Agentic AI: User Empowerment or Enclosure?

David Gamba[1], **Daniel M. Romero**[1], **Grant Schoenebeck**[1]

[1]University of Michigan
Ann Arbor, MI, USA
gamba@umich.edu, drom@umich.edu, schoeneb@umich.edu

Abstract

Agentic AI promises a more flexible form of digital agency: 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 we argue that the answer depends on more than the technology. We conduct a comparative case analysis of four more mature domains where similar forms of agency arose: browser-based ad blockers, platform recommender systems, financial robo-advisors, and email spam governance. Across the cases, decisions 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 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 adversarial challenge, user-aligned agency proved more durable; where proprietary infrastructure and closed standard-setting absorbed contestation, displacement compounded. We apply this to agentic AI, where governance arrangements consolidating around the Model Context Protocol and the Agentic AI Foundation are settling these configurations before the choices that define what agents can do move outside the reach of users and the public.

1 Introduction

In 2019, Google redesigned the extension API for its Chrome browser. 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 (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.

Earlier technologies that promised to act on users' behalf raise the same kind of question. 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 the boundaries of possible agency shifted over time.

The four cases span three domains and two decades. Each initially appeared explicable by a simpler account (technological arms race, regulatory capture, or market selection) and the comparison turns on whether those accounts hold across the variation the cases provide: organized collective contestation that formed and persisted, that collapsed under pressure, and that was never constituted at all.

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). Our analysis shows that durable user empowerment depends on the collective capacity to contest such choices. Across the cases, that capacity was sustained where intermediary institutions provided adversarial governance and independent infrastructure (the SEC's rulemaking, Firefox's extension API) and collapsed where proprietary infrastructure and gated standard-setting foreclosed participation (M3AAWG, Chrome's Manifest V3). When foreclosure accumulates across all three dimensions at once, the footholds from which contestation could be rebuilt are removed together, making displacement qualitatively harder to reverse.

By *user empowerment* we mean the capacity to meaningfully influence how an agent acts in ways that align with the user's own interests. 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*.

Agentic AI is now consolidating across a far broader set of domains, including commerce, information retrieval, and task automation, through a single protocol infrastructure. Anthropic released the Model Context Protocol in late 2024 and, after major providers adopted it within twelve months (Linux Foundation 2025), donated it to the Agentic AI Foundation at the Linux Foundation, where a Governing Board composed of platform providers carries

decision-making authority and no formal standing exists for user organizations or public-interest representatives. The governance arrangements, protocol standards, and default configurations that will shape what these agents can do are being settled in this venue now, before material embedding has accumulated.

We advance three contributions. First, using historical trajectories of user-facing technologies, we show that individual-level improvements and collective contestation capacity can move in opposite directions, and we specify conditions under which that divergence occurs. Second, we develop a constitutive politics framework identifying the infrastructural, epistemic, and teleological dimensions where user-aligned agency is foreclosed or sustained, specifying the mechanisms through which depoliticization operates in these domains. Third, we apply the framework to the protocol-level institutional arrangements consolidating around agentic AI, illustrating that the configuration emerging at the Agentic AI Foundation matches the foreclosure-prone patterns the cases identify, and we specify 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 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 (Winner 1980; Birch 2025; Mouffe 2005), 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.

Granted, there are of course other lenses that could 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 it, and each of the three dimensions denotes where those choices were made: in infrastructure, in what the system can know, in what it optimizes for.

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. When a platform encodes a particular metric as its recommendation objective, that is a political choice embedded in a loss function. When an advisory system defaults to particular allocations regardless of stated client preferences, that is a political choice embedded in a profiling algorithm.

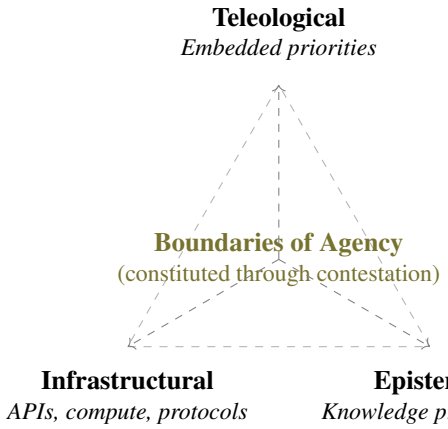

Figure 1: Dimensions of constitutive politics that define the boundaries of agency. Struggles over these dimensions constitute what agents in a domain can be and whose interests they can serve.

Across the trajectories, a common pattern emerges in how contestable choices get settled and why some configurations sustain collective challenge while others foreclose it. We derive it from the cases, not from the framework alone; it holds where the arms-race, regulatory-capture, and market-selection accounts do not.

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 moments and in different domains, as tools that would 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 how they developed: 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, such as what they mean, which were decisive, and whether they follow any common pattern, is a question 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 cancelled 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 the viability of both mechanisms depends on what browser extension APIs permit extensions to do—a question that 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). As adoption grew—one estimate placed Adblock Plus alone at approximately 100 million users by the late 2010s (Megali 2022)—the economic stakes for the publishing and advertising industries rose correspondingly. 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 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, which consolidated the Acceptable Ads arrangement as a legally legitimate 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 counter-measures. Specialized anti-adblock vendors such as PageFair, Taboola, and Outbrain 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-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 from consumer or public interest organizations, and its governance structure contained no adversarial challenge mechanism or independent review process (Blackburn, Ritala, and Keränen 2023). Beginning in 2019, Chrome enforced Coalition standards as a built-in browser feature, filtering ads that violated them automatically for all users regardless of whether they had installed any ad-blocking extension (Megali 2022). What counted as acceptable advertising, previously contested between publishers, advertisers, blocklist communities, and users, was now decided by a browser engine whose developer had co-founded the body that defined the standards.

By this point filter lists themselves had accumulated operational pressures that conditioned the next phase of the contest. Longitudinal analysis of EasyList found that

roughly 90% of resource-blocking rules were unused in typical browsing scenarios, contributing to computational overhead that mattered particularly 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, which allowed blocking tools to inspect and cancel network requests dynamically. 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 (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), framings that have been both reproduced and contested in critical analyses (Cyphers 2021; Hancock 2021).

The practical effect on network-level blocking capability has been debated among extension developers. The effect is less disputed on Firefox, which retained the older API and on which uBlock Origin continues to operate with its full functionality. The Coalition for Better Ads case and the Manifest V3 transition represent two distinct kinds of change to the ad-blocking terrain; whether one paved the way for the other or they are better understood as independent developments is a matter on which analysts have reached different conclusions.

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; and ongoing automation handles rebalancing, contributions, and, on some US platforms, tax-loss harvesting (Grealish and Kolm 2021a,b). Underneath sits established financial theory—modern portfolio theory and long-run expected-utility reasoning—which, for any specified risk preference, yields 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 it also communicates back 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 diversification, 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 questions were technical: how far could personalization 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 best-interest standards in both jurisdictions long predate the technology and constrain what kind of elicitation and recommendation 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 Šiková 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). Allocations tended toward conservative defaults that pushed investors toward lower equity exposure than their stated risk tolerance suggested, with evidence that some platforms nudged users into a specific default regardless of history (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 rates below prevailing market (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). Explainability in this trajectory has functioned as a trust-building surface that leaves the underlying recommendation logic outside the feedback loop.

In the US, the Securities and Exchange Commission proposed rules in July 2023 that would have required broker-dealers and investment advisers to identify and either eliminate or neutralize conflicts of interest arising from their use of predictive data analytics and similar technologies (Bearup 2023). Industry participants opposed the proposal broadly, arguing that it was unworkable in practice and would apply to a wide range of routine advisory functions, including communications encouraging clients to contribute to retirement accounts. By 2024, then-Chair Gensler had directed staff to revise the proposal and seek further public comment. On June 12, 2025, the SEC under incoming Chair Paul Atkins formally withdrew the proposal, along with thirteen other pending rulemakings from the prior administration, stating that any future regulatory action in the area would require beginning the rulemaking process anew (U.S. Securities and Exchange Commission 2025). The prior framework of fiduciary and best-interest standards remains in force, but the proposed extension to predictive analytics has been abandoned without replacement. European regulatory development continues on a different track. Cross-sectional evidence continues to document heterogeneity in what "suitable" or "in the client's best interest" means in practice across platforms and jurisdictions: adequate advice has no settled technical specification (Boreiko and Massarotti 2020; Gaspar and Oliveira 2024).

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's characterization of spam as a "shadow history" of the internet—a site where fundamental questions about legitimate communication, commercialization, and identity have been fought out but rarely made visible as political questions—sets a useful frame for what follows (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—covering who should be able to list, what criteria should apply, and what recourse listed parties had—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—a finding with significant implications for what kinds of interventions could be effective (Levchenko et al. 2011; Rao and Reiley 2012).

The defensive response to industrial-scale spam moved away from collaboratively maintained blocklists and toward 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 (SPF, Sender ID, and later DomainKeys/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 de-

liverability standards and anti-abuse best practices, but its membership is gated and its proceedings are not public. The criteria that determine whether a given sender can reach inboxes are decided 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, 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). Carmi characterizes spam as a historically produced category, where the line between "noise" and legitimate communication is co-produced by standards, legal regimes, interfaces, and measurement practices, not by protocol specifications alone (Carmi 2020).

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 requires different and 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 was open to any operator.

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 Instagram shows in a timeline. Their basic architecture involves retrieving a candidate set from a very large corpus and then ranking candidates using a scoring model, with additional layers for safety, integrity, and content policy (Cobbe and Singh 2019; Seaver 2018). The two-stage design reflects the scale of contemporary platforms. By the 2010s catalogs on the largest platforms had grown to hundreds of millions of items; a single ranking model cannot evaluate the entire corpus per request (Edelson, Haugen, and McCoy 2025; Rieder and Hofmann 2020). These systems run on proprietary infrastructure with no external access to either the candidate retrieval logic or the ranking models (Rieder and Hofmann 2020). Beyond the question of what any individual user sees, they govern the distribution of attention across entire media ecosystems: which creators are viable, which topics circulate, which commercial actors can reach audiences at scale.

Early recommender systems deployed on consumer platforms in the mid-2000s were oriented primarily toward relevance, returning content related to what a user had searched for or previously engaged with. YouTube's early recommendation logic, for example, used click-through rate as a primary signal, directing users toward content that others with similar search behavior had chosen to watch (Seaver 2018). Amazon's early collaborative filtering recommended products that users with similar purchase histories had bought, a logic that made the system's workings relatively legible to the people it affected.

The most widely noted shift in YouTube's history came around 2011–2012, when the platform moved from click-through rate to watch time and session length as its primary optimization targets (Seaver 2018; Xiang 2022). Two accounts of this shift coexist in the literature and in YouTube's own public statements, and they are not mutually exclusive. The first is a user-quality account: click-through rate rewarded misleading thumbnails and titles that generated clicks without delivering on their promise, whereas watch time was a more reliable signal that viewers found content worth completing. The second is a revenue account: watch time and session length convert directly to advertising impressions in a way that click-through rate, measured per video rather than per session, does not (Cobbe and Singh 2019). The shift coincided with a period in which YouTube was working to make its advertising model viable at scale. Both accounts can describe what happened simultaneously; they imply different assessments of whose interests the new optimization target primarily served. The shift sat within a broader technical convergence on behavioral data as the basis for inferring preference at scale. By that point the recommender systems literature had established that implicit behavioral signals carry information about preference that explicit ratings or click-throughs do not capture: they are generated at higher volume, do not suffer from rating-scale ambiguity, and reflect what users actually do rather than what they say (Hu, Koren, and Volinsky 2008). YouTube's own engineering account, published in 2016, documents the operationalization of these principles in a two-stage deep neural network pipeline trained on watch behavior (Covington, Adams, and Sargin 2016). Whether watch time is in fact a better proxy for user satisfaction, or whether it systematically promotes content that is psychologically engaging but not valued in retrospect, has been a persistent and unresolved question in subsequent research (Anwar, Dhillon, and Schoenebeck 2025).

Similar transitions occurred on other platforms. Instagram introduced an algorithmic feed in 2016, replacing reverse-chronological ordering with engagement-based ranking, and subsequently prioritized Reels, its short-form video product, in feed and discovery surfaces (Singh 2023). TikTok built its

core product around an interest-graph model, using engagement signals from each viewing session to construct individualized feeds independent of social connections, a design that proved highly effective at generating prolonged sessions (Wang 2022). Across platforms, the architecture converged on behavioral signals (dwell time, completion rate, shares, comments) as the primary basis for inferring user preference. The infrastructure required to process these signals at scale is, by construction, accessible only to the platforms that own the data.

Empirical audits of contemporary recommender systems have documented systematic effects shaped by platform commercial interests. Amazon's recommender systems favor the platform's own private-label products over comparable third-party offerings, a pattern shaped by Amazon's dual role as marketplace operator and seller on the same platform (Dash et al. 2021). On YouTube, integrity layers introduced under regulatory pressure have been reported to be designed in ways that limit their effect on watch time (Krüger 2023). Multi-objective approaches to recommender design, combining engagement with user-stated values, safety, and other objectives, have been proposed in the technical literature and partially implemented in industry integrity layers, with ongoing debate about whether such configurations meaningfully alter platform-level outcomes (Stray et al. 2024; Krüger 2023). External audits face significant limits when the underlying retrieval and ranking logic is proprietary and the platform controls what experimental access is available (Meßmer and Degeling 2023; Fabbri and Boratto 2025).

Regulatory attention began to accumulate in Europe in the early 2020s. The EU's Digital Services Act (DSA), adopted in 2022, requires very large online platforms to document ranking criteria, conduct risk assessments for recommender systems' societal effects, and offer users at least one recommendation option not based on behavioral profiling (Reviglio and Santoni 2023; Fabbri 2023). The Digital Markets Act (DMA), also adopted in 2022, targets the largest platforms directly with requirements for interoperability and prohibitions on self-preferencing. Early audits of DSA compliance have found limited implementation of the mandated user-choice provisions and significant gaps between documented ranking criteria and what external researchers can observe about recommendation behavior (Fabbri and Boratto 2025; Meßmer and Degeling 2023). Whether the regulatory framework will produce changes beyond initial compliance, and whether the audit tools it creates can access the underlying systems, remains an open question.

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 the statecraft tradition in political science (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 contrasting dynamic. EasyList's community governance in its early phase, the SEC's adversarial rulemaking over fiduciary obligations, and uBlock Origin's structural refusal of the Acceptable Ads framework each represent the active maintenance of those conditions, politicization in the same sense.

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.

Table 1 summarizes each trajectory across the three dimensions of constitutive politics. The agentic AI row is prospective: entries reflect current governance arrangements and deployment patterns rather than a completed trajectory.

Intermediary Institutions Settling Contestation

One might expect the decisive contests over whom agents serve to occur between users and platforms directly. 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 set through gated proceedings with non-public outputs, framed as expert consensus on abuse prevention, 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.

Table 1: Cases across the three dimensions of constitutive politics. The agentic AI row is prospective: entries reflect current governance arrangements and deployment patterns, not a completed trajectory.

Case	Infrastructural	Epistemic	Teleological	Contestation dynamic
Ad blockers	Browser extension APIs; Manifest V3; Chrome engine	Community blocklists (EasyList) vs. industry "acceptable ads" standard	User control vs. advertiser access to attention	Community governance → commercial formalization → embedded in Chrome; countercurrent: uBlock Origin on Firefox
Recommenders	Platform servers; proprietary algorithms; closed APIs	Behavioral surveillance; no community alternative	Engagement vs. well-being and diversity	Depoliticized by design; partial re-opening via DSA / DMA
Robo-advisors	Financial APIs; broker and compliance infrastructure	Risk questionnaires; opaque profiling	Fiduciary duty vs. commercial defaults	Re-opened by SEC / MiFID II; nested re-embedding at implementation level
Spam	SMTP; SPF / DKIM / DMARC; centralized providers	Collaborative blocklists → proprietary ML	Legitimate communication vs. provider stability	Epistemic commons displaced; M3AAWG legitimation of centralized regime
Agentic AI	*MCP / AAIF; open SEP process; no public interest mandate or adversarial mechanism*	*Partial observability; no epistemic commons; independent infrastructure requires open models, APIs, and compute simultaneously*	*Commercial defaults; no fiduciary equivalent; value choices through product decisions*	*Structural conditions being set now; AAIF reproduces absorbed-design features; EU AI Act not designed for agentic systems*

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 each counteract a specific foreclosure mechanism (Bearup 2023): open comment keeps the 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.

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 more aggressively and with greater accuracy than Chrome's Coalition-based enforcement (Snyder, Vastel, and Livshits 2020; Cyphers 2021). The paper's analytical focus is on the location of political capacity over agents. 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 material 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 (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 the first two 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). Public observability was lost; independent infrastructure became insufficient without it. Individual inbox quality improved substantially over this period. The organized capacity to produce shared knowledge about filtering criteria and contest the standards governing them collapsed in the same transition. 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 from their commercial launch onward: the infrastructural basis of the recommendation process, the behavioral data through which the epistemic task was performed, and the objectives the system 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 constituting the organized constituency that would 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 infrastructure through which such differences could have existed was configured by the platform from the outset.

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*: community knowledge is only consequential where it can be operationalized through infrastructure the platform does not control.
3. **Governance that sustains contestation** for collective knowledge to *persist* under adversarial pressure: governance structures without a formal challenge pathway absorb rather than sustain organized capacity to contest what agents do.

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. Across the trajectories, the dominant arrangements did not merely outperform alternatives on the relevant technical task (Nithyanand et al. 2016; Iqbal, Shafiq, and Qian 2017): they displaced the concrete arrangements through which alternatives had been organized. When Chrome redesigned its extension API, the infrastructure through which community filter lists operated was removed from the competitive field. 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 both cases there was nothing weakened to reactivate because the conditions for an alternative were removed alongside the consolidation of the dominant arrangement.

Foreclosure is durable because embedding is concrete: API specifications, data access regimes, institutional standards. Each dimension of contestable choice has two layers, the dimension's purpose and its implementation in specific artifacts and procedures. The two functions named in the preceding section, *material instantiation* and *discursive legitimation*, operate at this implementation layer. Reconfigure that layer and the substantive question the dimension was meant to keep open closes, whatever the formal rules say. This can happen through any dimension. In each of the cases, implementation-level displacement accompanied consolidation.

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 shows the same compounding: proprietary filtering, engagement-based reputation scoring, and centralized infrastructure combine such that engaging any dimension requires engaging the others (Ferrara 2019; Mathew and Cheshire 2017).

The concreteness of embedding also explains the limits of regulatory intervention. When adversarial institutions succeed in politicizing a dimension, actors with enough material control move the contestable choice down to an implementation level the intervention does not reach. 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 pattern extends to litigation: recent product liability claims targeting recommendation algorithms have reached the question of liability for system design while leaving implementation-level discretion over ranking and content exposure largely un-

touched (K.G.M. 2026).

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, arguing in community forums. After the same question was embedded in Chrome's Better Ads enforcement, the path 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). Access asymmetry reinforces the burden: Coalition governance was dominated by industry members, M3AAWG requires organizational resources most affected parties lack, and the Acceptable Ads Committee was funded by the actors it regulated (Mathew and Cheshire 2017). Foreclosure and the threshold for re-politicization accumulate together, through routine decisions that require no visible political act.

Adversarial institutional design prevents foreclosure only where it is coordinated with the material conditions identified in the preceding subsection. The account implies a falsifiability condition: trajectories where multi-dimensional foreclosure was readily reversed, or where 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 asks which features of emerging agentic AI arrangements correspond to configurations the cases show sustaining contestation and which to those under which it was foreclosed.

5 Looking Forward: Agentic AI

Protocol infrastructure across domains. Each of the trajectories analyzed above concerned a discrete layer of intermediation within a specific domain. Agentic AI is perhaps unique in that it could insert a single protocol infrastructure across all such domains at once: the same governance arrangements, default configurations, and epistemic conditions apply across commerce, information retrieval, and task automation. The protocol layer sets the boundaries of possible agency; the choice among agents operates within them. If the framework developed above were to hold for this technology, we would expect an industry-led intermediary institution converting contestable choices into apparently technical specifications, an access-asymmetric governance core layered beneath a discursively open surface, and no adversarial mechanism through which the openness can be defended once consolidation advances.

Institutional configuration in place. 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). On its discursive surface, the AAIF carries the features the historical cases showed such institutions to provide: an open specification, a Specification Enhancement Proposal (SEP) process through which anyone can submit changes, bi-weekly maintainer meetings with public notes, and a neutral institutional home (Parra 2025). Decision-making authority 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 for an affected party to contest an SEP outcome, and no public interest mandate the Board is accountable to. This is the configuration §4 identified as discursive openness without structural protection. It is the Coalition for Better Ads pattern at a higher layer of infrastructure.

Where consequential choices are being settled. At the protocol layer, the questions being decided remain publicly visible in a procedural sense: SEPs are open, meeting notes are published, comments are admissible. Consequential decisions, however, are taken by the Governing Board and the maintainers it endorses. 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 through this configuration. Downstream regulation will operate, but on boundaries already set upstream; the term "agentic systems" does not appear in the EU AI Act's legal text, and the technical standards under development for 2026–2027 will operate within the protocol-level settlements (Oueslati and Staes-Polet 2025). While individual users can configure their permissions, and affected publics can submit comments through the SEP process, the venue with binding authority over consequential outcomes sits inside an institution whose decision-making composition is determined by membership tier.

Preserving contestation. Our framework indicates that the concrete sites where contestation could be lodged are more consequential than broad normative declarations. In agentic AI we can start identifying these: who determines what information an agent may read and how, whether tool invocations and data flows are observable to external auditors, whether SEP outcomes can be challenged on user-harm grounds and by whom. Audit work on shopping agents has documented systematic choice homogeneity, position biases that persist across interface conditions, and model-dependent preferences, establishing that partial observability of agent behavior is empirically achievable and that value-embedding is already detectable in deployed systems (Allouah et al. 2025). Whether the protocol architecture preserves the conditions for such work to scale into collective binding force is an open question. The AAIF's governance charter is being finalized in 2026. The EU AI Act's technical standards for agentic systems are under de-

velopment for adoption in 2027. The default behaviors of commercially deployed shopping agents are accumulating user habituation now. Each of these is a decision that will be substantially harder to revisit once material adoption has advanced: the MiFID II trajectory shows that regulatory intervention arriving after material embedding has accumulated faces nested re-embedding at exactly the levels the regulation cannot reach.

Table 2 sets the arrangements described above against the conditions the trajectories identify. The three dimensions are assessed separately because foreclosure on each proceeds through a different mechanism, while the governance condition is stated once: it is institutional rather than dimension-specific, and in its absence degradation on any dimension can proceed without a venue in which it can be contested.

6 Discussion and Conclusion

Individual Benefits, Collective Capacity, and Contestation. Individual outcomes and collective contestation capacity can move in opposite directions across these trajectories. In spam filtering, inbox quality for ordinary users improved substantially at the same time that the epistemic commons needed for organized challenge collapsed. In recommender systems, recommendation relevance is high in a domain where no organized collective contestation capacity has ever formed. The right to be forgotten illustrates the pattern from a further angle: a user-aligned tool, when enacted as an individual right, can foreclose other parties' legitimate interests, including those who would search, those who would publish, and future users whose informational environment is being shaped.[1] These dual trajectories, and the tension between individual empowerment and other legitimate claims, are precisely what an individual-welfare framework misses. They become visible only when empowerment is understood as the sustained collective capacity to contest the terms under which agents operate, a capacity that necessarily involves conflict among competing collective demands.[2]

Implications for Alignment. One implication concerns alignment. Durable alignment that persists under competitive pressure, across multiple providers, and over time depends on the constitutive conditions that set which alignments are achievable across a domain. Individual alignment work matters, but it operates within boundaries that constrain its scope. Recent sophisticated approaches to alignment acknowledge the diversity of human values and ask how to encode them technically (Sorensen et al. 2024). Incomplete contracting theory identifies a related gap at the level of individual principal-agent relationships: because objectives can never be fully specified in advance, agents inevitably exercise discretion in ways that are sensitive to the structural conditions under which they operate (Hadfield-Menell 2021). The cases examined here show that the same structural problem recurs across entire domains, through the constitutive politics that set the boundaries within which any individual alignment effort works. The robo-advisor case illustrates the dynamic in operation: technically compliant alignment at the individual interaction level coexists with systematic displacement of user interests at the level of default configurations and fund selection. Individual-level and domain-level alignment problems are analytically distinct and call for distinct interventions.

The Limits of Transparency. A growing critical literature has questioned whether transparency alone is adequate 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, absent the structural complements the SEC model illustrates (standing for affected parties, adversarial proceedings, revision authority), can perform what §4 identifies as the discursive legitimation function of intermediary institutions: it makes the cage visible without providing a key.

Limitations. Several limitations bound the analysis. The cases examined are US- and EU-centric; 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. The three dimensions emerged from the trajectories we analyzed, and section 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 agentic AI section in §5 is prospective, identifying structural conditions and the stakes of early decisions while the trajectories themselves remain to unfold. The nested 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 conditions stated in §4 specify the kinds of evidence that would require the mechanism claim to be reconsidered.

Conclusion. The comparison shows that individual benefit and collective political capacity over agents can move in opposite directions, and that the divergence is sustained by institutional and material arrangements that can be located, described, and contested. Agentic AI is the domain in which those arrangements are currently being set. The framework developed here offers a way of identifying where they are being chosen, who participates in choosing them, and how the choices can be kept open to challenge.

[1] Search engines are referenced here through the right-to-be-forgotten case; full case study treatment is beyond the scope of this paper.

[2] The theoretical grounding draws on Mouffe's agonistic political theory, introduced in section 2.

References

Allouah, A.; Besbes, O.; Figueroa, J. D.; Kanoria, Y.; and Kumar, A. 2025. What Is Your AI Agent Buying? Evalua-

Table 2: Agentic AI across the three dimensions, with current arrangements assessed against what the trajectories show is required for collective contestation to be sustained. Entries in the third column are drawn from the cases; they describe configurations observed to sustain or foreclose contestation, not a prescription independent of the analysis.

	Current state	**Sustained contestation requires**
Infrastructural	MCP under the AAIF; capability boundaries set at the protocol layer; Governing Board composition determined by membership tier	*Independent infrastructure*: means of applying community knowledge that platform providers do not control (cf. uBlock Origin on Firefox). Interoperability and access mandates target this dimension (cf. DMA)
Epistemic	Agent behavior only partially observable; no epistemic commons; independent audit requires open-weight models, open APIs, and non-platform compute simultaneously	*Public observability*: the epistemic task performable with information accessible without platform cooperation. Transparency and algorithm-choice mandates target this dimension (cf. DSA), though disclosure alone has not constituted an organized constituency
Teleological	Commercial defaults in deployed shopping agents; no fiduciary equivalent; value commitments settled through product decisions rather than institutional process	Value commitments settled through a process that keeps them visible as contested (cf. SEC fiduciary rulemaking, MiFID II suitability requirements), with contestable revision of default configurations

Across dimensions, *governance that sustains contestation*: a statutory public interest mandate, formal standing for non-commercial parties, adversarial review, and revision authority. Absent these, the material conditions above can be degraded without a venue in which the degradation can be contested; the AAIF currently reproduces the discursive features of open governance without the structural ones.

tion, Biases, Model Dependence, & Emerging Implications for Agentic E-Commerce. arXiv:2508.02630.

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.

Anwar, M. S.; Dhillon, P. S.; and Schoenebeck, G. 2025. Recommendation and Temptation. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, 422–431.

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.

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.

Cyphers, A. M. a. 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.

Dash, A.; Chakraborty, A.; Ghosh, S.; Mukherjee, A.; and Gummadi, K. P. 2021. When the Umpire Is Also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 873–884. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8309-7.

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.

Edelson, L.; Haugen, F.; and McCoy, D. 2025. A Comparative Survey Of Algorithmic Feed Recommendation System Designs. *ACM Transactions on Recommender Systems*, 3757327.

Fabbri, M. 2023. Self-Determination through Explanation: An Ethical Perspective on the Implementation of the Transparency Requirements for Recommender Systems Set by the Digital Services Act of the European Union. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 653–661. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0231-0.

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.

Hadfield-Menell, D. J. 2021. *The Principal-Agent Alignment Problem in Artificial Intelligence*. Ph.D. thesis, UC Berkeley.

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.

Hu, Y.; Koren, Y.; and Volinsky, C. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *2008 Eighth IEEE International Conference on Data Mining*, 263–272.

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 Pusmaz, H. 2018. Designing a Robo-Advisor for Risk-Averse, Low-Budget Consumers. *Electronic Markets*, 28(3): 367–380.

K.G.M. 2026. K.G.M. v. Meta Platforms, Inc. No.23SMCV03371 (Cal. Super. Ct. Mar. 25, 2026) (jury verdict).

Krüger, S. 2023. Have Your Cake and Feed It Forward Too: YouTube, Oral Cravings and the Persistent Question of Media Addiction. *First Monday*.

Levchenko, K.; Pitsillidis, A.; Chachra, N.; Enright, B.; Felegyhazi, 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.

Model Context Protocol Blog. 2025. MCP Joins the Agentic AI Foundation.

Mouffe, C. 2005. *The Return of the Political*, volume 8. Verso.

Mrkývka, P.; and Šiková, 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.; Haddadi, 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)*.

Oueslati, A.; and Staes-Polet, R. 2025. Ahead of the Curve: Governing AI Agents under the EU AI Act. Technical report, The Future Society.

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.

Rieder, B.; and Hofmann, J. 2020. Towards Platform Observability. *Internet Policy Review*, 9(4).

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*.

Singh, D. P. 2023. The Algorithmic Bias of Social Media. *The Motley Undergraduate Journal*, 1(2).

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.; Mireshghallah, 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*.

Stray, J.; Halevy, A.; Assar, P.; Hadfield-Menell, D.; Boutilier, C.; Ashar, A.; Bakalar, C.; Beattie, L.; Ekstrand, M.; Leibowicz, C.; Moon Sehat, C.; Johansen, S.; Kerlin, L.; Vickrey, D.; Singh, S.; Vrijenhoek, S.; Zhang, A.; Andrus, M.; Helberger, N.; Proutskova, P.; Mitra, T.; and Vasan, N. 2024. Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Trans. Recomm. Syst.*, 2(3): 20:1–20:57.

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.

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-Adblocking 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.

A Case Timelines

The following timelines mark the principal technical and institutional developments in each trajectory. Open circles indicate changes to infrastructure, protocols, or technical architecture; filled circles indicate governance, regulatory, or institutional moves.

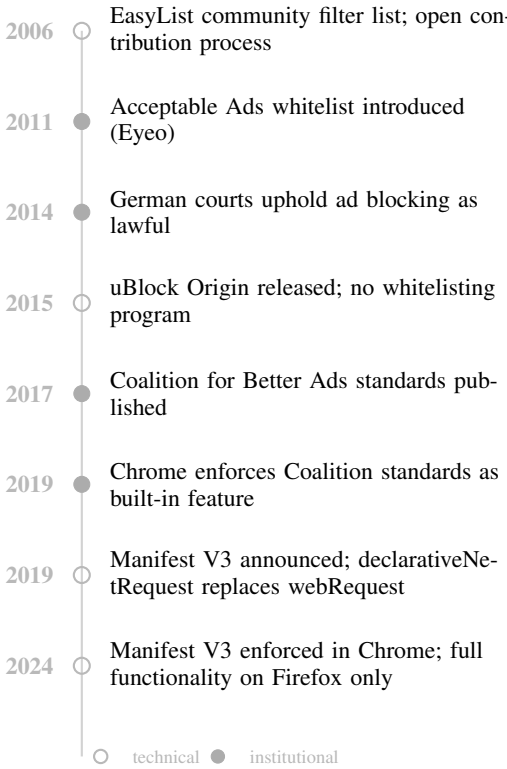

Figure 2: Ad blocker trajectory. The 2011 Acceptable Ads program and 2017 Coalition for Better Ads are the principal institutional structuring moves; Manifest V3 is the infrastructural one. uBlock Origin on Firefox represents the countercurrent maintained outside both.

B Pattern of Depoliticization

Figure 6 summarizes the pattern of institutional design and contestation outcomes observed across the trajectories: the fork is the intermediary institution in which structuring moves are settled, and the paths distinguish configurations where public interest mandates, adversarial challenge mechanisms, and revision authority are present from those where they are absent.

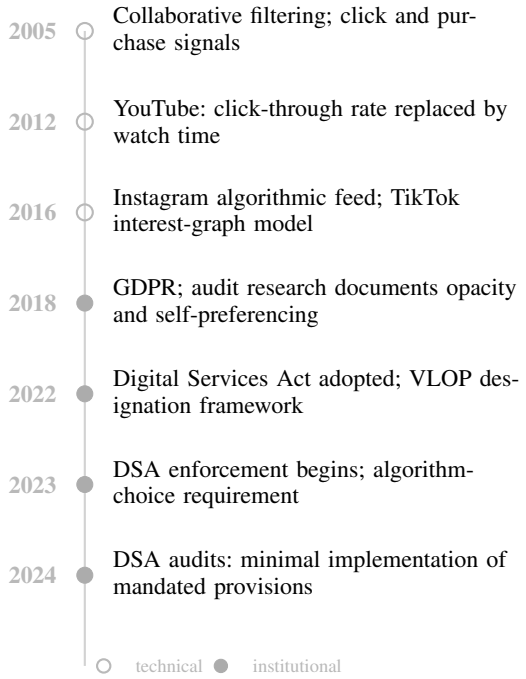

Figure 3: Recommender systems trajectory. Unlike the other three cases, no community-governed period precedes commercial consolidation. Regulatory intervention via the DSA began in 2023; early audits document limited effect on underlying recommendation logic.

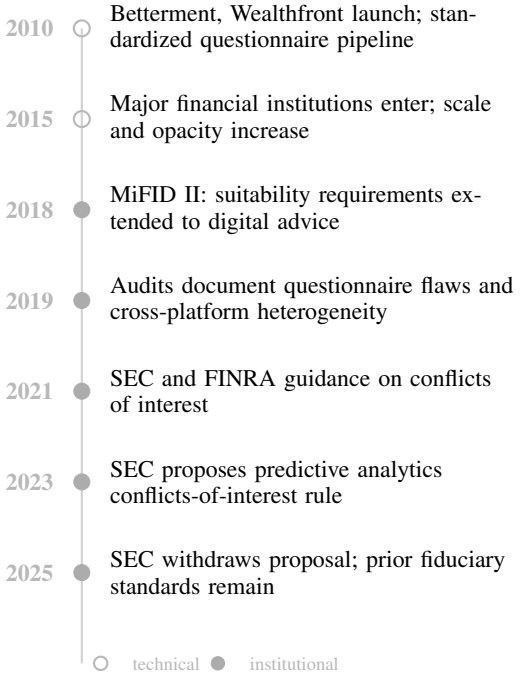

Figure 4: Robo-advisor trajectory. MiFID II in 2018 is the adversarial regulatory intervention the analysis identifies as the closest historical analogue to the structural features currently absent from AAIF governance. The 2025 SEC withdrawal illustrates the fragility of that intervention under changed administration.

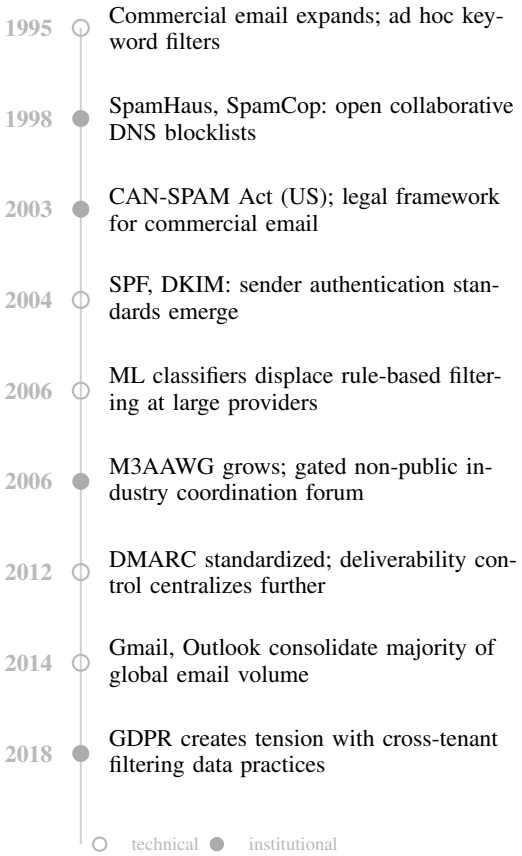

Figure 5: Spam governance trajectory. The 1998 collaborative blocklists represent the epistemic commons whose displacement the analysis traces. The co-occurrence of the 2006 ML classifier shift and M3AAWG institutionalization marks the period in which technical scale advantage and governance access consolidated simultaneously in the same set of large providers.

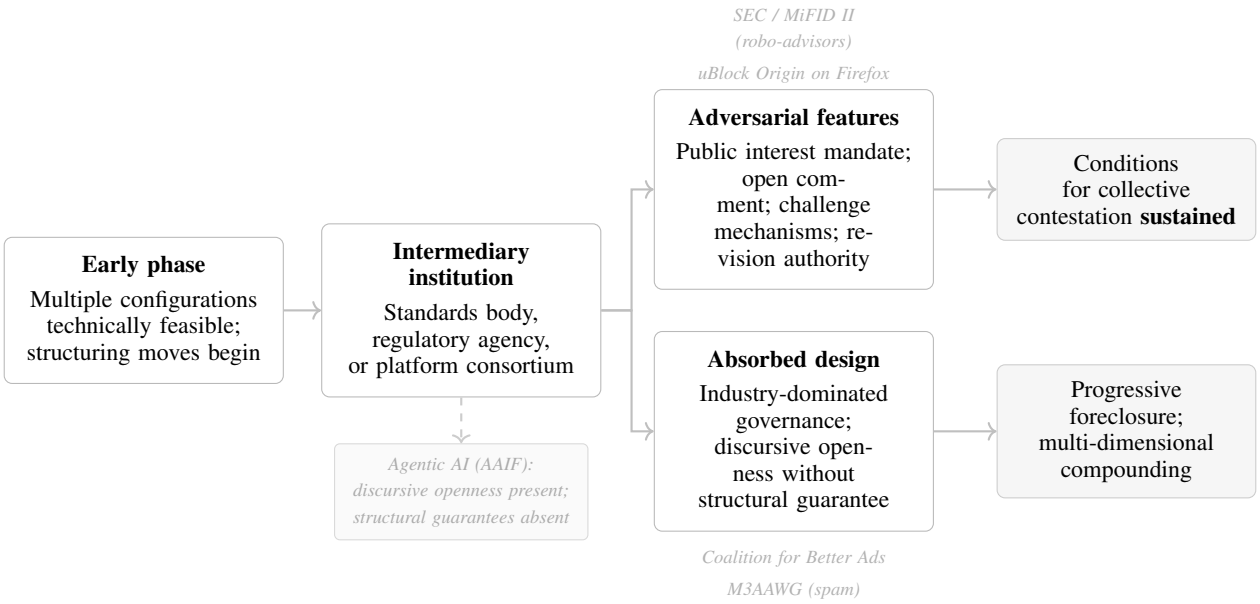

Figure 6: Pattern of institutional design and contestation outcomes across four historical trajectories. The structural features of the intermediary institution where structuring moves are settled determine whether conditions for collective contestation are sustained or progressively foreclosed. The two paths represent configurations observed in the cases; they are tendencies, not stages in a deterministic model. Where adversarial intervention did open a foreclosed dimension (SEC / MiFID II in the robo-advisor case), actors with sufficient infrastructure control moved the contestable choice down to a level the intervention did not reach, a pattern the analysis calls nested re-embedding, also visible in early DSA compliance. Agentic AI (AAIF) is annotated at the fork: the discursive features of the adversarial path are present; the structural features that would protect them are not.