Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers
Sheffield, UK / 30 Oct - 2 Nov 2024

TRUST IN ALTERNATIVE GOVERNORS: EXPLORING USER CONFIDENCE IN COMPANIES, STATES AND CIVIL SOCIETY IN PLATFORM CONTENT MODERATION

Dennis Redeker
University of Bremen

Introduction

Social media platforms such as Facebook, X and TikTok are the "new governors" or "custodians" of the Internet (Klonick 2018; Gillespie 2018). How they moderate global speech online affects the communication practices of billions of people and it can make or break social movements and political resistance, and generally be a critical risk factor for human rights violations. These platforms are increasingly joined by states, international organizations, civil society, journalists and others in defining and interpreting the limitations of speech online, be it through legislation, guidelines or by helping platforms to distinguish misinformation from legitimate content. In parallel, researchers ponder questions concerning the legitimacy of various approaches of content moderation (Haggart & Keller 2021; Suzor 2019), which must extend to the question of which actors ought to fulfill which function in the moderation of content. A legitimate content moderation constellation (and potentially division of labor) is arguably one that is perceived to be legitimate by the "governed" themselves (for whatever qualities are appraised by them). As of today, however, we have little empirical knowledge about what users actually think about content moderation in general. Even less so, we know what users think about different roles for states, international organizations or different civil society actors in platform content moderation.

The current paper presents novel empirical evidence on how users perceive platform content moderation and how they perceive content moderation roles of different governors of speech. "Roles" in content moderation are here defined to relate to the making of rules for platforms, the enforcement of rules and the adjudication of appeals as last-resort decisions. Among the seven potential "alternative governors" covered in this paper are platforms (Meta Inc.), state institutions (the parliament, the government and courts respectively), and international organizations (the UN, etc.). In addition, four

Suggested Citation (APA): Redeker, D. (2024, October). *Trust in alternative governors: Exploring user confidence in companies, states and civil society in platform content moderation*. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

groups representing civil society at large are included, including organized civil society (NGOs, etc.), journalists, academics, and users themselves. The quantitative analysis is based on a survey of more than 15,000 Facebook and Instagram users in 33 countries, which was conducted in six languages in late 2022 and early 2023. While with a focus on countries in Africa, Asia, Latin America, the data also allows comparison to attitudes in Europe through inclusion of Switzerland and six EU member states - Poland, Hungary, Romania, Croatia, Spain and Portugal - in the sample. Using this broad cross-sectional dataset, the paper shows that - far from observing one overarching trend of users entrusting specific (alternative) governors with functions of content moderation, significant differences exist between countries.

The paper inter alia shows that there are distinct differences between the “Global South” and Europe regarding the question of who should be in charge of which function in platform content moderation. As an illustration of this, for instance, respondents living in Switzerland, on average, tend to favor the Swiss government enforcing content moderation rules on social media platforms, but not the platforms themselves. In contrast, on average, users in Nigeria, Indonesia or Turkey heavily favor Meta to enforce the rules on Instagram and Facebook, but not the government. The paper discusses country- and regional differences and correlates these findings with more general responses to questions about the trust in various state and non-state institutions. Not unsurprisingly, there is a strong correlation between how much respondents from a country-level sample trust in their government or trust in Meta as a company and how they want these actors to be involved in (different) functions of the content moderation process. In addition, within-country differences regarding attitudes toward alternative governors - specifically those related to gender and age - are being discussed for selected countries.

Other questions that relate to the trust in (alternative) governors are also addressed in the paper. For instance, in a separate analysis, the paper examines how content moderation priorities in different countries relate to trust in different possible governors. The data shows that respondents from Belarus are on average most concerned about misinformation, government surveillance and censorship on social media platforms (in that order). In contrast, respondents from the Philippines are on average most concerned about misinformation, bullying and hate speech (in that order). The relatively varying concerns and priorities across countries is systematically related to preferences for who should be involved in content moderation, specifically with regard to the role of state institutions. This is specifically challenging amid the discussion of UNESCO’s recently published “Guidelines for Regulating Digital Platforms”, which has arguably correctly been criticized for overlooking the enormous cross-country differences in democratic capacity of state institutions. However, taking into account different levels of trust in alternative moderators is crucial for a meaningful discussion of how multistakeholder approaches and the inclusion of alternative governors can protect user rights.

While the paper discusses current proposals concerning alternative content moderation arrangements, including the trend toward greater involvement of state institutions in different functional phases of moderation, it also discusses the relevance of the findings in relation to the demand for greater decentralization of content moderation. This includes moves toward greater involvement of civil society and users themselves (again) in the moderation of content, amid the recent rise of the open-source platform Mastodon. Finally, the paper critically engages with the method chosen for data-collection, including a discussion about the limitations of the specific empirical approach. Deriving from this, possible avenues for future research are discussed, including a multi-platform approach to surveying content moderation attitudes.

References

Bygrave, L. A. (2015). *Internet Governance by Contract*. Oxford: Oxford University Press.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven: Yale University Press.

Haggart, B., & Keller, C. I. (2021). Democratic legitimacy in global platform governance. *Telecommunications Policy*, 45(6), 102152.

Kaye, D. (2019). Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression - A/HRC/38/35, available at documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf.

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review* 131, 1598–1670.

Suzor, N. (2019). *Lawless. The Secret Rules That Govern Our Digital Lives*. Cambridge: Cambridge University Press.