

Agentic AI: User Empowerment or Enclosure?

David Gamba, Daniel M. Romero, Grant Schoenebeck

University of Michigan
Ann Arbor, MI, USA

gamba@umich.edu, drom@umich.edu, schoeneb@umich.edu

Abstract

Agentic AI promises systems that can act on users' behalf, from filtering content to negotiating prices to selecting services. Whether it will empower users is an open question, and one that depends on more than the technology. We conduct a comparative case analysis of four earlier, more mature domains in which similar forms of agency emerged: browser-based ad blockers, platform recommender systems, financial robo-advisors, and email spam filtering. Across the cases, questions about whose interests agents would serve were resolved through technical arrangements: API choices, protocol governance, industry standards, and default configurations. Beyond their technical form, these were political decisions. We identify this settling of contestable questions in a technical form as depoliticization, a concept from political theory, here at work in technological systems. Its most consequential effect is that individual outcomes and collective contestation capacity can move in opposite directions: spam inbox quality improved substantially while the organized capacity to contest spam governance collapsed. Where intermediary institutions sustained formal channels for challenge, user-aligned agency proved more durable; where proprietary infrastructure and closed standard-setting absorbed contestation, the material basis for user-aligned alternatives was dismantled, and the loss proved hard to reverse. Applying this lens to agentic AI, we find a similar pattern forming: governance is consolidating around the Model Context Protocol and the Agentic AI Foundation, an industry-governed venue already deciding what agents will be able to do. Unlike in the completed trajectories, these decisions have not yet hardened, and remain open to challenge by users and the public.

Full version — <https://doi.org/10.48550/arXiv.2608.06510>

1 Introduction

In 2019, Google announced a redesign of the extension API for its Chrome browser, later shipped as Manifest V3. No ad blocker was banned. No filtering rule was invalidated. Instead, the operations available to browser extensions were redefined: the flexible `webRequest` API, which allowed extensions to intercept and modify network requests in real time, was replaced with the more constrained `declarativeNetRequest` API, which requires extensions to declare filtering rules in advance and

limits how many rules can be active simultaneously. The practical effect was to substantially reduce the capability of content-blocking extensions (Miagkov and Cyphers 2021). A contestable question (what should users be able to block in their browsers, and who decides) was settled by being embedded in browser architecture. What makes this a political act is that it was a choice among alternatives, made by an actor with interests in the outcome.

Agentic AI is a new technology that, like ad blockers, promises to act on users' behalf, though in more powerful ways. Could an ordinary user, for example, direct an agent to build and maintain their own Chromium-based browser that restores the blocking capability Chrome removed? Or, less ambitiously, could they build browser extensions that reorder or filter a social media feed or search results according to their personal instructions? Of course, the platforms may attempt to block such behavior. In the ensuing cat-and-mouse game, who wins?

To better understand the potential outcomes of this dynamic, we look to predecessors of agentic AI, systems that also promised to act in the interests of the user: ad blockers, recommender systems, robo-advisors, and spam filters. Each initially opened space for user-aligned action, and each has attracted a substantial critical literature. We perform a comparative analysis of these trajectories in detail to trace how that space shifted over time.

The focus of our study is *user empowerment*, by which we mean the capacity to meaningfully influence how an agent acts in ways that align with the user's own interests. That capacity is exercised individually but preserved, or lost, collectively. Our focus is on what kinds of agents are even possible in a domain: which optimization targets, data sources, and actions on behalf of users are available to any agent, regardless of which firm builds it. We call this the *boundaries of possible agency*. Whether an agent may build an ad-blocking browser, and whether an extension may reorder a feed, are questions about where those boundaries sit.

Comparing the trajectories reveals a structural tendency: *depoliticization*, the progressive removal of contestable choices from open challenge by embedding them in technical and institutional forms (Burnham 2001; Flinders and Buller 2006). These trajectories admit competing explanations, and for particular moments strong ones are available. The turn in spam filtering reads naturally as an arms race

won by better classifiers, and Chrome’s extension redesign was defended on security and performance grounds; market-selection and regulatory-capture readings fit other cases. We address these accounts where they bear and argue that none, though each may explain a moment, suffices to explain what the comparison reveals. Across four domains and more than two decades, the cases differ in whether organized collective contestation ever formed and, where it did, whether it survived.

Building on the depoliticization tendency the cases reveal, we advance three contributions. First, using historical trajectories of user-facing technologies, we show that improvements in individual user experience and collective contestation capacity can move in opposite directions, and we identify what drives the divergence.

Second, we develop a constitutive politics framework that locates where user-aligned agency is foreclosed or sustained along three dimensions: infrastructural (what agents run on), epistemic (what they can know), and teleological (what they optimize for). We argue that if user empowerment is to be durable, it must rest on the collective capacity to contest the choices that configure these dimensions. Across the cases, that capacity was sustained where intermediary institutions provided adversarial governance, i.e., formal channels through which their decisions can be challenged (the SEC’s rulemaking), and independent infrastructure existed (Firefox’s extension API), and collapsed where proprietary infrastructure (Chrome’s Manifest V3) and gated standard-setting (the email-deliverability forum M3AAWG) foreclosed participation. When foreclosure accumulates across all three dimensions at once, the footholds from which contestation could be rebuilt are removed together, and rebuilding them is far harder than defending them would have been.

Third, we apply the framework to agentic AI. It is now consolidating across a far broader set of domains than any of these predecessors, including commerce, information retrieval, and task automation, through a single protocol infrastructure: the Model Context Protocol. Anthropic released the protocol in late 2024, and major providers adopted it within twelve months (Linux Foundation 2025). Anthropic has since donated it to the Agentic AI Foundation at the Linux Foundation. There, a Governing Board composed of platform providers carries decision-making authority. User organizations and public-interest representatives have no formal standing. This configuration matches the foreclosure-prone patterns the cases identify. But the specifications are young, and their choices are not yet locked into deployed infrastructure or user habit. We set out what preserving contestation would require while the architecture remains malleable.

2 Constitutive Politics: A Lens

To function as a computational system, any “agent” of the kind examined in this paper requires three things: infrastructure to run on (compute, APIs, protocols), data from which to reason, and an objective that defines what it is optimizing for (Russell 2010; Wooldridge and Jennings 1995). This decomposition holds whether the system is a 1990s spam

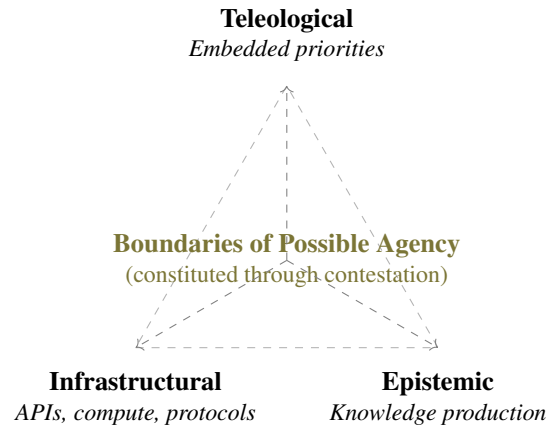


Figure 1: Dimensions of constitutive politics that define the boundaries of agency. Struggles over these dimensions constitute what agents in a domain can be and whose interests they can serve.

filter operating on fixed rules or a contemporary machine learning recommender. Each element is technically necessary; each is also a site of contestation (fig. 1). Drawing on Winner’s analysis of how artifacts embody political properties and Mouffe’s concept of the political as the domain of open contestation, which persists only where institutions grant opposed parties standing as adversaries (Winner 1980; Birch 2025; Mouffe 2005, 2011), we develop a *constitutive politics* lens organized around these three sites. The lens asks what agents in a domain can be, what they can know, and whose interests they can serve. It then asks where those possibilities are set, by whom, and whether they could be otherwise.

Other lenses could of course be used. Principal-agent theory, AI alignment, and participatory design each offer tools for improving agent behavior given the conditions under which agents operate (Duetting, Feldman, and Talgam-Cohen 2024; Sorensen et al. 2024; Muller and Kuhn 1993). The constitutive politics lens asks how those conditions came to be configured. At each turning point the cases examine, multiple configurations were technically feasible; the question is what forces selected the one that was adopted. The comparison across the trajectories answers this. Each of the three dimensions indexes where those choices were made: in infrastructure, in what the system can know, in what it optimizes for, respectively.

Infrastructural. The compute, APIs, protocols, browser engines, and data pipelines on which agents operate, along with the legal access rights that govern them. Control at any layer of this stack shapes what is possible at the layers above it, which is why a browser API redesign can redraw the boundaries for every blocking agent in a domain without targeting any one of them directly.

Epistemic. How agents produce knowledge about users and the world, and how that knowledge is represented. A community-maintained filter list and a proprietary behavioral classifier are both ways of knowing, yet they encode

different answers to who can challenge the judgments embedded in the system and on what basis.

Teleological. The priorities embedded in agents' concrete configurations, including optimization targets, default settings, fee structures, and legal standards. For instance, when a platform encodes a particular metric as its recommendation objective, that is a political choice embedded in a loss function.

The cases that follow are histories of how these three dimensions were defined and contested over time, organized around the actors involved, their incentives, and the turning points that shaped each trajectory.

3 Technological Trajectories

The technologies examined here were each introduced at different times and in different domains as tools to act on behalf of users for specific tasks. Each has operated long enough to be studied extensively, and a substantial literature has documented their development. We selected the trajectories to span variation in their development: from sustained community governance to complete foreclosure, with intermediate cases of collapse and regulatory absorption. We present them as historical narratives organized around the actors involved and the major turning points that shaped each trajectory. How these turning points should be interpreted and whether common patterns run across them are questions we take up in subsequent sections.

Browser-Based Ad Blockers

In the mid-2000s, browser extensions that intercept and filter advertising content emerged as a niche practice among technically adept users. By the 2010s, they had become one of the most widely deployed user-side tools on the web. At their core, ad blockers combine two mechanisms: network-level request blocking, in which outgoing HTTP requests to known advertising domains are canceled before loading, and cosmetic filtering, in which CSS rules hide ad-slot elements that survive request blocking (Storey et al. 2017; Nithyanand et al. 2016). The technical details matter because both mechanisms depend on what browser extension APIs permit, and what they permit has itself become contested.

The early ad-blocking ecosystem was organized around shared filter lists. EasyList, maintained by a small group of volunteers and open to crowdsourced contributions, became the de facto standard, encoding rules about which domains, URL patterns, and page elements should be blocked (Alrizah et al. 2019; Snyder, Vastel, and Livshits 2020). The list was publicly readable, forkable, and updated through a visible community process. Disputes about which ads should be blocked (including whether a given analytics domain belonged on a privacy list) were resolved through argument and revision within the community; anyone who disagreed with a decision could fork the list and maintain their own (Snyder, Vastel, and Livshits 2020). One estimate placed Adblock Plus alone at approximately 100 million users by the late 2010s (Megali 2022), and the economic stakes for publishers and advertisers rose with adoption. Web publishing had entered a sustained revenue crisis in the

early 2010s, with advertising the primary source of income for news and content sites, and rising ad-block adoption was experienced by publishers as a direct threat to that revenue base (Megali 2022; Zhao et al. 2020).

The first major institutional development came in 2011, when Eyeo GmbH, the company behind Adblock Plus, introduced the Acceptable Ads program (Gupta and Panda 2020). Rather than blocking all advertising, Acceptable Ads defined criteria for "non-intrusive" formats (governing size, placement, animation, and labeling) and allowed compliant ads through by default, charging fees to larger advertisers seeking inclusion on the whitelist. Publishers and advertisers who objected to ad blocking had pursued legal challenges in German courts; those courts upheld both the act of blocking and the practice of charging for whitelisting as lawful market behavior, thereby consolidating the Acceptable Ads arrangement as a legally valid settlement of the dispute (Nithyanand et al. 2016; Gupta and Panda 2020). A contested question, whether and by whom advertising could be blocked, was restructured into a different one: which advertising formats met the technical criteria for the whitelist, and at what price.

Not all projects accepted this settlement. uBlock Origin, developed independently by Raymond Hill, rejected the Acceptable Ads framework and maintained community-curated, user-modifiable filtering without a whitelisting program (Alrizah et al. 2019). It offered users direct control over which lists to enable, the ability to add custom rules, and the option to allow or block individual elements on a per-site basis. That choice split the ecosystem: commercial blockers now operate within the Acceptable Ads framework, while projects like uBlock Origin maintain a meaningfully different relationship to the advertising industry.

Alongside the Acceptable Ads negotiation, a more confrontational technical contest developed between blockers and publisher-side countermeasures. Specialized anti-adblock vendors such as PageFair and BlockAdblock developed detection techniques that publishers deployed on roughly 6–7% of high-traffic sites, using bait elements styled like ads, probes for missing ad-library objects, and behavioral network monitoring (Nithyanand et al. 2016; Gupta and Panda 2020). Filter-list maintainers responded with anti-adblock rules and scriptlets targeting known detection patterns; researchers formalized the dynamic as a security-style arms race and proposed perceptual, stealth, and active blocking techniques as next-generation responses (Storey et al. 2017; Iqbal, Shafiq, and Qian 2017; Alrizah et al. 2019). Publishers experimented with both hard walls (requiring whitelisting to access content) and soft cooperative approaches (delivering "acceptable" less-intrusive ads via Acceptable Ads exchanges). Field experiments on Forbes showed that hard walls reduced engagement overall, particularly among less-loyal users, while soft cooperation preserved a broader audience and gradually converted some users to whitelisting (Zhao et al. 2020).

In 2017, the Coalition for Better Ads published standards defining ad formats it deemed unacceptable, based on consumer experience surveys, and delegated enforcement to member platforms, primarily through Chrome's Ad Experience Report (Megali 2022). The Coalition included Google

among its co-founders alongside major advertising trade associations and publishing companies; its membership did not include formal representation from filter-list communities or consumer or public-interest organizations, and its governance structure lacked mechanisms for adversarial challenge or an independent review process (Blackburn, Ritala, and Keränen 2023). Beginning in 2019, Chrome enforced Coalition standards as a built-in browser feature, automatically filtering ads that violated them for all users, regardless of whether they had installed any ad-blocking extension (Megali 2022). What counted as acceptable advertising, previously contested among publishers, advertisers, blocklist communities, and users, now fell largely within Google's discretion, the largest seller of the advertising those standards govern.

The filter lists themselves had accumulated weaknesses that would later support the case for constraining them. A longitudinal analysis of EasyList found that roughly 90% of resource-blocking rules were unused in typical browsing scenarios, contributing to computational overhead that was particularly noticeable on mobile devices. False positives could persist for weeks or months under the list's highly centralized maintenance structure (Snyder, Vastel, and Livshits 2020; Alrizah et al. 2019). Against this background, the most recent significant development has been the transition from Chrome's Manifest V2 extension API to Manifest V3, announced in 2019 and progressively enforced in subsequent years. The older API permitted extensions to use a flexible `webRequest` interface, enabling blocking tools to dynamically inspect and cancel network requests. Manifest V3 replaced this with a `declarativeNetRequest` interface, under which extensions submit rule sets that the browser engine evaluates according to its own logic, with constraints on the number and complexity of rules (Miagkov and Cyphers 2021). The stated rationale invoked security (constraining the broad capabilities the old API granted to all extensions, including malicious ones) and performance (the declarative model permits engine-level optimization that arbitrary script-based filtering does not), both of which have been reproduced and contested in critical analyses (Miagkov and Cyphers 2021; Hancock 2021).

While the API constraints themselves are not in dispute, the extent to which they have reduced practical blocking capability has been debated among extension developers. The effect is less disputed on Firefox, which retained the older API and where uBlock Origin continues to operate with full functionality, and on Brave, a Chromium-derived browser built to apply the same community filter lists directly rather than through an extension (Brave Software 2023). Escape, however, now requires switching browsers. The Coalition for Better Ads case and the Manifest V3 transition represent two distinct kinds of change in the ad-blocking landscape; whether one paved the way for the other, or whether they are better understood as independent developments, is a matter on which analysts have reached different conclusions.

Platform Recommender Systems

Recommender systems organize what users encounter on large digital platforms: which videos appear in a YouTube

feed, which products Amazon surfaces in search results, which posts appear in a social media timeline. Their basic architecture involves retrieving a candidate set from a large corpus and ranking candidates using behavioral signals such as dwell time, completion rate, and shares (Cobbe and Singh 2019; Seaver 2018). Early systems used click-through rate as the primary signal; around 2011–2012 YouTube shifted to watch time and session length, and similar transitions followed on Instagram and TikTok (Xiang 2022; Wang 2022). Two accounts of these shifts coexist in the literature and in platforms' own public statements: a user-quality account, in which watch time better reflects whether viewers found content worth completing, and a revenue account, in which session-length metrics convert more directly to advertising impressions (Cobbe and Singh 2019). These shifts tracked the field's larger transition from collaborative filtering over explicit signals to deep learning over implicit behavioral ones, operationalized in YouTube's two-stage neural pipeline (Covington, Adams, and Sargin 2016). The individual outcome at stake is also broader than relevance. Personalization narrows what each user is shown, and the resulting filter bubbles are not observable from within the interface that produces them (Pariser 2011). The infrastructure and data required to operate and audit these systems are accessible only to the platforms that own them. The EU's Digital Services Act (DSA), adopted in 2022, requires major platforms to document ranking criteria and offer users an alternative not based on behavioral profiling; early compliance audits have found gaps between documented criteria and observed recommendation behavior (Fabbri and Boratto 2025; Meßmer and Degeling 2023). The full case narrative appears in the extended version.

Financial Robo-Advisors

Robo-advisors are digital platforms that provide automated investment advice and portfolio management to retail clients, typically through a web or mobile interface with minimal human adviser involvement (Grealish and Kolm 2021b; Rühr et al. 2019). The basic pipeline is standardized: a client completes a questionnaire about their financial situation, investment goals, time horizon, and self-assessed risk tolerance; an algorithm classifies the client into a risk category; that category maps to a model portfolio implemented through exchange-traded funds (ETFs); and ongoing automation handles rebalancing, contributions, and, on some US platforms, tax-loss harvesting (Grealish and Kolm 2021a,b). Underneath sit modern portfolio theory and long-run expected-utility reasoning, which for any specified risk preference yield a small set of efficient allocations across broad asset classes (Hayes 2021, 2020). The questionnaire that begins the process performs two functions at once: it elicits information from the client and, through its sliders, projections, and default allocations, communicates what kind of investor the client is (Hayes 2021; Tan 2020; Bräuer 2021).

Commercial robo-advisors emerged in the US around 2008–2012 with firms such as Betterment and Wealthfront. For clients with small balances or no prior access to a human adviser, these platforms offered systematic diversifica-

tion, low-cost ETF allocations, and automated services like rebalancing and lot-level tax-loss harvesting that would otherwise be difficult to perform manually (Jung et al. 2018; Grealish and Kolm 2021a). The platforms presented their technology as a neutral application of established financial theory made accessible at low cost through automation. The early design question was technical: how far personalization could extend beyond coarse risk categories without sacrificing scalability (Faloon and Scherer 2017).

Digital advice developed inside an existing regulatory framework. Suitability requirements, “know your customer” processes, and fiduciary or best-interest standards in both the US and Europe long predate the technology and constrain the kinds of elicitation and recommendations that can be performed. In Europe, MiFID II, PRIIPs, and the Insurance Distribution Directive specify how risk tolerance and capacity must be elicited, documented, and translated into recommendations; GDPR further regulates the use of personal data for automated profiling (Steennot 2021; Mrkývka and Šíková 2023; Bertrand, Eagan, and Maxwell 2023). In the US, SEC and FINRA rules play a similar role, with ERISA additionally governing retirement products (Grealish and Kolm 2021a). The short, structured questionnaires that became standard across the industry partly reflect this context: such formats produce documentation that supervisors can audit, backtest, and review against the firm’s stated suitability process (Rühr et al. 2019; Tertilt and Scholz 2018).

Between approximately 2018 and 2020, systematic audits of commercial platforms in the US and Europe produced a different picture of how the standardized pipeline was operating in practice. Risk questionnaires were often short, some questions had no observable effect on the resulting risk category, and platforms assigning the same risk label could end up with substantially different portfolios for the same client (Tertilt and Scholz 2018; Boreiko and Massarotti 2020). The mapping from questionnaire responses to risk category, and from risk category to portfolio weights, was opaque; the heterogeneity across providers exceeded what differences in portfolio theory could account for (Gaspar and Oliveira 2024). Defaults dominated. Investors who had previously held all-equity portfolios stayed at the platform’s default 50% equity exposure, while first-time users were the least likely to move away from it (Bräuer 2021). Qualitative work in the same period reframed robo-advisors as socio-technical assemblages that actively enact modern portfolio theory and Weberian rationality, “actively constructing passive investors” through automation, interface design, and behavioral defaults (Hayes 2021, 2020; Tan 2020).

Many platforms, particularly those operated by or affiliated with large financial institutions, generate revenue not only from asset-based fees but from product-related sources: affiliated ETFs, internal fund platforms, and cash sweep programs that deposit uninvested balances in interest-bearing accounts at below-market rates (Grealish and Kolm 2021b). These arrangements raised questions, in both the US and European regulatory contexts, about whether platforms operating in this way were meeting their legal obligations to act in clients’ best interests. Research on the implementation of explainability requirements under MiFID II and GDPR

found that feature-based explanations, introduced to meet legal expectations for transparency, sometimes increased user trust without improving user understanding of the underlying recommendation (Bertrand, Eagan, and Maxwell 2023). On this evidence, explainability functioned as a surface for building trust while leaving the underlying recommendation logic outside the feedback loop.

In the US, the SEC proposed extending conflict-of-interest rules to the algorithms themselves, then withdrew the proposal in 2025 after industry opposition, leaving the prior fiduciary framework in force (Bearup 2023; U.S. Securities and Exchange Commission 2025).

Spam Filters

Spam governance has a longer history than the other cases examined here, and a richer critical literature that has itself contested how the history should be read. Brunton reads spam as a “shadow history” of the internet, one where fundamental questions about legitimate communication, commercialization, and identity have been fought out but rarely made visible as political questions (Brunton 2013).

Email’s original protocol design made no provision for authenticating senders or distinguishing commercial bulk mail from other traffic (Banday and Qadri 2011). Through the mid-1990s, as commercial use of email expanded rapidly, unsolicited bulk mail grew from a nuisance to an operational problem for administrators, who responded with ad hoc keyword filters and heuristic rules. The first coordinated response took the form of collaborative blocklists: DNS-based databases of known abusive IP addresses maintained by volunteer organizations, most notably Spamhaus and SpamCop, to which any mail server operator could subscribe (Mathew and Cheshire 2017; Banday and Qadri 2011). These systems functioned as a shared knowledge infrastructure: listing criteria were relatively transparent, delisting procedures existed and were contestable, and participation was open to any operator. Early critics documented cases where entire IP address ranges were listed, excluding regions or providers rather than specific abusers, a pattern that Lueg and Twidale called “mystery meat” filtering and that others characterized as “digital redlining” (Lueg, Jeff, and Twidale 2007). The debates about blocklist governance in this period covered who could list a sender, on what criteria, and what recourse a listed party had. They were conducted in public, in technical forums, and through legal challenges.

From the mid-2000s, spam operations industrialized. Large botnets running on compromised consumer machines replaced open relays as the primary sending infrastructure, and spam monetization consolidated around affiliate marketing programs for pharmaceutical sites, counterfeit goods, and other gray or illegal markets (Levchenko et al. 2011; Pitsillidis 2013). Research tracing the spam value chain found that despite the apparent dispersion of sending infrastructure, revenue and operational control were concentrated at a relatively small number of monetization and payment nodes, the point in the chain the authors identified as most exposed to intervention (Levchenko et al. 2011; Rao and Reiley 2012).

The defensive response to industrial-scale spam shifted from collaboratively maintained blocklists to machine-learning classifiers trained on large volumes of message data aggregated across multiple tenants (Ferrara 2019; Banday and Qadri 2011). Classifier-based approaches scaled better against high-volume polymorphic threats than the rule-based and blocklist approaches they displaced; as cross-tenant training data accumulated at large providers, the performance advantage grew (Banday and Qadri 2011; Ferrara 2019). These classifier-based systems operate on proprietary cross-tenant data, and they do not provide senders with information about why a given message is or is not delivered (Ferrara 2019; Carmi 2020). During the same period, sender authentication standards (the Sender Policy Framework (SPF), Sender ID, and later DomainKeys Identified Mail (DKIM)) were developed and gradually adopted, establishing technical mechanisms to verify that a given host was authorized to send on behalf of a domain and thereby creating a new basis for sender reputation (Banday and Qadri 2011). Legal frameworks, including CAN-SPAM in the US and e-privacy directives in Europe, formalized a distinction between lawful commercial email and unlawful spam, though researchers consistently noted that legal frameworks lagged technical change and could not resolve the underlying economic dynamics (Rao and Reiley 2012; Banday and Qadri 2011).

The governance landscape underwent a parallel transformation. Mathew and Cheshire's ethnographic study of the anti-spam community documents a transition from a small group of administrators coordinating informally, through mailing lists, conferences, and direct contact with network operators, to a more institutionalized landscape of blocklists with formal procedures, industry forums, and best-practice bodies (Mathew and Cheshire 2017). M3AAWG (the Messaging, Malware and Mobile Anti-Abuse Working Group) became the primary venue for industry coordination on deliverability standards and anti-abuse best practices, but its membership is gated, and its proceedings are not public. The criteria determining whether a given sender can reach inboxes are set in a forum that the senders themselves cannot access.

Consolidation accelerated in the early 2010s as Gmail, Outlook, and Yahoo expanded to handle a substantial fraction of global email volume (Ferrara 2019). These large providers possessed data volumes sufficient to train filtering systems far more capable than anything available to smaller operators, and their filtering decisions effectively set de facto standards for deliverability across the ecosystem. DMARC (Domain-based Message Authentication, Reporting, and Conformance), standardized in 2012, extended the authentication framework to allow domain owners to specify policies for how receivers should handle messages that fail SPF or DKIM checks, providing domain owners with new leverage over how their sending infrastructure was represented but also further centralizing deliverability judgment at large providers who could implement and honor DMARC policies consistently (Banday and Qadri 2011).

Filtering effectiveness for ordinary users has improved substantially: inbox spam rates at major providers are much

lower than they were in the mid-2000s (Ferrara 2019). At the same time, smaller and independent senders face significant and often opaque barriers to deliverability. The GDPR has constrained how cross-service reputational data can be shared and retained, creating tension between the data-intensive methods that dominate modern filtering and the legal requirements of data minimization (Ferrara 2019). The most resourced abuse actors have adapted to the authentication and filtering environment, shifting from high-volume bulk spam to targeted phishing and business email compromise, which require different, more computationally intensive detection approaches and closer integration with broader security infrastructure (Rao and Reiley 2012). Access to that integrated security infrastructure requires the kind of organizational scale and industry relationships that smaller operators did not need in the blocklist era, when participation in the principal anti-spam coordination bodies was open to any operator.

4 Depoliticization and User Empowerment

In each case, the technology opened a space of contestable choices about infrastructure, knowledge, and objectives; in each case, that space was progressively narrowed through structuring moves that embedded those choices as technical facts. Drawing on accounts of how governments place contested decisions at one remove from democratic challenge (Burnham 2001, 2006), we call this pattern *depoliticization*: contestable questions are relocated into technical or expert forms, where they can no longer be collectively challenged (Flinders and Buller 2006; Mouffe 2005; Rancière 1999). The cases also show the reverse, *politicization*. EasyList's community governance kept the question of what to block open to challenge, and the SEC's rulemaking on fiduciary obligations kept open whose interests an advisor must serve.

A second throughline runs alongside the depoliticization dynamic: individual outcomes and organized collective capacity move independently. In spam, individual inbox quality at major providers improved substantially as classification consolidated around proprietary systems, while the organized capacity to challenge filtering criteria collapsed in the same transition. This paper's central analytical contribution is identifying what drives the two apart.

Intermediary Institutions Settling Contestation

One might expect the decisive contests over whom agents serve to occur directly between users and platforms. The ad blocker trajectory includes a period of direct technical competition: content delivery networks introduced obfuscated ad delivery, blockers refined selector syntax, and publishers deployed anti-blocking scripts (Iqbal, Shafiq, and Qian 2017). That contest did not produce the durable settlement. Across the trajectories, the structuring moves that shaped each outcome were often settled in intermediary institutions sitting between users and the actors who build and control technologies. These institutions exercise two functions that reinforce each other. They set the rules, specifying configuration choices that infrastructure actors then implement and

thereby constraining what is technically possible at the layers above. And they frame those rules as settled, presenting choices as technical specifications or expert consensus rather than open questions. We call these functions *material instantiation* and *discursive legitimation*. Where principal-agent accounts locate the source of misalignment in the dyadic relationship between user and firm (Dowding and Taylor 2024; Duetting, Feldman, and Talgam-Cohen 2024), the cases show that the decisive settlements occurred at this institutional level, prior to and independent of any individual user's relationship with any particular agent.

The Coalition for Better Ads and M3AAWG illustrate how this works in practice. The Coalition produced concrete filtering criteria enforced through Chrome and framed the question of acceptable advertising as industry consensus, with governance that included no adversarial challenge mechanism and no public interest representation (Megali 2022; Blackburn, Ritala, and Keränen 2023). M3AAWG operates on a similar logic: deliverability standards are set through gated proceedings with non-public outputs, framed as expert consensus on abuse prevention, and are inaccessible to most of the senders whose practices they govern (Mathew and Cheshire 2017). In both cases, the locus of contestation shifted from a publicly accessible forum to a closed institutional one. Organizing a challenge now requires standing in a body that was not designed to accommodate it.

The SEC's handling of fiduciary obligations and the EU's Digital Markets Act (DMA) and Digital Services Act (DSA) show the adversarial-enabling alternatives. A statutory public interest mandate, open comment periods, legal standing for public interest groups, judicial review, and revision authority counteract specific foreclosure mechanisms (Bearup 2023): for example, open comment keeps the value commitments contestable; revision authority prevents material embedding from accumulating unchecked. The DMA targets the infrastructural dimension through interoperability mandates; the DSA targets the epistemic dimension through transparency requirements and algorithm-choice mandates, both enforced through adversarial proceedings (Reviglio and Santoni 2023). Regulatory capture remains possible, and resource disparities and revolving door dynamics are well documented (Carpenter and Moss 2013), but the architecture raises the cost of capture and keeps the value commitment visible as a political choice.

A growing body of critical literature has questioned whether transparency alone is sufficient for algorithmic accountability (Ananny and Crawford 2018; Pasquale 2015). Transparency reports a system's current configuration. The constitutive choices are the ones that produced that configuration, and they are often not part of what gets reported. Chrome's Manifest V3 documentation accurately describes which extension APIs are available; it does not reveal that the API design was a contestable choice among alternatives, or that it was made by an actor with commercial interests in the outcome. This is why transparency without the structural complements that the SEC model illustrates risks performing the discursive legitimation function of intermediary institutions; it makes the cage visible without providing a key.

The cases show that technical performance for individual welfare and collective political capacity can move in opposite directions, but not as a necessary tradeoff. Centralized spam filtering improved individual inbox quality; whether that improvement required eliminating organized challenge capacity is a causal question the cases do not settle. The ad blocker trajectory is more direct: uBlock Origin on Firefox, the community-governed path, blocks a far wider range of advertising than Chrome's enforcement of Coalition standards, which filters only the formats the Coalition designates unacceptable (Megali 2022; Miagkov and Cyphers 2021). The cases provide no support for the view that centralized control is a precondition for adequate technical performance.

Conditions of Collective Contestation

Knowing where configurations get settled raises a further question: what conditions determine whether organized communities can contest them? The cases reveal three conditions that jointly determine whether that capacity can form.

Ad blocking provides the positive instance. When filter lists were maintained through community governance with publicly readable syntax and open contribution, three conditions were met: the epistemic task of identifying ad-serving domains was performable with publicly observable information (public observability); the browser extension API gave users independent infrastructure through which community knowledge could be applied (independent infrastructure) (Storey et al. 2017; Snyder, Vastel, and Livshits 2020); and EasyList's transparent rules and forkable structure kept the question of what to block visibly open to challenge (governance sustaining contestation). Manifest V3 degraded the infrastructure condition while leaving the knowledge base intact (Miagkov and Cyphers 2021). That uBlock Origin retained full capability on Firefox, where the API remained intact, demonstrates which condition was operative: the same knowledge base and governance produced different outcomes depending on which browser provided the material basis for applying them.

Spam governance traces the collapse of all three conditions. Early collaborative blocklists (Spamhaus, SpamCop, the Distributed Checksum Clearinghouse) satisfied them; the task was performable with publicly observable information, and decentralized email infrastructure provided the material basis for community knowledge to be applied (Mathew and Cheshire 2017; Banday and Qadri 2011). As spam industrialized, the task shifted to require behavioral data available only at large-provider scale (Ferrara 2019). This meant that public observability was lost; independent infrastructure became insufficient without it. Individual inbox quality improved substantially over this period. Yet at the same time, the organized capacity to produce shared knowledge about filtering criteria and to contest the standards governing them collapsed during the same transition. The governance condition failed with the other two. Spamhaus and its peers had settled filtering criteria in public exchange; the large providers that replaced them settled the same criteria within their own classifiers. What arms-race accounts read

as a technical victory for superior classifiers (Nithyanand et al. 2016) was, from the standpoint of political capacity, a displacement of the material conditions through which organized challenge had been possible.

Recommender systems present the case where none of the conditions were ever present. Major platforms controlling recommender systems for content and commerce have controlled all three dimensions since their commercial launch: the infrastructural basis of the recommendation process, the behavioral data through which the epistemic task was performed, and the objectives the systems optimized for. Because the choices were never publicly contested in the relevant sense, no neutral-appearing intermediary institution was ever established to settle them. The DSA transparency requirements have produced compliance documentation without forming the organized constituency needed to make those disclosures actionable (Fabbri and Boratto 2025; Meßmer and Degeling 2023). The market-selection account (that configurations reflect user preferences expressed through choice among alternatives) faces a prior difficulty here: users were never positioned to choose among meaningfully different recommendation configurations because the platform configured the infrastructure through which such differences could have existed from the outset. Browser extensions that hide or re-sort feed items, and research prototypes that demonstrate alternative rankings, work on the list the platform has already assembled. They cannot see which candidates it considered or what its ranking was optimizing for.

Robo-advisors isolate the governance condition: fiduciary standards, MiFID II, and open rulemaking keep the value question contestable, but the questionnaire-to-portfolio mapping stays opaque and platforms own the advice pipeline. Contestation persists at the value question while outcomes re-embed at the implementation layer (§4).

The comparison yields three conditions for organized collective contestation to form and persist.

1. **Public observability** for collective knowledge to *form*: the epistemic task must be performable with information accessible without platform cooperation.
2. **Independent infrastructure** for collective knowledge to be *used productively*, operationalized through infrastructure the platform does not control.
3. **Governance that sustains contestation** for collective knowledge to *persist* under adversarial pressure: a formal pathway through which decisions can be challenged. Without one, an institution absorbs contestation instead of sustaining it.

Whether rules formally permit challenge is a secondary question to whether the material conditions are in place.

Durability and the Burden of Re-politicization

These conditions, once lost, have proven hard to restore. Even where the dominant arrangements outperformed alternatives on the relevant technical task (Ferrara 2019), durability came from something else: they displaced the arrangements through which alternatives had been organized.

For instance, when Google redesigned Chrome's extension API, it removed the infrastructure through which community filter lists had operated. The lists survived, granted, but Chrome users could no longer run them in full. And while users could individually switch to other browsers that still support full functionality, most of the market is dominated not only by Chrome but also by browsers that use the same Chromium engine. Individual user choice counts for even less in spam. When spam classification consolidated around cross-tenant behavioral data, the publicly observable signals on which decentralized alternatives had depended were no longer sufficient, and the collaborative blocklist infrastructure lost the material basis it required. In this case, even a provider willing to open its classifier could not recreate the public observability that collaborative blocklists required, because the task now demands data no independent operator holds. In both cases, bringing an alternative back is not a choice users can make.

Foreclosure is durable because embedding is concrete: API specifications, data access regimes, institutional standards. Each dimension of contestable choice has two layers: the value question it poses and its material implementation in specific artifacts and procedures. The two functions named in §4, *material instantiation* and *discursive legitimation*, operate at this implementation layer. Whoever holds that layer decides the question, whatever the formal rules say. Google capped how many filtering rules an extension may load and settled what users may block; the large providers moved classification onto data they alone held and settled what filters could know. In ad blocking and spam, consolidation displaced working alternatives at this layer. Recommender systems never had alternatives here, because platforms held the layer from the moment the systems launched.

Single-dimensional foreclosure leaves cross-dimensional leverage. Manifest V3 constrains the infrastructural dimension, but the value question of what users should be able to block remains open, and Firefox preserves a site from which it can be pursued. When foreclosure accumulates across dimensions simultaneously, those footholds are removed together. Chrome's integration of Coalition for Better Ads standards embedded a value commitment within an institutional arrangement and fused both with browser infrastructure in a single sequence. Challenging the value standard requires overcoming the infrastructural embedding; the Coalition's governance structure provides no institutional pathway for either. Spam exhibits the same compounding: proprietary filtering, engagement-based reputation scoring, and centralized infrastructure combine in such a way that engaging in any dimension requires engaging the others (Ferrara 2019; Mathew and Cheshire 2017).

The concreteness of embedding also explains the limits of regulatory intervention. When institutions built for adversarial challenge succeed in politicizing a dimension, actors with enough material control move the contestable choice down to an implementation level the intervention does not reach, a pattern we call re-embedding. MiFID II forced risk-profiling criteria into public view and created a governance structure for ongoing challenge (Steennot 2021); platforms

responded by encoding commercial discretion in questionnaire design, default fund allocation, and profiling conventions that formally satisfy the requirements while preserving opacity at the implementation level (Boreiko and Massarotti 2020; Tertilt and Scholz 2018). Early DSA compliance patterns in recommender systems show the same dynamic, with transparency requirements producing ranking documentation that audits have struggled to reconcile with observed system behavior (Fabbri and Boratto 2025; Meßmer and Degeling 2023).

The shift in threshold for political challenge follows the same logic. Contesting a filtering decision in the EasyList period meant proposing a change, forking the list, and arguing in community forums. After the same question was embedded in Chrome's Better Ads enforcement, the path to changing what Chrome enforces requires demonstrating that "acceptable advertising" encodes a value commitment, identifying a challenge pathway the Coalition's governance does not provide, and overcoming material embedding in a browser with dominant market share (Megali 2022). That first step is what discursive legitimization costs a challenger. Once a standard circulates as expert consensus, anyone contesting it must first show that a choice was made at all. Switching to Firefox remains available. It changes one user's own configuration and leaves the standard governing the dominant browser in place. Access asymmetry reinforces the burden: Coalition governance was dominated by industry members (Blackburn, Ritala, and Keränen 2023), M3AAWG requires organizational resources most affected parties lack (Mathew and Cheshire 2017), and the Acceptable Ads Committee was funded by the actors it regulated (Gupta and Panda 2020). Foreclosure and the threshold for re-politicization accumulate together, through routine decisions that require no visible political act.

Adversarial institutional design prevents foreclosure only when coordinated with the material conditions identified in the preceding subsection. The account implies a falsifiability condition: trajectories in which multi-dimensional foreclosure was readily reversed, or in which sustained collective contestation persisted without those material conditions, would require it to be revised.

Depoliticization, across these trajectories, is a structural tendency in domains where the actors providing infrastructure have commercial interests in configuring what agents can do; this tendency drives the individual-collective divergence the opening named. The next section turns to the institutions now forming around agentic AI and asks whether they are built like those that sustained contestation or those that foreclosed it.

5 Looking Forward: Agentic AI

Shared structure, wider scope. Each technology examined above acted on users' behalf in a domain where another party held an interest potentially opposed to the user's (§3). Agentic AI shares this characteristic: agents are poised to transact on behalf of their users among merchants, platforms, and services whose interests the transaction affects. It is also poised to act across domains, and potentially beyond those analyzed here. The structure is the same; the scope is

wider. That widening means a single ecosystem could set the boundaries of possible agency across many domains at once. Because the structure is shared, the framework predicts where depoliticization will occur if this trajectory follows the others: in an industry-led intermediary that converts contestable choices into apparently technical specifications; through governance that is open at the surface while decision authority is held asymmetrically underneath; and without an adversarial mechanism to defend that openness once consolidation advances.

What is being settled. Anthropic's Model Context Protocol (MCP), released in November 2024, defines how AI agents communicate with external services through a standardized client-server architecture (Linux Foundation 2025). Adoption by OpenAI, Google DeepMind, and other major platforms within months of release, together with the December 2025 donation to the Agentic AI Foundation (AAIF) under the Linux Foundation, established MCP as the de facto standard within twelve months (Model Context Protocol Blog 2025). The scope of what is settled at this layer is broad: what counts as a tool an agent may invoke, what data flows are permitted under what authorization, and how personal data is delineated when an agent acts on a user's behalf. Each is being fixed in the specification and its defaults, not in any statute.

The breadth is striking, and it covers the domains of our case studies. In each, agentic AI could unsettle what the earlier trajectory settled. Recall the introduction's examples: an agent building and maintaining a personal Chromium-based browser, or writing extensions that put the user's instructions ahead of the platform's preferences.

Where consequential choices are being settled. On paper, the AAIF is open. The specification is public; anyone can submit changes through the Specification Enhancement Proposal (SEP) process, maintainer meetings are bi-weekly with public notes, and the Linux Foundation provides a neutral institutional home (Parra 2025). Binding authority sits elsewhere. Decision-making over strategic investments, budget, and membership criteria rests with a Governing Board composed of platinum members (AWS, Anthropic, Block, Bloomberg, Cloudflare, Google, Microsoft, OpenAI), with influence tied to financial commitment tier (Linux Foundation 2025). No formal standing exists for user organizations or public-interest representatives, no adversarial mechanism allows an affected party to contest an SEP outcome, and no public-interest mandate binds the Board. Individual users can configure their own permissions, and anyone can comment through the SEP process; adoption rests with the project's maintainers, and the Board that sets budget and membership sits above them. Peer standards bodies show that stronger commitments are available to protocol governance: the IAB directs that protocol decisions favor end users where interests conflict (Nottingham 2020), and the W3C's design principles rank users' interests above those of page authors and implementors (van Kesteren and Stachowiak 2007). Nor is regulation currently positioned to reach these choices: the term "agentic systems" does not appear in the EU AI Act's legal text, so the technical standards

under development for 2026–2027 will arrive downstream, operating within whatever the protocol layer has already settled on (Oueslati and Staes-Polet 2025). This is discursive openness without structural protection, the configuration §4 traced through the Coalition for Better Ads, now at a higher layer of infrastructure.

Is alignment enough? The dominant approach to making agents serve their users is technical alignment: training and steering a model so that an individual agent acts in accordance with its user’s values. The sophisticated end of this program takes value diversity seriously, asking how a model can represent many users’ values rather than average over them (Sorensen et al. 2024). The trajectories traced here suggest that even this cannot fully secure the goal. Encoding values, plural or not, is work the provider performs on infrastructure the provider runs. And beyond any single agent, what counts as a tool an agent may invoke, which data flows are authorized, and which servers a hosted agent may call are settled in the specification and its governance before any particular agent is trained. Those settlements decide which alignments are achievable in the domain at all. Take the robo-advisor case: advice satisfied suitability requirements at each interaction while defaults and fund selection displaced user interests across the platform (Boreiko and Massarotti 2020; Tertilt and Scholz 2018). Aligning an individual agent and keeping a domain contestable are distinct problems, and better training does not solve the second. Alignment that is to hold across providers and over time therefore depends on conditions the cases identified: an audit record lets anyone verify that an agent served its user, and a governance channel lets an affected party contest a misaligned default rather than merely document it.

Preserving contestation. Our analysis indicates that the sites where contestation can be lodged matter more than broad normative declarations. However, these sites have been rapidly solidifying: the AAIF has been standing up its governance since December 2025 (Linux Foundation 2025), the MCP specification is being iterated upon, with an updated version released in July 2026 (Model Context Protocol 2026), and the EU AI Act’s technical standards for agentic systems are due for adoption in 2027 (Oueslati and Staes-Polet 2025). This is a critical moment to establish formal standing for user and public-interest representatives, and an adversarial mechanism through which SEP outcomes can be challenged on user-harm grounds.

We now map to agentic AI the conditions that sustained contestation in the cases (§4): public observability, independent infrastructure, and governance that sustains contestation. Regarding public observability, the MCP project already records its decisions and the discussions behind them and makes them publicly available; the AAIF, whose budget and membership decisions carry no such obligation, would need to meet the same standard, and the agents themselves would need to be auditable, with tool calls and data flows readable by outside parties, including public bodies, rather than only by the provider running the agent. Regarding independent infrastructure, the specification is openly licensed, and anyone can run their own client and servers

today, but most agents are hosted by providers who decide which servers they may call, and keeping the protocol open to smaller players may require regulatory backing. Regarding governance, a duty to users comparable to the fiduciary duty the SEC imposed on advisors would give an affected party grounds to challenge a specification, and the audit record is what such a challenge would rest on.

6 Limitations and Conclusion

Limitations. A key limitation concerns the goal of user empowerment itself. We ask whether users will be empowered, not whether empowerment is beneficial. In some cases user empowerment is dangerous on its own terms: enabling a user to plan an attack, disrupt critical infrastructure, or target another person’s life or property. Relatedly, empowerment can put users’ interests in conflict with one another. The right to be forgotten illustrates this pattern: enacted as an individual right, it empowers the person seeking erasure at the expense of other parties’ legitimate interests, including those who would search, those who would publish, and future users whose informational environment is being shaped. We do not address how to collectively empower users with diverse and conflicting interests, how to adjudicate between them, or how viable coalitions among users might be built.

The analysis is bounded in other ways as well. The cases examined are US- and EU-centric and exclude other important global considerations. In particular, the Chinese contributions to agentic AI already loom large. Moreover, the mechanisms through which constitutive politics operate in regulatory contexts without equivalent administrative law traditions or civil society infrastructure are not addressed and require separate analysis.

At a methodological level, the three dimensions emerged from the trajectories we analyzed, and §2 argues they are necessary for any computational agent; whether they are exhaustive, and whether additional dimensions become constitutive in particular domains, remains open. The re-embedding dynamic (§4) and the interactions among the three dimensions are characterized through the cases, not formally modeled. The account is open to revision: the falsifiability condition stated in §4 specifies the kinds of evidence that would require reconsidering it. Additionally, the agentic AI section in §5 is prospective, identifying structural conditions and the stakes of early decisions while the trajectories themselves have yet to unfold.

Conclusion. We began by asking who wins the contest between users’ agents and the platforms they act on. By our analysis, the decisive moves are made before play begins across three dimensions: teleological, infrastructural, and epistemic. When institutions are able to convert political questions into technical settlements, users may benefit from an improved product, as spam was largely removed from inboxes, while losing the ability to affect the process and any future changes. Agentic AI’s rules are being written now and will simultaneously affect many domains. What matters most is not individual outcomes but an ecosystem that allows vigorous ongoing contestation.

Acknowledgments

We thank the members of the Political Economy and Algorithms Collective at the University of Michigan for their useful feedback. We also thank Professor David Temin at the University of Michigan for his feedback and suggestions on the political science approach, and Ali Alkhatib for his methodological feedback in the project's early stages.

Disclosure Statement

We used generative AI tools for editing and for searching sources, which we verified. Authors declare no competing interests.

References

- Alrizah, M.; Zhu, S.; Xing, X.; and Wang, G. 2019. Errors, Misunderstandings, and Attacks: Analyzing the Crowdsourcing Process of Ad-blocking Systems. In *Proceedings of the Internet Measurement Conference*, IMC '19, 230–244. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6948-0.
- Ananny, M.; and Crawford, K. 2018. Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20(3): 973–989.
- Banday, M. T.; and Qadri, J. A. 2011. SPAM – Technological and Legal Aspects. arXiv:1112.5621.
- Bearup, G. 2023. Quick Take - SEC Proposed Rule on Predictive Data Analytics.
- Bertrand, A.; Eagan, J. R.; and Maxwell, W. 2023. Questioning the Ability of Feature-Based Explanations to Empower Non-Experts in Robo-Advised Financial Decision-Making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 943–958. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0192-4.
- Birch, K. 2025. Do Artifacts Have Political Economy? *Science, Technology, & Human Values*, 01622439251352167.
- Blackburn, O.; Ritala, P.; and Keränen, J. 2023. Digital Platforms for the Circular Economy: Exploring Meta-Organizational Orchestration Mechanisms. *Organization & Environment*, 36(2): 253–281.
- Boreiko, D.; and Massarotti, F. 2020. How Risk Profiles of Investors Affect Robo-Advised Portfolios. *Frontiers in Artificial Intelligence*, 3.
- Bräuer, K. 2021. Nudged into Better Portfolios and Lower Risk: Robo-Advice and Savings Decisions. Social Science Research Network:3927860.
- Brave Software. 2023. Brave Shields. <https://brave.com/shields/>.
- Brunton, F. 2013. *Spam: A Shadow History of the Internet*. The MIT Press. ISBN 978-0-262-01887-6.
- Burnham, P. 2001. New Labour and the Politics of Depoliticisation. *The British Journal of Politics and International Relations*, 3(2): 127–149.
- Burnham, P. 2006. Depoliticisation: A Comment on Buller and Flinders. *The British Journal of Politics and International Relations*, 8(2): 303–306.
- Carmi, E. 2020. *Media Distortions: Understanding the Power Behind Spam, Noise, and Other Deviant Media*. Peter Lang. ISBN 978-1-4331-6691-4.
- Carpenter, D.; and Moss, D. A., eds. 2013. *Preventing Regulatory Capture: Special Interest Influence and How to Limit It*. New York: Cambridge University Press. ISBN 978-1-107-03608-6.
- Cobbe, J.; and Singh, J. 2019. Regulating Recommending: Motivations, Considerations, and Principles. *European Journal of Law and Technology*, 10(3).
- Covington, P.; Adams, J.; and Sargin, E. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198. Boston Massachusetts USA: ACM. ISBN 978-1-4503-4035-9.
- Dowding, K.; and Taylor, B. R. 2024. Algorithmic Decision-Making: A Principal Agent Framework.
- Duetting, P.; Feldman, M.; and Talgam-Cohen, I. 2024. Algorithmic Contract Theory: A Survey. <https://arxiv.org/abs/2412.16384v1>.
- Fabbri, M.; and Boratto, L. 2025. Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the Digital Services Act. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, RecSys '25, 51–61. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-1364-4.
- Faloon, M.; and Scherer, B. 2017. Individualization of Robo-Advice. *The Journal of Wealth Management*, 20(1): 30–36.
- Ferrara, E. 2019. The History of Digital Spam. *Commun. ACM*, 62(8): 82–91.
- Flinders, M.; and Buller, J. 2006. Depoliticisation: Principles, Tactics and Tools. *British Politics*, 1(3): 293–318.
- Gaspar, R. M.; and Oliveira, M. 2024. Robo Advising and Investor Profiling. *FinTech*, 3(1): 102–115.
- Grealish, A.; and Kolm, P. N. 2021a. Robo-Advisors Today and Tomorrow: Investment Advice Is Just an App Away. *The Journal of Wealth Management*, 24(3): 144–155.
- Grealish, A.; and Kolm, P. N. 2021b. Robo-Advisory: From Investing Principles and Algorithms to Future Developments. Social Science Research Network:3776826.
- Gupta, R.; and Panda, R. 2020. Block the Blocker: Studying the Effects of Anti Ad-blocking. arXiv:2001.09434.
- Hancock, A. 2021. Manifest V3: Open Web Politics in Sheep's Clothing. <https://www.eff.org/deeplinks/2021/11/manifest-v3-open-web-politics-sheeps-clothing>.
- Hayes, A. 2020. Enacting a Rational Actor: Roboadvisors and the Algorithmic Performance of Ideal Types. *Economy and Society*, 49(4): 562–595.
- Hayes, A. S. 2021. The Active Construction of Passive Investors: Roboadvisors and Algorithmic 'Low-Finance'. *Socio-Economic Review*, 19(1): 83–110.

- Iqbal, U.; Shafiq, Z.; and Qian, Z. 2017. The Ad Wars: Retrospective Measurement and Analysis of Anti-Adblock Filter Lists. In *Proceedings of the 2017 Internet Measurement Conference*, 171–183. London United Kingdom: ACM. ISBN 978-1-4503-5118-8.
- Jung, D.; Dorner, V.; Weinhardt, C.; and Puzmaz, H. 2018. Designing a Robo-Advisor for Risk-Averse, Low-Budget Consumers. *Electronic Markets*, 28(3): 367–380.
- Levchenko, K.; Pitsillidis, A.; Chachra, N.; Enright, B.; Felleggyhazi, M.; Grier, C.; Halvorson, T.; Kanich, C.; Kreibich, C.; Liu, H.; McCoy, D.; Weaver, N.; Paxson, V.; Voelker, G. M.; and Savage, S. 2011. Click Trajectories: End-to-End Analysis of the Spam Value Chain. In *2011 IEEE Symposium on Security and Privacy*, 431–446.
- Linux Foundation. 2025. Linux Foundation Announces the Formation of the Agentic AI Foundation (AAIF), Anchored by New Project Contributions Including Model Context Protocol (MCP), Goose and Agents.Md.
- Lueg, C.; Jeff, H.; and Twidale, M. 2007. Mystery Meat Revisited: Spam, Anti-Spam Measures and Digital Redlining. *Webology*, 4.
- Mathew, A.; and Cheshire, C. 2017. Risky Business: Social Trust and Community in the Practice of Cybersecurity for Internet Infrastructure. In *Hawaii International Conference on System Sciences*.
- Megali, T. 2022. Digital Platforms as Members of Meta-Organizations: A Case Study of the Online Advertising Market. *M@n@gement*, 25(2): 10–26.
- Meßmer, A.-K.; and Degeling, M. 2023. Auditing Recommender Systems – Putting the DSA into Practice with a Risk-Scenario-Based Approach. arXiv:2302.04556.
- Miagkov, A.; and Cyphers, B. 2021. Google’s Manifest V3 Still Hurts Privacy, Security, and Innovation. <https://www.eff.org/deeplinks/2021/12/googles-manifest-v3-still-hurts-privacy-security-innovation>.
- Model Context Protocol. 2026. Model Context Protocol Specification, Revision 2026-07-28. <https://modelcontextprotocol.io/specification/2026-07-28/>.
- Model Context Protocol Blog. 2025. MCP Joins the Agentic AI Foundation.
- Mouffe, C. 2005. *The Return of the Political*, volume 8. Verso.
- Mouffe, C. 2011. *On the Political*. routledge.
- Mrkývka, P.; and Šíková, Z. 2023. Robo-Advisory as a Driving Force of the Financial Market in the Light of Digitalization? *WSB Journal of Business and Finance*, 57(1): 66–77.
- Muller, M. J.; and Kuhn, S. 1993. Participatory Design. *Commun. ACM*, 36(6): 24–28.
- Nithyanand, R.; Khattak, S.; Javed, M.; Vallina-Rodriguez, N.; Falahrastegar, M.; Powles, J. E.; De Cristofaro, E.; Hadadi, H.; and Murdoch, S. J. 2016. Adblocking and Counter Blocking: A Slice of the Arms Race. In *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*.
- Nottingham, M. 2020. The Internet Is for End Users. RFC 8890, Internet Architecture Board.
- Oueslati, A.; and Staes-Polet, R. 2025. Ahead of the Curve: Governing AI Agents under the EU AI Act. Technical report, The Future Society.
- Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press. ISBN 978-1-59420-300-8.
- Parra, D. S. 2025. SEP-932: Model Context Protocol Governance.
- Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. ISBN 978-0-674-73606-1.
- Pitsillidis, A. 2013. *Spam Value Chain : Defensive Intervention Analysis*. Ph.D. thesis, UC San Diego.
- Rancière, J. 1999. *Disagreement: Politics and Philosophy*. U of Minnesota Press.
- Rao, J. M.; and Reiley, D. H. 2012. The Economics of Spam. *Journal of Economic Perspectives*, 26(3): 87–110.
- Reviglio, U.; and Santoni, G. 2023. Governing Platform Recommender Systems in Europe: Preliminary Insights from China.
- Rühr, A.; Streich, D.; Berger, B.; and Hess, T. 2019. A Classification of Decision Automation and Delegation in Digital Investment Management Systems. In *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS)*, 1435–1444.
- Russell, S. J. 2010. *Artificial Intelligence a Modern Approach*. Pearson Education, Inc.
- Seaver, N. 2018. Captivating Algorithms: Recommender Systems as Traps - Nick Seaver, 2019. *Journal of Material Culture*.
- Snyder, P.; Vastel, A.; and Livshits, B. 2020. Who Filters the Filters: Understanding the Growth, Usefulness and Efficiency of Crowdsourced Ad Blocking. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(2): 1–24.
- Sorensen, T.; Moore, J.; Fisher, J.; Gordon, M.; Miresghal-lah, N.; Rytting, C. M.; Ye, A.; Jiang, L.; Lu, X.; Dziri, N.; Althoff, T.; and Choi, Y. 2024. A Roadmap to Pluralistic Alignment. arXiv:2402.05070.
- Steennot, R. 2021. Robo-Advisory Services and Investor Protection. *Law and Financial Markets Review*, 15(3-4): 262–277.
- Storey, G.; Reisman, D.; Mayer, J.; and Narayanan, A. 2017. The Future of Ad Blocking: An Analytical Framework and New Techniques. *arXiv preprint arXiv:1705.08568*.
- Tan, G. K. S. 2020. Robo-Advisors and the Financialization of Lay Investors. *Geoforum*, 117: 46–60.
- Tertilt, M.; and Scholz, P. 2018. To Advise, or Not to Advise—How Robo-Advisors Evaluate the Risk Preferences of Private Investors. *The Journal of Wealth Management*, 21(2): 70–84.
- U.S. Securities and Exchange Commission. 2025. Withdrawal of Proposed Regulatory Actions.
- van Kesteren, A.; and Stachowiak, M. 2007. HTML Design Principles. W3C Working Draft, World Wide Web Consortium.

- Wang, P. 2022. Recommendation Algorithm in TikTok: Strengths, Dilemmas, and Possible Directions. *Int'l J. Soc. Sci. Stud.*, 10: 60.
- Winner, L. 1980. Do Artifacts Have Politics? In *Computer Ethics*, 177–192. Routledge.
- Wooldridge, M.; and Jennings, N. R. 1995. Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, 10(2): 115–152.
- Xiang, Y. 2022. YouTube and the Protocological Control of Platform Organisations. *Qualitative Research in Accounting & Management*, 19(3): 348–372.
- Zhao, S.; Kalra, A.; Borcea, C.; and Chen, Y. 2020. To Be Tough or Soft: Measuring the Impact of Counter-Ad-blocking Strategies on User Engagement. In *Proceedings of The Web Conference 2020, WWW '20*, 2690–2696. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-7023-3.