11

Introduction

Platform Governance

Kyle Langvardt

The term "content moderation," a holdover from the days of small bulletin-board discussion groups, is quite a bland way to describe an immensely powerful and consequential aspect of social governance. Today's largest platforms make judgments on millions of pieces of content a day, with world-shaping consequences. And in the United States, they do so mostly unconstrained by legal requirements. One senses that "content moderation" – the preferred term in industry and in the policy community – is something of a euphemism for content *regulation*, a way to cope with the unease that attends the knowledge (1) that so much unchecked power has been vested in so few hands and (2) that the alternatives to this arrangement are so hard to glimpse.

Some kind of content moderation, after all, is necessary for a speech platform to function at all. Gus Hurwitz's "Noisy Speech Externalities" (Chapter 12) makes this high-level point from the mathematical perspective of information theory. For Professor Hurwitz, content moderation is not merely about cleaning up harmful content. Instead, content moderation becomes most important as communications channels approach saturation with so much content that users cannot pick out the signal from the noise. In making this particular case for content moderation, Professor Hurwitz offers a striking inversion of the traditional First Amendment wisdom that the cure for bad speech is more speech. When speech is cheap and bandwidth is scarce, any incremental speech may create negative externalities. As such, he writes, "the only solution to bad speech may be *less* speech – encouraging more speech may actually be detrimental to our speech values." Professor Hurwitz therefore suggests that policymakers might best advance the marketplace of ideas by encouraging platforms to "use best available content moderation technologies as suitable for their scale."

Laura Edelson's "Content Moderation in Practice" (Chapter 13) provides some detail on what these technologies might look like. Through a survey of the mechanics of content moderation at today's largest platforms – Facebook, YouTube, TikTok, Reddit, and Zoom – Dr. Edelson demonstrates that the range of existing

https://doi.org/10.1017/9781009174411.015 Published online by Cambridge University Press

techniques for moderating content is remarkably diverse and complex. "Profound differences in content moderation policy, rules for enforcement, and enforcement practices" produce similarly deep differences in the user experience from platform to platform. Yet all these platforms, through their own mechanisms, take a hardline approach toward content that is "simply illegal" or that otherwise contravenes some strong social expectation.

In Chapter 14, "The Reverse Spider-Man Principle: With Great Responsibility Comes Great Power," Eugene Volokh examines the hazards that arise when private go-betweens assume the responsibility of meeting public expectations for content regulation. As companies develop technical capabilities that insinuate them more deeply into human decision-making and interaction, there is a natural temptation to require them to use their powers for harm prevention. But as seen in the case of online platforms, these interventions can create discomfiting governance dynamics where entities micromanage private life without clear guardrails or a public mandate. Volokh argues that courts do grasp this Reverse Spider-Man Principle at some level, and that they have worked to avoid its dangers in diverse settings. Tort law, for example, does not generally hold landlords responsible for screening out allegedly criminal tenants, even if such screening might help protect other tenants from violent crime. If it were otherwise, then the law would appoint landlords as narcotics officers, with likely disastrous consequences for individual liberty.

Alan Z. Rozenshtein's "Moderating the Fediverse: Content Moderation on Distributed Social Media" (Chapter 15) points toward an alternative social media architecture that would address the Reverse Spider-Man problem by dialing down the reach, responsibility, and power of any one community of moderators. This "Fediverse" does not rotate around any single intermediary in the way that today's mainstream social media architecture does. Instead, the Fediverse is held together by a common *protocol*, ActivityPub, that allows any user to found and operate their own "instance." In the case of Mastodon, the Fediverse's most popular social media platform, each instance works a bit like a miniature X platform with its own content policies and membership criteria. Groups of instances, in turn, can enter into federative agreements with each other: Instance A may allow its users to see content posted in instance B, but not content posted in instance C.

This architecture ensures that no one group of moderators has the scale – or the responsibility, or the power – to set content rules that control the shape of public discourse. But achieving this result would require great *effort* in the form of a distributed, almost Jeffersonian moderation culture in which a much larger group of users participates intimately in content decisions. Moreover, it is unclear that the Fediverse lends itself to ad-based monetization in the same way that platformed social media does. The seemingly natural behavioral and economic inclination toward market concentration and walled gardens indicates that public policy will have to play some role in encouraging the Fediverse to flourish. Professor Rozenshtein's chapter offers some suggestions.

https://doi.org/10.1017/9781009174411.015 Published online by Cambridge University Press