Separation of Function for AI: Restraining Speech Regulation by Online Platforms

Maayan Perel

The Free Speech Clause of the First Amendment of the U.S. Constitution restricts government regulation of private speech. However, it does not regulate private speech. New forms of speech regulation by online platforms disrupt this constitutional framework. Platforms, such as Google, Facebook and Twitter, are responsible for mediating much of the public discourse and governing access to speech and speakers around the world. These private businesses match users and content in whatever way best benefits their commercial interests. At the same time, however, they exercise regulatory power when they filter, block and remove content, at the request of governmental agents or state actors. Consequently, platforms effectively blend law enforcement and adjudication powers, and sometimes even lawmaking powers.

Courts and scholars who tackle speech regulation by platforms have basically relied on the well settled constitutional divide between private functions and governmental ones. To the extent that platforms exercise governmental powers in allowing or banning speech or speakers, platforms should be subject, as the argument goes, to public law principles of accountability, legitimacy, oversight, and power separation.

In this paper we question this approach. As a practical matter, the public/private framework presumes that public functions of a private entity could be neatly separated from its standard business affairs. We argue that with the increasing use of Artificial Intelligence (AI) by platforms for content moderation, the public, law enforcement functions are integrated with the private, business functions that are driven by

commercial interests. The same technical design which is used for targeted advertising and for curating personalized content is also deployed for monitoring and censoring online content. Using machine learning, the system is informed by the same labeling of users and content, and makes use of the same API, learning patterns, and software. Consequently, decisions on removal of speech, for (public) law enforcement purposes, are driven by the same data, algorithms and optimization logic which are also underlying all other functions performed by the platform. Therefore, the use of AI in content moderation calls for a fresh approach to restraining the power of platforms and securing fundamental freedoms in this environment.

This paper takes a system perspective to speech regulation. It contends that the normative distinctions between public and private functions could be upheld in online content moderation, provided that these distinctions are embedded in the system design. It introduces "separation of functions", a novel approach to restraining the power of platforms while enhancing the accountability in AI driven content moderation systems. We propose to facilitate independent tools embedding public policy. These tools would run on the platforms' data, and would include their own optimization processes informed by public policy. Such separation between independent public tools and private data may enhance public scrutiny of law enforcement speech restrictions which are a traditionally exclusive public function. This functional separation may also facilitate competition among different players who may enrich the design of speech regulation and mitigate biases. Finally, we explore the implications of this approach and discuss its possible limitations