Hao Wang & Marijn Sax

Transparency-as-a-Premium-Service: Algorithmic Justice and New Exploitative Commercial Practices

Transparency plays a key role in the literature on the societal impact of algorithms and artificial intelligence (AI). More specifically, the *lack of* transparency in the development and deployment of algorithmic and AI systems is often – for good reason – described as a problem. It is in this context that the metaphor of the black box figures prominently and is used as an explanation for the emergence of new societal harms and injustices resulting from algorithmic systems (see, e.g., Pasquale 2016, Leerssen 2020, Wang 2022).

The flipside of conceptualizing the lack of transparency as a cause or facilitator of societal harms and injustices, is the call of *more* transparency as a regulatory objective. 'Opening the black box' should provide policy makers, regulators, academics, journalists, and citizens with more information about the functioning of algorithmic systems. This, in turn, should help these various actors to both anticipate and mitigate risks. A prominent and recent example is the European Union's legislative agenda including the Digital Services Act, Digital Markets Act, and the AI Act. New transparency obligations play a key role in the EU's legislative push to hold provides of algorithmic systems accountable for societal harms and injustices (Eifert *et al.* 2021).

Despite the loud calls of various actors for better transparency practices around AI and algorithmic systems, most companies developing and deploying these systems are still very reluctant to do so for at least two reasons. First, they claim that 'the algorithm' is their secret sauce which secures their competitive edge vis-à-vis competitors (see, e.g., Dhinakaran 2021). Second, opening up their algorithms to public scrutiny could result, so they claim, in users using the new-found information to (learn to) *game* the algorithm for better personal outcomes (Cofone & Strandberg 2019). In the emerging tug of war between companies using opaque algorithms and other actors trying to pry open the black boxes, **we observe a possibly new market dynamic that can turn this ongoing transparency debate upside down**.

In this paper we explore a **new phenomenon in the marketplace**: Instead of worrying about users' potential manipulation of algorithms to secure better personal outcomes, some companies have been deliberately opening parts of its black-boxed algorithms to their users and inform/coach them how to game their own algorithmic recommendation and curation systems. Notedly, such a deliberate transparency of algorithmic systems is often packaged in a commercial service that is only accessible by those who pay a premium. As a result, **transparent AI, a desirable high-level objective for democratic societies, has turned into a business model**

for companies. A prominent example is Tinder, which announced a paid-for coaching service ('Tinder Concierge') to help users optimize their profiles to perform better vis-à-vis non-paying competitors on the dating 'market' in the opaque workings of Tinder's own algorithm. Such paid-for algorithm gaming tools not only provide some insight on how its algorithm works but also promise better outcomes to those who pay.

This new transparency-as-a-premium-service phenomenon raises, at least, **two sets of questions we seek to address**. First, this business model raises questions about (new?) incentives for companies. If lack of transparency serves as the necessary preconditions for offering a premium service that helps some – i.e., the paying – users to *overcome* a lack of transparency for better personal outcomes vis-à-vis non-paying users, then **companies have an incentive to artificially introduce even more opaqueness**. In exploring this dynamic, we can draw on literatures on for instance search engine optimization (e.g., Ziewitz 2019) and freemium gaming (Sax & Ausloos 2021) where a similar dynamic seems to be present.

Second, we want to explore ethical challenges introduced by transparency-as-a-premium-service. From an ethical perspective, the deliberate introduction or maintenance of a lack of transparency for the explicit purpose to capitalize on that very same lack of transparency, seems conceptually similar to cases of exploitation (e.g., Elster 1982, Roemer 1982, Mayer 2007, Zwolinski 2012, McKeowan 2016, Wollner 2019), parasitism (Van Donselaar 2009), and manipulation (Lanzing 2019, Susser *et al.* 2019, Sax 2022). So, one objective is to **use theories of exploitation, parasitism, and manipulation to conceptualize the relation between providers of transparency-as-a-premium-service and their (non-)users**. Another, related, objective is to use the conceptual analysis to **theorize wrong-making features** of such provider-user relations.

The overall aim of the paper is to provide a first conceptual and ethical exploration of the 'transparency-as-a-premium-service' phenomenon. The resulting analysis should reveal 1) what is – or isn't – distinct about transparency-as-a-premium-service; 2) what is – or isn't – problematic about this phenomenon; and 3) whether transparency-as-a-premium-service warrants regulatory attention.

References
Cofone, I.N., Strandburg, K.J. (2019). Strategic Games and Algorithmic Secrecy. *McGill Law Journal, 64*(4), 625-663.

Dhinakaran, A. (2021). Overcoming AI's Transparency Paradox. *Forbes*. < https://www.forbes.com/sites/aparnadhinakaran/2021/09/10/overcoming-ais-transparency-paradox/?sh=1a9113564b77 >

Donselaar, G. van (2009). *The Right to Exploit: Parasitism, Scarcity, Basic Income*. Oxford University Press.

Eifert, M., Metzger, A., Schweiter, H., & Wagner, G. (2021). Taming the Giants: The DMA/DSA Package. *Common Market Law Review, 58*, 987-1028.

Elster, J. (1982). Roemer versus Roemer: A Comment on 'New Directions in Marxian Theory of Exploitation and Class'. *Politics and Society, 55*(3), 363-373.

Pasquale, F. (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Lanzing, M. (2019). "Strongly Recommended:" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. *Philosophy & Technology, 32,* 549-568.

Leerssen, P.J. (2020). The Soap as a Black Box: Regulating Transparency in Social Media Recommender System. *European Journal of Law and Technology, 11*(2), 1-51.

Mayer, R. (2007). What's Wrong with Exploitation? *Journal of Applied Philosophy, 24*(2), 137-150.

McKeowan, M. (2016). Global Structural Exploitation: Towards an Intersectional Definition. *Global Justice: Theory and Practice, 9*(2), 155-177.

Roemer, J. (1982). *A General Theory of Exploitation and Class*. Harvard University Press.

Sax, M., & Ausloos, F. (2021). Getting Under Your Skin(s): A Legal-Ethical Exploration of Fortnite's Transformation into a Content Delivery Platform and Its Manipulative Potential. *Interactive Entertainment Law Review, 4*(1), 3-26.

Sax, M. (2022). *Between Empowerment and Manipulation: The Ethics and Regulation of For-Profit Health Apps*. Wolters Kluwer.

Susser, D., Roessler, B., Nissenbaum, H. (2019). Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review, 4*(1), 1-45.

Wang, H. (2022). *Algorithmic Colonization: Automating Love and Trust in the Age of Big Data*. PhD dissertation, University of Amsterdam.

Wollner, G. (2019). Anonymous Exploitation: Non-Individual, Non-Agential and Structural. *Review of Social Economy, 77*(2), 143-162.

Ziewitz, M. (2019). Rethinking Gaming: The Ethical Work of Optimization in Web Searches. *Social Studies of Science, 49*(5), 707-731.

Zwolinksi, M. (2012). Structural Exploitation. *Social Philosophy and Policy, 29*(1), 154-179.