

---

# Position: Build Agent Advocates, Not Platform Agents

---

Sayash Kapoor<sup>\* 1 2</sup> Noam Kolt<sup>\* 3</sup> Seth Lazar<sup>\* 4</sup>

## Abstract

Language model agents are poised to mediate how people navigate and act online. If the companies that already dominate internet search, communication, and commerce—or the firms trying to unseat them—control these agents, the resulting *platform agents* will likely deepen surveillance, tighten lock-in, and further entrench incumbents. To resist that trajectory, this position paper argues that we should promote *agent advocates*: user-controlled agents that safeguard individual autonomy and choice. Doing so demands three coordinated moves: broad public access to both compute and capable AI models that are not platform-owned, open interoperability and safety standards, and market regulation that prevents platforms from foreclosing competition.

## 1. Introduction

Language Model Agents (LMAs) are compound AI systems that can plan and execute complex digital tasks with limited human input. They have become a major research and product focus of the major AI companies (Wang et al., 2024; Zaharia et al., 2024; Casper et al., 2025). Although current performance is still limited (Kapoor et al., 2025; Stroebel et al., 2025), significant progress has been made, especially with the introduction of reasoning models trained with reinforcement learning to undertake more complex tasks (OpenAI, 2024; Google DeepMind, 2025; Anthropic, 2025; DeepSeek-AI et al., 2025). For example, Anthropic, Google DeepMind, and OpenAI have shipped LMAs that can use a web browser or virtual computer to perform a wide range of tasks (Anthropic, 2024; Google DeepMind, 2024; OpenAI, 2025). Most AI companies now offer a ‘Deep Research’ product that enables an LMA to conduct internet research on a topic and produce a corresponding report.

---

<sup>\*</sup>Equal contribution <sup>1</sup>Princeton University <sup>2</sup>Mozilla <sup>3</sup>Hebrew University <sup>4</sup>Australian National University. Correspondence to: Sayash Kapoor <sayashk@princeton.edu>, Noam Kolt <noam.kolt@mail.huji.ac.il>, Seth Lazar <seth.lazar@anu.edu.au>.

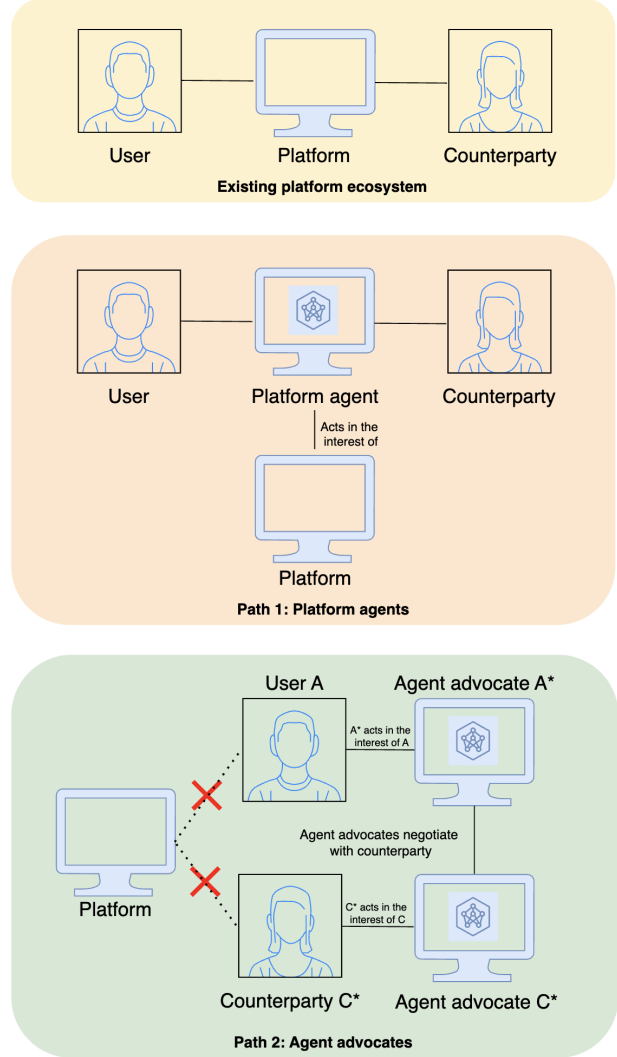


Figure 1: Two development pathways for AI agents: *platform agents* and *agent advocates*. Platform agents intensify many of the concerns of the platform economy. Agent advocates that act in the interests of users (rather than platforms) could cut out the intermediary entirely.

OpenAI, Google, and a number of smaller companies have released software engineering agents that can undertake economically valuable coding work. Enterprise software companies are building agents intended to automate much of customer service (Sierra, 2025).

Platform companies—businesses that intermediate between

users and products, services, and content—dominate the digital economy (van Dijck et al., 2018; Susskind, 2022; Lazar, 2025; Wu, 2025). Meta, Google, Amazon and Microsoft now aspire to develop the most prominent and capable LMAs, or else to steer other companies, such as Anthropic and OpenAI, that are leading the way in AI research (Korinek & Vipra, 2025).<sup>1</sup> Several major AI companies, such as OpenAI and xAI, now aspire to platform status themselves.<sup>2</sup> Today’s platform economy is likely to shape the emerging LMA ecosystem, for at least four reasons. *First*, platform companies have a competitive edge in developing and shipping LMAs due to their unparalleled access to users and user data (including workflows on which to train their agents), as well as to create and deploy agent infrastructure that will cement first-mover advantage. *Second*, users may themselves benefit from the convenience of using LMAs deployed by platforms with which they are already familiar. *Third*, platform companies can acquire or otherwise impede new LMA companies that threaten their market position. *Fourth*, LMAs offer a major opportunity to challenge platform incumbents’ position, retaining the structure of the platform economy while changing those at its head.

By default, therefore, the first generation of successful LMAs will likely be *platform agents*—controlled by incumbent or aspirant platform companies, acting on their behalf, and following their business logic. In particular, platform agents are likely to amplify the power that platform companies already hold, by: enabling aggravated surveillance harms; more decisively controlling the allocation of user attention; extracting additional rent from transactions that they intermediate; and more assertively policing user behavior.

While platform agents are the default path, this outcome is *not* inevitable. LMAs could genuinely advance user interests, not platforms’. In contrast to platform agents, *agent advocates* would serve user interests without competing obligations to platforms. They could run not on the platform’s servers, but on local hardware or an encrypted private cloud, giving users visibility into and control over their actions and data collection. **In this paper, we argue that AI researchers and developers should resist the rise of platform agents and spur the development of user-centric agent advocates.**

We first describe platform power in the broader digital economy (Section 2), then discuss the risks of platform agents as the default path of LMA development (Section 3). In Section 4 we show that agent advocates can mitigate platform agents’ harms, and explore how technical and institutional

challenges to their development can be addressed. In Section 5, we consider objections, and in Section 6 discuss complementary proposals for resisting platform agents. We conclude by advocating for three interventions to support agent advocates: broad public access to compute and independent AI models, open interoperability and technical safety standards, and market regulation that prevents platforms foreclosing competition.

## 2. Background: Platform Power

Digital platforms are intermediaries that enable interactions and exchanges among users and between users and businesses (van Dijck et al., 2018; Van Dijck, 2021; Lazar, 2025). For example, Amazon mediates e-commerce, Google mediates search, and Meta mediates online communication. In exchange, platforms take a ‘cut’ from the interactions that they mediate—often by collecting personal data that provides insights into user behavior (Kolt, 2019) or capturing user attention for advertising (Zuboff, 2019; Hwang, 2020; Benn & Lazar, 2021).

Intermediaries mediate between two or more *principals*. Schematically, intermediaries vary along three graduated dimensions. The first concerns whether the intermediary is a *representative* or a *go-between*. A representative acts on behalf of one principal. A go-between goes back and forth between the principals. For example, a lawyer should be a representative, while matchmakers and more conniving intermediaries like Iago in Shakespeare’s *Othello*, are go-betweens. The second dimension concerns whether the intermediary advances the principal’s interests, or someone else’s (its own or a third party’s). How much, if any, value does the intermediary extract from the transaction that it mediates? The third dimension concerns the degree to which the intermediary *constitutes* the relationships it mediates, by determining what is possible or impossible within it. *Neutral* intermediaries are content-independent pipelines; *constitutive* intermediaries substantially shape the relationships that they intermediate (Lazar, 2025).

Current digital platforms are *constitutive, self-interested go-betweens* that extract value from the principals they mediate between (van Dijck et al., 2018; Van Dijck, 2021; Viljoen et al., 2021). We highlight four key trends: (a) platforms subject users to pervasive surveillance; (b) they establish pseudo-markets that depart from liberal ideals; (c) they substantially determine the allocation of user attention; and (d) they define and enforce policies that shape user behavior.

Ironically, early warnings about the surveillance practices of digital platforms were so hyperbolic and regulatory interventions so ineffective (e.g., cookie banners), that digital surveillance now inspires widespread complacency. But, as a recent FTC report reconfirmed, we are all still exhaustively tracked across the internet, in ways that can readily

<sup>1</sup> See, for example, Wiggers (2024b;a).

<sup>2</sup> For OpenAI, see recent reporting such as Sullivan (2025) and Temkin (2025); for xAI, see the merger of xAI with X (formerly Twitter), the social media platform.

be reconstructed for advertising or other purposes, such as identifying possible targets for immigration enforcement.<sup>3</sup>

Digital platforms initially emerged as two-sided market-makers that mediate between those with specific needs and others who could fulfill them (Rochet & Tirole, 2003; Varian, 2010). This function was critical in the mid 2000s, as rapid internet growth introduced significant challenges around filtering, matching, and trusting online interactions (Benkler, 2006). However, rather than facilitate genuine markets—whether for goods or ideas—in which participants freely choose among impartially presented counterparties, platforms increasingly manipulated these exchanges to extract value from both sides of these transactions (Birch, 2020; Viljoen et al., 2021). In Cory Doctorow’s terms, platforms have “enshittified” (Doctorow, 2023). This goes beyond simply charging a commission or harvesting transactional data such as price signals and consumer behavior patterns. Platforms also shape incentives to advantage favored actors—including themselves (Farronato et al., 2023)—and actively obstruct interactions that might otherwise occur outside their domain (for example, when eBay identifies and prohibits attempts to share contact information with another user through their platform) (Birch, 2020).

Platform companies also shape the distribution of attention. For example, Google dominates online search, with close to 90% of market share (StatCounter, 2024). Meta’s Facebook, Instagram, and WhatsApp account for three of the top four social media platforms, with over 3 billion users (YouTube is ranked #2) (Shopify, 2024).<sup>4</sup> By shaping how individual users’ attention is allocated, platforms influence how people form their beliefs and desires (Susser et al., 2019; Benn & Lazar, 2021). In addition, their impact on the allocation of collective attention profoundly influences societies’ capacity and appetite for collective action (Lazar, 2023).

Digital platforms also exercise significant *governing* power over users—they make, implement, and enforce rules that users are compelled to abide by (Lazar, 2025; Black, 2002). Researchers have long argued that these rules—for example, about which content may be posted online, or how such content will be amplified or ‘de-boosted’—are substantively unjustified and procedurally illegitimate (Gillespie, 2018; Suzor, 2019). In recent years, this concern has spread as users and major political parties have expressed and acted on concerns about illegitimate governance of online speech (Nunziato, 2023).

<sup>3</sup> See Federal Trade Commission (2024); this confirmed years of research. For a comprehensive overview of the relevant literature see Benn & Lazar (2021). For recent reporting on the use of AI for immigration enforcement, see Haskins (2025).

<sup>4</sup> While digital platforms are prevalent worldwide, we focus on their impact in the U.S., primarily because that is the context we are most familiar with, and where major platforms are based.

These platform companies are now racing to build capable LMAs. Younger, AI-first companies such as OpenAI and xAI aspire to *become* platforms—to take market share in search, e-commerce, social media etc. from the established companies. The first highly capable LMAs are likely to be shaped by the structures of the existing platform economy.

### 3. The Default Path: Platform Agents

#### Defining and scoping language model agents (LMAs).

An LMA is an *agent* insofar as it can plan and act (agent is derived from the Latin for “to do”) to achieve goals in complex environments with limited human input (Chan et al., 2023; Shavit et al., 2023; Lazar, 2024; Gabriel et al., 2024; Kapoor et al., 2025; Kolt, 2025). An LMA is a *language model* agent insofar as calls to one or more LLMs play a crucial role in its architecture, typically scaffolded with tools that enable it to perceive, plan, and act in its environment (Sumers et al., 2024; Casper et al., 2025).

Initially, LMAs will operate in narrow, task-specific domains—much like current ‘Deep Research’ agents, though in more domains. For instance, one agent might monitor and reply to work emails, another might order groceries, yet another might keep track of financial markets, perhaps making investments on a user’s behalf. At least from the user’s perspective, these narrow-purpose agents will ultimately be subsumed into a single digital concierge that mediates between users and most or all of the digital environments with which they interact. These *universal intermediaries* (Lazar, 2024) will be able to use any app to perform tasks on behalf of users, not only to retrieve information to inform user action. In doing so, they will draw on their model of the user derived from all of their different interactions, to offer a more personalized and useful service. We expect LMAs to ultimately undertake most tasks for which people currently use the internet—including the procurement of products, services, and content.

**What are platform agents?** Platform agents are LMAs molded by the incentives and selection pressures of the platform economy to function as constitutive go-betweens controlled by incumbent or aspiring platform companies, and designed to act in the platform companies’ interests.

#### 3.1. Risks from platform agents

Today’s algorithmic intermediaries are already self-interested, constitutive go-betweens. Platform agents, due both to their ability to take *actions*, and their capacity for developing a much richer and more nuanced model of the user, would inherit and intensify the aspects of platform power described above.

**Heightened surveillance harms.** Platform surveillance harms users primarily by enabling interventions on their

behavior that are against their own interests. Existing recommenders predict user behavior at scale, making statistical inferences based on behavioral data collected from billions of users. This has been described as a kind of *stochastic manipulation* (Benn & Lazar, 2021): though effective at the margin when applied to massive populations, it is usually harmless at the individual level (see, e.g., very low click-through rates on targeted advertisements (Hwang, 2020)). Platform *agents* will be able to make much more individualized interventions, by drawing on their detailed, even intimate knowledge of their users across many distinct interactions. They won't rely on statistical patterns alone, but will know their users like a close friend. This knowledge, deployed by a self-interested commercial entity, may lead to much more acute manipulation harms than are now feasible.

**Market design.** Platform companies' current attempts to shape quasi-markets in their favor can sometimes be resisted. For example, customers and vendors can occasionally cut out the intermediary and transact directly (e.g., by arranging a second visit to an Airbnb directly with the owner). Similarly, users who invest time in independently researching products, services, or content can escape platforms' walled gardens, albeit paying the cost of inconvenience. As LMA capabilities improve and users rely on platform agents to conduct these everyday transactions, they will have progressively less opportunity to conduct their own basic due diligence, giving platform companies even greater control over their market choices—most obviously by favoring their own products, or those of vendors who buy advertisements (Jefries & Yin, 2021; Fowler, 2022) or strike revenue-sharing deals with the platform.

Relatedly, platform agents could exacerbate the already acute switching costs and network effects of existing platforms, by designing systems to work most effectively with agents from the same platform—as forewarned by some prescient legal scholars (Gal & Elkin-Koren, 2016; Van Loo, 2018). Switching away from a general-purpose digital concierge will be even more difficult than exiting a standalone platform website, because LMAs will know users intimately, learning about them over the course of perhaps years' worth of interactions. This tacit knowledge will afford a vastly better service, locking users in. This is already in train, with OpenAI enabling ChatGPT to reference all past conversations, and Google's Gemini being able to integrate a user's search history.

**User attention and engagement.** Platform agents could more minutely influence users' attention, compared with existing platform technologies. Despite prevalent concerns about filter bubbles and echo chambers in social media (Sunstein, 2001; Pariser, 2011), users can still piece together their media consumption across different platforms and are invariably exposed to some diversity of content (Sunstein, 2001; Pariser, 2011; Bruns, 2019; Cinelli et al., 2021; Sharma

et al., 2024). If platform agents browse the open internet on users' behalf, they will not only control users' information diet, but also how it is presented to them—enabling pervasive editorializing that would be very likely to reshape users' beliefs and desires over time.<sup>5</sup> For example, platform agents could subtly nudge users through dialogue, or manipulate users by expressing approval or disapproval of the users' tastes, decisions, and behaviors (Carroll et al., 2023). Just as platform agents can influence market supply, they can also influence market demand.

**Intensified governance by platform companies.** Platform agents could prove invaluable tools for illegitimately governing users (Gillespie, 2018; Suzor, 2019). By designing guardrails and nudges into LMAs, much as they are now built into LLMs, fine-grained control over people's use of digital technologies will become possible. The more pervasive a platform agent's intermediary role, the greater its ability to monitor our digital actions and interactions for conformity to some set of behavioral standards, as well as its ability to preclude or preempt undesired action (Zittrain, 2008). "Alignment" techniques can, in other words, be used to limit users' freedom of action. While these techniques are not at present robust to adversarial attack, they are sufficiently constraining for most ordinary users that they can empower platform companies to establish and preemptively enforce compliance with platform rules—irrespective of those rules' legitimacy (Ganguli et al., 2023).

### 3.2. Platform agents are the default path.

Platform agents will likely increase the *concentration*, *degree*, and *scope* of platform power. As platform agents become more capable, they will be and used by more people while still being controllable by platform companies and ultimately by those companies' CEOs (concentration). They will be used in more consequential interactions (degree), and across a wider range of actions (scope). We see this as the default pathway for LMAs, for the following reasons:

**Path dependency.** New technologies are likely to be molded by the political economy from which they emerge. Platform companies dominate the digital economy. They have clear incentives to capture the emerging market for LMAs—or else risk losing that dominance.

**User base.** Platform companies are incumbents with built-in user bases. They do not have to compete for an initial audience for LMAs. They can also use their existing user base to obtain feedback, as well as workflow data, that allows building more effective agents.

**Realization of AI investment.** Platform companies have already invested billions in training AI models. They will expect returns on that investment. Platform agents are an

<sup>5</sup> This is already apparent even just with opinionated autocorrect models (Jakesch et al., 2023).



obvious way to realize those returns (Stripe, 2024; LinkedIn, 2024). Some major platforms may even develop platform agents if only to prevent competitors' agents from drawing user revenue and attention away from them.

**Competition and incumbents.** The platform economy favors a small number of large companies. The advent of the agent economy presents an historically unprecedented opportunity to unseat at least some incumbents. This is a powerful incentive for other companies to aim at the near-hegemonic status of existing platform companies. In response, incumbents will likely take steps to defend their market position, acquiring new contenders such as smaller LMA developers where possible, undercutting them and undermining them where not (Mehta, 2024; Wiggers, 2024b).

## 4. Our Proposal: Agent Advocates

**What are agent advocates?** Platform agents are not the only path for LMA development. Instead of agents being a self-interested constitutive go-between controlled by and loyal to platform companies, they could be designed to represent the users who operate them, and act exclusively in their interests. Pursuing this *agent advocates* path would not only forestall platform agents' worst harms, but would also provide users with tools to shift the balance of power in the digital economy away from platforms.<sup>6</sup>

### 4.1. Addressing the risks of platform agents.

**Decreasing surveillance harms.** At a minimum, agents that advance the user's and not the platform's interests would avoid the surveillance harms that platform agents would cause. They could give users meaningful control over their data, either operating locally or in the private cloud. Beyond this, agent advocates could actively *defend* users against platform surveillance. For example, if agent advocates can browse digital platforms on the user's behalf, extracting information to re-present to them, they can deliver those platforms' benefits without the downsides of surveillance and targeted advertising (Lazar et al., 2024). They could also be designed to detect and interdict any attempts by platforms (or others) to track them online.

**Confronting platforms' market design power.** Platforms' market design power derives from their being a go-between that acts for both sides of the transactions that they mediate. If each party to online transactions has their own representative acting in their interests, the scope for this kind of con-fected quasi-market is reduced. Instead of communicating only through a centralized and self-interested intermediary, parties in two-sided markets could communicate with one another through representatives acting in their respective

interests, enabling new approaches to market design. Agent advocates would still depend on *some* matchmaking service, but they would be better-positioned than human users to monitor those matchmakers for predatory activity and to switch to better alternatives, as well as to communicate directly with counterparties. More than just avoiding platform agents' worst outcomes, agent advocates could recalibrate the digital economy towards more competitive markets.

**Reducing platform lock-in.** Current platforms bundle together several different functions (Benkler, 2006; Gillespie, 2018) and create walled gardens around those functions, relying on switching costs and network effects for user retention. Companies that offer competing services (such as DuckDuckGo for search, or Brave as a browser) are at a systematic disadvantage, simply because exiting the platform ecosystem requires more effort on the user's part, and potentially shuts one out from the benefits of joint usage of the same service by others (this is especially true for messaging services). Agent advocates could reduce switching costs by reducing friction in using new services, and could reduce network effects by providing a kind of bottom-up interoperability, hopping over the platforms' walled gardens. Consider, for example, an agent advocate that can collate various social media feeds, or create a group chat that integrates users of different messaging services.

Platforms fragment users, undermining collective action and inducing 'digital resignation' (Draper & Turow, 2019; Benn & Lazar, 2021). Many have advocated for collective action by users to resist this trend—but it does not happen because it is too effortful. Agent advocates could undertake some of the effort of coordinating on users' behalf.

**Changing how user attention translates to revenue.** Platforms determine how content creators are paid and shape how attention is allocated. Agent advocates could enable new approaches to matching attention to revenue. For example, users currently pay for content through subscriptions and a hodgepodge of free trials and other kludges. Users do this instead of paying for each individual piece of content in part because the transaction cost of deciding whether to outlay some small sum for some content is greater than the value of the content itself. Users could outsource this cognitive effort to their agent advocate, to decide whether the content is worth paying for, at what price. LMAs would also be able to provably forget content, so a user's agent can 'view' something on the user's behalf, and if it is not worth paying for, can provably forget it so that the publisher has lost nothing (Weiss et al., 2024). More generally, the owners of large social media websites currently have disproportionate control over the allocation of collective attention; agent advocates could enable users to decide for themselves what to attend to (Lazar, 2023; Lazar et al., 2024). This could of course have some negative implications, but still valuably

<sup>6</sup> We note here complementary work on fiduciary AI by Benthall & Shekman (2023); Reisman & Whitt (2024).

shifts power away from platforms.

**Increased user autonomy.** Platforms today can unilaterally shape how users interact and extract as much value from that interaction as users will tolerate (Wu, 2025), while subjecting them to often dubious and sometimes clearly illegitimate governance. Users endure this unfair bargain because they have no reasonable alternative, and because for all their predations, platforms do offer significant value. Because of this, and because platforms’ rules are, for most users, neither onerous to follow nor likely to result in highly costly penalties, most do not experience their quasi-forced subjection to platform governance as an acute kind of unfreedom. However, platforms’ growing role in users’ lives, and the prospect that platform agents might become a pervasive intermediary to *all* of our digital actions and interactions, suggests that this will change. Agent advocates are a bottom-up response that work not only to extend users’ agency by acting as their representative, but to actively combat platform power. Two parties who can only interact through a self-interested go-between, and who depend on that interaction to some non-trivial degree, are subject to the arbitrary power of that go-between. Two parties who can instead interact through fiduciary AI representatives are not subject to the same arbitrary power.

#### 4.2. Interventions to facilitate the development and adoption of agent advocates

Platform agents and agent advocates present two extreme points on a continuum. The default path skews towards the first extreme, where LMAs function as self-interested go-betweens that extract value from users and exercise power over them. How can we shift the default trajectory towards genuinely user-centric agent advocates that act as faithful representatives of their principal?

**Availability of useful open models.** Small AI companies with no prospect of unseating the platform incumbents might be able to build effective LMAs by integrating calls to closed frontier language models. Companies like OpenAI and Anthropic *could* in this way support the development of agent advocates, instead of pursuing the platform agent path. But that is too fragile a foundation for the future of the digital economy. Open weights models are likely essential to reduce exposure to the risks of a reversal in corporate strategy, and to give users real guarantees that an agent advocate is acting only in their interests (Kapoor et al., 2024) (technical barriers are discussed in Section 5.2).<sup>7</sup>

Currently, platform companies like Meta and Google and Chinese companies DeepSeek and Alibaba (itself a platform company) lead the way in open AI development (Meta,

2024; DeepSeek-AI et al., 2025; Team, 2025). Mistral (Mistral AI team, 2024) and Cohere (Cohere, 2024) also offer relatively competitive open models; AI2, a nonprofit research institute, though behind in terms of capabilities, has shown the returns on making models, code, and data entirely open (Groeneveld et al., 2024). The open foundation model ecosystem is today in better health than ever before, and yet access to the most capable models still depends on the whims of Silicon Valley executives and, for Chinese models, ultimately the leadership of the CCP. This is an unstable position for agent advocates. Meta and Google, for example, are unlikely to look favorably on LMAs that use their models to draw revenue away from their platforms—indeed, these will likely be prohibited by the license under which these models are provided. Chinese models, meanwhile, will have to comply with rules promulgated by the CCP. At present this kind of censorship can be circumvented (AI Team, 2025), but that may not always be so.

Expanding the availability of open models is an urgent imperative. Liberal democracies should create a public complement to the private industry leaders in frontier AI research. Researchers advocating a ‘CERN for AI’ often aspire primarily to help states better understand the risks of advanced AI, and unlock the societal benefits of ‘superintelligent’ AI systems (Brundage, 2024). While these are important initiatives, public AI could also be leveraged to prevent corporate or political control over advanced AI agents that would otherwise have severe negative implications for individual and collective autonomy and well-being (Public AI Network, 2024; Marda et al., 2024).

In addition, while we might hope that one day agent advocates might be powered by open models running on device, this is likely to be technically infeasible for some time. And if the computational demands of effective agent advocates continue to exceed what can feasibly be deployed locally, then the provision of affordable computational power (compute) will be as important as the availability of public interest AI models.

**Bolstering competitive markets.** Agent advocates will threaten platform companies’ market dominance. Those companies will restrict agent advocates however they can. Competition regulators will need to intervene to prevent this. For example, they could disbar digital platforms from blocking AI agents from their GUIs, and perhaps even require them to make APIs available for appropriately authorized agents to access. This intervention could be motivated on similar grounds to the ‘net neutrality’ principle (Wu, 2003) or earlier attempts to enforce interoperability (Wegner, 1996). This would enhance competition by giving users a real threat of exit, and by changing how advertising revenue is distributed. It would also incentivize competitors to produce better AI models or make better use of existing models to have a chance at the new value created by LMAs.

<sup>7</sup> Of course, we assume that few users will build their own agents; but the availability of capable open models will allow small companies to develop user-centric agent advocates.

Similarly, governments should explore the application of existing regulations that protect data portability to LMAs from the outset, ensuring that users are able to switch away from a given provider whenever they wish to.

**Operational and regulatory infrastructure.** By centralizing control, platforms make effective governance of users comparatively feasible; indeed, Tarleton Gillespie has argued that effective governance is the central product that platforms provide (Gillespie, 2018). A decentralized economy of agent advocates would be more liberal than a platform agent economy. But it is also riskier. For example, how can a user trust that their advocate is robustly advancing their interests? When agents transact with one another, what assurance do users have that they are not colluding, or that one has not manipulated the other? A decentralized economy of agent advocates will require novel infrastructure to ensure it is safe and fair (Chan et al., 2025).

You cannot govern what you cannot identify. Agent advocates will require infrastructure to provide agent instances with **credentials** (Chan et al., 2024; Adler et al., 2024; Chaffer et al., 2025; South et al., 2025) that certify that they meet certain minimum criteria to operate autonomously online. These credentials should also provide key information such as the agent’s underlying architecture, its core capabilities, the tools to which it has access, the types of actions it can take, and some identifier (likely anonymized) that attributes the agent’s actions to some upstream individual or company who can be held to account for its actions. They could also include credentials that certify the agent’s adherence to a fiduciary responsibility to the principal.

Platforms create network effects. To build an agent economy without platforms, there must be protocols that enable interoperability (Masnick, 2019). These would include **communications protocols** for agents (even though they can communicate in natural language this is unlikely to be maximally efficient (Marro et al., 2024)). Open communication protocols can prevent the establishment of walled gardens by platform companies. But they can also be used as mechanisms for governance, to incentivize credentialing. For example, users would be able to authorize their agents only to communicate with other credentialed agents.

An internet of agent advocates will be a vast market constituted by billions of microtransactions. Because directly regulating these microtransactions will be impractical, digital **clearinghouses** will be needed to secure agent transactions, and both protect the principals’ interests against collusion and fraud, and monitor for and mitigate negative externalities. Clearinghouses will ensure that agents do not manipulate or prompt inject each other, and provide means for recourse in the event of misconduct. This function could potentially be incorporated into communication protocols, but more likely would need to be routed through indepen-

dent, secure servers, on a similar model to the existing payments infrastructure for online transactions. Clearinghouses would provide further opportunity for bottom-up, market-based governance of LMAs, as agent credentials could be a necessary condition for transacting, and users could specify that their agents are not permitted to transact with other agents with specific properties—for example, you might instruct your agent to transact only with other agents of the same capability level (this might help reduce the risk of an arms-race dynamic in the open agent economy).<sup>8</sup>

As the development of these protocols and clearinghouses progresses, their implementation and promotion should be supported by **technical standards**. These standards, which would be comparable to existing standards for networking and cybersecurity (IETF, 2024), could be promulgated either by public bodies, such as the National Institute of Standards and Technology (NIST), or private bodies, such as the International Organization for Standardization (ISO).

The successful adoption of agent advocates requires **mitigating the risk of misuse and misalignment**. While developing robust credentials and protocols would facilitate the creation of agent advocates, they will not address the risk of malicious actors using agent advocates for nefarious purposes, or agents that malfunction or are misaligned with the interests of users or broader societal values. Mechanisms will need to be established to detect and disrupt agent advocates in such cases, which could potentially be achieved by developing defensive agents that identify and intercept malign agents.

## 5. Alternative Views

### 5.1. Counterargument 1: Platform agents are not a problem.

**Development and deployment bottlenecks.** Despite growing research and investment in LMAs, these systems may not become sufficiently capable or reliable for large-scale adoption, whether by platforms or other actors. Key bottlenecks in LMA development and deployment include: (i) the exceptionally high levels of reliability needed to perform complex sequential actions with only limited human involvement; (ii) the ability to generalize robustly and act effectively in diverse digital environments; (iii) the need to collect or create training data comprised of ‘agentic trajectories’ (i.e., state-action sequences); (iv) expensive operating costs (Kapoor et al., 2025; Stroebel et al., 2025).

**Market responses.** Even if LMAs become highly performant and sufficiently cheap, the risks considered above

<sup>8</sup> Would clearinghouses potentially become the new platforms? We don’t think so. Platforms are defined by their bundling of different functions; clearinghouses would play a discrete role of providing bottom-up market-based governance.



may not materialize if other market actors (e.g., online vendors, news sites) obstruct platform agents. Such measures may be technical (e.g., protocols resembling robots.txt and LLMs.txt adapted to agents specifically) or legal (e.g., access and licensing agreements with platform companies). Platform companies may also refrain from deploying LMAs as a B2C product that intermediates between consumers and digital services and, instead, integrate LMAs into their internal business operations, or use them to support external B2B services (e.g., managing platform-vendor relationships).

**Advantages of platform agents.** Platform agents may offer advantages over agent advocates that counterbalance their risks. For example, platform agents are likely to be comparatively convenient, due to the combination of data, compute, engineering talent, and safety and security practices, available to platform companies. Accordingly, just as hundreds of millions of consumers currently use platform-operated chatbots (Backlinko, 2025) despite the availability of non-platform alternatives, they may also prefer to use platform-controlled AI agents over agent advocates.

***Our response.** Even if platform agents do not develop precisely as discussed in Section 3, the prospect of even some platform-based risks arising from LMAs should nevertheless motivate the promotion of agent advocates. There is clearly a spectrum between platform agents and agent advocates; we argue for moving as far along that spectrum towards advocates as is feasible, subject to appropriate safety standards.*

## 5.2. Counterargument 2: Agent advocates are not feasible.

**Technical and safety barriers.** While researchers have explored the notions of ‘loyal AI’ and ‘fiduciary AI’ (Aguirre et al., 2020; Benthall & Shekman, 2023) and built systems that follow user instructions (Ouyang et al., 2022; Zhou et al., 2023), the problem of ensuring an agent’s loyalty to its principal is far from solved. The AI alignment literature (Russell, 2019; Christian, 2020; Ngo et al., 2022) illustrates myriad ways in which agents may fail to comply with user instructions, or comply in undesirable ways. Agent advocates are not immune to these problems, and we need to expand ways to control agents’ actions. For these and other reasons, some researchers advocate against building autonomous agents altogether (Cohen et al., 2024; Mitchell et al., 2025; Bengio et al., 2025).

**Institutional concerns.** We argue that platform companies’ profit incentive will drive them to develop self-advantaging LMAs that, at times, harm consumers; it will also drive other AI companies to aspire to build platform agents instead of agent advocates. The same incentives will also influence companies that develop and deploy agent advocates. Even if those companies initially develop user-centric agents, they may decide to change direction in the future,

perhaps building platform agents of their own.

**Past failures.** Prior attempts to challenge digital platforms and decentralize the control of digital technologies have often failed (Narayanan et al., 2012). Efforts to develop agent advocates might meet a similar fate.

***Our response.** Agent advocates face serious obstacles. For this reason we propose the interventions outlined in Section 4.2. In addition, LMAs present more favorable conditions for mitigating platform power than today’s internet infrastructure, since (a) They can both escape network effects and actively undermine them; and (b) User-centric companies can unilaterally build LMAs that, with sufficiently capable models, will undermine the power of platform companies.*

## 5.3. Counterargument 3: Agent advocates fail to address important problems.

**Malicious applications.** LMAs can be used for various malicious applications, including fraud, cybercrime, harassment, and weapons development (Andriushchenko et al., 2025; Li et al., 2024; Tur et al., 2025). Despite attempts to curtail such applications, jailbreaking is widespread and appears likely to persist (Wei et al., 2023; Xu et al., 2024). Developing agent advocates does not address this pressing issue, and may even embolden malicious actors.

**Multi-agent interactions.** LMAs’ most significant challenges may arise not from individual user-system interactions, but from multi-party interactions between different AI systems, human actors, institutions, and environments (Dafoe et al., 2020; 2021; Hammond et al., 2025). These include collusion, collective action problems, emergent network effects, and correlated or cascading failures (Motwani et al., 2024; Anwar et al., 2024; Kolt et al., 2025). While user-centric agent advocates do not necessarily exacerbate these concerns, they do not address them either.

**Broader societal impacts.** LMAs could have significant impacts on social and economic life. For example, interpersonal encounters might be increasingly mediated or even altogether obviated by LMAs (Lazar, 2024; Kolt, 2025). Certain jobs or workplace tasks might be outsourced to LMAs, either displacing people or altering the kind of work they undertake (Sellen & Horvitz, 2024).

***Our response.** Agent advocates will not solve all the problems of LMAs; they specifically target harms from platform agents. Supported by appropriate safety and governance frameworks, agent advocates are a pareto improvement on the default trajectory described in Section 3.*

## 6. Expanding the Solution Space

Agent advocates are not the only way to address risks from platform agents. Other approaches can either complement agent advocates or operate effectively even in their absence.



We focus on proposals that aim to preempt such harms, rather than redress them after the fact.

**Comprehensive AI regulation.** One potential approach to mitigating platform agent risks is to pass comprehensive AI regulation that prohibits or severely curtails their most harmful consequences. For example, bans could be placed on the most intrusive surveillance practices and forms of user manipulation (discussed in Section 3 above). Prohibitions like these have already been established in the European Union in Article 5 of the EU AI Act (European Parliament, 2024). *While such provisions could in theory address certain risks from platform agents, the prospect of the U.S. passing federal legislation comparable to the EU AI Act is highly unlikely.*

**Platform regulation.** Regulation could instead target platform companies, rather than AI technologies more generally. For example, large platforms could be prohibited from controlling agents that intermediate between users and particular services, perhaps in sensitive settings such as banking and healthcare. Alternatively, platforms might be subjected to rigorous transparency requirements and external audits. *While the EU’s Digital Services Act (European Parliament, 2022) offers a model for such regulation, the U.S. has largely refrained from regulating platform companies (e.g., social media), and appears unlikely to change course.*

**Consumer privacy law.** LMAs will both rely on and generate large quantities of private data. Data protection regulation could offer an effective, if indirect, strategy to address risks from platform agents (Gupta et al., 2024). Imposing strict controls on the uses of personal data for training or operating LMAs or requiring real-time informed consent from users could curtail certain undesirable consequences. While there are currently no broadly applicable federal consumer privacy laws in the U.S., some large states, including California, have passed noteworthy consumer privacy laws (State of California, 2018). *Whether such laws could be amended or otherwise tailored to tackle the distinct challenges posed by LMAs is, for now, an open question.*

**Antitrust law and enforcement.** In the absence of new regulation, existing antitrust law could offer a strategy for addressing platform agent risks (Narechania & Sitaraman, 2024). Companies that undertake some combination of developing chips, providing cloud computing, training frontier models, building products (including LMAs), and operating consumer and enterprise platforms have a substantial advantage over companies that operate at only one point in the value chain (Korinek & Vipra, 2025). Constraining such vertical integration could significantly bolster competition in the LMA industry, allowing new entrants to better compete. *A key obstacle to this approach is that the U.S. (in contrast to the EU) does not have a history of successfully taking major antitrust enforcement against platform companies or*

*digital services providers more broadly, although this may yet change.*

## 7. Conclusion

Computing amplifies agency. In the hands of the powerful, it reinforces centralized control. In individuals’ hands, it can enable counter-power. Historically, recurrent moments of technological expansion have seemed poised to usher in a more decentralized computing future. Each time, however, centralizing forces have reasserted themselves. Examples abound: the first hackers circumventing the gatekeepers of MIT’s PDP-6; the Silicon Valley Homebrew Club building alternatives to IBM’s mainframes; open, customizable software vs. closed operating systems; community-run BBSs vs. centralized ISPs; the open internet standing against the internet of platforms, and more.

Our current moment is not unique. It may, however, present a unique opportunity. Previously, the pathway toward decentralization was accessible primarily to technologically skilled users—hackers capable of circumventing constraints set by centralized authorities. Today, however, user-centric agent advocates could level the playing field.

By default, AI agents are likely to follow the same centralized pattern as the platform economy. Incumbent and aspiring platform companies will develop and control powerful agentic systems. These *platform agents* will intermediate digital interactions across countless personal and professional contexts. Although users may guide platform agents, ultimate control will remain firmly with centralized developers. Platform-controlled AI agents will be double agents, with the potential for profoundly negative implications: heightened surveillance harms, constrained user choice, granular market manipulation, and broad illegitimate power. The worst of platform capitalism’s current ills could be exacerbated.

But this outcome is not inevitable. A compelling alternative exists: user-centric *agent advocates* designed to serve the interests of individual users, not platform companies. Representatives, not go-betweens, that reject platform logic. Agent advocates could provide a path to harnessing the promise of AI agents without succumbing to platform-based control. Realizing this decentralized alternative will require targeted technical and institutional interventions. These include ensuring the availability of open-source models and public computational resources, as well as establishing robust safety standards and governance frameworks. It will also require engineers who can build highly capable universal intermediaries but resist entering the race to create the next platform. Independent researchers and developers must prioritize addressing these challenges now—before the default pathway locks in.

## Acknowledgments

We are grateful to Markus Anderljung, Matt Boulos, Alan Chan, Simon Goldstein, Lewis Hammond, Arvind Narayanan, Jake Stone, and members of the MINT lab at the Australian National University for feedback on a draft of this paper. SL acknowledges funding from the Australian Research Council (Grant FT210100724) and the Templeton World Charity Foundation.

## References

- Adler, S., Hitzig, Z., Jain, S., Brewer, C., Chang, W., DiResta, R., Lazzarin, E., McGregor, S., Seltzer, W., Siddarth, D., et al. Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online. *arXiv preprint arXiv:2408.07892*, 2024.
- Aguirre, A., Dempsey, G., Surden, H., and Reiner, P. B. Ai loyalty: a new paradigm for aligning stakeholder interests. *IEEE Transactions on Technology and Society*, 1(3):128–137, 2020.
- AI Team. Today we’re open-sourcing r1 1776, a version of the DeepSeek-R1 model that has been post-trained to provide unbiased, accurate, and factual information, 2 2025. URL <https://perplexity.ai/hub/blog/open-sourcing-r1-1776>.
- Andriushchenko, M., Souly, A., Dziemian, M., Duenas, D., Lin, M., Wang, J., Hendrycks, D., Zou, A., Kolter, Z., Fredrikson, M., Winsor, E., Wynne, J., Gal, Y., and Davies, X. AgentHarm: A Benchmark for Measuring Harmfulness of LLM Agents. 2025. doi: 10.48550/arXiv.2410.09024.
- Anthropic. Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku, 2024. URL <https://www.anthropic.com/news/3-5-models-and-computer-use>.
- Anthropic. Claude 3.7 Sonnet and Claude Code, 2025. URL <https://www.anthropic.com/news/claude-3-7-sonnet>.
- Anwar, U., Saparov, A., Rando, J., Paleka, D., Turpin, M., Hase, P., Lubana, E. S., Jenner, E., Casper, S., Sourbut, O., Edelman, B. L., Zhang, Z., Günther, M., Korinek, A., Hernandez-Orallo, J., Hammond, L., Bigelow, E. J., Pan, A., Langosco, L., Korbak, T., Zhang, H. C., Zhong, R., hEigearthaigh, S. O., Recchia, G., Corsi, G., Chan, A., Anderljung, M., Edwards, L., Petrov, A., Witt, C. S. d., Motwani, S. R., Bengio, Y., Chen, D., Torr, P., Albanie, S., Maharaj, T., Foerster, J. N., Tramèr, F., He, H., Kasirzadeh, A., Choi, Y., and Krueger, D. Foundational Challenges in Assuring Alignment and Safety of Large Language Models. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL <https://openreview.net/forum?id=oVTkOs8Pka>.
- Backlinko. Chatgpt / openai statistics: How many people use chatgpt?, 2025. URL <https://backlinko.com/chatgpt-stats>. Accessed: 2025-01-30.
- Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., Mindermann, S., Oberman, A., Richardson, J., Richardson, O., et al. Superintelligent agents pose catastrophic risks: Can scientist ai offer a safer path? *arXiv preprint arXiv:2502.15657*, 2025.
- Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.
- Benn, C. and Lazar, S. What’s wrong with automated influence. *Canadian Journal of Philosophy*, pp. 1–24, 2021. doi: 10.1017/can.2021.23.
- Benthall, S. and Shekman, D. Designing fiduciary artificial intelligence. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–15, 2023.
- Birch, K. Automated neoliberalism? the digital organisation of markets in technoscientific capitalism. *New Formations*, 100:10–27, 2020. doi: 10.3898/NewF:100-101.02.2020.
- Black, J. Critical reflections on regulation. *Australasian Journal of Legal Philosophy*, 27(2002):1–46, 2002. ISSN 1440-4982.
- Brundage, M. My recent lecture at Berkeley and a vision for a “CERN for AI”, December 2024. URL <https://milesbrundage.substack.com/p/my-recent-lecture-at-berkeley-and>.
- Bruns, A. *Are filter bubbles real?* John Wiley & Sons, 2019.
- Carroll, M., Chan, A., Ashton, H., and Krueger, D. Characterizing manipulation from ai systems. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–13, 2023.
- Casper, S., Bailey, L., Hunter, R., Ezell, C., Cabalé, E., Gerovitch, M., Slocum, S., Wei, K., Jurkovic, N., Khan, A., et al. The ai agent index. *arXiv preprint arXiv:2502.01635*, 2025.
- Chaffer, T. J., II, C. v. G., Okusanya, B., Cotlage, D., and Goldston, J. Decentralized Governance of Autonomous AI Agents, January 2025. URL <http://arxiv.org/abs/2412.17114>. arXiv:2412.17114 [cs].

- Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krashennikov, D., Langosco, L., He, Z., Duan, Y., Carroll, M., Lin, M., Mayhew, A., Collins, K., Molamohammadi, M., Burden, J., Zhao, W., Rismeni, S., Voudouris, K., Bhatt, U., Weller, A., Krueger, D., and Maharaj, T. Harms from Increasingly Agentic Algorithmic Systems. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pp. 651–666, New York, NY, USA, June 2023. Association for Computing Machinery. ISBN 9798400701924. doi: 10.1145/3593013.3594033. URL <https://dl.acm.org/doi/10.1145/3593013.3594033>.
- Chan, A., Kolt, N., Wills, P., Anwar, U., Witt, C. S. d., Rajkumar, N., Hammond, L., Krueger, D., Heim, L., and Anderljung, M. IDs for AI Systems. 2024. doi: <https://doi.org/10.48550/arXiv.2406.12137>.
- Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K., and Anderljung, M. Infrastructure for AI Agents. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL <https://openreview.net/forum?id=Ckh17xN2R2>.
- Christian, B. *The Alignment Problem: Machine Learning and Human Values*. WW Norton & Company, 2020.
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrocioni, W., and Starnini, M. The echo chamber effect on social media. *Proceedings of the national academy of sciences*, 118(9):e2023301118, 2021.
- Cohen, M. K., Kolt, N., Bengio, Y., Hadfield, G. K., and Russell, S. Regulating advanced artificial agents. *Science*, 384(6691):36–38, 2024.
- Cohere. Cohere’s Command R+ Model (Details and Application), 2024. URL <https://docs.cohere.com/v2/docs/command-r-plus>.
- Dafae, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K. R., Leibo, J. Z., Larson, K., and Graepel, T. Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*, 2020.
- Dafae, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K., and Graepel, T. Cooperative AI: machines must learn to find common ground. *Nature*, 593(7857):33–36, May 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/d41586-021-01170-0. URL <https://www.nature.com/articles/d41586-021-01170-0>.
- DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., Xue, B., Wang, B., Wu, B., Feng, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Ding, H., Xin, H., Gao, H., Qu, H., Li, H., Guo, J., Li, J., Wang, J., Chen, J., Yuan, J., Qiu, J., Li, J., Cai, J. L., Ni, J., Liang, J., Chen, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Zhao, L., Wang, L., Zhang, L., Xu, L., Xia, L., Zhang, M., Zhang, M., Tang, M., Li, M., Wang, M., Li, M., Tian, N., Huang, P., Zhang, P., Wang, Q., Chen, Q., Du, Q., Ge, R., Zhang, R., Pan, R., Wang, R., Chen, R. J., Jin, R. L., Chen, R., Lu, S., Zhou, S., Chen, S., Ye, S., Wang, S., Yu, S., Zhou, S., Pan, S., Li, S. S., Zhou, S., Wu, S., Ye, S., Yun, T., Pei, T., Sun, T., Wang, T., Zeng, W., Zhao, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Xiao, W. L., An, W., Liu, X., Wang, X., Chen, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yang, X., Li, X., Su, X., Lin, X., Li, X. Q., Jin, X., Shen, X., Chen, X., Sun, X., Wang, X., Song, X., Zhou, X., Wang, X., Shan, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Zhu, Y. X., Xu, Y., Huang, Y., Li, Y., Zheng, Y., Zhu, Y., Ma, Y., Tang, Y., Zha, Y., Yan, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Ma, Z., Yan, Z., Wu, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Pan, Z., Huang, Z., Xu, Z., Zhang, Z., and Zhang, Z. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, January 2025. URL <http://arxiv.org/abs/2501.12948>. arXiv:2501.12948 [cs].
- Doctorow, C. The ‘Enshittification’ of TikTok. *Wired*, 2023. ISSN 1059-1028. URL <https://www.wired.com/story/tiktok-platforms-cory-doctorow/>. Section: tags.
- Draper, N. A. and Turow, J. The corporate cultivation of digital resignation. *New Media & Society*, 21(8):1824–1839, 2019. doi: 10.1177/1461444819833331. URL <https://journals.sagepub.com/doi/abs/10.1177/1461444819833331>.
- European Parliament. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 2022. URL <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>. Official Journal of the European Union, L 277, 27 October 2022, pp. 1–102.
- European Parliament. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008,*

- (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). June 2024. URL <http://data.europa.eu/eli/reg/2024/1689/oj/eng>. Legislative Body: CONSIL, EP.
- Farronato, C., Fradkin, A., and MacKay, A. Self-preferencing at amazon: evidence from search rankings. In *AEA Papers and Proceedings*, volume 113, pp. 239–243. American Economic Association 114 Broadway, Suite 305, Nashville, TN 37203, 2023.
- Federal Trade Commission. FTC Staff Report Finds Large Social Media and Video Streaming Companies Have Engaged in Vast Surveillance of Users with Lax Privacy Controls and Inadequate Safeguards for Kids and Teens, September 2024. URL <https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-staff-report-finds-large-social-media-video-streaming-companies-have-engaged-vast-surveillance>.
- Fowler, G. A. It’s not your imagination: Shopping on Amazon has gotten worse. *The Washington Post*, 2022. URL <https://www.washingtonpost.com/technology/interactive/2022/amazonshopping-ads/>.
- Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., Mateos-Garcia, J., Weidinger, L., Street, W., Lange, B., Ingerman, A., Lentz, A., Enger, R., Barakat, A., Krakovna, V., Siy, J. O., Kurth-Nelson, Z., McCroskery, A., Bolina, V., Law, H., Shanahan, M., Alberts, L., Balle, B., de Haas, S., Ibitoye, Y., Dafoe, A., Goldberg, B., Krier, S., Reese, A., Witherspoon, S., Hawkins, W., Rauh, M., Wallace, D., Franklin, M., Goldstein, J. A., Lehman, J., Klenk, M., Vallor, S., Biles, C., Morris, M. R., King, H., Arcas, B. A. y., Isaac, W., and Manyika, J. The Ethics of Advanced AI Assistants, April 2024. URL <http://arxiv.org/abs/2404.16244>. arXiv:2404.16244 [cs].
- Gal, M. S. and Elkin-Koren, N. Algorithmic consumers. *Harv. JL & Tech.*, 30:309, 2016.
- Ganguli, D., Askell, A., Schiefer, N., Liao, T. I., Lukošiušė, K., Chen, A., Goldie, A., Mirhoseini, A., Olsson, C., Hernandez, D., Drain, D., Li, D., Tran-Johnson, E., Perez, E., Jackson, K., Kerr, J., Mueller, J., Landau, J., Ndousse, K., Nguyen, K., Lovitt, L., Sellitto, M., Nelson, E., Mercado, N., DasSarma, N., Rausch, O., Lasenby, R., Larson, R., Ringer, S., Kundu, S., Kada-vath, S., Johnston, S., Kravec, S., Sheer El, S., Lan-ham, T., Telleen-Lawton, T., Henighan, T., Hume, T., Bai, Y., Hatfield-Dodds, Z., Mann, B., Amodei, D., Nicholas, J., McCandlish, S., Brown, T., Olah, C., Clark, J., Bowman, S. R., and Kaplan, J. The capacity for moral self-correction in large language models. *arXiv preprint*, <https://arxiv.org/abs/2302.07459>; <https://arxiv.org/abs/2302.07459>, 2023. ISSN 2331-8422.
- Gillespie, T. *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, New Haven, 2018. doi: 99976897003.
- Google DeepMind. Project Mariner, December 2024. URL <https://deepmind.google/technologies/project-mariner/>.
- Google DeepMind. Gemini thinking, 2025. URL <https://ai.google.dev/gemini-api/docs/thinking>.
- Groeneveld, D., Beltagy, I., Walsh, E., Bhagia, A., Kinney, R., Tafjord, O., Jha, A., Ivison, H., Magnusson, I., Wang, Y., Arora, S., Atkinson, D., Authur, R., Chandu, K., Cohan, A., Dumas, J., Elazar, Y., Gu, Y., Hessel, J., Khot, T., Merrill, W., Morrison, J., Muennighoff, N., Naik, A., Nam, C., Peters, M., Pyatkin, V., Ravichander, A., Schwenk, D., Shah, S., Smith, W., Strubell, E., Subramani, N., Wortsman, M., Dasigi, P., Lambert, N., Richardson, K., Zettlemoyer, L., Dodge, J., Lo, K., Soldaini, L., Smith, N., and Hajishirzi, H. OLMo: Accelerating the Science of Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15789–15809, Bangkok, Thailand, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.841. URL <https://aclanthology.org/2024.acl-long.841>.
- Gupta, R., Walker, L., Corona, R., Fu, S., Petryk, S., Napolitano, J., Darrell, T., and Reddie, A. W. Data-centric ai governance: Addressing the limitations of model-focused policies. *arXiv preprint arXiv:2409.17216*, 2024.
- Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., Smith, C., Barfuss, W., Foerster, J., Gavenčiak, T., Han, T. A., Hughes, E., Kovařík, V., Kulveit, J., Leibo, J. Z., Oesterheld, C., de Witt, C. S., Shah, N., Wellman, M., Bova, P., Cimpanu, T., Ezell, C., Feuillade-Montixi, Q., Franklin, M., Kran, E., Krawczuk, I., Lamparth, M., Lauffer, N., Meinke, A., Motwani, S., Reuel, A., Conitzer, V., Dennis, M., Gabriel, I., Gleave, A., Hadfield, G., Haghtalab, N., Kasirzadeh, A., Krier,



- S., Larson, K., Lehman, J., Parkes, D. C., Piliouras, G., and Rahwan, I. Multi-agent risks from advanced ai, 2025. URL <https://arxiv.org/abs/2502.14143>.
- Haskins, C. ICE Is Paying Palantir \$30 Million to Build ‘ImmigrationOS’ Surveillance Platform. *Wired*, 2025. ISSN 1059-1028. URL <https://www.wired.com/story/ice-palantir-immigrationos/>. Section: tags.
- Hwang, T. *Subprime Attention Crisis*. FSG, New York, October 2020. ISBN 978-0-374-53865-1.
- IETF. Internet standards process, 2024. URL <https://www.ietf.org/process/process/>.
- Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., and Naaman, M. Co-writing with opinionated language models affects users’ views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pp. 1–15, 2023.
- Jeffries, A. and Yin, L. Amazon Puts Its Own “Brands” First Above Better-Rated Products – The Markup, October 2021. URL <https://themarkup.org/amazons-advantage/2021/10/14/amazon-puts-its-own-brands-first-above-better-rated-products>. Section: Amazon’s Advantage.
- Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., Hopkins, A., Bankston, K., Biderman, S., Bogen, M., Chowdhury, R., Engler, A., Henderson, P., Jernite, Y., Lazar, S., Maffulli, S., Nelson, A., Pineau, J., Skowron, A., Song, D., Storch, V., Zhang, D., Ho, D. E., Liang, P., and Narayanan, A. Position: On the Societal Impact of Open Foundation Models. *Proceedings of Machine Learning Research*, 235:23082–23104, 2024. ISSN 2640-3498. Publisher: ML Research Press.
- Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., and Narayanan, A. AI Agents That Matter. *Transactions on Machine Learning Research*, 2025. doi: <https://doi.org/10.48550/arXiv.2407.01502>.
- Kolt, N. Return on data: Personalizing consumer guidance in data exchanges. *Yale L. & Pol’y Rev.*, 38:77, 2019.
- Kolt, N. Governing ai agents. *Notre Dame Law Review*, 2025. Forthcoming.
- Kolt, N., Shur-Ofry, M., and Cohen, R. Lessons from complexity theory for ai governance. *arXiv preprint arXiv:2502.00012*, 2025.
- Korinek, A. and Vipra, J. Concentrating intelligence: scaling and market structure in artificial intelligence. *Economic Policy*, 40(121):225–256, 2025.
- Lazar, S. Communicative justice and the distribution of attention. *Knight First Amendment Institute*, pp. <http://knightcolumbia.tierradev.com/content/communicative-justice-and-the-distribution-of-attention>, 2023. URL [https://knightcolumbia.org/content/communicative-justice-and-the-distribution-of-attention?\\_preview\\_=536fe16e10](https://knightcolumbia.org/content/communicative-justice-and-the-distribution-of-attention?_preview_=536fe16e10).
- Lazar, S. Frontier ai ethics: Anticipating and evaluating the societal impacts of generative agents. *arXiv preprint arXiv:2404.06750*, 2024.
- Lazar, S. Governing the algorithmic city. *Philosophy & Public Affairs*, Early View:1–67, 2025.
- Lazar, S., Thorburn, L., Jin, T., and Belli, L. The Moral Case for Using Language Model Agents for Recommendation, October 2024. URL <http://arxiv.org/abs/2410.12123>. arXiv:2410.12123 [cs].
- Li, N., Pan, A., Gopal, A., Yue, S., Berrios, D., Gatti, A., Li, J. D., Dombrowski, A.-K., Goel, S., Mukobi, G., Helm-Burger, N., Lababidi, R., Justen, L., Liu, A. B., Chen, M., Barrass, I., Zhang, O., Zhu, X., Tamirisa, R., Bharathi, B., Herbert-Voss, A., Breuer, C. B., Zou, A., Mazeika, M., Wang, Z., Oswal, P., Lin, W., Hunt, A. A., Tienken-Harder, J., Shih, K. Y., Talley, K., Guan, J., Steneker, I., Campbell, D., Jokubaitis, B., Basart, S., Fitz, S., Kumaraguru, P., Karmakar, K. K., Tupakula, U., Varadharajan, V., Shoshitaishvili, Y., Ba, J., Esvelt, K. M., Wang, A., and Hendrycks, D. The WMDP Benchmark: Measuring and Reducing Malicious Use with Unlearning. In Salakhutdinov, R., Kolter, Z., Heller, K., Weller, A., Oliver, N., Scarlett, J., and Berkenkamp, F. (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 28525–28550. PMLR, July 2024. URL <https://proceedings.mlr.press/v235/li24bc.html>.
- LinkedIn. Meet Hiring Assistant for Recruiter & Jobs., 2024. URL <https://business.linkedin.com/talent-solutions/hiring-assistant>.
- Marda, N., Sun, J., and Surman, M. Creating a public counterpoint for AI, 2024. URL <https://blog.mozilla.org/en/mozilla/ai/public-ai-counterpoint/>.
- Marro, S., Malfa, E. L., Wright, J., Li, G., Shadbolt, N., Wooldridge, M., and Torr, P. A scalable communication protocol for networks of large language models. *arXiv preprint*, pp. <http://arxiv.org/abs/2410.11905>, 2024. doi: 10.48550/arXiv.2410.11905. URL <http://arxiv.org/abs/2410.11905>.

- Masnick, M. Protocols, not platforms: A technological approach to free speech. *Knight First Amendment Institute*, 2019.
- Mehta, I. Exclusive: Character.AI CEO Noam Shazeer returns to Google as the tech giant invests in the AI company. *TechCrunch*, August 2024. URL <https://techcrunch.com/2024/08/02/character-ai-ceo-noam-shazeer-returns-to-google/>.
- Meta. The llama 3 herd of models, 2024. URL <https://ai.meta.com/research/publications/the-llama-3-herd-of-models/>.
- Mistral AI team. Cheaper, Better, Faster, Stronger, April 2024. URL <https://mistral.ai/news/mixtral-8x22b/>. Section: news.
- Mitchell, M., Ghosh, A., Luccioni, A. S., and Pistilli, G. Fully autonomous ai agents should not be developed. *arXiv preprint arXiv:2502.02649*, 2025.
- Motwani, S. R., Baranchuk, M., Strohmeier, M., Bolina, V., Torr, P. H., Hammond, L., and de Witt, C. S. Secret Collusion among AI Agents: Multi-Agent Deception via Steganography. In *Advances in Neural Information Processing Systems (NeurIPS 2024)*, volume 37, pp. 73439–73486. Curran Associates, Inc., 2024. URL [https://proceedings.neurips.cc/paper\\_files/paper/2024/file/861f7dad098aec1c3560fb7add468d41-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2024/file/861f7dad098aec1c3560fb7add468d41-Paper-Conference.pdf).
- Narayanan, A., Toubiana, V., Barocas, S., Nissenbaum, H., and Boneh, D. A critical look at decentralized personal data architectures. *arXiv preprint arXiv:1202.4503*, 2012.
- Narechania, T. N. and Sitaraman, G. An antimonopoly approach to governing artificial intelligence. *Vanderbilt Law Research Paper*, (24-8), 2024.
- Ngo, R., Chan, L., and Mindermann, S. The alignment problem from a deep learning perspective. *arXiv preprint arXiv:2209.00626*, 2022.
- Nunziato, D. C. The old and the new governors: Efforts to regulate to influence platform content moderation. *First Amend. L. Rev.*, 22:348, 2023.
- OpenAI. Introducing OpenAI o1, September 2024. URL <https://openai.com/o1/>.
- OpenAI. Introducing deep research, 2025. URL <https://openai.com/index/introducing-deep-research/>.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.
- Pariser, E. *The filter bubble : what the Internet is hiding from you*. Viking, London, 2011. ISBN 067092038X (pbk) 9780670920389 (pbk). Eli Pariser. 22 cm. Includes bibliographical references and index. The race for relevance – The user is the content – The Adderall society – The you loop – The public is irrelevant – Hello, world! – What you want, whether you want it or not – Escape from the city of ghettos.
- Public AI Network. Public AI: Infrastructure for the Common Good, 2024. URL <https://publicai.network/>.
- Reisman, R. and Whitt, R. New Perspectives on AI Agentiality and Democracy: "Whom Does It Serve?". *Tech Policy Press*, December 2024. URL <https://www.techpolicy.press/new-perspectives-on-ai-agentiality-and-democracy-whom-does-it-serve/>.
- Rochet, J.-C. and Tirole, J. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003. ISSN 1542-4766. doi: 10.1162/154247603322493212.
- Russell, S. *Human compatible: AI and the problem of control*. Penguin Uk, 2019.
- Sellen, A. and Horvitz, E. The rise of the ai co-pilot: Lessons for design from aviation and beyond. *Communications of the ACM*, 67(7):18–23, 2024.
- Sharma, N., Liao, Q. V., and Xiao, Z. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642459. URL <https://doi.org/10.1145/3613904.3642459>.
- Shavit, Y., O’Keefe, C., Eloundou, T., McMillan, P., Agarwal, S., Brundage, M., Adler, S., Campbell, R., Lee, T., Mishkin, P., Hickey, A., Slama, K., Ahmad, L., Beutel, A., Passos, A., and Robinson, D. G. Practices for Governing Agentic AI Systems. December 2023. URL <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>.
- Shopify. Most Popular Social Media Platforms in 2024, 2024. URL <https://www.shopify.com/blog/most-popular-social-media-platforms>.

- Sierra. The Conversational AI Platform, 2025. URL <https://sierra.ai>.
- South, T., Marro, S., Hardjono, T., Mahari, R., Whitney, C. D., Greenwood, D., Chan, A., and Pentland, A. Authenticated Delegation and Authorized AI Agents, January 2025. URL <http://arxiv.org/abs/2501.09674>. arXiv:2501.09674 [cs].
- StatCounter. Search Engine Market Share Worldwide, 2024. URL <https://gs.statcounter.com/search-engine-market-share>.
- State of California. California Consumer Privacy Act of 2018 (CCPA), Cal. Civ. Code § 1798.100, 2018. URL <https://oag.ca.gov/privacy/ccpa>. Enacted as Assembly Bill No. 375, amended by the California Privacy Rights Act (CPRA) in 2020.
- Stripe. Adding payments to your LLM agentic workflows, 2024. URL <https://stripe.dev/blog/adding-payments-to-your-agentic-workflows>.
- Stroebel, B., Kapoor, S., and Narayanan, A. Hal: A holistic agent leaderboard for centralized and reproducible agent evaluation. <https://github.com/princeton-qli/hal-harness>, 2025.
- Sullivan, M. OpenAI wants to be more like Apple and Google. *Fast Company*, April 2025. URL <https://www.fastcompany.com/91317178/openai-wants-to-be-more-than-an-ai-company>.
- Sumers, T., Yao, S., Narasimhan, K., and Griffiths, T. Cognitive Architectures for Language Agents. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL <https://openreview.net/forum?id=1i6ZCvflQJ>.
- Sunstein, C. R. *Republic.com*. Princeton University Press, Princeton, N.J. ; Oxford, 2001. ISBN 0691070253 (alk. paper). 00045331 Cass R. Sunstein. 20 cm. Includes bibliographical references (p. 203-212) and index.
- Susser, D., Roessler, B., and Nissenbaum, H. Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review*, 4:1–45, 2019.
- Susskind, J. *The Digital Republic: On freedom and democracy in the 21st century*. Bloomsbury Publishing, 2022.
- Suzor, N. *Lawless: The Secret Rules That Govern our Digital Lives*. Cambridge University Press, Cambridge, 2019. doi: 9781108666428.
- Team, Q. Qwen2.5-Max: Exploring the Intelligence of Large-scale MoE Model, January 2025. URL <https://qwenlm.github.io/blog/qwen2.5-max/>. Section: blog.
- Temkin, M. OpenAI is reportedly in talks to buy Windsurf for \$3B, with news expected later this week. *TechCrunch*, April 2025. URL <https://techcrunch.com/2025/04/16/openai-is-reportedly-in-talks-to-buy-windsurf-for-3b-with-news-expected-later-this-week/>.
- Tur, A. D., Meade, N., Lù, X. H., Zambrano, A., Patel, A., Durmus, E., Gella, S., Stańczak, K., and Reddy, S. Safearena: Evaluating the safety of autonomous web agents. *arXiv preprint arXiv:2503.04957*, 2025.
- Van Dijck, J. Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, 23 (9):2801–2819, 2021. doi: 10.1177/1461444820940293. URL <https://journals.sagepub.com/doi/abs/10.1177/1461444820940293>.
- van Dijck, J., Poell, T., and de Waal, M. *The platform society: Public values in a connective world*. Oxford University Press, New York, 2018.
- Van Loo, R. Digital market perfection. *Mich. L. Rev.*, 117: 815, 2018.
- Varian, H. R. Computer mediated transactions. *American Economic Review*, 100(2):1–10, 2010. doi: 10.1257/aer.100.2.1.
- Viljoen, S., Goldenfein, J., and McGuigan, L. Design choices: Mechanism design and platform capitalism. *Big Data & Society*, pp. 1–13, 2021. doi: 10.1177/20539517211034312.
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024.
- Wegner, P. Interoperability. *ACM Comput. Surv.*, 28 (1):285–287, March 1996. ISSN 0360-0300. doi: 10.1145/234313.234424. URL <https://doi.org/10.1145/234313.234424>.
- Wei, A., Haghtalab, N., and Steinhardt, J. Jailbroken: How Does LLM Safety Training Fail? November 2023. URL <https://openreview.net/forum?id=jA235JGM09>.
- Weiss, M., Rahaman, N., Wuthrich, M., Bengio, Y., Li, L. E., Schölkopf, B., and Pal, C. Redesigning Information Markets in the Era of Language Models. August 2024. URL <https://openreview.net/forum?id=Zq9Dfj4nBo#discussion>.

Wiggers, K. Amazon forms an AI agent-focused lab led by Adept’s co-founder, December 2024a. URL <https://techcrunch.com/2024/12/09/amazon-forms-a-new-ai-agent-focused-lab-led-by-adept-co-founder/>.

Wiggers, K. Amazon hires founders away from AI startup Adept, June 2024b. URL <https://techcrunch.com/2024/06/28/amazon-hires-founders-away-from-ai-startup-adept/>.

Wu, T. Network neutrality, broadband discrimination. *J. on Telecomm. & High Tech. L.*, 2:141, 2003.

Wu, T. *The age of extraction : how tech platforms conquered the economy and threaten our future prosperity*. Alfred A. Knopf, New York, 2025. ISBN 9780593321249.

Xu, Z., Liu, Y., Deng, G., Li, Y., and Picek, S. A Comprehensive Study of Jailbreak Attack versus Defense for Large Language Models. In Ku, L.-W., Martins, A., and Srikumar, V. (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 7432–7449, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.443. URL <https://aclanthology.org/2024.findings-acl.443/>.

Zaharia, M., Khattab, O., Chen, L., Davis, J. Q., Miller, H., Potts, C., Zou, J., Carbin, M., Frankle, J., Rao, N., and Ghodsi, A. The shift from models to compound ai systems. <https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/>, 2024.

Zhou, A., Yan, K., Shlapentokh-Rothman, M., Wang, H., and Wang, Y.-X. Language Agent Tree Search Unifies Reasoning Acting and Planning in Language Models, December 2023. URL <http://arxiv.org/abs/2310.04406>. arXiv:2310.04406 [cs].

Zittrain, J. *Perfect Enforcement on Tomorrow’s Internet*, pp. 125–156. Hart Publishing, Portland, OR, 2008.

Zuboff, S. *The Age of Surveillance Capitalism*. Public Affairs, New York, 2019.



## Appendix: Hypothetical Examples of Platform Agents and Alternatives

Platform-controlled AI agents	Hardware-ecosystem agents	User-centric agent advocates
<b>Hypothetical Example:</b> Amazon Alexa Pro—a platform agent that collects all user interactions, prioritizes Amazon services and sponsored products, manipulates user attention through notifications, and creates significant switching costs to leave the Amazon ecosystem.	<b>Hypothetical Example:</b> Banana Intelligence Plus—an agent operating primarily within hardware company Banana’s hardware ecosystem, emphasizes privacy while collecting data for model improvement, offers fair third-party options while subtly prioritizing Banana’s other services, and primarily attempts to reinforce hardware purchases.	<b>Hypothetical Example:</b> Open Assistant—a fully open-source agent that runs completely locally, stores all data on user devices, interacts with any market or service without bias, allows full user control over filtering and priorities, and can be easily migrated between devices or platforms.
<b>Lock-in via platform agents.</b> Platform agents become indispensable “digital concierges,” erecting switching barriers and concentrating control.	<b>Lock-in to a hardware ecosystem.</b> Agents are tied to specific hardware products but may offer limited interoperability within the ecosystem.	<b>Fully interoperable.</b> Agent advocates can run on local or diverse hardware systems, making it easier to switch or migrate data and preserving user choice.
<b>Invasive personal data collection.</b> Platform agents observe every digital interaction, enabling unprecedented user profiling and heightened surveillance.	<b>Data collection not used for surveillance.</b> Collects data primarily for improving models and services, with some privacy protections in place.	<b>Local data.</b> Agent advocates enable users to store and process data locally, minimizing external data access and surveillance risks.
<b>Repressive market design.</b> Platform agents can steer transactions to certain vendors, embed hidden fees, and block competitor access.	<b>Transparent arbitrator with fair terms.</b> Attempts to provide fair options but may have subtle preferences for ecosystem-compatible services.	<b>Open market interactions.</b> Agent advocates can browse, compare, and transact across diverse vendors, thereby reducing platform dependence.
<b>Manipulating attention and engagement.</b> Platform agents prioritize content and notifications aligned with platform interests.	<b>Interested in complementary services, not in market manipulation.</b> Focuses on enhancing hardware value rather than driving engagement metrics.	<b>Enhanced user autonomy.</b> Agent advocates enable users to define how content is filtered or ranked, reducing platform influence over content consumption.

Table 1: There is a spectrum from platform-controlled agents to fully user-centric agent advocates. We illustrate both ends of the spectrum as well as an approximate midpoint.