Special Issue on Law and Technology
February 2025

European Journal of Legal Studies

www.ejls.eui.eu

Website: https://ejls.eui.eu/

Submissions

The European Journal of Legal Studies invites submissions of General Articles and New Voices Articles on a rolling basis in the fields of comparative law, European law, international law, and legal theory. Journal is committed to the promotion of linguistic diversity and accepts submissions in any language, subject to the competence of the editorial board. The EJLS is committed to double-blind peer review. We therefore ask authors to ensure that their submissions include an anonymised version, together with a title page in a separate file containing identifying information. We invite authors to consult our Style Guide and Author Guidelines, which are available at https://ejls.eui.eu/contributeto-ejls.

All submissions should be sent to: submissions.ejls@eui.eu.

Published 14 February 2025
Catalogue no.: QM-AZ-25-001-EN-N
ISSN: 1973-2937

EUI DEPARTMENT OF LAW

European Journal of Legal Studies

LT Special Issue

Table of Contents

EJLS Symposium Editorial: Is Fairness in Digital Governance a Trap?

Stefania Di Stefano [*], Rebecca Mignot-Mahdavi [†], Barrie Sander [‡], Dimitri Van Den Meerssche [§], Roxana Vatanparast [**]

Contemporary legal and policy discussions around the social and political implications of algorithmic decision-making and digital governance have increasingly revolved around an aspiration for 'fairness'.[1] While this aspiration may seem at first sight to be a relevant ideal for both law and technology in digital governance, there are political and distributive implications at stake in taking fairness as a given, especially when left undertheorized. The language of fairness is used by a multiplicity of actors in global digital governance and thus serves a wide variety of purposes, from the normalization and stabilization of problematic practices, to attempts at constraining and resisting them. Fairness is an inherently pluridimensional concept within the international law discipline: the debates and discussions

* PhD Researcher in International Law, Geneva Graduate Institute.

† Assistant Professor of International Law, Sciences Po Law School.

‡ Assistant Professor of International Law, Leiden University, Faculty of Governance and Global Affairs.

§ Senior Lecturer in Law and Fellow, Queen Mary University of London, Institute of Humanities and Social Sciences.

** Assistant Professor of Law, Capital University Law School.

[1] See, for example, Jed Meers, Simon Halliday, and Joe Tomlinson, 'Why We Need to Rethink Procedural Fairness for the Digital Age and How We Should Do It', in Bartosz Brożek, Olia Kanevskaia, and Przemysław Pałka (eds), *Research Handbook on Law and Technology* (Edward Elgar 2023) 468; Raphaële Xenidis, 'Beyond Bias: Algorithmic Machines, Discrimination Law and the Analogy Trap' (2023) 14 Transnational Legal Theory 378.

 doi: 10.2924/EJLS.2025.LT.001

that took place at the 18th Annual Conference of the European Society of International Law (ESIL), devoted to unpacking the question of whether international law is fair, demonstrate as much. Not only is fairness a difficult concept to delineate, but it also carries different connotations in different languages.[2] The definitional challenges of the concept of fairness are also reflected in the different connotations that people attach to it – including other concepts such as 'justice', 'equity', 'proportionality', 'democracy', but also 'procedural fairness'.[3] These difficulties may be further exacerbated in the context of law and technology, where the epistemic disconnect between the different communities involved in digital governance may further translate into different characterizations and meanings attributed to the concept of 'fairness' which may or may not be reconcilable.[4] In this particular domain of international law and technology, invocations of 'fairness', as associated with a more general turn to 'ethics' in this regulatory space,[5] risk reinforcing rather than counteracting forms of data extraction and configurations of corporate power.

[2] Federica Cristani, '"Is International Law Fair? Le droit international est-il juste?": A Few Remarks from the 2023 ESIL Conference in Aix-en-Provence', (2024) 35 EJIL 1; see also, Hubert Mayer, 'Is International Law Fair? A Conference Report on the 18th Annual Conference of the European Society of International Law in Aix-en-Provence' (2024) 17 Z Außen Sicherheitspolit 217.

[3] ibid Cristani.

[4] See, for example, Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-discrimination Law and AI' (2021) 41 Computer Law & Security Review 1; and Hilde Weerts et al., 'Algorithmic Unfairness Through the Lens of EU Non-Discrimination Law' (2023) FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency 805.

[5] For a nuanced account that cautions against both ethics washing and ethics bashing, see Elettra Bietti, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy' (2022) FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency 210.

Against this background, this symposium brings together a collection of perspectives which aim to unpack different facets and functions of the language of fairness in digital governance. The symposium aims to contribute to existing scholarship by moving beyond a concern with algorithmic fairness and liberal norms of non-discrimination to critically examine the broader functions of the concept of fairness in the digital governance landscape around the world – whether in terms of the limits of procedural fairness as a means of addressing questions of online platform governance, the concepts of fairness implicitly embedded in different narratives related to digital health governance, or the limits of the concept of fairness as a means of appraising the deployment of machine learning technologies in modern warfare. Each paper speaks from a distinct observational viewpoint, identifies different traps that accompany the vocabulary of fairness in particular technological contexts, and offers distinct outlooks for digital governance going forward.

The opening papers of the symposium explore the EU's Digital Services Act (DSA). As part of a wider package of regulations designed to enhance accountability in digital governance,[6] the DSA regulates online intermediaries and platforms with the aim of protecting fundamental rights and fostering innovation and competitiveness across the EU. Rachel Griffin opens the symposium with a paper that critically examines the notion of fairness that underpins the DSA. Griffin reveals how the DSA's regulation of content moderation is underpinned by 'procedural fetishism' – an approach

[6] See generally, Giovanni De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (CUP 2022). For a critical reading of this EU Digital Policy Framework and a series of recommendations on how it could be aligned with the concept of 'Digital Fairness' – particularly from the vantage point of consumer law – see Natali Helberger et al., 'Towards Digital Fairness' (2024) 13 Journal of European Consumer and Market Law 1 24. On the normative tension between this EU framework and other models of data governance and AI regulation, see Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (OUP 2023).

that prioritises procedural fairness over substantive justice.[7] Adopting a feminist lens, the paper reveals the normative and discursive effects of the DSA's emphasis on procedural fairness for users facing intersecting forms of structural disadvantage. This is a particularly urgent and important analysis in relation to current changes in the content moderation policies of major online platforms as a result of political changes in the US.

Focusing on the procedural safeguards against arbitrary moderation decisions elaborated in Articles 14-21 DSA, the paper offers a threefold critique of the DSA's proceduralist approach. First, drawing on empirical studies, the paper suggests that the DSA's procedural safeguards are unlikely to be widely used (particularly within marginalised communities), will in any event prove difficult to enforce, and are unlikely to significantly constrain content moderation decisions due to the indeterminacy of platform policies. Second, drawing on feminist theory, the paper reveals the inadequacy of the DSA's normative assumption that procedurally fair decisions will generate substantively fair outcomes, particularly given the disjuncture between the DSA's provisions offering procedural safeguards at the level of individual decisions and the need to address higher-level considerations and systemic biases that produce unjust moderation decisions in practice. Finally, the paper suggests that the DSA's proceduralist approach could divert resources from more effective regulatory reforms and ultimately stabilise existing structures of power by enabling platforms to appear more legitimate. For Griffin, therefore, the DSA's notion of procedural fairness is a trap – one which holds out the promise of a fairer content moderation landscape through individualised procedural protections, but which is structurally incapable of addressing systemic biases and inequalities.

At the same time, Griffin identifies a number of avenues within the DSA that gesture beyond procedural fetishism and offer the possibility for more systemic improvements to content moderation practices. First, a series of

[7] See generally, Monika Zalnieriute, 'Against Procedural Fetishism: A Call for a New Digital Constitution' (2023) 30 Indiana Journal of Global Legal Studies 2 227.

transparency provisions that may help bring systemic problems to the surface through independent public scrutiny. Second, a series of due diligence provisions that mandate very large online platforms to identify and mitigate systemic risks. And finally, recent caselaw from the European Court of Justice that could inform the interpretation of the DSA, suggesting that EU fundamental rights law requires platforms to implement *ex ante* safeguards to minimise biases rather than relying narrowly on procedural safeguards. What emerges from Griffin's paper, therefore, is a cautious optimism – the DSA contains 'footholds',[8] which could, if utilised effectively, provide a basis for advancing systemic reform of moderation processes; the risk remains, however, that the DSA's emphasis on procedural fairness may ultimately crowd out precisely these opportunities.

While Griffin examines how procedural fairness functions as a trap within the EU's digital governance, Moritz Schramm extends this analysis to consider the global implications of the DSA's fairness framework. Schramm's point of departure is the so-called 'Brussels Effect'[9] –a phenomenon where EU regulations shape global practices, which occurs either when States adopt EU regulatory frameworks or where companies adhere to such regulations globally. Reflecting on the potential Brussels Effect of the DSA, the paper emphasises the significance of the DSA's articulation of broad, context-dependent normative goals, including 'fairness', rather than concise and substantive standards. Similar to Griffin, Schramm reveals how the articulation of the normative aspiration of 'fairness' amounts to a potential trap since it is the private platforms, whose problematic behaviour triggered the need for the DSA in the first place, that are tasked with concretising such aspirations in practice. Drawing on organisational theory and a legal realist perspective, Schramm suggests that private platforms are likely to exercise

[8] See generally, Dianne Otto, 'Decoding Crisis in International Law: A Queer Feminist Perspective', in Barbara Stark (ed), *International Law and Its Discontents* (CUP 2015) 115, 129-136.

[9] See generally, Anu Bradford, 'The Brussels Effect' (2012) 107 Northwestern University Law Review 1.

the discretion granted to them under the DSA in what some would call a self-interested manner, with the risk that the EU's normative goals become mere 'constitutional metaphors' that fail to foster systemic reform and instead place an EU 'quality seal' on only mildly changed corporate practices and a structurally flawed *status quo*.

In order to guard against the DSA stabilising rather than constraining private power and diffusing mere 'ceremonial certification' of private forms of ordering around the world, Schramm suggests that the European Commission should aim to strike a better balance between normative specificity and broadness through its supervisory function. To this end, the Commission should strive to make more space within its lawmaking process for engineering expertise so that its shift towards normative specificity is formulated through norms that are technologically feasible. By relying on technological expertise and improving the granularity of the DSA's regulatory demands, Schramm posits that it may be possible for more stringent standards to eventually reverberate around the world.

The third contribution in the symposium by Tsung-Ling Lee explores digital health governance in the context of the Association of Southeast Asian Nations (ASEAN). Confronting critical questions related to who governs digital health technologies, who benefits from them, and how risks that arise from them are distributed across different communities, Lee adopts a narrative lens to make sense of digital health governance in the ASEAN context. The paper thereby distils and conceptualises distinct legal narratives, revealing how fairness may operate as a trap in different ways.

The narrative of *technological solutionism*, which appears as a critical target across several contributions in this symposium, portrays digital health innovation as a remedy for unfairness in terms of healthcare access, availability, and quality. Yet, Lee argues, this risks overlooking the structural causes of health disparities, understating the potential bias and discriminatory effects of digital technologies, and privatising public questions that require social and institutional changes. The narrative of *human rights law*, by

contrast, strives to draw attention to the potential biases and discriminatory effects that may be generated through digital innovation, but often struggles to address factors that influence the underlying political economy of digital health innovations and infrastructure – with the risk that fairness, according to a human rights vocabulary, may ratify rather than challenge existing structures of ownership. Finally, Lee observes how the narrative of *data sovereignty* has emerged as a counter to perceived Western imperialism in the digital sphere, whether through China's assertion of data sovereignty as a defence against foreign ownership and control of digital health infrastructure and services, or Indigenous assertions of data sovereignty as a defence against data colonialism by governments and private actors. Examining China's Digital Silk Road, in particular, Lee reveals that the pursuit of data sovereignty may, in certain circumstances, create dependencies and replicate colonial dynamics under the guise of promoting the transformation of the digital economy in the region. Data sovereignty narratives may overlook novel patterns of Global South to Global South data extraction as well as the impacts of China's interpretation of data sovereignty on the ASEAN region, as it becomes increasingly reliant on Chinese digital infrastructure.

The final contribution to the symposium by Henning Lahmann critically examines the extensive surveillance practices relied upon to feed machine learning technologies in military decision-support systems. Reflecting also on recent events in Gaza and the West Bank, Lahmann reveals how Israel has instrumentalised the law of targeting within international humanitarian law (IHL) as a 'justificatory rhetorical framework' for rationalising the entrenchment of increasingly pervasive surveillance practices that feed its AI-driven military decision-support systems. In a trenchant critique on the permissive nature of IHL, Lahmann argues that these AI systems are thereby not merely rationalised and legitimised but under certain circumstances even become legally mandated as part of a proportionality calculus and institutional process of precaution. The paper explores how such recourse to IHL has thereby obscured the problematic use of machine learning for

'anomaly detection' – the identification of patterns and relations in large datasets that stand out from what the algorithm determines to be the state of 'normality'. Importantly, rather than simply describing the legal reality, the algorithms involved in this anomaly detection may be understood to be performative in actively producing this reality.

Lahmann suggests that scholarly interventions to date have focused primarily on the privacy and data protection dimensions of these practices – a frame that helps guard against egregious misuse of personal data for the purposes of armed conflict, but which ultimately amounts to a fairness trap that serves to rationalise harm caused to communities affected by algorithmic warfare as an inevitable trade-off in the pursuit of protecting civilian lives during armed conflict. In an important intervention to the field of international law and technology, Lahmann argues that these traditional normative frameworks thereby deflect attention away from the ways in which anomaly detection structurally impacts data subjects' political agency. Here, Lahmann draws on the concept of 'spontaneous political action' developed by Rosa Luxemburg and Hannah Arendt to reveal the critical role of spontaneity and collective political will-formation for the exercise of the right to self-determination under international law. Since machine learning algorithms operate on the expectation that the future will look similar to the past (and that anything which falls outside this backward-looking pattern raises suspicion), Lahmann argues that systems of algorithmic warfare inevitably render spontaneous political action fraught with significant risk, thereby structurally undermining the exercise of the right to self-determination. This presents a powerful critique of technologies of algorithmic inference, pattern detection, and clustering, which exceeds the existing regulatory repertoire.

This critical examination of fairness comes at a crucial moment in digital governance. As a variety of actors worldwide grapple with deploying and regulating digital tools and modes of governance, the allure of fairness – like other regulatory paradigms such as transparency, accountability, or efficiency – remains strong. The papers in this symposium demonstrate how

the deployment of fairness as a normative tool often serves to reinforce rather than remedy structural inequalities. From procedural fairness in content moderation to fairness claims in digital health governance and military applications, regulatory narratives and frameworks risk becoming legitimizing devices that stabilize harmful technological practices. Yet rather than abandoning fairness altogether, or promoting an alternative framework, these contributions point to the need for more nuanced approaches that attend to power dynamics and specific contexts where negotiation and contestation can and do take place. The challenge ahead lies not in replacing one regulatory paradigm with another, but in continuing to offer insights into the complex dynamics between law, technology and power – insights that prove essential for meaningful engagement with digital governance.

PROCEDURAL FETISHISM IN THE DIGITAL SERVICES ACT

Rachel Griffin [*]

Dominant social media platforms' content moderation practices operate highly unequally, disproportionately censoring marginalised users, while inadequately protecting them against hate speech and harassment. The EU's main response to such issues has been the 2022 Digital Services Act (DSA), which aims to empower individuals to understand and contest moderation decisions. Analysing the DSA from a feminist perspective, I describe this approach in terms of 'procedural fetishism' and develop a three-level critique. First, available evidence as to how similar systems work in practice suggests these provisions may have relatively little practical impact, especially for less-privileged user groups. Second, reviewing individual decisions cannot address the higher-level decisions and systemic biases that produce unreliable and discriminatory moderation. Moreover, the DSA allows platforms discretion over substantive policies, provided they are applied in a procedurally fair way—including policies that demonstrably disadvantage marginalised communities. Third, by diverting resources away from potentially more effective interventions, and making platforms' existing moderation systems appear more legitimate, the DSA's fetishisation of procedure could actively exacerbate or reinforce unaccountable and unfair moderation. I conclude by identifying some elements of the DSA with the potential to enable more systemic reform of social media moderation, and thereby more effectively address arbitrary censorship.

Keywords: Social media regulation; platform regulation; platform governance; content moderation; procedural fairness; feminism

[*] PhD candidate, School of Law, Sciences Po Paris. Contact: rachel.griffin@sciencespo.fr

Table of Contents

I Introduction

'👋 @elonmusk In Europe, the bird will fly by our 🇪🇺 rules. #DSA'. Internal Market Commissioner Thierry Breton tweeted this at billionaire Elon Musk[1] in October 2022, immediately after Musk completed his tumultuous

[1] Thierry Breton, '👋 @elonmusk In Europe, the bird will fly by our 🇪🇺 rules. #DSA' (Twitter, 28 October 2022) <https://twitter.com/thierrybreton/status/1585902196864045056?lang=en> accessed 18 January 2023. The waving hand emoji is not just a visual flourish. Placing another character before Musk's username ensured the tweet would be visible to all Breton's followers, indicating that Breton's tweet was intended for public consumption, not as a message for Musk personally.

acquisition of Twitter[2] (now renamed X[3]). Breton's message, also emphasised in a video with Musk[4] and reiterated on Twitter/X's decentralised rival Mastodon,[5] was that in Europe, social media platforms must respect regulations – notably the Digital Services Act (DSA), passed earlier in 2022[6] – which protect the public interest and prevent the arbitrary exercise of power.

Following the acquisition, Musk stated his intention to comply with the DSA.[7] However, he subsequently implemented numerous changes which raised concerns about site integrity and user safety. This includes introducing new content policies that appeared capricious and self-interested, like banning accounts sharing public information about his

[2] Kate Conger and Lauren Hirsch, 'Elon Musk Completes $44 Billion Deal to Own Twitter' (*The New York Times,* 27 October 2022) <https://www.nytimes.com/2022/10/27/technology/elon-musk-twitter-deal-complete.html> accessed 18 January 2023.

[3] Wes Davis, 'Twitter is being rebranded as X' (*The Verge*, 24 July 2023) <https://www.theverge.com/2023/7/23/23804629/twitters-rebrand-to-x-may-actually-be-happening-soon> accessed 5 December 2023.

[4] Thierry Breton, 'Today @elonmusk and I wanted to share a quick message with you on platform regulation 🇪🇺 #DSA' (Twitter, 9 May 2022) <https://twitter.com/ThierryBreton/status/1523773895974612992> accessed 18 January 2023.

[5] Thierry Breton, 'The DSA Checklist' (Mastodon, 30 November 2022) <https://social.network.europa.eu/@EC_Commissioner_Breton/109438646322670493> accessed 18 January 2023.

[6] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance) [2022] OJ L277/1 ('Digital Services Act').

[7] Javier Espinoza and others, 'EU and US turn up the heat on Elon Musk over Twitter' (*Financial Times,* 30 November 2022) <https://www.ft.com/content/a07ca1ae-9f9a-46ee-9457-27bb30e18ed2> accessed 18 January 2023.

private jet;[8] firing Twitter/X's entire AI ethics and accessibility teams,[9] as well as many policy staff and content moderators;[10] and introducing various design changes thought to exacerbate misinformation issues, like removing contextual information from hyperlinks.[11] Twitter/X's management since 2022 has been unconventional to say the least. However, some of these changes reflect broader industry trends. For example, during a wave of layoffs across the tech industry in 2023, most major platforms followed Twitter/X's lead and fired swathes of trust and safety, AI ethics and moderation staff.[12] Journalists have documented instances where other major platforms like Facebook and Instagram rolled out or refused to change design features raising well-known safety concerns.[13]

---

[8] Mitchell Clark, 'Elon Musk re-enabled Twitter accounts for several journalists banned over @ElonJet' (*The Verge*, 18 December 2022) <https://www.theverge.com/2022/12/17/23513620/elon-musk-suspended-journalists-twitter-reinstated-elonjet> accessed 18 January 2023.

[9] Taylor Hatmaker, 'Elon Musk just axed key Twitter teams like human rights, accessibility, AI ethics and curation' (*TechCrunch,* 4 November 2022) <https://techcrunch.com/2022/11/04/elon-musk-twitter-layoffs/> accessed 18 January 2023.

[10] Emma Roth, 'Twitter reportedly cut thousands of contractors without warning' (*The Verge,* 13 November 2022) <https://www.theverge.com/2022/11/13/23456554/twitter-reportedly-cut-thousands-contractors-without-warning-layoffs-elon-musk> accessed 18 January 2023.

[11] David Gilbert, 'The Israel-Hamas War Is Drowning X in Disinformation' (*Wired*, 9 October 2023) <https://www.wired.com/story/x-israel-hamas-war-disinformation/> accessed 5 December 2023.

[12] J.J. McCorvey, 'Tech layoffs shrink 'trust and safety' teams, raising fears of backsliding efforts to curb online abuse' (*NBC News*, 10 February 2023) <https://www.nbcnews.com/tech/tech-news/tech-layoffs-hit-trust-safety-teams-raising-fears-backsliding-efforts-rcna69111> accessed 5 December 2023.

[13] Karen Hao, 'How Facebook got addicted to spreading misinformation' (*MIT Technology Review*, 11 March 2021) <

This all raises questions about whether Breton's optimistic perspective on platform regulation is satisfactory. Can the new rules in the DSA effectively address harmful practices by social media companies, such as arbitrary moderation and inadequate investment in safety measures? And does being rule-bound inherently make these companies' power less harmful or more legitimate? In this article, I argue that the answer to both questions is no.

Focusing on how the DSA regulates content moderation,[14] I analyse its approach in terms of 'procedural fetishism'. This term was originally used in contexts such as international law[15] and urban planning[16] to argue against an excessive focus on decision-making procedures over substantive outcomes, highlighting that established procedural fairness norms often coexist comfortably with substantive injustice. In technology law, it has notably

---

https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> accessed 5 December 2023; Jeff Horwitz and Katherine Blunt, 'Instagram's Algorithm Delivers Toxic Video Mix to Adults Who Follow Children' (*Wall Street Journal*, 27 November 2023) < https://www.wsj.com/tech/meta-instagram-video-algorithm-children-adult-sexual-content-72874155> accessed 5 December 2023.

[14] Academic definitions of content moderation vary. Some authors include any governance mechanisms through which platforms structure user content and communications (see e.g. Tarleton Gillespie, 'Do Not Recommend? Reduction as a Form of Content Moderation' (2022) 8(3) *Social Media + Society* <https://doi.org/10.1177/20563051221117552>). However, the DSA provides a narrower definition: 'activities, whether automated or not…aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with [intermediaries'] terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that illegal content or that information': Art 3(t), Digital Services Act (n 6).

[15] David Clark, 'Iraq has wrecked our case for humanitarian wars' (*The Guardian*, 12 August 2003) <https://www.theguardian.com/politics/2003/aug/12/iraq.iraq1> accessed 18 January 2023.

[16] Peris S. Jones, 'Urban Regeneration's Poisoned Chalice: Is There an Impasse in (Community) Participation-based Policy?' (2003) 40(3) *Urban Studies* 581.

been developed by feminist and queer legal theorist Monika Zalnieriute.[17] Breton's tweet aptly sums up the ethos of procedural fetishism: that following rules and procedures automatically makes things better. Drawing on perspectives from feminist legal theory, I argue that this is not only practically flawed, but normatively unconvincing. Specifically, it fails to address substantive injustice and systemic biases against marginalised groups in content moderation.

Systematic quantitative analyses of moderation outcomes are rare, due largely to difficulties in accessing platform data.[18] However, mounting

[17] Monika Zalnieriute, 'Procedural Fetishism and Mass Surveillance under the ECHR' (*Verfassungsblog*, 2 June 2021) <https://verfassungsblog.de/big-b-v-uk/> accessed 18 January 2023; Monika Zalnieriute, '"Transparency-Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism' (2021) 8(1) *Critical Analysis of Law* 39; Monika Zalnieriute, 'Against Procedural Fetishism: A Call for a New Digital Constitution' (2023) 30(2) *Indiana Journal of Global Legal Studies* 227.

[18] This is likely to change gradually, as the DSA requires platforms to provide internal and public data to independent researchers: see Arts 40(4) and 40(12), Digital Services Act (n 6). However, many unresolved questions remain about ease of access and data quality: see e.g. Philipp Darius and others, *Implementing Data Access of the Digital Services Act: Collaboration of European Digital Service Coordinators and Researchers in Building Strong Oversight over Social Media Platforms* (Hertie School Centre for Digital Governance, 2023) <https://www.hertie-school.org/en/news/detail/content/hertie-school-researchers-present-recommendations-for-the-implementation-of-the-digital-services-act> accessed 5 December 2023; Julian Jaursch, Jakob Ohme & Ulrika Klinger, *Enabling Research with Publicly Accessible Platform Data: Early DSA Compliance Issues and Suggestions for Improvement* (Weizenbaum Institute, 2024) <https://www.weizenbaum-library.de/server/api/core/bitstreams/e589a831-f910-42e5-a8dd-4ccc9f00b9ca/content> accessed 29 September 2024; Philipp Darius, 'Researcher Data Access Under the DSA: Lessons from TikTok's API Issues During the 2024 European Elections' (*Tech Policy Press*, 24 September 2024) <https://www.techpolicy.press/-researcher-data-access-under-the-dsa-lessons-from-tiktoks-api-issues-during-the-2024-european-elections/> accessed 29 September 2024.

evidence suggests that major platforms' moderation systems are rife with familiar heterosexist, racist, classist and other biases. For example, ethnographic and survey research has found that experienced professional creators perceive moderation and recommendation systems as pervasively biased against marginalised groups.[19] Qualitative research suggests policies banning 'adult' content are applied particularly strictly to LGBTQIA+ creators and other minorities, and more leniently to 'mainstream' content from conventionally attractive white women and celebrities.[20] Activists and ordinary users are regularly censored when attempting to challenge prejudice and discrimination.[21] At the same time, moderation systems fail to

[19] Zoë Glatt, 'Precarity, discrimination and (in)visibility: An Ethnography of "The Algorithm" in the YouTube Influencer Industry' in Elisabetta Costa and others (eds), *The Routledge Companion to Media Anthropology* (Routledge 2022); Brooke Erin Duffy and Colten Meisner, 'Platform governance at the margins: Social media creators and algorithmic (in)visibility' (2022) 45(2) *Media, Culture & Society* 285; Jordan Foster, '"It's All About the Look": Making Sense of Appearance, Attractiveness, and Authenticity Online' (2022) 8(4) *Social Media + Society* <https://doi.org/10.1177/20563051221138762>

[20] Carolina Are and Susanna Paasonen, 'Sex in the shadows of celebrity' (2021) 8(4) *Porn Studies* 411; Ari Ezra Waldman, 'Disorderly Content' (2022) 97(4) *Washington Law Review* 907; Alexander Monea, *The Digital Closet: How the Internet Became Straight* (MIT Press 2022).

[21] Chloé Nurik, '"Men Are Scum": Self-Regulation, Hate Speech, and Gender-Based Censorship on Facebook' (2019) 13 *International Journal of Communication* 2878; 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; Thibault Grison and Virginie Julliard, 'Les enjeux de la modération automatisée sur les réseaux sociaux numériques : les mobilisations LGBT contre la loi Avia' (2021) 10 *Communication, technologies et développement* <https://doi.org/10.4000/ctd.6049>

adequately protect marginalised users from hate and harassment,[22] further limiting their ability to express themselves.[23]

These findings are unsurprising, given available knowledge about how commercial content moderation works. Platforms' policies reflect commercial pressures to cater to wide mainstream audiences, and to advertisers' branding goals.[24] This incentivises them to ban content deemed offensive, 'toxic' (likely to make people leave conversations[25]) or unsuitable for children,[26] even where such forms of expression may be important and valuable to minorities.[27] Poorly-paid and poorly-trained moderators working under intense pressure, often lacking relevant linguistic and cultural context, make snap decisions which are inevitably influenced by

[22] Rachel Griffin, 'The Sanitised Platform' (2022) 13(1) *JIPITEC* 36.

[23] Eugenia Siapera, 'Online Misogyny as Witch Hunt: Primitive Accumulation in the Age of Techno-capitalism' in Debbie Ging and Eugenia Siapera (eds) *Gender Hate Online* (Springer International 2019); Mary Anne Franks, 'Beyond the Public Square: Imagining Digital Democracy' (2021) 131 *Yale Law Journal Forum* 427.

[24] Rachel Griffin, 'From brand safety to suitability: advertisers in platform governance' (2023) 12(3) *Internet Policy Review* <https://doi.org/10.14763/2023.3.1716>

[25] Zeerak Talat, '"It ain't all good": Machinic abuse detection and marginalisation in machine learning' (PhD thesis, University of Sheffield 2021) <https://etheses.whiterose.ac.uk/30950/> accessed 18 January 2023.

[26] Ben Wagner, *Global Free Expression: Governing the Boundaries of Internet Content* (Springer Nature 2016), 111.

[27] For example, many LGBTQIA+ communities particularly value openly expressing their sexuality and gender identity in ways which resist mainstream norms about 'appropriateness', making them particularly likely to be moderated: Oliver L. Haimson and others, 'Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies' (2019) 21(3) *Feminist Media Studies* 345; Clare Southerton and others, 'Restricted modes: Social media, content classification and LGBTQ sexual citizenship' (2021) 23(5) *New Media & Society* 920; Rachel Griffin, 'The Heteronormative Male Gaze: Experiences of Sexual Content Moderation Among Queer Instagram Users in Berlin' (2024) 18 *International Journal of Communication* 1266; Waldman, 'Disorderly Content' (n 20).

(un)conscious prejudices.[28] Automated moderation systems frequently indiscriminately censor keywords deemed offensive,[29] which often particularly affects marginalised users: for example, where minority communities use reclaimed slurs,[30] or LGBTQIA+-related keywords are deemed offensive due to their use in heterosexual pornography.[31] More sophisticated AI classifiers are trained on past decisions which reflect human moderators' prejudices,[32] and on industry-standard datasets pervaded by biases and stereotypes.[33] Unsurprisingly, then, they exhibit familiar forms of algorithmic bias, tending to associate marginalised identities with 'toxicity' and negative connotations.[34]

This article presents a feminist critique of the DSA's 'procedural fetishist' approach, arguing that it cannot address these issues adequately. In

[28] 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>; Monea (n 20); ACLU and Daphne Keller, 'Daphne Keller and ACLU File Comment to Meta Oversight Board in "UK Drill Music" Case' (*Stanford Cyber Policy Center*, 23 August 2022) <https://cyber.fsi.stanford.edu/news/daphne-keller-and-aclu-file-comment-uk-drill-music-case> accessed 18 January 2023.

[29] 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; Grison and Julliard (n 21).

[30] Grison and Julliard (n 21).

[31] Monea (n 20).

[32] 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>

[33] Monea (n 20).

[34] Nicolas Kayser-Bril, 'Automated moderation tool from Google rates People of Color and gays as "toxic"' (*AlgorithmWatch,* 19 May 2020) <https://algorithmwatch.org/en/automated-moderation-perspective-bias/> accessed 18 January 2023; Talat (n 25).

particular, it draws on Adam Romero's framing of feminism as methodology.[35] For Romero, feminist legal research is characterised not only by addressing substantive topics related to gender and inequality, but also by methodological approaches which are interdisciplinary, empirically-informed, and critical of the values implicit in legal frameworks.[36] In this sense, feminist approaches are not limited to asking the 'woman question'[37] but aim to analyse and critique multiple, intersecting forms of inequality.[38] Such methods have informed influential critiques of legal frameworks focused on formal legal protections, procedural fairness and individual rights, showing that they are unsuited to addressing systemic and institutional disadvantage.[39] Feminist theorists have also criticised the normative assumptions and discursive effects of such legal frameworks, arguing that they reinforce liberal, pro-market framings of policy problems and solutions,[40] and obscure particular interests of women and other structurally

[35] Adam P. Romero, 'Methodological Descriptions: "Feminist" and "Queer" Legal Theories' in Martha Albertson Fineman, Jack E. Jackson and Adam P. Romero (eds), *Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations* (Routledge 2010).

[36] See also Lyn K.L. Tjon Soei Len, 'On politics and feminist legal method in legal academia', in Marija Bartl and Jessica C. Lawrence, *The Politics of European Legal Research* (Elgar 2022).

[37] Cochav Elkayam-Levy, 'A Path to Transformation: Asking "The Woman Question" in International Law' (2021) 42(3) *Michigan Journal of International Law* 429.

[38] As Romero (n 35) notes, these methods are characteristic of but far from unique to feminist research. Intersectional feminist scholarship overlaps both substantively and methodologically with other critical approaches to legal research, such as critical race theory, Marxist legal theory and queer legal theory.

[39] Kimberlé Crenshaw, 'Race, Reform, and Retrenchment: Transformation and Legitimation in Antidiscrimination Law' (1988) 101(7) *Harvard Law Review* 1331; Anna Lauren Hoffmann, 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse' (2019) 22(7) *Information, Communication & Society* 900.

[40] Wendy Brown, 'Suffering Rights as Paradoxes' (2000) 7(2) *Constellations* 208.

disadvantaged groups.[41] Drawing on the methodological approaches discussed by Romero, this article aims to show the relevance of such critiques to the DSA, arguing that it fails to address the disparate impacts of content moderation on users facing various intersecting forms of structural disadvantage.

The article proceeds as follows. Section II describes the DSA's approach to regulating content moderation. Section III then develops a critique of the DSA, informed by the feminist methodological traditions discussed above: on the one hand, it critically considers the values animating the DSA's regulation of content moderation – in particular, its focus on formal equality and procedural fairness – and on the other, by drawing on empirical literature from various disciplines, it shows how pursuing these values through procedural protections fails to address systemic disadvantage. This critique proceeds in three stages. First, it argues that the DSA's procedural protections face practical limitations which undermine their ability to prevent arbitrary and discriminatory moderation. Second, even were these practical limitations overcome, these provisions remain normatively unsatisfactory, as they fail to address systemic inequalities in content moderation. Finally, the DSA's procedural fetishist approach could actively worsen unaccountability and bias, by legitimising corporate power and diverting resources from more effective interventions. Section IV concludes by highlighting some aspects of the DSA which go beyond proceduralism, and could offer better avenues to address injustice in content moderation. In particular, these include provisions on transparency and research data access, systemic risk mitigation, and *ex ante* fundamental rights obligations.

[41] Hilary Charlesworth, 'Feminist Methods in International Law' (1999) 93(2) *American Journal of International Law* 379; Hoffmann (n 39).

II THE DSA'S PROCEDURALIST APPROACH

Experts have described the DSA as 'in its essence…a digital due process regulation'[42] and as centring 'procedure over substance'.[43] Its provisions on content moderation (set out in Chapter III, sections 1-3) primarily aim to strengthen accountability by regulating platforms' decision-making procedures, rather than substantive decisions. For example, Article 14 requires platforms to publish clear and accessible moderation policies,[44] and apply them in a 'diligent, objective and proportionate manner'.[45] Article 15 requires publication of transparency reports outlining how moderation systems work (e.g. how human review and automated moderation are used) and how much content is moderated for different reasons. Article 16 requires an easy-to-use interface to report illegal content on the platform.

Subsequent provisions specifically aim to provide procedural safeguards against arbitrary moderation decisions. Article 17 requires 'a clear and specific statement of reasons' for users whose content is restricted (which includes exclusion from revenue-sharing and demotion in recommendations, as well as removal or account deletion).[46] Article 20 requires internal appeals processes for users to challenge moderation decisions. Platforms must review complaints 'in a timely, non-discriminatory, diligent and non-arbitrary manner',[47] with some involvement from human reviewers, though the exact wording – '*under the supervision* of appropriately qualified staff, and not *solely* on the basis of

[42] Martin Husovec, 'Will the DSA work?' (*Verfassungsblog*, 9 November 2022) <https://verfassungsblog.de/dsa-money-effort/> accessed 3 January 2023.

[43] Pietro Ortolani, 'If You Build It, They Will Come' (*Verfassungsblog,* 7 November 2022) <https://verfassungsblog.de/dsa-build-it/> accessed 18 January 2023.

[44] Art 14(1), Digital Services Act (n 6).

[45] Art 14(4), Digital Services Act (n 6).

[46] Art 17(1), Digital Services Act (n 6).

[47] Art 20(4), Digital Services Act (n 6).

automated means'[48] – implies reviews can also involve automated decision-making tools.[49] If users show that decisions have no basis in the law or in platforms' stated content policies, platforms must reverse them.[50] Article 21 allows users to appeal further to certified independent dispute settlement institutions; their decisions are non-binding, but platforms must engage with them in good faith.[51]

Giovanni De Gregorio suggests that these obligations (and similar procedural rights in EU data protection law) are united by an underlying liberal ethos, aiming to protect autonomy and dignity by enabling individuals to understand and contest decisions which affect them.[52] This approach also reflects prevalent views in academic literature on social media governance. Scholarship raising concerns about censorship, discrimination and unaccountable moderation frequently turns to procedural fairness or 'due process' norms as a solution.[53] These proposals generally draw on

[48] Art 20(6), Digital Services Act (n 6) (emphasis added).

[49] For a detailed analysis see Rachel Griffin & Erik Stallman, 'A Systemic Approach to Implementing the DSA's Human-in-the-Loop Requirement' (*Verfassungsblog*, 22 February 2024) <https://verfassungsblog.de/a-systemic-approach-to-implementing-the-dsas-human-in-the-loop-requirement/> accessed 29 September 2024.

[50] Art 20(4), Digital Services Act (n 6).

[51] Art 21(2), Digital Services Act (n 6).

[52] Giovanni De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Cambridge University Press 2022).

[53] Nicolas Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms' (2018) 4(3) *Social Media + Society* 4; Giovanni De Gregorio, 'Democratising online content moderation: A constitutional framework' (2019) 36 Computer Law & Security Review 105374; Rory Van Loo, 'Federal Rules of Platform Procedure' (2020) 88 *University of Chicago Law Review* 829; Torben Klausa, 'Graduating from "new-school"–Germany's procedural approach to regulating online discourse' (2022) 26(1) *Information, Communication & Society* 54; Evelyn Douek, 'Content Moderation as Systems Thinking' (2022) 136 *Harvard Law Review* 526.

principles from public law and liberal constitutionalism:[54] while some authors suggest platforms should emulate judicial institutions,[55] others argue moderation should draw on transparency and consultation procedures in administrative law.[56] The DSA displays elements of the administrative approach, with provisions mandating public-facing transparency and stakeholder consultations.[57] However, its regulation of moderation is primarily characterised by a judicial approach: it aims to ensure decisions in particular cases follow applicable rules, include reasoned explanations and are open to challenge by affected individuals.[58]

Some provisions also address substantive moderation policies and practices, notably Article 14(4) (requiring platforms to have regard to users' fundamental rights) and Articles 34-35 (requiring the largest platforms to assess and mitigate systemic risks). These are considered in detail in section IV. However, the procedural protections set out in Articles 14-21 should be understood as the DSA's primary safeguard against arbitrary and unaccountable content moderation. This is true for several reasons. First, they create far more detailed, specific and stringent obligations[59] than the systemic risk provisions, which are far more abstract, vague and flexible,[60] and the rather ambiguous mandate to have 'due regard to' fundamental

[54] Edoardo Celeste, 'Digital constitutionalism: a new systematic theorisation' (2019) 33(1) *International Review of Law, Computers & Technology* 76.

[55] Van Loo (n 53).

[56] Douek, 'Systems Thinking' (n 53).

[57] See e.g. Arts 15, 40, 42 (on transparency), Recital 90 and Arts 35(3) and 45 (on consultations), Digital Services Act (n 6).

[58] Douek, 'Systems Thinking' (n 53).

[59] Daphne Keller, 'The DSA's Industrial Model for Content Moderation' (*Verfassungsblog*, 24 February 2022) <https://verfassungsblog.de/dsa-industrial-model/> accessed 24 December 2022.

[60] Article 34(1) mentions nine risk areas, several of which (e.g. 'fundamental rights' and 'public security') are extremely broad and open to interpretation.

rights in Article 14(4).[61] Second, insofar as risk management involves changes to content moderation systems (mentioned as one potentially-relevant type of mitigation measure in Article 35(1)), certain aspects of the provisions themselves and their interpretation thus far suggest that this will focus more heavily on removing more content deemed harmful than on addressing risks of over-removal.[62] Finally, the systemic risk provisions only apply to designated 'very large online platforms' – those with over 45 million EU users[63] – while the procedural protections apply to all online platforms (with some exceptions for micro and small enterprises[64]). At its core, then, the EU's regulatory framework for content moderation is focused on procedural fairness, with substantive changes playing a more minor role.

---

[61] See Naomi Appelman, João Pedro Quintais and Ronan Fahy, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' (2023) 24(5) *German Law Journal* 881; Rachel Griffin, 'Rethinking rights in social media governance: human rights, ideology and inequality' (2023) 2(1) *European Law Open* 30.

[62] For example, 'dissemination of illegal content' is the first risk area mentioned in Article 35(1). The first model risk assessment published by the Commission, which focused on mitigating risks of Russian disinformation operations, also clearly suggested that more such content (which is not necessarily illegal) should be removed: European Commission, 'Digital Services Act study: Risk management framework for online disinformation campaigns' (European Commission, 30 August 2023) <https://digital-strategy.ec.europa.eu/en/library/digital-services-act-study-risk-management-framework-online-disinformation-campaigns> accessed 10 April 2024. In contrast, while over-removal and discriminatory moderation are certainly within the scope of Articles 34-35 (since they affect fundamental rights like freedom of expression and non-discrimination) they are not explicitly mentioned, suggesting they might be a lower priority.

[63] Platforms with over 45 million monthly active users in the EU can be designated as very large online platforms by the Commission, meaning they are subject to the DSA's risk mitigation and assessment obligations, as well as certain other additional obligations: see Art 33, and more generally Chapter III Section 5 of the Digital Services Act (n 6).

[64] Arts 15(2) and 19, Digital Services Act (n 6).

These DSA provisions build on earlier EU legislation which also introduced procedural protections as safeguards against excessive censorship by online platforms – notably the 2019 Copyright Directive (CD) and 2021 Terrorist Content Regulation (TCR), both of which require platforms to allow users to appeal content removals.[65] However, the DSA not only introduces more detailed procedures, but significantly expands their scope. The CD and TCR's appeals procedures apply to content which platforms are legally required to remove, whereas the DSA's rules apply to all moderation – whether the content is removed due to illegality, or under platforms' in-house content policies.

Thus, in this regulatory model, platforms are free to set their own policies regarding what content they allow and how it is promoted and organised. However, they must transparently explain to users – both in general terms in their published policies, and in specific cases – how these rules are applied, and allow users to challenge decisions as inconsistent with them. The substance of moderation remains up to the platforms; procedural protections aim to ensure users can understand and challenge decisions that may deviate from stated policies.

III A FEMINIST CRITIQUE OF PROCEDURAL FETISHISM

This section presents a three-level critique of the DSA's approach, characterising it in terms of a 'procedural fetishism' which places inadequate weight on systemic problems and substantive justice. First, informed by available empirical evidence relating to procedural rights in the context of

[65] Art 17(9), 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 (Text with EEA relevance) [2019] OJ L130/92 ('Copyright Directive'); Art 10, Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (Text with EEA relevance) [2021] OJ L172/79 ('Terrorist Content Regulation').

content moderation, it highlights several practical limitations. Second, more fundamentally, it argues a legal framework based on individual rights and formally fair policies is inherently incapable of addressing the most consequential issues in content moderation. Finally, it suggests proceduralism is not only ineffective, but might actively exacerbate unaccountability and bias. Not only could it divert resources from other interventions; fetishising 'procedure over substance'[66] could also legitimise current practices, entrenching systemically unfair approaches to platform governance.

1. *Practical deficiencies*

If we accept the normative basis of the DSA's approach – that is, assuming that legitimacy comes from treating individuals fairly according to clear and consistent rules, and empowering them to understand and contest decisions – this section argues that these procedural safeguards fail on their own terms. They cannot offer users meaningful, effective or equal protection, and will not significantly constrain arbitrary decision-making.

First, several factors suggest they may not be widely used in practice. In US copyright law, similar 'counter-notice' appeals processes have been available for decades, but studies consistently find they are rarely used, even where there appear to be high proportions of mistaken decisions.[67] Uptake may be

---

[66] Ortolani (n 42).

[67] Jennifer M. Urban and Laura Quilter, 'Efficient Process or Chilling Effects – Takedown Notices under Section 512 of the Digital Millennium Copyright Act' (2006) 22(4) *Santa Clara High Technology Law Journal* 621; Annemarie Bridy and Daphne Keller, 'U.S. Copyright Office Section 512 Study: Comments in Response to Second Notice of Inquiry' (*SSRN*, 2017) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2920871> accessed 2 September 2022; Jennifer M. Urban, Brianna L. Schofield and Joe Karaganis, 'Takedown in Two Worlds: An Empirical Analysis' (2017) 64 *Journal of the Copyright Society of the USA* 483; Alexandra Kuczerawy, 'From 'Notice and Takedown' to

significantly different in contexts other than copyright infringement; however, many of the factors these studies identify as discouraging appeals also have broader relevance. Challenging moderation decisions not only requires time, energy and motivation, but also quite a detailed understanding of the platform's content policies and the relevant legal framework. Empirical research shows that many people have little knowledge of social media platforms' content guidelines and appeals policies.[68] Even users aware of appeals procedures may be intimidated by needing to assert whether their content falls within legal or policy categories, such as copyright infringement or hate speech, which are not straightforward for non-experts.[69]

Abilities to utilise procedural protections will also be unequally distributed. Reading a platform's explanation for a decision, in combination with its moderation policies and the relevant legal provisions, and then deciding whether to appeal, requires time, informational resources and (digital) literacy skills. These generally track broader social and economic inequalities,[70] so more privileged users will be better able to challenge

"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).

[68] Tom Tyler and others, 'Social media governance: can social media companies motivate voluntary rule following behavior among their users?' (2022) 17 *Journal of Experimental Criminology* 109.

[69] Rachel Griffin, 'New school speech regulation as a regulatory strategy against hate speech on social media: The case of Germany's NetzDG' (2022) 46 *Telecommunications Policy* 102411; Kuczerawy (n 64).

[70] Matthew T. McCarthy, 'The big data divide and its consequences' (2016) 10(12) *Sociology Compass* 1131; Simeon J. Yates et al, 'Who are the limited users of digital systems and media? An examination of U.K. evidence' (2020) 25(7) *First Monday* <https://doi.org/10.5210/fm.v25i7.10847> accessed 2 September 2022; Kenny Jacoby, 'Facebook fed posts with violence and nudity to people with low digital literacy' (*USA Today*, 23 November 2021)

arbitrary moderation.[71] One US study also suggests women may be less likely than men to appeal.[72] It is difficult to generalise without more quantitative evidence, but anecdotal evidence from Germany suggests similar patterns may exist in Europe.[73]

Where people do use such protections, their practical utility is also questionable. Users may struggle to effectively argue against decisions. In a study simulating Facebook appeals, users were more likely to challenge the moderation system's overall goals as biased or pointless, or make general claims about their own character and motivations, than to offer concrete arguments that their posts did not violate Facebook's policies.[74] Since Article 20 does not require platforms to consider these broader criticisms, many appeals will likely be ineffective.[75]

More fundamentally, there is no simple way to define a mistaken, unfair or unfounded decision. Content policies written to govern platforms with

---

<https://eu.usatoday.com/story/tech/2021/11/23/facebook-posts-violence-nudity-algorithm/6240462001/> accessed 22 March 2022.

[71] Hoffmann (n 38).

[72] Jonathon W. Penney, 'Privacy and Legal Automation: The DMCA as a Case Study' (2019) 22(2) *Stanford Technology Law Review* 412.

[73] Daniel Holznagel, 'Enforcing the Rule of Law in Online Content Moderation: How European High Court decisions might invite reinterpretation of CDA §230' (*Business Law Today*, 9 December 2021) <https://businesslawtoday.org/2021/12/rule-of-law-in-online-content-moderation-european-high-court-decisions-reinterpretation-cda-section-230/> accessed 22 March 2022.

[74] Kristen Vaccaro, Christian Sandvig and Karrie Karahalios, '"At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation' Vol 4 CSCW2 Article 167 *Proceedings of the ACM on Human-Computer Interaction* 1.

[75] Such perspectives are arguably more relevant to the legitimacy of moderation systems than narrow policy-based arguments. As such, the inability of appeals procedures to facilitate these forms of contestation is another major limitation, discussed further in section III(2).

millions or billions of users, spanning diverse social and cultural contexts, are necessarily highly indeterminate.[76] For example, Facebook's hate speech policy (also applicable to Instagram[77]) includes 'harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal'.[78] These are fundamentally ambiguous, contestable and subjective categories. Consequently, there is no baseline of objectively 'correct' interpretations of such policies which can be used to identify and correct particular 'incorrect' decisions. In practice, this means platforms will have plenty of leeway to apply policies in biased and inconsistent ways without making decisions which are demonstrably incorrect or unfounded – meaning Article 20(4) DSA will not oblige them to change their decisions.

Relying on human review to correct errors in automated moderation is also over-optimistic. Empirical evidence indicates that humans are generally bad at identifying and correcting biased algorithmic decisions.[79] That will be particularly true in this context, given that moderators work under intense time pressure and follow highly standardised rulebooks which are not designed for nuanced consideration of individual cases.[80] Unsurprisingly, given this context, existing appeals procedures (implemented voluntarily by

[76] Paddy Leerssen, 'An End to Shadow Banning? Transparency rights in the Digital Services Act between content moderation and curation' (2023) 48 *Computer Law & Security Review* 105790.

[77] Instagram, 'Community Guidelines' (Instagram Help Centre, n.d.) <https://help.instagram.com/477434105621119> accessed 5 December 2023.

[78] Meta, 'Hate Speech' (Meta Transparency Centre, n.d.) < https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/ > accessed 5 December 2023.

[79] Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 *Computer Law & Security Review* 105681.

[80] Sana Ahmad and Maximilian Greb, 'Automating social media content moderation: implications for governance and labour discretion' (2022) 2(2) *Work in the Global Economy* 176; Oversight Board, 'Reclaiming Arabic words' (*Oversight Board*, 2022) https://www.oversightboard.com/decision/IG-2PJ00L4T/> accessed 20 January 2023.

major platforms) have been described as 'dysfunctional', involving seemingly arbitrary results and little meaningful communication.[81] While Article 20 requires 'appropriately qualified staff' to oversee appeals,[82] in light of what is currently known about content moderators' working conditions – and the lack of clear regulatory incentives in the DSA for platforms to significantly overhaul these labour processes – it should not be expected that staff will have the time and resources to carefully reconsider each decision.[83] As Mathieu Fasel notes, the terminology of 'appeals' is somewhat misleading when it merely refers to a review by the same institution, without any mechanisms to ensure the same mistakes are not repeated.[84]

Users can further appeal to independent institutions, which are set up to place most costs on platforms and be relatively attractive to users.[85] However, platforms are not bound to follow their decisions.[86] Presumably only particularly motivated and informed users will pursue appeals this far, but when they do, platforms could disagree with the external interpretation and refuse to change their decision. In any case, since much social media content is topical and time-sensitive, successfully demanding reinstatement after an appeals process, which is likely to take at least some weeks, will often be practically irrelevant.

---

[81] Carolina Are, '"Dysfunctional" Appeals and Failures of Algorithmic Justice in Instagram and TikTok Content Moderation' (2024) *Information, Communication & Society* <https://doi.org/10.1080/1369118X.2024.2396621>

[82] Art 20(6), Digital Services Act (n 6).

[83] Keller, 'Industrial Model' (n 59); Griffin & Stallman (n 49).

[84] Mathieu Fasel, 'Sanctions and appeals by social media companies – arbitrariness or adequate user protection?' (Law, AI & Regulation Conference, Rotterdam, June 2023).

[85] Daniel Holznagel, 'A Self-Regulatory Race to the Bottom through Art. 18 Digital Services Act' (*Verfassungsblog*, 16 March 2022) <https://verfassungsblog.de/a-self-regulatory-race-to-the-bottom-through-art-18-digital-services-act/> accessed 18 January 2023.

[86] Art 21(2), Digital Services Act (n 6).

Finally, enforcing compliance with these provisions may be challenging in practice – in particular regarding 'shadowbanning', demotion, geoblocking and other interventions which restrict visibility without removing content entirely.[87] These measures are expressly included in Articles 17 and 20. However, they are not always apparent to users, and even where users suspect visibility has been restricted, it is neither practically nor conceptually easy to demonstrate. Recommendation algorithms' inputs and outputs are hugely complex, incorporating many thousands of data points and ranking content differently for each potential audience member.[88] There is no default or correct level of visibility as a baseline for comparison, making any definition of demotion somewhat arbitrary.[89] Users cannot easily know or prove that their content 'should' have been more popular in a counterfactual situation where it was not demoted.[90]

Paddy Leerssen has argued that, in principle, Articles 17 and 20 can coherently be interpreted as applying where platforms intervene to change the default outcome of algorithmic recommendation systems, using discrete subsystems which apply policy-based criteria.[91] However, in practice this will not be simple. As Leerssen notes, how recommendation systems break down into stages is manipulable: to avoid triggering costly procedural obligations, platforms could simply integrate policy considerations leading

[87] These are both contested terms without clear definitions, and may overlap in practice: see Gillespie (n 14) and Leerssen, 'Shadow Banning' (n 76).

[88] Zhuoran Liu and others, 'Monolith: Real Time Recommendation System With Collisionless Embedding Table' (2022) arXiv <https://arxiv.org/pdf/2209.07663.pdf> accessed 18 January 2023.

[89] Daphne Keller, 'Amplification and Its Discontents' (2021) Knight First Amendment Institute Occasional Papers <https://knightcolumbia.org/content/amplification-and-its-discontents> accessed 2 September 2022; Gillespie (n 14).

[90] Arvind Narayanan, *Understanding Social Media Recommendation Algorithms* (Knight Foundation Essays & Scholarship, 9 March 2023) <https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms> accessed 5 December 2023.

[91] Leerssen, 'Shadow Banning' (n 76).

to demotion in earlier stages. Moreover, given the opacity of these processes, if platforms simply ignore Article 17 and do not notify users of visibility restrictions, this may not be noticed or challenged.

Overall, while the DSA's impacts in practice remain to be seen, the evidence reviewed here suggests that its procedural protections might not be very effective safeguards against arbitrary moderation. First, they may not be widely used, particularly among users from marginalised communities. This is not to say they will never be useful: particularly informed and motivated users, like journalists and (semi-)professional content creators, will probably find them valuable. However, second, even where users do utilise appeals, enforcing compliance may be challenging. And third, in any case, the inevitable indeterminacy of platforms' policies suggests that they will not significantly constrain platforms' moderation decisions.

2. Normative deficiencies

Conversely, if we ignore these practical limitations and assume procedural protections would be effectively, fairly and consistently enforced, the normative assumption that they will make content moderation more legitimate is still flawed. Feminist theorists have argued that in an unequal society, promises of fair and equal treatment are often implausible, but – crucially – that even if this were not the case, substantive equality still requires systemic and institutional reforms beyond 'fairly' applying the same rules to everyone.[92] These arguments are particularly salient in the social media context, because access to online media platforms is now essential to participate in many areas of social, cultural, economic and political life. That makes it particularly important to value substantively just outcomes, not only fair procedures.

[92] Iris Marion Young, *Justice and the Politics of Difference* (Princeton University Press 1990); Jenny E. Goldschmidt, 'New Perspectives on Equality: Towards Transformative Justice through the Disability Convention?' (2017) 35(1) *Nordic Journal of Human Rights* 1; Crenshaw (n 38); Hoffmann (n 38).

First, procedural protections fail to represent all relevant interests. Like the equivalent provisions in the CD and TCR,[93] Article 17 DSA only requires explanations for users whose content is removed. However, many others might be affected. Often, the user sharing content is not its original author, but the author's interests are still relevant: for example, journalists have an interest in readers being allowed to share their articles. Moreover, potentially millions of users have an interest in access to social media content.[94] Article 20 attempts to address this by allowing anyone to challenge decisions, including people who reported content which was not ultimately removed. Nonetheless, this does not adequately protect the interests of audiences or other interested parties. Typically they will not even know content has been moderated; if they do, the diffuse and collective nature of interests in access to information makes it less likely that any individual or group will appeal.

More generally, individual procedural protections fail to represent social and collective interests. Discriminatory censorship does not just harm particular users whose content is suppressed. It has broader implications for whose voices are heard in public discourse, who sees themselves represented online, and who benefits from economic and cultural opportunities. Young LGBTQIA+ people who turn to social media for support, advice and community are affected by platforms' decisions to permit only sanitised, 'family-friendly' representations of queer identity.[95] Moderation that disproportionately targets marginalised groups hinders political activism and organisation.[96]

Importantly, Article 86 permits appeals by NGOs and associations on behalf of user groups, and, indeed, requires platforms to handle these complaints as

[93] Art 17(9), Copyright Directive (n 65); Art 10, Terrorist Content Regulation (n 65).

[94] Daphne Keller, 'Facebook Filters, Fundamental Rights, and the CJEU's *Glawischnig-Piesczek* Ruling' (2020) 69(6) *GRUR International* 616.

[95] Waldman, 'Disorderly Content' (n 20); Southerton and others (n 27).

[96] Danielle Blunt and others, 'Deplatforming Sex: A roundtable conversation' (2021) 8(4) *Porn Studies* 420; Grison and Julliard (n 21).

a priority.[97] Such bodies could facilitate and aggregate complaints and strategically select key cases as a way to highlight and challenge collective issues facing marginalised groups. However, this raises questions about where the necessary resources will come from and which social groups will be best placed to make use of these procedures: often, relying on civil society organisations to represent stakeholder interests does not favour the most marginalised.[98]

Even allowing for the possibility of collective challenges, a more fundamental limitation of these procedures is their focus on individual decisions, affecting specific pieces of content. In large-scale moderation systems, the most salient questions are not about individual posts, but about how systems are designed, what objectives they are optimised to pursue, and – since errors are inevitable – which types of error are preferred.[99] Moderation decisions reflect broader social, institutional and technical structures, taking in everything from the staffing and working conditions of moderation teams to the datasets and benchmarks used in designing automated moderation software.[100] In particular, AI classifiers are essentially probabilistic:[101] they predict a probability that content matches existing data or categories, and can be calibrated to intervene once particular probability thresholds are met. These thresholds are set based on normative choices

[97] Art 86(2), Digital Services Act (n 6).

[98] Rachel Griffin, 'Public and private power in social media governance: multistakeholderism, the rule of law and democratic accountability' (2023) 14(1) *Transnational Legal Theory* 46.

[99] Douek, 'Systems Thinking' (n 53).

[100] Douek, 'Systems Thinking' (n 53); Monea (n 16).

[101] Mike Ananny, 'Probably Speech, Maybe Free: Toward a Probabilistic Understanding of Online Expression and Platform Governance' (2019) Knight First Amendment Institute: Essays and Scholarship <https://knightcolumbia.org/content/probably-speech-maybe-free-toward-a-probabilistic-understanding-of-online-expression-and-platform-governance> accessed 18 January 2023.

about how many false positives and false negatives and which types of errors and biases can be tolerated across the system as a whole.[102]

The implications of these choices cannot sensibly be understood at the individual level. A moderation system that is a few percentage points more likely to censor people of colour than white people is unfair, but it does not obviously involve particular 'incorrect' decisions. Empowering users to challenge moderation decisions they dislike, but not to contest higher-level decisions about how moderation systems are designed and operated, is aptly described as 'accountability theater rather than accountability itself'.[103] It signals to users (and other stakeholders) that their interests are being considered, without addressing the most pervasive forms of unfairness or restricting companies' freedom to design their moderation systems in ways that systemically disadvantage marginalised users.[104]

Since the DSA's procedural obligations operate at the level of individual platforms, they are particularly unsuited to addressing systemic issues spanning multiple platforms. One prominent example is the use of shared hash databases to coordinate automated removal, so that content identified by one platform can be automatically removed across all participating platforms.[105] This has most notably been used for terrorism-related content and child sexual abuse material, but it has also expanded to other areas,

[102] Douek, 'Systems Thinking' (n 53).

[103] Douek, 'Systems Thinking' (n 53), 533.

[104] See Tom R. Tyler & Kathleen M. McGraw, 'Ideology and the interpretation of personal experience: Procedural justice and political quiescence' (1986) 42(2) *Journal of Social Issues* 115.

[105] Evelyn Douek, 'The Rise of Content Cartels' (2020) Knight First Amendment Institute: Essays and Scholarship <https://knightcolumbia.org/content/the-rise-of-content-cartels> accessed 18 January 2023. Hashing generates a unique code for an image or video, and produces the same code again each time the hashing algorithm is run. This enables platforms to efficiently identify content that has already been moderated, by matching content to databases of existing hash codes, and remove it automatically. For a detailed explanation see Gorwa and others (n 32).

including the moderation of content which is not alleged to be illegal.[106] Users appealing to one platform that removes their content may thus be unable to challenge the ultimate rationale of this decision. Civil society groups have documented cases where users wrongly classed as posting terrorist content successfully appealed, only to see the content automatically removed again shortly afterwards.[107]

Furthermore, unjust outcomes do not only result from 'mistakes' where policies are incorrectly or inconsistently applied. As discussed above, often there is no single correct way to apply policies, which gives platforms plenty of latitude for inconsistent and self-interested decisions which are difficult to qualify as mistaken. However, even if this were not the case, feminist theory shows that substantive rules can have unequal and unjust impacts, even if they are consistently applied. In social media, a well-studied example is Facebook's 'real name policy'. Experts have repeatedly shown that it endangers LGBTQIA+ people, abuse victims, activists and other groups with particular needs to remain anonymous or use multiple identities, and that it leads to disproportionate censorship of such users for policy violations.[108] In such cases, 'fairly' applying the same policy to everyone will still systematically disadvantage marginalised groups.

---

[106] Kalhan Rosenblatt and Maya Eaglin, 'Meta teams up with Snap and TikTok to address self-harm content' (*NBC News*, 12 September 2024) <https://www.nbcnews.com/tech/social-media/meta-teams-snap-tiktok-address-self-harm-content-rcna170838> accessed 29 September 2024.

[107] WITNESS, 'Content Regulation in the Digital Age: Submission to the United Nations Human Rights Council Special Rapporteur for Freedom of Expression' (*OHCHR,* June 2018) <https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/ContentRegulation/Witness.pdf> accessed 18 January 2023.

[108] Rena Bivens, 'The gender binary will not be deprogrammed: Ten years of coding gender on Facebook' (2015) 19(6) *New Media & Society* 880; Oliver L. Haimson and Anna Lauren Hoffmann, 'Constructing and enforcing 'authentic' identity online:

Such users will not benefit from challenging decisions as inconsistent with platforms' stated policies: rather, they would benefit from being able to collectively contest the policies themselves and their underlying objectives. Corporate platforms' policies are primarily set up to pursue commercial goals – deflecting regulatory and reputational risks by taking action on issues constructed as security threats, like terrorist content[109] and disinformation,[110] and protecting advertisers' 'brand safety' from associations with negative content.[111] These objectives are not necessarily conducive to – and will often conflict with – creating open, inclusive and egalitarian online environments.[112]

For example, despite well-established criticisms of Facebook's real name policy, the company's moralistic commitment to 'authenticity'[113] has obvious commercial advantages: it ensures accounts can be linked to real consumers and to external sources of data about their buying behaviour.[114] Equally, addressing systemic biases in moderation would require expensive investments in staff and technical resources that commercial platforms simply are not incentivised to make, especially outside their core, most

Facebook, real names, and non-normative identities' (2016) 21(6) *First Monday* <http://dx.doi.org/10.5210/fm.v21i6.6791>; Soraya Chemaly, 'Demographics, Design, and Free Speech: How Demographics Have Produced Social Media Optimized for Abuse and the Silencing of Marginalized Voices' in Susan J. Brison and Katherine Gelber (eds) *Free Speech in the Digital Age* (Oxford University Press 2019).

[109] Marguerite Borelli, 'Social media corporations as actors of counter-terrorism' (2023) 25(11) *New Media & Society* 2877.

[110] 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.

[111] Griffin, 'Brand safety' (n 24).

[112] Are and Paasonen (n 20); Roberts (n 28); Griffin, 'The Sanitised Platform' (n 22).

[113] Haimson and Hoffmann (n 108).

[114] Bivens (n 108).

profitable markets (the US and Canada, and to a lesser extent, Europe[115]). Substantively fairer moderation calls for regulations that enable accountability and contestation regarding these higher-level choices and priorities. Procedural fairness norms cannot achieve this.

3. Disadvantages and risks

Not only is the DSA's proceduralist approach practically flawed and normatively unsatisfactory, it could also have unintended consequences. There is a real risk that its centrality in EU regulation and surrounding policy debates will substitute for more effective reforms, for two reasons. First, procedural fairness obligations are likely to divert resources away from alternative approaches within individual companies, regulators and the industry generally. Second, they may stabilise dominant companies' power by making them appear more legitimate.

On the first point, compliance with the DSA's procedural protections is expected to be resource-intensive for companies.[116] For platforms with large user bases, if even a small fraction of users appeal decisions under Article 20 – which requires supervision by human staff – this will entail substantial working time for moderators, and correspondingly large costs. Article 17's scope is even more enormous, since notices must be sent for every single moderation decision. While this could be done automatically, building and maintaining such automated notification systems will in and of itself require significant technical and human resources.

---

[115] Oversight Board, 'Policy advisory opinion on Meta's cross-check program' (*Oversight Board,* 2022), 31 <https://oversightboard.com/news/501654971916288-oversight-board-publishes-policy-advisory-opinion-on-meta-s-cross-check-program/> accessed 18 January 2023.

[116] Daphne Keller, 'The EU's new Digital Services Act and the Rest of the World' (*Verfassungsblog,* 7 November 2022) <https://verfassungsblog.de/dsa-rest-of-world/> accessed 18 January 2023; Keller, 'Industrial Model' (n 59).

An obvious possibility is that to compensate for these expenditures, platform companies will invest less in other trust and safety measures. Scholarship on inequality and bias in content moderation points to many substantive improvements that platform companies could make (and could be incentivised to make by regulation). This could, for example, include proactively addressing and mitigating algorithmic biases;[117] hiring more content moderators and improving their working conditions (e.g. by increasing pay and recognising labour unions), which would not only be desirable in itself but could lead to more careful and reliable moderation decisions;[118] introducing more quality control measures in moderation systems;[119] research into safer technological design;[120] and changing content policies shown to structurally disadvantage marginalised user groups.[121] These kinds of interventions might bring more benefit to marginalised users, but the DSA's focus on procedural fairness creates few regulatory incentives to invest in them and instead incentivises platform staff to prioritise reviewing individual decisions.[122] Thus, to the extent that procedural rights

[117] On the potential, limits and practical challenges of mitigating algorithmic biases, see Agathe Balayn and Seda Gürses, *Beyond Debiasing: Regulating AI and Its Inequalities* (2021) EDRi <https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/> accessed 18 January 2023.

[118] Investing more resources in content moderation and recognising it as a skilled profession has been advocated by representatives of moderators' unions: see e.g. Hendrix and others, 'Checking on the Progress of Content Moderators in Africa' (*Tech Policy Press*, 3 December 2023) <https://www.techpolicy.press/checking-on-the-progress-of-content-moderators-in-africa/> accessed 26 September 2024.

[119] Douek, 'Systems Thinking' (n 53); Griffin & Stallman (n 49).

[120] Nicolas Suzor and others, 'Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online' (2019) 11(1) *Policy & Internet* 84.

[121] Waldman, 'Disorderly Content' (n 20).

[122] Articles 14-21 apply to all except the very smallest platform companies (see section II). However, the companies which own some of today's largest social media platforms (e.g. Meta-owned Facebook and Instagram, Microsoft-owned LinkedIn,

disproportionately benefit more privileged users, their distributional effects will be regressive.

---

or Google-owned YouTube) are some of the largest and wealthiest in the world. It might therefore be argued that in these particular cases, resource constraints are irrelevant. In practice, however, it is clear that even these companies have other strong incentives – such as shareholder value maximisation – to minimise expenditures on moderation, except where they face strong regulatory, commercial or reputational incentives to invest more in specific areas. For example, in 2023 several 'big tech' companies laid off large numbers of staff to cut costs, resulting in noticeable share price increases: see Subrat Patnaik and Ryan Vlastelica, 'Big Tech's Job Cuts Spur Rallies Even as an Economic Slowdown Looms' (*Bloomberg*, 25 January 2023) <https://www.bloomberg.com/news/articles/2023-01-25/big-tech-s-job-cuts-spur-rallies-even-as-economic-slowdown-looms> accessed 26 September 2024. More generally, reporting has consistently documented how regulatory compliance and 'trust and safety' teams at VLOPs are perennially understaffed and have to carefully prioritise resource allocation: see e.g. Julia Carrie Wong, 'How Facebook let fake engagement distort global politics: a whistleblower's account' (*Guardian*, 12 April 2021) <https://www.theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang> accessed 18 January 2023; Justin Scheck, Newley Purnell and Jeff Horwitz, 'Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show' (*Wall Street Journal*, 16 September 2021) <https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953> accessed 18 January 2023; Donie O'Sullivan, Clare Duffy and Brian Fung, 'Ex-Twitter exec blows the whistle, alleging reckless and negligent cybersecurity policies' (*CNN*, 23 August 2022) <https://edition.cnn.com/2022/08/23/tech/twitter-whistleblower-peiter-zatko-security/index.html> accessed 18 January 2023; Jason Koebler, 'Where Facebook's AI Slop Comes From' (*404 Media*, 6 August 2024) <https://www.404media.co/where-facebooks-ai-slop-comes-from/> accessed 26 September 2024. Thus, even in the case of VLOPs, it should be expected that if they are forced to invest more some areas of trust and safety to comply with legal obligations (like the DSA's procedural safeguards), this may lead them to compensate by investing less in other areas.

Regulatory agencies overseeing DSA compliance will also face resource constraints.[123] The DSA's procedural obligations are just one aspect of a broader regulatory framework; as the following section discusses, it also includes provisions regulating content moderation at a more systemic level, as well as other important issues like research data access, ad targeting, and recommendations.[124] However, the procedural obligations are among the most detailed, specific and concrete provisions. Overseeing and enforcing them will be resource-intensive for regulators,[125] given the scale of commercial moderation systems and the complex technical systems, rules and policies, and business processes involved.[126] This will leave less capacity for proactive investigations and oversight of other aspects of the DSA – including those with greater potential to address systemic issues and substantive inequalities, like the risk management obligations discussed in section IV.

Finally, procedural fairness obligations will have distributional effects across the social media industry as a whole. As mentioned above, compliance with Articles 14-21 will be burdensome for smaller companies. Daphne Keller, a leading expert on platform regulation with years of industry experience as an associate general counsel for Google, has suggested that this could threaten the financial viability and scalability of smaller and medium-sized platforms.[127] In contrast, companies like Meta and Google already have many

[123] Husovec (n 42).

[124] See respectively Arts 40(4), 26(3) and 27, Digital Services Act (n 6).

[125] Thierry Breton, 'Sneak peek: how the Commission will enforce the DSA & DMA' (LinkedIn, 5 July 2022) <https://www.linkedin.com/pulse/sneak-peek-how-commission-enforce-dsa-dma-thierry-breton/> accessed 18 January 2023.

[126] Platform companies frequently outsource moderation labour to independent contractors and third-party firms, so overseeing compliance along these supply chains introduces additional complexity: see Ahmad and Greb (n 80).

[127] Articles 15(2) and 19 DSA create exemptions for micro or small enterprises, defined as those with under 250 employees and annual turnover under €50 million or annual

elements of the required procedures in place, and have the resources to expand them. As such, the DSA could reinforce existing structural advantages favouring today's dominant platforms.[128] This would not only weaken their accountability by limiting consumer choice but could also hinder the emergence of other business models and approaches to moderation, which might better serve inclusion and equality.

At the same time, procedural fetishism could strengthen currently-dominant platform companies by making them appear more legitimate to policymakers and the public. Legitimacy – both actual and perceived – is a recurring theme in the academic literature on procedural fairness in platform governance.[129] Advocates of proceduralism argue that there are no correct solutions to questions about online speech regulation, which inevitably involve competing interests and values, so widely-accepted procedural fairness norms offer the best path to legitimacy.[130] Moreover, this approach offers a regulatory strategy that can achieve some degree of consensus,[131] especially since it enjoys wide support from civil society.[132] Advocates also

balance sheet total under €43 million (referring to Art 2, Annex 1, Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises). However, Keller suggests these thresholds are too low to address the DSA's anticompetitive effects. Medium-sized and rapidly-scaling platforms, which have the greatest potential to present alternatives to today's dominant platforms, could still face significant barriers. See Keller, 'Industrial Model' (n 51); Keller, 'Rest of the World' (n 116).

[128] Néstor Duch-Brown, 'The Competitive Landscape of Online Platforms' (2017) JRC Digital Economy Working Paper 2017-04 <https://joint-research-centre.ec.europa.eu/system/files/2017-06/jrc106299.pdf> accessed 18 January 2023.

[129] Evelyn Douek, 'The Meta Oversight Board and the Empty Promise of Legitimacy' (2024) 37(2) *Harvard Journal of Law & Technology* 373.

[130] Van Loo (n 53); Klausa (n 53).

[131] Klausa (n 53).

[132] 'The Santa Clara Principles on Transparency and Accountability in Content Moderation' (*Santa Clara Principles*, 2021) <https://santaclaraprinciples.org/> accessed 2 September 2022.

draw on psychological research showing that people generally strongly value procedural fairness and will see even unfavourable decisions as more legitimate if they get an explanation and a hearing.[133] Evidence for this in the social media context is mixed. One study found that Twitter users who perceived moderation processes as procedurally fair considered them more legitimate and were less likely to violate rules again,[134] whereas another study found that appeals processes actually decrease perceived legitimacy, probably because users find them frustrating and do not feel their input is taken seriously.[135]

More importantly, however, increasing perceived legitimacy is not necessarily positive. Scholarship on social media governance has often raised concerns about 'accountability theater',[136] describing various rhetorical strategies and superficial reforms that companies use to deflect criticism and regulatory and reputational risks without actually changing harmful business practices or meaningfully strengthening accountability.[137] In turn, this could undermine arguments for further regulatory interventions aiming to address unfair and discriminatory moderation.[138] Through this lens, insofar as it increases platform companies' perceived legitimacy, the DSA's emphasis on

[133] Van Loo (n 53), citing Tom Tyler, 'The Psychological Consequences of Judicial Procedures: Implications for Civil Commitment Hearings' (1992) 46(2) *SMU Law Review* 433.

[134] Tyler and others (n 68).

[135] Vaccaro and others (n 74).

[136] Douek, 'Systems Thinking' (n 53), 533.

[137] Thomas Kadri, 'Juridical Discourse for Platforms' (2022) 136 *Harvard Law Review Forum* 163; Josh Cowls and others, 'Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance' (2022) 26(5) *New Media & Society* 2448; Moritz Schramm, 'Emulated Guardians—Can the Oversight Board and the DSA's Out-of-Court Dispute Settlement Bodies Control Platform Power?' (PlatGov Research Network Conference, virtual, April 2023); Zalnieriute, 'Against Procedural Fetishism' (n 17); Appelman and others (n 61).

[138] Zalnieriute, 'Against Procedural Fetishism' (n 17).

procedural fairness could be seen as not only ineffective but potentially counterproductive.

Notably, leading scholar on the psychology of procedural justice (and co-author of the aforementioned Twitter study) Tom Tyler has argued that procedural fairness norms can induce 'political quiescence', socialising people into accepting injustice by giving the (misleading) impression that they have meaningful input.[139] Similar effects can operate at the institutional level. Sociolegal research shows how companies can use procedure to perform regulatory compliance without meaningfully changing business practices; this performance can, in turn, influence how regulators and the public understand the goals of regulation, with correct procedures ultimately elevated over substantive improvements.[140] In the context of content moderation, as noted at the start of this section, there exist many possible ways to improve moderation systems and address systemic inequalities, which platforms have largely not implemented so far because they are not commercially appealing. Regulations could force them to invest in such measures, but by focusing exclusively on 'fair' treatment of individual users according to existing policies, Articles 14-21 DSA suggest that further reforms are unnecessary.

This reinforces the relevance of Zalnieriute's account of procedural fetishism. Highlighting political barriers to more substantive platform regulation in the US and EU (notably economic pressures, and geopolitical concerns about the need for globally competitive tech businesses), she suggests that procedural fetishism 'limits pressure for stricter laws by convincing citizens and institutions that their interests are sufficiently protected without inquiring into the substantive legality of corporate

[139] Tyler & McGraw (n 104).

[140] Laura Edelman, *Working Law: Courts, Corporations and Symbolic Civil Rights* (University of Chicago Press 2016); Ari Ezra Waldman, 'Privacy Law's False Promise' (2020) 97(3) *Washington University Law Review* 773; Ari Ezra Waldman, 'Privacy, Practice and Performance' (2022) 110 *California Law Review* 1221.

practices'.[141] Similarly, in the regulation of state surveillance, she argues that focusing on procedure over substance 'strengthens the negotiating position of law enforcement agencies and governments…by affirming the *prima facie* legality of mass surveillance'.[142] The concept of procedural fetishism thus calls attention not only to what the DSA does *not* do – effectively address substantive inequalities – but also to what it *does* achieve: stabilising existing business practices and governance structures by providing a veneer of legitimacy and diverting attention from more substantive reforms.

As such, despite promising consensus, focusing on procedure over substance is far from politically neutral.[143] The DSA's procedural fetishism affirms the legitimacy of markets and business interests as the primary forces shaping online content governance. Corporate platforms can determine what content to allow, how their policies are enforced and how they organise and promote content. Procedural rights simply promise users fair access to these market services, by allowing them to enforce platforms' stated policies. As shown above, these rights in practice fail to offer equal protection and are structurally incapable of addressing systemic biases and inequalities. Even more fundamentally, however, they assume the legitimacy of a marketised system of online speech governance, where moderation systems operate within certain regulatory constraints but are ultimately primarily shaped by the business imperatives of platforms and their advertising clients.

IV Beyond procedural fetishism

There are some aspects of the DSA that go beyond the procedural fetishist approach and could offer avenues for more systemic improvements. First, transparency provisions offer promising ways to identify and address problems through independent public scrutiny. Second, certain provisions could be interpreted as mandating platforms to address systemic issues and

[141] Zalnieriute, 'Against Procedural Fetishism' (n 17), 249.

[142] Zalnieriute, 'Mass Surveillance' (n 17).

[143] Zalnieriute, 'Against Procedural Fetishism' (n 17).

structural injustice, although how effectively they will achieve that in practice remains uncertain. Finally, recent ECJ case law acknowledges (some) limitations of procedural protections and establishes that EU fundamental rights law, in some circumstances, requires more substantive restrictions on content moderation.

1. Transparency

Articles 14 and 17 DSA attempt to strengthen transparency towards individual users. As section III(2) showed, procedures for contesting individual decisions are fundamentally unsuited to addressing systemic issues. However, the DSA also contains several provisions aiming to strengthen transparency towards researchers, journalists, policymakers and the public about the overall functioning of moderation systems.[144] These provisions aim to strengthen accountability via public scrutiny of platforms' higher-level decisions and business strategies. Improving researchers' and policymakers' understanding of these systems could also enable more effective regulatory interventions in the future.[145]

Article 15 DSA requires all intermediary services to publish yearly reports detailing how much content they remove following reports from users, trusted flaggers and state institutions. They must also include:

> meaningful and comprehensible information about the content moderation engaged in at the providers' own initiative, including the use of automated tools, the measures taken to provide training and assistance to persons in charge of content moderation, the number and type of measures taken that affect the availability, visibility and accessibility of information provided by

[144] Paddy Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (2020) 11(2) *European Journal of Law and Technology* <https://ejlt.org/index.php/ejlt/article/view/786> accessed 18 January 2023.

[145] Douek, 'Systems Thinking' (n 53).

> the recipients of the service and the recipients' ability to provide information through the service, and other related restrictions.[146]

Additionally, under Article 24(5), all statements of reasons for moderation decisions that online platforms send to users (as required by Article 17) must also be uploaded to a searchable public database maintained by the Commission.[147]

While information about technical tools and moderation processes could be useful to researchers, these obligations may not provide much meaningful information about disparate outcomes, since the data involved is generally highly aggregated and/or standardised and does not indicate details about specific content.[148] For example, the number of posts removed as hate speech would include an unknown proportion of non-hateful content removed by mistake, but would not reveal how much hate speech was not identified and remains online. Further, it would not indicate which types of speech fall into each category: for example, whether certain types of hate speech are more likely to be overlooked than others, or what kinds of speech are most likely to be mistakenly removed.[149]

However, under Article 40(4), 'very large online platforms' (those with over 45 million EU users) must also provide internal data on request to researchers vetted by national regulators. This is an important step because it allows

---

[146] Art 15(1)(c), Digital Services Act (n 6).

[147] 'DSA Transparency Database' (European Commission, 2024) <https://transparency.dsa.ec.europa.eu/> accessed 29 September 2024.

[148] Rishabh Kaushal and others, 'Automated Transparency: A Legal and Empirical Analysis of the Digital Services Act Transparency Database´(2024) *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1121–1132. Research has also raised questions about the accuracy and consistency of the database and transparency reports: Amaury Trujillo, Tiziano Fagni & Stefano Cresci, 'The DSA Transparency Database: Auditing Self-reported Moderation Actions by Social Media' (2024) arXiv <https://arxiv.org/html/2312.10269v3> accessed 29 September 2024.

[149] Ananny (n 101).

researchers to request whatever data they need to investigate particular issues, rather than relying on the predefined, aggregated categories in transparency reports. Questions remain about how easy data access will be in practice,[150] but over the medium to long term, such research will undoubtedly significantly improve public understanding of content moderation.[151] Public criticism substantiated by independent research could effectively pressure platforms to better address systemic issues.

Nonetheless, insofar as regulation focuses on making platforms' activities more transparent instead of substantively regulating them,[152] it is open to some of the same critiques as procedural protections. Zalnieriute sees both as 'part of a wider phenomenon of procedural fetishism', arguing that transparency reporting is a favoured solution of platform companies because it can strengthen perceived legitimacy without requiring substantive change.[153] Better understanding injustice in content moderation is one step towards redressing it, but overemphasising transparency and research can misleadingly suggest that if these problems were better understood, consensus solutions would present themselves. In fact, addressing them would require large investments – for example, because they would involve significant labour time, including from highly-skilled software engineers – and/or structural reforms of the social media market that would likely meet significant resistance.

2. Due diligence obligations

The DSA also includes some provisions which could be interpreted as mandating systemic and structural changes to moderation systems.

---

[150] Paddy Leerssen, 'Platform research access in Article 31 of the Digital Services Act' (*Verfassungsblog*, 7 September 2021) <https://verfassungsblog.de/power-dsa-dma-14/> accessed 18 January 2023.

[151] Mathias Vermeulen, 'Researcher Access to Platform Data: European Developments' (2022) 1(4) *Journal of Online Trust & Safety* <https://doi.org/10.54501/jots.v1i4.84>.

[152] Leerssen, 'Platform research access' (n 150).

[153] Zalnieriute, 'Transparency-Washing' (n 17), 151.

However, the relevant provisions are fairly abstract and open-ended, making it difficult to predict their effects in practice.

First, Article 14(4) provides that platforms must not only publish clear and transparent content policies, but also apply them in a 'diligent, objective and proportionate manner' with 'due regard to the rights and legitimate interests of all parties involved'. This includes all fundamental rights in the EU Charter, including non-discrimination, so issues around biased moderation could be within scope. Moreover, Article 14(4) specifically mentions media freedom and pluralism, which suggests that it is intended to encompass more structural and collective issues, not only individual rights.

Whether Article 14(4) creates enforceable individual rights remains unclear, but this is a possibility, since the DSA is a regulation and thus directly effective in national courts.[154] Undoubtedly, national regulators can investigate and enforce compliance (as can the Commission, for very large online platforms[155]). Furthermore, Article 53 provides that individuals and associations can complain to national regulators about non-compliance with the DSA. Thus, Article 14(4) could enable both individual and collective challenges to discriminatory moderation. For example, it could be argued that operating automated moderation systems which are probabilistically biased against people of colour does not have due regard to fundamental rights and thus violates Article 14(4).

How this will play out in practice remains to be seen. The relevant Charter rights and the obligation to have regard to them are expressed in vague and abstract terms and could be open to many interpretations.[156] Almost any decision about content moderation involves multiple, competing rights, which could reasonably be balanced in different ways.[157] As such, Article

[154] Appelman and others (n 61).

[155] Art 56, Digital Services Act (n 6).

[156] Griffin, 'Rethinking Rights' (n 61).

[157] Evelyn Douek, 'The Limits of International Law in Content Moderation' (2021) 6 *UC Irvine Journal of International, Transnational and Comparative Law* 37, 44.

14(4) may not meaningfully constrain platforms' decisions, as they could generally make a case for their preferred interpretation. This will be particularly true if having 'due regard to' rights is interpreted in procedural terms: that is, as requiring platforms to show that they considered relevant rights in decision-making processes rather than substantively prohibiting policies and practices which do not balance rights appropriately. National regulators, which have primary responsibility for enforcing Article 14,[158] may approach this question differently. While the European Board for Digital Services (established under Article 62 DSA) will provide a forum for them to discuss and align on such interpretative questions, with most large social media platforms based in Ireland, the view of the Irish Media Commission will ultimately be most consequential.

Finally, as mentioned in the introduction, the DSA creates much more extensive due diligence obligations for designated 'very large online platforms'.[159] Here, the focus is not on individual users but on 'systemic risks' to collective interests, including, for example, public health, electoral integrity and fundamental rights.[160] Article 34 requires platforms to conduct regular risk assessments focusing on these areas. Under the heading of fundamental rights, it explicitly mentions freedom of expression and information, non-discrimination, and media pluralism – again suggesting that fundamental rights should here be understood as encompassing structural conditions and collective values.[161] Article 35 requires platforms to take 'reasonable, effective and proportionate' measures to mitigate identified risks.[162] An indicative list of possible measures encompasses various areas beyond content moderation, such as advertising and recommendation systems, as well as addressing moderation at a systemic level:

[158] Art 56, Digital Services Act (n 6).

[159] Chapter III Section 5, Digital Services Act (n 6).

[160] Art 34(1), Digital Services Act (n 6).

[161] Art 34(1)(b), Digital Services Act (n 6).

[162] Art 35(1), Digital Services Act (n 6).

> adapting content moderation processes, including the speed and quality of processing notices related to specific types of illegal content and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence, as well as adapting any relevant decision-making processes and dedicated resources for content moderation'.[163]

References to expeditious removal and illegal content suggest that the envisaged adaptations to content moderation systems would primarily involve ensuring more content is removed, as opposed to minimising mistaken removals or disparate impacts. Nonetheless, efforts in the latter areas could certainly also be within the scope of this provision. Article 35(1) also mentions 'testing and adapting…algorithmic systems', which could, for example, encompass efforts to reduce algorithmic bias.

Again, much will ultimately depend on interpretation and enforcement. That compliance with Article 35 could, in principle, involve addressing systemic biases does not necessarily mean it will actually be implemented in this way.[164] Importantly, responsibility for identifying and mitigating risks rests primarily with platforms themselves, and secondarily with independent auditors, who must review their risk assessments and mitigation measures.[165] The Commission only plays a general oversight role – which, given the resources and technical expertise demanded by other parts of the DSA, not to mention the rest of the EU's Digital Single Market legislation,[166] may ultimately be rather hands-off.[167] Existing literature suggests that delegating regulatory interpretation to private companies via compliance documentation, risk assessment and auditing obligations often shifts focus away from systemic issues requiring significant reform, and towards risks

[163] Art 35(1)(c), Digital Services Act (n 6).

[164] Griffin, 'Rethinking Rights' (n 61).

[165] Art 37, Digital Services Act (n 6).

[166] Breton, 'Sneak peek' (n 125).

[167] The Commission has sole responsibility for enforcing systemic risk obligations: see Art 56(2), Digital Services Act (n 6).

that threaten industry interests and demand only superficial adjustments.[168] Since this is a novel field lacking widely-accepted standards on risk assessment, compliance and auditing methodologies, the DSA's auditing system may be particularly susceptible to 'capture' by platforms' interests.[169]

Much will depend on the regulatory strategy adopted by the Commission, which could significantly shape the interpretation of Articles 34-35 and could put pressure on platforms to implement more systemic improvements to content moderation. For example, it can issue guidance on how it will interpret obligations and evaluate compliance,[170] and can help develop codes of conduct setting out best practices for compliance.[171]

As mentioned in section III, however, early signs suggest the Commission's enforcement strategy may be oriented towards pressuring platforms to remove more content deemed harmful, rather than addressing risks of over-removal or discriminatory censorship. For example, its first model risk assessment[172] and interpretative guidance[173] both focused on risks associated with political disinformation, clearly implying that a primary risk mitigation measure should be removing more such content. Such regulatory pressure on platforms to remove more content deemed harmful by public authorities

---

[168] Julie E. Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019); Waldman, 'False Promise' (n 140); Waldman, 'Performance' (n 140).

[169] Johann Laux, Sandra Wachter and Brent Mittelstadt, 'Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA' (2021) 43 *Computer Law & Security Review* 105613.

[170] Art 35(3), Digital Services Act (n 6).

[171] Art 50, Digital Services Act (n 6).

[172] European Commission, 'Risk management framework' (n 62).

[173] European Commission, 'Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes' (European Commission, 26 March 2024) <https://digital-strategy.ec.europa.eu/en/library/guidelines-providers-vlops-and-vloses-mitigation-systemic-risks-electoral-processes> accessed 10 April 2024.

raises evident risks of deliberate political influence and repression, as well as inadvertent over-removal – both of which are likely to most heavily affect marginalised and stigmatised social groups.[174]

In this context, independent research and advocacy could also influence how policymakers and industry actors understand systemic risks and appropriate mitigation measures. Ideally, they would encourage a more holistic approach to regulatory enforcement – one which emphasises the risks of censorship, discrimination and overinclusive moderation and the need to mitigate these risks through systemic improvements to moderation rather than correction of individual 'errors'.[175]

3. Ex ante fundamental rights protection

Some interesting indications of how platforms' fundamental rights obligations could be interpreted, which could also encourage proactive approaches to addressing inequality, were recently provided by the ECJ decision in *Poland v Parliament and Council*,[176] an unsuccessful fundamental rights challenge to the obligations in Article 17 CD for platforms to automatically filter copyright-infringing content. Advocate General Øe's Opinion – which was largely followed by the ECJ – discusses at some length the capacity of notification and appeals procedures to protect users' rights, acknowledging some – though not all – of the limitations discussed in section III(1). He clearly acknowledges that appeals procedures are insufficient to protect freedom of expression and information: realistically, many users affected by overinclusive filtering will not appeal, and even

[174] Griffin, 'The Sanitised Platform' (n 22).

[175] Husovec (n 42).

[176] Case C-401/19 *Poland v Parliament and Council* (ECJ, 26 April 2022).

successful appeals may be too late for the reinstated content to have its intended impact.[177]

In Advocate General Øe's interpretation, protecting fundamental rights requires mistaken removals to be strictly minimised *ex ante*, if not entirely eliminated. Platforms should, therefore, only be required to block content that 'manifestly' infringes copyright.[178] The judgment did not use the word 'manifest', but largely followed this approach, ruling that filtering systems should not be used where they 'might not distinguish adequately between unlawful content and lawful content, with the result that [their] introduction could lead to the blocking of lawful communications',[179] or where determining infringement requires independent (manual) assessment.[180] Neither judgment nor opinion details how to ensure this, instead stating that member states should provide clear safeguards in their implementing legislation.[181] So far, a minority of member states (notably Germany) have included clear *ex ante* restrictions on automated filtering;[182] whether other

[177] He does not address inequalities between users, or the failure to represent audiences and other collective interests – even though freedom of information, more than other fundamental rights, is classically understood as a collective or social interest: Sarah Eskens, Natali Helberger and Judith Moeller, 'Challenged by news personalisation: five perspectives on the right to receive information' (2017) 9(2) *Journal of Media Law* 259.

[178] *Poland* (n 176), Opinion of AG Øe, para 201.

[179] *Poland* (n 176), para 86.

[180] *Poland* (n 176), para 90.

[181] *Poland* (n 176), paras 96-99, and Opinion of AG Øe, paras 193 and 209-13.

[182] Felix Reda and Paul Keller, 'CJEU upholds Article 17, but not in the form (most) Member States imagined' (*Kluwer Copyright Blog*, 28 April 2022) <http://copyrightblog.kluweriplaw.com/2022/04/28/cjeu-upholds-article-17-but-not-in-the-form-most-member-states-imagined/> accessed 28 June 2022.

member states must introduce further safeguards remains debated,[183] and there have not so far been indications that they will do so.

Nonetheless, the underlying principle established by *Poland* is significant. The ECJ held that procedural protections like notifications and appeals are insufficient to protect users' fundamental rights in the context of automated filtering. To prevent excessive censorship, such filtering must be strictly limited *ex ante*, to content so obviously illegal that it can reliably be identified automatically. Interpreting the judgment narrowly, these principles only apply to state-mandated filtering, which is restricted by the prohibition of general monitoring obligations in the 2000 E-Commerce Directive.[184] Platforms routinely engage in generalised, automated monitoring and filtering of user content on a (semi-)voluntary basis,[185] both for commercial reasons and to comply with self-regulatory commitments and due diligence obligations,[186] and this has never been considered to violate the prohibition of general monitoring obligations.

However, Articles 14 and 35 DSA make little sense if fundamental rights are understood only as constraints on state regulation: they clearly envisage fundamental rights as more generally-applicable guiding principles for platform companies. Accordingly, insofar as *Poland* provides guidance on the scope and meaning of users' fundamental rights in the context of social media moderation, it should be taken into account when interpreting

[183] Eleonora Rosati, 'What does the CJEU judgement in the Polish challenge to Article 17 (C-401/19) mean for the transposition and application of that provision?' (*The IPKat*, 11 May 2022) <https://ipkitten.blogspot.com/2022/05/what-does-cjeu-judgement-in-polish.html> accessed 28 June 2022.

[184] Art 6, 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 L178/1 ('E-Commerce Directive'). This provision is replaced but maintained essentially unchanged by Art 8, Digital Services Act (n 6).

[185] Gorwa and others (n 32).

[186] Borelli (n 109).

platforms' due diligence obligations. Importantly, the requirements the ECJ established for Article 17 CD might not be exactly applicable in other contexts: the decision is based on a proportionality assessment, and greater interferences with users' rights might be proportionate to deal with harms more severe than copyright infringement.[187] This could, for example, suggest that automated moderation with significant error rates should be permitted for seriously harmful illegal content like child sexual abuse material, hate speech and harassment, but not for things like nudity and 'brand unsafe' content. On the other hand, investments in mitigating algorithmic biases would be highly desirable from a fundamental rights perspective for all types of content – and indeed especially where the gravity of the harms involved means that automated tools with higher error rates are tolerated.

Poland, therefore, suggests that having regard to fundamental rights under Article 14 requires platforms to proactively invest in addressing systemic issues: for example, by implementing *ex ante* safeguards to minimise false positives and biases, rather than relying only on procedural protections, and possibly also by limiting automated filtering to types of content that can accurately and reliably be identified automatically. Similarly, Article 35 could be interpreted as mandating very large online platforms to take measures like these to address systemic risks to freedom of expression and non-discrimination.

V Conclusion

Overall, then, elements of the DSA could enable the kind of regulatory approach Evelyn Douek terms 'content moderation as systems thinking',[188] focused on designing sociotechnical systems that – to the extent possible – are reliable, treat people fairly and prioritise public interests over private

[187] Willemijn Kornelius, 'Prior filtering obligations after Case C-401/19: balancing the content moderation triangle' (2023) 14 *JIPITEC* 123.

[188] Douek, 'Systems Thinking' (n 53).

profit. Hopefully, as the institutional architecture of DSA enforcement develops, regulators and civil society will make use of these opportunities: for example, by using soft law tools like codes of conduct to push for systemic reform of moderation processes and taking collective action to challenge discriminatory systems as incompatible with fundamental rights.

However, there is a real risk that the DSA's individualistic and formalistic approach to procedural fairness will crowd out or overshadow such opportunities: by entrenching existing approaches to moderation, making dominant platform companies appear superficially more accountable, and diverting attention and resources away from more systemic reforms. These risks are well illustrated by the intensely relaxed attitude Breton publicly adopted towards Musk's acquisition of Twitter. The governance of a major piece of communications infrastructure by a super-rich individual based on his personal ideological views and interests is hardly less concerning because the company's decisions will comply with some procedural formalities.[189]

Procedural protections are equally inadequate to address systemic biases, unfair outcomes, and lack of accountability at other major platforms. A feminist analysis, drawing on the limited empirical evidence available and considering how the DSA's rules are likely to operate in practice in their social and institutional context, suggests that their impact will be limited and unequal. Feminist theory also points to more fundamental normative flaws in the assumption that procedural fairness towards individuals can achieve legitimacy in content governance. It fails to address the most consequential decisions about how moderation systems are designed and reinforces a liberal, marketised paradigm where moderation is primarily designed to serve business interests.

---

[189] Victor Pickard, 'When Oligarchs Control the Media: An Interview with Victor Pickard' (*Sciences Po Digital, Governance & Sovereignty Chair,* 29 April 2022) <https://www.sciencespo.fr/public/chaire-numerique/en/2022/04/29/when-oligarchs-control-the-media-an-interview-with-victor-pickard/> accessed 19 January 2023.

On this view, subjecting content moderation to procedural fairness norms does not automatically make it more legitimate. Rules, regulations and procedures are not inherently valuable. They should be evaluated based on how they can achieve substantive goals like making companies accountable for their most consequential business practices, goals and policies; preventing arbitrary and biased moderation, and redressing disadvantages currently faced by marginalised groups. Looking ahead, policymakers, researchers and civil society should endeavour to build on those aspects of the DSA framework which can best achieve these substantive changes, rather than fetishising procedure as an end in itself.

CURB YOUR ENTHUSIASM: WHY EUROPE'S DIGITAL REFORMS MAY NOT BECOME A GLOBAL STANDARD

Moritz A. Schramm *

The European Union is widely perceived and presents itself as the global vanguard in the struggle to regulate digital corporations. The Union's regulatory schemes, especially the Digital Services Act, are widely hailed as the continent's – and from the EU's perspective, the world's – best shot at taming digital capitalism. The EU designed many of those measures to become a 'global standard'. Yet, drawing from organization theory and a legal realist analysis of several of the key provisions of the DSA and their potential implementation, I claim that crucial parts of Europe's reforms will not become a global normative standard – or, if they do, in ways fundamentally different to what many would expect. That is for two reasons. First, while the DSA does establish a few concise and objective substantive standards, it also grants extensive discretion to private organizations. Second, if private actors will, as we must assume, exercise this discretion in an autonomous (some might say self-serving manner), many publicly acclaimed provisions of Europe's digital governance reforms may yield globalized private ordering carrying the legitimizing label of EU supervision. Consequently, some current European reforms may stabilize rather than constrain private power and diffuse, if at all, only European ceremonies and labels but not necessarily the full substance of EU law.

* Moritz A. Schramm is a Research Scholar and Adjunct Professor of Law at New York University. At NYU, Moritz is affiliated with Guarini Global Law & Tech, the Institute for International Law and Justice, and the Information Law Institute. This article is based on research conducted in the context of the author's PhD and monograph *Governance by Emulation: Platform Adjudication, the Oversight Board, and the Digital Services Act* (Cambridge University Press, forthcoming 2025). The author would like to thank the members and organizers of the European Society of International Law's 'International Law and Technology' Interest Group, especially Barrie Sander, and the team of the European Journal of Legal Studies for their generous and thoughtful support. All views are exclusively my own. All the usual disclaimers apply. This article was submitted in December 2023 and accepted in May 2024.

Keywords: Brussels Effect, Regulation, European Union, European Commission, Digital Services Act, DSA, Extraterritoriality, Private Ordering, Private Governance, Content Moderation, Meta, X, Instagram, Twitter, BlueSky, Organization Theory, Neoinstitutionalism, Organization Sociology

TABLE OF CONTENTS

I. INTRODUCTION

In recent years, we saw a frenzy of regulatory and "self"-regulatory experiments to advance accountability in digital governance. In general discourse, the European Union emerged as a global vanguard in that struggle. The Union's regulatory schemes in the field of privacy and, especially, its recently enacted Regulation (EU) 2022/2065, commonly known as the Digital Services Act (DSA), and Regulation (EU) 2024/1689, commonly known as the Artificial Intelligence Act (AIA), are widely hailed as our best shot at making data-intense social media platforms, search engines, and other remnants of the information economy more transparent, accountable, fundamental rights-oriented, and, of course, 'fairer'.[1] The European Union (EU) explicitly designed many of those measures, most notably the DSA, to become a 'global standard'. In that context, commentators, and EU officials alike reference Anu Bradford's hugely influential 'Brussels Effect'.[2] The Brussels Effect describes how the EU leverages its formidable market size and regulatory capability to exert global normative force over some products and industries.

[1] See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) and Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[2] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020); see already Anu Bradford, 'The Brussels Effect' (2012) 107 Northwestern University Law Review 1.

Yet, drawing from organization theory and a legal realist analysis of several of the DSA's key norms, I claim that crucial parts of Europe's reforms will *not* become a global normative standard – or, if they do, in ways fundamentally different to what many would expect. Rather than look at the economics and politics of policy diffusion, my argument focuses on how norms of the DSA will be implemented in practice. Here, I focus on the Digital Services Act (DSA), as its application of public law principles to private governance is particularly striking. However, similar observations can arguably be made about other EU tech regulations, especially in the areas of data protection and artificial intelligence.

Beyond the European Commission, national institutions, and eventually the European Court of Justice, a pivotal actor in the implementation of EU tech regulations is, somewhat ironically, the tech companies themselves. Their role is ensnared in a familiar yet intricate double bind. On the one hand, regulatory expectations like proportionality and respect for fundamental rights are lofty and steeped in the ethos of public law. On the other, these ideals must navigate the terrain of organizations governed not by civil servants but by executives, subject to the often-conflicting imperatives of corporate governance and securities law.

In this setting, the EU's high-minded demands might struggle to fully bloom. Managerial constraints, legally enshrined corporate autonomy, and the inevitable self-interest of for-profit enterprises can subtly reshape these public-law-inspired requirements. This dynamic could lead to a gradual reinterpretation of Europe's normative aspirations as they are filtered through the machinery of corporate implementation. Put differently, the downstream application of Europe's digital reforms—values such as due process, review, proportionality, and fundamental rights—will inevitably be influenced by the private, profit-driven nature of the organizations now tasked with upholding these principles. To their credit, these companies possess formidable bureaucratic and infrastructural capacities (after all, Meta

essentially pioneered large-scale content moderation). Yet, their overriding goal remains what it has always been: to function as profitable enterprises rather than as stewards of public law ideals.

This tension is neither shocking nor inherently problematic. Indeed, corporate autonomy and the right to conduct business are explicitly protected under the EU's Charter of Fundamental Rights. But for those invested in seeing European normative ideals resonate globally, it would be unwise to assume that corporate reinterpretations of Brussels' regulatory aspirations will automatically translate into a worldwide embrace of public-law classics within corporate frameworks.[3] In other words, this piece explores the challenges of implementing regulation—or, to borrow from Bradford, how the reality of the '*de facto* Brussels Effect' proves far more complex than the initial euphoria surrounding Europe's so-called 'digital constitution' might have anticipated.[4]

When we look closely at the norms stipulated in the DSA, two arguments undermine – or complicate – the widely held perception that the DSA becomes a 'global standard'.

[3] For another sceptical perspective on the DSA's global reach see Martin Husovec and Jennifer Urban, 'Will the DSA Have the Brussels Effect?' (*Verfassungsblog*, 21 February 2024) <https://perma.cc/DCX4-6FYK>. For alternative (and, at times, competing) analyses of global policy diffusion in general see esp. Eleanor Westney, *Imitation and Innovation - The Transfer of Western Organizational Patterns to Meiji Japan* (Harvard University Press 1987); Beth A Simmons, Frank Dobbin and Geoffrey Garrett (eds), *The Global Diffusion of Markets and Democracy* (Cambridge University Press 2008); Charles R Shipan and Craig Volden, 'The Mechanisms of Policy Diffusion' (2008) 52 American Journal of Political Science 840; Mark Lawrence Schrad, *The Power of Bad Ideas: Networks, Institutions, and the Global Prohibition Wave* (Oxford University Press 2010); Fabrizio Gilardi and Fabio Wasserfallen, 'The Politics of Policy Diffusion' (2019) 58 European Journal of Political Research 1245; critical of the Brussels Effect: Abraham L Newman and Elliot Posner, 'Putting the EU in Its Place: Policy Strategies and the Global Regulatory Context' (2015) 22 Journal of European Public Policy 1316.

[4] De facto Brussels Effect refers to companies adopting EU norms, de jure Brussels Effect refers to third states emulating EU legislation like the GDPR, see in detail regarding the former Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press 2023) 324 et seq; see in general already Bradford, *The Brussels Effect* (n 2) 142 et seq.

First, upon closer inspection, the DSA only establishes a handful of concise and substantive standards.[5] Mostly, the Union's regulatory technique seems to be to formulate legitimacy-carrying but broad and context-dependent normative goals like 'fairness', 'transparency', 'due regard for relevant interests such as fundamental rights', and 'proportionality'.[6] What these terms mean concretely remains unclear. Responsible for breathing life into those broad goals are the same actors whose practices triggered the regulation in the first place. We know that such 'supervised self-regulation' can lead to superficially rather than substantively satisfying results.[7] However, the DSA's overall thrust to pressure digital corporations into explaining and justifying their actions is in itself a big achievement that may very well positively shape corporate practices around the globe.[8] Second, private actors, as is natural given their corporate autonomy, will presumably exercise the discretion granted to them by the EU in ways aligned with their organizational priorities. This structural difference from public law idealism means that many of the publicly celebrated provisions of Europe's digital governance reforms might, in practice, function as a form of public

[5] Terms like standard, norm, rule, or regulation are understood broadly. Unless otherwise indicated I do not attach distinct meaning to them but use them interchangeably. That is because many legal theoretical distinctions like that between rules and standards in US American jurisprudence collapse once one takes the actual normative practice into account, which is necessarily individualized yet general. Therefore, the overarching term may be 'norm', which Christoph Möllers describes as a combination of a possibility and a 'realization marker'. In our case, most norms would further be explicit and stipulated through a pre-defined procedure by an organized authority (e.g., the EU). For the underlying praxeological conception of norms see Christoph Möllers, *The Possibility of Norms: Social Practice beyond Morals and Causes* (Oxford University Press 2020) 71 et seq.

[6] For example, the term 'fundamental right(s)' appears 39 times in the recitals and articles of the DSA as published in the official journal of the European Union, cf OJ L 277, 27.10.2022, p. 1–102.

[7] Perhaps the most interesting case study of the vices and virtues of self-regulation are financial services. For an overview (which rightly points out the informational advantage of self-regulatory regimes over classic top-down regulation) see Saule T Omarova, 'Rethinking the Future of Self-Regulation in the Financial Industry' (2010) 35 Brooklyn Journal of International Law 665. For a more recent yet particularly vivid example see Joanna R Schacter, 'Delegating Safety: Boeing and the Problem of Self-Regulation' (2021) 30 Cornell Journal of Law and Public Policy 637.

[8] See in that sense Bradford, *Digital Empires* (n 4) 337 et seq.

legitimization – serving as ceremonial 'certifications' that leave core corporate practices largely untouched.

These shortcomings may undermine the efficacy of many regulatory projects, especially in the field of private power and technology. Regulators and scholars must invest more effort to translate normative demands into organizationally and technologically feasible solutions. But how? The article concludes with two recommendations. On the one hand, when implementing the DSA, the Commission should strive for a better balance of normative specificity and normative broadness. This would significantly narrow companies' leeway when implementing regulatory demands. Only clearly defined and empirically testable norms enable meaningful reforms of technology companies. On the other hand, translating legislative ideas into (privately owned) technological structures is phenomenally difficult. To at least improve that translation, the EU should further include engineering perspectives into its lawmaking process. So far, the legislative discourse about the DSA and, to a lesser degree, even the AIA seem to be dominated by actors – lawyers, political scientists, etc. – who focus on how a specific technology or service *should* function (normative dimension) but less how to *achieve* this (sociological and technological dimension).

This article progresses as follows. It briefly outlines, first, the Brussels Effect's political saliency for the EU and, second, the key elements of the Brussels Effect as described in the literature. Third, the article discusses the regulatory approach permeating many current EU digital regulations: formulating broad normative goals and some outer procedural bounds to then let tech companies devise the specifics. Fourth, the article questions whether that approach undercuts at least one of the Brussels Effect's premises, namely that the EU enacts stringent standards which then reverberate around the globe. Fifth, in examining the DSA, the article draws on organization theory and prior – currently unpublished – empirical research to explore the complexities and ambivalence of organizational compliance efforts. These efforts may serve to enhance platforms' public legitimacy while leaving core

practices largely intact.[9] Sixth, the article looks at possible solutions, namely increased regulatory specificity and heightened independent technological expertise in lawmaking and enforcement.

II. WHY IS THE BRUSSELS EFFECT IMPORTANT?

In the past couple of years, the EU championed – or rather branded – a new way of thinking about its politics.[10] Even though the EU is a sprawling organization, yearslong political drag and media onslaught on its public legitimacy left Brussels craving for a new, fresh narrative. For more than ten years, there have been ongoing problems with stalled reforms, the Euro crisis, populism, and Brexit.[11] Many EU officials started thinking that the Union needed to reinvent itself. This time, however, it was not the Court of

[9] This unpublished empirical work refers to qualitative interviews conducted with EU officials, civil society activists, and executives of digital platforms as well as staffers and members of Meta's Oversight Board. That work was conducted for the monograph *Governance by Emulation: Platform Adjudication, the Oversight Board, and the Digital Services Act* (Cambridge University Press forthcoming 2025).

[10] Two disclaimers are warranted. Firstly, many criticisms directed at 'the EU' are exaggerated. The Council stands as the Union's primary political entity, with 'the EU' in the Council effectively representing the Member States. Consequently, Member States bear the primary responsibility for contentious policies, such as managing the Euro crisis, the dysfunctional migration regime, NextGenerationEU, or addressing authoritarian tendencies within Member States. Thus, a deeper understanding reveals that the situation may not be as dire as portrayed. However, secondly, this disclaimer underscores a fundamental issue in EU politics. If grasping the nuances of EU law is necessary to discern that much of the criticism aimed at the EU, and particularly most anti-EU rhetoric, is unfounded, it indicates a significant political challenge. The EU may struggle to garner (further) public legitimacy if it lacks, whether justifiably or not, the political capacity to attract broader public support.

[11] In chronological order: David Vogel, 'The New Politics of Risk Regulation in Europe' [2001] Centre for Analysis of Risk and Regulation, London School of Economics and Political Science; André Sapir, *Fragmented Power: Europe and the Global Economy* (Bruegel 2007); Elliot Posner, 'Making Rules for Global Finance: Transatlantic Regulatory Cooperation at the Turn of the Millennium' (2009) 63 International Organization 665; David Vogel, *The Politics of Precaution: Regulating Health, Safety, and Environmental Risks in Europe and the United States* (Princeton University Press 2012); Bradford, 'The Brussels Effect' (n 2); Joanne Scott, 'Extraterritoriality and Territorial Extension in EU Law' (2014) 63 American Journal of Comparative Law 87; Alasdair R Young, 'The European Union as a Global Regulator? Context and Comparison' (2015) 22 Journal of European Public Policy 1233; Newman and Posner (n 4).

Justice calling the shots, but the Union's regulatory machinery, combining lawmakers and agencies. Over time, these regulatory bodies have become powerful globally. The EU, in turn, incrementally became perhaps the most influential regulator on matters of global commerce. This potent regulatory posture is by now widely known as the Brussels Effect, as Anu Bradford coined it, nodding to the well-established California Effect in US literature on regulation and policy diffusion.[12] The Brussels Effect, Bradford argues, manifests either when non-member-states adopt the EU's regulatory *acquis* (usually for economic reasons) or when companies abide by that *acquis* globally, simply because it is more efficient to produce one globally sellable product than an extra product for the huge European market (or miss out on that market entirely). According to this idea, the EU has become important in setting rules about things like food safety, privacy, and now, free speech.[13] This phenomenon, especially the presumed moral high ground attached to fighting for online rights and against private, seemingly uncontrollable companies, provided a renewed sense of purpose to a Union grappling with uncertainty and introspection.[14] Recently, Bradford doubled down and argued that the DSA 'which was adopted in 2022, may further increase the EU's ability to shape tech companies' global business practices, and regulate the global digital economy in ways that the US and China are not able to do'.[15] Evidently, this is music to the ears of many working either for or on the European Union.

Bradford suggests that the Brussels Effect evolved gradually, initially arising as a consequence of internal market regulations before expanding into a broader external agenda.[16] Once again, news broke from across the

[12] Bradford, 'The Brussels Effect' (n 2); Bradford, *The Brussels Effect* (n 2).
[13] For the latter see esp. recently Bradford, *Digital Empires* (n 4).
[14] Bradford, 'The Brussels Effect' (n 2); Bradford, *The Brussels Effect* (n 2).
[15] Bradford, *Digital Empires* (n 4) 340.
[16] Cf Bradford, *The Brussels Effect* (n 2) 7 et seq, 25 et seq.

Atlantic.[17] Discussions surrounding the Brussels Effect appeared to instill in the EU a renewed sense of confidence, providing yet another source of output legitimacy through its (alleged) global regulatory reach.[18] Who needs, one might think, the sword or the purse if one can regulate worldwide commerce? EU institutions demonstrated adeptness in navigating intricate normative frameworks and political maneuvers to promote and protect 'European values'.[19]

Despite the theorem's discursive prevalence, scepticism prevailed. Numerous scholars cast doubt on the Union's capacity to wield normative influence across the globe, with critiques of the Brussels Effect becoming more pronounced as its underlying assumptions, empirically derived claims, and extrapolations from specific instances faced growing scrutiny.[20] Nonetheless,

[17] As it already was, famously, with Eric Stein, 'Lawyers, Judges, and the Making of a Transnational Constitution' (1981) 75 American Journal of International Law 1. Not all authors on the Brussels Effect are Americans or work in the US (e.g. Scott), but many (e.g. Vogel, Posner, Bradford) are/do.

[18] Many EU officials seem to have widely embraced the Brussels Effect. The Union's highest echelons – including the Commission President, the President of the Council, and the High Representative for Foreign Affairs – reference their global ambitions in speeches, talks, and press releases.

[19] Cf Fritz W Scharpf, *Demokratietheorie Zwischen Utopie Und Anpassung* (Universitätsverlag Konstanz 1970) 21 et seq, 66 et seq. For reflections on the EU's sometimes shaky 'social legitimacy' see Ulrich Haltern, 'Finalität' in Armin von Bogdandy and Jürgen Bast (eds), *Europäisches Verfassungsrecht* (Springer 2003) 283–285; pointing especially to the ECJ and the Union's growing executive power Dieter Grimm, 'Auf der Suche nach Akzeptanz - Über Legitimationsdefizite und Legitimationsressourcen der Europäischen Union' (2015) 43 Leviathan 325, 328 et seq; critical in that respect Christoph Möllers, 'Krisenzurechnung und Legitimationsproblematik Der Europäischen Union' (2015) 43 Leviathan 339, 341 et seq, 343 et seq, 356 et seq; highlighting the legitimizing potential of the EU's bureaucratic expertise Enrico Peuker, *Bürokratie und Demokratie in Europa: Legitimität im europäischen Verwaltungsverbund* (Mohr Siebeck 2011) 218–224.

[20] Insightful and comprehensive: Alasdair R Young, 'The European Union as a Global Regulator? Context and Comparison' (2015) 22 Journal of European Public Policy 1233, 1236–1343; criticizing the 'reductionist and monocausal' brushstrokes of the Brussels Effect: Abraham L Newman and Elliot Posner, 'Putting the EU in Its Place: Policy Strategies and the Global Regulatory Context' (2015) 22 Journal of European Public Policy 1316, 1321. Arguing that the EU is a global actor 'past its peak' Charlotte Bretherton and John Vogler, 'A Global Actor Past Its Peak?' (2013) 27 International Relations 375, 386. See already much earlier assessments of the EU's lacking ability (or willingness) to 'export' human rights norms

EU policymakers pressed forward, undeterred by the sceptics, initiating several political projects, mostly in the digital realm, using the language of and aiming for the results described in Bradford's Brussels Effect.[21] The prime example is the DSA. Another one would be the AIA.

This article focuses mainly on the DSA, as it has already entered into force, but will make sidenotes where necessary on the AIA and Regulation (EU) 2022/1925, commonly known as the Digital Markets Act (DMA).[22] As Ursula von der Leyen said in her State of the Union address in 2020, the Commission 'envisages the Digital Services Act as a standard-setter at global level'.[23]

III. HOW DOES THE BRUSSELS EFFECT WORK?

But how does the Brussels Effect work in practice? According to Bradford, not every large jurisdiction or big market exerts the type of global regulatory force the EU does.[24] It is the interplay of five elements that enables the Brussels Effect: 'market size, regulatory capacity, stringent standards, inelastic targets, and non-divisibility'.[25]

1. Market and Product

Two elements, market size and regulatory capacity, are self-explanatory. Inelastic target means that the Brussels Effect only transpires if regulated products or actors cannot escape the regulation through forum shopping.[26]

via trade agreements Frank Hoffmeister, *Menschenrechts- und Demokratieklauseln in den vertraglichen Außenbeziehungen der Europäischen Gemeinschaft* (Springer 1998).

[21] See above n 18.

[22] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[23] European Commission, 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 2000/32/EC 2020 [COM(2020) 825 final] explanatory memorandum at 2.

[24] Bradford, *The Brussels Effect* (n 2) 37 et seq.

[25] Ibid 25.

[26] Ibid 48 et seq.

This can be relatively easily achieved through tying regulatory jurisdiction to engagement with EU customers. If goods or services are sold to EU consumers, EU regulations apply, irrespective of where those goods and services come from. Non-divisibility, in turn, means that the regulated product, service, or actor cannot easily be divided into one version for the EU market and another for the rest of the world.[27] That is often the case for economies of scale and services relying on network effects like social media. The bumpy start of Meta's new network Threads – which did not launch in the EU, as it did not comply with EU data protection and antitrust law – highlights this.[28] We can assume that all these four elements will play out in favour of globalizing the regulatory effects of recent EU regulations, especially the DSA.

2. Stringent Standards

One last element, the 'stringent standards', may be trickier. As Bradford writes,

> even significant regulatory capacity by a large market does not guarantee regulatory influence unless such regulatory capacity is supplemented by the political will to deploy it. Thus, the Brussels Effect requires that the jurisdiction also has the propensity to promulgate stringent regulatory standards.[29]

However, the exact contours of each element emerge only reflexive to the other elements. Simply put, a stringent standard in food safety law is, in practice, not the same thing as a stringent standard in data protection law or, as in our example, platform governance and content moderation. The permissible amount of certain ingredients in food may be defined in clear units (e.g., only amount x of bisphenol, which leads to menstrual and even fertility problems, per unit food). In contrast, moderating speech and

[27] Ibid 53 et seq.

[28] Anu Bradford, 'Meta vs the EU: Who Governs the Digital Economy?' (*ukandeu.ac.uk*, 30 August 2023) <https://perma.cc/TWL2-86HX>.

[29] Bradford, *The Brussels Effect* (n 2) 37.

designing the organizational safeguards of a social media platform escapes such clearly specified units. Whether communication acts are offensive or illegal, hate speech or mockery, ironical or appalling cannot be defined easily through stringent substantive standards. Instead, it requires organizational reform and a cultural shift within those companies.

3. Global Diffusion of Private Ordering?

Therefore, the DSA – and other technology-focused regulations – only outlines (broad) normative demands for private ordering. The DSA formulates abstract goalposts like 'fair procedures' or 'due regard' for fundamental rights while companies may figure out the normative, procedural, and organizational path towards reaching these goalposts. This is something fundamentally different from the Brussels Effect's earliest occurrences in areas like food safety, which often come with much more granular and concise – and in that sense 'stringent' – standards.[30] Because coming up with a stringent substantive standard is increasingly complicated in digital governance, the EU devised the DSA as a 'turn to process'.[31] Mandating platforms to establish fair and transparent procedures is thought to assure 'good' behaviour of platforms.

Crucially, in this model, the norms applied vis-à-vis users are usually not laws enacted by the EU or its Member States but private ordering like social media companies' so-called community standards. These are the rules that govern what can be uttered on social media websites. Those private norms are stipulated by the companies themselves, largely as they please.[32] These

[30] See in that regard ibid 171 et seq.

[31] See e.g. Martin Husovec and Irene Roche-Laguna, 'Digital Services Act: A Short Primer' (SSRN, 5 July 2022) <https://perma.cc/JE3Q-55XA>; Martin Husovec, *Principles of the Digital Services Act* (Oxford University Press 2024).

[32] See e.g. Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019); Luca Belli, Pedro Augusto Francisco and Nicolo Zingales, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Luca Belli and Nicolo Zingales (eds), *Platform Regulations - How Platforms are Regulated and How they Regulate Us [Official Outcome of the UN IGF Dynamic Coalition on Platform*

are numerous layers of explicit or even technical norms regulating user behaviour, some labelled as terms of service or community standards, others unnamed. Also of paramount importance are infrastructures and technological set-ups.[33] The EU does not regulate these norms directly. Those terms of service must not fully reflect the EU *acquis* of fundamental rights (the abstract definition of which would be another formidable challenge). Instead, platforms must only pay 'due regard' to EU fundamental rights when enforcing and applying platform-made rules. I will discuss this in detail in the following section. For the moment, we may nonetheless anticipate that such mediation shifts the authority from the EU to other actors, especially tech companies. As Jennifer Daskal put it:

> forms of unilateral, global rulemaking are mediated through private sector actors rather than states or international institutions, making the private sector a central player in deciding whose rules apply and thus the scope of privacy and speech rights on a global scale.[34]

This shift may have far-reaching implications on the quality of the EU's global regulatory influence.[35] There are not one but two sets of standards that globally diffuse. Only one is formulated by the EU, the other remains within the remit of transnational corporations. The latter standard may differ

Responsibility] (2017); from an American perspective Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harvard Law Review 1598.

[33] See in general Paul N Edwards, 'Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems' in Thomas J Misa, Philip Brey and Andrew Feenberg (eds), *Modernity and Technology* (MIT Press 2002); Benedict Kingsbury and Nahuel Maisley, 'Infrastructures and Laws: Publics and Publicness' (2021) 17 Annual Review of Law and Social Science 353. For specific case studies see especially Eerie's and Streinz's analysis of how the People's Republic of China leverages such technical infrastructures to expand its (normative) influence outside of the arena of formal law, Mathew S Erie and Thomas Streinz, 'The Beijing Effect: China's "Digital Silk Road" as Transnational Data Governance' (2021) 54 New York University Journal of International Law and Politics 1. See further Evelyn Douek, 'Content Moderation as Systems Thinking' (2022) 136 Harvard Law Review 526.

[34] Jennifer Daskal, 'Borders and Bits' (2018) 71 Vanderbilt Law Review 179, 235.

[35] See further from the perspective of political philosophy Linnet Taylor, 'Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector' (2021) 34 Philosophy & Technology 897.

substantially from the former. Unfortunately, as the next section shows, the standards set by the DSA are not as stringent as we might want. As a consequence, the DSA only rather loosely confines the private ordering of platform companies.

Thus, in a scenario many might find more dire, the actually 'stringent' standards are devised by transnational corporations. Those private standards might very well be global in reach – whether they are indeed European in substance (whatever that means) remains to be seen. In conclusion, the one thing from the DSA that might reverberate globally are some structural governance decisions as to how platforms moderate content – not which content is substantively legitimate. From a critical perspective, even the DSA's purported focus on process appears exaggerated, as the DSA mostly only outlines what processes shall achieve – not what they shall look like. Such mission control tactics require agents who share the regulators' intentions, which is not necessarily the case here.

This might be no news to everyone who read the DSA's fine print. The Commission never claimed that it would substantively regulate what can be said online. Yet, one cannot help but notice a stark difference between the grandiose expectations towards the DSA's normative influence and the leeway it grants to platforms.

IV. WHAT DOES THE EU WANT?

The previous section highlighted that the Brussels Effect requires regulations to be 'stringent'. But what would be a 'stringent' standard in the context of Brussels' reforms of the digital single market? Or, more abstractly spoken, what is the overall regulatory aim of acts like the DSA?

1. *Make Platforms Better Administrators*

Simply put, the EU wants to tame the private power structures that control technology, infrastructure, data, and services in the digital single market.[36] Currently, among the most dominant *leitmotifs* structuring these company dealings are profit and scalability.[37] In principle, these are classic traits of any business. However, as in any other business, the pursuit of profit might yield externalities. In the case thoroughly entrenched products like social media and other digital infrastructure, those externalities can be particularly high. In some cases, journalists, activists, and politicians therefore highlighted problematic or outright illegal practices ranging from data breaches, impacts on societal and political discourse, to fundamental rights abuse.[38] Since platforms (and other tech companies) run infrastructures that govern societal discourse and the exercise of public rights, the EU (and other actors) want platforms to be more accountable, transparent, and 'fair'. In other words, platforms should become better administrators. The DSA seeks to inject certain administrative law principles – like non-discrimination, proportionality, and a duty to give reasons for individual decisions – into platforms' private governance structures.[39] In that sense, content

[36] See already Moritz Schramm, 'Platform Administrative Law: A Research Agenda' (SSRN, 24 July 2024) <https://perma.cc/JH2A-8ZW2> and idem, 'Administrification of the Digital Single Market: A New Role for the European Ombudsman in the DSA Framework?', in Deirdre Curtin, Tanja Ehnert, Anna Morandini, Sarah Tas (eds), *The Evolving Role of the European Ombudsman* (Hart forthcoming 2025).

[37] See further Adrian Daub, *What Tech Calls Thinking: An Inquiry into the Intellectual Bedrock of Silicon Valley* (Farrar, Straus and Giroux 2020).

[38] See e.g. Jeremy B Merril and Will Oremus, 'Five Points for Anger, One for a "Like": How Facebook's Formula Fostered Rage and Misinformation' *Washington Post* (26 October 2021) <https://perma.cc/7E55-RPZL>; Sheera Frenkel and Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination* (Harper 2021); Rune Karlsen and others, 'Echo Chamber and Trench Warfare Dynamics in Online Debates' (2017) 32 European Journal of Communication 257.

[39] Public law metaphors of content moderation abound, with some highlighting constitutional perspectives as others more administrative and bureaucratic aspects. See further, Giovanni De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Cambridge University Press 2022) 157 et seq. In the US, Jack Balkin referred to these phenomena as 'private bureaucracies', which strikes me as a fitting metaphor, see Jack M

moderation, platform governance, and many other digital reforms can be conceptualized as a new form of Global Administrative Law (GAL).[40] GAL – once conceptualized by Benedict Kingsbury, Richard Steward, and Nico Krisch in the 2000s – described administrative spaces beyond the state.[41] Potentially, it also includes private and profit-oriented administrators, but never really followed through on those.[42]

Because platforms deal with 'public goods' like public communication and fundamental rights at a massive scale, they are increasingly regulated to abide to public law norms. These norms regulate how platforms shall deal with their users. The normative model for many such rules – fundamental rights, due process, proportionality, non-discrimination, hearing rights, duties to give reasons etc. – is public law, or more precisely administrative law. The whole endeavour could be called the 'administrification' of content moderation.[43] Administrification of content moderation and platform governance is a relatively recent but not entirely new phenomenon.[44] In 2016, several social media platforms, including Facebook and X, agreed with

Balkin, 'Free Speech Is a Triangle' (2018) 118 Columbia Law Review 2011, 2021 et seq. See further Schramm, 'Platform Administrative Law' (n 36).

[40] See further Hannah Bloch-Wehba, 'Global Platform Governance: Private Power in the Shadow of the State' (2019) 72 SMU Law Review 27; Schramm, 'Platform Administrative Law' (n 36).

[41] For an introduction to GAL see Benedict Kingsbury, Nico Krisch and Richard B Stewart, 'The Emergence of Global Administrative Law' (2005) 68 Law and Contemporary Problems 15; Sabino Cassese, *Advanced Introduction to Global Administrative Law* (Edward Elgar Publishing 2021); Benedict Kingsbury, 'Frontiers of Global Administrative Law in the 2020s' in Jason NE Varuhas and Shona Wilson Stark (eds), *The Frontiers of Public Law* (Hart 2020).

[42] However, see recently esp. Bloch-Wehba (n 41); Rodrigo Vallejo Garretón, 'After Governance?: The Idea of a Private Administrative Law' in Poul F Kjaer (ed), *The Law of Political Economy* (Cambridge University Press 2020); Douek analogizes content moderation to administration but then focuses on the (somewhat elusive) concept of 'systems thinking', Douek (n 33).

[43] See further Schramm, *Governance by Emulation* (n 9); idem, 'Administrification of the Digital Single Market: A New Role for the European Ombudsman in the DSA Framework?'(n 36); idem, 'Platform Administrative Law' (n 36).

[44] Comparisons between platforms and administrators emerged roughly since 2018, see speaking of 'privatized bureaucracies' Balkin (n 41); comparing content moderation to independent agencies Klonick (n 32); making a comparison to Global Administrative Law Bloch-Wehba (n 40); see further Douek (n 33).

the European Commission on a code of conduct when it came to countering illegal hate speech online.[45] The social media platforms agreed to 'have in place clear and effective processes to review notifications regarding illegal hate speech' and committed 'to review such requests against their rules and community guidelines'.[46] Since platforms operate at scale, their commitments towards EU authorities are likely to affect, and perhaps water down, their global standards. In this context, Anu Bradford argued that platforms' normative material (the so-called community standards) increasingly mirrors the normative *acquis* of EU law.[47]

2. *The Reality of Content Moderation*

However, it remains unclear whether platforms' terms of service indeed 'reflect *the* European standard of hate speech',[48] or merely echo the overall tone of EU law without reflecting the EU *acquis* in a substantively meaningful way. That is primarily because the Union's normative material remains much more coarse than the material implemented by platforms. For example, the Charter of Fundamental Rights establishes a right to free expression but does not detail what type of speech this right covers in practice. Certainly, doctrine and jurisprudence establish an overarching framework to refine more specific norms. However, platforms practically do not apply norms like 'everyone enjoys freedom of expression'. As Klonick convincingly argued, platforms cannot work (only) with coarse normative standards for the sheer scale of the task.[49] Rather, the norms employed by platforms are thickly layered, convoluted, adjective-laden descriptions of what content is concretely prohibited.[50] For example, at the time of writing Meta defined a 'tier 1 direct attack' within its standard on hate speech as

[45] EU Code of Conduct on countering illegal hate speech online 2016 <https://perma.cc/7ABV-TNCQ>.
[46] ibid 2.
[47] Bradford, *The Brussels Effect* (n 2) 161 et seq.
[48] ibid 158 (emphasis added).
[49] See illuminatingly Klonick (n 32) 1631.
[50] In that sense, Klonick differentiates between 'standards' and 'rules', ibid.

> [d]ehumanizing speech or imagery in the form of comparisons, generalizations, or unqualified behavioural statements […] to or about […] insects, animals that are culturally perceived as intellectually or physically inferior, […] sub-humanity.[51]

The practical complexity of such a definition is obvious. How to draw a line between 'direct' and 'indirect' attacks against people? How to handle necessarily subjective, value-laden, and context-dependent adjectives like 'violent', 'harmful', or 'dehumanizing'? Which diseases are 'serious'? Does only the content of a communication act make it 'hate speech' or also the attitudinal stances of the person who expresses it?[52] What about the temporal and territorial space in which an utterance was made? Are we even capable of identifying the past, current, and future social, political, and cultural contexts of an utterance with sufficient certainty? What about contexts that are culturally foreign to Meta's predominantly North American rulemakers? Practitioners I interviewed for earlier, currently unpublished work described content moderation as 'phenomenally difficult',[53] and scholars[54] and tech journalists[55] have even pronounced content moderation an impossible endeavour.[56] Platforms' increasing reliance on automated moderation

[51] See: <https://perma.cc/GGB3-XST8> (last accessed 4 October 2024).

[52] In essence, these questions point to one of the central aporias of the philosophy of language. See Judith Butler's critique of J.L. Austin's 'total situation' approach. In his foundational work *How to Do Things with Words*, Austin argued that context – meaning mostly spatial and temporal aspects – would, if 'totally comprehended' arguably allow for an absolute interpretation of certain speech acts. Butler disagrees. According to them, nothing can be 'totally' understood, cf *Judith Butler, Excitable Speech: A Politics of the Performative* (Routledge 1997) 2–13. See further John L Austin, *How to Do Things with Words* (Oxford University Press 1962); John R Searle, *Speech Acts: An Essay in the Philosophy of Language* (Cambridge University Press 1969).

[53] This was phrasing used by a person working for Meta's Oversight Board who I interviewed for other work, see above in n 9.

[54] Douek (n 33) 533, 568.

[55] Mike Masnick, 'Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well' (*TechDirt*, 20 November 2019) <https://perma.cc/Q97F-CFHB>.

[56] Cf Schramm, *Governance by Emulation* (n 9).

(despite the well-known shortcomings of natural-language processing[57]) only exacerbates these difficulties.

Thus, in short, the core problem of content moderation is not that platforms lack an overall commitment to freedom of speech or other fundamental rights. The problem is the large organizational and normative gap between committing to high-flying norms and effectively safeguarding them in thousands of specified norms, typified decisions, and implementation measures.

Bradford points out that we do not know how these commitments are put into practice; especially since platform rules consist of many sublayers, and platform lawyers 'ultimately use their own judgment on what constitutes illegal hate speech. The outcome is therefore less likely to be perfectly aligned with the Commission's regulatory approach […]'.[58] Building on and further cementing this *division du travail* between social media platforms and European regulators, the Commission proposed a major overhaul of the regulatory framework for social media platforms in Europe in December 2020.

3. The Digital Services Act

This overhaul resulted in the DSA. Contrary to the Union's political marketing and widespread belief, it reiterates the relative freedom of platforms to devise their own procedures, decision-making processes, and substantive rules if they only abide by overarching and normatively inspired criteria like 'objectivity', 'proportionality', and 'due regard for fundamental rights'.[59] To keep this article lean, I will focus mostly on one of the key

[57] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 1. See also Paul Friedl, 'Dis/Similarities in the Design and Development of Legal and Algorithmic Normative Systems: The Case of Perspective API' (2023) 15 Law, Innovation and Technology 25.

[58] Bradford, *The Brussels Effect* (n 2) 162.

[59] See in detail ibid.

provisions, Article 14, although others would warrant equally close attention (we will briefly get to them below). As mentioned above, the DSA imposed several administrative law-inspired duties on platforms. Yet, the concrete path to get there remains unclear. This is partly because these duties are, upon closer inspection, more an ill-defined region than a clear point on the map. Take the enthusiastically discussed Article 14(4) DSA as an example. It stipulates that platforms should act with

> due regard to the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service, such as the freedom of expression, freedom, and pluralism of the media, and other fundamental rights and freedoms as enshrined in the Charter.

Some commentators consider Article 14 DSA as the gateway to directly bind platforms to EU fundamental rights.[60] However, from a more critical perspective, the standard only reaffirms private might. Paying 'due regard to the legitimate interests of all parties involved' is as wobbly a standard as it gets. Think of the snippet from Meta's hate speech standard presented above. Does that pay 'due regard' to fundamental rights? What would a specific rule that pays such 'due regard' look like? The DSA does not answer these questions. It also does not provide more normative specificity on how companies may reach such standards in any way that is different from what they are doing already.

Quite the contrary, in its impact assessment the Commission even argued that platforms would not have to change much in their internal procedures. In many respects the DSA does not explicate procedures platforms should implement but only the targets that should be reached through procedures, which are designed by the platforms themselves.

Consequently, Article 14 DSA never strikes a definitive equilibrium but requires case-by-case assessments. That is because the involved parties change from case to case and what interests are 'legitimate' and what 'regard'

[60] João Pedro Quintais, Naomi Appelman and Ronan Ó Fathaigh, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' (2023) 24 German Law Journal 881.

is 'due' evades stringent cognition.[61] Practically, it will be the platforms who rebrand (and perhaps improve) their already existing content moderation bureaucracies as doing exactly that: providing procedures that do a little better.

Strikingly, if one digs deep enough, the Commission itself does not assume that provisions like Article 14 will initiate meaningful reform in the platforms. Buried deep in the DSA's impact assessment, one finds hints that the Commission might have underestimated the costs of – and, therefore, platforms' hesitancy towards – making platforms better administrators. As discussed, one of the DSA's key goals is to make platforms more responsive to the fundamental rights of their users. One crucial mechanism to facilitate this is to introduce notice and action obligations, information duties vis-à-vis users, procedural balancing obligations, and redress mechanisms. This is why we have, among others, Article 14. These are key provisions of the DSA. Nonetheless, the Commission argues that compliance to such allegedly groundbreaking and new normative goals of fairness, due process, and fundamental rights protection would come at zero costs. Verbatim, the Commission states regarding potential costs for platforms implementing crucial DSA provisions on fundamental rights and content moderation:

> These are indicative costs and, for most companies, they do *not* represent an additional cost compared to current operations, but require a process *adaptation* in the receipt and processing of notices and streamline costs stemming from fragmented obligations currently applicable.[62]

[61] Ironically, these developments in EU regulatory law seem to run counter to the overall thrust of platforms' internal rules, which become ever more specific and detailed. Already in 2018, Kate Klonick argued that platforms modified their approach from establishing standards to enacting much more concise and detailed rules, cf Klonick (n 50) 1631–1635.

[62] European Commission, 'Commission Staff Working Document, Impact Assessment, Accompanying the Commission Proposal for the DSA, Part 1/2. SWD (2020) 348 Final' para 197, table 4 at row 2. Further, the EU's focus on 'streamline cost' omits that according to Articles 2 and 20, platforms are also required to delete content illegal under Member State law. This necessitates at least partially fragmented enforcement structures.

But why would a 'process adaptation' not incur costs? Especially, when processes shall transform from their prior largely unaccountable state to full alignment with EU law? Put reversely, if transforming content moderation takes so little effort that it is essentially free, what do we need the DSA for? Yet, as argued above, a deep-rooted reform of content moderation and platform governance would require a deep overhaul of procedures and overall considerable resource allocation.[63] The best things in life are free – good content moderation certainly is not.

Therefore, one may be reasonably doubtful whether the DSA in its current form brings strict enough rules to change content moderation in Europe and on a global scale. Eventually, the Commission and, ultimately, the European Court of Justice might fill some of these relational and vague provisions with concise meaning. Yet, as the next section explains in detail, platforms seem to remain in the driving seat at least for some of the DSA's key provisions.

V. Challenges of Platform Reform

The DSA's regulatory goal – making platforms better administrators – faces one big problem: presumably, platforms do not want to play along entirely. Acting like an administrator means, slightly exaggerated, to act cautiously rather than quick, proportionately rather than tough, and formalistic rather than efficient. To a degree, bureaucratization is an inevitable side-effect of growth, and furthermore well described for content moderation.[64] Having a degree of bureaucracy may align with platforms' profit interest, as globally uniform internal rules enable the company to further expound a business model that predominantly aims for scale. However, in contrast, many of the regulatory demands devised by European regulators – fairness, proportionality, due regard for fundamental rights, duty to give reasons,

[63] See e.g. Jack M Balkin, 'To Reform Social Media, Reform Informational Capitalism' in Lee C Bollinger and Geoffrey R Stone (eds), *Social Media, Freedom of Speech, and the Future of our Democracy* (Oxford University Press 2022).

[64] See especially Klonick (n 32). From a sociological perspective see already Philip Selznick, *Law, Society, and Industrial Justice* (Russel Sage Foundation 1969).

hearing rights, review, etc. – would, if implemented with fully-fledged public law idealism, likely jeopardize this business model. That is for two reasons: cost and vagueness.

1. Cost

On the one hand, a classic public-law-inspired understanding of core DSA ideas like proportionality, procedural fairness, and respect for fundamental rights would require an enormous expansion of the resources poured into making individual content decisions (not to speak of the systemic undercurrent of said decisions).[65] As we have just seen, Article 14(4) DSA could be understood, in its most ambitious interpretation, as requiring platforms to make individual balancing decisions with much greater care in every single case. Greater care requires more resources, which drives up costs and curbs profit. Balancing fundamental rights is intricate and, most importantly, extremely costly. Content moderation decisions are made under significant time and resource constraints. For instance, within a three-month period in 2022, YouTube removed over 737 million comments and more than five million videos globally, with human moderators, typically employed in South-East Asia under low-wage conditions, having only a few seconds per enforcement decision.[66] Similarly, internally handled appeals might be processed by different teams, potentially operating under slightly improved material and institutional circumstances.[67] In essence, the reality

[65] For the latter see Douek (n 33).

[66] See Google Transparency Report: YouTube Community Guideline Enforcement, accessible via: <https://perma.cc/67GV-R8HZ>. For updated numbers on Meta see the respective transparency center accessible via: <https://perma.cc/67GV-R8HZ>.

[67] Insightful: Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2019) 111-139 (esp. 120–124). See also, highlighting the often precarious working conditions for many outsourced human moderators Sarah T Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (Yale University Press 2019) 173, 176–183; MacKenzie F Common, 'Fear the Reaper: How Content Moderation Rules Are Enforced on Social Media' (2020) 34 International Review of Law, Computers & Technology 126, 127 et seq.

of content moderation diverges from the ideal envisioned in Article 14(4) DSA.

Further, the DSA seeks platforms to enforce a multitude of EU and Member State laws, which again requires setting up, in theory, 28 different enforcement schemes. One for the EU and 27 for the Member States. At least for smaller Member States, platforms may find this exceedingly costly and, perhaps even more important, contrary to their internal goal of unifying standards globally.

2. Vagueness

On the other hand, many of the DSA's normative goals are vague. Taking Article 14 as an illustration, Martin Eifert and his colleagues contend that the DSA grants platforms 'unfettered discretion' to engage in private regulation of public discourse.[68] It is worth noting that the due diligence obligations outlined in the DSA cover the individual application and enforcement of platform-made rules, rather than the making of these rules. But can any bureaucracy be legitimate if its rulemaking lacks any meaningful input legitimacy?[69] Similar criticism was voiced regarding the Digital Markets Act, whose provisions are, according to Anne Witt 'not as rigid as they may appear at first sight'.[70]

Viewed through the lens of public law, it seems logical to bind individual enforcement actions to fundamental rights. Requiring digital corporations to 'pay due regard' to fundamental rights when enforcing terms of service should, theoretically, resolve the problem that content moderation may

[68] Martin Eifert and others, 'Taming the Giants: The DMA/DSA Package' (2021) 58 Common Market Law Review 987, 1013.

[69] Speaking of 'authoritarian' structures at Facebook: Evelyn Douek, 'Facebook's "Oversight Board": Move Fast with Stable Infrastructure and Humility' (2019) Volume 21 North Carolina Journal of Law & Technology 1, 10.

[70] Overall, Witt's assessment of the DMA is however rather favourable, especially highlighting that lobbying efforts presumably did not manage to water down many core provisions, see Anne C Witt, 'The Digital Markets Act - Regulating the Wild West' (2023) 60 Common Market Law Review 625, 651 et seq, 665.

negatively impact fundamental rights.[71] In almost orthodox fashion, Article 14(4) DSA appears to mandate platforms to engage in ongoing balancing, aiming to prevent infringements of individual rights. However, public law orthodoxy cannot be easily transplanted to a for-profit, corporate context like content moderation. Doctrinal manoeuvres like 'horizontal effect', at the end of the day, deal only with a doctrinal understanding of the problem at hand. Additionally, the DSA's introductory 'recitals' fail to clarify how and to what extent platforms should consider fundamental rights when restricting communication. Consequently, the notion of individuals consciously balancing conflicting fundamental rights for each moderation decision appears impractical.

This undermines the EU's argument that the DSA establishes potentially globalizing 'stringent standards'. What exactly should that standard be? That platforms act 'nicely' vis-à-vis users? Unless pressed with concise demands of what platforms should actually *do* and not merely what they should seek to achieve, the DSA may trigger a global wave of sugarcoating – not one of deep-running reforms.

3. Another Example: The 'Compliance Function'

These resource constraints on behalf of platforms, paired with the vagueness of the goals, open a grey area in which platforms may comply with the letter of the law but are far from achieving its high-flying intention. The DSA provisions enabling such merely ceremonial adaptation are, especially, Articles 14 (terms and conditions), 20 (internal complaint handling system), 21 (out-of-court dispute settlement bodies), 25 (online interface design and organisation), 34 (risk assessment), 35 (mitigation of risks), and 41 (compliance function). These provisions oblige platforms to establish

[71] However, even this language allows for interpretation, as it does not straightforwardly mandate platforms to safeguard fundamental rights but only to pay 'due' respect to such rights. As any first-year law student understands, any additional word or modification appended to an otherwise robust obligation can be exploited to undermine an otherwise seemingly robust legal assurance. The precise meaning of 'due' in this context remains subject to interpretation and application.

procedures or formal structures to achieve a specific normative target. Examples of such targets are 'fair' and 'proportionate' procedures, a 'compliance function with sufficient authority, structure, and resources', or the wholly unspecified demand that platforms should design their infrastructure not to 'deceive, manipulate, or otherwise impair their users' ability to make free and informed decisions'. All these provisions focus on important and publicly discussed issues like fundamental rights and terms of service (Article 14), platforms' internal proceedings (Article 20), individual rights adjudication (Article 21), the 'infrastructure design' of platforms (Article 25), and arguably even pick up on scholarly demands for a 'separation of functions' within platforms (Article 41). For example, Article 41(1) of DSA asserts:

> Providers of very large online platforms […] shall establish a *compliance function*, which is independent from their operational functions and composed of one or more compliance officers, including the head of the compliance function. That compliance function shall have sufficient authority, stature and resources, as well as access to the management body of the provider of the very large online platform […] to monitor the compliance of that provider with this Regulation. (emphasis added)

Here again, the DSA ostensibly echoes public law principles like the separation of powers or, in a more contemporary reading, the separation of functions. Notably, Article 41(1) of the DSA does not refer to compliance officers or a compliance department but to a compliance *function*.[72] The provision was expanded very late on in the legislative procedure. The Commission proposal only referred to compliance officers, which remained the official language until at least December 2021. Arguably, the ultimately legislated 'compliance function' seeks to implement expert input, as it

[72] See again for a critique of such metaphorical allusions to constitutional law: Josh Cowls, Philipp Darius, Dominiquo Santistevan, Moritz Schramm, 'Constitutional Metaphors: Facebook's "Supreme Court" and the Legitimation of Platform Governance' (2024) 26 New Media & Society 2448. Also, regarding the discussions about the Oversight Board, Thomas Kadri, 'Juridical Discourse for Platforms' (2022) 136 Harvard Law Review (Forum) 163.

mirrors recent proposals for the reform of online platforms.[73] In general, it seems fair to assume that these problems are too complex and the organizations causing them too big to hope for sweeping one-size-fits-all solutions. Remedying, or at the very least softening, the harm caused by deeply embedded governance flaws tends to occur through step-by-step reforms. However, the DSA's regulatory design risks watering down public sway over platform governance because, in practice, it will be the platforms who determine whether their compliance function has 'sufficient authority, stature and resources'.

Further, a mere compliance function falls far short of the recommendations from the original expert theorization, which demanded that platforms separate their business function from their governance function.[74] Especially Evelyn Douek demanded severing rulemaking and rule enforcement from advertising and product design.[75] Except for its reference to an independent compliance function, the DSA does too little to deliver on the substance of that expert theorization. Currently, the Union's publicly advocated normative goals risk remaining constitutional metaphors that fail to foster internal change but instead publicly legitimize a mildly updated version of a structurally flawed status quo.[76] Thus, in its current form, the DSA might even stabilize problematic governance practices because platforms can now legitimately claim that they abide by a regulatory framework that imposes

[73] Infusing a separation of functions was one of Evelyn Douek's key reform ideas, cf Douek (n 33) 586 et seq.

[74] The term 'expert theorization' stems from organization theory and describes the role of experts in policy diffusion and emulative practices, especially in lawmaking and institution building, see further Beth A Simmons, Frank Dobbin and Geoffrey Garrett, 'Introduction: The Diffusion of Liberalization' in Beth A Simmons, Frank Dobbin and Geoffrey Garrett (eds), *The Global Diffusion of Markets and Democracy* (Cambridge University Press 2008) 34 et seq.

[75] Douek (n 33) 586 et seq. Note, however, that this distinction is a bit artificial because content moderation and product design are inherently intertwined (e.g., regarding amplification and demotion of content). See in that regard Jing Zeng and D Bondy Valdovinos Kaye, 'From Content Moderation to Visibility Moderation : A Case Study of Platform Governance on TikTok' (2022) 14 Policy & Internet 79.

[76] Cowls, Darius, Santistevan, Schramm (n 72); Kadri (n 72).

on them a variety of public-law principles. However, as sociological research tells us, formally conforming to these public-law principles is relatively easy – in contrast, substantively advancing them would entail a fundamental reshuffle of platform's structures, procedures, norms, and business models.

VI. Ambiguities of Compliance

These observations point to a tricky reality. Clearly, many dynamics of the Brussels Effect may still play out for Europe's digital reforms. Except for stringent standards, all elements seem to be there. Even with loose standards, at least good ideas and intentions might diffuse. Also, the Commission may further refine some DSA provisions through guidelines and delegated legislation. Yet, from its whole regulatory focus on broad goals and loosely-defined process, the DSA legally stabilizes rather than restricts platform authority. Simply put, platforms were doing all these things anyway. The DSA now says: keep going but improve your customer relations.

But what does this tacit but indeed massive shift in rulemaking authority mean for the global relevance of EU norms? Will it really be EU norms that globalize, or rather private ordering cloaked in the legitimizing guise of ceremonial compliance to EU law. As my earlier empirical work mentioned above and organization theory indicate, we may expect a form of ceremonial compliance on behalf of platforms but only limited substantive change. Platforms may devise several public-facing formal structures to interact with individual users. The lasting and structural effect of such ceremonial structures remains however unclear. This would be an ironic result of the Union's global ambitions. In such a scenario, largely unchanged private power structures can lay valid claim to comply with EU law and would be seen, thereby, as legitimate.

1. *What is Ceremonial Compliance?*

The term ceremonial compliance warrants further explanation. It derives from the use of the term 'ceremony' in organization theory, which describes

how large organizations – including, perhaps especially, corporations – orient their formal structures not only according to efficiency considerations, but also conforming to public expectations about what makes for a publicly legitimate company.[77] Formal structures in that sense can be many things: for example a new human rights officer, a new compliance function, or a council of free speech advisors. Aligning formal structures with public demands for accountability is most visible with Meta, which invested – with some success – considerable resources into establishing its Oversight Board.[78] In other words, responding, even if only incrementally, to public demands for better governance once again highlights that such powerful and socially relevant organizations are more than a business. They govern, and therefore quite naturally must navigate demands of governance that aligns with basic principles commonly known from public law.

However, as the organization theorists tell us, aligning formal structures with public expectations may remain de-coupled from the organizations' actual practice.[79] Sociologists call this 'ceremonial' compliance.[80] Such ceremonial reforms benefit their creators by appearing to effectuate functional, and thereby publicly legitimizing, accountability – even though, as many have pointed out, in reality that accountability remains fairly limited. From a public law perspective, ceremonial compliance with EU law, that is compliance with the law's letter rather than its intent, would disappoint. These questions, of course, defy binary answers. Actors that begin as largely ceremonial may gradually accrue stature over time, while mechanisms hailed as powerful at their inception may later reveal themselves to be toothless. These are inherently complex, context-dependent issues.

[77] Foundationally: John W Meyer and Brian Rowan, 'Institutionalized Organizations: Formal Structure as Myth and Ceremony' (1977) 83 American Journal of Sociology 340; see in general Walter W Powell and Paul DiMaggio (eds), *The New Institutionalism in Organizational Analysis* (University of Chicago Press 1991).

[78] See also here Schramm, *Governance by Emulation* (n 9).

[79] See in detail Meyer and Rowan (n 77) 348 et seq.

[80] ibid *passim*.

However, in general, mere ceremonies cannot normatively legitimize the exercise of power, because they are constitutively unable to exercise control or demand accountability from the body they claim control over.

2. *Ceremonies in the Tech Sector*

Recent empirical scholarship brought to attention the complexities that large, regulated private organizations in the tech sector face when incorporating public law norms into managerial processes and mindsets.[81] Ari Ezra Waldman argued, based on extensive fieldwork regarding tech companies' struggle to make good on their public promises and protect their users' privacy:

> Tech companies […] routinize antiprivacy norms and practices in privacy discourse, compliance, and design. Those bureaucracies constrain workers directly by focusing their work on corporate-friendly approaches to privacy. As information industry workers perform these antiprivacy routines and practices, those practices become habituated, inuring employees to data extraction, even as they earnestly profess to be privacy advocates. The result is a system in which the rank and file have been conscripted into serving the information industry's surveillant interests, and in which the meaning of privacy has been subtly changed, often without them even realizing what's happened.[82]

Similar phenomena have been observed for decades in the field of antidiscrimination law. Lauren Edelman famously argued that companies all too easily devise 'visible symbols' of compliance but fail to vehemently further the legislated course. Scholarship suggests that such visible symbols rather than actual compliance emerge

[81] See seminally Meyer and Rowan (n 77); other 'ceremonial' structures could be, for example greenwashing or tokenism. See e.g. Magali Delmas and Vanessa Cuerel Burbano, 'The Drivers of Greenwashing' (2011) 54 California Management Review 64; Mariateresa Torchia, Andrea Calabrò and Morten Huse, 'Women Directors on Corporate Boards: From Tokenism to Critical Mass' (2011) 102 Journal of Business Ethics 299; Ari Ezra Waldman, *Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power* (Cambridge University Press 2021).

[82] ibid 5.

> [w]here legal ambiguity, procedural constraints, and weak enforcement mechanisms leave the meaning of compliance open to organizational construction.[83]

As argued above, the DSA brings such ambiguity. Further, the whole 'procedure focused' regulatory approach 'leave[s] the meaning of compliance' to the construction of whatever the companies say it is. Hence, the 'procedures' are indeed not devised by the regulator but by the regulated entity itself.

Nonetheless, the DSA vests the Commission with the authority to impose hefty fines on non-compliant companies. However, from a rule of law perspective, such fines can hardly be based directly on vague provisions demanding only 'due regard' and fair procedures. Or, put reversely, it would be easy to comply with those vague provisions by establishing formal structures. Therefore, a serious enforcement of all the institution-building foreseen by the DSA will require the Commission to specify the content and scope of many DSA provisions through guidelines, delegated legislation, and the like.[84] Further, crucially, the Commission must remain vigilant and well-staffed vis-à-vis platforms to then incrementally developed 'stringent standards'. However, for the short to medium term, it might not be the DSA or the yet to be crafted specification thereof but the tech companies' shiny but ceremonial compliance efforts that will reverberate globally.

If that were to happen, we might have reached a dead end. To harvest the legitimizing fruits of its political labour, the EU will presumably trumpet how the DSA now forces platforms to 'fly by our […] rules' as Thierry Breton put it.[85] Consequentially, and rightly so, industry actors might pick up that narrative and argue vis-à-vis other regulators and the public that they

[83] Lauren B Edelman, 'Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law' (1992) 97 American Journal of Sociology 1567 (emphasis added).

[84] See in that sense for the DMA also Witt (n 70) 651 et seq.

[85] The full tweet was: '[waving hand emoji] @elonmusk In Europe, the bird will fly by our [EU flag emoji] rules. #DSA', Thierry Breton, 29 October 2022, see via: <https://perma.cc/CPH9-LZL5>. At the time, Breton was the Commissioner for Internal Market.

indeed already comply (globally) with the new, purportedly strict rules from Brussels. As described above, when it comes to the (wobbly) letter of the law that may very well even be accurate in most cases. This effectively places an EU seal of approval on the very structural dynamics of digital power relations that the EU aimed to reform. Whether this is a desirable outcome for the EU, or indeed for anyone, remains to be seen.

VII. Possible Solutions

To end this article on an optimistic note, I now briefly present two potential solutions to the problems described above. These solutions are well-balanced regulatory specificity and a stronger focus on technological rather than legal expertise in the legislative process. Clearly, neither solution is a silver bullet. Yet, they might be a start. Effectively devising them warrants its own article and can only be addressed briefly in the remaining paragraphs.[86]

1. Balance Specificity and Broadness through Supervision

The most potent but also most complicated device against ceremonial compliance may be to better balance specificity and broadness in the normative material applied to tech companies. Or, to use American terminology, effective regulation thrives on the interplay of broad standards and specific rules.[87] The clearer the normative demands of a regulation are, the easier it is to part actual compliance from window dressing. However, in highly complex and constantly changing fields like technology

[86] Ironically, demanding the EU to balance specificity and broadness could also be considered a rather vague goal as the actual means to achieve it are organizational rather than substantive.

[87] For the distinction between standards and rules see Pierre Schlag, 'Rules and Standards' (1985) 33 University of California Los Angeles Law Review 379. Kate Klonick used the distinction also to emphasize the increasing specificity the normative material platforms enforce vis-à-vis their users, cf Klonick (n 32) 1631 et seq.

regulation, such specificity is elusive.[88] To remain adaptive towards the subject they regulate, norms must strike a delicate balance between broadness and specificity, shall neither be overinclusive (as this waters down their normative guidance) nor underinclusive (which would limit their applicability).[89] Being overly specific may lock in antiquated understandings of the regulated phenomenon.[90] Further, too specific rules also invite evasion as regulated entities may claim that a specific rule does not fit their specific phenomenon.

Especially the latter phenomenon can be theorized as a form of 'regulatory arbitrage'.[91] Regulatory arbitrage describes how financially potent actors evade the spirit of, for example, tax laws by 'exploiting the gap between the economic substance of a transaction and its legal and regulatory treatment'.[92] Entirely eradicating such avoidances appears impossible. Nonetheless, we

[88] For a theoretical overview of regulation (not only regulatory law) see generally Peter Drahos, *Regulatory Theory: Foundations and Applications* (Australian National University Press 2017); and famously, Lessig who emphasizes that other norms (social, technological) and infrastructures 'regulate' human behaviour Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books 1999).

[89] The predominantly US-American distinction between 'overinclusive' and 'underinclusive' is here understood as describing only whether a norm manages to regulate the phenomena it intends to regulate. This narrow understanding does not concern whether a norm's particularly broad or narrow approach appears justified regarding other (higher ranking) norms. In the US, these two issues of, on the one hand, the functionality of the norm itself, and on the other hand, the justification thereof are often discussed interchangeably. However, from a European and – admittedly– predominantly German perspective, the latter issue would be addressed in terms of proportionality. For the US American perspective see Kenneth Simmons, 'Overinclusion and Underinclusion: A New Model' (1989) 36 University of California Los Angeles Law Review 447.

[90] Such 'locking in' is Douek's criticism of reforms like the DSA. Whether any large regulation would escape the problem as Douek describes it is unclear. The experimentalist 'systems thinking' approach proposed by Douek seems not to lock in any strict regulatory guidelines – and may therefore be equally criticized as too lenient with actors that, in the words of Meta whistleblower Frances Haugen, known to 'pick profit over safety every time', cf Douek (n 33) 564 et seq; 'Statement of Frances Haugen' <https://perma.cc/3QN6-JMY4>.

[91] Viktor Fleischer, 'Regulatory Arbitrage' (2010) 89 Texas Law Review 227.

[92] ibid 229.

can learn from tax law that it makes sense to establish – ex post if necessary – legal constraints on specific circumventions if detected.[93]

Regulatory norms must be broad enough to cover all relevant practices yet specific enough to clearly outlaw unwanted practices. For example, legal norms criminalizing theft do not differentiate whether the thief steals an item by putting it into their backpack or back pocket but define and criminalize theft as such, which may come in many forms. Yet, rules covering complex phenomena like content moderation, artificial intelligence, or privacy cannot be easily distilled into crisp, one-size-fits-all norms. Lastly, supervisors must enforce regulatory norms with a realistic picture of the willingness and ability of tech companies to meaningfully reform their practices.[94] As Viktor Fleischer quipped:

> policy makers [and supervisors] should not rely on moral suasion or ethical or professional constraints on [regulatory] arbitrage. Lawyers have a professional obligation to help their clients manage regulatory costs, and the idea that lawyers would discourage their clients from engaging in behaviour that is legal and profitable would not likely be effective, even if all lawyers were saints, which we are not.[95]

Therefore, balancing specificity and comprehensiveness requires ongoing revisions, implementations, interpretations, reviews, and guidelines.[96] In that sense, maintaining the balance between specificity and comprehensiveness is a process rather than a state. Giving platforms leeway to design procedures and institutions to ensure effective compliance within such a framework may be sensible. Especially from an informational perspective, industry self-regulation can be simply inevitable as regulators often lack the capacity to

[93] Fleischer (n 91).

[94] See in that regard rather pessimistically Waldman (n 81).

[95] Fleischer (n 91) 289.

[96] Such practices can be theorized as 'experimentalism', see Charles F Sabel and William H Simon, 'Minimalism and Experimentalism in the Administrative State' (2011) 100 Georgetown Law Journal 53. For an illuminating perspective regarding financial services supervision see Niamh Moloney, *EU Securities and Financial Markets Regulation* (Fourth, Oxford University Press 2023) 944 et seq.

keep up with all the data, business models, innovation, and so on.[97] However, for the reasons mentioned above, we must assume that companies will utilize this leeway to their own benefit. And unsurprisingly so, since their 'freedom to conduct a business' is protected by the charter of fundamental rights.

The ensuing predicament of, on the one hand, lofty goals for fair procedures and, on the other hand, internalized motivations and practices hampering the fulfilment of said goals warrants a specific normative corridor to guide implementation and enforcement.

The DSA's current broad normative goals are not enough. Yet, hope is certainly not lost. Immediately after the ink of the DSA dried, the Union and the Member States entered the implementation phase. For the Commission, this means to supervise very large online platforms.

Thus, we are now – halfway through the 2020s – at a crucial crossroads. The Commission assumes its supervisory posture vis-à-vis very large online platforms and has hired new staffers to that end.[98] Member States establish Digital Services Coordinators that shall police all the other platforms. The DSA's broadness allows these actors to further specify and adapt the act's normative framework. To an extent, we know this from other regulatory fields like financial services, which fruitfully combine very specific obligations (e.g., capital requirements in absolute or relative numerical terms) with broad and comprehensive normative goals (e.g., preserving the stability of the financial system).[99] It is this practical, incremental balancing of specificity and broadness in the process of supervision that makes regulation effective. As Niamh Moloney noted,

[97] See instructively Omarova (n 7).

[98] See further Suzanne Vergnolle, 'Putting Collective Intelligence to the Enforcement of the Digital Services Act: Report on Possible Collaborations between the European Commission and Civil Society' Conservatoire National des Arts et Métiers (CNAM) Paper Series, 2023.

[99] See e.g. Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012 (Credit Requirement Regulation).

> [r]egulation does not operate in a vacuum; it must be operationalized through supervision, which is a 'hands on' business. Supervision requires granular engagement with firms and the taking of decisions which carry risk to the markets, the supervisor, and the tax-payer.[100]

Naturally, such skills are not learned over night and the required institutions to effectuate such supervision need time to grow. Thus, in conclusion, the Commission and the Digital Services Coordinators are in a good position to increase the DSA's specificity wherever needed.

2. Hire Engineers!

To guarantee such normative specificity, it is furthermore decisive to formulate – or clarify through subsequent guidelines and more – norms that can be implemented *through* technology. Currently, many rules imposed by the EU on technology companies originate predominantly from debates among lawyers and political scientists in the Commission, the Council, and the Parliament.[101] Doubtless, those norms drew on various hearings and consultations. Nonetheless, the lawmaking process remains dominated by lawyers.[102] To a large extent, lawyers, political scientists and, most importantly, elected politicians are necessary craftspeople moulding political will into legislative material. However, judging from the impact assessment and the legislative material of the DSA, many legislative discussions focus(ed) perhaps too much on wishful thinking rather than their 'technological

[100] Moloney made this remark regarding the EU's financial markets supervision. Although that is a distinctly different field to that covered by the DSA, the former may hold valuable lessons. After all, financial markets and its institutions were the big topic in the decade following the 2008 financial crisis. There, the EU apparently managed to establish a potent regulatory and supervisory regime, see further Niamh Moloney, *EU Securities and Financial Markets Regulation* (Oxford University Press 2023) 944.

[101] The EU lawmaking process is not exactly known for its transparency, so it is hard to assess which voices were indeed decisive, but the dominance of lawyers in the Commission and the Council directorate – especially the powerful legal services – arguably plays a large role. See in general Päivi Leino-Sandberg, *The Politics of Legal Expertise in EU Policy-Making* (Cambridge University Press 2021).

[102] ibid.

substance', to borrow Fleischer's terminology.[103] By now European lawmakers produced hundreds of pages of text, speeches, policy briefs and legislation, detailing how technology, platforms, and content moderation should function. However, we still know relatively little about if and how the proposed regulations are supposed to be implemented. As shown above, the Commission itself thinks implementation of many provisions will come at zero costs – which begs the question either, whether the provisions are indeed ambitious enough or, why platforms are not already compliant if abiding by allegedly 'strict' European standards comes so cheap.

To ensure the DSA functions effectively, regulators and supervisors may need to adopt a more pragmatic approach to faithfully implementing its regulatory demands. If the legal language proves overly convoluted, they should instead focus on aligning implementation with the publicly stated objectives and purposes of the legislation. Achieving this, however, may require a greater reliance on independent technological expertise. Ideally, independent technological expertise already informs the legislative process and yields norms that are technically easily to implement and whose implementation is easily verifiable.

At this point I shall note what I mean by expertise and how its expansion would improve regulation and implementation. That is because expertise itself is not an objective given but emerges continually through processes and is subject to constant reconfiguration. These reconfigurations may be intentionally captured, e.g. through concerted industry efforts, or unintentionally shaped through normative predispositions, e.g. value-judgments about which population groups count as vulnerable. Bluntly put, knowledge is not simply there but emerges through means of production, it is, therefore, never independent, neutral or objective.[104] That is particularly

[103] Fleischer speaks of 'economic substance' in the context of regulatory arbitrage, see idem (n 91) 229.

[104] Michel Foucault, *The Archaeology of Knowledge* (Pantheon Books 1972); Robert Proctor, *Value-Free Science?: Purity and Power in Modern Knowledge* (Harvard University Press 1991);

relevant whenever knowledge is concentrated in, or perhaps even monopolized by the very organization one intends to regulate. Therefore, calling for more and better 'independent' expertise in regulating digital corporations is a simplified way of calling for a more reflexive and continuous engagement with the type of experts and expertise we want in our regulatory actors.[105] Exploring this deeper, however, goes beyond the scope of this article.

Also, later implementation may further specify some unclear normative goals. Therefore, the Commission is in a crucial position to ramp up its own technical expertise and critically scrutinize whatever platforms sell as DSA compliant. In that sense, the somewhat unassuming Article 40 DSA, which obliges platforms to grant 'vetted' researchers access to internal data, may end up being one the most consequential provisions of the whole regulation.

VIII. Conclusion

This article scrutinized whether a prevalent European regulatory trend – formulating lofty, public-law inspired normative goals and procedural demands and leaving regulated actors do the rest – formulates strict enough standards to reverberate globally or, to use Anu Bradford's term, manifest the 'Brussels Effect'. Focusing on the DSA and using one of its key provisions, Article 14, as an example, the article argues that many standards are indeed anything but strict. Instead, they leave the decisive norm-making and institution building to the regulated entities themselves. While understandable for practical reasons, the norms and institutions that diffuse globally through such a procedure are not really those of the EU but rather whatever platforms construe as their compliance to those rules. Potentially,

Sheila Jasanoff, 'Beyond Epistemology: Relativism and Engagement in the Politics of Science' (1996) 26 Social Studies of Science 393.

[105] See especially the work of Slayton and Clark-Ginsberg who argue that better regulation requires 'negotiation and creation of new forms of expertise, which must balance several distinct public goods, including economy, reliability, and security', Rebecca Slayton and Aaron Clark-Ginsberg, 'Beyond Regulatory Capture: Coproducing Expertise for Critical Infrastructure Protection' (2018) 12 Regulation & Governance 115, 116.

some of said compliance practices may be superficial or, in the terminology adopted here, 'ceremonial'.

Currently, the Commission builds up its enforcement infrastructure vis-à-vis very large online platforms. It hires people, who will write frameworks, guidelines, and delegated norms. If those enforcers tap into technological and organizational expertise and increase the granularity and specificity of the DSA's regulatory demands, we may eventually see stringent standards that may, over time, reverberate globally.

DIGITAL HEALTH GOVERNANCE: ASEAN AND THE THREE NARRATIVES OF DIGITAL (IN)JUSTICE

Tsung-Ling Lee *

International law has provided limited formal responses to digital health technology challenges, as evidenced by the absence of binding legal agreements. Yet international law continues to shape digital health and influence global perspectives on digital health innovation. This article distills the complex global digital health discourse by presenting a conceptual framework of three competing narratives in digital health governance: technological solutionism, human rights, and data sovereignty. The technological solutionism narrative—frequently employed by international development agencies—portrays digital health innovations as solutions to healthcare access disparities and digital divides. While the human rights narrative critically challenges this view, the international human rights law framework has not adequately addressed power dynamics in digital health infrastructure ownership despite its aim to fill technological solutionism's normative gaps. Meanwhile, data

* Associate professor of Law, Graduate Institute of Biotechnology and Health Law, Taipei Medical University, Taiwan. The author thanks Stefania di Stefano, Rebecca Mignot-Mahdavi, Barrie Sander, Dimitri van den Meerssche, and Roxane Vatanparast for organizing the 2023 ESIL Interest Group on International Law and Technology and the participants of the workshop "Is Fairness in Digital Governance a Trap?" held at Aix-en-Provence, France. The author is especially grateful to Barrie Sander and Dimitri Van Den Meerssche for thoroughly reviewing the previous drafts, the two anonymous reviewers for the insightful comments and to the editorial team of the European Journal of Legal Studies—particularly Michael Widdowson and Cielia Eckardt—for their meticulous editorial work and support. The project also received generous support from Taiwan Ministry of Science and Technology MOST110-2636-H038-002. As usual, all errors are the sole responsibility of the author. The author can be reached at: tl265@georgetown.edu.

sovereignty has emerged as a counterforce to perceived Western dominance in the digital sphere and the US hegemony more broadly, with China playing a significant role in shaping this narrative. Through examining how the digital health discourse unfolds in the Association of Southeast Asian Nations (ASEAN), this article demonstrates that as the global digital health landscape grows in complexity, there is a pressing need to understand the discursive patterns that shape digital health innovations through international law—and the distribution of negative and positive externalities—within the global context.

Keywords: Digital health, Digital technology, ASEAN, International human rights law, Data sovereignty, Technological solutionism,

TABLE OF CONTENTS

I INTRODUCTION

The COVID-19 pandemic, one of the most significant global health challenges in recent history, affected all aspects of human life. The pandemic notably accelerated digital health technologies' use and social acceptance worldwide. Digital transformation of health care through the Internet of Things, telemedicine, digital health apps, smart wearables, electronic health records, artificial intelligence, and big data analytics have proven potential to improve health outcomes. It has been documented that these technologies have been pivotal in combating the pandemic, from tracking the spread of the virus to facilitating remote health consultations and treatments via telemedicine.[1] Likewise, digital health innovations play a significant role in health systems by using electronic health records and population health data to monitor and inform evidence-based public health policies and integrate digital tools such as telemedicine into routine clinical care.

Digital transformation is commonly viewed as a catalyst for advancing Sustainable Development Goals (SDGs) by governments and international developmental agencies.[2] According to the World Health Organization

[1] Dinesh Visva Gunasekeran and others, 'Digital Health during COVID-19: Lessons from Operationalising New Models of Care in Ophthalmology.' (2021) 3 The Lancet Digital Health; Daniel Shu Wei Ting and others, 'Digital Technology and COVID-19' (2020) 26 Nature Medicine 459.

[2] See e.g. 'Transforming for a Digital Future: 2022 to 2025 Roadmap for Digital and Data - Updated September 2023' (*GOV.UK*, 29 November 2023) <https://www.gov.uk/government/publications/roadmap-for-digital-and-data-2022-to-2025/transforming-for-a-digital-future-2022-to-2025-roadmap-for-digital-and-data> accessed 23 November 2024; Digital Transformation Agency, 'Digital Transformation Agency' (2024) <https://www.dta.gov.au/> accessed 23 November 2024; Asian Development Bank, 'Strategy 2030 Digital Technology Directional Guide: Supporting Inclusive Digital Transformation for Asia and the

(WHO), broadening the availability of digital technologies can bridge the digital divide and mitigate health disparities.[3] A considerable body of literature has argued that digital health technologies can advance universal health coverage by strengthening health systems.[4] Many commentators see digital health innovations as accelerating the progressive realization of health rights and improving resource allocation and coordination in health programs.[5] However, there are human rights concerns that disparities in access and quality of these digital health services could result in uneven health outcomes.[6] Furthermore, as control and ownership of digital health infrastructures largely belong to big technology companies or, in some

Pacific' <https://www.adb.org/documents/strategy-2030-digital-technology-directional-guide> accessed 23 November 2024.

[3] World Health Organization, *Global Strategy on Digital Health 2020-2025* (1st ed, World Health Organization 2021).

[4] See e.g. World Health Organization, 'Harness Digital Health for Universal Health Coverage' (*WHO,* 20 March 2023) <https://www.who.int/southeastasia/news/detail/20-03-2023-harness-digital-health-for-universal-health-coverage> accessed 24 January 2025; Steven van de Vijver and others, 'Digital Health for All: How Digital Health Could Reduce Inequality and Increase Universal Health Coverage' (2023) 9 Digital Health. 'Digital Health and Universal Health Coverage: Opportunities and Policy Considerations for Pacific Island Health Authorities' (*WHO,* 3 August 2022) <https://apo.who.int/publications/i/item/digital-health-and-universal-health-coverage-opportunities-and-policy-considerations-for-pacific-island-health-authorities> accessed 21 January 2024; 'The Case for Digital Health: Accelerating Progress to Achieve UHC' (*AI for Good*, 30 July 2021) <https://aiforgood.itu.int/event/the-case-for-digital-health-accelerating-progress-to-achieve-uhc/> accessed 21 January 2024.

[5] See e.g., Kasoju, N., Remya, N.S., Sasi, R. and others, 'Digital health: trends, opportunities and challenges in medical devices, pharma and biotechnology', (2023) 11 CSIT 11.

[6] The Lancet Digital Health, 'Digital Technologies: A New Determinant of Health' (2021) 3 The Lancet Digital Health 11 e684.

instances, governments, there are concerns about the power imbalance between providers and consumers.[7]

Crucial and frequently overlooked questions lie at the intersection of these competing interests: who holds the power to make critical decisions that shape the implementation and deployment of these technologies? Who reaps the benefits—both intended and unintended—from these technological interventions? Do relationships between governments, individuals, and the private sector reconfigure as these digital technologies permeate the social fabric? Should these relationships transform in response to these technologies? How are decisions about the risks that stem from digital health technologies distributed across populations? This paper distills the complexity surrounding the global digital health discourse by offering a conceptual framework that sheds light on these questions. Specifically, this paper delves into three distinct narratives within the digital health discourse: technological solutionism, human rights law, and data sovereignty. Individually, they shed light on the power dynamics that underpin crucial questions about decision-making processes in digital transformation, addressing a different aspect of the distribution of potential benefits and risks from digital health technologies across different populations and countries. Each narrative provides a unique analytical lens to grasp the complexity of the issue. Collectively, these narratives distill the complex landscape of global digital health governance.

While a significant body of international law literature has examined its influences on the global digital space, less attention has focused on its role in shaping the digital health landscape.[8] This oversight stems partly from

[7] Yang Chen and Amitava Banerjee, 'Improving the Digital Health of the Workforce in the COVID-19 Context: An Opportunity to Future-Proof Medical Training' (2020) 7 Future Healthcare Journal 189.

[8] See e.g. Emily Lee Jones, 'Digital Disruption: Artificial Intelligence and International Trade Policy' (2023) 39 Oxford Review of Economic Policy 70; Pedro A Villarreal,

international law's limited impact on digital health transformation discourse thus far. However, international law's influence is growing due to increased global economic integration and greater public and commercial interest in the expanded distribution of digital health technologies worldwide. The expansion of digital health applications and the digitalization of healthcare could lead to a more prominent role in international law in the future. These three narratives highlight different aspects of governance contestation as digital health technologies evolve, set against the backdrop of an evolving global legal order.

This paper clarifies the emerging and intersecting rationalities and practices of digital health governance by providing a conceptual framework and offers an original contribution to the global digital health discourse.

The paper uses the Association of Southeast Asian Nations (ASEAN) as a case study to illustrate whether international law facilitated these narratives, thereby shaping perceptions of digital health innovations.

The technological solutionism narrative, while not an international legal concept, portrays digital health innovations as remedies for healthcare access disparities and digital divides, commonly deployed by international developmental agencies. According to this view, disparities in access, availability, and quality of digital health tools and services are central concerns in advancing digital health. The World Bank, for instance, highlights the need to scale up digital infrastructure to maximize benefits

'International Law and Digital Disease Surveillance in Pandemics: On the Margins of Regulation' (2023) 24 German Law Journal 603; Colin J Carlson and Alexandra Phelan, 'International Law Reform for One Health Notifications' (2022) 400 The Lancet 462; Mira Burri, 'The Impact of Digitalization on Global Trade Law' (2023) 24 German Law Journal 551; Nina Sun and others, 'Human Rights and Digital Health Technologies' (2020) 22 Health and Human Rights 21; Stephen Gilbert and others, 'Citizen Data Sovereignty Is Key to Wearables and Wellness Data Reuse for the Common Good' (2024) 7 NPJ Digital Medicine.

from digital health.[9] It also underlines the potential discrimination that could arise from these technologies' usage.

Conversely, the human rights narrative challenges this view. Despite promises of the international human rights law narrative to address the normative gap in the technological solutionism narrative, it falls short in tackling the power imbalance in the ownership of digital health infrastructure. Ownership and control of these technologies increasingly reflect and extend the political ideologies of global powers, particularly between the United States and China. Consequently, the data sovereignty narrative has emerged as a counterforce to the perceived Western dominance in the digital sphere, with China shaping the narrative.[10]

---

[9] World Bank, 'DIGITAL-IN-HEALTH: Unlocking the Value for Everyone' (*World Bank*, 19 August 2023) <https://www.worldbank.org/en/topic/health/publication/digital-in-health-unlocking-the-value-for-everyone> accessed 21 January 2024.

[10] Aynne Kokas, *Trafficking Data: How China Is Winning the Battle for Digital Sovereignty* (Oxford University Press 2022). See also Johannes Thumfart, 'The norm development of digital sovereignty between China, Russia, the EU, and the US: From the late 1990s to the Covid-crisis 2020/21 as catalytic event' in Ronald Leenes, Paul De Hert and Dara Hallinan (eds), *Data Protection and Privacy, Volume 14: Enforcing Rights in a Changing World* (Bloomsbury 2023)1-44. Johannes Thumfart's analysis presents contrasting approaches between the United States and China regarding digital technologies. He notes that the United States adopts a liberal stance which assumes that unrestricted information flow will weaken authoritarian dictatorship and promote democratic values. This perspective contrasts sharply with China's position, which frames digital sovereignty as a matter of national security and strategic importance. China's distinctive approach, Thumfart explains, is deeply rooted in its self-perception as a post- and anti-colonial power. This identity profoundly shapes how China conceptualizes sovereignty, drawing from traditional Confucian philosophy—specifically the 'Tianxia' system of governance. This framework fundamentally departs from Western notions of sovereignty, where independent secular powers maintain peaceful coexistence through mutual recognition and respect for territorial boundaries.

These three distinct narratives related to digital health shed light on the different ways international law serves as a battleground for power, control, and knowledge. They reveal how various actors, each with unique perspectives and interests, compete over ideologies and control of the digital health realm, highlighting the complexity and fluidity of this field. This dynamic interplay shapes the digital health discourse and ultimately steers the course of digital health governance on a global scale.

Every narrative reflects the political-economic context from which it emerges, and the impacts of international law on these narratives vary.

For instance, intellectual property rights—such as patents and trade secrets—as a branch of international law have reinforced technological solutionism by establishing a legal framework for major technology companies, supporting their global growth since the 1990s.[11] Likewise, intellectual property rights have established and sustained the digital ecosystem by enabling the creation of digital health infrastructure and applications.[12] Governments have capitalized on the allure of digital health as the next transformative technology, which often neglects the power and control of these transnational technology companies that control, own, and set standards for how these digital health technologies are used.[13]

---

[11] Stephen Hilgartner, 'Intellectual Property and the Poitics of Emerging Technology: Inventors, Citizens, and Powers to Shape the Future', (2009) 84 Chi.-Kent L. Rev. 197.

[12] Sharifah Sekalala and Tatenda Chatikobo, 'Colonialism in the New Digital Health Agenda' (2024) 9 BMJ Global Health <https://gh.bmj.com/content/9/2/e014131> accessed 24 January 2025

[13] For example, Singapore's Healthier SG initiative, along with the Health Hub and Healthy 365 apps, form the country's digital health strategy. This strategy aims to promote preventive care and lifestyle changes by encouraging citizens to use digital health apps and third-party wearables, personalized e-services are provided through these digital technologies. See e.g., 'Speech by MDM Rahayu Mahzam Minister of State, Ministry of Health, At the Asia New Vision Forum, 26 Sep 2024' (*Singapore*

On the other hand, international human rights law seeks to counterbalance technological solutionism, emphasizing the diverse health impacts of digital technology usage. Meanwhile, data sovereignty has emerged from escalating geopolitical tensions over access to and control of digital health services and infrastructure. In this context, the concept of sovereignty in international law has been reinterpreted to counter the dominance of the liberal West.[14]

Ministry of Health, 26 September 2024) <https://www.moh.gov.sg/newsroom/speech-by-mdm-rahayu-mahzam--minister-of-state--ministry-of-health--at-the-asia-new-vision-forum--26-sep-2024> accessed 24 January 2025.

[14] Legal scholar Henry Gao observed that data sovereignty has undergone several iterations as part of China's global ambition, with the concept evolving from physical control of information technologies to control of the digital software layer and digital infrastructure. See Henry Gao, 'Data Regulation with Chinese Characteristics' in Mira Burri (ed) *Big Data and Global Trade Law* (Cambridge University Press 2021) 245, 248. Legal scholar Anqi Wang also notes that grounding data sovereignty in traditional notions of territorial sovereignty strengthens the Communist Party's control over digital technologies and infrastructure to protect national and ideological security. See Anqi Wang, 'Cyber Sovereignty at Its Boldest: A Chinese Perspective' (2020), 16 Ohio St. Tech. L.J. 395, 403; Protecting Internet Security, (*China.org*), <http://www.china.org.cn/government/whitepaper/2010-06/08/content_20207978.htm> last accessed Dec 11, 2024. Similarly, Anupam Chander and Haochen Sun argue that China invented 'digital sovereignty' as a way to consolidate Communist party control, maintain social order, and reinforce Socialist and Confucian values. According to them, China's concept of digital sovereignty is rooted in traditional notions of territorial sovereignty, functioning as a defense against the perceived hegemony of the West in cyberspace—where 'information freedom' is seen as a threat to 'Chinese ideological security' and a form of cultural imperialism. Similarly, Anupam Chander and Haochen Sun argue that China invented 'digital sovereignty' as a way to consolidate Communist party control, maintain social order, and reinforce Socialist and Confucian values. According to them, China's concept of digital sovereignty is rooted in traditional notions of territorial sovereignty, functioning as a defense against the perceived hegemony of the West in cyberspace—where 'information freedom' is seen as a

By using ASEAN as a case study, this article considers the complex history of colonial influences on the Southeast Asian region, the assertion of Asian values as resistance to perceived Western influence, and China's growing role in providing digital health infrastructure. ASEAN provides an interesting case study, especially considering the region's complex history of colonial influences, the assertion of Asian values as resistance to perceived Western influence, and China's growing role in providing digital health infrastructure.

The article consists of six parts. Part II sketches the technological solutionism narrative, which positively portrays digital health innovations as an essential component of the international development discourse. Part III provides an overview of the right to health vis-à-vis digital health, where the human rights narrative emerged as a counterbalance to the technological solutionism narrative. Part IV sketches the data sovereignty narrative where major powers compete over digital health infrastructures, leading to a reinterpretation of sovereignty. Part V focuses on ASEAN as a case study to demonstrate how these three narratives individually present an aspect of digital health governance. Part VI focuses on technological barriers and ICT infrastructure landscape in ASEAN and analyzes how these three different narratives unfold. Part VII concludes.

II TECHNOLOGICAL SOLUTIONISM NARRATIVE

The public intellectual Evgeny Morozov introduced 'technological solutionism' to describe the misplaced optimism in modern technologies to

threat to 'Chinese ideological security' and a form of cultural imperialism. Anupam Chander and Haochen Sun point out that in 2010, the Chinese State Council linked the control of internet access and content regulation as part of sovereignty, linking territorial with cyberspace as a nod to international law. See Anupam Chander and Haochen Sun, 'Introduction: Sovereignty 2.0' in Anupam Chander and Haochen Sun (eds), *Data Sovereignty: From the Digital Silk Road to the Return of the State* (OUP 2023).

combat complex, multifaceted social issues. He defines it as a perspective that frames 'all complex social situations as either neatly defined problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimized.'[15]

Over the past decade, technology solutionism has emerged as a powerful force driving international development policies, with its promises to address complex social, economic, and political challenges. Through the global reach of intergovernmental agencies, technological solutionism has shaped the global development policy discourse and influenced how countries approach digital health policies.

International Policy Discourse

Although technological solutionism is not an international legal concept per se, international organizations and governments worldwide have embraced and embedded forms of the concept into their policy documents. Accordingly, digital health technologies are often seen as a solution to social ills, which include alleviating poverty and advancing sustainable development goals. For example, the World Health Organization's (WHO) Global Initiative on Digital Health argues that 'Digital health is a proven accelerator to advance health outcomes and achieve Universal Health Coverage (UHC) and the health-related Sustainable Development Goals (SDGs).'[16] Similarly, the Group of Seven (G7) strongly advocates for governments to embrace the transformative effects of digital health in

[15] Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (First edition, PublicAffairs 2013)

[16] WHO, 'Executive Summary, Global Initiative on Digital Health' (*Global Initiative on Digital Health*, 31 July 2023) <https://cdn.who.int/media/docs/default-source/digital-health-documents/global-initiative-on-digital-health_executive-summary-31072023.pdf?sfvrsn=5282e32f_1> accessed 3 February 2025.

healthcare services, affirming that '[d]ata and cost-effective digital technology are key drivers of innovations in health services'.[17]

Technological solutionism often perpetuates the perception that technology is inherently value-neutral; this perception can sometimes mask and exacerbate underlying social issues that express discriminatory effects of digital health. Yet, intergovernmental agencies—entities created through treaties between state parties to solve common issues— often rely on the rhetoric of technological solutionism as a basis for advocating for harmonization of ASEAN's digital health policies. Similar reasoning is also found in international development financing institutions such as the Asian Development Bank (ADB), which depicts emerging technologies as opportunities to leapfrog traditional industrial development phases. For instance, an ADB report highlights how drones could help deliver medical supplies to remote regions with poor transport infrastructure.[18] However, without tackling the structural causes of ill health, overreliance on digital health tools could inadvertently exacerbate health inequalities rather than narrow them as promised by techno-solutionism.

While there is an increased awareness as well as a recognition of potential bias and discriminatory effects in deploying these digital technologies in policy documents, these issues are not addressed legally. Instead, international development banks tend to focus on technical and regulatory issues that hinder the operations of the digital economy. The World Bank, for instance, has called for 'international regulatory cooperation' as a method

[17] G7 Research Group, 'G7 Nagasaki Health Ministers' Communiqué' (University of Toronto, 14 May 2023) <http://www.g7.utoronto.ca/healthmins/230514-communique.html> accessed 24 January 2025.

[18] Asian Development Bank, 'ASEAN 4.0: What Does the Fourth Industrial Revolution Mean for Regional Economic Integration?' (*Asian Development Bank,* November 2017) <https://www.adb.org/publications/asean-fourth-industrial-revolution-regional-economic-integration> accessed 5 August 2024.

to strengthen the digital economy through the development of principles in areas such as taxation, cybercrime, and cybersecurity.[19]

A World Bank policy document, for example, acknowledges that 'digital technology can exacerbate existing inequalities due to prevailing gaps in access and availability' and also notes that social norms constrain women's access to digital technologies.[20] Yet, it recommends a 'customized approach (…) to develop coalitions with local leaders and engage with household dynamics'.[21] Thus, the access question is treated as a private matter rather than a public one that would require social and institutional changes.

Treating unequal access and availability of digital health tools as a private matter is, however, turning these supposedly emancipatory tools into instruments of constraint. For example, by ignoring the gender dynamics that impact digital access, we can miss opportunities to reach key populations and worsen the gendered effects of digital tools. Without tackling the social and physical barriers that prevent women from having equal access to these tools, it could further undermine women's agency and autonomy.[22]

Nonetheless, optimism in technology is not a new phenomenon. Scholars have traced the origins of techno-solutionism to the 'New Economy' of the late 1990s and 2000s.[23] In the time of the New Economy, the neoliberal ideology, which advocated for market-based technology policies,

[19] Anita Gurumurthy and Nandini Chami, 'Towards a Global Digital Constitutionalism: A Radical New Agenda for UN75' (2021) 64 Development (Society for International Development) 29.

[20] Clara Aranda Jan and Qursum Qasim, *Increasing Access to Technology for Inclusion* (World Bank, Washington DC 2023) <https://hdl.handle.net/10986/39495> accessed 24 January 2025.

[21] ibid.

[22] AS George and others, 'Gender Dynamics in Digital Health: Overcoming Blind Spots and Biases to Seize Opportunities and Responsibilities for Transformative Health Systems' (2018) 40 Journal of Public Health ii6.

[23] Dillon Wamsley and Benjamin Chin-Yee, 'COVID-19, Digital Health Technology and the Politics of the Unprecedented' (2021) 8 Big Data & Society.

flourished.[24] Private sector innovations became the driving force behind the digital economy during this period. In the 1990s, US-led international trade agreements further cemented global technology infrastructure by reinforcing intellectual property rights, benefitting US-based technology companies. During this era, technology was widely viewed as a solution to complex, multilayered social issues.[25]

Digital health tools became widely embraced by governments during the COVID-19 pandemic chiefly because of their ease of use. This adoption further embedded technological solutionism in national digital policies and gained salience among international development agencies.

The COVID-19 Pandemic

This belief persists today, as seen during the COVID-19 pandemic when governments worldwide have used digital tools to manage and mitigate the spread of SARS-CoV-2.[26] The common rhetoric of techno-solutionism has further encouraged countries to speed up digitalization even after the pandemic has ended. During the pandemic, various digital health tools were utilized, such as mobile phones for digital contact tracing, remote care through telemedicine, and the adoption of electronic medical records to support social distancing and health systems. For example, in Vietnam, the

[24] A Fremstad and Mark Paul, 'Neoliberalism and Climate Change: How the Free-Market Myth Has Prevented Climate Action' (2022) 197 Ecological economics 107353.

[25] Gurumurthy and Chami (n 19) 35.

[26] See e.g. Cristina Valencia and others, 'Adoption of Digital Tools in the Context of the COVID-19 Pandemic in the Region of the Americas - The Go.Data Experience' (2022) 16 The Lancet Regional Health Americas; Amitabh B Suthar and others, 'Lessons Learnt from Implementation of the International Health Regulations: A Systematic Review' (2018) 96 Bulletin of the World Health Organization 110; Steve Davis and Pardis Sabeti, 'Digital Health Tools for Pandemic Preparedness', (*Brookings*, 28 December 2021) <https://www.brookings.edu/articles/digital-health-tools-for-pandemic-preparedness/> accessed 21 January 2024.

government used digital tools for public health surveillance, telemedicine, public health communication, and artificial intelligence to aid in diagnosis and treatment.[27] However, despite the wide adoption of digital tools, the effectiveness and utilities of these interventions remain contested.[28]

In recent years, digital health has gained significant attention as a means of improving healthcare outcomes globally. International intergovernmental organizations, including the World Bank, actively promote digital transformation as part of international development discourse. For instance, the World Bank advocates for governments to embrace digitalization in healthcare, where '[d]igital technologies have the potential to bring significant value to health systems (…) Technology and data (…) are catalytic components of the current wave of health system changes'.[29] These global entities assist countries in adopting digital technologies, services, and infrastructure, as well as open, universal governance frameworks. [30] Developmental agencies accomplish this through a variety of initiatives aimed at modernizing societies. [31] For instance, in ASEAN, the Asia Development Bank funds the Community of Interoperability Lab to address the challenge of health information interoperability. [32] This technical

[27] Emnet Getachew and others, 'Digital Health in the Era of COVID-19: Reshaping the next Generation of Healthcare' (2023) 11 Frontiers in Public Health 942703.

[28] United Nations Development Programme, 'United Nations Development Programme - Digital Strategy 2022-2025' (*UNDP,* 2022) <https://digitalstrategy.undp.org/> accessed 24 January 2025.

[29] World Bank, 'Digital in Health: Unlocking the Value for Everyone - Summary (English)' (*World Bank*, 28 November 2023) <http://documents.worldbank.org/curated/en/099112823142531926/P1750750fb6ae70740acd203c7a25e55a43> accessed 24 January 2025.

[30] 'DIGITAL-IN-HEALTH: Unlocking the Value for Everyone' (n 10).

[31] World Bank, 'Digital Transformation' (*World Bank,* 14 October 2024) <https://www.worldbank.org/en/topic/digital/overview > accessed 3 February 2025.

[32] 'What Is an Interoperability Lab? (*Standards and Interoperability Lab – Asia*, 30 March 2020) <http://sil-asia.org/what-is-an-interoperability-lab/> accessed 15 January 2024.

assistance can be viewed as an expression of technological solutionism—where technical assistance is offered without consideration of broader ethical or human rights impacts.

Similarly, at the global level, technological solutionism permeates the work of other international organizations. One prime example is the WHO, where health promotion lies at the core of its constitutional mandate and defines its institutional identity within the UN system.[33] Yet, strikingly, the WHO's promotion of digital health acknowledges the digital divide but does not actively address the structural factors that underpin the digital divide across and within countries.[34] Instead, the organization tends to promote the use of information and communication technologies (ICTs) to advance the global adoption of digital health technologies. This preference for techno-solutionism, where digital health tools are widely perceived as a path to universal health coverage, aligns with the WHO's technical culture. However, this approach misses an opportunity to address not only the digital divide affecting access but also fails to acknowledge the geopolitical tension that underpins major powers in the battle for control and access to critical digital technology infrastructure. Such avoidance may be attributable to the diplomatic nature of the WHO, which often avoids—or directly conforms with—politically charged issues.[35] In the same way, the UN Human Rights Council has made considerable efforts to tackle the digital divide through

[33] International Health Conference, 'Constitution of the World Health Organization 1946' (2002) 80 Bulletin of the World Health Organization 983-984.

[34] Bernardo Mariano, 'Towards a Global Strategy on Digital Health' (2020) 98 Bulletin of The World Health Organization 231.

[35] See e.g. Tsung-Ling Lee, 'Informal Rulemaking at the World Health Organization: Technocratic, Iterative, and Political Constraints' International Organizations Law Review, (forthcoming 2025).

special rapporteurs,[36] but the geopolitical aspect of digital infrastructure remains unaddressed.

As such, the international development discourse surrounding digital transformation often presents it in an emancipatory language, with promises of addressing existing health inequalities by offering technological solutions to deeply rooted social and political conditions.[37] While it is true that digital health innovations have the potential to revolutionize healthcare, it is important to recognize that this portrayal may conceal the underlying social and political factors that create health disparities.[38]

An AI system, including machine learning, uses statistical and mathematical modeling to analyze data. For the purposes of this paper, I adopt the 2023 version of the updated OECD definition of an AI system, which is defined as 'a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments'.[39]

---

[36] United Nations Human Rights Council, 'Digital innovation, technologies and the right to health, Report of the Special Rapporteur on the Right to Health' UN Doc. A/HRC/53/65 (2023).

[37] See e.g. Poonam Khetrapal Singh and Mark Landry, 'Harnessing the Potential of Digital Health in the WHO South-East Asia Region: Sustaining What Works, Accelerating Scale-up and Innovating Frontier Technologies' (2019) 8 WHO South-East Asia Journal of Public Health 67-70.

[38] James Shaw and Wiljeana J Glover, 'The Political Economy of Digital Health Equity: Structural Analysis' (2024) 26 Journal of Medical Internet Research; Tereza Hendl and Ayush Shukla, 'Can Digital Health Democratize Health Care?' (2024) 38 Bioethics 491; Hannah E Knight and others, 'Challenging Racism in the Use of Health Data' (2021) 3 The Lancet Digital Health e144.

[39] Stuart Russell, Karine Perset, and Marko Grobelnik, 'Updates to the OECD's Definition of an AI System Explained' (*OECD.AI*, 29 November 2023) <https://oecd.ai/en/wonk/ai-system-definition-update> accessed 24 September 2024.

In contrast, generative AI refers to machine learning models that are trained to make new data that generate new, original texts, content, images, or objects based on patterns and structures of existing data.[40]

Studies have shown that machine learning could more precisely diagnose tumors, improve surgery accuracy, and, as a result, lead to better population health outcomes.[41] It is important to note that when using AI algorithms in healthcare, the outcome of the prediction is probabilistic, not deterministic. Furthermore, the prediction and precision of machine learning capabilities are limited by the quality and availability of existing health datasets, which often lack data for women and marginalized groups, leading to skewed diagnoses for these populations.[42] Using an algorithm as part of healthcare operations can also mirror the programmers' hidden assumptions and biases, potentially perpetuating these biases and, at worst, discriminatory practices.

In essence, the use of machine learning that draws from health data carries a potential risk of perpetuating bias and discrimination if not appropriately addressed early during the development and deployment phase.[43] For instance, the deployment of software in HIV response has already been impacted by data gaps, particularly for marginalized, stigmatized, and

[40] Tara Templin and others, 'Addressing 6 Challenges in Generative AI for Digital Health: A Scoping Review' (2024) 3 PLOS Digital Health.

[41] See e.g., Dow-Mu Koh and others, 'Artificial Intelligence and Machine Learning in Cancer Imaging' (2022) 2 Communications Medicine 133.

[42] Morgan King, 'Harmful biases in artificial intelligence' (2022) 9 The Lancet Psychiatry 48.

[43] However, it should be noted that other AI systems that operate without solely relying on data, such as personalized or precision medicine that uses AI systems combined with genomic technology, provide accurate results but do not predict diseases. See Shawneequa L Callier and others 'Ethical, Legal, And Social Implications Of Personalized Genomic Medicine Research: Current Literature And Suggestions For the Future' (2016) 30 Bioethics 698.

criminalized communities that are not recorded in official health data.[44] In the United States, data collected through HIV/AIDS surveillance serves as a basis for resource allocation. However, this data might be incomplete because structural barriers prevent individuals from accessing healthcare.[45] As the world moves towards digitalization, digital health technologies, such as AI and smart wearable devices, which rely on data to power their application, may inadvertently perpetuate health inequalities instead of reducing them. Using datasets that are not representative of populations could lead to inaccurate generalization by AI systems, which might be prone to bias against women and minorities due to a lack of data.

The promises and allures of digital technologies can overshadow their potential negative effects, to which the article turns next.

Unintentional Digital Health Policy Impacts

While public health crises often speed up technology adoption, these tools have historically been introduced in states that lack the capacity to invest in public health infrastructure and to implement proven public health interventions: policy measures implemented by governmental health departments that aim at improving the population's mental and physical health. Viewing digital health technologies from a technological solutionist perspective can shift attention and resources away from publicly funded, proven public health interventions in resource-limited ASEAN countries. Similarly, the allure of emerging digital technologies, such as AI systems or generative AI, can lead to over-estimation of benefits and dismissal of potential challenges, resulting in unbalanced healthcare policies and misguided investments. This optimism about the benefits of digital health can particularly affect resource-limited countries and high-income countries that are under pressure to cut healthcare spending.

---

[44] Benjamin K Ngugi and others, 'Data Quality Shortcomings with the US HIV/AIDS Surveillance System' (2019) 25 Health Informatics Journal 304.

[45] ibid.

For instance, a significant body of research has shown that the COVID-19 pandemic has disproportionately impacted vulnerable and marginalized populations due to gaps in wealth, housing, and access to healthcare.[46] Thus, the increasing adoption of digital health technologies is not without ethical, regulatory, and legal challenges. These arise, in part, from the disparate impacts of these digital health innovations.[47] The shift towards digital health has created new health vulnerabilities and risks, potentially leading to differential health outcomes. For example, the use of digital contact tracing apps during the pandemic has been linked to increased stigma and discrimination against women and marginalized populations.[48] In South Korea, the government published information about travel histories of confirmed, anonymous individuals, including limited information on gender and age groups. However, once this information was available in the public domain, it inadvertently exposed private lives that fueled social stigma

[46] Don Bambino Geno Tai and others, 'Disproportionate Impact of COVID-19 on Racial and Ethnic Minority Groups in the United States: A 2021 Update' (2021) Journal of racial and ethnic health disparities 2334.

[47] Various implications of new technologies have been addressed by the Human Rights Council, including the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, see e.g. Frank La Rue and UN Human Rights Council Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (20213) <https://digitallibrary.un.org/record/756267> accessed 11 August 2023.

[48] Sara L M Davis, 'Contact Tracing Apps: Extra Risks for Women and Marginalized Groups' (*Health and Human Rights Journal*, 29 April 2020), <https://www.hhrjournal.org/2020/04/contact-tracing-apps-extra-risks-for-women-and-marginalized-groups/> accessed 3 August 2023.

as speculations and public imagination grew over the activities and behaviors of those infected.[49]

The digital divide in health has also led to disparities in health outcomes, reflecting differences in digital access, digital skill levels, and the availability of digital health infrastructure within and across countries. Unequal access to health technologies and inadequate digital literacy can exacerbate health disparities, further marginalizing vulnerable populations.[50]

If these underlying structural causes of ill health are not adequately addressed, digital health innovations could exacerbate the health disparities they aim to alleviate. However, addressing these structural health determinants is a political choice in liberal democracies—a point to which human rights law draws attention, which the ensuing section explores.

III International Human Rights Law & The Right to Health

Relevant to the discussion on the digital health discourse is the right to health, embodied in the WHO Constitution, which envisages 'The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition'.[51] Article 25(1) of the Universal Declaration of Human Rights affirms: 'Everyone has the right to a standard of living adequate for the health of himself and of his family, including food, clothing, housing and medical care and necessary social services'.[52]

---

[49] Nemo Kim, 'More Scary than Coronavirus": South Korea's Health Alerts Expose Private Lives' (*The Guardian,* 6 March 2020) <https://www.theguardian.com/world/2020/mar/06/more-scary-than-coronavirus-south-koreas-health-alerts-expose-private-lives> accessed 22 October 2024.

[50] Elisabeth Beaunoyer, Sophie Dupéré and Matthieu J Guitton, 'COVID-19 and Digital Inequalities: Reciprocal Impacts and Mitigation Strategies' (2020) 111 Computers in Human Behavior 106424.

[51] Preamble, Constitution of the World Health Organization 1948.

[52] Universal Declaration of Human Rights 1948 art 25(1).

Articles 2(2) and 12 of the International Covenant on Economic, Social and Cultural Rights (ICESCR) proscribe any discrimination in access to health care and recognize states' obligations to address the underlying determinants of health. Furthermore, the right to health is recognized in Article 24 of the Convention on the Rights of the Child, Article 12 of the Convention on the Elimination of All Forms of Discrimination against Women[53], and Article 25 of the Convention on the Rights of Persons with Disabilities.[54]

Digital Divide and the AAAQ

Fair digital health transformation requires an equitable distribution of the benefits of technology, which can mitigate the digital divide that reflects socioeconomic disparities across populations, education levels, generations, and genders. Studies have shown that individuals with higher education levels have higher uptake of digital technologies and use of health-related applications.[55] Strikingly, health disparities also widen as digital technologies advance—individuals who are more adaptable to new technologies are more likely to reap health benefits from these digital tools.[56]

The health divide can be narrowed by making health facilities, goods, and services widely available, accessible, acceptable, and of good quality (AAAQ).

---

[53] UN General Assembly, Convention on the Elimination of All Forms of Discrimination Against Women, United Nations, Treaty Series, vol. 1249, p. 13, 18 December 1979, art 12.

[54] UN General Assembly, Convention on the Rights of Persons with Disabilities: resolution / adopted by the General Assembly, A/RES/61/106, 24 January 2007 art 25.

[55] Dorothy Szinay and others, 'Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review' (2020) 22 Journal of Medical Internet Research.

[56] See e.g. Sarah Elgazzar, Joanne Yoong and Eric Finkelstein, 'Digital Health as an Enabler of Healthy Aging in Southeast Asia' *(Duke NUS,* 28 September 2020) <https://www.duke-nus.edu.sg/core/think-tank/news/publications/digital-health-as-an-enabler-of-healthy-aging-in-southeast-asia> accessed 11 December 2024.

The Committee on Economic, Social and Cultural Rights states that functioning public health and healthcare facilities, goods, services, and programs must be sufficiently available within a state party.[57] While the nature of the facilities, goods, and services may vary, depending on various considerations, digital health technologies can increase points of contact between healthcare providers and patients, thereby expanding the availability of health services.

Likewise, digital accessibility can be understood in terms of non-discrimination, physical accessibility, economic accessibility, and information accessibility. These four dimensions help address the challenge of digital exclusion, broadly classified into three categories: Lack of access due to economic costs, lack of motivation to connect, and lack of digital skills and education. The principle of non-discrimination demands that health services be available and accessible irrespective of social and economic circumstances.

As healthcare systems expand to the digital realm, older persons might be excluded from accessing care due to a lack of connectivity, digital literacy, or ability to view and navigate online platforms.[58] Lack of perceived benefits from digital health can hinder older people from taking up digital health tools.[59] Limited digital healthcare applications, platforms, and technologies are designed specifically with older users in mind, which can further restrict access and diminish the willingness of older people to use these digital tools.[60] While many jurisdictions provide intermediaries to assist older persons with

[57] Mariano (n 34).

[58] See e.g. Tshepo Mokuedi Rasekaba and others, 'Exploring Telehealth Readiness in a Resource Limited Setting: Digital and Health Literacy among Older People in Rural India (DAHLIA)' (2022) 7 Geriatrics 28.

[59] Emma Kainiemi and others, 'Perceived Benefits of Digital Health and Social Services among Older Adults: A Population-Based Cross-Sectional Survey' (2023) 9 Digital Health.

[60] Chuanrui Chen, Shichao Ding and Joseph Wang, 'Digital Health for Aging Populations' (2023) 29 Nature Medicine 1623.

access to these digital services, the services are generally unavailable in resource-poor areas.[61]

Similarly, individuals with disabilities often encounter digital exclusion due to limited access to digital services and infrastructure. Research has shown that digital health interventions can improve access to care, mitigate psychosocial distress, improve quality of care, and alleviate caregiver burden for patients living with chronic illnesses and their informal caregivers.[62] However, the failure to address issues such as digital access, connectivity, and literacy may result in older individuals, people with disabilities, and their informal caregivers being unable to reap the benefit of digital health technologies.

According to the UN, digitalization refers to the process of converting analog information into digital form.[63] Varying levels of digitalization can create uneven digital access across populations, as rural and remote areas tend to receive less investment and bear higher costs for digital infrastructure. As public and health services become increasingly digitalized, those who lack digital access face disadvantages that can negatively affect their health and well-being. The digital exclusion of older people and their caregivers, as well as people living with disabilities, can be concerning on a population level.[64]

---

[61] ibid.

[62] Yunhuan Li and others, 'The Effectiveness of E-Health Interventions on Caregiver Burden, Depression, and Quality of Life in Informal Caregivers of Patients with Cancer: A Systematic Review and Meta-Analysis of Randomized Controlled Trials' (2022) 127 International Journal of Nursing Studies.

[63] Gregory Smith and Wyatt Achong, 'From Digitisation to Transformation: Understanding Digital Government Part 1' (*UNDP*, 2024) <https://www.undp.org/trinidad-and-tobago/blog/digitisation-transformation-understanding-digital-government-part-1> accessed 11 December 2024.

[64] Siddig Fageir, Omer Osman and Clifton Addison, 'A Closer Look at Dementia Patients' Barriers to Telemedicine Utilization during the COVID-19 Pandemic' (2023) 7 European Journal of Environment and Public Health.

The disparity of digitalization also exists across countries. The level of digitalization varies significantly across countries, where digital exclusion mirrors these variations. For example, digital exclusion affects 24% of the population in Denmark, 65.5% in Mexico, and 96.9% in China. Empirical studies have found that digital exclusion correlates with poorer health outcomes, with older persons facing higher risks because they are less likely to use digital tools.[65] This pattern holds true across both middle-income countries like Mexico and China and high-income countries like Denmark.

The Special Rapporteur on Health cautions that the allure of digital innovations and technologies in health services should not lead to decreased investment in physical medical facilities and services.[66] Instead, digital tools should be utilized to better meet the specific needs of individuals who require reasonable assistance due to their disabilities. Health services, such as telemedicine or other types of remote care, should be made affordable to ensure economic accessibility. This affordability should extend to digital devices, such as mobile phones, to which women of lower economic status tend to have less access. In terms of informational accessibility, sensitive health data, such as certain sexual orientations and practices, healthcare procedures that are criminalized, and health status, such as HIV/AIDS, must be kept confidential.[67] Insofar as digital health tools may perpetuate inequities and bias due to a lack of diversity in the innovation processes, more inclusive design processes that reflect local needs are desirable.

The Draft General Recommendation no. 37 on Racial Discrimination in the Enjoyment of the Right to Health specifically cautions on the use of artificial intelligence and racial discrimination in the context of health, where

[65] Xin Lu, Yao Yao and Yinzi Jin, 'Digital Exclusion and Functional Dependence in Older People: Findings from Five Longitudinal Cohort Studies' (2022) 54 eClinical Medicine 101708.

[66] Mofokeng (n 36) para 15.

[67] 'What is an Interoperability Lab?' (n 31) para 40.

> racial discrimination may permeate artificial intelligence through electronic health records and machine learning algorithms, while their use in health is increasing. As in other fields, details about their development remain largely unknown and lack of transparency does not allow any adjustment by health providers in practice. [68] However, studies demonstrate how clinical algorithms reproduce structural inequalities outcomes in hospitals by translating them into health indicators. As just one of the examples, studies show that the algorithm using health costs as a proxy for health needs, reproduces bias based on how money is spent on patients of African descent who have the same level of need, and the algorithm thus falsely concludes that they are healthier than other equally sick patients. Due to missing data, an AI algorithm that depends on genetic test results is more likely to mischaracterize the risk of breast cancer for patients protected under the Convention. Bias is also instilled through studies that do not challenge embedded racial ideologies and fail to assess the synergies between psychosocial, genetic, and environmental factors in explaining differences in health outcomes, such as hypertension.[69]

In short, these two competing narratives—technological solutionism and human rights—have been part of framing the adoption of digital health technologies, particularly the use of AI in healthcare, where medical decisions can be made in a black box. Under the technological solutionism narrative, digital innovations are widely seen as an opportunity to bridge the health divide and accelerate the implementation of universal health coverage. One common scenario used in the technological solutionism narrative is the use of digital health technologies in rural areas, where it is argued that digital

[68] European Union Agency for Fundamental Rights., *Getting the Future Right :Artificial Intelligence and Fundamental Rights : Report* (EU Publications Office 2020) <https://data.europa.eu/doi/10.2811/58563> accessed 17 January 2025.

[69] Insa M Schmidt and Sushrut S Waikar, 'Separate and Unequal: Race-Based Algorithms and Implications for Nephrology' (2021) 32 Journal of the American Society of Nephrology 529. However, it should be noted that the challenges for genetic testing using AI algorithms lie in data interpreting, not the quality of data available. Data interpretation remains a technical challenge for the genetic testing industry.

health innovations could help alleviate the chronic shortage of healthcare workers, improve healthcare quality, and strengthen health services.[70] On the other hand, without addressing the broader human rights implications of digital health technologies, it could undermine these genuine benefits of digital health and perpetuate existing structural inequalities, an aspect that is highlighted in the human rights narrative. AI in healthcare warrants further discussion as the field has attracted considerable investment, but regulations around the world remain patchworked and fragmented, which is also a consistent concern in ASEAN. [71] In addition to the potentially discriminatory impacts of digital health solutions mentioned above, addressing the healthcare worker shortage would need to examine why the shortage occurs, including, for instance, working conditions and whether there are sufficient legal protections for healthcare workers. Similarly, improving healthcare quality and strengthening health services would depend on whether a country invests adequate and sustainable funding into the health system. Indeed, several human rights reports and resolutions on technology have highlighted the nefarious implications of digital health, including concerns over the right to privacy and the right to health, as well as questions over good governance, such as transparency, accountability, discrimination, and more.[72]

Despite the concerted efforts aimed at reducing health disparities by drawing attention to the potential discriminatory effects and biases generated through digital innovations, it is crucial to acknowledge that human rights law, while providing an important framework, has its own set of conceptual and normative limitations. This is primarily due to the fact that human rights law

---

[70] OECD, *Health Data Governance for the Digital Age: Implementing the OECD Recommendation on Health Data Governance* (OECD Publishing 2022).

[71] Jessica Morley and others, 'Governing Data and Artificial Intelligence for Health Care: Developing an International Understanding' (*JMIR formative research*, 1 January 2022) <https://formative.jmir.org/2022/1/e31623/PDF> accessed 11 December 2024.

[72] 'What is an Interoperability Lab?' (n 31)

could fail to address the larger global structural factors that give rise to ill health, such as an oppressive political system, a government's austerity measures, the negative impacts of international trade agreements on public health, or broader geopolitical power dynamics. These political, economic, and social dynamics influence control over and access to digital technologies within the complex landscape of the political economy of digital health.

For instance, in the realm of data governance that underpins digital health, legal scholars Angelina Fisher and Thomas Streinz argue that data inequality is a function of infrastructural control.[73] Data is missing, not only because of the lack of infrastructure but because the absence reflects a deliberate byproduct of economic, social, and political choices. Feminist scholars Lauren F. Klein and Miriam Posner note that 'data sets never arrive in the world fully formed but are assembled from tangles of historical forces and ideological motivations, as well as practical concerns'.[74]

Scholars Catherine D'lgnazio and Lisa F. Klein observe that the 'phenomenon of missing data is a regular and expected outcome in all societies characterized by unequal power relations'.[75] Likewise, professors Fisher and Streinz warn that attempts to produce data to overcome the lack of data might also risk 'reproducing and accelerating inequalities of power relations that are embedded in the choices about what has become (and what was excluded from becoming) data'.[76]

This is particularly concerning, as the digital infrastructure responsible for producing and collecting this data is privately owned. The former Special

[73] Angelina Fischer and Thomas Streinz, Confronting Data Inequality 60(3) Columbia Journal of Transnational Law 829-956 (2022) World Development Report 2021 background paper, IILJ Working Paper 2021/1, NYU School of Law, Public Law Research Paper No. 21-22,

[74] Miriam Posner and Lauren F Klein, 'Editor's Introduction' (2017) 3 Feminist Media Histories 1.

[75] Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (The MIT Press 2020) 18.

[76] Fischer and Streinz (n 75) 12.

Rapporteur on the Rights of Persons in Extreme Poverty, Philip Alston, argued that digitization might become a Trojan Horse where industry takes on public health services.[77] Privatization of digital infrastructure becomes concerning mainly as these commercial actors are driven by profits rather than public interests. As a result, public services that lack commercial appeal are often neglected. For instance, it is unlikely that big technological companies will invest time and finances in developing sexual and reproductive health information in the local languages for sex workers.[78] As the world embraces digitalization, reducing the information asymmetry of the public remains a much-neglected endeavor.

Data politics

Furthermore, the notion of 'data colonialism' has attracted significant attention in the literature regarding claims of health data ownership. The Special Rapporteur on the Right to Health has observed that extracting data without consent from the Global South to the Global North for profit-seeking purposes has perpetuated 'colonial dynamics in technology and digital tools that extend to the present'.[79] Dr. Kadija Ferryman argues that data colonialism is exploitative, mainly because data functions as a valuable resource that is removed from its places of origin, altered, and transformed into products with marginal benefits to those at its place of creation.[80]

Large global digital platforms and search engines, such as Google and Facebook, significantly influence how data is gathered, accumulated, stored, transferred, and used. This fact highlights the importance of states,

[77] Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, 'Report of the Special Rapporteur on extreme poverty and human rights (2019) A/74/493.

[78] Nina Sun and others, 'Human Rights and Digital Health Technologies' (2020) 22 Health and Human Rights 21.

[79] 'Digital innovation, technologies and the right to health, Report of the Special Rapporteur on the Right to Health' (n 36).

[80] Kadija Ferryman, 'The Dangers of Data Colonialism in Precision Public Health' (2021) 12 Global Policy 90.

development agencies and individuals to consider data and its constitutive infrastructure. In particular, these big technology companies can influence health-related decision-making. This was seen during the COVID-19 pandemic when social media accelerated the spread of misinformation and disinformation about the safety and efficacy of vaccines.[81] Misinformation—information shared without intent to cause harm—can adversely affect health outcomes through inaccurate portrayals of vaccinations and side effects, for instance.[82] One study found that COVID-19 vaccine misinformation spread faster than those that were fact-checked because providers of misinformation were able to connect and form co-sharing networks through different social media platforms.[83] As Facebook, Twitter, Instagram, and YouTube did not prohibit users from posting inaccurate information on vaccinations during the COVID-19 pandemic,[84] empirical studies have found that exposure to

[81] Ingjerd Skafle and others, 'Misinformation About COVID-19 Vaccines on Social Media: Rapid Review' (2022) 24 Journal of Medical Internet Research. See also Tsung-Ling Lee, 'Pandemic Accord, Digital Health Literacy, and Human Rights in the Era of Infodemic' (2023) 18 Asian Journal of WTO & International Health Law and Policy 397.

[82] Sahil Loomba and others, 'Measuring the Impact of COVID-19 Vaccine Misinformation on Vaccination Intent in the UK and USA' (2021) 5 Nature Human Behaviour 337-348.

[83] Aimei Yang and others, 'The Battleground of COVID-19 Vaccine Misinformation on Facebook: Fact Checkers vs. Misinformation Spreaders' [2021] Harvard Kennedy School Misinformation Review <https://misinforeview.hks.harvard.edu/article/the-battleground-of-covid-19-vaccine-misinformation-on-facebook-fact-checkers-vs-misinformation-spreaders/> accessed 13 December 2024.

[84] Leo Kelion, 'Coronavirus: Facebook, Twitter and YouTube "Fail to Tackle Anti-Vaccination Posts"' (*BBC*, 3 September 2020) <https://www.bbc.com/news/technology-54001894> accessed 13 December 2024.

misinformation that was factually accurate but with deceptive content about vaccination has reduced vaccination intentions.[85]

As more technology giants enter the healthcare industry and introduce digital health applications, the proliferation of digital health apps is capturing various aspects of individuals' lives. This trend could lead to data migration from the Global South to the Global North. Such migration could deprive individuals of control over their data, including how it is used and whether the migration of data would benefit them in any way. Similarly, as the health data migrates from the Global South to the Global North, how the data is governed, accessed, stored, and used raises additional concerns.[86] However, such concerns are not restricted to Global South to Global North extraction but also extend to Global South to Global South digital cooperation.

Professors Angelina Fisher and Thomas Streinz go further and argue that control over data infrastructure is a form of control over social, political, and economic organizations.[87] The UN Human Rights Council notes that how data is created and used often reflects the values and biases of the companies or individuals that created them. As digital technologies blur the boundaries between physical and digital environments, they create a 'physical-digital-physical loop' where the flow of data from the real world to the digital space and back into the real world potentially entrenches those biases.[88] As control over data is no longer solely within the province of individuals, unequal control over data could be a pervasive form of digital inequality that could

[85] Jennifer Allen, Duncan J Watts and David G Rand, 'Quantifying the Impact of Misinformation and Vaccine-Skeptical Content on Facebook' (2024) 384 Science eadk3451.

[86] Sekalala and Chatikobo (n 11).

[87] Corey H Basch and others, 'Social Media, Public Health, and Community Mitigation of COVID-19: Challenges, Risks, and Benefits' (2022) 24 Journal of Medical Internet Research e368044.

[88] Human Rights Council Advisory Committee, 'Possible impacts, opportunities and challenges of new and emerging technologies with regards to the promotion and protection of human rights' (2021) A/HRC/47/52.

undermine economic development, human agency, and collective self-determination more broadly.[89]

Human rights law, at both the individual and state levels, could be a pragmatic remedy to technological solutionism. However, the invocation of human rights as an emancipatory language might overlook the broader geopolitical tension that shapes the political economy of digital health innovations and infrastructure. Even though human rights law aspires to be an emancipatory language, without confronting the political and geopolitical structures that influence power dynamics underpinning the political economy of digital health, it risks perpetuating existing power structures, which is significantly problematic. As such, despite international human rights law drawing attention to the issues of accessibility, availability, and quality of digital services and flagging the issue of data colonialism, I argue that human rights law does not go far enough to critically interrogate the geopolitical powers that shape the distribution and ownership of digital health infrastructure—to which the paper now turns.

IV Data Sovereignty Narrative

The unfolding narrative of digital health in ASEAN is increasingly infused with a geopolitical undertone. Data sovereignty has emerged as an alternate narrative that encapsulates the state-driven regulatory model on the use, collection, export, management, and extrapolation of data. Legal scholar Anu Bradford foregrounds the era of digitalization as power contestation amongst the world's powerful states, which poses challenges to the global legal order.[90] States now extend political influence to the digital sphere, shaping the digital environment to reflect their interests and values. Thus, it could be argued that 'data sovereignty' as a legal concept is an extension of

[89] Shawneequa L Callier and others (n 43).

[90] Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press 2023).

states' geopolitical influences in the digital space as powerful states seek their global technological dominance.[91]

China and Data Sovereignty

This concept asserts a state's territorial authority over data produced within its borders.[92] China's interpretation of data sovereignty is particularly noteworthy. Despite its widespread use, there is no universally agreed-upon definition for data sovereignty.[93] Nonetheless, the term commonly refers to the autonomy and control over data at the individual, population, or national levels. Conceptually, cyber sovereignty refers to control over cyberspace, while data sovereignty refers to control over data. Cyber sovereignty can be seen as a broader concept than data sovereignty as it includes states conducting actions in cyberspace using information technology

[91] Renata Avila Pinto, 'Digital Sovereignty or Digital Colonialism?' (*International Journal on Human rights,* 16 July 2018) <https://sur.conectas.org/en/digital-sovereignty-or-digital-colonialism/> accessed 12 December 2024; Benjamin Cedric Larsen, 'The Geopolitics of AI and the Rise of Digital Sovereignty' (Brookings, 2022) <https://www.brookings.edu/articles/the-geopolitics-of-ai-and-the-rise-of-digital-sovereignty/> accessed 12 December 2024; Sharinee Jagtiani, 'The Geopolitics of Data Governance and Digital Power Play, GJIA' (*Georgetown Journal of International Affairs*, 10 August 2023) <https://gjia.georgetown.edu/2023/08/10/the-global-cloudscape-the-geopolitics-of-data-governance-and-digital-power-play/> accessed 12 December 2024; Martina Francesca Ferracane, *Data governance models and geopolitics: insights from the Indo-Pacific region* (European University Institute Publications Office 2022).

[92] Chien-Liang Lee, 'Normative Implications of Digital Sovereignty: Conceptual Framework' (2024) The Taiwan Law Review 355.

[93] Ilona Kickbusch and others, 'The Lancet and Financial Times Commission on Governing Health Futures 2030: Growing up in a Digital World' (2021) 398 The Lancet 1727.

infrastructures that include the internet, telecommunications networks, computer systems, and internet-connected devices.[94]

China has previously asserted the concept of 'cyber sovereignty' as a counter to U.S. dominance in cyberspace and a means to balance global internet regulation.[95] Along with Russia, China proposed a Code of Conduct for state behavior in the United Nations General Assembly in 2011 and 2014, which aimed to embed the principle of sovereignty with international cooperation in cyberspace. [96] Strikingly, despite presenting cyber sovereignty as resistance to Western influence, China paradoxically adopts a Westphalian state definition of sovereignty—traditionally associated with European powers. While the precise meaning and content of China's cyber sovereignty remains intentionally vague, President Xi explained in 2015, 'respecting cyber sovereignty' meant 'respecting each country's right to choose its own internet development path, its own internet management model, its own public policies on the internet, and to participate on equal basis in the governance of international cyberspace—avoiding cyber-hegemony and avoiding interference in the internal affairs of other countries'.[97] China's vision of cyber sovereignty originated in a 2010 white

[94] 'Application of International Law to States' Conduct in Cyberspace: UK Statement' (*GOV.UK*, 3 June 2021) <https://www.gov.uk/government/publications/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement/71358e6f-f834-4d09-88be-d5a13f8bf1df> accessed 13 December 2024.

[95] Justin Sherman, 'How Much Cyber Sovereignty Is Too Much Cyber Sovereignty?' (*Council on Foreign Relations*, 30 October 2019) <https://www.cfr.org/blog/how-much-cyber-sovereignty-too-much-cyber-sovereignty> accessed 12 December 2024.

[96] Dennis Broeders and Bibi van den Berg, *Governing Cyberspace Behavior, Power and Diplomacy* (Rowman & Littlefield Publishers Incorporated 2020).

[97] People's Republic of China, 'Jointly Build a Community with a Shared Future in Cyberspace' (*State Council Information Office of the People's Republic of China*, November 2022) <http://english.scio.gov.cn/node_8033411.html> accessed 13 December 2024.

paper titled 'The Internet in China' which declared, '[w]ithin Chinese territory the Internet is under the jurisdiction of Chinese sovereignty'[98] linking content regulations and state's censorship of internet with territoriality as recognized in international law. Extending the traditional notion of territorial sovereignty that expresses Westphalian notion of statehood to the cyberspace, China asserts the principle of non-interference which further justifies Communist Party's content regulations and strengthen the country's participation as an equal in cyberspace governance. This enabled China to advocate for a state-centric multilateralism model as opposed to the 'bottom-up multi-stakeholders' model endorsed by the US and other Western Countries. In short, this adherence to sovereignty principles can be seen as a strategic move to assert China's right to develop its cyber model within its borders.[99]

In many ways, data sovereignty is an extension of China's broader concept of cyber sovereignty. This concept, introduced by China's President Xi in 2015, represents a state's assertion of autonomy in controlling its digital technologies, content, and infrastructure within its territories.[100] China's cyber sovereignty provides a basis for the government to take a more

[98] Information Office of the State Council of the People's Republic of China, 'White Paper on the Internet in China' (*Information Office of the State Council of the People's Republic of China*, 8 June 2010) <http://www.china.org.cn/government/whitepaper/node_7093508.htm> accessed 13 December 2024.

[99] Steven Feldstein, 'New Digital Dilemmas: Resisting Autocrats, Navigating Geopolitics, Confronting Platforms' [2023] Carnegie Endowment for International Peace; Nicholas Zúñiga and others, 'The Geopolitics of Technology Standards: Historical Context for US, EU and Chinese Approaches' (2024) International Affairs 1635; Huotari and others, 'Decoupling: Severed Ties and Patchwork Globalisation' (*European Chamber of Commerce in China*, 1 January 2020) <https://merics.org/en/report/decoupling-severed-ties-and-patchwork-globalisation> accessed 11 December 2024.

[100] Lizhi Liu, 'The Rise of Data Politics: Digital China and the World' (2021) 56 Studies in Comparative International Development 45. See also Changer and Sun (n 14).

interventionist role in controlling information content, data storage, and market access in cyberspace. Thus, cyber sovereignty—while not universally accepted—poses dual critically challenges: how the internet is used currently, and the United States' hegemony in cyberspace.

China's increasing involvement in this area of digital health is notably ambitious. In recent years, China has been asserting its influence in the digital health sector, a vital component of healthcare provision that is part of advancing universal health coverage. China's ambitions go beyond service providing; some scholars argue that the Chinese government is motivated by the desire to set technical standards that reflect its ideological values. [101]

China's distinct interpretation and approach to data sovereignty holds significant importance in the global era of digitalization. As a global power, China is strategically redefining the concept of 'data sovereignty' to set itself apart from liberal democracies.[102] This strategy could arguably be a response to what China views as inherent imperialistic tendencies in the growing phenomenon of data colonialism. Consequently, the term 'data sovereignty' becomes imbued with a potent political undercurrent, transforming it into a political resistance tool against the ideologies of the liberal West, rallying support from the Global South amidst escalating geopolitical tension.[103]

Significantly, China's assertion of data sovereignty serves as a defense against foreign ownership and control of digital health services and infrastructure.[104] More broadly, this assertion presents an alternative data paradigm, posing a critical challenge to the political ideologies of liberal democracies. Whether

[101] Stacie Hoffmann, Dominique Lazanski and Emily Taylor, 'Standardising the Splinternet: How China's Technical Standards Could Fragment the Internet' (2020) 5 Journal of Cyber Policy 239.

[102] Kokas (n 10).

[103] Thumfart (n 10) 4.

[104] ibid.

China's (re)interpretation of data sovereignty will gain political traction in countries that rely on China's digital capacities remains to be seen.

Data Sovereignty as Emancipation

In addition to data sovereignty as a legal concept, the term has also emerged as a powerful emancipatory language, increasingly resonating with and being invoked by various Indigenous communities.[105] Indigenous Data Sovereignty (IDSov) consists of a network of Indigenous academics, innovators, and knowledge-holders in the United States, Canada, Aotearoa (New Zealand), Australia, the Pacific, and Scandinavia which advocates the right of Indigenous people to own their data.[106] Arguably, data sovereignty is used as a pushback against 'data colonialism'—where governments and private sector entities claim ownership over individual data without their consent or involvement. Such a scenario could reinforce existing power dynamics and worsen health disparities. Thus, data sovereignty becomes a crucial defense strategy for these communities, providing them with a means to contest and resist unwanted digital data extraction. In this context, data sovereignty functions as empowerment, enabling these communities to protect their digital autonomy and assert control over their data, thus countering potential infringements on their digital rights. Data sovereignty

[105] See e.g. Maui Hudson and others, 'Indigenous Peoples' Rights in Data: A Contribution toward Indigenous Research Sovereignty' (2023) 8 Frontiers in Research Metrics and Analytics 1173805; Kelsey Leonard, Stephanie Russo and others, 'Our Common Agenda Global Digital Compact March 2023: CARE Statement for Indigenous Data Sovereignty' (2023) <https://www.un.org/digital-emerging-technologies/sites/www.un.org.techenvoy/files/GDC-submission_WAMPUM_Lab_and_the_Collaboratory_for_Indigenous.pdf>. See also First Nations Information Governance Centre, 'Home' (*FNIGC/CGIPN*) <https://fnigc.ca/> accessed 3 February 2025.

[106] Tahu Kukutai, 'Indigenous Data Sovereignty—A New Take on an Old Theme' (2023) 382 Science eadl4664.

is often invoked as an emancipatory tool by Native tribes in the United States[107] and Indigenous populations in New Zealand and India[108].

V ASEAN Digital Health Landscape

ASEAN includes ten Southeast Asian countries and nearby archipelagos: Brunei, Singapore, Malaysia, Thailand, the Philippines, Indonesia, Vietnam, Lao PDR (Laos), Cambodia, and Myanmar (Burma). The region has a population of over 650 million, approximately twice that of the United States and three times the size of Western Europe. ASEAN is marked by an immense diversity among and within countries regarding health systems, political structures, geography, cultures, sociodemographic traits, languages, religions, and history. Economies range from Singapore, a high-income country, to low-income countries like Lao PDR.

Over the past two decades, differences in economic powers in the region have been narrowing. In 1997, Singapore's GDP was 57 times that of Lao PDR.[109] By 2016, changes in economic situations have reduced this gap to less than 19 times.[110] However, growing inequalities within countries could overshadow this promising trend. With the impacts of digital technologies that blur the lines between the physical, digital, and biological spheres, concerns over widening economic inequalities across countries remain. More broadly, as ASEAN thrives on economic development to achieve social and economic stability at the regional level, the widening economic

[107] Rebecca Tsosie, 'Tribal Data Governance and Informational Privacy: Constructing "Indigenous Data Sovereignty"' (2019) 80 Montana Law Review 229.

[108] Mana Raraunga, Data Sovereignty' (*Royal Society Te Apārangi*, 2023) <https://www.royalsociety.org.nz/what-we-do/our-expert-advice/all-expert-advice-papers/mana-raraunga-data-sovereignty/> accessed 13 January 2024.

[109] Fukunari Kimura, Venkatachalam Anboumozho and Hidetoshi Nishimura *ASEAN Vision 2040: Transforming and Deepening the ASEAN Community* (Economic Research Institute for ASEAN and East Asia 2019)

[110] ibid.

disparities could threaten regional integration and reduce public trust in government.[111]

Established in 1967, ASEAN was formed partly in response to the withdrawal of colonial powers from the region.[112] The power vacuum and the spread of communism in Vietnam and the People's Republic of China provided the political impetus for the founding member countries of ASEAN, Indonesia, Thailand, the Philippines, Malaysia, and Singapore, to form a regional organization with the support of the United States.[113] The creation of ASEAN was not only a response to the threat of communism, which then was deeply entrenched in Vietnam and the People's Republic of China, but also against the imperialistic impulse of Western nations to interfere in domestic political affairs at the regional level and to preserve individual member states' economic stability at the national level.[114]

As such, one of ASEAN's primary objectives was to reduce the influence of foreign powers in the region and promote the national identities of its member states. ASEAN countries sought to improve welfare through economic development to pursue these objectives.[115] The Declaration of ASEAN Concord embodies this common aspiration where regional cooperation in economic and social development would facilitate the 'elimination of poverty, hunger, disease, and illiteracy, with particular

[111] ibid.

[112] See e.g. Ooi Kee Beng and others, *The 3rd ASEAN Reader* (ISEAS–Yusof Ishak Institute 2015) <https://muse.jhu.edu/pub/70/edited_volume/book/42016> accessed 3 February 2025.

[113] ibid.

[114] Council on Foreign Relations, 'What Is ASEAN?' (*Council on Foreign Relations*, 18 September 2023) <https://www.cfr.org/backgrounder/what-asean> accessed 13 December 2024.

[115] ibid.

emphasis on the promotion of social justice and the improvement of the living standards'.[116]

Economic development was widely regarded as a bedrock to build national resilience. The founders of the ASEAN believed that poverty and economic discontent would be an incubator for communist insurgencies. Aligning the region's economic development with the West would encourage internal political stability, engender confidence from international donors and investors, and encourage regional stability.[117]

Since the establishment of ASEAN, economic prosperity and relative political stability have seen increases in life expectancy across the region. Paradoxically, this celebratory trend also brings new health challenges as the region experiences demographic and epidemiological shifts. [118] Rising healthcare costs, increased prevalence of chronic diseases, and surged demands for care for older adults have also begun to strain healthcare systems. [119] While cross-ASEAN populations are aging at an uneven rate, incidences of dementia amongst older adults are projected to triple from 23 million in 2015 to 71 million individuals by 2050. As more countries implement universal health coverage, public expenditures on healthcare have also risen considerably.

Even before the outbreak of the COVID-19 pandemic, digital health was prominently featured as a policy priority for the region.[120] Interest in digital

[116] ASEAN, 'The Declaration of ASEAN Concord, Bali, Indonesia, 24 February 1976' (*ASEAN Main Portal*, 14 May 2012) <https://asean.org/the-declaration-of-asean-concord-bali-indonesia-24-february-1976/> accessed 3 August 2023.

[117] Beng (n 112).

[118] Elgazzar, Yoong and Finkelstein (n 56).

[119] One study shows that deaths from non-communicable diseases resulted in 3 million deaths in 2017, making NCDs one of the top causes of preventable mortality and morbidity in the region. See ibid.

[120] The ASEAN Secretariat, 'ASEAN Health Protocol for Pandemic Preventive Measures in Public Places' (2022) <https://asean.org/wp-

health and regional coordination over digital health policy has since intensified as the technologies have demonstrated their importance during the pandemic. In particular, telemedicine has become a viable option that complements traditional health care and has facilitated the practice of social distancing during the pandemic. Strict quarantine requirements and social distancing saw a surge in the region's use of digital health tools. Three policy documents pertaining to digital health in the region—the ASEAN Post-2015 Health Development Agenda (2016-2020), the ASEAN Economic Community Blueprint 2025, and the ASEAN Digital Masterplan 2025—set out the region's policy objectives in digital technology in health and non-health sectors.[121] These policy documents also embrace digital health as part of the region's development discourse.

Despite these regional policy documents, several regulatory challenges relating to digital health still need to be addressed. As the digital health industry expands its services, some ASEAN members lack a clear and well-defined legal framework for data protection to govern the collection, storage, process, and sharing of sensitive health data.[122] Brunei, Lao, and Thailand lack laws and regulations specific to digital health, and Cambodia and Myanmar are developing data protection laws and regulations. Likewise, telemedicine frameworks are at different levels of development within the

content/uploads/2022/11/ASEAN-Health-Protocol_Final_20221215.pdf> accessed 14 December 2024.

[121] ASEAN Secretariat, 'The ASEAN Post-2015 Health Development Agenda (2016-2020)' (2018) <https://asean.org/wp-content/uploads/2018/12/16-ASEAN-Post-2015-Health-Development-Agenda-1.pdf> accessed 14 December 2024; ASEAN Secretariat, 'The ASEAN Economic Community Blueprint 2025' (2015) <https://asean.org/book/asean-economic-community-blueprint-2025/> accessed 14 December 2024; ASEAN Secretariat, 'The ASEAN Digital Masterplan 2025' (2021) <https://asean.org/book/asean-digital-masterplan-2025/> accessed 14 December 2024.

[122] Katrina Navallo and Keith Detros, 'Assessing Digital Health Adoption in ASEAN' (ASEAN-Japan Centre 2024).

region. Indonesia, Malaysia, the Philippines, and Vietnam have specific regulations governing the conduct of telemedicine sectors, while other countries rely on general codes or guidelines.[123] Researchers in Vietnam have found that using artificial intelligence-enhanced applications in health care could benefit from a clear legal and regulatory framework.[124] Privacy and data protection require legislation and a legal framework to safeguard privacy,[125] confidentiality, and access to health information.[126] However, the Global Digital Health Index shows that many ASEAN countries still lack laws and regulations on privacy and agreed rules on the migration of health data and sharing.[127] Concerns over privacy and data protection can deter patients from adopting digital health services.[128] The absence of clear guidelines for adopting artificial intelligence in health settings can create confusion about healthcare providers' responsibilities and compromise patient privacy. In the absence of applicable laws and regulations, low-resource countries may rely on guidelines developed by high-income

[123] OECD, *Economic Outlook for Southeast Asia, China and India 2021: Reallocating Resources for Digitalisation* (OECD 2021) 145.

[124] Ho Quang Chanh and others, 'Applying Artificial Intelligence and Digital Health Technologies, Viet Nam' (2023) 101 Bulletin of The World Health Organization 487.

[125] Magdalena Słok-Wódkowska and Joanna Mazur, 'Between Commodification and Data Protection: Regulatory Models Governing Cross-Border Information Transfers in Regional Trade Agreements' (2024) 37 Leiden Journal of International Law 111.

[126] Singapore has a robust legal framework for data protection, but the country is not immune from cybersecurity attacks. In 2018, Singapore Health Services had a massive data breached where the personal health information of 1.5 million patients was leaked. This data breach included outpatient data, including Singaporean Prime Minister Lee Hsien Loon. See Ministry of Health Singapore, 'Singhealth's IT System target of cyberattack. Singapore' (*Singapore Ministry of Health*, 20 July 2018) <https://www.moh.gov.sg/news-highlights/details/singhealth's-it-system-target-of-cyberattack>.

[127] OECD (n 123) 35.

[128] Elgazzar, Yoong and Finkelstein (n 56).

countries. However, the direct applicability of these guidelines may be limited as they may only partially reflect local needs and settings.

VI TECHNOLOGICAL BARRIERS – ICT INFRASTRUCTURE

Within ASEAN states, the development and adoption of digital services are hindered by technical and infrastructural barriers, particularly in rural and remote areas and among older populations. Insufficient technological infrastructure, including unreliable and unaffordable access to the internet, mobile phones, computers, and even electricity, continues to limit access to digital health services. For instance, Cambodia relies on power imported from Thailand, Vietnam, and Lao PDR.[129] The unstable power supply source poses a primary barrier to developing digital health services. In Indonesia, information and communication technologies (ICT) infrastructure is weak, and connection speed and internet bandwidth remain slow. In the least developed countries—Lao PDR, Cambodia, and Myanmar—a large proportion of the population has no stable internet access or no internet access. In fact, according to a joint OECD and WHO report, only 14% of the population in ASEAN states have access to affordable high-speed internet.[130]

Yang Chen and Amitava Banerjee argued that, in order to promote the broader adoption of telemedicine, it is critical to upskill health professionals with the necessary digital health skills.[131] These skill sets include navigating and operating on digital platforms and interacting virtually with patients while observing privacy and data protection requirements. Offering telemedicine

[129] Sudarshan Varadhan, 'Cambodia to Boost Power Import Capacity to Improve Flexibility' (*Reuters*, 22 October 2024) <https://www.reuters.com/business/energy/cambodia-boost-power-import-capacity-by-over-50-next-two-years-2024-10-21> accessed 14 December 2024.

[130] OECD and the World Health Organization, *Health at a Glance: Asia/Pacific 2020: Measuring Progress Towards Universal Health Coverage* (OECD 2020).

[131] Chen and Banerjee (n 7).

alongside in-person visits could help address the healthcare worker shortage. ASEAN states suffer from a chronic shortage of trained health and healthcare workers. The WHO advises a minimum of 4.45 skilled medical workers per 1000 population, but most ASEAN countries fall below this requirement.[132] As healthcare systems become more digitalized, it also places additional demands on healthcare workers to upgrade their digital skills. Training and adapting these new digital technologies often require strategic planning, additional institutional resources, time, and financial support, which countries often lack.[133] Even though countries increasingly adopt electronic systems for recording medical information, which could reduce the amount of paperwork and enable healthcare workers to dedicate more time to patient care, potentially improving the quality of health services, the inadequate training and uptake of digital healthcare services by healthcare workers can impede the integration of technology into healthcare operations and programs effectively.[134]

Yet, according to the ASEAN policy documents, digital health technologies offer a promising public health solution for reducing and managing healthcare costs at the population level.[135] These policy documents present a more optimistic viewpoint, where, by increasing access and improving health service delivery, policymakers argue that these technologies could reach previously underserved populations and lower healthcare costs. With

[132] World Health Organization, 'Global Health Workforce Statistics Database, Medical Doctors' (*World Health Organization*, 2019) <https://www.who.int/data/gho/data/themes/topics/health-workforce> accessed 13 December 2024.

[133] Israel Júnior Nascimento and others, 'Barriers and Facilitators to Utilizing Digital Health Technologies by Healthcare Professionals' (2023) 6 npj Digital Health 1.

[134] ibid.

[135] ASCC Research and Development Platform, 'Transforming the Digital Health Landscape in ASEAN' (*ASEAN Social-Cultural Community Knowledge Hub* 2023) <https://knowascc.asean.org/publication/transforming-the-digital-health-landscape-in-asean/> accessed 13 December 2024.

over half of the populations in ASEAN countries digitally connected—about 360 million users in Southeast Asia, 90% of whom are connected through mobile devices - the potential to increase access, expand service coverage, improve service quality, reduce health disparities, and lower healthcare costs is tremendous according to these policy documents. [136]

During the pandemic, telemedicine has become increasingly popular in the region. It spurred investment in telehealth and predictive analytics, among the highest capital invested in Southeast Asia health technology.[137] One estimation shows that the Internet economy is worth over $100 billion in ASEAN countries, attracting private investments in digital health innovations.[138] Patients and health providers have relied on commercial video conferencing services, such as Zoom, Microsoft Teams, and FaceTime, which are not specifically designed for health consultations. However, using these third-party applications may carry privacy risks for patients and providers.

Generally, the burgeoning health innovations in the region can be classified into the following five areas: (1) monitoring and tracking health outcomes and behaviors, (2) nudging devices or platforms to support individual self-management, (3) telehealth services that integrate with existing health systems, (4) crowdsource health information apps or platforms, and (5) AI/machine learning (ML).[139] All of these require a robust digital health infrastructure.

---

[136] ibid.

[137] ibid.

[138] Marc Mealy and others, 'Southeast Asia's Digital Economy Projected to Hit US$100 Billion in Revenue in 2023' (*US-ASEAN* 28 November 2023) <https://www.usasean.org/article/southeast-asias-digital-economy-projected-hit-us100-billion-revenue-2023#:~:text=Southeast%20Asia's%20digital%20economy%20is,to%20reach%20US%24218%20billion> accessed 13 December 2024.

[139] Elgazzar, Yoong and Finkelstein (n 56).

As a regional economic bloc, ASEAN promotes the digitization of healthcare. The ASEAN Post-2015 Health Development Agenda (2016-2020), the ASEAN Economic Community Blueprint 2025, and the ASEAN Digital Masterplan 2025on digital health exhibit a sense of technological solutionism.[140] The policy papers adopted by ASEAN often describe digital health innovations that could significantly transform the healthcare sector, including addressing the shortage of healthcare workers and keeping healthcare expenditures down.[141] Such optimism in digital health solutions is not new. The ASEAN e-Government Strategic Action Plan, developed in 2011, identified e-health as a tool to improve the health conditions of people in the region as well as expand healthcare access to the populations. This optimism has been maintained for one decade, with the most recent 'ASEAN Digital Master Plan' providing an ambitious roadmap to transform the regional bloc into a digital society.[142]

With the ambition to become a leading economy, the Digital Master Plan acknowledges that '[t]he digital divide has been highlighted as a critical barrier to the mitigation value of digitalization. In particular, populations unserved or partially served by broadband cannot benefit from home-based learning for children, telecommuting, access to e-commerce and healthcare information'.[143] However, the Master Plan does not go further to address other human rights issues relating to the rapid implementation of digital infrastructure, despite the fact that all ASEAN member states have ratified the Convention on the Rights of the Child (CRC), the Convention on Rights of Persons with Disabilities (CRPD) and the Convention on the Elimination on All Forms of Discrimination Against Women. The absence of the articulation of human rights consideration in these ASEAN policy

[140] ASEAN Secretariat (n 121).

[141] 'Transforming The Digital Health Landscape in ASEAN' (n 135); ASEAN Secretariat (n 121).

[142] 'ASEAN Digital Masterplan 2025' (ASEAN Main Portal) <https://asean.org/book/asean-digital-masterplan-2025/> accessed 2 August 2024.

[143] ibid.

documents is perhaps unsurprising as the region has been historically reluctant to embrace international human rights law.[144] Despite this, the UN General Assembly adopted a resolution underscoring the negative impact of technological surveillance on human rights in 2013,[145] through their separate mandates, Special Rapporteur on the Right to Health,[146] Special Rapporteur on Extreme Poverty and Human Rights,[147] the Special Rapporteur on the Right to Freedom of Opinion and Expression[148] each have identified and cautioned that digital health can have disproportionate impacts on different population groups because of gender, race, and their potential negative impacts on the youth and individuals with disabilities.

Some scholars are critical of the lack of attention given by governments in ASEAN to mitigating these human rights concerns.[149] Professor Sarah Davis, for instance, voices concerns that '[w]hile there is potential to protect human rights within the realm of digital technologies and cyberspace, the current

[144] John Arendshorst, 'The Dilemma of Non-Interference: Myanmar, Human Rights, and the ASEAN Charter' (2009) 8 Northwestern Journal of Human Rights 102.

[145] United Nations General Assembly, 'Resolution adopted by the General Assembly on 18 December 2013 on the right to privacy in the digital age' (2014) Res. 68/147, UN Doc. A/RES/68/167.

[146] 'Digital innovation, technologies and the right to health, Report of the Special Rapporteur on the Right to Health' (n 36).

[147] United Nations Human Rights Council, Report of the Special Rapporteur on Extreme Poverty and Human Rights, (2019) UN Doc. A/74/493.

[148] United Nations Human Rights Council, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN Doc. A/HRC/41/35 (2019).

[149] Andrew Lowenthal, 'Upholding Human Rights and Privacy in the Digital Age' (*EngageMedia* 14 January 2021) <https://engagemedia.org/2021/human-digital-rights-asean/> accessed 13 December 2024; Adam Poulsen and others, 'Digital Rights and Mobile Health in Southeast Asia: A Scoping Review' (2024) 10 Digital Health 1.

ASEAN literature does not reveal concern with these issues'.[150] While the human rights narrative has, thus far, played a relatively limited role in informing digital health policies in ASEAN, While the human rights narrative has, thus far, played a relatively limited role in informing digital health policies in ASEAN, the availability and accessibility of these technologies remain central focuses in ASEAN's digital health discourse.[151] The negative implications of digital health technologies have not gone unnoticed by human rights advocates and UN human rights bodies, including the Special Rapporteurs.[152] For example, the UN Human Rights Office of the High Commissioner has voiced concerns about the misuse of digital surveillance technologies during the pandemic.[153] At least six ASEAN countries have implemented some form of digital contact tracing, with unclear policies regarding their implementation or use.[154] This has raised security concerns, posing potential risks to user data and undermining privacy protection.[155]

Although the human rights narrative has not significantly shaped the digital health discourse in ASEAN states, data sovereignty has played a more prominent role. In particular, in the realm of digital health, China is creating a level of reliance and dependency among other Global South countries by

[150] Sara (Meg) Davis, Nerima Were and Tara Imalingat, 'Digital Health Rights: Initial Analysis' (2021) 27 Global Health Centre Working Paper 1.

[151] ASEAN Secretariat (n 121).

[152] Davis (n 158), United Nations Human Rights Council (n 156).

[153] United Nations Human Rights Office of the High Commissioner, 'Human Rights Should Be at the Heart of Tech Governance' (*United Nations Human Rights Office,* 1 September 2022) <https://www.ohchr.org/en/stories/2022/09/human-rights-should-be-heart-tech-governance> accessed 13 December 2024.

[154] Singapore, Indonesia, Thailand, Vietnam, Malaysia, and the Philippines adopted digital contract tracing as part of their pandemic response. See DigitalReach, 'Recommendations for AICHR on FOEI in ASEAN Member States' (*DigitalReach*, 22 January 2020) <https://digitalreach.asia/advocacy/recommendations-for-aichr-on-foei-in-asean-member-states/> accessed 3 February 2025.

[155] ibid.

providing crucial digital infrastructure such as cables, satellites, and smart cities. China's strategic exportation of digital infrastructure may create reliance on Chinese digital capabilities rather than encouraging local innovations and ownership.[156] While China actively engages in knowledge transfer through training local personnel to enhance their digital skills, China does not engage in technology transfer. As a result, a host country is unable to acquire the technical expertise needed to build, own, and maintain its own digital infrastructure.

ASEAN and China initiated their cooperation on digital economy infrastructure with the commencement of the Digital Economy Framework Agreement in September 2023.[157] Even prior to this agreement, China's influence in ASEAN countries began as early as 2010, with private companies catering to the price-sensitive market in ASEAN countries by selling budget mobile phones.[158]

ASEAN governments are increasingly partnering with Chinese IT companies such as Huawei and Alibaba to advance digitalization. Huawei is aiding 5G network development, data center construction, smart city building, and IT human resource development in Cambodia, Malaysia, Thailand, and Indonesia.[159] Alibaba is also collaborating with ASEAN

---

[156] Matthew S Erie and Thomas Streinz, 'The Beijing Effect: China's "Digital Silk Road" as Transnational Data Governance' (2021) 54 New York University Journal of International Law and Politics 1.

[157] China Global Television Network, 'China, ASEAN Countries Hail Digital Economy Cooperation, See Pragmatic Results' (*CGTN,* 6 September 2023) <https://news.cgtn.com/news/2023-09-06/China-ASEAN-countries-hail-digital-economy-cooperation-1mSoYri57X2/index.html> accessed 13 January 2024.

[158] Kaori Iwasaki, 'Chinese Firms Driving Digitalization in the ASEAN Region' <https://www.jri.co.jp/en/reports/rim/2023/90/> accessed 23 October 2024.

[159] Joseph Sipalan, 'Huawei and Malaysia's 5G Future: Unpacking the Embrace of Chinese Tech' (*South China Morning Post*, 19 October 2024) <https://www.scmp.com/week-asia/economics/article/3282981/huawei-spies-second-chance-shape-malaysias-5g-future> accessed 28 October 2024.

governments, notably signing a Memorandum of Understanding—a nonbinding agreement between two parties—with the Thai government to support the 'Thailand 4.0' vision. These collaborations purportedly offer mutual benefits: Chinese IT companies gain acceptance and business opportunities, while ASEAN countries receive much-needed technical support to enhance their digital capacities.

Globally, Chinese companies like ZTE and Huawei are major suppliers of 5G products and services. While developed countries are increasingly excluding Chinese companies from 5G network development,[160] Vietnam has expressly supported Chinese-owned companies operating within its territory, and Cambodia is partnering with Huawei to develop a 5G network involving the country's top three telecom operators. Singapore's major telecom operators did not select Chinese suppliers, neither did they actively exclude Chinese suppliers. Despite warnings from the US and EU about potential security risks and negative investment impacts if Huawei wins the tender, Malaysia plans to transition from a monopoly to a two-company structure for 5G development in 2024 involving Huawei.[161]

Chinese IT firms have increased their presence in ASEAN countries, some with government support and others independently. Whether voluntary or government-influenced, these actions advance the digitalization of ASEAN

[160] Aparna Divya, 'Germany Is Phasing out Chinese Tech From 5G Networks. Is It the Right Call?' (*The Diplomat,* 26 July 2024) <https://thediplomat.com/2024/07/germany-is-phasing-out-chinese-tech-from-5g-networks-is-it-the-right-call/> accessed 12 December 2024; UK Government, 'Huawei to Be Removed from UK 5G Networks by 2027' (*GOV.UK*, 13 October 2022) <https://www.gov.uk/government/news/huawei-to-be-removed-from-uk-5g-networks-by-2027> accessed 12 December 2024.

[161] Joseph Sipalan, 'Huawei and Malaysia's 5G Future: Unpacking the Embrace of Chinese Tech' (*South China Morning Post,* 19 October 2024) <https://www.scmp.com/week-asia/economics/article/3282981/huawei-spies-second-chance-shape-malaysias-5g-future> accessed 12 December 2024.

countries and contribute to China's Digital Silk Road (DSR) Initiative.[162] This increased presence of Chinese IT firms in ASEAN countries is in part because of the Chinese government's stronghold on Chinese IT companies through domestic regulations. Notably, from 2020 to 2022, the Chinese government significantly tightened regulations to weaken the influence of Chinese IT firms, including Alibaba. Companies were either accused of antitrust violations or had their IPOs suspended.[163] Consequently, these Chinese IT firms were more willing to comply with demands from the Chinese government as state ownership within these companies grew.[164]

Given ASEAN's diverse legal systems and socioeconomic arrangements, analyzing individual member countries is beyond this paper's scope. However, Indonesia serves as a useful example to illustrate the foreign ownership dimension, which is typically present in the region.[165]

Indonesia has a weak ICT infrastructure. While foreign ownership of digital infrastructure is nothing new, the potential implications of cross-border data usage are of particular concern in China's context. It involves data generated overseas being processed in China under the Chinese government's oversight. This situation, combined with limited transparency on how the

[162] Singapore Institute of International Affairs, 'Understanding the "Digital Silk Road": Implications for ASEAN' (*Singapore Institute of International Affairs,* 27 August 2018) <https://siiaonline.org/understanding-the-digital-silk-road-implications-for-asean/> accessed 12 December 2024.

[163] Daisuke Wakabayashi, 'Alibaba, China's e-commerce giant, will split into 6 units' (*The New York Times*, 28 March 2023) < https://www.nytimes.com/2023/03/28/business/alibaba-china-e-commerce.html> accessed 24 January 2025.

[164] Luisetta Mudie, 'China Moves Towards Nationalization With Probe Into Alibaba' (*Radio Free Asia*, 25 December 2020) <https://www.rfa.org/english/english/news/china/alibaba-probe-12252020170303.html> accessed 28 October 2024.

[165] 'Chinese Firms Driving Digitalization in the ASEAN Region' (n 158).

Chinese government utilizes and stores data, raises concerns about privacy and security.[166]

Like other countries, Indonesia is concerned about foreign governments using ICT infrastructure for political and economic leverage. Indonesia's law requires that data from the public sector must be stored, managed, and processed in the country, but its sector-specific laws do not extend to other areas.[167] Moreover, Indonesia also faces more immediate issues, such as cybercrimes and the rapid spread of mis- and disinformation, which pose threats to domestic social and political stability. As such, Chinese technology firms that offer immediate technology solutions to these concerns in the digital information domain have become pragmatic choices. For many countries, Chinese IT companies are pragmatic choices as they lack the technical expertise in building digital infrastructure; these Chinese companies offer an opportunity to leap into the digital economy.

As China possesses advanced knowledge in digital software, hardware, and infrastructure compared to recipient countries, it is also exploiting its relatively lax regulations. While China champions the rhetoric of data sovereignty, the country is also taking advantage of the host countries' relatively loose data protection laws in ASEAN countries, where these host countries do not have as stringent data localization laws as China. As such, cross-border transfer of data is possible with minimal protection over data

[166] Kokas, (n 10). In 2016, Beijing Kunlun Tech Co. Ltd. acquired Grindr, gaining access to sensitive user data, including voice, video, and HIV status, through servers accessible to the Chinese government. While Kokas' examples are drawn from the United States, countries with weak data protection laws are vulnerable to data unintendedly flowing to China without supervision.

[167] Yogesh Hirdaramani, 'Why Data Localisation May Not Be a Panacea for Data Privacy Woes in ASEAN' (*GovInsider*, Oct 03, 2022) <https://govinsider.asia/intl-en/article/why-data-localisation-may-not-be-a-panacea-for-data-privacy-woes-in-asean> accessed 3 February 2025.

and privacy rules, as in the case of Indonesia.[168] During the COVID-19 pandemic, the Chinese government also advanced the 'Health Silk Road, an iteration of the Digital Silk Road, which saw the Chinese government advancing its version of digital health surveillance and its political practices of mass surveillance.[169]

China's Digital Silk Road, a component of the Belt and Road Initiative (BRI), strives to accelerate digitalization in BRI countries. Its objectives encompass promoting the export of Chinese digital products and services, ensuring China-led standardization of next-generation digital technologies, and constructing a cross-border digital network with China as its core.[170] Policymakers in Myanmar and Malaysia have expressed concerns over sovereignty and excessive debt related to Digital Silk Road projects.[171] Beijing's nationalist diplomacy is causing unease among policymakers in these countries about the potential downsides of the Digital Silk Road and Belt and Road Initiative. However, concerns over being left behind in the benefits of transforming into a digital economy have seen countries making pragmatic choices.

As the world becomes more digitalized, China's Digital Silk Road offers an alluring promise of leapfrogging development and enabling recipient countries to benefit from access to digital technologies. China's technical

[168] Esther Sri Astuti, 'The Indonesian Digital Policy: Lessons from PRC's Experiences' (2021) 10 ECIDC Project Paper 1 <https://unctad.org/system/files/official-document/BRI-Project_RP10_en.pdf> accessed 12 December 2024.

[169] Hirdaramani (n 167).

[170] Kaori Iwasaki, 'Chinese Firms Driving Digitalization in the ASEAN Region' (2023) XXIII Pacific Business and Industries 90.

[171] Council on Foreign Relations, 'Assessing China's Digital Silk Road Initiative A Transformative Approach to Technology Financing or a Danger to Freedoms?' (*Council on Foreign Relations*) <https://www.cfr.org/china-digital-silk-road/> accessed 12 December 2024.

assistance is exported along with its concept of data sovereignty, which enables the country to extend its reach and influence.

The concept of data sovereignty strengthens a state's authority over the collection, processing, and storage of data by affirming its autonomy in these areas.[172] It also exerts territorial control over corporations, providing the necessary infrastructure and services for data management. China's approach to sovereignty ensures the government maintains control over digital infrastructure managed by corporations. Thus, data sovereignty becomes a lever for China to dominate its digital infrastructure and corporations within its jurisdiction. When this digital infrastructure is exported to other Global South countries, China's control and influence over these host countries are extended. As China has long advocated for multipolar internet governance, the Digital Silk Road enables China to align and promote its concept of data sovereignty, extending its version of internet governance that advocates for state-controlled internet access. China's concept of data sovereignty thus is exported through the control of digital infrastructure, where China also aims to control data produced within these systems.

According to Professors Matthew Erie and Thomas Streinz, 'data sovereignty is illusory for most developing countries as the power to govern data effectively is dependent on controlling all relevant digital infrastructure, most of which is increasingly being supplied by Chinese technology companies, which are, in turn, operating — to varying degrees — under the influence of the Chinese Communist Party (CCP)'.[173] They argue that data sovereignty is largely unattainable for many developing countries, which paradoxically reinforces the colonial dynamics that China claims to resist and challenge. Coined as "the Beijing Effect," Professors Erie and Streinz argue that China's growing influence in data governance is driven by emerging

[172] Chander and Sun (n 14) 4-5.

[173] Szinay and others (n 55) 5.

economies' demand for digital infrastructure as well as propelled by China-led resistance against the liberal West.[174]

Such a perspective challenges the assumption that data colonialism exclusively concerns exploitation by the Global North of the Global South. Data colonialism, conventionally seen as the extraction by the Global North from the Global South, is also becoming a Global South to Global South phenomenon due to the rising influence of China. While existing literature has documented China's influence over Africa in the digital sphere, it remains to be seen whether such dynamics will be replicated in ASEAN, the world's fifth-largest economy. ASEAN, as an organization, has immense potential in the digital economy, attracting Chinese-owned firms to the region.

Further, this situation also underscores the differences in legal protections over data across countries. Indonesia does not have a general data protection law, whereas China has strong data localization requirements that enable the country to take advantage of the cross-flow of data.[175] Unlike China, which enforces strict data localization laws that restrict cross-border data transfers, Indonesia has fewer of these restrictions. The Chinese Data Security Act applies to Chinese companies overseas and mandates data processing within China. This law is enforced by the Chinese government's Cyber Administration Agency.[176] Consequently, Indonesian personal data could be accessible to third parties, and data gathered by Chinese businesses in Indonesia must be returned to China for processing.

Chinese home-grown companies such as Alibaba, Tencent, and Huawei have increased investments in ASEAN countries due to political uncertainty

[174] ibid.

[175] Mariano (n 33).

[176] Gusty da Costa, 'Chinese investors pour in billions of dollars to Indonesia's digital market' (*Indonesia Business Post*, 24 May 2022) <https://indonesiabusinesspost.com/insider/chinese-investors-are-flooding-indonesias-digital-market-with-billions-of-dollars/> accessed 24 January 2025.

and a saturated market at home.[177] The shift of Chinese information and communication factories to this region has driven a surge in exports; from 2018 to 2022, exports rose by 78 percent.[178]

This increased presence of Chinese companies in ASEAN countries not only signifies a reorganization of the global supply chain but also reflects the impact of the Washington-Beijing tension—where the US is taking actions to reduce economic dependency on China, a move that has caused tension between the two great powers—which is prompting a shift in supply chains to ASEAN countries.[179]

However, this approach poses a significant challenge to the established values and interests of the US and Europe, namely human rights, democracy, and a liberal global legal order.[180] These Western powers, traditionally dominant on the global stage, may find their influence challenged and their values at odds with those pushed by China. This changing dynamic weaves a complex web of interests and power conflicts, placing ASEAN at the center of this competing narrative. It is too early and outside the scope of the article to

[177] Kwangyin Liu, Shu-ren Koo and Silva Shih, 'Five Years on, ASEAN a Winner in U.S.-China Trade war' (*CommonWealth Magazine*, 19 October 2023) <https://english.cw.com.tw/article/article.action?id=3543> accessed January 15, 2024.

[178] ibid.

[179] 'Malaysia Says Asean Is a Winner from Shifting Supply Chains' (*The Straits Times,* 14 November 2024) <https://www.straitstimes.com/asia/se-asia/malaysia-says-asean-is-a-winner-from-shifting-supply-chains> accessed 3 February 2025.

[180] 'United States International Cyberspace & Digital Policy Strategy' (United States Department of State) <https://www.state.gov/united-states-international-cyberspace-and-digital-policy-strategy/> accessed 12 December 2024; Directorate-General for Communication, 'A Safer Digital Future: New Cyber Rules Become Law - European Commission' (*European Commission*, 10 December 2024) <https://commission.europa.eu/news/safer-digital-future-new-cyber-rules-become-law-2024-12-10_en> accessed 12 December 2024.

assess and evaluate the extent to which ASEAN countries will become dependent on China's digital capacities.

While American-based technologies continue to dominate the ASEAN market, the growing influence and cooperation between ASEAN governments and Chinese-owned companies have attracted political attention amid escalating geopolitical tensions between the two great powers. Arguably, these initiatives, through their ties to the Chinese government, function as an extension of the Belt and Road Initiative and China's expanding influence. Consequently, China's export of digital infrastructure also serves as a reconfiguration of sovereignty that supports China's assertion of and claims to a multipolar world.

It is too early to assess how and the extent to which China's interpretation of data sovereignty will impact ASEAN countries, particularly in terms of privacy protection and data mining. The ASEAN region, nonetheless, presents a compelling case study, especially considering its complex history of colonial influences and China's growing role in providing digital health infrastructure. The region's historical context, shaped by different colonial powers, has led to a unique socio-legal-political landscape that continues to develop. This includes the region's earlier invocation— particularly by former Singaporean Prime Minister Lee Kuan Yew—of Asian values as a form of resistance to perceived Western dominance. Interestingly, China's expanding market presence in ASEAN countries could similarly establish and replicate those colonial power dynamics.

VII CONCLUSION

As the global digital health landscape evolves and becomes more complex, there is a need to identify and understand discursive patterns amid the complexities of emerging technologies and the infrastructure supporting their innovations. These three narratives—technological solutionism, international human rights law, and data sovereignty—taken together demonstrate the different facets of international law in facilitating a

particular version of digital health. Each of these narratives sheds light on how international law plays (or does not play) a role in sculpting and shaping the field of digital health. Individually, these narratives do not fully capture the complex nature of digital health governance at national, regional, and global levels. However, when considered collectively, these three narratives cover a particular aspect of digital health governance that is otherwise overlooked.

Arguably, international law's formal response to the challenges brought by digital health technologies has been limited, as shown by the lack of binding legal agreements on digital health. The prevailing narrative of technological solutionism resurfaced during the pandemic, which further accelerated government and international development agencies to devote resources and technical knowledge in this area. However, the narrative of technological solutionism entails several threats in terms of privacy erosion and discriminatory impacts. Several scholars have described these risks with reference to international human rights law. They warn that if the disparate impacts of digital health technologies are not adequately addressed, they could exacerbate inequalities.[181]

However, international human rights law addresses only one facet of digital health, neglecting the political economy of innovations and the development and ownership of digital health infrastructures. While this narrative aims to challenge power disparities through emancipatory language, international human rights law might inadvertently reinforce existing geopolitical power structures. These structures are particularly significant in relation to the ownership and control of digital health

[181] See e.g., United Nations Office of the High Commissioner for Human Rights, 'A/HRC/51/17: The Right to Privacy in the Digital Age' (*OHCHR*) <https://www.ohchr.org/en/documents/thematic-reports/ahrc5117-right-privacy-digital-age> accessed 11 August 2023; 'Digital innovation, technologies and the right to health, Report of the Special Rapporteur on the Right to Health' (n 36).

infrastructure, which has become a proxy for geopolitical power amongst the world's major powers.

Additionally, this issue is particularly significant when considering the extraterritorial application of a country's law over data processing. New lexicons such as 'data colonialism' and 'data sovereignty' have surfaced as the global legal order is being challenged. As lexicons have become a proxy for geopolitical power, they also lead to the redefinition and reinterpretation of long-standing concepts in international law. These new terms are reshaping and redefining traditional concepts of international law. They reflect a growing awareness and concern about the control and use of data by big technology entities and foreign governments and the subsequent impact on national sovereignty to govern data by connecting to the long-standing international law concepts.

As the global landscape of digital health continues to progress and evolve, creating a more intricate web of complexities, there is an increased need to pinpoint and comprehend discursive patterns in understanding the impacts of digital health technologies and their governance for those populations most impacted by health divides.

SELF-DETERMINATION IN THE AGE OF ALGORITHMIC WARFARE

Henning Lahmann [*]

The paper advances the claim that the pervasive surveillance practices employed for the purpose of feeding AI-supported decision-support systems prevent spontaneous and collective political action, thus violating the right to self-determination. Analysing recent events in Gaza and the West Bank, the article describes Israel's utilisation of algorithmic systems in armed encounters with Palestinians, in particular for the purpose of the detecting 'anomalous behaviour'. It claims that because the Israeli security apparatus can point to the legal strictures of IHL targeting rules to rationalise the further entrenchment of surveillance architectures that are necessary for the increasing deployment of machine-learning algorithms, the law of armed conflict functions as a justificatory rhetorical framework for the perpetuated, structural denial of the exercise of the right to self-determination by the Palestinian people. This claim is defended through the conceptualisation of spontaneous political action as advanced by Rosa Luxemburg and Hannah Arendt. Spontaneity is inherent in the idea of collective political agency, which in turn is presupposed in the concept of self-determination as a procedural right to political action. As the algorithmic rationalities of the military and security context inevitably

[*] Henning Lahmann is Assistant Professor at eLaw – Center for Law and Digital Technologies, Leiden University Law School, The Netherlands. I am grateful to Genevieve Lipinsky de Orlov, Kerttuli Lingenfelter, Klaudia Klonowska, and Dimitri van den Meerssche; the discussants Roxana Vatanparast and Thomas Kleinlein at the ESIL Interest Group on International Law and Technology Pre-Conference Workshop at the Annual Conference of the European Society of International Law at Aix-Marseille University, Aix-en-Provence, 30 August 2023; the participants at the DILEMA 2023 Conference at Asser Institute, The Hague, 12 October 2023; the two anonymous reviewers; and the editorial team of the EJLS for their very helpful comments on earlier drafts of the paper that helped to sharpen my arguments. All remaining mistakes and ambiguities are mine.

inhibit the possibility to act spontaneously, the deployment of such systems will thus violate the right to self-determination.

Keywords: Algorithmic warfare; Gaza; Palestine; Israel; Self-determination; International humanitarian law; Military artificial intelligence; Spontaneity

TABLE OF CONTENTS

> And then there is that other assumption, which is terribly dangerous – that we are constant, and that our reactions can be predicted.
>
> Olga Tokarczuk, *Flights*[1]

The senior officer of the Israel Defence Force (IDF) Intelligence Corps was, evidently, rather pleased with himself and his subordinates: in May 2021, Israel's armed forces had just ceased another round of pummelling Gaza with rockets and missiles for eleven days, a campaign during which they had killed, according to the United Nations, around 245 Palestinians, of whom 128 were believed to be civilians, including 63 children.[2] Yet something had been different this time, the officer insisted: 'For the first time, artificial intelligence was a key component and power multiplier in fighting the enemy. [...] We implemented new methods of operation and used technological developments that were a force multiplier for the entire IDF.'[3] Laying claim to having just fought the world's 'first AI war', Israel's military maintained that it had deployed algorithmic systems to conduct and support intelligence, surveillance, and reconnaissance (ISR) activities as well as targeting, using platforms that fused and analysed data from signals, visual, human, and geospatial intelligence to generate predictive recommendations for targets in Gaza in real time.[4] Algorithms for combat drones with names such as 'Alchemist' and 'Gospel', all devised by Intelligence Corps Unit 8200,

---

1 Olga Tokarczuk, *Flights* (Riverhead Books 2018), 15.

2 United Nations Office for the Coordination of Humanitarian Affairs, Occupied Palestinian Territory (oPt): Response to the escalation in the oPt, Situation Report No. 1: 21-27 May 2021.

3 Anna Ahronheim, 'Israel's Operation against Hamas Was the World's First AI War' *The Jerusalem Post* (Jerusalem, 27 May 2021) <https://www.jpost.com/arab-israeli-conflict/gaza-news/guardian-of-the-walls-the-first-ai-war-669371> accessed 16 July 2023.

4 Ibid; Judah Ari Gross, 'IDF Intelligence Hails Tactical Win in Gaza, Can't Say How Long Calm Will Last' *The Times of Israel* (Jerusalem, 27 May 2021) <https://www.timesofisrael.com/idf-intel-hails-tactical-win-over-hamas-but-cant-say-how-long-calm-will-last/> accessed 16 July 2023.

enabled the IDF to strike purported Hamas infrastructure and combatants with increasingly reduced human intervention.[5]

Since this 'Operation Guardian of the Walls' in May 2021, Israel has further expanded the use of AI in its military operations. The IDF reports that the entirety of Gaza is now covered at all times by surveillance balloons[6] and a squadron of unmanned aerial vehicles (UAVs),[7] allegedly enabling intelligence units to constantly produce and locate new targets in a process that now takes a month, rather than the years it took before.[8] Israel has recently begun to extend its drone surveillance programme to the West Bank,[9] an area that had already been blanketed with increasingly "smart" cameras equipped with facial recognition software.[10] Aside from real-time aerial footage and CCTV, Israeli intelligence personnel also deploy

---

[5] Carma Estetieh, 'Israel's Push Towards a "Frictionless" Occupation: A Blessing or a Dystopian Nightmare?' (*Euro-Med Human Rights Monitor*, 3 October 2022) <https://www.euromedmonitor.org/en/article/5358/Israel%E2%80%99s-push-towards-a-%E2%80%9Cfrictionless%E2%80%9D-occupation:-A-blessing-or-a-dystopian-nightmare?>; Yuval Abraham, '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza' (*+972 Magazine*, 3 April 2024) <https://www.972mag.com/lavender-ai-israeli-army-gaza/> accessed 11 April 2024.

[6] Emad Moussa, 'Israeli AI Is Turning Palestine into a Dystopian Reality' (*The New Arab*, 22 June 2023) <https://www.newarab.com/opinion/israeli-ai-turning-palestine-dystopian-reality> accessed 8 August 2023.

[7] Emanuel Fabian, 'Armed Drones Gave IDF "Surgical" Precision During Recent Gaza Fighting, Officers Say' *The Times of Israel* (Jerusalem, 17 August 2022) <https://www.timesofisrael.com/armed-drones-gave-idf-surgical-precision-during-recent-gaza-fighting-officers-say/> accessed 8 August 2023.

[8] Sophia Goodfriend, 'How AI Is Intensifying Israel's Bombardments of Gaza' (*+972 Magazine*, 6 June 2023) <https://www.972mag.com/israel-gaza-drones-ai/> accessed 8 August 2023.

[9] Sophia Goodfriend, 'Drones Terrorized Gaza for Years. Now They'll Do the Same in the West Bank' (*+972 Magazine*, 13 October 2022) <https://www.972mag.com/drones-idf-west-bank-gaza/> accessed 26 March 2023.

[10] Elizabeth Dwoskin, 'Israel Escalates Surveillance of Palestinians with Facial Recognition Program in West Bank' *Washington Post* (Washington, DC, 8 November 2021) <https://www.washingtonpost.com/world/middle_east/israel-palestinians-surveillance-facial-recognition/2021/11/05/3787bf42-26b2-11ec-8739-5cb6aba30a30_story.html> accessed 27 June 2023.

algorithms to continuously monitor Palestinians' online activities[11] and routinely collect cell phone location data.[12] In June 2023, the head of the IDF's cyber division voiced his expectation that in a few years' time, 'every area of warfare [conducted by the IDF] will be based on generative AI information'.[13]

Taking events in Gaza and the West Bank between May 2021 and October 2023 as the principal point of departure for its analysis, this article provides a detailed description of the salient points of Israel's utilisation of algorithmic systems in armed encounters with Palestinians. Based on this account, the article claims that because the Israeli security apparatus can invoke the legal requirements of international humanitarian law (IHL) targeting rules to rationalise pervasive and constant surveillance to sustain the deployment of machine-learning algorithms, the law of armed conflict has assumed the function of a justificatory rhetorical framework for the perpetuated, structural denial of the exercise of the right to self-determination by the Palestinian people.[14] I base this claim on the conceptualisation of spontaneous political action as advanced in the works of Rosa Luxemburg and Hannah Arendt. I demonstrate that spontaneity is inextricable from the idea of collective political agency, which in turn is presupposed in self-determination as a procedural right to political action. As the algorithmic

[11] Melanie Swan, 'Israel Develops "Cyber Iron Dome" to Find Terrorists Online' *The Times* (London, 8 August 2023) <https://www.thetimes.co.uk/article/israel-develops-cyber-iron-dome-to-find-terrorists-online-h8gwsxjpw> accessed 8 August 2023.

[12] Goodfriend (n 9).

[13] Yonah Jeremy Bob, 'IDF Will Run Entirely Generative AI Very Soon – Israeli Cyber Chief' *The Jerusalem Post* (Jerusalem, 28 June 2023) <https://www.jpost.com/israel-news/defense-news/article-748028> accessed 8 August 2023.

[14] It bears noting at the outset that this argument in no way intends to interfere with the more general, and correct, observation that Israel's indefinite occupation of Palestinian lands violates the Palestinian people's right to self-determination in and of itself; see on this only Ralph Wilde, 'Using the Master's Tools to Dismantle the Master's House: International Law and Palestinian Liberation' (2021) 22 The Palestine Yearbook of International Law Online 1.

rationalities of the military and security context inevitably inhibit the possibility to act spontaneously, it follows that the deployment of such systems will come to violate the collective right to self-determination.

The first draft of this article was finalised and submitted on 9 August 2023. Almost exactly two months later, Hamas and the Palestinian Islamic Jihad breached the highly fortified outer perimeter of Gaza, launching a devastating attack against IDF military bases, kibbutzim and other communities in southern Israel, as well as a music festival, killing approximately 1,139 people (including 36 children, 71 foreign nationals, and 373 members of Israeli security forces) and taking around 250 hostages.[15] Shortly thereafter, Israel responded with overwhelming and, at the time of finalising a revised version, still ongoing military force through the air and by means of a ground invasion of Gaza that began on 27 October 2023. Up until 7 October 2024, according to the Hamas-controlled health ministry in Gaza, the IDF's all-out campaign had killed at least 41,870 Palestinians, the overwhelming majority of them civilians.[16] While several aspects concerning the terrorist attack itself and the reaction to it did make a careful re-evaluation of the arguments advanced in this article necessary, both the core premises and the principal conclusions derived from the theoretical framework conceived in the following sections have lost none of their validity or explanatory power.

The argument unfolds in the following four steps. Section 1 begins by describing the increasing use of machine learning technologies in military decision support systems. While the focus is on Palestinian territories as a salient case to expose the particulars and intentionalities of such technologies and the related data practices, it also points to the broader implications of

[15] France 24, 'Israel Social Security Data Reveals True Picture of Oct 7 Deaths' (15 December 2023) <https://www.france24.com/en/live-news/20231215-israel-social-security-data-reveals-true-picture-of-oct-7-deaths> accessed 15 March 2024.

[16] Al Jazeera, 'One Year of Israel's War on Gaza: Key Moments Since October 7' (7 October 2024) <https://aje.io/crs9jl> accessed 8 October 2024.

such developments. After laying out how the current regime of IHL, especially the law of targeting, can be used to rationalise the further use of algorithms and big data, the third sub-section explains how recourse to the rules of IHL has helped to obscure one of the principal use cases of machine learning in this context, which is the process of anomaly detection as opposed to "simple" target identification and verification.

Section 2 critiques emerging scholarly interventions that have responded to the algorithmic data practices by militaries and intelligence agencies as described in Section 1 by applying the conceptual framework of privacy and data protection. Although such attempts are helpful in shedding light on some of the more egregious and consequential misuses of personal data for the purposes of warfare, the basic principles of machine learning render this particular analytical lens ultimately futile while deflecting from the more fundamental and problematic aspects of the described uses of machine learning algorithms.

Building on this assessment, Section 3 analyses the consequences of the workings of warfare algorithms through the concept of spontaneous political action as developed by Rosa Luxemburg and Hannah Arendt. After reappraising the collective right to self-determination as (also) amounting to a primordial procedural right to political practice, the paper explicates the critical role of spontaneity for any emancipatory politics in the understanding of the two political theorists. Based on this investigation, Section 4 synthesises the previous findings by advancing the argument that systems of algorithmic warfare suppress the spontaneous and collective political will-formation that is the condition of possibility of the exercise of self-determination.

I. THE VISION OF ALGORITHMIC WARFARE IN PALESTINE

Israel has erected vast and ever-expanding surveillance architectures that constantly collect new data to feed the models of an array of algorithmic military decision-support systems to sustain the administration and control

of the occupied Palestinian territories.[17] This mode of security governance is pursued with the objective of detecting threats before they can materialise, directly furthering the security interests of the occupying power both in the territories under its control and in its own adjacent territory. This practice is rhetorically justified by recourse to the rationales of the core rules of the law of armed conflict – a critical connection that will emerge more clearly from the following explication.

1. War Algorithms

The idea that AI – understood as any system 'capable of learning, reasoning and problem-solving'[18] – is set to revolutionise all facets of military affairs has already become a cliché.[19] Among many other armed forces, the IDF has started working towards systematically incorporating AI-based applications across the entire organisation.[20] So-called decision-support systems (DSS) have assumed a particularly prominent position in the strategic considerations on the integration of AI technologies in light of the increasing complexity of contemporary armed conflicts. AI-based DSS are broadly understood as algorithmic systems that are capable of assisting

[17] I will omit a deeper discussion of whether Gaza is to be considered remaining under military occupation, as this question is immaterial for the arguments presented here. For treatments of this question, see most recently the ICJ in *Legal Consequences Arising from the Policies and Practices of Israel in the Occupied Palestinian Territory, Including East Jerusalem*, Advisory Opinion of 19 July 2024, at paras. 86–94; for earlier scholarly examinations see e.g. Shane Darcy and John Reynolds, 'An Enduring Occupation: The Status of the Gaza Strip from the Perspective of International Humanitarian Law' (2010) 15 Journal of Conflict and Security Law 211; Yuval Shany, 'Binary Law Meets Complex Reality: The Occupation of Gaza Debate' (2008) 41 Israel Law Review 68.

[18] International Organization for Standardization (ISO), 'What Is Artificial Intelligence (AI)?' <https://iso.org/artificial-intelligence/what-is-ai>.

[19] Paul Scharre, 'AI's Inhuman Advantage' (*War on the Rocks*, 10 April 2023) <https://warontherocks.com/2023/04/ais-inhuman-advantage/> accessed 10 April 2023.

[20] Seth Frantzman, 'Israel Unveils Artificial Intelligence Strategy for Armed Forces' (*C4ISRNet*, 11 February 2022) <https://www.c4isrnet.com/artificial-intelligence/2022/02/11/israel-unveils-artificial-intelligence-strategy-for-armed-forces/> accessed 26 March 2023.

military decision-makers at every step, from gathering and analysing intelligence and suggesting possible courses of action, to identifying and marking military objectives in armed engagements.[21] The emergent technologies underlying AI-based DSS are big data and machine learning.

Whereas the concept of "big data" broadly describes the accumulation and analysis of massive swathes of data from a variety of sources,[22] machine learning is the currently prevalent methodology of training algorithms tasked to parse these large databases. Machine learning systems are trained on vast amounts of data that allow them to build their own models to effect certain outcomes instead of operating on the processing of pre-programmed rules, as was the case with earlier generations of AI. This means that the output depends on a number of variant and interdependent factors, such as the type of learning process and the resulting model, which is a function of the data with which the algorithm is fed. In other words, machine learning is a type of statistical analysis based on the principle of induction.[23] It follows that the output is always a prediction based on the discovery of patterns, that is, links and correlations between data points. Machine learning algorithms attempt to 'detect the mathematical target function that properly describes a dataset, hoping that the function will apply to new data'.[24] One crucial distinction is between supervised and unsupervised learning. For the former, human operators will first label input data (e.g. pictures of cats) to indicate

---

[21] Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare' (17 March 2021), 15 <https://papers.ssrn.com/abstract=3823881> accessed 26 June 2023.

[22] Shiri Krebs, 'Predictive Technologies and Opaque Epistemology in Counterterrorism Decision-Making' in Arianna Vedaschi and Kim Lane Scheppele (eds), *9/11 and the Rise of Global Anti-Terrorism Law: How the UN Security Council Rules the World* (Cambridge University Press 2021), 205.

[23] Erik J Larson, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (The Belknap Press of Harvard University Press 2021), 118; first-generation AI was based on deductive frameworks.

[24] Mireille Hildebrandt, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20 Theoretical Inquiries in Law 83, 85.

the patterns that constitute the desired predictive output (e.g. the correct identification of cats in a large set of visual data). Such classification tasks are common types of machine learning algorithms. When the system is programmed to find patterns in the data on its own, this is called unsupervised learning.[25]

Military decision support has long been marked as especially fit for exploiting the purported advantages of machine learning technologies. Ever since the 9/11 attacks prompted the 'datafication of counter-terrorism',[26] the amount of data recording the behaviour of individuals collected by intelligence agencies has grown to such an extent that human analysts have simply lost the capacity of making sense of the amassed information.[27] Personal data scraped from social media and other online communication is combined with visual or audio-visual feeds from sensors mounted on satellites in geostationary or low earth orbit or drones that autonomously cover a wide range of territory, complemented by a rapidly expanding array of internet-of-things devices that effectively act as remote sensors.[28] In effect,

[25] Larson (n 23), 133–4.

[26] Fionnuala D Ni Aolain, 'The Datafication of Counter-Terrorism' in Laura A Dickinson and Edward W Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023) <https://papers.ssrn.com/abstract=4083433> accessed 7 July 2023. See in this context also the revelations by Edward Snowden, about them e.g. David Lyon, 'Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique' (2014) 1 Big Data & Society 2053951714541861.

[27] Adam Frisk, 'What Is Project Maven? The Pentagon AI Project Google Employees Want out Of' (*Global News*, 5 April 2018) <https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/> accessed 8 August 2023.

[28] Nishwan S Smagh, 'Intelligence, Surveillance, and Reconnaissance Design for Great Power Competition' (Congressional Research Service 2020) R46389 <https://fas.org/sgp/crs/intel/R46389.pdf> accessed 8 August 2023; Ed Stacey, 'Future Warfighting in the 2030s: An Interview with Franz-Stefan Gady' (*Strife*, 9 September 2020) <https://www.strifeblog.org/2020/09/09/future-warfighting-in-the-2030s-an-interview-with-franz-stefan-gady/> accessed 8 August 2023; Richard H Schultz and Richard D Clarke, 'Big Data at War: Special Operations Forces, Project Maven, and Twenty-First Century

everything has become a prospective source to continuously feed the deluge of big data.[29] It takes machine learning algorithms to parse such amounts of data to put out predictions aimed at raising strategic, operational, or situational awareness for military commanders.[30]

The latest generation of these technologies, so-called platform-independent fusion architectures, can integrate an even greater variety of sensors and other sources whose data streams are dynamically analysed in real time, instantly providing a large network of connected military assets and units with suggested courses of action in the theatre of conflict.[31] Such 'battlefield management systems' are imagined to produce an accurate and comprehensive operating picture at all times, crucially including the ability to reliably classify and identify objects and persons encountered in the field.[32]

Israel has been among the first states to fully embrace the promises of machine learning for its own security purposes. For years now, Israel's intelligence services have penetrated Palestinian communications networks to surveil all online activities by the population located in the territories and to build models for algorithms based on the data streams being constantly

---

Warfare' (*Modern War Institute*, 25 August 2020) <https://mwi.usma.edu/big-data-at-war-special-operations-forces-project-maven-and-twenty-first-century-warfare/> accessed 8 August 2023.

[29] Jessica Bayley, 'Transforming ISR Capabilities through AI, Machine Learning and Big Data: Insights from Dr. Thomas Killion, Chief Scientist, NATO' (*Defence IQ*, 30 July 2018) <https://www.defenceiq.com/defence-technology/news/transforming-isr-capabilities-through-ai-machine-learning-and-big-data> accessed 8 August 2023.

[30] Margarita Konaev, 'With AI, We'll See Faster Fights, But Longer Wars' (*War on the Rocks*, 29 October 2019) <https://warontherocks.com/2019/10/with-ai-well-see-faster-fights-but-longer-wars/> accessed 18 July 2023.

[31] Arthur Holland Michel, 'There Are Spying Eyes Everywhere – And Now They Share a Brain' (*Wired*, 4 February 2021) <https://www.wired.com/story/there-are-spying-eyes-everywhere-and-now-they-share-a-brain/> accessed 18 July 2023.

[32] Jackson Barnett, 'Air Force Moving Project Maven into Advanced Battle Management System Portfolio' (*FedScoop*, 10 August 2020) <https://www.fedscoop.com/project-maven-air-forces-advanced-battle-management-system/> accessed 18 July 2023.

collected.[33] Big data analysis of young Palestinians' behaviour on social media combined with other intelligence sources was allegedly the critical factor in ending a string of knife attacks by individuals acting alone over the course of 2015: the algorithmic assessment led to the preventative detention of a large number of minors accused of having planned assaults.[34] Recently, Isarel's domestic intelligence service Shin Bet has begun talking about a comprehensive 'cyber Iron Dome' that will further expand such online monitoring.[35] In the West Bank in particular, these measures are complemented by a vast network of cameras that are now equipped with facial recognition software, a technology that is likewise based on machine learning.[36] UAVs, such as the Elbit Hermes 450 drone, and balloons provide a permanent feed of visual data from the Palestinian territories.[37] Since representatives of the IDF have recently revealed the existence of fusion architectures that use 'all of our existing advanced sensors and sources'[38] to train models for the generation of 'a common operating picture for the armed forces',[39] one may assume that all of these different data practices now

[33] Asaf Lubin, 'The Duty of Constant Care and Data Protection in War' (2022) <https://papers.ssrn.com/abstract=4012023> accessed 15 June 2023, 6.

[34] Amos Harel, 'How Israel Stopped a Third Palestinian Intifada' *Haaretz* (4 October 2019) <https://www.haaretz.com/israel-news/2019-10-04/ty-article/.premium/how-israel-stopped-a-third-palestinian-intifada/0000017f-e355-df7c-a5ff-e37f99d30000> accessed 8 August 2023.

[35] Swan (n 11).

[36] Moussa (n 6); Keren Weitzberg, 'Biometrics and Counter-Terrorism: Case Study of Israel/Palestine' (Privacy International 2021) <https://privacyinternational.org/report/4527/biometrics-and-counter-terrorism-case-study-israelpalestine> accessed 8 August 2023.

[37] Fabian (n 7).

[38] Yaakov Lappin, 'IDF Identifies "As Many Targets in a Month as It Did in a Year"' (*Jewish News Syndicate*, 4 December 2022) < https://www.jns.org/idf-identifies-as-many-targets-in-a-month-as-it-did-in-a-year/> accessed 8 August 2023.

[39] Frantzman (n 20).

feed the same assembled systems.[40] Most recently, revelations published in early 2024 about the use of the 'Lavender' system for the production of targets in Israel's war against Hamas in Gaza has confirmed these suspicions,[41] despite some scholars with purported inside knowledge disputing some of the factual assertions and inferences.[42]

While the 2021 campaign against Gaza may have established Israel as the avant-garde in developing and actively deploying these capabilities,[43] an assessment reinforced by reports on the widespread reliance on algorithmic decision support during its 2023/24 campaign against Hamas, other recent events have shown states' growing incentives to exploit scientific progress in AI for battlefield applications. A salient catalyst for the wider acceptance and creeping normalisation of algorithmic practices in contemporary warfare has proven to be Ukraine's desperate attempt to fend off Russian military aggression since Russia's full-scale invasion in March 2022.[44] A June

---

[40] To be sure, not all of the predictions put out by these machine learning algorithms lead to targeting decisions. Some of them "merely" result in detention. See Orr Hirschauge and Hagar Shezaf, 'How Israel Jails Palestinians Because They Fit the "Terrorist Profile"' *Haaretz* (31 May 2017) <https://www.haaretz.com/israel-news/2017-05-31/ty-article-magazine/.premium/israel-jails-palestinians-who-fit-terrorist-profile/0000017f-f85f-d044-adff-fbff5c8a0000> accessed 21 July 2023.

[41] Abraham (n 5); Christopher Elliott, 'Expedient or Reckless? Reconciling Opposing Accounts of the IDF's Use of AI in Gaza' (*Opinio Juris*, 26 April 2024) <https://opiniojuris.org/2024/04/26/expedient-or-reckless-reconciling-opposing-accounts-of-the-idfs-use-of-ai-in-gaza/> accessed 30 April 2024.

[42] Tal Mimran and Gal Dahan, 'Artificial Intelligence in the Battlefield: A Perspective from Israel' (*Opinio Juris*, 20 April 2024) <https://opiniojuris.org/2024/04/20/artificial-intelligence-in-the-battlefield-a-perspective-from-israel/> accessed 6 May 2024.

[43] Avi Kalo, 'AI-Enhanced Military Intelligence Warfare Precedent: Lessons from IDF's Operation "Guardian of the Walls"' (*Frost & Sullivan*, 9 June 2021) <https://www.frost.com/frost-perspectives/ai-enhanced-military-intelligence-warfare-precedent-lessons-from-idfs-operation-guardian-of-the-walls/> accessed 5 December 2022.

[44] See Bruno Maçães, 'How Palantir Is Shaping the Future of Warfare' (*Time*, 10 July 2023) <https://time.com/6293398/palantir-future-of-warfare-ukraine/> accessed 1 August 2023; Will Knight, 'The AI-Powered, Totally Autonomous Future of War Is Here' [2023] *Wired*

2023 article in the Atlantic approvingly noted notorious U.S.-based tech company Palantir's cooperation with Kiev to provide Ukrainian forces with its latest software for targeting assistance based on various machine-learning algorithms.[45] In turn, Palantir has begun to aggressively promote the product to a wider future customer base.[46] Debates to start harnessing the potentials of AI-based applications in the armed forces have also been ongoing among Member States of the European Union (EU) since at least 2019, when Finland, Estonia, France, Germany, and the Netherlands issued the joint discussion paper 'Digitalization and Artificial Intelligence in Defence'.[47] At the same time, the EU has been trying to position itself as a leading voice in the emphasis on the ethically and legally responsible development of the technology,[48] including by way of government-funded research projects in various Member States.[49]

<https://www.wired.com/story/ai-powered-totally-autonomous-future-of-war-is-here/> accessed 1 August 2023.

[45] Anne Applebaum and Jeffrey Goldberg, 'Zelensky's Plan to Defeat Russia—And Take Back Crimea' [2023] *The Atlantic* <https://www.theatlantic.com/magazine/archive/2023/06/counteroffensive-ukraine-zelensky-crimea/673781/> accessed 4 May 2023.

[46] Matthew Gault, 'Palantir Demos AI to Fight Wars but Says It Will Be Totally Ethical Don't Worry About It' (*Vice*, 26 April 2023) <https://www.vice.com/en/article/qjvb4x/palantir-demos-ai-to-fight-wars-but-says-it-will-be-totally-ethical-dont-worry-about-it> accessed 1 May 2023. Another dubious company that has seized on the opportunity provided by the invasion to mend its image is Clearview AI, see <https://www.clearview.ai/ukraine> accessed 1 May 2023.

[47] See on this Justinas Lingevicius, 'Military Artificial Intelligence as Power: Consideration for European Union Actorness' (2023) 25 Ethics and Information Technology 18.

[48] See Vincent Boulanin et al., 'Responsible Military Use of Artificial Intelligence: Can the European Union Lead the Way in Developing Best Practice?' (*SIPRI*, November 2020) <https://sipri.org/publications/2020/policy-reports/responsible-military-use-artificial-intelligence-can-european-union-lead-way-developing-best> accessed 13 August 2023.

[49] See e.g. in the Netherlands the ELSA Lab Defence <https://elsalabdefence.nl> accessed 13 August 2023.

2. *Imperative Surveillance: The Law of Targeting as a Justificatory Rhetorical Framework for AI*

Ostensibly, the new age of algorithmic warfare is to the benefit of everyone. Taking political and military decision-makers at their word, one might be forgiven for concluding that the advancement of AI technologies in military decision-support systems is almost exclusively motivated by the universal expectation that their widespread deployment will soon usher in a new era of completely sanitised warfare.[50] In reporting on Israel's recent technological gains, virtually no news outlet forgot to echo what IDF representatives have been repeating ad nauseam: that the algorithmically enabled targeting processes are ultimately being pursued only with the Palestinians' best interests in mind. In the Israeli armed forces' telling, the use of advanced AI will greatly enhance the precision of weapon systems, and thus minimise any unintended consequences of strikes against militants.[51] As one senior IDF official alleged, '[w]e always aim for low collateral damage. That is our assumption. Keeping that as a constant, and doing a lot more, means you *have* to be using advanced algorithms'.[52] According to the IDF Chief of Staff, it is thanks to such technological advantages that recent engagements with Palestinians in Gaza prior to October 2023 allegedly had 'the lowest combatant-to-civilian casualty ratio in the world'.[53]

---

[50] In this, the AI narrative of course only further reinforces the already familiar trope in favour of unrestricted drone warfare against terrorist suspects, see only Daniel L Byman, 'Why Drones Work: The Case for Washington's Weapon of Choice' (*Brookings*, 30 November 1AD) <https://www.brookings.edu/articles/why-drones-work-the-case-for-washingtons-weapon-of-choice/> accessed 18 June 2023.

[51] Kalo (n 43).

[52] Frantzman (n 20) (emphasis added).

[53] Lappin (n 38); after the start of Israel's campaign against Hamas in Gaza in October 2023, several scholars suggested that the IDF to a large extent dispensed with all pretences of being guided by the principle of minimising civilian harm, see only Janina Dill, 'Law and Survival in Israel and Palestine' (*Just Security*, 26 October 2023) <https://www.justsecurity.org/89767/law-and-survival-in-israel-and-palestine/> accessed

That the deployment of machine learning algorithms in targeting systems will save many civilian lives is not an argument advanced exclusively by Israel. Quite the contrary, the claim has already assumed the status of received wisdom. Deeply impressed by the latest technological progress, the consensual outcome document of the 2023 high-level global Summit on Responsible Artificial Intelligence in the Military Domain (REAIM) explicitly recognises 'the potential of AI applications in the military domain for a wide variety of purposes, at the service of humanity, including AI applications to reduce the risk of harm to civilians and civilian objects in armed conflicts'.[54] Such optimistic official declarations are now regularly underwritten by emphatic academic engagement. Given the inescapable limitations of human cognitive capabilities and psychological flaws, one recent paper contends that *not* exploiting the potential of machine learning in warfare 'would be irresponsible and unethical'.[55] Indeed, on this view even the most contentious of such technologies, fully autonomous weapon systems, 'will eventually be able to use lethal force far more humanely than human soldiers ever have or ever will'.[56]

The rationale guiding such evaluations is not simply based on ethical positioning but directly flows from a particular framing of applicable legal

26 October 2023. Instead, in the weeks after Hamas' massacres in Southern Israel it quickly became apparent that the existing AI-powered decision-support systems like "Gospel" and "Lavender" were appreciated primarily for their ability to vastly accelerate the production of new targets during the ongoing campaign rather than to increase precision for the benefit of civilian lives in Gaza, see Abraham (n 5); Yuval Abraham, '"A Mass Assassination Factory": Inside Israel's Calculated Bombing of Gaza' (*+972 Magazine*, 30 November 2023) <https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/> accessed 2 December 2023.

[54] REAIM 2023 Call to Action (16 February 2023), para. 2 <https://www.government.nl/documents/publications/2023/02/16/reaim-2023-call-to-action> accessed 8 August 2023.

[55] HW Meerveld and others, 'The Irresponsibility of Not Using AI in the Military' (2023) 25 Ethics and Information Technology 14.

[56] Kevin Jon Heller, 'The Concept of "The Human" in the Critique of Autonomous Weapons' (30 January 2023) <https://papers.ssrn.com/abstract=4342529> accessed 3 February 2023.

standards. The body of international humanitarian law mandates the protection of civilians in armed conflict to the greatest extent possible. If machine learning algorithms can ensure such outcomes, as more and more observers contend,[57] then for its proponents it follows that the widespread use of AI is not a matter of choice but is necessary for a state to comply with its legal duties.[58]

The set of legal obligations that provides this justificatory rhetorical framework can be found in Additional Protocol I to the Geneva Conventions (AP I),[59] as well as in corresponding customary international law. At its foundation lies the principle of distinction, set out in Article 48 AP I:

> In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.

Accordingly, civilians may never be directly targeted unless and for such time as a civilian takes direct part in hostilities. The obligation to distinguish is complemented by the principle of proportionality, which prohibits attacks that are 'expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated', as stipulated in Article 51(5)(b) AP I.

The principle of distinction is further fleshed out by the third pivotal rule of IHL targeting law, the principle of precautions in attack. Article 57(1) AP I

---

[57] See only ibid.

[58] For a detailed discussion of the IHL aspects in the context of AI and machine-learning see only Shivam Kumar Pandey and Anditya Narayan, 'Means and Methods of Warfare and International Humanitarian Law in the Age of Artificial Intelligence and Machine Learning' (2021) 5 International Journal of Legal Science and Innovation 160.

[59] Protocol Additional of 10 June 1977 to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (AP I), Article 51(5)(b).

provides that 'in the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects', thus setting up an 'obligation of conduct, i.e. a positive and continuous obligation aimed at risk mitigation and harm prevention and the fulfilment of which requires the exercise of due diligence'.[60] As the reference to the broad category of 'military operations' implies, the obligation should be interpreted as applying not only to 'attacks' but conduct by armed forces more generally.[61] For attacks specifically, Article 57(2)(a)(i) AP I obliges military commanders planning or deciding on an attack to do 'everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives'. This provision is usually interpreted as mandating the collection of reliable intelligence as well as the conduct of surveillance and reconnaissance in the theatre of conflict to ensure that only legitimate targets are attacked.[62] The purpose of the rule is to spare civilians to the furthest extent possible.[63] A corresponding duty follows from the principle of proportionality: any reasonable calculation of possible harm to civilians requires a detailed and up-to-date picture of the target area regarding the presence of any legally protected persons or objects.[64] For the particular context of the practice of so-called 'targeted killings' carried out by the IDF in the Palestinian territories, in 2006 the High Court of Justice of Israel likewise clarified that

[60] International Law Association Study Group on the Conduct of Hostilities in the 21st Century, 'The Conduct of Hostilities and International Humanitarian Law. Challenges of 21st Century Warfare' (2017) 93 International Legal Studies 322.

[61] Lubin (n 33), 10; Eliza Watt, 'The Principle of Constant Care, Prolonged Drone Surveillance and the Right to Privacy of Non-Combatants in Armed Conflicts' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (CCDCOE 2022), 169.

[62] Watt (n 61), 168; Asaf Lubin, 'The Reasonable Intelligence Agency' (2021) 47 The Yale Journal of International Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3805700> accessed 26 July 2023.

[63] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (1987) 680.

[64] Watt (n 61), 168.

'[i]nformation which has been most thoroughly verified is needed' when determining whether a civilian is actively taking part in hostilities and can thus be considered a legitimate target.[65]

While the obligation stemming from the principle of precautions does not require a hundred percent certainty before an attack may be carried out, the duty to verify targets by means of intelligence, surveillance and reconnaissance (ISR) is contingent on the technological capabilities and resources of the forces.[66] As put by Rosen, advanced equipment 'must be used if it is available, makes good military sense and will minimize civilian impact'.[67] To the extent that it is true that machine learning algorithms deployed in military assets – for example, facial recognition technologies – in fact 'significantly enhance a military commander's capacity to identify the enemy and comply with their humanitarian obligations',[68] it may be inferred that their use to the extent that is possible and feasible forms part of the obligation under Article 57(2)(a)(i) AP I. Even if this implication is considered too far-reaching, in any case the rules on targeting can be invoked to justify the deployment of such systems even if their primary objective is an increase in military efficiency and not in fact the protection of civilians.[69]

In turn, this alleged legal imperative to deploy algorithmic decision support systems necessarily entails the requirement to ramp up the collection of data. Since the foundational principle of machine learning is the detection of a

---

[65] The Public Committee against Torture in Israel et al. v. The Government of Israel et al., HCJ 769/02, 14 December 2006, para. 40.

[66] Jean-François Quéguiner, 'Precautions under the Law Governing the Conduct of Hostilities' (2006) 88 International Review of the Red Cross 793, 797.

[67] Frederik Rosen, 'Extremely Stealthy and Incredibly Close: Drones, Control and Legal Responsibility' (2014) 19 Journal of Conflict & Security Law 113, 127.

[68] Leah West, 'Face Value: Precaution versus Privacy in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (CCDCOE 2022) 140.

[69] See the recent discussions surrounding the IDF's 'Lavender' system, Abraham (n 5).

target function that accurately describes a dataset so as to be applicable to new data, the likelihood of this to be the case rises with the volume of available data.[70] Despite the more recent experimentation with different approaches, so far it remains true that the only consistently and demonstrably reliable method to ensure that machine learning systems are validated against the widest possible degree of variance in data is to increase the size of the data sets on which they are trained and tested.[71] Critically, the reliability of predictive outputs does not simply improve with increasing the sheer amount of input data but heavily depends on the quality of the data that is used to train the model, which above all must be representative of the eventual operational environment.[72] In other words, a decision support algorithm for targeting recommendations that was trained on data from Afghanistan will be highly error-prone when deployed in Mali. If data on the particular context is non-existent, the related output will necessarily fail to produce meaningful predictions.[73] To avoid such a situation, militaries that consider relying on machine learning are incentivised to at all times 'preserv[e] all relevant data in useable form for future algorithms', as Deeks recommended a few years ago.[74] It follows that the more decision support tasks are handed over to machine learning algorithms, the more states can invoke the argument that the success of such operations, in the sense of both the meeting of military objectives and the sparing of civilian lives to discharge the legal obligations imposed by IHL, is directly contingent on the collection of contextually relevant, accurate, and high-quality data. And

[70] Hildebrandt (n 24) 85.

[71] Arthur Holland Michel, 'Known Unknowns: Data Issues and Military Autonomous Systems' (United Nations Institute for Disarmament Research 2021), 27.

[72] Klonowska (n 21) 19.

[73] Avi Goldfarb and Jon R Lindsay, 'Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War' (2022) 46 International Security 7, 19–20.

[74] Ashley S Deeks, 'Detaining by Algorithm' (*Humanitarian Law & Policy*, 25 March 2019) <https://blogs.icrc.org/law-and-policy/2019/03/25/detaining-by-algorithm/> accessed 17 July 2023.

for technical reasons, this can only be achieved through constant and pervasive multi-source surveillance of the population in the target area.[75]

3. The Logic of Anomaly

The above-described prevalent framing of AI-based DSS enabling militaries to enhance the protection of civilians and thus to increase compliance with the core rules of IHL, however, obscures what is in fact one of the primary purposes of algorithmic security governance by means of pervasive surveillance. Prior to assigning a machine learning system with the task of verifying the identity of an object or person of interest in order to distinguish protected entities from military objectives, the object or person must have been discovered and identified – or *suspected* – as a potential target in the first place.[76] In a technical sense, different ways are conceivable for an algorithm to accomplish such a task. The 'Lavender' system deployed by the IDF during its campaign against Hamas in Gaza after 7 October 2023, for instance, works by finding markers in the input data that designate a person as a Hamas member based not on visual identifiers such as uniforms or the carrying of weapons but on generated ratings made up of '"hundreds and thousands" of features' detected in the data, for example 'being in a Whatsapp group with a known militant, changing cell phone every few months, and changing addresses frequently'.[77] Another, even more striking variety of AI-supported security governance is the use of machine learning algorithms that parse the masses of data collected through multi-source surveillance to engage in an operation that has come to be known as 'anomaly detection'. Amply utilised in other contexts, such as the uncovering of fraudulent bank transactions, anomaly detection is based on an analysis of frequencies, exploiting the fact that models can establish what is assumed to be a state (or

[75] Henning Lahmann, 'The Future Digital Battlefield and Challenges for Humanitarian Protection: A Primer' (2022) 21.

[76] Klonowska (n 21), 18.

[77] Abraham (n 5).

pattern) of normality in a large dataset and then identify patterns that diverge from that state.[78]

As scholars have previously pointed out, attempting to algorithmically detect correlations between points in large sets of data that somehow stand out from what the algorithm has, through machine learning, determined to be the 'normal' state of things has become the principal means to discover suspicious persons or objects.[79] The idea is that once the algorithm has flagged an anomaly, this first suspicion can be analysed further,[80] which now usually implies not human intervention but the system itself seamlessly translating the anomaly into a suspicious pattern of behaviour that suggests a potential 'lawful target'.[81]

In the eyes of military and intelligence agencies, the genius of this method is that what makes a pattern stand out is inherently impossible to predetermine – the algorithm can detect anomalies that a human would never notice. Such deviations from the state (or pattern) of normality as described above may be as 'mundane and even absurd'[82] as 'the time or length of a phone call, an overnight stay, or rare use of a mobile device';[83] they may be some insignificant display of 'hostile or benign intent of individuals in a

[78] Larson (n 23) 150–1. See in the context of the EU Passenger Name Record directive CJEU, Judgment of 21 June 2022, *Ligue des Droits Humains v. Conseil des Ministres*, C-817/19, EU:C:2022:491, at paras. 194–5.

[79] Claudia Aradau and Tobias Blanke, *Algorithmic Reason: The New Government of Self and Other* (Oxford University Press 2022) 71.

[80] Ashley S Deeks, 'Predicting Enemies' (2018) 104 Virginia Journal of International Law 1529, 1560.

[81] Nicola Perugini and Neve Gordon, 'Distinction and the Ethics of Violence: On the Legal Construction of Liminal Subjects and Spaces' (2017) 49 Antipode 1385, 1386; Geoff Gordon, Rebecca Mignot-Mahdavi and Dimitri Van Den Meerssche, 'The Critical Subject and the Subject of Critique in International Law and Technology' (2023) 117 AJIL Unbound 134, 135.

[82] Tasniem Anwar and Klaudia Klonowska, 'Co-Production of Terrorist Suspects: The Role of Law in Associating Security Assemblages' (2023) 15.

[83] Aradau and Blanke (n 79) 71.

crowded street' detected by an 'emotional prediction' algorithm,[84] perhaps by registering 'facial expressions, characteristics, involuntary gestures, and estimated heart rate' that somehow do not correspond with whatever is supposed to be normal in the system's model of the world.[85] The promise of this approach has long been recognised by Israeli security agencies in its suppression of violent Palestinian resistance.[86] As we have seen above, it was the detection of 'unusual activity' by young Palestinians on social media that allegedly allowed the Shin Bet to pre-empt the continuation of knife attacks carried out by lone perpetrators in 2015 by detaining a large number of suspects thus 'identified'.[87] The logic of anomaly was also at the heart of the algorithmic early warning systems as part of Israel's separation barrier with Gaza; obviously, a terrorist attack such as the one unfolding on the morning of 7 October 2023 was precisely the type of incident that the vast and pervasive surveillance architectures in and above Gaza were supposed to render virtually impossible. But as experts were quick to point out, there is such a thing as too much surveillance:[88] whether the algorithms did not pick up on any anomalies,[89] or whether the machine outputs were ignored or misinterpreted is not (yet) clear, though early media reports suggested that pervasive misogyny within the IDF was among the principal reasons why correctly identified clues, probably first flagged by algorithms parsing surveillance video footage, got stuck in the chain of command because the

[84] Goldfarb and Lindsay (n 73) 37.

[85] Watt (n 61) 136.

[86] See e.g. David Siman-Tov, 'How Artificial Intelligence Is Transforming Israeli Intelligence Collection' (*The National Interest*, 28 April 2022) <https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/how-artificial-intelligence> accessed 5 December 2022.

[87] Harel (n 34).

[88] Matt Burgess and Lily Hay Newman, 'Israel's Failure to Stop the Hamas Attack Shows the Danger of Too Much Surveillance' *Wired* <https://www.wired.com/story/israel-hamas-war-surveillance/> accessed 31 October 2023.

[89] See Sophia Goodfriend, 'Israel's High-Tech Surveillance Was Never Going to Bring Peace' (*Foreign Policy*, 30 October 2023) <https://foreignpolicy.com/2023/10/30/israel-palestine-gaza-hamas-war-idf-high-tech-surveillance/> accessed 31 October 2023.

women 'spotters' picking up those algorithmic outputs were not taken seriously.[90]

The algorithmic creation of potential targets on the basis of the constant mass collection of data through surveillance, either by way of identifying connections with known militants or through anomaly detection, is consistently framed as the necessary first step of distinction and precaution in targeting. The data practices of target detection, target identification, and target verification thus become inextricably linked. Yet, as pointed out by Shiri Krebs, what the foregoing makes clear is that rather than just describing the legal reality by strictly applying the core rules of targeting to the dataset, the algorithms in fact actively produce this reality to begin with.[91] In this way, the increasing deployment of machine learning algorithms serves to further Israel's narrative of the IDF as the 'most moral army in the world'[92] through the recourse to IHL, while it dictates pervasive surveillance practices that in turn produce more and more potentially 'lawful targets' that inevitably emerge from the masses of collected data.[93]

II. Applying the Privacy Lens to Military Data Practices

The increasing deployment of machine learning algorithms in military applications has prompted a flurry of multi-disciplinary academic

[90] Maya Lecker, 'On October 7, Sexism in Israel's Military Turned Lethal' *Haaretz* (Tel Aviv, 20 November 2023) <https://www.haaretz.com/israel-news/haaretz-today/2023-11-20/ty-article/.highlight/on-october-7-sexism-in-israels-military-turned-lethal/0000018b-ee5b-ddc3-afdb-fe5b25be0000> accessed 1 February 2024; Alice Cuddy, 'They Were Israel's "Eyes on the Border" – But Their Hamas Warnings Went Unheard' *BBC* (London, 15 January 2024) <https://www.bbc.com/news/world-middle-east-67958260> accessed 8 October 2024.

[91] Shiri Krebs, 'Drone-Cinema, Data Practices, and the Narrative of IHL' (2022) 82 Zeitschrift für ausländisches öffentliches Recht und Völkerrecht / Heidelberg Journal of International Law 309, 331.

[92] See James Eastwood, *Ethics as a Weapon of War: Militarism and Morality in Israel* (Cambridge University Press 2017).

[93] See also Gordon, Mignot-Mahdavi and Meerssche (n 81) 135.

engagement trying to grapple with the ramifications of this development. To date, the majority of scholars has been focused on the implications for the life and physical integrity of civilians present in theatres of armed conflict, attempting to solve the question of adherence to IHL targeting rules through elaborations on the concept of 'meaningful human control', both from a legal and an ethical perspective,[94] as well as questions of responsibility and accountability for the employment of such systems.[95] Only a few have turned their attention toward the large-scale data practices that sustain the

[94] See only Berenice Boutin and Taylor Woodcock, 'Aspects of Realizing (Meaningful) Human Control: A Legal Perspective' in Robin Geiß and Henning Lahmann (eds), *Research Handbook on Warfare and Artificial Intelligence* (Edward Elgar 2024) 179; Tsvetelina van Benthem, 'Responsible Deployments of Militarised AI – The Power of Information to Prevent Unintended Engagements' (*Opinio Juris*, 2 April 2024) <https://opiniojuris.org/2024/04/02/symposium-on-military-ai-and-the-law-of-armed-conflict-responsible-deployments-of-militarised-ai-the-power-of-information-to-prevent-unintended-engagements/> accessed 4 July 2024; Georgia Hinds, 'A (Pre)Cautionary Note About Artificial Intelligence in Military Decision Making' (*Opinio Juris*, 4 April 2024) <https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-a-precautionary-note-about-artificial-intelligence-in-military-decision-making/> accessed 4 July 2024; Ingvild Bode and Anna Nadibaidze, 'Human-Machine Interaction in the Military Domain and the Responsible AI Framework' (*Opinio Juris*, 4 April 2024) <https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-human-machine-interaction-in-the-military-domain-and-the-responsible-ai-framework/> accessed 4 July 2024; Gary P Corn, 'De-Anthropomorphizing Artificial Intelligence – Grounding Notions of Accountability in Reality' (*Opinio Juris*, 5 April 2024) <https://opiniojuris.org/2024/04/05/symposium-on-military-ai-and-the-law-of-armed-conflict-de-anthropomorphizing-artificial-intelligence-grounding-notions-of-accountability-in-reality/> accessed 4 July 2024; Marta Bo and Jessica Dorsey, 'The "Need" for Speed – The Cost of Unregulated AI Decision-Support Systems to Civilians' (*Opinio Juris*, 4 April 2024) <https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-the-need-for-speed-the-cost-of-unregulated-ai-decision-support-systems-to-civilians/> accessed 4 July 2024.

[95] See only Bérénice Boutin, 'State Responsibility in Relation to Military Applications of Artificial Intelligence' (2023) 36 Leiden Journal of International Law 133; Rebecca Crootof, 'Front- and Back-End Accountability for Military AI' (*Opinio Juris*, 2 April 2024) <https://opiniojuris.org/2024/04/02/symposium-on-military-ai-and-the-law-of-armed-conflict-front-and-back-end-accountability-for-military-ai/> accessed 4 July 2024.

algorithms employed for the military DSS themselves, as described in detail in the previous section. Those scholars have attempted to address the issue of entrenched surveillance to train and deploy machine learning systems by invoking principles from privacy and data protection frameworks, correctly pointing out that such questions remain insufficiently considered in the existing rules of IHL.[96] As these examinations are relevant for the larger issues explored in this article, this section briefly reproduces three salient scholarly interventions deploying this line of argumentation before concluding that these accounts fail to sufficiently capture the larger societal implications of the military data practices under study.

Departing from the premise that the drafters of the IHL frameworks were not in a position to anticipate the role that the collecting and processing of (personal) data would come to play in military operations, some authors have recently sought to find sites within the existing rules to anchor obligations to respect privacy and data protection principles. From this corpus of norms, the principle of constant care has emerged as the most probable candidate to provide the desired legal safeguards. Arguing that Article 57(1) AP I should be understood as governing all surveillance and other data collection activities carried out to support military operations, even if performed outside of the temporal and spatial limits of armed conflict,[97] Lubin identifies the rule as reflecting 'a primeval and elementary data protection rule',[98] in fact 'truly a data protection regime in disguise'.[99] His approach is largely pragmatic. Given that the body of IHL does not contain any specific rules

[96] Watt (n 61) 159. Note in this context Rohan Talbot, 'Automating Occupation: International Humanitarian and Human Rights Law Implications of the Deployment of Facial Recognition Technologies in the Occupied Palestinian Territory' (2020) 102 International Review of the Red Cross 823, who analysed Israel's use of facial recognition technologies in the Palestinian territories under the law of belligerent occupation and applicable human rights instruments, concluding that the practice constitutes a violation of Palestinians' right to privacy.

[97] Lubin (n 33) 10.

[98] Ibid 8.

[99] Ibid 13.

to protect privacy, yet technological progress clearly calls for one, we only have Article 57 AP I as a reasonable normative lead to this effect.[100]

Further advocating for such a progressive interpretation of the duty of constant care, Eliza Watt claims that the concept of 'constant care' itself accounts for more than simply avoidance of physical harm to protected persons or objects. Instead, it extends to the protection of the rights of civilians against arbitrary interference during military operations generally, including their rights to privacy and data protection.[101] In practice, this amounts to an obligation for military commanders to observe 'fairness' by always weighing the need to gather intelligence for target verification against the obligation to respect the privacy of the civilians present in the theatre of conflict 'by imposing geographical and temporal limits on the surveillance and the amount of collected data'.[102] The author derives the legal considerations that should guide such balancing directly from data protection frameworks in civilian contexts, arguing for an application of the principles of legality, fairness, and transparency to data collection and processing practices.[103] As suggested by Gianclaudio Malgieri, compliance with the principle of fairness specifically involves not just observance of procedural safeguards but a substantial balancing of interests between the data controller and the data subject with the aim of mitigating unfair imbalances that lead to situations of 'vulnerability'.[104] To this effect, Watt understands fairness as dictating that personal data ought to be 'relevant' and 'not excessive' in relation to the purpose for which it is processed.[105]

Finally, focusing on the more specific obligation to take precautions in attack pursuant to Article 57(2)(a)(i) AP I in the context of facial recognition

[100] Ibid 16.

[101] Watt (n 61) 175–7.

[102] Ibid 176.

[103] See e.g. Article 5(1)(a) GDPR.

[104] Gianclaudio Malgieri, 'The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation' Proceedings of FAT* '20 (2020).

[105] Watt (n 61) 177–8.

technologies, Leah West seeks to develop practical guidance for military commanders to incorporate measures and processes that integrate privacy concerns into their operational protocols when using such equipment for target identification and verification.[106] She invokes the standard commentary to Additional Protocol I to support her claim that the rule's 'everything feasible' standard does indeed not imply that a commander must make use of advanced technology 'in all cases' but instead observe 'common sense and good faith' in doing so.[107] This effectively implies that an algorithmic system should be deployed only to the extent that it in fact assists in clarifying existing uncertainty as to the potential target's legal status while considering 'any potential risks associated with its deployment', including any privacy implications for present civilians.[108] Consequently, whenever the analysis suggests that less intrusive means suffice to verify the target in a way that satisfies the requirements of the precautions in attack obligation, it follows from the principles of necessity and proportionality in respect of the right to privacy that these means must be used. According to West, this will particularly apply to periods of less intense conflict when military commanders are under decreased pressure and time constraints.[109]

It makes sense to scrutinise existing rules in the law of armed conflict to uncover at least some preliminary legal instruments for limiting the unconstrained data practices that militaries and intelligence agencies currently engage in. However, ultimately the existing rules on targeting cannot provide a satisfying solution. For one, from a doctrinal perspective, rooting data protection obligations in the principle of constant care stands on shaky ground. Even if we accept the more expansive interpretation of the rule's protective scope, the problem remains that for whatever else Article 57 AP I might be invoked, its primary purpose remains to support and bolster

[106] West (n 68), 137–8; in civilian uses of AI, such constructions are usually discussed under the concept of "privacy by design".

[107] Ibid 141–2; referring to Sandoz, Swinarski and Zimmermann (n 63) 680.

[108] West (n 68) 142.

[109] Ibid 150–1.

the foundational principle of distinction so that civilian casualties and damage to civilian objects be reduced to a minimum. Any further values that may reasonably be read into the provision's scope of protection, such as privacy, must come second in the case of a conflict with the overarching aim of protecting the physical integrity of civilians. The rule may indeed be open to encompass values other than life and limb, yet it is not obvious how 'good faith' considerations lead to an outcome that de-prioritises the avoidance of physical harm. If the purpose of the processing of data is the disposal of uncertainties through target verification, it is unclear how the data practices necessary to achieve that could ever fail to meet the 'fairness' requirement by being 'excessive' or 'irrelevant'. It is the principles of machine learning that seem to call for surveillance activities that cannot simply be switched on and off at will – for AI-based DSS to work reliably at all, their models must be trained on context-specific, timely, and by default large datasets. According to this rationale, the alternative would be the deployment of poorly adjusted systems that risk ill-considered targeting decisions and consequently rising civilian casualties, the very outcome the regime of Article 57 AP I was created to prevent. If that is the case, however, it is doubtful whether privacy considerations dictating a reduction of data collection practices within the framework of existing IHL could ever prevail.

To be sure, with a view to the Palestinian situation it must be conceded that human rights frameworks, with their unambiguous inclusion of the right to privacy,[110] have an important role to play due to their general applicability in situations of a state's effective control over territory, which is at least the case with regard to illegally annexed East Jerusalem and the prolonged belligerent occupation of the West Bank, which arguably also continues in Gaza.[111] Nevertheless, most of Israel's surveillance practices are carried out with a more or less direct nexus to armed engagements with militant

[110] See only Article 17(1) ICCPR.

[111] Talbot (n 96); Watt (n 61) 170; see on this question *Legal Consequences Arising from the Policies and Practices of Israel in the Occupied Palestinian Territory, Including East Jerusalem*, Advisory Opinion of 19 July 2024, at paras. 86–94.

resistance in the territories, which are primarily governed by the principles of the law of armed conflict.

Either way, attempts to tackle excessive data practices and surveillance for the purpose of algorithmic warfare by way of applying principles of privacy and data protection ultimately fall short of accounting for the deeper harms such practices bring about. Even if one correctly understands privacy as the fundamental right underpinning political freedoms such as free expression, assembly, and association, and thus recognises its realisation as the condition of possibility of these freedoms' actualisation,[112] privacy as the principal lens through which to appraise algorithmic warfare fails to capture the essence of the relationship between these data practices and the subjects' political agency. To substantiate this critique, the following section explores this relationship in more detail.

III. Machine Rationalities and Political Action

Whereas the previous section interrogated attempts to capture the wider harms caused by large-scale data practices by militaries for the purpose of training AI systems, this section turns toward implications for collective political rights. The de-politicising effects of both algorithmic security[113] and of drone warfare have already been the subject of scholarly scrutiny.[114] To further deepen these lines of inquiry, the following deliberations reappraise the consequences of algorithmic rationalities in the security realm, in the specific case of Palestine but also more generally, through an

[112] Talbot (n 96) 845.

[113] Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Duke University Press 2020).

[114] See only Alex Edney-Browne, 'The Psychosocial Effects of Drone Violence: Social Isolation, Self-Objectification, and Depoliticization' (2019) 40 Political Psychology 1341; International Human Rights and Conflict Resolution Clinic, Stanford Law School and Global Justice Clinic, NYU School of Law, 'Living Under Drones: Death, Injury, and Trauma to Civilians from US Drone Practices in Pakistan' (2012) <https://chrgj.org/wp-content/uploads/2016/09/Living-Under-Drones.pdf>.

application of Rosa Luxemburg's and Hannah Arendt's concepts of spontaneity. To do so, I first anchor the capability to spontaneous political action in the collective right to self-determination, which is clarified and differentiated in its procedural alongside its substantive normative dimension. Before finally explaining how systems of algorithmic warfare prevent the collective formation of political will in the exercise of self-determination in Section 4, it is then first necessary to inquire the role of spontaneity for political agency in the thinking of both Luxemburg and Arendt.

1. *The Principle of Self-Determination as a Right to Collective Political Action*

That the Palestinian people are the legitimate bearer of the right to self-determination within the Palestinian territories is not in dispute.[115] What is less straightforward is the precise content of such a right. Traditional international legal doctrine has focused on material outcomes, which can partly be explained by looking at the right's historical position within the nexus of non-self-governing territories and post-World War II processes of decolonialisation under the auspices of the United Nations.

According to this framing, self-determination is achieved once a certain legal status has been realised, be it autonomy within the structures of an existing state as an expression of 'internal' self-determination, on the one hand, or independence – through the termination of a colonial relationship to a metropolitan state or secession from a larger state – as the quintessential form of 'external' self-determination, on the other. Under existing international law, the precise manifestation of the right that the self-determination unit is entitled to depends on the specifics of the situation. In the context of

[115] See only *Legal Consequences Arising from the Policies and Practices of Israel in the Occupied Palestinian Territory, Including East Jerusalem*, Advisory Opinion of 19 July 2024, at para. 230; *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory*, Advisory Opinion of 9 July 2004, ICJ Rep 2004, 136, at para. 118.

decolonisation and in other situations of foreign occupation,[116] the people in question have an enforceable right to form their own state. Whether there is a right to 'external' self-determination in the form of secession outside of this context remains contentious and is in any case not settled law.[117] Within this approach, in regard to non-self-governing territories, such a result was mostly for an outside entity to bring about. Accordingly, Article 73(3) UN Charter obliged colonial powers to seek to 'develop self-government' and 'to assist [the people in non-self-governing territories] in the progressive development of their free political institutions', while Article 76(b) UN Charter urged administering authorities within the trusteeship system to

> promote the political, economic, social, and educational advancement of the inhabitants of the trust territories, and their progressive development towards self-government or independence as may be appropriate to the particular circumstances of each territory and its peoples.

This framing had its precursor in Article 22 of the Covenant of the League of Nations,[118] which even more starkly put responsibility on the 'advanced nations' to promote the 'well-being and development' of 'peoples not yet able to stand by themselves under the strenuous conditions of the modern world'. As late as 2004, the International Court of Justice (ICJ) approvingly cited this provision in its *Wall* Advisory Opinion as implying that the 'ultimate objective' of the trusteeship system was the self-determination of the peoples concerned.[119]

The highly paternalistic notion of self-determination as expressed in these rules prompted some states in the Third Committee of the UN General Assembly to speak out against the inclusion of a provision on self-determination in the two principal UN human rights instruments, the

---

[116] *Policies and Practices of Israel in the Occupied Palestinian Territory* (n 115), para. 233.

[117] Daniel Thürer and Thomas Burri, 'Secession', in Rüdiger Wolfrum and Anne Peters (eds.) *Max Planck Encyclopedia of Public International Law* (Oxford University Press 2009).

[118] Covenant of the League of Nations, (adopted 28 June 1919) 108 LNTS 188.

[119] Ibid, at para. 88, with reference to the previous decisions *South West Africa*, *Western Sahara*, and *East Timor*.

International Covenant on Civil and Political Rights (ICCPR)[120] and the International Covenant on Economic, Social and Cultural Rights (ICESCR).[121] The argument was that while Articles 1 and 55 UN Charter clarified that the self-determination of peoples formed the basis of friendly relations among states, the *granting* of independence and self-government 'could only be achieved progressively and in line with the development of the peoples of these Territories and their readiness to govern themselves'.[122] And despite having found its positive manifestation as a (collective) human right in Article 1 common to the ICCPR and ICESCR, its third paragraph is still read as directing all states to 'take positive action to facilitate realization of and respect for the right of peoples to self-determination'.[123]

In contrast to this patronising account of self-determination, which ultimately implies that 'peoples do not actually possess a veritable right to self-determination' but are merely 'beneficiaries' of the right conferred by the two Covenants to the state parties,[124] stands an understanding that takes seriously the principle as reflecting and actualising 'the wishes of the people concerned'.[125] Among international legal instruments, this is expressed most succinctly in the African Charter on Human and Peoples' Rights, whose Article 20(2) unambiguously sets out that '[c]olonized or oppressed peoples shall have the right to *free themselves* from the bonds of domination *by resorting to any means* recognized by the international community'.[126]

---

[120] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171.

[121] International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 1 March 1976) 993 UNTS 3.

[122] UN Doc. A/3077 (1955), at para. 30.

[123] Office of the High Commissioner for Human Rights, CCPR General Comment No. 12: Article 1 (Right to Self-determination) – The Right to Self-determination of Peoples, 13 March 1984, at para. 6.

[124] See, not supporting this view, Antonio Cassese, *Self-Determination of Peoples: A Legal Reappraisal* (Cambridge University Press 1995) 143.

[125] Ibid 242.

[126] Emphases added.

This understanding refers back to the concept's intellectual roots in Enlightenment thought; in this tradition, it was originally devised as meaning principally that 'the form of government in a state should be determined by the collective will of the people who are subject to it'.[127] Such notion, in turn, necessarily implies that the very process of forming authority and political will that enables the people to express their choice freely forms an integral part of the right itself.[128] This has been – in the very different context of external interference in elections – noted by Jens David Ohlin, who contends that true self-determination cannot be sustained without protecting the 'deliberations of the public' that precede electoral processes as the periodic actualisation of the right in democratic societies.[129] In other words, rather than merely stipulating a claim to a material-legal outcome in terms of political status, the right to self-determination would be incomplete, and indeed contradictory, without a corresponding procedural component that provides the right to form the political will that is a precondition for achieving the desired outcome in the first place.

Applying this reading to the situation in Palestine, it further bears mentioning that according to a correct interpretation of the law of occupation as a transitory and exceptional regime, the collective right to self-determination of the population of an occupied territory is implied within the ambit of Article 43 of the 1899 Hague Regulations.[130] The provision bestows on the occupying power the authority to 're-establish and insure

---

[127] Tom Sparks, *Self-Determination in the International Legal System: Whose Claim, to What Right?* (Hart Publishing 2023) 20–1.

[128] See Nicholas Tsagourias, 'Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace' (17 August 2019) 14 <https://papers.ssrn.com/abstract=3438567> accessed 8 July 2023.

[129] Jens David Ohlin, *Election Interference: International Law and the Future of Democracy* (Cambridge University Press 2020) 100–2 <https://www.cambridge.org/core/books/election-interference/62027877A63505C5B6D93F485C5208B5> accessed 12 June 2023.

[130] Eliav Lieblich and Eyal Benvenisti, *Occupation in International Law* (Oxford University Press 2022) 85.

public order and safety' while maintaining respect for sovereignty that remains vested in the occupied people.[131] It thus follows from the above that the law of occupation itself imposes on Israel, expressed in Hohfeldian terms,[132] a correlative duty to tolerate or even facilitate political will-formation among the Palestinian people as the holders of the right[133] (unless, that is, it directly interferes with public order and safety in the territory under occupation).[134]

Again, while there is no denying that the Palestinian people have an enforceable claim to a concrete material-legal outcome – political, 'external' self-determination in the guise of their own, fully formed state[135] – that claim must encompass the right to realise a set of procedures that together constitute the conditions of possibility of achieving such outcome through political action for the right itself to be at all meaningful. This reading acknowledges what should be self-evident not least with recourse to the concept's historical roots as a 'polity-based' claim, to borrow Sparks' terminology:[136] that self-determination is not something to be realised primarily through outside forces and processes but by the people as the claim's bearers themselves. More precisely, self-determination is primarily a procedural right, or it is nothing; as a procedural right, it is directed at enabling collective political agency and will-formation. As such, it is neither congruent with nor exhausted by the individual political rights of freedom of information, expression, assembly, association, and the rights to vote and

---

[131] Orna Ben-Naftali, 'Belligerent Occupation: A Plea for the Establishment of an International Supervisory Mechanism' in The Late Antonio Cassese (ed), *Realizing Utopia: The Future of International Law* (Oxford University Press 2012) 543.

[132] See Wesley Newcomb Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 The Yale Law Journal 16.

[133] Cassese (n 124) 143.

[134] As the ICJ observed in its *Wall* advisory opinion (n 115), although Israel is not party to the Fourth Hague Convention of 1907, to which the Hague Regulations are annexed, the provisions are reflective of customary international law, see at para. 89.

[135] Wilde (n 14).

[136] Sparks (n 127) 19.

to be elected, although it is in the rights' nature that infringement will frequently occur concurrently. As opposed to these individual rights, the principle of self-determination captures and protects the distinctive and critical *collective* dimension of political struggle. This understanding, finally, raises the question of what conditions must exist for a people to be able to actualise that political will-formation, which the next section investigates further.

2. *Spontaneity and Collective Political Agency*

> All the above great and partial mass strikes and general strikes (...) originated for the most part spontaneously, in every case from specific local accidental causes, without plan and undesignedly, and grew with elemental power into great movements (...).[137]

If the exercise of self-determination is contingent on a collective practice to form a directed political will, then what conditions must be present for the latter to become possible? One answer, as will emerge from the following, lies in a nuanced understanding of the concept of spontaneity as developed in the writings of Rosa Luxemburg and Hannah Arendt.

In her analysis of the struggles of workers' movements at the beginning of the twentieth century, Rosa Luxemburg put great and persistent emphasis on the significance of spontaneous action to bring about true political change.[138] As she wrote in her famous 1906 essay *The Mass Strike, the Political Party, and the Trade Unions*, which assessed the course of the Russian

[137] Rosa Luxemburg, 'The Mass Strike, the Political Party, and the Trade Unions' in Helen Scott (ed), *The Essential Rosa Luxemburg* (Haymarket 2008) 142.

[138] In this context, it bears noting at the outset that Luxemburg was famously opposed to the idea of 'national self-determination' as she conceived it as an obstacle to the universal cause of the working class, which could be achieved not within the boundaries of a state but only in an international movement, see Rosa Luxemburg, 'The National Question' in Le Blanc and Helen Scott (eds), *Socialism or Barbarism: Selected Writings by Rosa Luxemburg* (Pluto Press 2010). As will become clear, however, this does not prevent us from fruitfully using her insights on the role of spontaneity for political agency.

Revolution that had begun in January of the previous year and the contrast between (organised) political strikes and (spontaneous) mass strike as the principal instruments of revolutionary struggle, 'in the mass strikes in Russia the element of spontaneity plays such a predominant part not because the Russian proletariat is "uneducated", but because revolutions do not allow anyone to play the schoolmaster with them'.[139] Luxemburg's insistence on the substantial importance of the spontaneity of the masses has traditionally been interpreted as her implying that it constituted the pivotal factor for the eventual success of revolution at the expense of considered direction and leadership as embodied by the social democratic party and the labour organisations.[140] In this, she found fierce opposition not only among the socialist and communist leaders in Germany and elsewhere,[141] but also from theorists such as Antonio Gramsci who, while not dismissing the utility of spontaneity entirely, argued that it needed to be combined with 'conscious leadership' and 'discipline' to become 'the real political action of the subaltern classes, insofar as it is mass politics and not a mere adventure by groups that appeal to the masses'.[142] Without such coordination, he claimed, the political struggle would remain ineffective and even regressive.[143]

Several writers, however, have since noted that this is not the only, and indeed not the most persuasive, way to conceive Luxemburg's understanding of spontaneity. What she had in mind instead was the 'capacity for producing change' that spontaneous political action

[139] Luxemburg, 'The Mass Strike, the Political Party, and the Trade Unions' (n 137) 148.

[140] See Ottokar Luban, 'Rosa Luxemburg's Concept of Spontaneity and Creativity in Proletarian Mass Movements – Theory and Practice' (2019) 9 International Critical Thought 511, 512.

[141] See Sidonia Blättler and Irene M Marti, 'Rosa Luxemburg and Hannah Arendt: Against the Destruction of Political Spheres of Freedom' (2005) 20 Hypatia 88, 90–2.

[142] Antonio Gramsci, *Prison Notebooks, Vol. II, Notebook 3* (Joseph A Buttigieg ed, 1996) §48.

[143] Marcus E Green, 'Gramsci and Subaltern Struggles Today: Spontaneity, Political Organization, and Occupy Wall Street' in Mark McNally (ed), *Antonio Gramsci* (Palgrave 2015) 156.

generates.[144] Rather than focusing on concrete outcomes, Luxemburg emphasised the 'creative spirit'[145] of such activity that first makes visible[146] and produces critical consciousness of the people's objective conditions,[147] as 'the stormy gesture of the political struggle causes [them] to feel with unexpected intensity the weight and the pressure of [their] economic struggle'.[148] Consequently, spontaneous action intensifies 'the inner urge of the workers to better their position, and their desire to struggle',[149] and thus acts as a catalyst that engenders the collective conditions that must exist to *initiate* a transformative politics. Spontaneous mass action is thus not about tangible practical 'success', but the experience and knowledge gained about the political situation and the next steps in the sense of a 'self-enlightenment' of the people without which a struggle moving toward emancipation remains impossible.[150] As put by Paulina Tambakaki, with its inherent connection to *initiative*, spontaneity makes an 'opening to change' by creating and honing a 'capacity for resistance'.[151]

Expanding upon Luxemburg's considerations, whose work she admired and had studied intensively, Hannah Arendt further advanced our understanding of the pivotal role that spontaneity plays in political affairs.[152] For Arendt,

---

[144] Paulina Tambakaki, 'Why Spontaneity Matters: Rosa Luxemburg and Democracies of Grief' (2021) 47 Philosophy & Social Criticism 83, 83–4.

[145] Luban (n 140) 515.

[146] Tambakaki (n 144) 92.

[147] Alex Levant, 'Rethinking Spontaneity Beyond Classical Marxism: Re-Reading Luxemburg through Benjamin, Gramsci and Thompson' (2012) 40 Critique 367, 371–2.

[148] Luxemburg, 'The Mass Strike, the Political Party, and the Trade Unions' (n 137) 146.

[149] Ibid 144.

[150] Blättler and Marti (n 141) 91.

[151] Tambakaki (n 144) 98–9.

[152] Maria Tamboukou, 'Imagining and Living the Revolution: An Arendtian Reading of Rosa Luxemburg's Letters and Writings' (2014) 106 Feminist Review 27, 32; Blättler and Marti (n 132) 90. See also Arendt's review of J.P. Nettl's biography of Luxemburg, Hannah Arendt, 'A Heroine of Revolution' [1966] *New York Review of Books* <https://www.nybooks.com/articles/1966/10/06/a-heroine-of-revolution/> accessed 6 August 2023.

political freedom as such can only be actualised through *action*, the highest form of human activity within the hierarchy of the *vita activa*, which she distinguishes from the two lower activities *labour* and *work*. Whereas labour only serves the purpose of sustaining a person's biological functions through eating, drinking, and other such activities,[153] the notion of work describes the fabrication of objects, which above all involves imposing a preconceived model upon the world and using the physical world as material.[154]

Action, in contrast, 'is not forced upon us by necessity, like labor, and it is not prompted by utility, like work'.[155] It is the only 'truly political'[156] of the human activities and implies both *initiating* something new that interrupts the course of events and *interaction* as it occurs in the public sphere of politics.[157] When people act 'in concert', they engender power;[158] in Jürgen Habermas's reading of Arendt, 'the fundamental phenomenon of power is (...) the formation of a *common* will in a communication directed to reaching agreement'.[159] Such communicative power of the people, however, can only be sustained for the 'fleeting moment of action',[160] vanishing 'the moment [the people] disperse'.[161] With its capacity to initiate the unexpected and

[153] Hannah Arendt, *The Human Condition (HC)* (2nd edition, The University of Chicago Press 1958) 79 ff.; see on this further Paul Voice, 'Labour, Work and Action' in Patrick Hayden (ed), *Hannah Arendt: Key Concepts* (Routledge 2014) 36.

[154] Arendt, HC (n 153) 140; see on Arendt's conception of "work" further Pritika Nehra, 'Judging Work: The Concept of "Work" in Hannah Arendt's "The Human Condition"' in Dominika Polkowska (ed), *The Value of Work in Contemporary Society* (Brill 2014).

[155] Arendt, HC (n 153) 177.

[156] Marieke Borren, 'Plural Agency, Political Power, and Spontaneity' in Christopher Erhard and Tobias Keiling (eds), *The Routledge Handbook of Phenomenology of Agency* (Routledge 2020) 164.

[157] Ibid 165.

[158] Hannah Arendt, *On Violence (OV)* (Harcourt Brace Jovanovich 1970) 44.

[159] Jürgen Habermas, 'Hannah Arendt's Communication Concept of Power' (1977) 44 Social Research 3, 4.

[160] Arendt, HC (n 153) 201.

[161] Ibid 200.

incalculable,[162] action is consequently also the only human activity that is fully defined by spontaneity. Spontaneity, for Arendt, is 'a man's power to begin something new out of his resources, something that cannot be explained on the basis of reactions to environment and events'.[163] Through its spontaneous character, action is intrinsically creative, contingent, unpredictable, and 'boundless' – as opposed to work, which is always directed at producing a certain material outcome – not least because action takes place in 'an already existing web of human relationships, with its innumerable, conflicting wills and intentions'.[164] While the human capacity to spontaneous action itself is conceived as pre-political, Arendt insisted that 'all political freedom would forfeit its best and deepest meaning without this freedom of spontaneity'.[165] In other words, political freedom deprived of spontaneity is effectively meaningless.[166]

Both Luxemburg and Arendt understood the significance of spontaneity for a truly emancipatory politics through collective action that fosters the creative potential and generates the political will that is necessary to take the initiative.[167] It is only through spontaneous activity that individuals can relate themselves to the world[168] and consequently, as a collective, bring about political change.[169] If the initiative to such political action prevails and sparks a catalysing event, the people can be said to exercise a genuinely self-determined politics even if the action fails to succeed, as insinuated in Arendt's emphatic *Reflections on the Hungarian Revolution*:

---

[162] Hannah Arendt, *The Origins of Totalitarianism (OT)* (Penguin Classics 1951) 598.

[163] Ibid 596; in this, spontaneity is intimately related to Arendt's concept of *natality*; see Hildebrandt (n 24) 89.

[164] Arendt, HC (n 153) 183–4; 190–1.

[165] Hannah Arendt, *The Promise of Politics* (2005) 127–8.

[166] Katarzyna Eliasz, 'The Structure of the Concept of Political Freedom in Hannah Arendt's Philosophy' (2019) 10 Avant 29, 33.

[167] See Blättler and Marti (n 141) 94.

[168] Erich Fromm, *Escape from Freedom* (Holt Paperbacks 1941) 261.

[169] Suzanne Jacobitti, 'Hannah Arendt and the Will' (1988) 16 Political Theory 53, 65.

> If there was ever such a thing as Rosa Luxemburg's "spontaneous revolution" – the sudden uprising of an oppressed people for the sake of freedom and hardly anything else, without the demoralizing chaos of military defeat preceding it, without coup d'état techniques, without a closely knit apparatus of organizers and conspirators, without the undermining propaganda of a revolutionary party, something, that is, which everybody, conservatives and liberals, radicals and revolutionists, had discarded as a noble dream – then we had the privilege to witness it.[170]

As Heba Raouf Ezzat and Artemy Magun have observed, more recent upheavals such as the initial phase of the Arab Spring in Egypt and Tunisia in 2011 or the Maidan Revolution in Ukraine in 2014 may be taken as further examples of catalysing events that demonstrated the merit of Arendt's theory, demonstrating how '[s]pontaneity can create windows of political opportunities'.[171] At the same time, Arendt's observation also explains why the terrorist attacks by Hamas on 7 October 2023, contrary to some early interpretations that likened them to a 'pogrom',[172] cannot be conceived as a 'spontaneous' political uprising – unlike, arguably, the First Intifada that began in 1987.[173] The operation was launched after years of meticulous

[170] Hannah Arendt, 'Totalitarian Imperialism: Reflections on the Hungarian Revolution' (1958) 20 The Journal of Politics 5, 8.

[171] Heba Raouf Ezzat, 'Palimpsests of Civicness: Spontaneity and the Egyptian Uprising/Cairo 2011' (2022) 18 Journal of Civil Society 239, 256; Artemy Magun, 'Spontaneity and Revolution' (2017) 116 The South Atlantic Quarterly 815, 822–3. However, Magun, 828–9, claims that in such situations, spontaneity is difficult to prove and thus ultimately "in the eye of the beholder".

[172] Jonathan Dekel-Chen, 'Does the Hamas Massacre of October 7 Echo the Holocaust?' *Haaretz* (Tel Aviv, 30 January 2024) <https://www.haaretz.com/opinion/2024-01-30/ty-article-opinion/.premium/does-the-hamas-massacre-of-october-7-echo-the-holocaust/0000018d-5abf-d997-adff-dfffbc90000> accessed 31 January 2024.

[173] Ibrahim Al-Marashi, 'What the World Can Learn from the History of Hamas' (*TIME*, 17 October 2023) <https://time.com/6324221/hamas-origins-history/> accessed 31 January 2024.

planning.[174] Relatedly, to contend that the atrocities were somehow the inevitable outcome of the suppression of any other type of political expression by the citizens of Gaza is equally insufficiently nuanced an explanation, if only as it fails to account for the guiding ideology of Hamas and the other involved militant groups, an ideology that by itself – aside from calling for the destruction of Israel – has set up political structures in Gaza that systematically deny the exercise of political rights by anyone other than the organisation itself.[175]

The work of Luxemburg and Arendt reveals the capacity to spontaneous initiative as the condition of possibility to enact an emancipatory politics, which is intrinsically linked to the collective exercise of the right to self-determination. In Luxemburg's words, for a people to form the political will to determine its own political future, it must be able to creatively shape 'the forms that will carry the revolutionary movements to a successful outcome'[176] without preconceived external direction, in a voluntary, impromptu, and not priorly predictable manner. Spontaneity is, as the essential expression of political freedom,[177] diametrically opposed to, as Erich Fromm put it, the 'activity of the automaton, which is the uncritical adoption of patterns suggested from the outside'.[178] The next section investigates what happens when the postulates of the 'automaton' are imposed on spontaneous political action by machine learning algorithms.

---

[174] Sophia Goodfriend, 'Israel's High-Tech Surveillance Was Never Going to Bring Peace' (*Foreign Policy*, 30 October 2023) <https://foreignpolicy.com/2023/10/30/israel-palestine-gaza-hamas-war-idf-high-tech-surveillance/> accessed 31 October 2023.

[175] See: Shaul Bartal, 'Ḥamās: The Islamic Resistance Movement' in Muhammad Afzal Upal and Carole M Cusack (eds), *Handbook of Islamic Sects and Movements* (Brill 2021) <https://www.jstor.org.ezproxy.leidenuniv.nl:2048/stable/10.1163/j.ctv1v7zbv8.23> accessed 1 February 2024.

[176] Rosa Luxemburg, 'The Junius Pamphlet: The Crisis in German Social Democracy' in Peter Hudis and Kevin B Anderson (eds), *The Rosa Luxemburg Reader* (Monthly Review Press 2004) 329.

[177] Arendt, *The Promise of Politics* (n 165) 127.

[178] Fromm (n 168) 257.

IV. FREEZING THE PAST

Having established the critical function of spontaneous action as a precondition to form the will necessary for the exercise of a self-determined emancipatory politics, the question about the nature and consequence of the relationship between such behaviour and the inner workings of machine learning algorithms under conditions of perpetual surveillance remains to be answered. The point that the increasing use of algorithmic surveillance negatively impacts the ways in which politics is performed and actualised in the public sphere has been made before.[179] The argument I want to advance here is that this effect is a direct and inevitable consequence of how spontaneity, as conceived by Luxemburg and Arendt, interacts with machine learning algorithms in the security context.

Although 'the outputs of predictive technologies are often perceived as objective, complete, and neutral',[180] and indeed increasingly as omnipotent and superior to human cognitive faculties,[181] all evidence suggests that such trust in their capabilities is misguided. For one, despite recent advances with large language models that some take as seeming to suggest otherwise, even the latest generations of machine learning algorithms continue to lack any sense of contextual understanding[182] or the faculty of common sense (abductive) reasoning.[183] Expectations of imminent breakthroughs toward

[179] See: Amoore (n 113). In respect to facial recognition technologies in Palestine see Talbot (n 96). For very instructive qualitative research on this issue see most recently Daragh Murray and others, 'The Chilling Effects of Surveillance and Human Rights: Insights from Qualitative Research in Uganda and Zimbabwe' [2023] Journal of Human Rights Practice 1.

[180] Krebs (n 22) 201.

[181] Katja Grace and others, 'Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts' (2018) 62 Journal of Artificial Intelligence Research 729.

[182] Klonowska (n 21) 18; Larson (n 23) 130, 139.

[183] Gaël Gendron and others, 'Large Language Models Are Not Abstract Reasoners' (arXiv, 31 May 2023) <http://arxiv.org/abs/2305.19555> accessed 23 July 2023; Xiang Lorraine Li and others, 'A Systematic Investigation of Commonsense Knowledge in Large Language Models'

'artificial general intelligence' (AGI)[184] or that at least take as a given that 'autonomous technology is far more likely to improve than human decision-making'[185] are frequently based on category errors,[186] whereas any actual progress is far off.[187]

Because machine learning is based on the principles of statistical analysis and inductive reasoning, the lack of contextual 'world knowledge' means that

(arXiv, 31 October 2022) <http://arxiv.org/abs/2111.00607> accessed 23 July 2023. For an earlier argument that it is possible to provide computer systems with abductive reasoning capabilities see John R Josephson and Susan G Josephson (eds), *Abductive Inference: Computation, Philosophy, Technology* (Cambridge University Press 1994).

[184] The concept of AGI is generally understood as describing software at the same level as or beyond human-like intelligence, see Reece Rogers, 'What's AGI, and Why Are AI Experts Skeptical?' (*Wired*, 20 April 2023) <https://www.wired.com/story/what-is-artificial-general-intelligence-agi-explained/>.

[185] Heller (n 56) 67. Pointing to inherent limitations of human capacities to contrast them with machine abilities, as Heller does, is of no avail insofar as time and again it has been demonstrated that reliance on "AI" does not compensate for but exacerbates human cognitive insufficiencies.

[186] See: Arjun Ramani and Zhengdong Wang, 'Why Transformative Artificial Intelligence Is Really, Really Hard to Achieve' (*The Gradient*, 26 June 2023) <https://thegradient.pub/why-transformative-artificial-intelligence-is-really-really-hard-to-achieve/> accessed 9 July 2023. Recently, Rylan Schaeffer, Brando Miranda and Sanmi Koyejo, 'Are Emergent Abilities of Large Language Models a Mirage?' (arXiv, 22 May 2023) <http://arxiv.org/abs/2304.15004> accessed 23 July 2023 have suggested that any perceived hints of emergent capabilities toward abductive reasoning in large language models "appear due to the researcher's choice of metric rather than due to fundamental changes in model behavior with scale".

[187] There is some talk of developing "third wave AI", with the current state of the art of machine learning algorithms, including large language models, by combining the rules-based (of first wave AI) and statistical approach of machine learning to create models that are capable of understanding context, see Brandi Vincent, 'How DARPA's AI Forward Program Seeks "New Directions" on the Path to Trustworthy AI' (*DefenseScoop*, 31 March 2023) <https://defensescoop.com/2023/03/31/how-darpas-ai-forward-program-seeks-new-directions-on-the-path-to-trustworthy-ai/> accessed 13 April 2023. Whether such attempts will be successful, or are even promising, remains to be seen. Either way, such technology requires a paradigm shift in how the mainstream of the scientific field approaches the idea of artificial intelligence.

these algorithms are intrinsically incapable of dealing with situations that fall outside of what is represented within the dataset fed to it during training. By definition, 'no algorithm can be trained on future data',[188] yet the 'real world generates datasets all day long, twenty-four hours a day, seven days a week, perpetually', so that 'any given dataset is only a very small time slice representing, at best, partial evidence of the behavior of real-world systems'.[189] It is for this reason that any prediction as the output of machine learning systems is necessarily based on the premise that the future will 'look like the past',[190] i.e. the corpus of data on past events that was used to build the model. Concerning areas of conflict, one salient problem with this is that such environments have a tendency to generate only limited data in the first place.[191] More importantly, and fundamentally, it follows that the algorithms proceed on the baseline assumption that human behaviour remains constant, that is consistent with whatever patterns and frequencies have been detected in the dataset.[192] But the real world is inherently surprising, and the new data generated by such unexpected events 'can always disrupt the predictive accuracy of the hypothesis target function'.[193]

It is important to note that these inherent limitations of machine learning affect both classification tasks, such as image recognition, visual object recognition, or in the guise of sequence classification used in natural language processing,[194] and anomaly detection as the inverse of frequency assumptions. As far as the latter is concerned, as explained, the frequencies in the past data are used to determine the state of 'normality' which is set against any unexpected events subsequently picked up by the system, which will accordingly mark them as suspicious. In this context, critical observers have

[188] Hildebrandt (n 24) 92.

[189] Larson (n 23) 139.

[190] Ibid 119.

[191] Goldfarb and Lindsay (n 73) 26.

[192] Klonowska (n 21) 21.

[193] Hildebrandt (n 24) 99.

[194] Larson (n 23) 135.

pointed to the so-called base-rate fallacy, which means that 'if you are looking for very rare instances or phenomena in a very large dataset, you will inevitably obtain a very high percentage of false positives in particular – and this cannot be remedied by adding more or somehow 'better' data: by adding hay to a haystack'.[195]

How this plays out can be explained by the example of SKYNET. In 2007, the U.S. National Security Agency deployed its machine learning model to uncover terrorist suspects in Pakistan. Parsing metadata from 55 million domestic mobile phone users, the algorithm tried to detect usage patterns matching that of a few known individuals working as couriers for al-Qaeda, which reportedly generated a false positive rate of merely 0.008 percent – yet while that figure may seem very low, it still implies that approximately 15,000 people were wrongly marked as potential terrorist couriers by the model.[196] Claudia Aradau and Tobias Blanke have noted that this is hardly accidental, as public surveillance algorithms are set to tolerate high false positive rates in order to detect or identify persons of interest.[197] Consequently, even small, completely innocuous 'anomalies' of human behaviour, incidental correlations between data points that do not match existing patterns, will be registered and flagged as suspicious. This is another way of saying that these models do not simply *discover* potential targets; they *produce* them. And it is in the nature of the 'opaque epistemologies'[198] such models engender that it will be impossible to comprehend or reproduce the reason for the algorithm to have arrived at a certain predictive output.[199]

---

[195] Douwe Korff, 'The Limitations of and Flaws in Algorithmic/AI-Based Technologies' (3 May 2023) 2 <https://papers.ssrn.com/abstract=4437110> accessed 29 June 2023.

[196] See: Kathleen McKendrick, 'Artificial Intelligence Prediction and Counterterrorism' (Chatham House 2019) 11. Notoriously, SKYNET marked the Al Jazeera reporter Ahmad Muaffaq Zaidan as a potential al-Qaeda courier, see Klonowska (n 21) 21.

[197] Aradau and Blanke (n 79) 170.

[198] Krebs (n 22) 220.

[199] Hildebrandt (n 24) 100. See on this also *Ligue des Droits Humains* (n 78) paras. 194–5.

At this point, it bears emphasising that nothing we have come to know to date suggests that the described issues could be mitigated through the concept of 'meaningful human control'.[200] For one, the problem of over-reliance has been demonstrated over and over again.[201] One of the psychological phenomena inevitably at play is interpretation bias, describing the situation in which a human operator misunderstands the implications of the model's prediction.[202] Closely related is the problem of selective adherence, a type of confirmation bias, meaning 'the strong tendency of decision-makers to selectively seek and interpret information in light of pre-existing stereotypes, beliefs, and social identities', with the consequence that they 'assign greater weight to information congruent with prior beliefs and contest inputs that contradict them'.[203] In the case of Israel and Palestine, add to this reports that the IDF puts great pressure on its personnel to constantly produce new targets and add them to the database, further incentivising

[200] See for a good overview of the concept Boutin and Woodcock (n 94).

[201] See for the latest large language models e.g. OpenAI, 'GPT-4 Technical Report' (arXiv, 27 March 2023) 59–60 <http://arxiv.org/abs/2303.08774> accessed 21 July 2023.

[202] UNIDIR, 'Algorithmic Bias and the Weaponization of Increasingly Auonomous Technologies: A Primer' (United Nations Institute for Disarmament Research 2018) 5 <https://unidir.org/sites/default/files/publication/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf> accessed 21 July 2023.

[203] Saar Alon-Barkat and Madalina Busuioc, 'Human-AI Interactions in Public Sector Decision-Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice' (2023) 33 Journal of Public Administration Research and Theory 153. How such selective adherence plays out in reality could be witnessed in the case of the 2021 drone strike in Kabul in which the U.S. killed 10 civilians, see Matthieu Aikins and others, 'Times Investigation: In U.S. Drone Strike, Evidence Suggests No ISIS Bomb' *The New York Times* (10 September 2021). Available information strongly suggests that the incident started with an algorithm analysing visual surveillance data collected from satellites and UAVs and detecting an anomaly that flagged the Afghan aid worker as a potential ISIS terrorist.

intelligence analysts to de-emphasise whatever safeguards exist that could amount to *meaningful* human control.[204]

If machine learning algorithms function on the basic expectation that the future will look like the past, and that whatever does not fit this backward-looking pattern is raising suspicion, then it becomes manifest how such processes interrelate with Luxemburg's and Arendt's understanding of emancipatory political action as intrinsically linked to spontaneity. With its 'transformative potential'[205] that Luxemburg so strongly advocated for, it lies in the very nature of spontaneous political action that it generates rifts in the dominant fabric; more to the point, that it creates anomalies. It is always directed at *initiating* something new by disrupting the predetermined course of events.[206] For spontaneity, as Arendt has shown, is 'the human capacity to begin, to initiate something that did not exist before and which cannot be deduced from precedents'.[207] Emancipatory politics is messy, unruly, and *disorderly* – resisting the order of the regime it encounters and resists. With its intrinsic 'incalculability',[208] then, spontaneous action can by definition not find representation in the dataset and will thus be registered as an anomaly by the algorithm. As Arendt reminds us, '[t]he new always happens against the overwhelming odds of statistical laws and their probability'.[209]

This is what Arendt meant with the 'inherent boundlessness of action': its 'inherent unpredictability' not just in the sense of an 'inability to foretell all the logical consequences of a particular act', because if it were not more than

---

[204] Abraham (n 5); Yaniv Kubovich, 'Vacation Days for New Targets: Israeli Officers on Bombing Gaza, Casualties and Political Pressure' *Haaretz* (15 December 2019) <https://www.haaretz.com/israel-news/2019-12-15/ty-article/.premium/vacation-days-for-new-targets-how-israel-builds-its-gaza-target-database/0000017f-ef50-df98-a5ff-effdbcdc0000> accessed 26 July 2023.

[205] Tambakaki (n 144) 92.

[206] Borren (n 156) 165; see Arendt, HC (n 153) 189.

[207] Ibid 169.

[208] Arendt, OT (n 162) 598.

[209] Arendt, HC (n 153) 178.

that, then 'an electronic computer would be able to foretell the future':[210] if simple logical complexity were the issue, an algorithm would indeed be the right instrument to calculate human action. But no, the unpredictability of spontaneous action by definition reaches beyond the capacities of any algorithm. The very idea that big data analysis with machine learning algorithms could ever generate valuable and reliable predictions about collective politics is based on a conflation of 'action' with 'work' – algorithmic rationalities unfold according to Arendt's concept of 'work', meaning the imposition of a preconceived model upon the world,[211] which in the case of machine learning algorithms was created by means of analysing the training data. In the case of such 'fabrication' (i.e., work), 'the light by which to judge the finished product is provided by the image or model perceived beforehand'.[212] Arendt warned that attempting to apply this approach to the world of politics, i.e. the realm of 'action', betrays either 'the delusion that we can 'make' something in the realm of human affairs' or 'the utopian hope that it may be possible to treat men as one treats other "material"'.[213]

This explains why the models engendered by machine learning algorithms are incapable of accounting for the intrinsic unpredictability of human action. However, while this must have a direct impact on the accuracy and reliability of predictive outcomes, it does not follow that the systems will simply cease to operate. On the contrary, such spontaneous activities will all the more be registered, yet with unpredictable outcomes for those individuals who are subjected to the predictive technologies. These individuals can never trust that acting in concert politically will not cause the emergence of spurious correlations in the data that raise suspicion and suggest activities that provoke a security intervention; as Krebs has pointed

[210] Ibid 191–2.

[211] Ibid 140–4.

[212] Ibid 192.

[213] Ibid 188.

out, 'anybody and everybody can become a target'.[214] In their quest to uncover 'unknown unknowns' through anomaly detection,[215] the rationalities of machine learning algorithms in AI-based military DSS thus render spontaneous political action fraught with great risk for anyone involved.[216]

Arendt makes clear that the 'inherent boundlessness of action' means that such activity is always and inevitably risky to some extent, but in normal societal configurations of modernity, such risk is at least somewhat mitigated through the protections offered by 'the various limitations and boundaries we find in every body politic',[217] which is necessary for individuals to be able to fully express their humanity.[218] In democratic societies, the arrangement that fulfils this function is that of the rule of law, which defines the limits of tolerated action and thus establishes a sense of predictability for the subjects, who as a result are mostly able to rely on the given legal determinations to guide their behaviour.

By now it has become apparent how the 'opaque epistemologies' of machine learning undercut any such sense of reliability. The unpredictability of spontaneous political action – the input data – renders algorithmic processes – the output – unpredictable. As a consequence of the 'lack of control and inability to predict the next violent episode',[219] the – frequently lethal – security interventions are experienced by those subjected to them as random

[214] Krebs (n 22) 206.

[215] Aradau and Blanke (n 79) 76.

[216] In the specific context of U.S.-led drone warfare as part of the "war on terror" see already Edney-Browne (n 114) 1349: "[C]ongregating in busy communal spaces and socially interacting with new people is considered risky. Drone attacks on 'gatherings', 'parties', and 'jirgas' (...) create fear about entering these spaces and participating in these activities". Also see International Human Rights and Conflict Resolution Clinic, Stanford Law School and Global Justice Clinic, NYU School of Law (n 114) 95 ff.

[217] Arendt, HC (n 153) 191.

[218] Voice (n 153) 46.

[219] Edney-Browne (n 114) 1350.

and arbitrary acts of violence. Needless but important to add for the context of Palestine in this regard is that there are no legal remedies available for potentially affected Palestinians, not least as the algorithmically produced target databases by the IDF are secret, but also because targeting decisions will often be made instantaneously based on incidental correlations and 'emergent patterns'[220] becoming visible within the dataset.

To the extent that it thus follows that constant algorithmic surveillance for the purposes of warfare does not simply render spontaneous political action fraught with risks but effectively suppresses the potentiality of imagined political futures that may arise from spontaneous acting in concert, finally, it follows that the technology is totalitarian – perhaps not in its intent but in its impact. As Arendt reminds us, the primary mode for any totalitarian regime to establish and exert control is the elimination of spontaneous action.[221] In a situation where individuals are unable to predict the reaction to their actions due to the randomness of violence, they will not only be frightened 'into impotence',[222] but it will indeed be a rational response to 'avoid all intimate contacts'[223] if any spontaneous association or assembly might be picked up by the algorithm and marked as suspicious. Such self-isolation out of necessity, in turn, prevents the actualisation of any emancipatory politics directed at engendering genuine change – recall that according to Arendt, political power to form a common will is contingent on the ability to act 'in concert',[224] indeed that 'to be isolated is to be deprived of the capacity to act'.[225] By modifying their behaviour in an attempt to

[220] Gordon, Mignot-Mahdavi and Meerssche (n 81) 134.

[221] Arendt, OT (n 162) 598; see on this Michal Aharony, 'Hannah Arendt and the Idea of Total Domination' (2010) 24 Holocaust and Genocide Studies 193.

[222] Lon Fuller, *Morality of Law* (Yale University Press 1969) 40; see on this further Colleen Murphy, 'Lon Fuller and the Moral Value of the Rule of Law' (2005) 24 Law and Philosophy 239.

[223] Arendt, OT (n 162) 423.

[224] Arendt, OV (n 158) 44.

[225] Arendt, HC (n 153) 188.

mitigate the risks originating with the algorithmic data practices, individuals become conditioned and *calculable*, which in Arendt's theory is the last step toward achieving '[t]otal domination'[226] even if the potential for spontaneity itself can never be extinguished entirely.[227] Ultimately, this is how, by freezing the past and treating it as a model that is imposed on collective human behaviour to generate predictions about the future that lead to targeting decisions or other security interventions, algorithmic surveillance practices corrode the possibility of spontaneous political action and thus of the exercise of the right to self-determination.

V. Concluding Remarks

In this article, I have defended the claim that the pervasive surveillance practices employed for the purpose of training and feeding AI-based military DSS negate the conditions of possibility of spontaneous and collective political action, a practice that is both a precondition of and legally secured by the right to self-determination. I have argued that the political theory of Rosa Luxemburg and Hannah Arendt provides the conceptual tools to understand how the intrinsically backward-looking principles of machine learning cannot but stifle a practice that is determined by spontaneity as required to initiate a transformative and emancipatory politics of change. This far-reaching consequence of the increasing proliferation of the use of machine learning algorithms in the conduct of military operations has so far been largely overlooked in the prevalent discourse in international legal scholarship. The article, in contrast, has demonstrated how the focus on the rules of IHL makes the use of such technologies seem legally imperative once we accept the premise that technological progress will soon and inevitably lead to the superiority of machines when it comes to targeting precision and thus the sparing of the lives of civilians.

[226] Hannah Arendt, 'Social Science Techniques and the Study of Concentration Camps' (1950) 12 Jewish Social Studies 49, 60.

[227] Arendt, *The Promise of Politics* (n 165) 128.

One might be tempted to look at the issue, then, through the prism of an ostensible conflict of rules of *jus cogens*: after all, in its Draft conclusions on identification and legal consequences of peremptory norms of general international law (*jus cogens*), the International Law Commission referred to both the 'right of self-determination' and the 'basic rules of international humanitarian law' – whose content the ILC Study Group on the fragmentation of international law had described as amounting to 'the prohibition of hostilities directed at civilian populations'[228] – as having peremptory status.[229] If that is the case, a doctrinal approach might call for an attempt to disentangle and then somehow resolve such a normative conflict.[230] But this would mean to already have bought into the false dichotomy the prevailing IHL narrative engenders and entrenches. Ultimately, however, we must reject the insinuation that we *need* machine learning algorithms in decision support systems in order to improve IHL compliance and that all it will take to preserve the rights of affected populations is to inject some considerations borrowed from privacy and data protection principles and the contested notion of meaningful human control.

In the realm of warfare, fairness is no appropriate category to appraise the deployment of machine learning technologies. As the article has demonstrated, doing so fails to account for and will only further entrench the larger harms to communities affected by algorithmic warfare by rationalising that harm and presenting it as an inevitable trade-off in the pursuit to protect the life of civilians in armed conflict with the assistance of

[228] Report of the Study Group on the fragmentation of international law (finalised by Martti Koskenniemi), UN Doc. A/CN.4/L.682 and Add.1, 13 April 2006, at para. 374.

[229] International Law Commission, Draft conclusions on identification and legal consequences of peremptory norms of general international law (*jus cogens*), with commentaries, 2022, at 87.

[230] See e.g. João Ernesto Christófolo, *Solving Antinomies Between Peremptory Norms in Public International Law* (Schulthess 2016); Valentin Jeutner, 'Rebutting Four Arguments in Favour of Resolving Ius Cogens Norm Conflicts by Means of Proportionality Tests' (2020) 89 Nordic Journal of International Law 453.

cutting-edge technology. In that respect, such a fairness narrative revolving around privacy and data protection can be seen as yet another building block in the larger, much-scrutinised account that upholds the virtues of humanitarian law to sanitise warfare at the expense of avoiding war in the first place.[231] For the case of Palestine in particular, it furthermore helps to bolster 'Israel's liberal democratic investment in humanitarian gestures of "let live"'[232] while obscuring the fact that any technological improvement to spare civilians in the name of the laws of armed conflict will only legitimise and reinforce the continued control of the Palestinian people. At the same time, while it is important to acknowledge and emphasise the specific situation and lived experience of Palestinians, the arguments made in this paper point to the larger implications for the possibility of an emancipatory politics in the algorithmic age beyond the 'Palestine Laboratory'[233] of occupation and modern warfare. In this, despite Geoff Gordon, Rebecca Mignot-Mahdavi and Dimitri van den Meerssche recently having compellingly deemed 'reinvigorated ideals of liberal subjectivity to be ill-suited in curtailing technoscopic regimes, especially for those historically made vulnerable',[234] I nevertheless want to insist on preserving the ability to act spontaneously in concert as the precondition to create the 'capacity for resistance'[235] that opens up the potential to imagine an alternative future.

[231] On this only Samuel Moyn, *Humane: How the United States Abandoned Peace and Reinvented War* (Verso 2022); Craig Jones, *The War Lawyers: The United States, Israel, and Juridical Warfare* (Oxford University Press 2020).

[232] Jasbir K Puar, *The Right to Maim: Debility, Capacity, Disability* (Duke University Press 2017) 141.

[233] Antony Loewenstein, *The Palestine Laboratory* (Verso 2023).

[234] Gordon, Mignot-Mahdavi and Meerssche (n 81) 138.

[235] Tambakaki (n 144) 99.