed 17 November 2021.

malicious reporting being used against victims of discrimination under the German NetzDG system[45].

Moreover, most EU regulation overlooks the structural factors which contribute to policy problems like online racism, disinformation and discrimination. In general, it targets the level of individual pieces of content – not only by requiring illegal content to be removed, but also by providing safeguards for freedom of expression which largely involve individual users complaining that their individual posts should be reinstated[46]. Focusing only on the content level fails to take into account how the harmfulness of content can differ widely depending on its context[47]. For example, one of the most harmful aspects of online harassment is its networked nature: users can easily incite others to join them in harassing a target with large numbers of abusive messages and other harmful actions, such as revealing personal information[48]. In such cases, examining the legality of individual messages may entirely overlook the primary harm they cause, as well as being practically unlikely to address enough of the harassment to have a significant impact.

Focusing on the content level also ignores important contextual and structural factors. Even in instances where harm is inflicted by individual pieces of content and can be remedied by content removal, considering contextual factors such as platform design and user cultures is crucial to ensure effective moderation. For example, much harmful behaviour is not reported to moderators because platform interfaces make it laborious for users to report it or because they do not expect a helpful response[49]. More broadly, structural factors such as platform algorithms, architectures and business models can contribute to significant social harms which cannot be resolved by removing individual pieces of content.

The typical social media business model, which is based on maximising user engagement and time on site in order to gather as much data and sell as much advertising space as possible, is frequently criticised for exacerbating social harms such as hate speech and disinformation. In particular, algorithms optimised for maximum user engagement are accused of promoting divisive, extremist and sensationalist content, and driving users towards harmful content and ideologies by showing them more extreme versions of whatever they are interested in[50]. Systematic studies of this

---

[45] Janosch Delcker, 'Germany's balancing act: Fighting online hate while protecting free speech' (*Politico*, 24 February 2020) <https://www.politico.eu/article/germany-hatespeech-internet-netzdg-controversial-legislation/> accessed 18 November 2021; Nicole Shephard, 'Digitale Gewalt an Frauen: Was kann das NetzDG?' (Heinrich Böll Stiftung, Gunda-Werner-Institut für Feminismus und Geschlechterdemokratie, 3 March 2020) <https://www.gwi-boell.de/de/2020/03/03/digitale-gewalt-frauen-was-kann-das-netzdg> accessed 18 November 2021.

[46] For comments on the inadequacy of individual user appeals as a safeguard against overblocking see Keller, 'Facebook Filters' (n 14), Frosio and Mendis (n 21).

[47] Richard Ashby Wilson and Molly K. Land, 'Hate Speech on Social Media: Content Moderation in Context' (2021) 52 *Connecticut Law Review* 1029 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3690616> accessed 11 January 2022; Owen Bennett, 'The promise of financial services regulatory theory to address disinformation in content recommender systems' (2021) 10(2) *Internet Policy Review* <https://doi.org/10.14763/2021.2.1558> accessed 18 November 2021.

[48] Sarah Jeong, *The Internet of Garbage* (The Verge 2018); Cynthia Khoo, *Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence* (Women's Legal Education and Action Fund (LEAF), 2021) <https://www.leaf.ca/publication/deplatforming-misogyny/> accessed 11 January 2022; Mary Anne Franks, 'Beyond the Public Square: Imagining Digital Democracy' (2021) 131 *Yale Law Journal Forum* < https://www.yalelawjournal.org/forum/beyond-the-public-square-imagining-digital-democracy> accessed 11 January 2022.

[49] Duguay et al (n 44); Rachel Griffin, 'New School Speech Regulation and Online Hate Speech: A Case Study of Germany's NetzDG' (GigaNet Symposium, Warsaw, December 2021) <https://www.giga-net.org/2021SymposiumPapers/GigaNet%20paper%20NetzDG.pdf> accessed 11 January 2022. n

[50] Siva Vaidhyanathan, *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy* (Oxford University Press 2018); Lance Bennett, Alan Borning, Martin Landwehr, Daniela Stockmann and Volker Wulf, *Treating Root Causes, not Symptoms: Regulating Problems of Surveillance and Personal Targeting in the Information Technology Industries* (G20 Insights, 2021)

phenomenon are lacking (and are hampered by the inaccessibility of platform data to independent researchers[51]). However, there is some evidence to support these claims. Journalistic investigations have found that Facebook, Instagram, YouTube and TikTok (all of which rely heavily on algorithmic content ranking and recommendations) actively recommend extremist content, as well as other harmful content such as self-harm, and show increasingly extreme content to users based on their previous interests[52].

Platforms' profiling and categorisation of users can also have more subtle impacts, such as reinforcing social inequalities. To target content and ads, platforms commonly profile users based on sensitive identity categories like gender and race, often using simplistic and offensive categorisations (e.g. imposing binary gender categories irrespective of user preferences[53]). These tend to symbolically further marginalise historically oppressed groups, by positioning them as deviations from a default "normal" user who is white, straight, etc.[54] They can also materially harm such groups in various ways: for example, by exposing sensitive information to advertisers[55], allowing advertisers to deliberately target vulnerable groups[56], or excluding them from economic opportunities.

A particularly well-studied example which obviously replicates historical patterns of discrimination is when marginalised users are excluded from adverts for jobs or housing. Facebook in the past allowed advertisers to explicitly exclude certain "ethnic affinities" from their ad audiences, which

<https://www.g20-insights.org/policy_briefs/treating-root-causes-not-symptoms-regulating-problems-of-surveillance-and-personal-targeting-in-the-information-technology-industries/> accessed 18 November 2021.

[51] Mathias Vermeulen, *The Keys to the Kingdom* (Knight First Amendment Institute, 2021) <https://knightcolumbia.org/content/the-keys-to-the-kingdom> accessed 11 October 2021.

[52] Jonas Kaiser and Adrian Rauchfleisch, 'Unite the Right? How YouTube's Recommendation Algorithm Connects The U.S. Far-Right' (*Medium*, 11 April 2018) <https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd> accessed 18 November 2021; Jeff Horwitz and Deepa Seetharaman, 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive' (*Wall Street Journal*, 26 May 2020) <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499> accessed 18 November 2021; Rob Barry, Georgia Wells, Joanna Stern and Jason French, 'How TikTok's Algorithm Serves Up Sex and Drug Videos to Minors' (*Wall Street Journal*, 8 September 2021) <https://www.wsj.com/articles/tiktok-algorithm-sex-drugs-minors-11631052944> accessed 18 November 2021; Keach Hagey and Jeff Horwitz, 'Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead' (*Wall Street Journal,* 15 September 2021) <https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215> accessed 18 November 2021; Center for Countering Digital Hate, *Malgorithm: How Instagram's Algorithm Publishes Misinformation and Hate to Millions During a Pandemic* (Center for Countering Digital Hate, 2021) <https://www.counterhate.com/malgorithm> accessed 11 October 2021; Brandy Zadrozny, '"Carol's Journey": What Facebook knew about how it radicalized users' (*NBC News*, 23 October 2021) <https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581> accessed 23 October 2021.

[53] Rena Bivens, 'The gender binary will not be deprogrammed: Ten years of coding gender on Facebook' (2015) 19(6) *new media & society* 880 <https://doi.org/10.1177/1461444815621527> accessed 17 November 2021.

[54] Kelley Cotter, Mel Medeiros, Chankyung Pak and Kjerstin Thorson, '"Reach the right people": The politics of "interests" in Facebook's classification system for ad targeting' (2021) 8(1) *Big Data & Society* <https://doi.org/10.1177%2F2053951721996046> accessed 11 January 2022.

[55] Eduard Fosch-Villaronga, Adam Poulsen, Roger A. Søraa and Bart Custers, 'Gendering Algorithms in Social Media' (2021) 23(1) *ACM SIGKDD Explorations Newsletter* 24 <https://doi.org/10.1145/3468507.3468512> accessed 11 January 2022.

[56] Nadine Bol, Joanna Strycharz, Natali Helberger, Bob van de Velde and Claes H. de Vreese, 'Vulnerability in a tracked society: Combining tracking and survey data to understand who gets targeted with what content' (2018) 22(11) *new media & society* 1996 <https://doi.org/10.1177%2F1461444820924631> accessed 11 January 2022. See also McMillan Cottom's theoretical work on 'predatory inclusion': Tressie McMillan Cottom, 'Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society' (2020) 6(4) *Sociology of Race and Ethnicity* 441 <https://doi.org/10.1177%2F2332649220949473> accessed 11 January 2022.

attracted heavy criticism[57]. However, researchers have shown that even without using criteria referring to race or other protected characteristics, advertisers can use proxies such as language or place of residence to exclude certain groups[58]. Moreover, even where there is no intention to discriminate, predictive targeting may automatically select audiences which are heavily skewed by race, gender and other protected characteristics[59]: this may, for example, reinforce the disadvantage women face in many professions by preventing them from seeing adverts for jobs that have historically been more appealing to men[60]. The use of predictive "affinity profiling" rather than concrete data about how users identify may allow such profiling to escape the ambit of anti-discrimination and data protection law[61].

The failure to address structural issues such as these, and the near-exclusive focus on illegal content as the key vector for harm, is a major flaw of the current European approach to platform regulation. It should be noted that the Digital Services Act represents a partial shift away from this approach, in that it introduces new obligations for platforms to assess and take action on "systemic risks stemming from the functioning and use of their services"[62]. Article 27 explicitly encourages them to make structural changes in order to mitigate these risks, such as altering platform design and algorithms, or reforming internal processes and organisation[63]. This represents a positive step away from a narrowly content-focused approach.

However, these changes should not be overstated. First, the relevant obligations apply only to the category of "very large online platforms", those with over 45 million EU users[64]. Smaller platforms also have new obligations, but these mostly address the content level (e.g. complaints and redress mechanisms for individual content removal decisions). Second, how effective the new regulations for very large online platforms will be in practice remains to be seen. They rely heavily on self-regulation and privatised enforcement: while the Commission will have new oversight powers including the right to require disclosure of information from very large online platforms and to conduct on-site inspections[65], the primary procedures intended to identify and address systemic risks will be platforms' internal risk assessments and voluntary measures, and yearly independent expert audits[66]. These types of privatised regulatory enforcement are intransparent and prone to

---

[57] Julia Angwin and Terry Parris Jr., 'Facebook Lets Advertisers Exclude Users By Race' (*ProPublica*, October 28 2016) <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race> accessed 11 January 2022; Thao Phan and Scott Wark, 'What Personalisation Can Do for You! Or: How to Do Racial Discrimination Without "Race"' (2021) 20 *Culture Machine* <https://culturemachine.net/vol-20-machine-intelligences/what-personalisation-can-do-for-you-or-how-to-do-racial-discrimination-without-race-thao-phan-scott-wark/> accessed 17 November 2021.

[58] Phan and Wark (n 57); Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove, 'Potential for discrimination in online targeted advertising' (2018) 81 *Proceedings of Machine Learning Research* 1 <http://proceedings.mlr.press/v81/speicher18a/speicher18a.pdf> accessed 11 January 2022.

[59] Jinyan Zang, 'How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity' (2021) 2021101901 *Technology Science* <https://techscience.org/a/2021101901/> accessed 17 November 2021; Phan and Wark (n 57).

[60] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove and Aaron Rieke, 'Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes' (2019) Vol 3 CSCW Article 199 *Proceedings of the ACM on Human-Computer Interaction* 1 <https://doi.org/10.1145/3359301> accessed 17 November 2021.

[61] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioral Advertising' (2020) 35 *Berkeley Technology Law Journal* 367.

[62] Article 26 Digital Services Act (n 7).

[63] Article 27 Digital Services Act (n 7).

[64] Article 25(1) Digital Services Act (n 7).

[65] Articles 50-66 Digital Services Act (n 7).

[66] Article 26-8 Digital Services Act (n 7).

capture[67], especially in complex, high-tech, information-based industries – such as social media – where external oversight is difficult[68].

In another influential critique of Title VII, Edelman theorised a process of "legal endogeneity" whereby formalities used to demonstrate compliance come to eclipse the substantive goals of regulation[69]. This allows businesses to influence the law to their own advantage, as courts and regulators increasingly defer to industry "best practices" when deciding whether legal standards have been met. Edelman's theory has been applied to technology regulation by Waldman[70], who finds ample evidence for similar processes taking place in privacy law enforcement. The Digital Services Act's regulatory approach may create similar problems, with formalities like risk assessments taking precedence over meaningful change in industry practices and ultimately reinforcing the status quo. A regulatory focus on mitigating discrete risks also overshadows broader questions about how technologies are used and for whose benefit[71]. Typically, harms that diverge from what powerful industry actors deem "normal" are classified as risks, while harms that stem from underlying structural features of an industry are not[72]. As will be discussed in more detail in section D, when private actors are charged with the definition and identification of risks, they will tend to construct those risks in the ways that best serve their own business interests.

II. Overinclusivity

Equally, EU regulation of social media content is overinclusive in significant respects. Like the American sex discrimination jurisprudence that Schultz criticises, it incentivises platforms to delete and suppress a wide range of content and behaviour which should not be considered harmful. Perhaps the best-documented example is the suppression by almost all major platforms of content which is sexually suggestive and/or related to sex work[73]. This causes significant material harm to sex workers by cutting off income sources, driving them towards more dangerous offline work and preventing them from advocating politically for their interests[74]. Blanket bans on sexual content also affect other users' wellbeing, for example by hampering access to sexual health advice[75], and lead to much broader policing of online art, culture and self-expression. For example, museums have regularly been blocked from posting images of nude art when promoting exhibitions[76]. Such policies could aptly be described as creating a "sanitised" internet.

---

[67] Michael Power, *The Audit Society: Rituals of Verification* (Oxford University Press 1999).

[68] Julie Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019).

[69] Laura Edelman, *Working Law: Courts, Corporations and Symbolic Civil Rights* (University of Chicago Press 2016).

[70] Ari Ezra Waldman, 'Privacy Law's False Promise' 97(3) *Washington University Law Review* 773.

[71] James Wilsdon and Rebecca Willis, *See-Through Science: Why Public Engagement Needs to Move Upstream* (Demos, 2004) <http://sro.sussex.ac.uk/id/eprint/47855/1/See_through_science.pdf> accessed 17 November 2021.

[72] Cohen (n 68).

[73] Jillian C. York, *Silicon Values: The Future of Free Speech Under Surveillance Capitalism* (Verso, 2021); Reina Sultan, 'Inside Social Media's War on Sex Workers' (*Bitch Media,* 23 August 2021) <https://www.bitchmedia.org/article/inside-social-medias-war-on-sex-workers> accessed 17 November 2021.

[74] Sophie K. Rosa, 'Sex Workers Denounce Instagram's "Puritanical" New Rules' (*Novara Media*, 21 November 2020). <https://novaramedia.com/2020/11/21/sex-workers-denounce-instagrams-puritanical-new-rules/ accessed 17 November 2021; York (n 73); Danielle Blunt and Zahra Stardust, 'Automating whorephobia: sex, technology and the violence of deplatforming – An interview with Hacking//Hustling' (2021) 8(4) *Porn Studies* 350 <https://doi.org/10.1080/23268743.2021.1947883> accessed 11 January 2022; Are (n 2).

[75] Danielle Blunt, Stefanie Duguay, Tarleton Gillespie, Sinnamon Love and Clarissa Smith, 'Deplatforming Sex: A roundtable conversation' (2021) 8(4) *Porn Studies* 420 <https://doi.org/10.1080/23268743.2021.2005907> accessed 11 January 2022.

[76] Elle Hunt, 'Vienna museums open adults-only OnlyFans account to display nudes' (*The Guardian*, 16 October 2021) <https://www.theguardian.com/artanddesign/2021/oct/16/vienna-museums-open-adult-only-onlyfans-account-to-display-nudes> accessed 17 November 2021.

These strict policies are significantly influenced by the US' 2018 FOSTA/SESTA legislation, which removed platforms' intermediary liability exemptions for content related to sex work[77]. Many platforms which did not already ban sexual content for commercial reasons responded to the legislation by implementing strict bans on sexual content worldwide, including in countries where sex work is legal[78]. However, the impact of European regulatory choices in this context should not be overlooked. First, if European legal systems did not grant platforms near-unfettered discretion to remove legal content under their contractual terms of service, they would not be able to arbitrarily impose US standards worldwide. Second, an important factor driving platforms to ban sexual content is app store policies: social media platforms rely heavily on users accessing them through mobile apps, and Apple (one of the two dominant app stores) is particularly notorious for banning apps that permit any kind of sexual content[79]. While the Commission is currently investigating Apple's App Store for anticompetitive behaviour relating in particular to its enforcement of in-app payments from which it takes a commission[80], European authorities have chosen not to intervene in Apple's use of its infrastructural power to enforce content policies which suppress sexual content across a wide swathe of the internet. Finally, some European countries have similar laws restricting pornography and advertising for sex work: examples include the German *Jugendschutzgesetz*, which sets broad requirements for online media accessible to under-18s to be child-friendly[81], and Article 380ter of the Belgian Criminal Code, which criminalises all advertising of sex work[82]. These would in any case incentivise platforms to take a restrictive approach.

Similar over-enforcement can be seen in regard to other types of content. Over-removal in copyright cases, based on obviously spurious notices from rights-holders, has been extensively documented[83]. Commentators have raised particular concerns about the inability of automated classifiers to identify legally protected uses of a work such as parody and quotation[84]. Copyright notices have also been abused to effect the removal of political content[85]. Attempts by platforms to remove terrorist content regularly censor activists aiming to challenge extremism or document violent incidents[86]. There have also been numerous documented instances of social media posts in

---

[77] An Act to amend the Communications Act of 1934 to clarify that section 230 of such Act does not prohibit the enforcement against providers and users of interactive computer services of Federal and State criminal and civil law relating to sexual exploitation of children or sex trafficking, and for other purposes [2018] Public Law 115–164 ('FOSTA-SESTA').

[78] Catherine Barwulor, Allison McDonald, Eszter Hargittai and Elissa M. Redmiles, "'Disadvantaged in the American-dominated Internet": Sex, Work and Technology' (2021) *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 563 <https://doi.org/10.1145/3411764.3445378> accessed 17 November 2021.

[79] Katrin Tiidenberg, 'Sex, power and platform governance' (2021) 8(4) *Porn Studies* 381 <https://doi.org/10.1080/23268743.2021.1974312> accessed 11 January 2022.

[80] European Commission, 'Antitrust: Commission opens investigations into Apple's App Store rules' (European Commission, 2020) <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1073> accessed 11 January 2022.

[81] Jugendschutzgesetz [2002] BGBl. I S. 2730 ('JuSchG').

[82] Article 380ter, Code Pénal [1867].

[83] Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List* (Center for Internet and Society at Stanford Law School, 2021) <http://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws> accessed 17 November 2021.

[84] Dan L. Burk, 'Algorithmic Faire Use' (2019) 86 *University of Chicago Law Review* 283; Bloch-Wehba, 'Automation in Moderation' (n 34); Montagnani (n 21).

[85] Julia Reda, 'How Copyright Bots Are Governing Free Speech Online' (*Digital Freedom Fund Blog*, 3 May 2021) <https://digitalfreedomfund.org/how-copyright-bots-are-governing-free-speech-online/> accessed 11 October 2021.

[86] WITNESS, *Content Regulation in the Digital Age Submission to the United Nations Human Rights Council Special Rapporteur for Freedom of Expression* (OHCHR, 2018) <https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/Witness.pdf> accessed 11 October 2021; Ellery Roberts Biddle, "'Envision a new war": the Syrian Archive, corporate censorship and the struggle to

which people of colour describe their experiences of racism being tagged as racist hate speech and deleted[87], or reclaimed slurs that are widely used in a positive sense in LQBTQ+ communities being indiscriminately censored[88].

Schultz highlights that the over-enforcement of harassment law is not evenly distributed, but reflects existing inequalities and power structures. She describes cases where sexual harassment claims were used to target LGBTQ+ employees, or where sexualised behaviour which was tolerated from white employees was treated as inappropriate when it came from people of colour. Similarly, the disproportionate impact of online content moderation on minorities and marginalised groups has been well documented. Policies on sexual content and nudity not only frame female and non-binary bodies as problematic[89]; they have also consistently been disproportionately enforced against women of colour and people who do not meet normative beauty standards, while celebrities and conventionally attractive white women are treated more leniently[90]. Waldman has also comprehensively detailed how the suppression of sexual content disproportionately affects LGBTQ+ users, maintaining social media platforms as "straight spaces"[91]. Major platforms often permit queer visibility only where it is desexualised, unthreatening and integrated into heteronormative family structures and values[92]. In the context of terrorist content – a regulatory priority for the EU – moderation unfolds through close cooperation between platforms and European security agencies[93], which primarily target Islamist terrorism and have long histories of racist and Islamophobic discrimination[94]. Bloch-Wehba has shown how the way platforms define

---

preserve public history online (*Global Voices*, 1 May 2019) <https://globalvoices.org/2019/05/01/envision-a-new-war-the-syrian-archive-corporate-censorship-and-the-struggle-to-preserve-public-history-online/> accessed 17 November 2021; Mathew Ingram, 'Social networks accused of censoring Palestinian content' (*Columbia Journalism Review,* 19 May 2021) <https://www.cjr.org/the_media_today/social-networks-accused-of-censoring-palestinian-content.php> accessed 17 November 2021; Isabella Barroso, 'Colombians "save the evidence" as they denounce social media censorship of protests' (*Global Voices*, 29 May 2021) <https://globalvoices.org/2021/05/29/colombians-save-the-evidence-as-they-denounce-social-media-censorship-of-protests/> accessed 17 November 2021.

[87] Jessica Guynn, 'Facebook while black: Users call it getting "Zucked," say talking about racism is censored as hate speech' (*USA Today,* 24 April 2019) <https://eu.usatoday.com/story/news/2019/04/24/facebook-while-black-zucked-users-say-they-get-blocked-racism-discussion/2859593002/> accessed 17 November 2021; Kishonna L. Gray and Krysten Stein, '"We 'said her name' and got zucked": Black Women Calling-out the Carceral Logics of Digital Platforms' (2021) 35(4) *Gender & Society* 538 <https://doi.org/10.1177%2F08912432211029393> accessed 17 November 2021.

[88] Dottie Lux and Lil Miss Hot Mess, 'Facebook's Hate Speech Policies Censor Marginalized Users' (*Wired*, 14 August 2017) <https://www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/> accessed 11 January 2022; Oliver L. Haimson, Daniel Delmonaco, Peipei Nie and Andrea Wegner, 'Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas' Vol 5 CSCW2 Article 466 *Proceedings of the ACM on Human-Computer Interaction* 1 <https://dl.acm.org/doi/10.1145/3479610> accessed 11 January 2022.

[89] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018); Ysabel Gerrard and Helen Thornton, 'Content Moderation: Social Media's Sexist Assemblages' (2020) 22(7) *new media & society* 1286 <https://doi.org/10.1177%2F1461444820912540> accessed 17 November 2021.

[90] Alex Peters, 'Nyome Nicholas-Williams took on Instagram censorship and won' (*Dazed Digital*, 28 August 2020) <https://www.dazeddigital.com/life-culture/article/50273/1/nyome-nicholas-williams-instagram-black-plus-size-censorship-nudity-review> accessed 17 November 2021; Carolina Are and Susanna Paasonen, 'Sex in the shadows of celebrity' (2021) *Porn Studies* <https://doi.org/10.1080/23268743.2021.1974311> accessed 17 November 2021.

[91] Waldman, 'Disorderly Content' (n 44).

[92] Clare Southerton, Daniel Marshall, Peter Aggleton, Mary Lou Rasmussen and Rob Cover, 'Restricted modes: Social media, content classification and LGBTQ sexual citizenship' (2021) 23(5) *new media & society* 920 <https://doi.org/10.1177%2F1461444820904362> accessed 11 January 2022.

[93] Rocco Bellanova and Marieke de Goede, 'Co-Producing Security: Platform Content Moderation and European Security Integration' (2021) *Journal of Common Market Studies* <https://doi.org/10.1111/jcms.13306> accessed 11 January 2022.

[94] Liz Fekete, 'Anti-Muslim Racism and the European Security State' (2004) 46(1) *Race and Class* 3 <https://doi.org/10.1177/0306396804045512> accessed 11 January 2022; Marie Martin, *Growing racism not just a*

and identify terrorist content is heavily shaped by security discourses which have consistently stigmatised and targeted Muslims, while downplaying threats from the extreme right[95]. This appears to be one reason that Arabic social media users – including activists and journalists – are particularly vulnerable to indiscriminate censorship[96].

D. What can we learn from Schultz's analysis?

As the previous section showed, there are clear parallels between Schultz's account of the sanitised workplace and the failings of current European platform regulation. However, her theory is not only useful in framing a descriptive account of these failings. This paper contends that Schultz models a feminist and sociolegal approach to legal scholarship which can sharpen our understanding and critique of current regulatory approaches.

I. Ambiguous categories and the power of interpretation

In the tradition of feminist and queer legal theory, Schultz problematises the supposedly clear legal categories on which the allocation of liability is based. She argues that clearly defining sexuality and walling it off from other aspects of social life is impossible; attempts to do so typically enforce dominant norms around sexual conduct and are imbued with bias against minority groups. The same could be said of defining "terrorist content", a broad and vague category which has long been used to legitimise anti-Muslim bias[97]; or even of "hate speech", a category which is meant to protect marginalised groups. Hate speech remains a deeply contested concept, and its interpretation is influenced by established social norms and power structures. As Post highlights, the term is rarely applied to elite discourse, even where it has evident discriminatory effects[98]. In practice, it has been used by social media platforms to suppress marginalised groups' challenges to oppressive social structures[99]. Overall, when European policymakers exhort platforms to be "responsible"[100] and act in accordance with European values[101], they are strategically glossing over the contested nature of these values.

Schultz also makes a forceful case for a sociolegal approach which highlights the gaps between how the "law on the books" allocates liability and how businesses respond to liability incentives in

---

member state issue (Statewatch, 2012) <https://www.statewatch.org/media/documents/analyses/no-196-eu-racism.pdf> accessed 11 January 2022.

[95] Bloch-Wehba, 'Automation in Moderation' (n 34).

[96] Marwa Fatafta, 'Facebook is bad at moderating in English. In Arabic, it's a disaster' (*Rest of World*, 18 November 2021) <https://restofworld.org/2021/facebook-is-bad-at-moderating-in-english-in-arabic-its-a-disaster/> accessed 11 January 2022.

[97] Van Hoboken (n 17); Bloch-Wehba 'Automation in Moderation' (n 34).

[98] Robert Post, 'Hate Speech', in Ivan Hare and James Weinstein (eds) *Extreme Speech and Democracy* (Oxford University Press 2009).

[99] Chloé Nurik, '"Men Are Scum": Self-Regulation, Hate Speech, and Gender-Based Censorship on Facebook' (2019) 13 *International Journal of Communication* 2878 <https://ijoc.org/index.php/ijoc/article/viewFile/9608/2697> accessed 17 November 2021; Gray and Stein (n 87); Human Rights Watch, 'Israel/Palestine: Facebook Censors Discussion of Rights Issues' (*Human Rights Watch*, 8 October 2021) <https://www.hrw.org/news/2021/10/08/israel/palestine-facebook-censors-discussion-rights-issues> accessed 18 October 2021; Elizabeth Dwoskin, Nitasha Tiku and Craig Timberg, 'Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show' (*Washington Post,* 21 November 2021) <https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/> accessed 11 January 2022.

[100] Ursula Von der Leyen, 'Speech by President von der Leyen at the Lisbon Web Summit. European Commission' (European Commission, 2 December 2020) <https://ec.europa.eu/commission/presscorner/detail/en/speech_20_2266> accessed 11 October 2021.

[101] European Commission, 'Europe fit for the Digital Age: Commission proposes new rules for digital platforms' (European Commission, 15 December 2020) <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347> accessed 17 November 2021.

practice. This is closely related to the former point, since ambiguous legal categories give businesses greater latitude for selective and self-interested enforcement. In the context of social media regulation, Frosio and Husovec have highlighted how formal legal liabilities are just one factor influencing platforms' content governance: "The real responsibility landscape is equally determined by a mixture of voluntary agreements, self-regulation, corporate social responsibility, and ad hoc deal-making."[102] This is especially and increasingly the case as the EU promotes private ordering measures such as self-regulation and flexible legal obligations based on industry "best practices", as outlined in section B(I).

This has implications for the normative orientation of the law. Edelman and Waldman's work on legal endogeneity shows empirically how, when the law charges private actors with enforcing vaguely-defined legal standards, they are likely to be interpreted in a way that serves corporate interests and dominant social norms more than the nominal goals of the regulation – even where these are supposedly progressive and egalitarian[103]. Schultz further argues that corporate actors will interpret the law in simplified ways to streamline enforcement processes, over-enforce to minimise liability risks, and focus on suppressing economically unproductive behaviour over behaviour which is actually harmful.

These problems are equally present in social media regulation. Speech rules must be simplified and streamlined to enable industrial-scale content moderation for global platforms[104]: the injustices that can result from such reductive interpretations are exemplified by the 2017 leak revealing that Facebook's content moderation guidelines defined as "hate speech" invective against white men, but not black children[105]. This dynamic is exacerbated by increasing reliance on algorithmic enforcement, given the limitations of currently-existing technology in understanding the meaning and context of expressions[106]. Speech rules shift to reflect what algorithms are capable of assessing, rather than what is actually considered desirable on policy grounds: for example, when all nudity is treated as pornography because it is what can most easily be identified by image recognition software[107].

Overblocking to minimise liability risks is also a much-discussed problem[108], and the influence of platforms' economic interests on their content moderation practices is evident. Content moderation experts point out that apparent inconsistencies in moderation policies tend to line up with whether the content in question is valuable for advertisers[109]. Recalling the ambiguities of the term "hate speech" discussed above, it is notable that major social media companies have openly negotiated with the World Federation of Advertisers to align the definition of hate speech in their platform content policies with what advertisers consider harmful to their "brand safety"[110].

---

[102] Frosio and Husovec (n 6), 614.

[103] Edelman (n 78); Waldman, 'Privacy Law's False Promise' (n 79).

[104] Robyn Caplan, *Content or Context Moderation? Artisanal, Community-Reliant and Industrial Approaches* (Data & Society, 2018) <https://datasociety.net/library/content-or-context-moderation/> accessed 11 January 2022; Sarah T. Roberts, 'Digital detritus: "Error" and the logic of opacity in social media content moderation' (2018) 23(3) *First Monday* <https://doi.org/10.5210/fm.v23i3.8283> accessed 17 November 2021.

[105] Julia Angwin and Hannes Grassegger, 'Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children' (*ProPublica,* 28 June 2017) <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms> accessed 17 November 2021.

[106] Gorwa et al. (n 13).

[107] Gillespie (n 89).

[108] Jack Balkin, 'Old-School/New-School Speech Regulation' 127 *Harvard Law Review* 2329; Keller, 'Facebook Filters' (n 14).

[109] Roberts (n 104); Are and Paasonen, (n 90).

[110] World Federation of Advertisers, 'WFA and platforms make major progress to address harmful content' (*World Federation of Advertisers*, 23 September 2020) <https://wfanet.org/knowledge/item/2020/09/23/WFA-and-platforms-make-major-progress-to-address-harmful-content> accessed 11 October 2021.

Considering how commercial priorities shape the application of the law is particularly important given the increasing turn towards private ordering in EU platform regulation. Platforms are not only being co-opted to enforce state speech regulation[111]. They are required to make "best efforts" on enforcement[112], choose the appropriate "specific measures" to respond to harmful content[113], utilise contractual terms and conditions to forbid harmful behaviour[114], and agree self-regulatory industry codes and best practices[115]. As Land suggests, these very broad discretionary powers over how the law is interpreted and how offline norms are adapted to the online context effectively amount to legislative power[116].

In the context of sexual harassment, Schultz shows that this delegation of power leads to a wide discrepancy between what the law states is illegal and what is actually banned in workplaces in practice. Similarly, delegating the interpretation of speech laws to platforms can significantly change what they are understood to mean. For example, contextual factors which are traditionally considered relevant in applying the law but which are harder to incorporate into industrial-scale moderation processes may be excluded entirely[117]. As noted above, this is exacerbated by automated enforcement, as standards shift to reflect the limited evaluative capabilities of software[118]. Safeguards provided by law – such as appeals systems for users, which the EU relies upon heavily in the Terrorist Content Regulation, Copyright Directive and Digital Services Act[119] – may not be effective or widely used in practice[120]. For example, Bloch-Wehba argues that where regulations heavily incentivise automated removal but stipulate that appeals should involve human review, in practice this will mean that the former takes place at scale but the latter cannot[121].

As well as disproportionately affecting marginalised groups through specific enforcement decisions, such private ordering is likely to more broadly reinforce mainstream or dominant norms about permissible views, discourse and sexual expression. Regulators' appeals for platforms to act "responsibly" and in accordance with public values[122] may risk incentivising a majoritarian approach, where platforms simply try to regulate content in line with dominant tastes and ideologies, while suppressing controversial or non-mainstream viewpoints – as observed by Waldman in his study of platforms as "straight spaces"[123].

Moreover, the EU's reliance on private ordering measures means that enforcement of regulatory objectives is in practice inseparably intertwined with platforms' pursuit of their own commercial goals. As discussed in section B(II), platforms are encouraged by the Copyright Directive and Terrorist Content Regulation (as well as the *Glawischnig-Piesczek* ruling) to design and deploy automated moderation solutions, and by the Terrorist Content Regulation and Digital Services Act to use their contractual terms and conditions to forbid undesired behaviour. These regulatory

---

111 Rory Van Loo, 'The New Gatekeepers: Private Firms as Public Enforcers' 106 *Virginia Law Review* 467; Balkin (n 108).

112 Copyright Directive (n 16).

113 Terrorist Content Regulation (n 15).

114 Terrorist Content Regulation (n 15); Digital Services Act (n 7).

115 European Commission, 'Code of Conduct on Hate Speech' (n 37); European Commission, 'Code of Practice on Disinformation' (n 37); Copyright Directive (n 16); Digital Services Act (n 7).

116 Land (n 35).

117 Land (n 35).

118 Gillespie (n 89); Burk (n 84).

119 Article 10 Terrorist Content Regulation (n 15); Article 17(9) Copyright Directive (n 16); Article 17 Digital Services Act (n 7).

120 Keller, 'Facebook Filters' (n 14); Frosio and Mendis (n 21); Senftleben (n 21).

121 Bloch-Wehba, 'Automation in Moderation' (n 34).

122 Frosio and Husovec (n 6).

123 Waldman, 'Disorderly Content' (n 44).

devices mean that there will be little distinction between content moderation for law enforcement purposes and commercial purposes. Platforms typically remove content under their contractual terms and conditions where possible, in order to apply consistent standards worldwide, even where it would anyway have to be removed based on applicable national law[124]. Legally-mandated moderation, voluntary moderation, and content curation more generally are all based on the same technical tools and classifications[125]. In practice, this means that any automated tools developed for law enforcement will likely also be deployed more widely in platforms' voluntary and commercially-motivated content governance[126]. The increasing use of automated content moderation tools subjects all online communication to the distorting influence of platforms' commercial goals[127]. This is likely to exacerbate the issues of overinclusivity and discrimination discussed in section C.

II. Whose interests does the law serve?

As with Schultz's analysis of sexual harassment law, we should not only observe that content regulation is over- and underinclusive, but ask who benefits from this state of affairs. Schultz argues that bright-line rules aiming to eliminate any kind of sexual conduct resonate with corporate interests and managerialist ideologies, which aim to make workplaces maximally efficient and rational[128]. Feminists arguing for a ban on sexual harassment found it politically expedient to put forward arguments that aligned with these perspectives, arguing that harassment made female employees less productive[129].

Similarly, we should question whose interests are served by the current approach to platform regulation. It is first relevant to note big tech companies' gargantuan lobbying expenditures in the EU, which outstrip all other sectors[130]. They also influence broader academic and policy debates by funding think tanks, research centres etc.[131] Leading platforms have been willing to accept greater regulation, as long as it strengthens dominant market players and does not demand fundamental changes to their business models[132]. These lobbying and advocacy efforts are not only about getting the regulatory results that they want, but shifting regulators' attention to the topics that are least threatening by amplifying "the criticism they can structurally live with"[133]. In this context, we should be attentive to potential ways that the orientation and priorities of European

---

[124] Heldt (n 41); Liesching (n 30).
[125] Niva Elkin-Koren and Maayan Perel, 'Separation of Functions for AI: Restraining Speech Regulation by Online Platforms' 24 *Lewis & Clark Law Review* 857.
[126] Bloch-Wehba, 'Automation in Moderation' (n 34).
[127] Jennifer Cobbe, 'Algorithmic Censorship by Social Platforms: Power and Resistance' (2020) *Philosophy & Technology* <https://doi.org/10.1007/s13347-020-00429-0> accessed 18 November 2021.
[128] Schultz's interpretation of early twentieth-century managerialist philosophies has been challenged by Lee: Rebecca K. Lee, 'The Organization as a Gendered Entity: A Response to Professor Schultz's The Sanitized Workplace' (2006) 13 *Columbia Journal of Gender and Law* 609. However, I believe Schultz's overall argument – that employers and managers will frequently see advantages in suppressing sexual conduct, which could be seen as undermining efficiency and discipline – is convincing.
[129] Abigail C. Saguy, 'Sexual harassment in France and the United States: activists and public figures defend their definitions' in Michèle Lamont and Laurent Thévenot (eds) *Rethinking Comparative Cultural Sociology: Repertoires of Evaluation in France and the United States* (Cambridge University Press 2010).
[130] Corporate Europe, 'Big Tech takes EU lobby spending to an all time high' (Corporate Europe Observatory, 31 August 2021) <https://corporateeurope.org/en/2021/08/big-tech-takes-eu-lobby-spending-all-time-high> accessed 11 October 2021.
[131] Laurie Clarke, Oscar Williams and Katharine Swindells, 'How Google quietly funds Europe's leading tech policy institutes' (*New Statesman,* 30 July 2021) <https://www.newstatesman.com/science-tech/2021/07/how-google-quietly-funds-europe-s-leading-tech-policy-institutes> accessed 18 November 2021.
[132] Aaron Sankin, 'What Does Facebook Mean When It Says It Supports "Internet Regulations"?' (*The Markup*, 16 September 2021) <https://themarkup.org/ask-the-markup/2021/09/16/what-does-facebook-mean-when-it-says-it-supports-internet-regulations> accessed 18 November 2021.
[133] Clarke et al. (n 116).

regulation might align with platforms' interests, even if individual regulatory requirements are unwelcome and burdensome.

Just as the focus on individual sexual misconduct in sex discrimination law excuses businesses from considering organisational context and structural inequality, European regulation arguably gives platforms obligations that are easy for them to "live with" instead of demanding structural changes that might discourage harmful speech and create more equal and inclusive online environments. European regulation has been criticised for focusing on the content of individual posts, rather than contextual factors like platform design[134]. However, this orientation serves platforms' interests insofar as it aligns with their current moderation practices[135], and with their commercial priorities. Irrespective of regulatory considerations, platforms have incentives to find and remove the most obviously offensive or illegal content, which is likely to repel users and advertisers[136]. They have much less incentive to redesign recommendation algorithms and platform architectures that contribute to social harms, given that these architectures in their current form are optimised for profit. In focusing on moderation at the content level rather than broader contextual, structural and design considerations, EU regulation effectively aligns with platform priorities more than the public interest.

It also reflects the influence of other powerful stakeholders. The new forms of private ordering that the EU has promoted in areas like terrorist content and disinformation involve close cooperation between platforms and national authorities. This not only enables those authorities to censor content online while circumventing formal legal channels and the checks and balances they entail[137], but also facilitates security agencies' collection of data on platform users and their activities[138]. EU regulation has also been particularly heavily influenced by lobbying from the copyright industries[139] – so much so that platforms are now, rather counterintuitively, subject to stricter intermediary liability for copyright infringement than for any other type of content, including terrorist content or child sexual abuse material[140]. Copyright owners are primarily interested in restricting the availability of specific content in which they have an economic interest, not in broader considerations about how online environments are constructed. This natural tendency towards a content-level orientation in one of the EU's highest-priority policy areas may have influenced its approach in other areas of social media regulation: an example is the notice and

---

134 Wilson and Land (n 47); Bennett (n 47).

135 Land (n 35).

136 Kate Klonick, 'The New Governors: The People, Rules and Processes Governing Online Speech' (2018) 131 *Harvard Law Review* 1598 <https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/ > accessed 18 November 2021; Roberts (n 104).

137 Land (n 35); Daphne Keller, 'Who Do You Sue? State and Platform Hybrid Power Over Online Speech' (Hoover Institution Aegis Series Paper No. 1902, 2019) <https://www.hoover.org/research/who-do-you-sue> accessed 11 January 2022.

138 Hannah Bloch-Wehba, 'Content Moderation As Surveillance' 36 *Berkeley Technology Law Journal* (forthcoming) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3872915> accessed 11 January 2022; Joris van Hoboken and Ronan Ó Fathaigh, 'Regulating Disinformation in Europe: Implications for Speech and Privacy' (2021) 6 *UC Irvine Journal of International, Transnational and Comparative Law* 9 <https://scholarship.law.uci.edu/ucijil/vol6/iss1/3/> accessed 11 January 2022.

139 Corporate Europe, 'Copyright Directive: how competing big business lobbies drowned out critical voices' (Corporate Europe Observatory, 10 December 2018) <https://corporateeurope.org/en/2018/12/copyright-directive-how-competing-big-business-lobbies-drowned-out-critical-voices> accessed 11 January 2022; Lucia Bertuzzi, 'Guidance on copyright law the result of "hefty lobbying", campaign groups say' (*Euractiv,* 8 June 2021) <https://www.euractiv.com/section/copyright/news/guidance-on-copyright-law-the-result-of-hefty-lobbying-stakeholders-say/> accessed 11 January 2022.

140 Folkert Wilman, 'The EU's system of knowledge-based liability for hosting service providers in respect of illegal user content – between the e-Commerce Directive and the Digital Services Act' (2021) 12(3) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 317 <https://www.jipitec.eu/issues/jipitec-12-3-2021/5343> accessed 11 January 2022.

takedown system, which was originally developed in the US Digital Millennium Copyright Act but now applies to all illegal content in the EU[141]. Thus, the over- and underinclusivity of the EU's platform regulation seems to reflect the interests of a variety of state and corporate actors in prioritising content-level regulation and surveillance over designing safe and egalitarian online spaces.

E. Conclusion

Schultz's theory of the sanitised workplace invites us to question whether the law as implemented in practice actually serves the goals it nominally pursues; whether the legal and semantic categories we use to delimit unacceptable behaviour can really be clearly and stably defined; and how the delegation of law enforcement to private actors can result in the law being twisted to serve commercial goals. These questions are highly relevant in the context of European social media regulation – especially at the present moment, when the regulatory landscape is rapidly shifting and new systems of privatised governance are being developed.

This paper contends that European regulation is functioning in tandem with, and actively reinforcing, commercial pressures to create "sanitised platforms". As section C shows, the tendencies towards under- and overinclusive regulation are already visible, as are its unevenly distributed effects. A wide range of content classed as illegal must be rapidly deleted, sweeping up significant portions of legal and harmless content along with it, and disproportionately suppressing marginalised groups and non-mainstream views. At the same time, beyond the limited provisions on systemic risk in the proposed Digital Services Act, platforms have few regulatory incentives to consider the broader social harms associated with their profit-optimised design choices and surveillance-based business models. We may end up with sterile social media platforms, increasingly empty of unconventional self-expression, creative uses of copyright works, and controversial political views – even while hate speech, disinformation and more insidious social harms, such as the discriminatory effects inherent in data-based profiling and ad targeting, continue to thrive.

As Schultz's analysis shows, these over- and underinclusive effects are connected with underlying regulatory structures. Where liability incentives are used to delegate the interpretation and enforcement of ambiguous and contested legal categories to private companies, there is an inherent risk that they will target behaviour which is unprofitable, rather than behaviour and organisational structures which are actually harmful. The turn to private ordering in European social media regulation exacerbates this risk further. By encouraging platforms to develop their own organisational and technical systems for enforcing speech law, and then to use the same enforcement systems to enforce their private, commercially-driven speech policies, European law effectively subordinates all social media communications to commercial priorities.

Schultz's policy prescriptions for workplace harassment focus on how work environments influence sexist behaviour, and gender equality more broadly. She advocates a tiered liability system, with reduced liability risks for companies which create more egalitarian and less gender-segregated workplaces. The feasibility of these detailed proposals in the employment context has been questioned[142], but the focus on structural and environmental factors could certainly provide a useful orientation for European platform regulation in the future. Instead of demanding "sanitised platforms" which indiscriminately suppress non-normative content, European regulators should

---

[141] Wilman (n 140).

[142] In particular, Williams suggests that creating blunt incentives for employers to have a gender-balanced workforce overlooks the complexity and durability of gender segregation in employment and the ways that women's work is frequently undervalued: Christine L. Williams, 'The Unintended Consequences of Feminist Legal Reform: Commentary on The Sanitized Workplace' (2006) 26 *Thomas Jefferson Law Review* 101.

be asking how the law can ensure social media platforms are incentivised to mitigate the harmful effects of advertising-driven business models – or to adopt different business models entirely – and to design diverse and inclusive online public spaces.