Centre for International Governance Innovation

This policy brief is part of CIGI's project on freedom of thought: Legitimate Influence or Unlawful Manipulation?

Find out more at: www.cigionline.org/fot

Policy Brief No. 11 — April 2024

New Logics for Governing Human Discourse in the Online Era

Richard Reisman

Key Points

- The democratization of access to online media tools is driving a transformation of human discourse that is disrupting freedom of thought.
- This shift in the flow of thought is being encoded into a global infrastructure dominated by commercial platform companies whose operations co-opt individual, collective and governmental agency.
- Attempts to govern these tools are relying on "yesterday's logic." The new logic, largely unrecognized, relates to acceleration of word-of-mouth propagation, much like rumouring, putting the listener's freedom of impression, rather than the speaker's freedom of expression, at the fore.
- Governance is needed to restore individual and community agency, which could re-energize the vision of technology as "bicycles for our minds," enabling individuals and society to flourish and maintain resilience in an increasingly challenging world.

Introduction

The democratization of access to online media is the most transformative change in human discourse since Gutenberg's press democratized access to the written word — and it is by far the most sudden ever.

New logics to guide governance are urgently needed. "Platforms" are as much the symptom as the cause of current problems. Society faces a *socio-technical* problem, to be solved by society and technology in concert. At the core of this problem is the cyclical process of thought and discourse. Yesterday's logic was based on amplifying one-to-many flows, first via the printing press, then via broadcast media. Freedom of expression was pre-eminent — or sometimes balanced, to varying degrees, with the other rights of individuals and society.

The new logic, still largely unrecognized, relates to the radical transformation of how this new medium "mediates" human communication messaging. Media theorist Marshall McLuhan taught that "we shape our tools and thereafter our tools shape us."[1] Social media massively accelerate the primal logic of word-of-mouth propagation through each individual and all those they interact with directly or indirectly. This reflexive cycle of thought as a social process, through the stages

1 Whether he first expressed it in those terms or one of his colleagues did — in any case, the phrase is on his recording: http://ubusound.memoryoftheworld.org/mcluhan_marshall/Mcluhan-Marshall_The-Medium-Is-The-Massage_01-Stereo.mp3 at 6:27.

About the Author

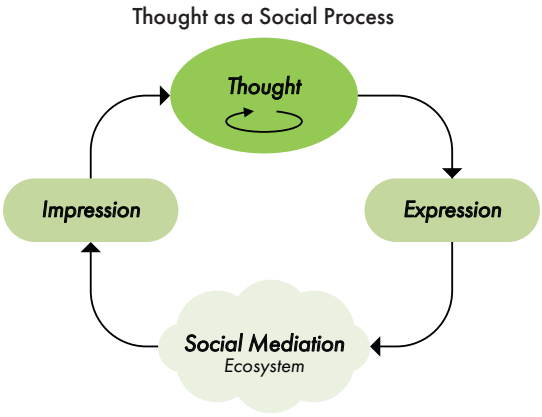
Richard Reisman is a non-resident senior fellow at the Foundation for American Innovation, and a frequent contributor to Tech Policy Press. He blogs on human-centred digital services and related tech policy at SmartlyIntertwingled.com, and his work was cited in a US Federal Trade Commission report to Congress in 2022, *Combatting Online Harms Through Innovation*.

At time of publication, he is part of the team convening the symposium "Shaping the Future of Social Media with Middleware" sponsored by the Foundation for American Innovation and the Stanford Cyber Policy Center. The symposium's discussions will be published in a comprehensive white paper later in 2024.

His book, *FairPay: Adaptively Win-Win Customer Relationships*, and related blog, *The FairPay Zone*, introduce customer-value-first revenue strategies for digital services and were described in 2013 in *Harvard Business Review*. He has managed and consulted for businesses of all sizes, developed pioneering online services, and holds more than 50 media-tech patents licensed by more than 200 companies to serve billions of users (now all in public domain). He has an A.B. in applied math from Brown University and an M.S. in operations research from Lehigh University.

of expression, social mediation and impression, before feeding back to more thought (see Figure 1) now happens far more extensively and rapidly.

Figure 1: Thought as a Social Process



Source: Reisman (2023a). Copyright Richard Reisman; adapted and reproduced with permission.

As society moves online, the shape of these flows among individuals and the ecosystems of social groups becomes encoded into, and further mediated by, software and decision algorithms. Current thinking on managing these flows often suffers from a narrow focus on the individual thought and its expression, without recognizing that computer mediation of communications changes how our human social mediation works — that is, that technological tools shape and influence the patterns and practices of our discourse. How our *expression* of thought (what we speak) is socially mediated — received by listeners and then reflected back into our *impression* (what we listen to) — is changed by software and algorithmic influences on those human interactions, thus shaping our ongoing thought process of sense-making both individually and collectively (Reisman and Riley 2022b, 2022c).

Social media platforms have disrupted the patterns and pace of human social mediation by spreading them far wider but with less context — and doing that so quickly that emotional reflexivity outruns deliberation. Current efforts to mitigate that disruption often focus on limiting *speakers' freedom of expression* instead of on providing support for more considered involvement of our social mediation ecosystem (see Box 1). Far better to apply tools to empower *listeners' freedom of impression*. That would be more effective —

Box 1: What Is a Social Mediation Ecosystem?

- The idea of a social mediation ecosystem integrating with social media feeds is a re-visioning of how things used to work. Society has been organically building on such sense-making ecosystems for millennia.
- The groups that comprise the social mediation ecosystem have historically served as a "public square," or "public sphere," ranging from informal gathering places such as coffee shops and taverns to social and civic associations, the press, academia, churches, unions, workplaces and other communities of interest.
- This square or sphere is not unitary but an ecosystem, a polycentric web of interlinked groups in a multidimensional space (Reisman and Riley 2022c).
- Such associations develop norms and contexts for discourse. Our participation in a network of them shapes what we see and hear of the world. *Both kinds of social influence nudge us to speak "freely," but with sensitivity to those norms and values, so others will choose to listen to us.*
- Online media technology can enable restoration of that mediating role through enterprise-class middleware affordances that support community operation and let users interact both within and across the diverse communities they opt into.
- Middleware can facilitate and enrich user-community interactions, and enable us to steer our feeds to blend content favoured by any mix of communities we choose to include at a given time — depending on our tastes, objectives, tasks and moods.
- For example, current curators of news could become attention agent services. Users might select a set of such services — for example, *The New York Times*, CNN, MSNBC, Fox, *The Atlantic*, *People* — to play a role in composing their feeds, assigning them different relative weights in ranking. Other groups in the social media ecosystem, such as civic, political, faith communities and special interest associations, could also be selected by the user to function as attention agents. Content ranking inputs could come from each community's expert curators/editors or be crowdsourced from the user population that follows those curators, or from a combination of both.
- Importantly, and as it has been historically, this ecosystem must be open and diverse, and users must be able to draw on combinations of many mediation sources to maintain an open and balanced understanding of the world.
- Many fear that the involvement of attention agents or middleware might increase fragmentation and partisan sorting. That may be a concern while there are just one or a few mediators, but being able to selectively combine exposure to many loosely connected communities is how open societies have always limited that ever-present risk.

and more democratically legitimate — than platform control of speech (Reisman 2023a).

Compounding that disruption, concentration in ownership of social media platforms impedes not only the restoring of user agency over algorithmic decisions but also the restoring and enriching of the social mediation ecosystem of communities and institutions. There is now urgent need to restore both user agency and the nuanced role of the kind of social mediation ecosystem that society developed over centuries.

As We Now Think: The Dilemma of Social Media "Moderation"

The blessing of radically powerful media tools that offer almost everyone global reach for their speech has proven to also be a curse. Major social media platforms can feed users far more speech than their attention can handle — much of it undesirable. Centralized platforms seek to remove (censor) harmful speech, but this "platform law" is inflexible, overbroad and underinclusive (Land 2020) — lacking in the legitimacy of agency, subsidiarity and context (Reisman and Riley 2022a).

The problems of truly illegal speech (child sexual abuse, terrorism and the like) are relatively simple compared to the intractability of "lawful but awful" speech — when "awful" is often in the eye of the beholder, and highly dependent on context and community (Keller 2022; Chou 2023).

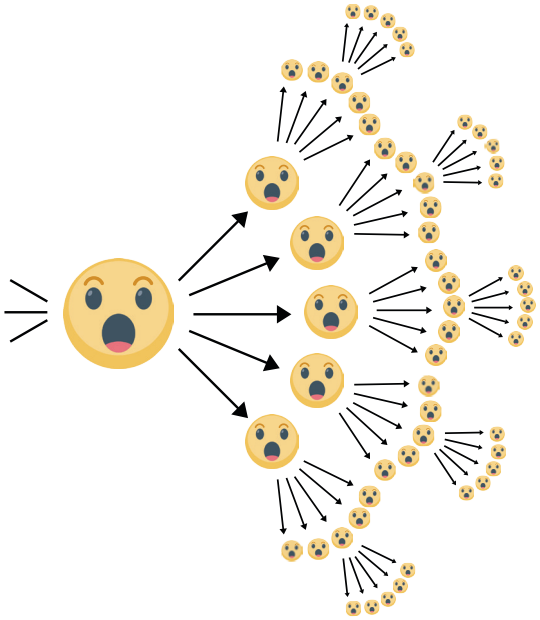
What speech should be selected as desirable or deselected as harmful? By whom? In what way? How can that decision-making process be regulated to serve citizens and society? How should it be governed along the spectrum of democracy, paternalism and authoritarianism?

Misunderstanding Massively Online Media

We are reacting to the turbulent virality of online speech with the logic of yesterday's mass media, viewing it as "amplification" or a "megaphone." But speech in social media is not a matter of single-stage, one-to-many *amplification* but of multi-stage *propagation* from person to person, much like word-of-mouth rumouring (see Figure 2). This propagation happens at an accelerated cycling rate that makes this new medium hyper-reflexive (at the micro level of individual speech interactions), feeding speech back on itself at speeds and scales we do not yet understand how to manage.

The governance challenge is to understand how these feedback loops are driven by *both* algorithmic objective functions *and* human behaviour to reinforce positive or negative spirals, or to seek balance (Johnson 2023). The societal challenge is how to shape it — and ourselves — to augment human discourse, not de-augment it.

Figure 2: Multi-Stage, Word-of-Mouth Propagation of Speech



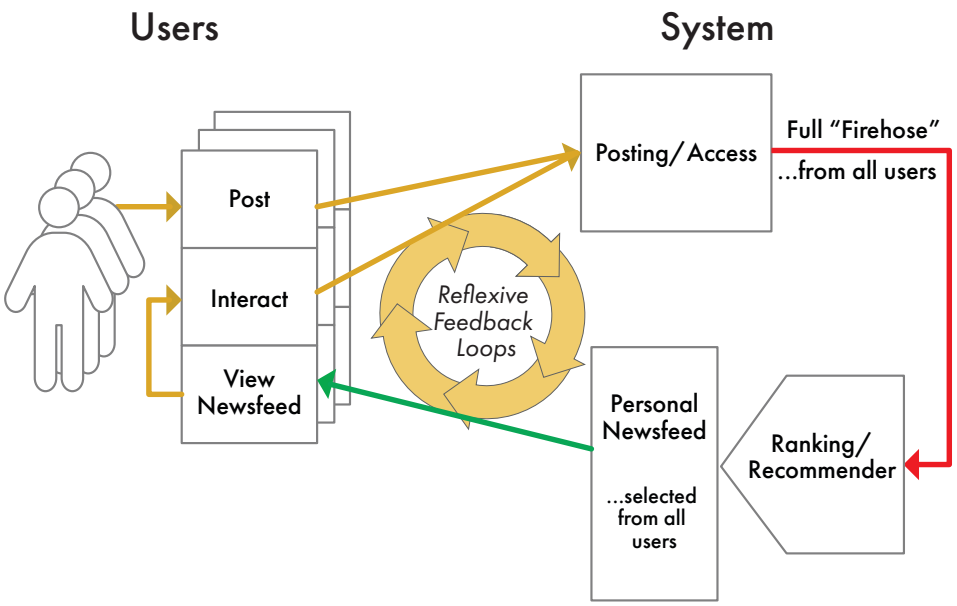
Source: Rose-Stockwell (2023). Copyright Tobias Rose-Stockwell; adapted and reprinted with permission.

Consider the media history of how "we shape our tools and thereafter our tools shape us" (macro-level reflexivity of our media tools). Social media harks back to traditional word-of-mouth processes — but with immensely greater scale and speed. Over centuries, society built a context for understanding, managing and enhancing speech through communities and institutions that grew into a rich social mediation ecosystem (see Box 1).

Early societies were hierarchical and feudal, mediating from the top down. Gutenberg catalyzed centuries of gradual democratization and globalization, giving speakers more reach, and listeners more freedom of impression — deciding who to listen to.

Broadcast technology brought mass media that does amplify in a single stage, like a megaphone. A limited number of costly channels led to pragmatic limits on freedom of speech — tied to the rights and discretion of "publishers" (as editors and curators) to control reach, and subject to new provisions for "equal time" or "fairness." But it

Figure 3: Understanding Social Media — Reflexivity



Source: Reisman and Riley (2022b). Copyright Richard Reisman; reproduced with permission.

remained easy for listeners to change channels. Openness and generativity were constrained, but listener freedom of impression, the ability to direct our own attention, remained taken for granted.

Social media brought back word-of-mouth style propagation. User-generated content on social media has brought freedom of expression and potentially global reach to everyone, everywhere, almost all at once — in a way that feeds reflexively on a backchannel of likes, shares, comments and so forth (user-generated reflexivity that drives algorithmic selections at each stage) (Reisman and Riley 2022b; see Figure 3).

But the way the platforms control "news feeds" is curtailing freedom of impression. Feeds must be selective to avoid attention overload because the full stream of speech quickly grew into a "firehose," rushing too fast to drink from. Sophisticated attention allocation systems now select a manageable number of items to present in each user's feed at a given time, largely based on how other users have reacted to them (Ovadya and Thorburn 2023; Narayanan 2023).

Platforms took on this attention allocation function to keep their users engaged. They decide for each user what they will see from the firehose (with only minimal user controls). Giving users direct control of this would be beyond the skill and patience of most users. Platforms also manage user "groups" with similarly rudimentary control by group "owners." But platform optimization for engagement has proved highly problematic, generating significant blowback.[2]

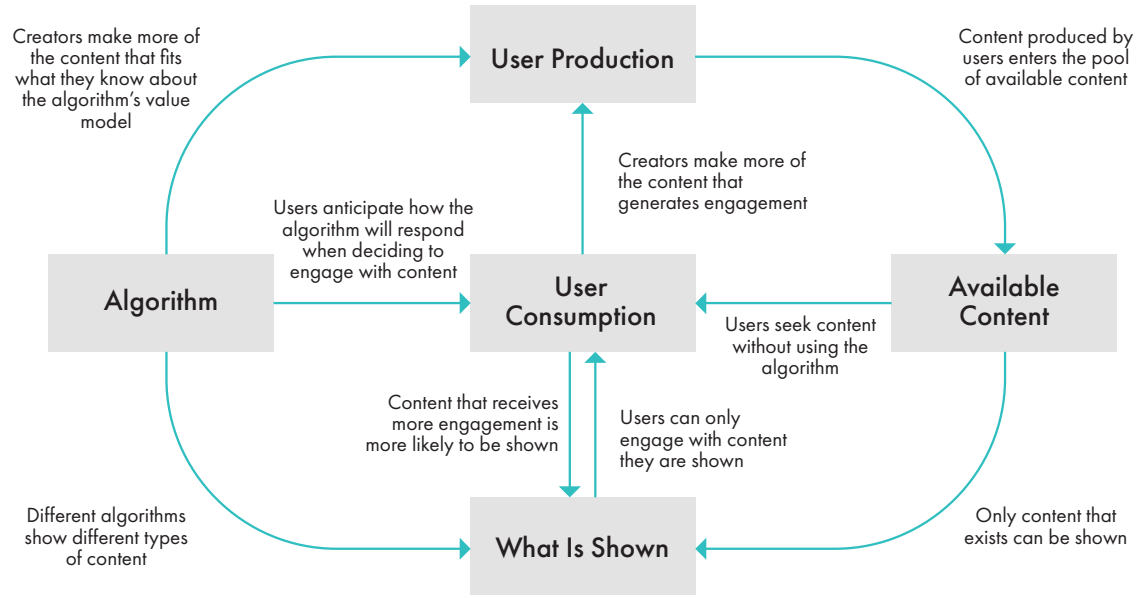
Misdirected by the old logic of mass media, that blowback caused platforms (and policy makers) to double down on the wrong remedy — moderation-as-removal — a solution that is simple, neat and wrong (Hendrix 2022; Gillespie 2022; Goldman 2021; Stray 2022). More nuanced remedies are required: ones that leverage both user agency and the social component of social media to recreate mediation processes to avoid draconian censorship.

The New Hyper-Reflexivity

This reflexivity is shaped by a complex socio-technical system that makes simple notions of "amplification" misleading because "changing the algorithm will also cause user behavior to change" (Stray, Thorburn and Bengani 2023b, para. 14; Eckles 2021; Matias 2023) (see Figure 4).

2 See https://facebookpapers.com/.

Figure 4: User-Algorithm Interaction Patterns



Source: Stray, Thorburn and Bengani (2023b). Copyright Stray, Thorburn and Bengani; adapted and reproduced with permission.

This is an issue not only of individual freedom but of the health of democratic society. As social creatures, humans have evolved a social mediation ecosystem (see Box 1) — a web linking all of our communities and institutions. That was not designed, but emerged and evolved dynamically as the social mechanism that helps us shape both our lives and our thoughts.

Each of us gives legitimacy to the communities and institutions of that ecosystem as we determine what voices to listen to, how they reflect back on what we think, and what and how we express back to others. In terms of recognized human rights, this flow of our ideas through sense-making social mediating influences evokes the online counterpart to freedom of association and assembly.

We underappreciate the extent of our "epistemic dependence" — that is, most of what we know comes not from our own direct experience but from that of others (Hutson 2020). In "Mediating Consent," Renee DiResta (2019) surveys the evolution of social mediation, from how Gutenberg's new media technology liberated speakers and listeners and on into the ages of mass media and social media. Jürgen Habermas (1991) looks back to the emergence of a "public sphere" and how it broke the top-down hierarchy of feudalism, creating a broad public eager for democracy — and how mass media began partly undoing that.

Social media are now breaking up and distorting the web of this nuanced mediation ecosystem, threatening freedom of thought by regimenting how individuals use media to socially construct reality (Couldry and Hepp 2017) under a *new* top-down hierarchy of undemocratic "platform feudalism" (Brennan-Marquez and Susser 2022) that lacks nuance. Emphasizing the importance of "epistemic process," Benjamin Laufer and Helen Nissenbaum (2023) diagnose this as "algorithmic displacement of social trust." For still-broader perspectives on social sense-making and value formation in an open society, see Fukuyama (2022) and Haidt (2013).

Simplistic Thinking about Socio-Technical Problems

Current social media platforms are severely limited in how they support the constellations of communities and institutions that their users participate in. That is where the entanglement of technology, sociology and economics becomes most problematic.

Thus, platforms are not only co-opting human agency over the tools we think with collectively but also disintermediating this pre-existing social mediation ecosystem — *the traditional driver of social cohesion and a shared reality* — with no adequate replacement in sight. Instead of enhancing this ecosystem, social media platforms have been disintermediating and neglecting the interdependence of individuals with their communities and institutions. To support those existing communities, or new ones, the platforms would need to support the wide diversity of human social structures.

It is inevitable that our online social networks will be increasingly "intertwingled" — Ted Nelson's portmanteau evoking the intertwining and intermingling of elements of human thought that he sought to support as non-linear "hypermedia" (Nelson 1974). This is just as our real-life communities and institutions — and thus our patterns of discourse and its governance — are intertwingled in many levels and dimensions as a "federalist public sphere" (Newitz 2023). That federalism echoes the governance design paradigm that the "federalist" founders of the United States government envisioned — a theme that is central to the design of the internet and its early social use (Zuckerman and Rajendra-Nicolucci 2023) and pervades the author's work (Reisman and Riley 2022c), including this brief.

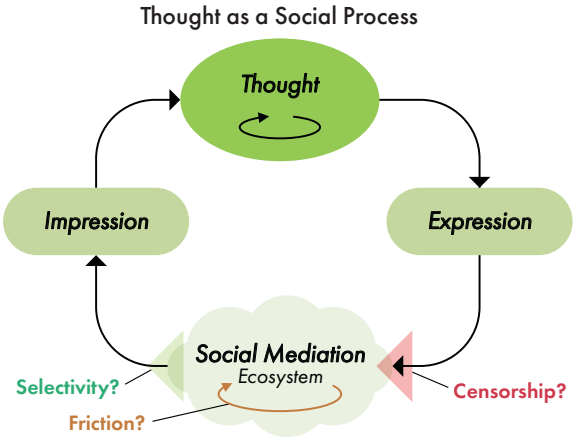
Existential Questions for Democracy and Society

Even if large platforms were motivated to manage individual users' feeds well, that would be difficult for them. It is essentially impossible to do moderation well at scale because of the huge diversity in norms and values across the individuals and communities being served (Masnick 2019; see also Gillespie 2020). There will always be too many false negatives (harmful content propagated) and false positives (valuable content missed). Unlike the traditional social mediation process, automated moderation is based on crude rules and simplistic data that result in "context collapse" (boyd 2013) and "nuance destruction" (Jeff Bezos, quoted in Canales 2020). There is little of the social tracking of reputation and authority that humans naturally rely on.

Even with its huge potential, artificial intelligence (AI) and algorithms using it will not have a reliable understanding of human values and nuance across diverse communities any time soon (Kissinger, Schmidt and Huttenlocher 2023). Nor will these new media tools, working outside of a broader organic social context, be able to duplicate the "epistemic processes" that Laufer and Nissenbaum (2023) see as needed to develop and apply needed levels of "social trust." As they explain, "algorithmic amplification is problematic because, like an invasive species, it chokes out trustworthy processes that we have relied on for guiding valued societal practices and for selecting, elevating, and amplifying content. Problematic content is a symptom, only partially addressed when disassociated from its underlying causes."

These turbulent times call for the wisdom often attributed[3] to Peter Drucker: "The greatest danger in times of turbulence is not the turbulence, it is to act with yesterday's logic." The need is to shift from yesterday's mass media logic of moderation-as-removal of harmful expression to a new logic of reliance on both social mediation and individual judgment (see Figure 5).

Figure 5: Control Points for the Social Mediation of Thought



Social Mediation: Management of Control Points
(Who decides, at what point?)

Source: Reisman (2023a). Copyright Richard Reisman; reproduced with permission.

3 See www.druckerforum.org/blog/leadership-everywhere-means-reversed-leadership-by-jane-mcconnell/.

That new logic — more accurately a reinvention of a still older, but largely unconscious logic — would reinvigorate our collective human intelligence and augment it with new technology. In the words of Emma S. Spiro and Kate Starbird (2023), the still older logic of the "informational and emotional drivers of rumoring can support more empathetic — and perhaps more effective — interventions" that can "engage with an information cascade before veracity or intent can be determined." Fundamentally, determinations of veracity and intent often remain fraught and can unnecessarily increase distrust and affective polarization. Pragmatically, if possibly harmful speech reaches few listeners, it does little harm. And, from a speech rights perspective, if speech does not reach listeners because they choose not to listen, free speech is not infringed. As DiResta (2018) pithily observed, "free speech does not mean free reach."

The old logic of moderation-as-removal is inherently antithetical to freedom of thought. It fails because it is a constraint on expression that is hard to legitimize. In an insightful analysis of the issues in terms of US First Amendment law, former Google counsel Daphne Keller (2021) notes that while platforms are free to censor speech, "governments can't restrict speech *or* reach." Keller highlights giving users control over what content they see as a promising solution — noting legal precedent for doing that, in the ways explained below, as analogous to the "unbundling" requirements in US anti-monopoly telecommunications law that separated out distinct Bell System telephone company functions and opened them to competition, diversity and innovation.

That approach based on user control is broadly preferable, even in territories where governments are less restricted. Consider that moderation-as-removal, as now done by dominant platforms, is effectively "platform law," a pseudo-governmental activity but one with questionable democratic legitimacy. Relying more on user choice would be less restrictive of the rights of users as both speakers and listeners (Fukuyama 2023; Keller and Fukuyama 2024; Reisman 2024b).

Most fundamentally, removal is a draconian loss to the commons of humanity, even when an item's value seems far-fetched. Much better to just downrank from casual view in accord with user preferences (and restrict access if truly illegal). And pragmatically, removal often has perverse effects within and beyond media platforms: quashing speech tends to make it pop up elsewhere as a "Streisand effect" (Masnick 2015) that actually increases attention to it.

A more nuanced strategy is especially important for these increasingly complex socio-technical systems in a global environment composed of many interacting and overlapping communities and jurisdictions — all facing unprecedented challenges and rates of change. As Divya Siddarth, Danielle Allen and E. Glen Weyl (2022) have described, underlying the management of such networks of communities is the principle of "subsidiarity" as a control architecture for enabling "composable local control…leveraging a core principle of both markets and democracy: Those closest to a problem usually have the most knowledge and the greatest stake in its resolution, and it is by aggregating, federating, and filtering this knowledge that the best collective decisions are made" (see also Couldry 2024; Douek 2022; Hasinoff and Schneider 2022; Tambini 2019).

This moves beyond simplistic extremes of centralization or decentralization to distribute and blend controls in ways that are adaptive to both global and local needs — to gain legitimacy and context from the bottom up, while applying needed levels of guidance from the top down. In that way, the basic communications network functionality that truly belongs at a platform level can be open and generic, in the form of utility services that support subsidiarity of specific messaging mediation choices and values — as an open network of distributed function and control layers that ride on top of the platform.

As We May Think: Steering Our Own Attention

In the spirit of McLuhan's view of humans as tool makers, and media as tools, Steve Jobs, in a video from 2006, described humans' relatively inefficient locomotion compared to other animal species, with the condor excelling above all others for speed and energy use. But when the condor was compared to a human on a bicycle, the human "blew the condor away." He

concluded that computers are like "a bicycle for our minds" (Michael Lawrence Films 2006).

Social media tools (in the broad sense) should extend that, but that potential is being squandered. As early Facebook employee Jeff Hammerbacher famously put it, "The best minds of my generation are thinking about how to make people click ads" (Rosenberg 2011).

Instead of Skinner boxes seeking to engage us to click ads to profit the platforms, social media modalities and systems could become bicycles for our minds. That more positive vision is why individual agency is a core focus of this brief — would anyone choose to use a bicycle that is steered externally by some corporation that wants us to view their billboards or visit their sponsored roadside attractions? Or, by a government seeking to herd us like sheep?

Beyond that simple question of agency, consider what that agency might enable. A user's objectives in viewing their feed can vary depending not only on who they are, their interests and values, and what communities they belong to, but also on the tasks and subjects currently of interest to them, as well as their moods.

So where does that individual agency fit in? Consider how current social media platforms compose information "news" feeds and make recommendations of people or groups of potential interest. The democratization of global speech has produced an oversupply, so that we are no longer limited by the supply of speech but by the limits of our attention. The platforms took on the role of "attention agent" for us because the "firehose" of messages and the universe of people and groups that may or may not be of potential interest became overwhelming. We now need tools to help guide our attention. But…should the platforms be the ones to control an individual's attention agents?

Unlike the relatively mechanical task of routing messages over a network, even at global scale, being an agent for guiding a person's attention is one of the most challenging tasks — and one of the most consequential threats to freedom of thought. No one simple attention agent can handle the diversity, nuance and variability of attention needs — it takes a customized blend of agents that balance a wide range of objectives and criteria.

Those objectives and criteria are best served by user-driven composition, orchestration and real-time steering of multiple attention agents to compose feeds as each user chooses (Reisman 2023b). Agency requires that most of these attention agents be user-controlled — but, to add balance, some blend of community- and government-controlled attention agents might be worked into the mix of what goes into our feeds.

The power to steer our own agents also relates to advertisers' claims on our attention and how platforms seek to maximize value for that, rather than for serving their users. Imagine that instead of being controlled by platforms, ad placements could be negotiated by specialized attention agents tailored to suit individuals' willingness to allocate attention to ads — regarding what to show, in what style, and how much it intrudes on our attention. Some might pay to get no ads, some might pay less to accept very selective and unobtrusive ads, while others might choose to pay nothing and accept more ads, but still retain some control over what kind they were (Reisman 2018b).[4]

If new forms of media are to become the boon to humanity they promise to be, governance of online discourse must be not just a matter of preventing manipulation but of supporting a much broader positive objective: ensuring that our media tools actively align with and augment our desires and ability to think freely, productively and collaboratively.

Need for User-Delegated Attention Agents ("Middleware")

The wrong turn of online media makes the half-century-old vision of new media tools that flexibly augment human thought and discourse now seem hopelessly lost (Reisman 2021d), but the tide may be turning. Dissatisfaction with centralized platforms is widespread, and interest is growing for creating an open market in attention agents as essential to the preservation of democracy (Fukuyama et al. 2020) — even if there is also full recognition that creating that open market will take significant development (Reisman 2021b). The idea is to manage this critical and highly individualized

4 For deeper discussion of governance issues in the relationship between attention agents, business models (advertising), and making social media less harmful and more beneficial, see Reisman (2019) and other items in his blog *The FairPay Zone*.

function as a separate mediating layer between the user and the platform — as enabled by software services now often referred to as "middleware."

The power of advanced attention agents *that serve users* is in how they could be orchestrated to capture nuance. This is especially relevant to the all-but-intractable problem of managing attention to "lawful but awful" speech, when one person's awful may properly be another person's wonderful. The attention agent's essential task should be to rank items to be considered worthy of attention in a multidimensional space of attributes — not just the singular dimensions of "engagement" or "awfulness" (Chou 2023). They could be used to uprank or downrank an item in terms of that full set of attributes. Any mix of rankings, in multiple dimensions, from multiple sources, could be combined to create a composite ranking for a user's attention at a given time.

Moderation-as-removal (rather than just as downranking) lacks that multidimensional nuance. Any triggering flaw, in any single sensitive dimension, will completely remove an item from consideration by anyone, no matter how much that item might also properly be upranked as desirable in one or more other dimensions — at least for some users. The problem with moderation-as-removal is that one lowest-common-denominator censor rules for all, leaving discourse improperly impoverished.

While it is possible that a major platform will see the merit in introducing user-directed attention agents voluntarily — or enabling others to do that for them[5] — there is emerging legislative effort to mandate a right of delegation to attention agents (Whitt 2020; 2021) in the United States, via the Communications Decency Act of 1996 and the (pending) ACCESS Act, and still more deeply in (pending) New York State Senate Bill S6686,[6] and in the European Union, in the Digital Markets Act.[7]

Meanwhile, the market context is rapidly evolving. Users and system builders are becoming fed up with the centralized control, governance failures and lock-in-enabled extractive degradation[8] of monopolistic platforms — and are increasingly migrating to a diversity of smaller community platforms that offer greater user agency and less extractive business models. But the global imperative of the internet is leading many of these "decentralized" services to use open protocols for selective interconnection to form a growing fediverse (*fed*erated un*iverse*; see Rozenshtein 2022) of communities running on Mastodon or other compatibly networked platform software (largely free, open sourced). More advanced distributions of function — and support for still-greater, and yet more scalable, diversity and subsidiarity — are emerging in the proof-of-concept Gobo "pluriverse,"[9] and in the now-rapid growth of Bluesky.[10] Thus a market for middleware services is beginning to emerge organically.

With the proliferation and rapid scaling of these smaller but still globally connected platforms, simple usability drives an intensifying need for user-controlled agents that can simplify, harmonize and manage each user's experience across those diverse communities and platforms. That has already begun in the basic form of multi-homing and cross-posting services (Lane 2022; Reisman 2022).

Even more compelling, the challenges of "moderation," mediation and filtering to eliminate illegal content — and manage the more nuanced issues of "lawful but awful" content in these increasingly diverse but interconnecting

5 See Tracy Chou's remarks about Block Party, an early middleware service for Twitter (Hendrix 2021), as well as Jack Dorsey's US House Committee testimony that "Twitter is also funding Bluesky…to develop open and decentralized standards for social media….Bluesky will eventually allow Twitter and other companies to contribute to and access open recommendation algorithms that promote healthy conversation and ultimately provide individuals greater choice" (Dorsey 2021, 3).

6 See, respectively, *Communications Decency Act of 1996*, 47 USC § 230 (2000); US, S.4309, *ACCESS Act of 2022*, 117th Cong, 2022; US, S.6686, *An act to amend the general business law, in relation to social media open application programming*, 2023–24, Reg Sess, NY, 2023.

7 See EC, *Regulation (EU) 2022/1925 of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)*, [2022] OJ, L 265/1.

8 See Cory Doctorow's incisive analysis of this process, encapsulated in his coinage of a 2023 "word of the year" in "The 'Enshittification' of TikTok" (Doctorow 2022).

9 See Rajendra-Nicolucci and Sugarman (2023); Lane (2022).

10 See Graber (2023a, 2023b, 2023c, 2023d); The Bluesky Team (2024a, 2024b); Masnick 2024.

services — are growing far beyond the ability of small community operators. That necessitates shared services that can be managed in a way that accommodates this diversity of norms and "community standards" (Roth and Lai 2024; Rozenshtein 2023; Thiel and DiResta 2023). Bluesky and Gobo are working to demonstrate how the solution can be enabled by middleware. Recently, Mike Masnick (2024) insightfully explained "Why Bluesky Remains The Most Interesting Experiment In Social Media, By Far."

Given all these issues, it seems a new layer of attention agents between users and network platforms may thus evolve from the centre out, or from the edge in, or very likely both — as suggested by the commitment of Meta's new Threads service to federate with Mastodon and other decentralized services (Peters 2023; Hogg and Schönander 2023).

This is not to say that user-agent solutions are a panacea, as concern remains that giving users greater autonomy might leave them even more at risk of falling into harmful "filter bubbles," "echo chambers" and "rabbit holes" — all forms of thought-narrowing feedback loops (Stray, Thorburn and Bengani 2023a). The real mechanism and impact of this danger remain unclear and require more structured study. But in any case, the counter-argument is that compartmentalizing a global network's reflexive mediation processes into a multiplicity of diverse community-based attention agents should reduce the contagious virality of harmful content driven by cascading feedback loops — much as compartmentalizing an ice-cube tray reduces sloshing, a fluid dynamics analogy that is more fitting than it may seem (Wilhelm 2023).

Balancing that, these competing attention agents can appeal to users by providing a positive service of "bridging" that supports mediating crossover. Users might actively choose such services to productively expose them to different communities that share some of their interests and values. Underlying this objective are the normative principles that societies are healthier by such measures as robustness, creativity, generativity and resilience when they remain open, diverse and polycentric (New_ Public 2024).

Attention agents that support bridging could provide new ways to reduce the current rise of *affective polarization* (as emotional dislike and distrust of persons, versus more cerebral *ideological polarization* as differences in policy positions) and the related "partisan sorting" effect. Bridging agents might shift attitudes from spirals of antagonism toward a more accepting agonism of openness to diverse views (Reisman 2012; Reisman and Riley 2022c, 2022d; Ovadya and Thorburn 2023; Törnberg 2022; Bail et al. 2018). Letting users choose to opt-in to such bridging agents could provide these benefits in ways that focus on the rights of listeners without restricting the rights of speakers in ways that might not pass legal scrutiny (Reisman 2024a).

In any case, user agents should be applied in concert with other highly scalable non-removal remedies,[11] as well as via more labour-intensive complements.[12] And as a socio-technical problem,[13] technical solutions must be complemented by "trust and safety" teams at both community and attention agent levels. Broader remedies such as education, public service media and greater support for the social mediation ecosystem, including reduction of structural inequities that marginalize disadvantaged communities, are also very necessary, and require the combined efforts of government, industry and civil society.

A Full Spectrum of Ranking Methodologies and Data, Only Partially Exploited

The art of composing news feeds for individual users of interactive media that have global reach requires balancing human and AI support. Purely human moderation becomes intractable at scale because handling context and nuance requires frequent, skilled and thoughtful exception handling (Masnick 2019). Direct artificial natural language understanding of content has proven useful for triage — but is limited beyond that, because of its high error rates and failure to comprehend context and nuance (Gillespie 2020). Even the recently emerging large language models such as ChatGPT suffer from high error rates, bias, lack of common sense and outright fabulation (so-called hallucination) (Kissinger, Schmidt and Huttenlocher 2023).

11 See Goldman (2021); Gillespie (2022); Ovadya (2021); Goodman (2021); DiResta and Rose-Stockwell 2021; Goodman, Slater and Hand (2022); and Rajendra-Nicolucci and Sugarman (2023). See also https://help.twitter.com/en/using-x/community-notes.

12 See www.newsguardtech.com/solutions/newsguard/ and Ovadya (2021).

13 See Task Force for a Trustworthy Future (2023), in particular, annex 5.

Thus, attention agents should rely on a hybrid of human and machine intelligence, much as was suggested decades ago as a way of "augmenting human intellect" (Engelbart 1962). Large platforms already apply such hybrids in optimizing for advertisers, but the largely untapped power of similar augmentation for user objectives has been suggested by this author (Reisman 2004, 2012, 2018a), and more recently by Facebook (Allen 2019) and the Integrity Institute (Allen 2022).

The idea is that instead of expecting AI technologies to understand quality and value by understanding content themselves, they instead can collect and distill *signals of human judgment* of quality and value. A robust strategy for augmenting human wisdom in attention agents would draw on both explicit and implicit signals of human judgments. Explicit signals, such as expert or crowdsourced ratings and tags, can often be most valuable, but they are costly in human effort, and user participation is often insufficient. More scalable, and widely used by platforms, are the implicit signals derived from routine behaviours that incidentally imply a judgment, such as likes, shares and comments in social media. While these implicit signals are typically less accurate and nuanced than explicit ratings, their abundance can more than compensate for those shortcomings through judicious analysis.

When done well by favouring reputable users, this can augment the human "wisdom of crowds." When done badly as a popularity contest that ignores user reputation, as happens now in optimizing for engagement, they can de-augment the "madness of crowds."

Mediation, Reputation and a Cognitive Immune System

Forgetting the lessons of how traditional society handles rumours as raw data to be validated, current online social media have little context of reputation — and how reputation ties to a persistent identity. Platforms now generally count attention signals from every user equally — whether thoughtful and wise, immature, deranged, criminal, fraud or bot. That is a major factor in the loss of social context.

Perhaps the most promising strategy for effective attention agents is to reapply prior lessons of social context to develop equivalent reputation and trust metrics for how peers view both a given item of speech and the speaker more generally. That is how attention works in real life — one listens to those reputed to be worth listening to — by the listener and by other members of the communities the listener participates in. Automating that may seem hopelessly complex, but Google solved the problem of search decades ago using just such a method in its original PageRank algorithm (Brin and Page 1998).[14] Some platforms have gotten some of this by using social graph proximity (degrees of separation), but sadly, the trend has been toward more engagement-driven factors (Heath 2022).

Importantly, reputation-based methods can factor in not only the reputation of a specific content item, and of its creator, but also the reputations of all who reflexively "amplify" (confer weight to) them through their interactions with them. Reputation can also be partitioned to consider reputations within specific communities, in terms of community contexts — knowledge, tastes and values. Users could compose feeds to include a blend of communities, and might shift that blend as they steer into different activity contexts. Such communities may have explicit membership — others may be implicitly derived by clustering. Partitioning reputations by topic domain can also add nuance for cases where a source may be very authoritative in some contexts but not at all in others.

Design issues in reputation systems relate to validating the authority of the raters and their authenticity, to avoid questions of legitimacy and of gaming the system. Here again, a networked, web-like ecosystem offers a way to apply user agency to bestow legitimacy from the bottom up, while also applying subsidiarity to bring in a mix of top-down institutional authorities as appropriate.

Gaming of reputation and trust systems is an ongoing cat and mouse game — literally a "confidence game" — but a variety of established solution strategies exist for managing this as well. Google's PageRank algorithm came under attack from those who would try to falsify signals of authority, but supplementary newer tools such as TrustRank (Gyöngyi, Garcia-Molina and Pedersen 2004) were applied to use known authorities as

14 This author may have been first to propose extending similar strategies to social media (Reisman 2004, 2012, 2018a) – which has since been advocated by Jeff Allen, while at Facebook (Allen 2019), and later at the Integrity Institute (Allen 2022).

reference points for assessing unknown sources. Other tools to assess reputation and trust use additional methods (some specifically intended for social media), such as a "persona" clustering method (Cai, Wang and Gong 2020); the CoRank or "Coupled Dual Networks Trust Ranking" method, which assesses behaviour patterns (Li et al. 2020); Click-Gap, an algorithm Facebook developed to weed out content from sources identified as low-quality (Rodriguez 2019) and Topical TrustRank, which uses topical information to demote spam sites or pages (Wu 2006).

Reputation works as a repeated game, in which trust can be slow to build but rapidly lost. When that game is designed well, it motivates care in maintaining one's reputation. Because of that, tracking reputation requires a persistent identity — but here again, context has been collapsed. Identity is not canonical but pragmatic. Proving one is tied to a reputation is essential, but tying a legal name to that reputation is rarely essential. Details of reputations can be earned or validated. They can apply to a legal name or a human alias, or even a bot. New users might start with no reputation, and anonymous users build no reputation. Aliases might build reputation within recognized confines, and bots might as well. Governance should support those diverse types of identity and reputation.

Looking ahead, might advances in generative AI and hopes for artificial general intelligence, as well as extensions of social media into the "metaverse" of virtual and augmented reality, weaken the arguments in this brief? No; indeed, the reverse is true: the same principles of drawing on user agency and our social mediation ecosystem should be central to governance of these technologies as well (see Box 2).

A Synthesis: Three Pillars of Thought as a Social Process

In considering the key threads of this brief, three elements this author has been advocating for many years (Reisman 2021a) have an importantly synergistic effect. Each has faced concerns that have limited their uptake (Reisman 2021c), but importantly, the best way to counter those concerns is to build on *how these pillars work in combination* — so that they reinforce one another and serve as a foundation for a full suite of complementary remedies.

This represents a significant broadening of common thinking about "middleware" services as intermediaries between users and platforms. That combined strategy becomes far more powerful and central to civil discourse, and counters concerns that have hindered their acceptance as a way to preserve democracy in the online era.

Specifically, as suggested in Figure 6, middleware can support three essential pillars of discourse that synergize with each other to restore the human context that currently dominant online platforms have collapsed:

- **Individual agency** (the current focus of middleware). This agency creates speaker/listener context. The idea is to return power to users to steer our online "bicycles for our minds" for ourselves, through agency over our feeds and recommenders. This is the most obvious pillar of freedom of thought, and one that most platforms conspicuously co-opt.
- **A social mediation ecosystem** (currently seen as fragmentary, apart from and even conflicting with individual agency — see Box 1). This often informal, loosely structured, web-like network of communities, institutions and informal social groups mediates context collectively — by cooperatively applying collective intelligence, wisdom, judgment and values that reflect back on their members. This is a less obvious, but equally critical pillar of freedom of thought (drawing on freedom of association and assembly). Current social media systems have largely failed to integrate and support the traditional roles of more or less organized social groups in creating "social trust" (Laufer and Nissenbaum 2023). This has limited the ability of social media to leverage our social associations to promote "bridging" of the divides that online media now seems to highlight and reinforce (Ovadya and Thorburn 2023).
- **Reputation and trust** (now generally considered only in primitive form online, primarily to exclude bad actors). These are aspects of *relationship strength* that underpin the interplay of individual agency and social mediation

Box 2: What about Emerging Technologies — AI, the Metaverse and Neural Interfaces?

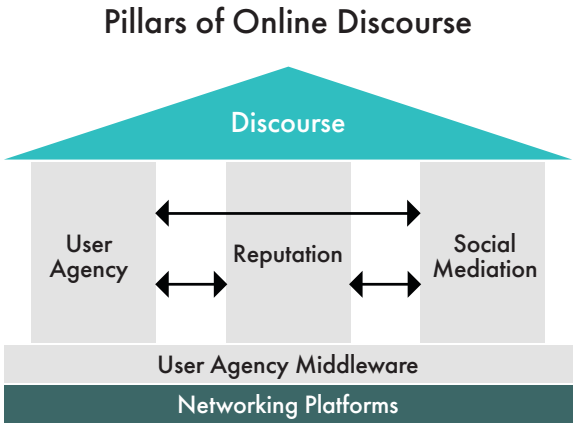
Looking ahead to radically new technologies, the principles of human individual agency and its extension into our social mediation ecosystem should continue to be central to governance.

- AI will certainly change discourse, as a new kind of mediation — and as a tool — but it should leverage and broaden the role of middleware and agency, not eliminate it. There is a symmetrical relationship between AI and social media:
  - AI will be increasingly used within social media services to improve attention agent algorithms and support and apply human mediation activities.
  - AI will itself become "social" media as data corpuses become increasingly dynamic, real-time and reflexive.
- In both social media and AI, it will be increasingly important to apply both artificial and human intelligence to distinguish authentic and trustworthy human content from inauthentic and untrustworthy content — both human and artificial. Normatively, only humans can be trusted to oversee that, and we, the users of these technological tools, must retain agency over them. That, again, requires all three "pillars of discourse" (see Figure 6), as enabled through middleware.
- A direct extension of the ideas in this brief to AI governance is to develop "personal AI agents" that serve individuals through a kind of middleware interoperation in their dealings with institutional AI agents (including those of social media platforms). This can level the power asymmetry between individuals (and their groups) and institutional AIs: "Have your AI agent talk to my AI agent" (for more, see Whitt [2024] and Friston et al. [2024]).
- The "metaverse" of virtual, augmented and extended reality (VR/AR/XR) technologies will add new challenges as more immersive, intimate modalities of social media. As Brittan Heller (2020) illuminates, the same threats to freedoms of association and expression, and the same principles for creating trust through individual and social agency, will apply and gain urgency.
- Also affecting freedom of thought is the broad category of neural interfaces and sensors, still-emerging threats of *direct* interference *inside our heads* (in contrast to the indirect forms of manipulation presented by social media). The principles of individual agency and social trust outlined in this brief will be essential to governance of both indirect and direct threats to our freedom of thought.

with regard to how humans evaluate speaker/mediator context and trustworthiness. This evaluation is done both individually and collectively to allocate our attention to what ideas, individuals and groups to attend to, or to shun — thus motivating a desire to maintain a good reputation for being trustworthy. Current platform attention agent services give little consideration to this self-motivating principle of trustworthiness in upranking and downranking items into our feeds and recommendations.

Most current discussion of middleware focuses on the first of these pillars (important as it is), thus understating its true potential and raising concerns that synergies with the other pillars might reduce. In a recent working outline, this author (Reisman 2023c) proposes a broader vision in which these three pillars reinforce one another, through middleware, to restore the lost matrix of context, thus making the nuance of online discourse clearer to both humans and the algorithms that assist them.

Figure 6: Middleware as the Foundation for Three Pillars of Online Discourse



Source: Reisman (2023c). Copyright Richard Reisman; reproduced with permission.

The fundamental synergy here is the dialectic of a flexibly optimized blend of human freedom, gently balanced by a degree of social nudging toward responsibility. Underlying that synergy is the collective wisdom that humans embed in reputation. From this perspective, middleware is the technology that supports this traditional human context in the online world of computer-mediated discourse. Think of it as *contextware* — and of the related question of legitimacy: *whose context?*

Conclusion

The digitization of human discourse is an immense undertaking that will radically reshape humanity and the nature of human thought, unfolding over many decades. Online propagation in a global network open to the entire population is socially driven. It is best regulated not by censoring expression out of all flows, but by enabling listener agency to select what flows in to them, as freedom of impression. That freedom can be exercised through delegation — to obtain ranking guidance from chosen elements of the social mediation ecosystem with consideration of reputation and trust. This is a digital augmentation of the same social dynamic that enabled open democratic societies to flourish and develop over centuries with minimal government restriction of freedom of expression, drawing on explicitly protected rights of association and assembly.

Shifting agency toward users enables not only individual freedom but also restoration and augmentation of the social mediation ecosystem for driving toward the shared reality and common welfare that are now being lost. The return of agency to citizens, and to their mediating communities, can restore the essential role of the social mediation ecosystem that the platforms have disintermediated. By restoring agency to thinkers as listeners, not just as speakers — and by integrating the sense-making support role of mediating organizations — our marketplace of ideas can be largely self-regulated by citizens and the communities they choose to be participants in. The marketplace of ideas maintains order through a marketplace of mediators.

Regardless of how quickly this can be accomplished, the task of governance is to help achieve balance and legitimacy from subsidiarity based on the delegation of listener agency:

- Long-term objectives should focus on, first, listener agency over individual feeds and recommendations, and second, enablement and support of social mediation ecosystem development to guide that agency and nudge it toward agonism (bridging) over antagonism (affective polarization).
- Speaker-side removals and "platform law" should be used sparingly, as stopgaps — to be relaxed as social mediation and listener controls are restored, as the democratic way to modulate virality.
- The social process of discourse should be lightly guided, with nudging to limit affective polarization and support bridging, while keeping it open, fair and innovative.

Achieving these changes will not be quick or easy, but turning back toward listener agency, and the web-like social ecosystem it opts in to, can quickly begin to apply human collective intelligence to limit current dysfunctions. Individual platforms lack the openness, agility and diversity to do that — it takes an open ecosystem — and now, in the digital era, that ecosystem must be technically enabled. That can enable an ongoing whole-of-society effort to build toward a future collaborative and generative public sphere that fully supports the healthy give and take of human freedom of thought.

Acknowledgements

My thanks to those who have provided stimulating feedback on these ideas, including Chris Riley, Luke Thorburn, Aviv Ovadya, Daphne Keller, Francis Fukuyama, Zoelle Egner, Ethan Zuckerman, Chand Rajendra-Nicolucci, Zach Graves, Luke Hogg, Harold Feld, Pri Bengani, Gabe Nicholas, Renée DiResta, Justin Hendrix, Richard Whitt, Ellen Goodman. And thanks to CIGI for their valuable support in structuring and editing this, including Aaron Shull, Lynn Schellenberg.

Works Cited

Allen, Jeff. 2019. "How Communities Are Exploited On Our Platforms: A Final Look at the 'Troll Farm' Pages." October 4. Contributed by *MIT Technology Review*. www.documentcloud.org/documents/21063547-oct-2019-facebook-troll-farms-report.

———. 2022. "The Integrity Institute's Analysis of Facebook's Widely Viewed Content Report (Q4 2021–Q1 2022)." Integrity Institute, June 15. https://integrityinstitute.org/widely-viewed-content-analysis-tracking-dashboard.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, et al. 2018. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences* 115 (37): 9216–21. https://doi.org/10.1073/pnas.1804840115.

boyd, danah. 2013. "How 'context collapse' was coined: my recollection." *Apophenia* (blog), December 8. www.zephoria.org/thoughts/archives/2013/12/08/coining-context-collapse.html.

Brennan-Marquez, Kiel and Daniel Susser. 2022. "Privacy, Autonomy, and the Dissolution of Markets." Knight First Amendment Institute, August 11. https://knightcolumbia.org/content/privacy-autonomy-and-the-dissolution-of-markets.

Brin, Sergey and Lawrence Page. 1998. "The anatomy of a large-scale hypertextual Web search engine." *Computer Networks and ISDN Systems* 30: 107–17. https://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf.

Cai, Mei, Yiming Wang and Zaiwu Gong. 2020. "An Extension of Social Network Group Decision-Making Based on TrustRank and Personas." *International Journal of Computational Intelligence Systems* 13 (1): 332–40. https://doi.org/10.2991/ijcis.d.200310.001.

Canales, Katie. 2020. "Amazon CEO Jeff Bezos said social media is a 'nuance-destruction' machine when asked about his views on 'cancel culture.'" *Business Insider*, July 29. www.businessinsider.com/jeff-bezos-amazon-nuance-destruction-machine-social-media-2020-7.

Chou, Tracy. 2023. "Coming to terms with the messy spectrum of online speech." Block Party, January 10. www.blockpartyapp.com/blog/coming-to-terms-with-the-messy-spectrum-of-online-speech/.

Couldry, Nick. 2024. "Resonance, Not Scalability." Harvard Kennedy School Ash Center for Democratic Governance and Innovation, February 2. https://ash.harvard.edu/publications/resonance-not-scalability.

Couldry, Nick and Andreas Hepp. 2017. *The Mediated Construction of Reality*. Cambridge, UK: Polity Press.

DiResta, Renee. 2018. "Free Speech Is Not the Same as Free Reach." *Wired*, August 30. www.wired.com/story/free-speech-is-not-the-same-as-free-reach/.

———. 2019. "Mediating Consent." *Ribbonfarm* (blog), December 17. www.ribbonfarm.com/2019/12/17/mediating-consent/.

DiResta, Renee and Tobias Rose-Stockwell. 2021. "How to Stop Misinformation Before It Gets Shared." *Wired*, March 26. www.wired.com/story/how-to-stop-misinformation-before-it-gets-shared/.

Doctorow, Cory. 2023. "The 'Enshittification' of TikTok." *Wired*, January 23. www.wired.com/story/tiktok-platforms-cory-doctorow/.

Dorsey, Jack. 2021. "Written Testimony of Twitter CEO Jack Dorsey (@Jack) U.S. House Committee on Energy & Commerce." March 25. www.congress.gov/117/meeting/house/111407/witnesses/HHRG-117-IF16-Wstate-DorseyJ-20210325.pdf.

Douek, Evelyn. 2022. "Content Moderation as Systems Thinking." *Harvard Law Review* 136 (2): 526–607. https://harvardlawreview.org/print/vol-136/content-moderation-as-systems-thinking/.

Eckles, Dean. 2021. "Algorithmic transparency and assessing effects of algorithmic ranking. Testimony before the Senate Subcommittee on Communications, Media, and Broadband. Hearing of December 9th, 2021." Washington, DC: US Senate Committee on Commerce, Science, and Transportation. www.commerce.senate.gov/services/files/62102355-DC26-4909-BF90-8FB068145F18.

Engelbart, Douglas C. 1962. *Augmenting Human Intellect: A Conceptual Framework*. SRI Summary Report AFOSR-3223. Menlo Park, CA: Stanford Research Institute. www.dougengelbart.org/pubs/augment-3906.html.

Friston, Karl J., Maxwell J. D. Ramstead, Alex B. Kiefer, Alexander Tschantz, Christopher L. Buckley, Mahault Albarracin, Riddhi J. Pitliya et al. 2024. "Designing ecosystems of intelligence from first principles." *Collective Intelligence* 3 (1). https://doi.org/10.1177/26339137231222481.

Fukuyama, Francis. 2022. *Liberalism and Its Discontents*. New York, NY: Farrar, Straus and Giroux.

Fukuyama, Francis, Barak Richman, Ashish Goel, Roberta R. Katz, Douglas Melamed and Marietje Schaake. 2020. *Report of the Working Group on Platform Scale*. White Paper, November 17. Stanford, CA: Stanford Cyber Policy Center. https://cyber.fsi.stanford.edu/publication/report-working-group-platform-scale.

Gillespie, Tarleton. 2020. "Content moderation, AI, and the question of scale." *Big Data & Society* 7 (2): July–December. https://doi.org/10.1177/2053951720943234.

———. 2022. "Do Not Recommend? Reduction as a Form of Content Moderation." Social Media & Society 8 (3). https://doi.org/10.1177/20563051221117552.

Goldman, Eric. 2021. "Content Moderation Remedies." *Michigan Technology Law Review* 28 (1): 1–60. https://doi.org/10.36645/mtlr.28.1.content.

Goodman, Ellen P. 2021. "Digital Fidelity and Friction." *Nevada Law Journal* 21 (2): 623-53. https://scholars.law.unlv.edu/nlj/vol21/iss2/6.

Goodman, Ellen P., Derek Slater and Marc Hand. 2022. "Imagining a New Public Sphere: Public Media and a New Resilient Future with Public Broadcasting." Tech Policy Press, July 29. www.techpolicy.press/imagining-a-new-public-sphere-public-media-and-a-new-resilient-future-with-public-broadcasting/.

Graber, Jay. 2023a. "Algorithmic choice." *Bluesky* (blog), March 30. https://blueskyweb.xyz/blog/3-30-2023-algorithmic-choice.

———. 2023b. "Composable Moderation." *Bluesky* (blog), April 13. https://blueskyweb.xyz/blog/4-13-2023-moderation.

———. 2023c. "Federation Architecture Overview." *Bluesky* (blog), May 5. https://blueskyweb.xyz/blog/5-5-2023-federation-architecture.

———. 2023d. "Moderation in a Public Commons." *Bluesky* (blog), June 23. https://blueskyweb.xyz/blog/6-23-2023-moderation-proposals.

Gyöngyi, Zoltán, Hector Garcia-Molina and Jan O. Pedersen. 2004. "Combating Web Spam with TrustRank." In *Proceedings of the Thirtieth International Conference on Very Large Data Bases*, 576–87, edited by Mario A. Nascimento, M. Tamer Özsu, Donald Kossmann, Renée J. Miller, José A. Blakeley and K. Bernhard Schiefer. Boston, MA: Morgan Kaufmann. https://dblp.org/db/conf/vldb/vldb2004.html.

Habermas, Jürgen. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. First published in German in 1962. Translated by Thomas Burger. Cambridge, MA: MIT Press.

Haidt, Jonathan. 2013. *The Righteous Mind*. New York, NY: Vintage Books.

Hasinoff, Amy A. and Nathan Schneider. 2022. "From Scalability to Subsidiarity in Addressing Online Harm." *Social Media + Society*, July-September. https://doi.org/10.1177/20563051221126041.

Heath, Alex. 2022. "Facebook is changing its algorithm to take on TikTok, leaked memo reveals." *The Verge*, June 15. www.theverge.com/2022/6/15/23168887/facebook-discovery-engine-redesign-tiktok.

Heller, Brittan. 2020. "Watching Androids Dream of Electric Sheep: Immersive Technology, Biometric Psychography, and the Law." *Vanderbilt Journal of Entertainment and Technology Law* 23 (1): 1–52. https://scholarship.law.vanderbilt.edu/jetlaw/vol23/iss1/1/.

Hendrix, Justin. 2021. "Reconciling Social Media & Democracy: Tracy Chou and Mike Masnick." Tech Policy Press, October 16. www.techpolicy.press/reconciling-social-media-democracy-tracy-chou-and-mike-masnick/.

———. 2022. "Evaluating Alternatives to Content Removal on Social Media Platforms." Tech Policy Press, April 13. www.techpolicy.press/evaluating-alternatives-to-content-removal-on-social-media-platforms/.

Hogg, Luke and Lars Erik Schönander. 2023. "Meta's Interoperable Reversal." Tech Policy Press, July 19. www.techpolicy.press/metas-interoperable-reversal/.

Hutson, Matthew. 2020. "Why you don't really know what you know." *MIT Technology Review*, October 21. www.technologyreview.com/2020/10/21/1009445/the-unbearable-vicariousness-of-knowledge/.

Johnson, Charley. 2023. "How the Netflix show 'Ultimatum' explains complex algorithmic systems." Untangled with Charley Johnson, June 11. https://untangled.substack.com/p/how-the-netflix-show-ultimatum-explains.

Keller, Daphne. 2021. "Amplification and Its Discontents." Knight First Amendment Institute, June 8. https://knightcolumbia.org/content/amplification-and-its-discontents.

———. 2022. "Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users." *University of Chicago Law Review Online Archive* (blog), June 28. https://lawreviewblog.uchicago.edu/2022/06/28/keller-control-over-speech/.

Keller, Daphne and Francis Fukuyama. 2024. "Don't Let Texas and Florida Chew Up the Internet." *The Wall Street Journal*, February 29. www.wsj.com/us-news/law/dont-let-texas-and-florida-chew-up-the-internet-1080a5af.

Kissinger, Henry, Eric Schmidt and Daniel Huttenlocher. 2023. "ChatGPT Heralds an Intellectual Revolution." *Wall Street Journal*, February 24. www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6.

Land, Molly K. 2020. "The Problem of Platform Law: Pluralistic Legal Ordering on Social Media." In *The Oxford Handbook of Global Legal Pluralism*, edited by Paul Schiff Berman, 974–94. New York, NY: Oxford University Press. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3454222.

Lane, Spencer. 2022. "Gobo 2.0: All Your Social Media in One Place." Initiative for Digital Public Infrastructure at UMass Amherst, November 9. https://publicinfrastructure.org/2022/11/09/gobo-2-0-all-your-social-media-in-one-place/.

Laufer, Benjamin and Helen Nissenbaum. 2023. "Algorithmic Displacement of Social Trust." Knight First Amendment Institute, November 29. https://knightcolumbia.org/content/algorithmic-displacement-of-social-trust.

Li, Peiyao, Weiliang Zhao, Jian Yang, Quan Z. Sheng and Jia Wu. 2020. "Let's CoRank: trust of users and tweets on social networks." *World Wide Web* 23: 2877–901. https://doi.org/10.1007/s11280-020-00829-4.

Masnick, Mike. 2015. "For 10 Years Everyone's Been Using 'The Streisand Effect' Without Paying; Now I'm Going To Start Issuing Takedowns." *Techdirt* (blog), January 8. www.techdirt.com/2015/01/08/10-years-everyones-been-using-streisand-effect-without-paying-now-im-going-to-start-issuing-takedowns/.

———. 2019. "Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well." *Techdirt* (blog), November 20. www.techdirt.com/2019/11/20/masnicks-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well/.

———. 2024. "Why Bluesky Remains The Most Interesting Experiment In Social Media, By Far." *Techdirt* (blog), March 27. www.techdirt.com/2024/03/27/why-bluesky-remains-the-most-interesting-experiment-in-social-media-by-far/.

Matias, J. Nathan. 2023. "Humans and algorithms work together — so study them together." *Nature*, May 10. www.nature.com/articles/d41586-023-01521-z.

Michael Lawrence Films. 2006. "Steve Jobs, 'Computers are like a bicycle for our minds.'" June 1. YouTube video, 1:39. www.youtube.com/watch?v=ob_GX50Za6c.

Narayanan, Arvind. 2023. "Understanding Social Media Recommendation Algorithms." Knight First Amendment Institute, March 9. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

Nelson, Theodore H. 1974. *Computer Lib/Dream Machines*. https://computerlibbook.com/collections/all.

New_ Public. 2024. *Dynamic Polycentricism: A Way to Measure Pluralism in Online Communities without Excessive Surveillance*. New York, NY: New_ Public. https://newpublic.org/uploads/2024/02/Dynamic-Polycentrism.pdf.

Newitz, Annalee. 2023. "Ben Franklin Would Have Loved Bluesky." *The Atlantic*, May 12. www.theatlantic.com/technology/archive/2023/05/social-media-centralism-fediverse/674041/.

Ovadya, Aviv. 2021. "Towards Platform Democracy: Policymaking Beyond Corporate CEOs and Partisan Pressure." Harvard Kennedy School Belfer Center Paper, October 18. www.belfercenter.org/publication/towards-platform-democracy-policymaking-beyond-corporate-ceos-and-partisan-pressure.

Ovadya, Aviv and Luke Thorburn. 2023. "Bridging Systems: Open problems for countering destructive divisiveness across ranking, recommenders, and governance." Knight First Amendment Institute, October 26. https://knightcolumbia.org/content/bridging-systems.

Peters, Jay. 2023. "Adam Mosseri spells out Threads' plans for the fediverse." The Verge, December 15. www.theverge.com/2023/12/15/24003435/adam-mosseri-threads-fediverse-plans.

Rajendra-Nicolucci, Chand and Michael Sugarman. 2023. "The Three-Legged Stool: A Manifesto for a Smaller, Denser Internet." Initiative for Digital Public Infrastructure at UMass Amherst, March 29. Amherst, MA: University of Massachusetts Amherst. https://publicinfrastructure.org/2023/03/29/the-three-legged-stool/.

Reisman, Richard. 2004. "Method and apparatus for an idea adoption marketplace." United States Patent Application Pub. No. US 2004/0186738Al September 23. www.teleshuttle.com/ftp/Reisman%20Collab-Rep-Social%20Media%20Patent%20App%20(public%20domain).pdf.

———. 2012. "Filtering for Serendipity – Extremism, 'Filter Bubbles' and 'Surprising Validators.'" *Smartly Intertwingled* (blog), October 8. https://ucm.teleshuttle.com/2012/10/filtering-for-serendipity-extremism.html.

———. 2018a. "The Augmented Wisdom of Crowds: Rate the Raters and Weight the Ratings." *Smartly Intertwingled* (blog), July 22. https://ucm.teleshuttle.com/2018/07/the-augmented-wisdom-of-crowds-rate.html.

———. 2018b. "Reverse the Biz Model! — Undo the Faustian Bargain for Ads and Data." *The FairPay Zone* (blog), December 3. www.fairpayzone.com/2018/12/reverse-biz-model-undo-faustian-bargain.html.

———. 2019. "Regulating our Platforms — A Deeper Vision." *The FairPay Zone* (blog), December 30. https://ucm.teleshuttle.com/2019/12/regulating-our-platforms-deeper-vision.html.

———. 2021a. "The Internet Beyond Social Media Thought-Robber Barons." Tech Policy Press, April 22. www.techpolicy.press/the-internet-beyond-social-media-thought-robber-barons/.

———. 2021b. "Scholars Reckon with Democracy and Social Media." Tech Policy Press, August 9. www.techpolicy.press/scholars-reckon-with-democracy-and-social-media/.

———. 2021c. "Unbundling Social Media: A Taxonomy of Problem Areas." Tech Policy Press, August 9. www.techpolicy.press/unbundling-social-media-a-taxonomy-of-problem-areas/.

———. 2021d. "The Roots of My Thinking on Tech Policy." *Smartly Intertwingled* (blog), September 24. https://ucm.teleshuttle.com/p/tech-policy.html.

———. 2022. "Into the Plativerse...Through Fiddleware?" Tech Policy Press, December 20. www.techpolicy.press/into-the-plativerse-through-fiddleware/.

———. 2023a. "From Freedom of Speech and Reach to Freedom of Expression and Impression." Tech Policy Press, February 14. www.techpolicy.press/from-freedom-of-speech-and-reach-to-freedom-of-expression-and-impression/.

———. 2023b. "Context Lost; Context Regained – Comments on iDPI's 'A Manifesto for Smaller, Denser Internet.'" *Smartly Intertwingled* (blog), April 5. https://ucm.teleshuttle.com/2023/04/context-lost-context-regained-comments.html.

———. 2023c. "A New, Broader, More Fundamental Case for Social Media Agent 'Middleware.'" *Smartly Intertwingled* (blog), November 9. https://ucm.teleshuttle.com/2023/11/a-new-broader-more-fundamental-case-for.html.

———. 2024a. "We Can Protect Freedom of Thought by Deciding What We Feed Our Brains." Opinion, Centre for International Governance Innovation, January 17. www.cigionline.org/articles/we-can-protect-freedom-of-thought-by-deciding-what-we-feed-our-brains/.

———. 2024b. "The Supreme Court Can Help Fix Social Media Governance With NetChoice Ruling." Tech Policy Press, February 9. www.techpolicy.press/the-supreme-court-can-help-fix-social-media-governance-with-netchoice-ruling/.

Reisman, Richard and Chris Riley. 2022a. "Delegation, Or, The Twenty Nine Words That The Internet Forgot." Tech Policy Press, February 28. www.techpolicy.press/delegation-or-the-twenty-nine-words-that-the-internet-forgot/.

———. 2022b. "Understanding Social Media: An Increasingly Reflexive Extension of Humanity." Tech Policy Press, April 27. www.techpolicy.press/understanding-social-media-an-increasingly-reflexive-extension-of-humanity/.

———. 2022c. "Community and Content Moderation in the Digital Public Hypersquare." Tech Policy Press, June 17. www.techpolicy.press/community-and-content-moderation-in-the-digital-public-hypersquare/.

———. 2022d. "Contending for Democracy on Social Media and Beyond." Tech Policy Press, September 22. www.techpolicy.press/contending-for-democracy-on-social-media-and-beyond/.

Rodriguez, Salvador. 2019. "Facebook is taking a page out of Google's playbook to stop fake news from going viral." CNBC, April 10. www.cnbc.com/2019/04/10/facebook-click-gap-google-like-approach-to-stop-fake-news-going-viral.html.

Rosenberg, Eli. 2011. "Quote: The Ad Generation." *The Atlantic*, April 15. www.theatlantic.com/technology/archive/2011/04/quote-ad-generation/349689/.

Rose-Stockwell, Tobias. 2023. *Outrage Machine: How Tech Amplifies Discontent, Disrupts Democracy—And What We Can Do About It*. New York, NY: Hachette.

Roth, Yoel and Samantha Lai. 2024. "Securing Federated Platforms: Collective Risks and Responses." *Journal of Online Trust & Safety* 2 (2). https://doi.org/10.54501/jots.v2i2.171.

Rozenshtein, Alan Z. 2023. "Moderating the Fediverse: Content Moderation on Distributed Social Media." University of Minnesota Legal Studies Research Paper No. 23-19. https://dx.doi.org/10.2139/ssrn.4213674.

Siddarth, Divya, Danielle Allen and E. Glen Weyl. 2022. "The Web3 Decentralization Debate Is Focused on the Wrong Question." *Wired*, May 12. www.wired.com/story/web3-blockchain-decentralization-governance/.

Spiro, Emma S. and Kate Starbird. 2023. "Rumors Have Rules." *Issues in Science and Technology* XXXIX (3). https://issues.org/rumors-research-misinformation-spiro-starbird/.

Stray, Jonathan. 2022. "Designing recommender systems to depolarize." *First Monday* 27 (5). https://doi.org/10.5210/fm.v27i5.12604.

Stray, Jonathan, Luke Thorburn and Priyanjana Bengani. 2023a. "From 'Filter Bubbles', 'Echo Chambers', and 'Rabbit Holes' to 'Feedback Loops.'" Tech Policy Press, April 17. www.techpolicy.press/from-filter-bubbles-echo-chambers-and-rabbit-holes-to-feedback-loops/.

———. 2023b. "Making Amplification Measurable." Tech Policy Press, April 28. www.techpolicy.press/making-amplification-measurable/.

Tambini, Damian. 2019. "Rights and Responsibilities of Internet Intermediaries in Europe: The Need for Policy Coordination." In *Models for Platform Governance*, 91–96. Waterloo, ON: CIGI. www.cigionline.org/publications/models-platform-governance.

Task Force for a Trustworthy Future Web. 2023. *Scaling trust on the web*. Digital Forensic Research Lab, June 21. Washington, DC: Atlantic Council. www.atlanticcouncil.org/event/report-launch-scaling-trust-on-the-web/.

The Bluesky Team. 2024a. "Bluesky: An Open Social Web." *Bluesky* (blog), February 22. https://bsky.social/about/blog/02-22-2024-open-social-web.

———. 2024b. "Bluesky's Stackable Approach to Moderation." *Bluesky* (blog), March 12. https://bsky.social/about/blog/03-12-2024-stackable-moderation.

Thiel, David and Renée DiResta. 2023. *Child Safety on Federated Social Media*. July 24. Stanford, CA: Stanford Internet Observatory. https://doi.org/10.25740/vb515nd6874.

Törnberg, Petter. 2022. "How digital media drive affective polarization through partisan sorting." *Proceedings of the National Academy of Sciences* 119 (42): 1–11. https://doi.org/10.1073/pnas.2207159119.

Warner, Mark R. 2022. "Lawmakers Reintroduce Bipartisan Legislation to Encourage Competition in Social Media." Press release, May 25. www.warner.senate.gov/public/index.cfm/2022/5/lawmakers-reintroduce-bipartisan-legislation-to-encourage-competition-in-social-media.

Whitt, Richard S. 2020. "Old School Goes Online: Exploring Fiduciary Obligations of Loyalty and Care in the Digital Platforms Era." *Santa Clara High Technology Law Journal* 36 (1): 75–131. https://digitalcommons.law.scu.edu/chtlj/vol36/iss1/3/.

———. 2021. "Hacking the SEAMs: Elevating Digital Autonomy and Agency for Humans." *Colorado Technology Law Journal* 19 (1): 135–212. https://drive.google.com/file/d/118IAxowzYIq1tH_BfA0y2yKY3z2ydCMu/view.

———. 2024. "Aligning Two Dimensions of AI Agency." Google Slides deck. GLIA Foundation, February. www.glia.net/.

Wilhelm, Elisabeth. 2023. "A New Science for Describing Unhealthy Online Environments." *Physics*, June 5. https://physics.aps.org/articles/v16/89.

Wu, Baoning, Vinay Goel and Brian D. Davison. 2006. "Topical TrustRank: using topicality to combat web spam." In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, edited by Leslie Carr, David De Roure, Arun Iyengar, Carole Goble and Mike Dahlin, 63–72. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1135777.1135792.

Zuckerman, Ethan and Chand Rajendra-Nicolucci. 2023. "From Community Governance to Customer Service and Back Again: Re-Examining Pre-Web Models of Online Governance to Address Platforms' Crisis of Legitimacy." *Social Media + Society* 9 (3). https://doi.org/10.1177/20563051231196864.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Managing Director and General Counsel Aaron Shull
CIGI Senior Fellow and Project Co-Leader Susie Alegre
Director, Program Management Dianna English
Program Manager Jenny Thiel
Publications Editor Lynn Schellenberg
Publications Editor Susan Bubak
Graphic Designer Abhilasha Dewan

This policy brief was made possible thanks to the financial support of the Konrad-Adenauer-Stiftung (KAS) Canada.

For publications enquiries, please contact publications@cigionline.org.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org