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.