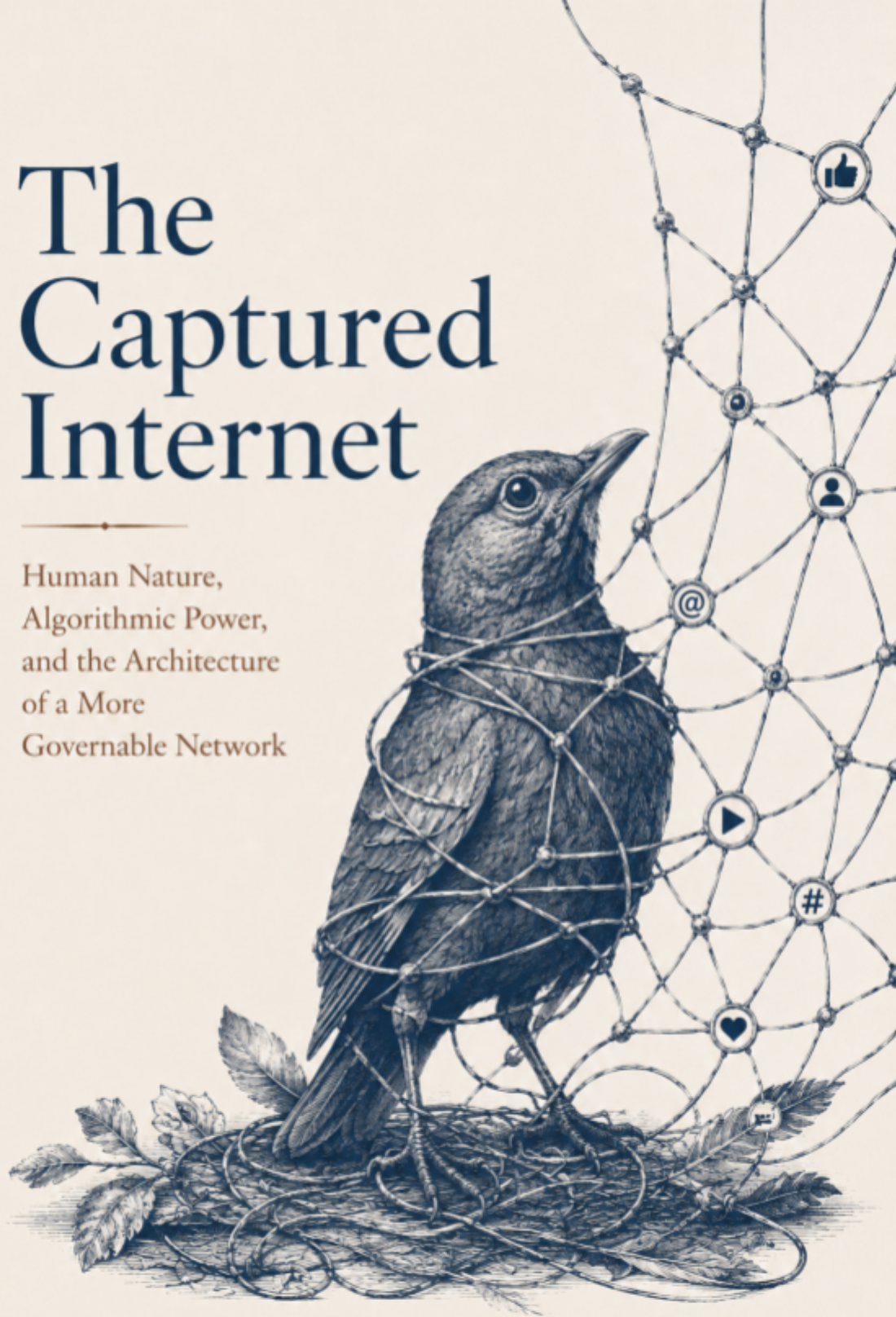


The Captured Internet

Human Nature,
Algorithmic Power,
and the Architecture
of a More
Governable Network



The Captured Internet

*Human Nature, Algorithmic Power, and the
Architecture
of a More Governable Network*

Axiologic Research

July 2026

*Imagine that the Internet continues to function perfectly
while becoming progressively less worth entering.*

HISTORY • GOVERNANCE • RECOMMENDATION • IDENTITY • INSTITUTIONAL DESIGN

Copyright © [2026] Sînică Alboaie

All rights reserved.

KDP publishing rights held by Outfinity SRL (Iași, Romania).

Usage rights granted to Outfinity Sagl (Caslano, Switzerland).

Available for free on the Outfinity Venture Studio website (www.outfinity.ch).

Artificial intelligence systems were used as editorial and research instruments.

The argument, selection, and authorial responsibility belong to Sînică Alboaie.

Author's Note:

This work was created within the framework of the overarching Outfinitism research program. Although artificial intelligence language models were leveraged to assist with text generation, the underlying philosophy, thematic structure, and analytical frameworks stem entirely from the author's decades of dedicated research into social phenomena and social technologies.

Analyzing the fundamental issues of today's internet, this book outlines key evolutionary trajectories compatible with the rise of AI, forming a cornerstone of the investment thesis at Outfinity Venture Studio.

Table of contents

The Whole Book in Ten Pages.....	5
A Network That Still Works.....	5
What a Better Internet Would Actually Require.....	9
Chapter 1: Before Platforms.....	15
1.1 Open Protocols, Public Money, and Small Communities.....	15
1.2 Email, Forums, Directories, and Search.....	18
Chapter 2: Platforms and Hidden Constitutions.....	22
2.1 Why Platforms Won.....	22
2.2 Corporate Capture, Moderator Power, and Free Labor.....	25
Chapter 3: The Human Adversary.....	29
3.1 Anonymity, Disinhibition, Trolling, and Coordinated Abuse.....	29
3.2 Moderation as Government.....	32
Chapter 4: Relevance Is a Constitution.....	36
4.1 Search, Feeds, Engagement, and the Drift Toward Idiocracy.....	36
4.2 Misinformation, Echo Chambers, and Election Manipulation.....	39
Chapter 5: The Synthetic Web and the Internet That Still Works.....	42
5.1 Dead Internet Theory, Bots, and Generative Abundance.....	42
5.2 Wikipedia, Open Source, Science, and Durable Commons.....	45
Chapter 6: Beyond Platform Sovereignty.....	48
6.1 Polycentric Communities and Decentralized Brands.....	48
6.2 Economics, Licensing, and Anti-Capture Design.....	51
Chapter 7: Personal AI as the User's Recommender.....	54
7.1 From Hidden Profiles to User-Governed Relevance.....	54
7.2 Privacy, Bias, Pluralism, and the Risk of a Personal Echo Chamber.....	57
Chapter 8: Rebuilding Communication from the Protocol Layer.....	61
8.1 WebRTC, Email-Like Addressing, DIDs, and Trust-Gated Connections.....	61
8.2 A Smaller Attack Surface for Humans and Agents.....	64
References.....	69

The Whole Book in Ten Pages

A Network That Still Works

Imagine that the Internet continues to function perfectly while becoming progressively less worth entering. Search is instant. Messages arrive. Video never stops. Personal assistants summarize thousands of pages before breakfast. Yet much of what appears on the screen was not created to inform a person, selected by anyone answerable to that person, or offered by an institution that can be challenged. The machinery works. The public world carried by the machinery is thinning.

This opening contains the entire argument in compressed form. A reader who stops here should understand the diagnosis, the strongest evidence, and the proposed direction. The rest of the book exists for readers who want the history, qualifications, and design details.

The standard story begins with an innocent decentralized Internet and ends with corporate capture. It is partly right and too simple. The early network arose from public funding, military and university research, standards bodies, telecommunications infrastructure, and small communities with unusually high technical competence. Openness was an institutional and engineering achievement, not the absence of power. Early harmony also depended on selective access, manageable scale, and administrators whose authority was visible because the communities were small (Clark 1988; Abbate 1999; Turner 2006).

Open protocols made heterogeneous machines interoperate and made copying, publishing, and communication extraordinarily cheap. They did not solve identity, reputation, discovery, spam, moderation, financing, or legitimate collective decision-making. As participation expanded, those deferred problems returned. Email acquired filters and sender reputation. Usenet and forums acquired administrators and conflict. Directories acquired editors. Search

engines acquired the power to decide which of billions of pages would become visible.

Platforms won because they solved real problems: coherent interfaces, hosting, search, persistent accounts, mobile access, payments, security, recommendation, and moderation. Centralization was useful before it became constitutional. Capture began when the same provider came to own identity, relationships, archives, ranking, revenue, and the right to change every rule governing them. Users and moderators continued to create the community, but the community became an asset on somebody else's balance sheet.

Table 1. The diagnosis in six layers

Element	Meaning or design consequence
Protocol layer	Interoperability and cheap publication were achieved; identity, moderation, financing, and legitimacy remained unresolved.
Platform layer	Usability, hosting, search, and security improved, but identity, ranking, relationships, archives, and revenue became vertically concentrated.
Human layer	Low friction protected vulnerable speech and widened participation while also reducing the cost of harassment, fraud, and coordinated manipulation.
Epistemic layer	Search and recommendation solved scarcity of attention by turning relevance into a private and often opaque constitutional power.
Synthetic layer	Automation reduced production costs, but also weakened authenticity, provenance, and the economic position of careful human work.
Institutional layer	Open participation created immense value without reliably granting contributors durable rights of voice, exit, or economic benefit.

Corporate capture is only one component of the story. The Internet also changed the price of human behavior. Pseudonymity can protect dissidents, patients, whistleblowers, and vulnerable minorities. Low-cost identity can also reduce reputational consequences, permit rapid return after exclusion, and make harassment or simulated consensus cheaper. The online disinhibition effect does not turn

every anonymous person into a monster, but it changes the incentives around impulsive conduct (Suler 2004). Experimental evidence further suggests that trolling is partly situational: negative mood, hostile examples, and local norms can make ordinary participants imitate abuse (Cheng et al. 2017).

The amplification loop is equally important. Provocation produces reaction; reaction is recorded as engagement; engagement produces distribution; repeated exposure changes norms; creators learn that contempt, certainty, humiliation, and factional conflict travel farther than qualification. Studies of moral-emotional language and out-group hostility support this mechanism (Brady et al. 2017; Rathje, Van Bavel, and van der Linden 2021). The result is often presented as revealed user preference even though the system helped manufacture the preference it measured.

This is the defensible meaning of the metaphor of online “Idiocracy.” Users did not suddenly become less intelligent. The environment selects for material that can generate a cheap, rapid signal. A false novelty can be written in seconds; verification may require hours. A dramatic accusation fits a feed better than a careful uncertainty estimate. Engagement optimization therefore subsidizes immediate certainty and taxes slow epistemic work. The problem is adverse selection against complexity, not a biological decline in intelligence.

Moderation cannot simply be removed. Nonintervention transfers control to persistent aggressors, organized factions, automated swarms, or owners who intervene only when legal and commercial risk becomes intolerable. The real choice is between governance that is visible, bounded, reviewable, and legitimate, and governance that is improvised, opaque, or captured (Gillespie 2018; Suzor 2019). Volunteer moderators perform indispensable civic labor, often without pay or protection (Matias 2019), but can also turn stewardship into status or ideological rule. Their local power must be distinguished from the platform company’s structural control over databases, accounts, monetization, and constitutional change.

Wikipedia and Reddit show that the outcome is not predetermined. Wikipedia demonstrates that a global knowledge commons can

function without behavioral advertising because revisions, sources, discussions, and reversals are inspectable. It also shows how anti-vandalism tools, complex norms, and contributor inequalities can calcify a successful institution (Halfaker et al. 2013; Wagner et al. 2016). Reddit grants communities local rule-making on corporate infrastructure. Research suggests that strong sanctions can reduce hate speech rather than merely move it, while explanations and appeals improve procedural legitimacy (Chandrasekharan et al. 2017; Jhaver, Bruckman, and Gilbert 2019). Institutions matter more than slogans about openness or control.

Political manipulation requires the same discipline. False information can diffuse farther than true information in some networks, bots can amplify low-credibility sources, and coordinated influence operations are real (Vosoughi, Roy, and Aral 2018; Shao et al. 2018). But operation, exposure, persuasion, turnout effects, agenda change, and electoral outcome are different claims. The investigation of Russia's Internet Research Agency documented a deliberate campaign during the 2016 United States election; it did not identify a defensible number of converted votes. Echo chambers also exist unevenly, and forced exposure to hostile opposing views can intensify rather than reduce polarization (Bail et al. 2018; Cinelli et al. 2021).

The strongest Dead Internet Theory claims that most online activity is already artificial and centrally coordinated. Evidence does not support that total conspiracy. Its weaker warning is credible. Automated traffic is substantial; generative systems can flood search and social feeds with plausible material optimized for ranking; agents increasingly summarize sources that users never visit; recursive training on synthetic data can degrade models when original human data disappear (Shumailov et al. 2024). The Internet is not dead, but authenticity and human legibility are becoming scarce resources.

The counterevidence is equally important. Open-source software, Wikipedia, OpenStreetMap, scientific preprints, patient communities, humanitarian mapping, and specialist forums have generated immense public value. The conclusion is not that openness failed. It

is that institutions creating value are often separated from institutions controlling the resulting assets. The Internet's successes came from open technical foundations combined with norms, roles, histories, and repair mechanisms. The next design must preserve that capacity while correcting its constitutional weaknesses.

What a Better Internet Would Actually Require

The alternative is neither a return to small forums nor a universal blockchain. It is a polycentric system: open protocols underneath; communities with real constitutions above them; several competing service providers around them; and rights allowing members to move providers without abandoning identity, relationships, reputation evidence, or archives. The objective is not to eliminate power. It is to make power visible, divided, temporary, and contestable.

A community requires more than a chat room and terms of service. It needs a stated purpose, membership boundaries, known powers, qualified moderators, sanctions that escalate proportionately, reasons for consequential decisions, an appeal path, conflict-of-interest rules, and financing for infrastructure and difficult labor. Ostrom's work on durable commons remains relevant because it rejects both naive privatization and naive openness: successful communities match rules to context, let affected members participate in changing them, monitor conduct, apply graduated sanctions, resolve disputes, and operate through nested levels of governance (Ostrom 1990).

A decentralized brand can provide the missing institutional layer. A brand is not merely a logo; it is a public promise about quality, compatibility, conduct, and responsibility. In a conventional platform, one company owns that promise. In a purely open project, the artifact may be forkable while reputation and economic coordination remain fragile. A decentralized brand places the name, standards, treasury, and certification process under a constitution that no founder, investor, moderator group, or infrastructure provider can alter alone.

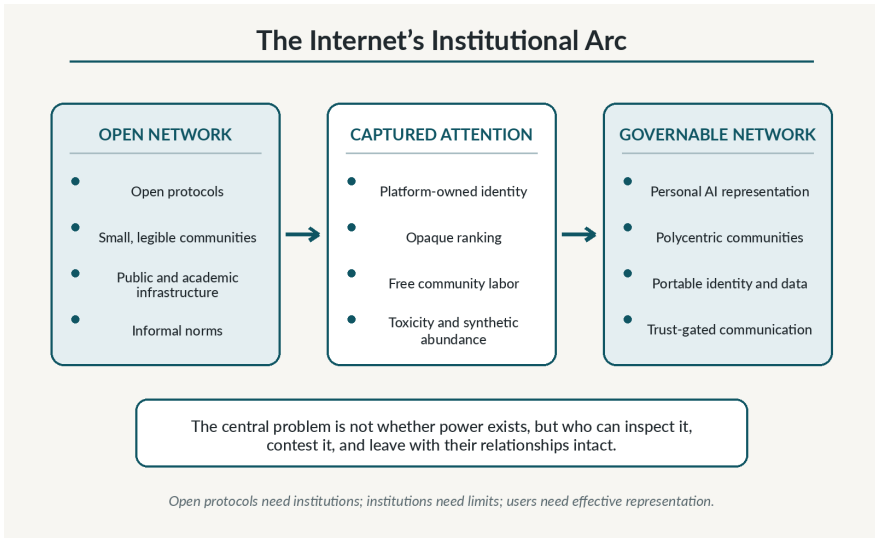


Figure 1. The Internet's institutional arc: from open protocols, through concentrated attention, toward governable networks.

The instruments must remain distinct. A software or content license governs copying and modification. A brand license governs use of the shared identity. A constitution allocates decision, audit, amendment, emergency, and appeal powers. A contribution agreement recognizes labor, maintenance, and capital. Service contracts constrain hosting, search, recommendation, and data use. Neither a token nor an open-source license can perform all these jobs.

Decentralization must be tested in practice. Can members export relationships and archives, not merely download a file? Can a community change hosting without disappearing? Can appeal reverse a moderator? Can an operator be replaced without breaking identity? Do contributors share economic benefits? Can any provider redefine access unilaterally? A repository or distributed ledger is not decentralization when the social institution remains captive.

Recommendation needs a similar reversal. Today platforms build hidden profiles and use mixed objectives - retention, advertising, safety, commercial priorities, and strategy - to decide what users notice. Even sophisticated ranking treats the model of the user as the platform's asset. A personal AI could instead maintain a private, revisable representation of interests, obligations, biases, and limits. It

could gather candidates from many sources, apply explicit policies, explain trade-offs, remember corrections, prefer primary evidence, reduce outrage, introduce credible disagreement, or reject undisclosed sponsorship. Recent research describes this shift as movement from hidden profiles toward governable personalization (Liu et al. 2026).

The important change is constitutional, not merely predictive. Platforms already predict taste. The new possibility is that the representation of the user becomes inspectable, portable, and answerable to the user. Recommender research has long shown that experience depends on control, diversity, privacy, effort, and explanation as well as accuracy (Cramer et al. 2008; Knijnenburg et al. 2012; Helberger, Karppinen, and D'Acunतो 2018). The agent should expose alternative rankings and counterevidence rather than quietly becoming a more intimate echo chamber.

Such an agent also creates risks. It may flatter temporary impulses, preserve obsolete preferences, leak private memory, depend on a cloud provider, or be manipulated by prompt injection and poisoned sources. Its objectives, uncertainty, commercial relationships, update path, and delegated powers must therefore be visible. Preferences should be revisable and sometimes expire. Several ranking policies should be comparable. Local or confidential processing, provenance, capability limits, and auditable action logs are not optional details; they are constitutional safeguards.

A second reversal can occur beneath the platform layer. The current Internet exposes services to a global address space and then spends enormous effort rejecting unwanted traffic. Email accepts attempted delivery from strangers. Public services defend themselves against scanning, spam, credential attacks, scraping, and denial of service. Openness remains valuable, but global receptivity should not be the default for every personal service and AI agent.

A future system could retain familiar interfaces - addresses, inboxes, contacts, groups, shared documents, and assistants - while changing the trust boundary. A public address need not reveal a permanent personal endpoint. Initial contact could pass through a narrow

introduction service. A successful introduction could issue a revocable capability defining who may signal, send, invoke an agent, or open a channel. After that, humans and agents could communicate through encrypted direct or relayed sessions using WebRTC data channels or related transports.

WebRTC is only one component. It supplies APIs and protocols for interactive media and data, with connectivity negotiated through ICE and often assisted by relays (IETF 2018; IETF 2021a; W3C 2025a). It does not define signaling, offline delivery, archives, social identity, or trust. An email-like readable address could remain useful, while a DID or equivalent identifier provides portable cryptographic control and Verifiable Credentials provide selective claims (W3C 2022; W3C 2025b; W3C 2026). These technologies prove control or issuer statements, not honesty. Trust still comes from relationships, institutions, endorsements, history, and context.

The resulting network would be trust-gated rather than globally receptive by default. Unknown actors might require an introduction token, referral, bond, or relevant credential. Pairwise identifiers could reduce correlation across relationships. Agents could negotiate rate limits, accepted formats, confidentiality, and automation before a session begins. A user could revoke one relationship without replacing a global identity. The exposed surface would shrink because most personal services would never accept arbitrary unsolicited traffic.

Table 2. A different set of defaults

Element	Meaning or design consequence
Hidden platform profile	A user-governed personal AI model that can be inspected, corrected, constrained, exported, or deleted.
Globally exposed inbox	Trust-gated introductions, capability-scoped contact, and relationship-specific channels.
One opaque ranking system	Plural, inspectable ranking providers selected by users and communities for declared purposes.
Corporate community	Polycentric communities operating under a shared decentralized brand and a contestable constitution.

Element	Meaning or design consequence
Invisible free maintenance	Explicit financing for moderation, security, infrastructure, repair, and long-term stewardship.
Nominal data export	Portable identity, relationships, reputation evidence, archives, and practical continuity of service.
Authenticity by intuition	Signed provenance, accountable sponsorship, and clear labeling of automated participation.

Servers would remain. WebRTC often needs signaling and relay infrastructure; asynchronous communication needs store-and-forward services; secure groups need protocols such as Messaging Layer Security (IETF 2023; IETF 2025). Providers would still sell storage, availability, discovery, backup, computation, and abuse response. The difference is that no provider would need to own identity, relationships, recommendation, and archives as one indivisible package.

Personal agents would act as local mailrooms, recommenders, security mediators, and representatives. They could screen introductions, verify credentials, select relays, summarize messages, enforce privacy policy, and track delegated authority. Because this role is powerful, the agent must be auditable and replaceable. Its model, memory, permissions, and commercial incentives cannot become another hidden constitution.

The most realistic future is a mixed topology. Some communities will remain centralized because clear ownership or liability serves their purpose. Others will federate. Others will share standards and reputation through decentralized brands while using independent operators. Pluralism is not a temporary defect. It is the defense against a new universal sovereign.

The final test is simple. A living Internet should help people form communities, create durable public goods, discover surprising information, and use intelligent assistance without surrendering identity and relationships to whoever owns the dominant interface. It should make abuse expensive without making dissent impossible; permit moderation without unreviewable moderators; reward

founders and operators without granting hereditary rule; and let recommendation help a person think rather than silently define what the person is allowed to see.

Human nature will not be removed from the network. Curiosity, generosity, vanity, tribalism, courage, cruelty, and care will appear in every architecture. Progress means building institutions in which the better capacities are easier to coordinate, the worse capacities are harder to weaponize, and temporary usefulness does not become permanent ownership of the social world.

***Evidence note.** Claims are supported primarily by peer-reviewed research, academic books, standards, and institutional reports. Technical proposals in Chapters 7 and 8 are presented as design directions built from mature components, not as claims that the integration problem is already solved.*

Chapter 1: Before Platforms

The early Internet is often remembered as a lost republic of equals. It was more interesting than that. Open architecture emerged from institutions, expert communities, and deliberate engineering choices. Its success created freedoms that its social design was not prepared to govern.

1.1 Open Protocols, Public Money, and Small Communities

ARPANET and the Internet were not spontaneous products of a market or a counterculture. They emerged from public research programs, universities, contractors, standards communities, and engineers who needed heterogeneous networks to interoperate. Packet switching, host protocols, and later TCP/IP were solutions to technical and organizational problems. The system had to survive equipment diversity, administrative boundaries, and changing applications. The resulting architecture distributed functions toward the edges and avoided requiring one network to understand every future use (Clark 1988; Leiner et al. 1997).

This technical history matters because later political interpretations often treat decentralization as the original moral purpose of the network. Resilience, interoperability, and permissionless application development certainly had political consequences, but they were not a complete political theory. The end-to-end tradition constrained what the network core should decide; it did not specify how communities should handle harassment, how contributors should be paid, or how a public ranking system should justify its choices (Saltzer, Reed, and Clark 1984).

The early network's culture was shaped by selective access. Universities, research laboratories, and technically sophisticated amateurs supplied many participants. Shared professional norms and repeated interaction made reputation meaningful. Informality worked because communities were small enough for history to be

remembered and because participation itself required effort. This did not eliminate hierarchy. Administrators controlled machines, mailing lists, and accounts. Expertise conferred status. People who did not share the dominant technical and cultural assumptions could be ignored or excluded.

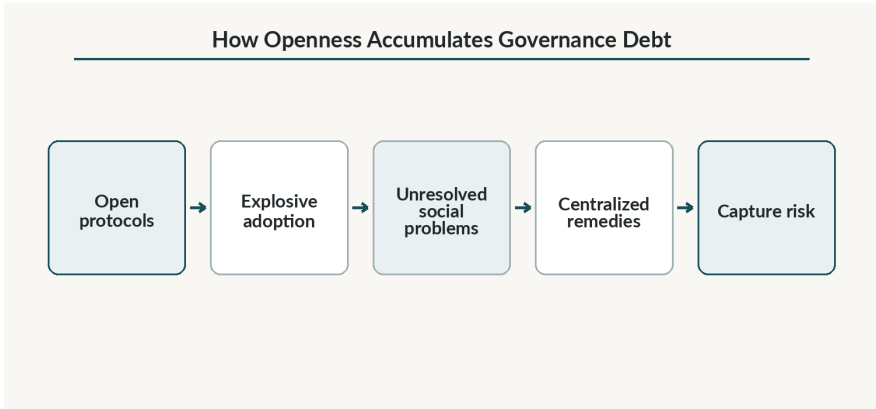


Figure 2. Open protocols reduced technical dependence while leaving a growing debt in identity, discovery, moderation, and financing.

The freedom associated with the early Internet was therefore both real and subsidized. Public money supported infrastructure. Universities absorbed costs. Volunteers maintained software and documentation. Employers allowed skilled staff to contribute. The network appeared to float above ordinary economics because much of its economy was hidden in institutions that had reasons other than direct profit to support it. When access expanded, the cost of operating social systems became harder to conceal.

The release of World Wide Web software by CERN in 1993 exemplified the power of institutional generosity. By allowing royalty-free use, CERN enabled competing browsers, servers, and sites to appear without asking permission from a central owner (CERN 1993). The decision did not predetermine a noncommercial Web. It made commercial and noncommercial innovation possible on the same base. Openness created a field in which both commons and corporations could grow.

The first contradiction was already present. A protocol can be open while the communities using it are closed. A service can be

decentralized while access to expertise, hardware, and time remains unequal. A public standard can enable private concentration if complementary assets such as discovery, identity, and distribution become scarce. The history of the Internet is not a fall from technical purity. It is the movement of power from one layer to another as each bottleneck is solved.

This explains why nostalgia is a poor guide. Recreating old protocols or chronological forums would not recreate the social conditions under which they seemed manageable. Billions of users, automated adversaries, commercial incentives, and generative systems change the scale of the problem. What remains valuable from the early Internet is not its innocence but its institutional achievement: multiple networks and implementations could cooperate because standards were open and authority was limited at the core.

The standards culture itself contained a useful institutional clue. Requests for Comments were public technical arguments rather than decrees, and the ideal of rough consensus and running code made authority depend partly on implementable results. The process was never perfectly democratic: participation required expertise, time, and access, while key resources such as names and addresses still needed coordination. Yet it demonstrated that common rules can emerge without a single vendor owning every implementation. The legitimacy came from open specification, competing implementations, and the possibility of technical challenge.

Even the supposedly decentralized core contained concentrated functions. The Domain Name System introduced hierarchical delegation; address allocation and protocol registries required stewardship; large backbone providers occupied strategic positions. Decentralization meant that no single concentration necessarily controlled every layer, not that concentration vanished. This layered view is essential for contemporary reform. A system may tolerate a central registry if identities, data, clients, and communities can move around it. It becomes dangerous when control of one layer is leveraged into ownership of all the others.

The unfinished task is to extend that principle upward. Identity should not be inseparable from one application. Reputation should not be trapped inside one database. Communities should be able to hire infrastructure without transferring sovereignty. Recommendation should be replaceable. Moderation should operate through declared powers and appeal. The early engineers created an open communications substrate. The next generation must create an open institutional substrate.

1.2 Email, Forums, Directories, and Search

Email is the clearest demonstration that federation can survive at global scale. A person can send a message across organizational boundaries without both parties belonging to the same provider. Domains allow local administration while common protocols preserve interoperability. This model is less elegant in practice than in diagrams, but it has resisted complete ownership by any single company for decades.

Its weakness is equally instructive. Email allows unsolicited contact at extremely low marginal cost. The sender externalizes filtering, storage, attention, and security costs to recipients and providers. The Spamalytics study found that a campaign could remain economically viable despite an extraordinarily low conversion rate because millions of messages were cheap to send through compromised machines (Kanich et al. 2009). Authentication and reputation systems reduced abuse, but they also concentrated practical power in large providers able to operate sophisticated filters and influence deliverability.

Usenet and web forums created richer local cultures. Their topic-centered organization preserved long conversations and allowed reputation to accumulate within a domain. A good technical forum could become a living archive more useful than any formal manual. Yet most forums were private governments. Owners controlled the database and domain. Moderators interpreted rules, often without appeal. Communities could be generous and expert while remaining constitutionally fragile.

Directories such as Yahoo's early catalogue and the Open Directory Project made selection visible. Human editors classified sites and implicitly vouched for relevance. The model could express judgment and context, but it did not scale with the Web. Editors became bottlenecks, categories hardened, and decisions could be captured by small groups. Still, the user could see that the directory represented an editorial structure rather than an objective map of reality.

Search engines solved scale by automating selection. Link analysis and large indexes made the Web usable, but ranking acquired a political role. A search engine does not merely retrieve what exists. It allocates visibility, which often determines whether a source has practical existence for most users. Introna and Nissenbaum argued early that search engines inevitably favor some sites and values, even when their choices are presented as technical relevance (Introna and Nissenbaum 2000).

Table 3. What each early layer solved and left unresolved

Element	Meaning or design consequence
TCP/IP and end-to-end design	Connected heterogeneous networks and limited dependence on intermediate owners; did not create social identity or community governance.
Email	Created global federated messaging; exposed recipients to spam, spoofing, reputation concentration, and abuse at negligible sending cost.
Usenet and forums	Enabled durable topic communities; relied on fragile local norms, owners, administrators, and socially costly conflict resolution.
Directories	Made human editorial judgment visible; failed to scale and created power around inclusion, categories, and editor factions.
Search engines	Scaled discovery across the open Web; made ranking an opaque intermediary between publication and practical visibility.

The move from directory to search changed the appearance of power. Human categories looked subjective. Algorithmic results looked discovered. Yet every ranking system chooses signals, handles

manipulation, defines freshness, and balances commercial and public objectives. Search also created an adversarial industry. Publishers optimized for the ranking system; the ranking system adapted to optimization; content increasingly addressed the classifier before the reader.

Email, forums, directories, and search therefore contain the full future in miniature. Federation preserves exit but attracts abuse. Local communities produce knowledge but can become private oligarchies. Human curation provides context but cannot cover everything. Algorithmic ranking scales but hides judgment. None of these systems is a failure. Each reveals a different institutional trade-off.

The social breakdown associated with “Eternal September” made the scaling problem visible before modern platforms. New cohorts had always entered Usenet and learned its norms, but commercial access converted a periodic influx into permanent mass arrival. Old participants often blamed newcomers, sometimes with elitist contempt, yet the deeper problem was institutional: onboarding, moderation, and memory had not been designed for continuous growth. Norms that can be transmitted only through patient mentoring are vulnerable when arrivals outnumber mentors.

The Open Directory Project later attempted to combine human editorial judgment with distributed volunteer work. It proved that thousands of editors could maintain a public catalogue, but also that category ownership, slow review, and uneven standards could become bottlenecks. Search engines displaced directories not because human judgment was useless, but because automated indexing could cover orders of magnitude more material. The trade was coverage for inspectability.

Email also became operationally more centralized than its protocols suggest. Independent servers remain possible, yet deliverability depends on reputation, authentication, abuse response, and relationships with major providers. A small operator can conform to standards and still find messages filtered. This is a recurring pattern: open protocol at the formal layer, concentrated trust at the

operational layer. The cure is not to deny the usefulness of professional filtering but to make its rules, portability, and appeals less dependent on a few companies.

The platform era did not replace these systems because it had a superior philosophy. It combined their useful functions behind one account and one interface. Messaging, identity, discovery, reputation, distribution, and monetization became easier because one operator could coordinate them. The cost of that convenience was that the operator also acquired the power to redefine every function at once.

Chapter 2: Platforms and Hidden Constitutions

Platforms won by making the Internet comprehensible. They became dangerous when integration turned into sovereignty and when communities discovered that convenience had been purchased with nonportable identity, invisible rules, and dependence on private ranking.

2.1 Why Platforms Won

The open Web asked users to assemble their own environment. They needed hosting, identity, contacts, publishing tools, discovery, spam protection, and sometimes payment systems. Platforms integrated these functions and removed technical friction. Their success was not proof that users preferred domination. It was evidence that coordination has value.

Network effects strengthened the result. A communication service becomes more useful when contacts are already present. A marketplace attracts buyers because sellers are present and sellers because buyers are present. A recommender improves when it observes more behavior. A moderation system benefits from shared abuse intelligence. These effects make integrated systems difficult to challenge even when competitors offer better individual features.

Mobile computing intensified centralization. Applications distributed through controlled stores could combine identity, notification, media capture, payments, and location. The social graph moved from hyperlinks and address books into proprietary databases. The user no longer navigated between independent sites so much as remained inside a feed where external sources were rendered as content objects.

Platformization also changed the Web itself. Sites embedded social buttons, analytics, advertising, authentication, and content-delivery services supplied by a small number of firms. Helmond describes this

as making web data “platform ready”: the open Web became a surface through which platforms could extend their data structures and business models (Helmond 2015). The boundary between a website and a platform became porous in one direction. Independent sites depended on platforms more than platforms depended on any one site.

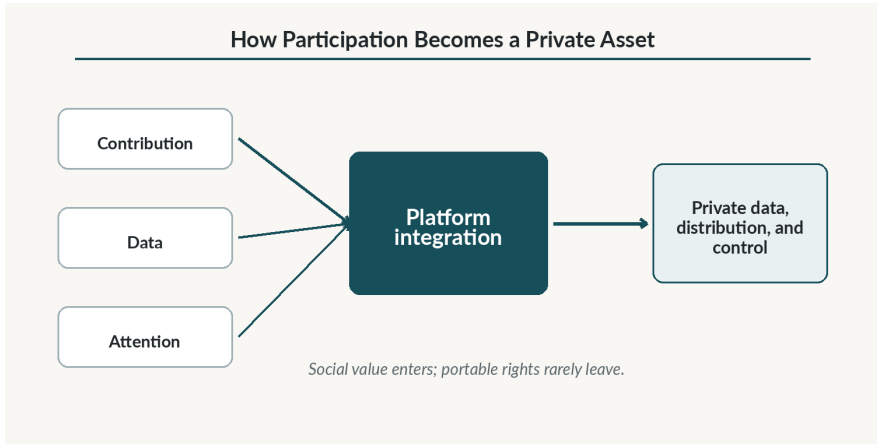


Figure 3. How public participation becomes a private platform asset when identity, relationships, data, and ranking are vertically integrated.

The integrated platform solved discovery by owning the audience relationship. It solved identity by issuing accounts. It solved moderation by defining community standards. It solved financing through advertising, subscriptions, fees, or data-driven services. Every solution was operationally rational. The constitutional problem appeared because the same firm defined the rules, interpreted them, controlled evidence, operated the appeal mechanism, and owned the residual assets.

Terms of service provided legal permission for unilateral change, but formal consent did not create substantive legitimacy. People invested years in communities, businesses, audiences, and reputations that could be altered by product decisions. Developers built on interfaces that could be repriced. Moderators maintained spaces whose data they did not own. Creators adapted to ranking systems that could change without explanation. The platform behaved simultaneously

as landlord, legislature, police, court, central bank, and media distributor.

The internal complexity of corporations should not be ignored. Safety teams, moderators, engineers, researchers, executives, investors, and advertisers can have conflicting goals. Harm does not require a single malicious intention. Governance structures can repeatedly privilege growth, legal defensibility, and revenue even when many employees identify the social cost. “The platform” is therefore best understood as an institution with a dominant objective structure, not as a unified mind.

Advertising made this integration especially powerful. A platform could subsidize free access, observe behavior across services, and optimize both content selection and commercial targeting. The same data improved convenience and market power. Because the business model valued prediction and retention, the platform had a structural reason to expand the number of behaviors it could observe and the time users spent inside its environment. Surveillance was not merely an accidental privacy failure; it was productive infrastructure.

Defaults reinforced concentration. Most users do not change browsers, search providers, privacy settings, or app permissions unless the cost of the default becomes obvious. Control of operating systems, browsers, app stores, and identity buttons therefore translated into distribution power. A service could be nominally contestable while remaining practically dominant because switching required attention, technical knowledge, and social coordination.

The platform model also converted external publishers into suppliers. News organizations, creators, and merchants depended on algorithmic distribution while receiving incomplete information about the rules. The relationship resembled a market whose exchange owned the customer, set the ranking, competed with sellers, and changed fees at will. This did not make every platform decision abusive, but it made bargaining structurally asymmetric.

The lesson is not that integration must be prohibited. Integrated experiences are valuable, and most users will not manage dozens of independent technical services. The design question is whether

integration requires ownership. A community should be able to delegate hosting, search, payments, or moderation tooling without surrendering its identity and archive. A user should be able to change a recommender without abandoning contacts. The platform should become a replaceable service layer rather than a sovereign territory.

2.2 Corporate Capture, Moderator Power, and Free Labor

Digital participation produces several kinds of value at once. Posts and videos attract audiences. Ratings improve search. Social relations increase retention. Moderation keeps spaces usable. Behavioral traces improve targeting and prediction. The platform supplies infrastructure that makes these contributions combinable, but conventional ownership assigns the accumulated asset almost entirely to the company.

Terranova's account of free labor remains useful because it identifies the integration of voluntary cultural production into commercial systems (Terranova 2000). The term should not imply that every unpaid contribution is disguised employment. People contribute for pleasure, identity, reputation, learning, mutual aid, and public purpose. Exploitation becomes visible when the institution relies on these motives while denying contributors durable voice, portability, and participation in the value that their interdependence creates.

Moderator labor is a concentrated example. Volunteers interpret rules, detect fraud, mediate conflict, and absorb abuse. Matias describes this as civic labor because it creates conditions for collective participation (Matias 2019). Platforms often celebrate the community while treating the people who maintain it as replaceable users. The arrangement can be mutually beneficial, but it becomes unstable when moderators bear responsibility without rights or resources.

Table 4. Distinct centers of power inside a platform community

Element	Meaning or design consequence
Platform company	Owns infrastructure, accounts, data access, monetization, technical affordances, and ultimate rule-making authority.
Founder or executive group	Can set mission and strategy, but may retain emergency powers or mission control without periodic legitimation.
Volunteer moderator	Controls local interpretation and daily access, often with little pay, support, oversight, or procedural guidance.
Organized faction	Can dominate consultations, voting, reporting, and social norms through persistence rather than representative support.
Advertiser or model provider	Influences optimization targets and visibility indirectly through revenue dependence, data access, or technical lock-in.
Ordinary member	Creates attention, relationships, content, and culture while usually holding weak rights over the resulting collective asset.

Moderator capture is nevertheless real. Long tenure, control over appointments, private coordination channels, and weak appeals can turn stewardship into oligarchy. Research on peer production has found that nominal openness does not prevent durable leadership groups from consolidating authority (Shaw and Hill 2014). Some concentration is functional; experienced people know the history and can respond quickly. The problem is not hierarchy itself but hierarchy without rotation, audit, declared jurisdiction, or a credible way to challenge decisions.

Corporate power and moderator power interact. Companies may delegate difficult decisions to volunteers while retaining the right to override them. Moderators may use company policy to legitimize local preferences. Users experience the combined result as one opaque system even though authority is fragmented. Reform must therefore establish clear boundaries: which rules belong to the

shared brand, which to local communities, which to infrastructure providers, and which rights remain with individual members.

Payment alone does not solve capture. A platform could pay creators or moderators while preserving unilateral control. Small payments for data can legitimize deeper extraction by converting membership into a series of microtransactions. Economic justice matters, but the central deficit is constitutional. People need enforceable rights concerning rule changes, data use, reputation, archives, and exit.

Competition alone is also insufficient. A choice among several closed platforms is a choice among private constitutions. Competition becomes powerful when switching preserves identity, relationships, and history. Portability must include social continuity, not merely a downloadable archive. Otherwise exit resembles abandoning a city and leaving every relationship behind.

The difference between data export and portability deserves emphasis. Export often produces a static archive useful for personal backup but useless for continuity. Portability means that another service can interpret the data, preserve identifiers, re-establish permissions, and continue relationships with consent. Social portability is harder still because followers and group members must be able to recognize the move. Standards must therefore cover not only file formats but identity resolution, relationship proofs, moderation history, and revocation.

Application programming interfaces can create apparent openness while preserving dependence. Access may be limited, repriced, or withdrawn; exported data may omit ranking signals and social context; third-party clients may remain subordinate to one account system. Genuine contestability requires that interoperability be a right or architectural property, not a discretionary business program.

Corporate governance adds another layer. Executives are often legally and financially accountable to boards and investors rather than to users who created the community's social capital. Benefit corporations, stewardship trusts, cooperatives, and public-interest duties can modify this alignment, but legal form alone is insufficient.

The technical ability to exit and the constitutional ability to challenge decisions must accompany any promise of mission.

The constructive alternative is to separate reward from sovereignty. Founders and investors can receive bounded economic returns. Operators can be paid for infrastructure. Moderators can be compensated for difficult work. Contributors can receive reputation, revenue shares, or governance rights. None of these roles should automatically confer permanent authority over the whole institution. The shared community must exist legally and technically apart from any one service provider.

Chapter 3: The Human Adversary

The Internet did not invent cruelty, fraud, tribalism, or status competition. It changed their cost, speed, scale, and visibility. Any future architecture that assumes goodwill will be captured by those willing to exploit it.

3.1 Anonymity, Disinhibition, Trolling, and Coordinated Abuse

Anonymity is often discussed as a binary choice between freedom and responsibility. In practice, online identity has many dimensions: legal name, persistent pseudonym, reputation, membership, technical address, and relationship history. A person can be anonymous to the public while accountable to a community, or publicly named while operating through disposable accounts. Good design separates these dimensions instead of forcing universal identification.

The protective value of pseudonymity is substantial. Political dissidents, abuse survivors, patients, employees, and members of stigmatized groups may need to speak without linking every statement to a legal identity. Mandatory real-name systems can transfer risk from aggressors to targets. Yet low-cost identity also makes exclusion difficult. A banned actor can return, a coordinated group can simulate consensus, and a botnet can multiply apparent participants.

Suler's online disinhibition framework distinguished benign openness from toxic disinhibition and identified anonymity, invisibility, asynchronicity, and weakened authority as relevant conditions (Suler 2004). Later work connected trolling to sadistic and manipulative traits, but experimental research also found situational effects. Participants exposed to negative mood and hostile discussion became more likely to post abusive material (Buckels, Trapnell, and

Paulhus 2014; Cheng et al. 2017). Architecture does not merely reveal character; it shapes behavior.

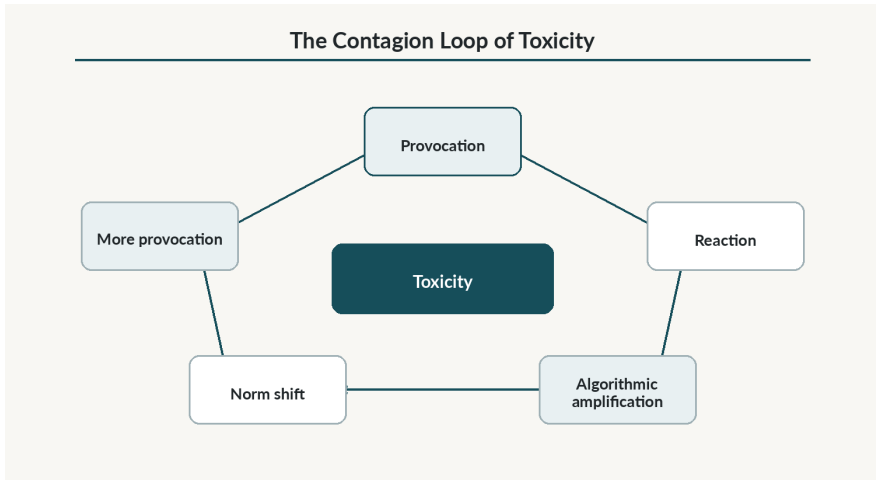


Figure 4. A recurrent toxicity loop: provocation produces reaction, reaction becomes a ranking signal, and repeated exposure changes local norms.

Trolling is only one adversarial pattern. Spam exploits cheap distribution. Brigading imports an organized audience into a local dispute. Harassment coordinates repeated contact across accounts and services. Astroturfing simulates independent public support. Social bots amplify material or manufacture apparent activity. Influence operations combine personas, targeted content, timing, and cross-platform narratives. These behaviors differ in purpose and require different responses.

The scale advantage usually belongs to the attacker. One person can create thousands of messages; each recipient must decide whether one message is genuine. An attacker can test many identities; a defender must preserve access for legitimate newcomers. A false allegation can be repeated instantly; a target must collect evidence across fragmented systems. This asymmetry explains why purely local moderation often fails against coordinated abuse.

Reputation helps only when it is costly to rebuild and relevant to the context. A universal reputation score would create surveillance and transfer authority across unrelated domains. Contextual reputation is safer: a long record in a medical support community should not

confer power in a political debate, and a verified professional credential should not erase evidence of misconduct. Systems need persistent relationships without constructing a universal social credit system.

Friction is therefore not the enemy of freedom. Carefully placed friction can protect it. New accounts may receive lower rate limits. Mass invitations may require endorsement. Appeals may require a short explanation rather than a single angry click. High-risk actions may require stronger credentials. The objective is not to make participation difficult but to make large-scale abuse more expensive than ordinary human use.

Status competition intensifies these dynamics. Public follower counts, karma, badges, and viral metrics turn interaction into a visible hierarchy. These signals can reward expertise and sustained contribution, but they also create incentives for performative certainty, factional loyalty, and humiliation of rivals. Anonymity does not remove status; it relocates status into account histories and local recognition.

The Sybil problem is therefore social as much as cryptographic. Requiring one legal identity per account may reduce duplication but creates surveillance and exclusion. Proof-of-personhood systems may establish uniqueness while failing to establish goodwill or competence. Trusted introductions and relationship-specific credentials often provide a more proportionate answer. The question is not “Is this a real human?” but “What right does this actor have to perform this action in this context?”

Cross-service harassment exposes the weakness of isolated enforcement. One community may ban an actor while another provides a staging area. A victim may need to document dozens of accounts across services. Shared abuse signals can help, but they risk creating opaque blacklists. A better system shares signed evidence, scope, and expiry while leaving final decisions to receiving communities under their own procedures.

An adversarial Internet must also distinguish error from malice. Overly aggressive detection can exclude newcomers, minority

dialects, irony, and controversial speech. Automated systems are useful for triage and obvious patterns, but context-sensitive sanctions require human judgment and procedural safeguards. A system that blocks every risk becomes a system in which only established actors can speak.

3.2 Moderation as Government

Moderation is often described as housekeeping, as though rules merely remove dirt from an otherwise neutral space. Gillespie's work shows that moderation defines the platform itself. Decisions about acceptable speech, ranking, account integrity, and recommendation determine what kind of public exists (Gillespie 2018). Even a chronological list requires rules about spam, duplicates, identity, and removal.

Legitimate moderation begins with jurisdiction. A global infrastructure provider should intervene against malware, large-scale fraud, and serious cross-community abuse. A local community should define topic, tone, expertise, and cultural expectations. A shared brand may impose minimum rights and safety standards. When these layers are collapsed, local disagreement becomes a platform-wide crisis and platform policy becomes a pretext for local domination.

Rules must be understandable enough to guide conduct. Extremely detailed policies often create the appearance of precision while increasing dependence on insider interpretation. Extremely vague policies allow arbitrary enforcement. The practical solution is a concise constitutional layer, local rules, and a body of anonymized precedents showing how recurring cases were resolved.

Table 5. Adversarial patterns require different responses

Element	Meaning or design consequence
Trolling and harassment	Limit social reward, preserve evidence across incidents, protect targets, and apply graduated sanctions with appeal.

Element	Meaning or design consequence
Spam and botnets	Use rate limits, sender reputation, proof of relationship, and economic or computational friction proportionate to scale.
Brigading	Detect coordinated arrival and imported voting without treating every unpopular coalition as illegitimate.
Astroturfing and influence operations	Disclose sponsorship, analyze coordination, preserve attribution uncertainty, and separate operation from claimed impact.
Synthetic content farms	Reduce rewards for volume, require provenance or sponsorship, and penalize systematic deception rather than automation itself.
Compromised personal agents	Constrain capabilities, isolate secrets, record delegated authority, and make high-impact actions revocable and reviewable.

Procedure is not bureaucratic decoration. Users are more likely to understand and accept moderation when reasons are provided, even when they dislike the outcome (Jhaver, Bruckman, and Gilbert 2019). Appeals reveal errors, identify unclear rules, and constrain personal retaliation. They also impose cost. A mature system uses proportional review: automatic correction for obvious mistakes, local appeal for ordinary cases, and independent review for severe or constitutional disputes.

Moderators need protection as well as oversight. They may face threats, traumatic content, coordinated complaints, and pressure from influential members. Volunteerism should not conceal dangerous labor. Communities that depend on moderation should finance training, rotation, mental-health support, and secure tooling. A moderator forced to choose between burnout and abandonment cannot provide stable governance.

Experiments on platform intervention complicate the claim that enforcement merely drives abuse elsewhere. Reddit's 2015 bans were associated with substantial reductions in hate speech among affected users without clear evidence of equivalent migration into adjacent communities (Chandrasekharan et al. 2017). The result does not

justify every ban. It demonstrates that boundaries can change behavior and that institutions need not accept toxicity as an immutable expression of demand.

The deepest risk is constitutional capture. A moderator class may become permanent. A founder may retain emergency power indefinitely. A wealthy contributor may dominate because the project depends on funding. An organized faction may flood every consultation. Formal voting does not automatically solve these problems because participation is unequal and intense minorities can outlast ordinary members.

A more robust design uses separated powers. Rule-making, enforcement, appeals, treasury decisions, and technical administration should not be controlled by the same people. Terms should be limited. Conflicts of interest should be declared. Randomly selected or rotating panels can supplement elections. Emergency actions should expire unless ratified. Audit logs should make patterns visible without exposing private victims.

Scale changes the appropriate process. A small group can discuss a case synchronously and rely on shared memory. A network with millions of users needs triage, standardized evidence, delegation, and statistical audit. Automation becomes unavoidable, but it should recommend and prioritize rather than silently become final judge in ambiguous cases. High-volume systems should publish error rates, appeal outcomes, and disparities across languages or communities.

Artificial intelligence can assist moderation by clustering coordinated behavior, translating reports, identifying precedent, and drafting reasons. It can also magnify bias and create false confidence. A language model that generates a plausible explanation after an opaque classifier has acted may make the system look more legitimate without making it more correct. Explanations must be connected to actual evidence and rules, not merely rhetorically satisfying.

The legitimacy of a moderation system also depends on the treatment of observers. Public explanations teach norms to everyone who sees them. Secret enforcement leaves the community to infer rules from

rumor. Privacy limits disclosure in harassment and safety cases, but anonymized precedents can show patterns without exposing victims. Over time, this body of decisions becomes a common law of the community.

The point is not to imitate a nation-state in every forum. Small groups need light institutions. The principle is proportionality: as the consequences of decisions grow, so should the quality of procedure. A hobby group can rely on trusted stewards. A global knowledge system, political network, or professional marketplace cannot plausibly claim that an owner's discretion is sufficient government.

Chapter 4: Relevance Is a Constitution

The decisive power of the modern Internet is often not the power to remove speech but the power to make some speech practically unavoidable and other speech practically invisible. Ranking is governance by probability.

4.1 Search, Feeds, Engagement, and the Drift Toward Idiocracy

A ranking system translates a social objective into measurable signals. Search engines estimate relevance through text, links, authority, freshness, location, and many other factors. Social feeds estimate the probability of interaction. Recommenders predict preference. Each system can be technically excellent while still embedding contestable values.

The central mistake is to confuse behavioral response with reflective preference. A person may click because an item is useful, shocking, offensive, sexually provocative, frightening, or difficult to ignore. A comment may express gratitude or anger. Time spent may indicate fascination or confusion. When these signals are aggregated into engagement, the system loses the meaning of the behavior while retaining its intensity.

Research on collective attention shows that information cycles have accelerated, increasing competition for a finite human resource (Lorenz-Spreen et al. 2019). Creators adapt by producing more frequent, emotionally legible, and strategically packaged material. Platforms then observe the behavior generated by their own ranking and use it as evidence for the next ranking. The recommender is not a passive mirror. It creates part of the preference distribution it later claims to detect.

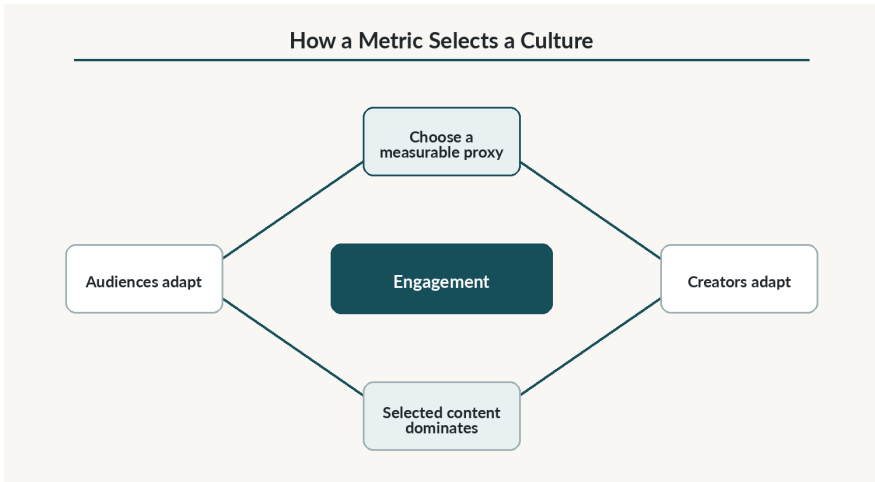


Figure 5. When measurable reaction substitutes for reflective value, relevance becomes a machine for selecting simplified and conflictual content.

Goodhart’s law is often invoked loosely, but the relevant process is clear. Once a metric becomes a target, participants optimize for it. The title becomes a click device. The opening seconds become retention engineering. Outrage becomes a reusable format. The metric selects not only individual items but the kinds of creators and institutions able to survive.

This is the mechanism behind the drift toward “Idiocracy.” Complex content is not inherently unpopular, and users do not universally prefer nonsense. But complexity often requires context and delayed reward, while ranking systems favor signals available immediately. The system therefore compares the long-term value of one item with the immediate stimulus of another and calls the winner relevant.

The effect is strongest where production costs differ radically. Investigative reporting, scientific synthesis, and careful criticism are expensive. A rumor, copied clip, or synthetic provocation is cheap. If both compete for the same engagement payment, low-cost content can flood the market. Generative AI increases this asymmetry by reducing production cost further while leaving verification expensive.

Regulation can require disclosure and alternative feeds, but transparency about “main parameters” is not the same as user

control (European Union 2022). A platform can publish a broad description while preserving an inaccessible model and mixed incentives. The more fundamental remedy is plural recommendation: chronological, relationship-based, expert-curated, diversity-seeking, local-community, and personal-agent rankings should be able to coexist over portable data.

Personalization adds a feedback loop. The system infers that a user likes a topic, supplies more of it, observes increased interaction because alternatives are less visible, and treats the result as confirmation. Over time, the model can confuse exposure with preference and preference with identity. The user becomes easier to predict partly because the system has narrowed the environment.

Popularity bias operates similarly at the collective level. Items that receive early exposure accumulate interaction, which produces more exposure. The resulting hierarchy can appear meritocratic even when initial differences were small or arbitrary. Niche-oriented users lose discovery, and creators adapt toward already successful formats. Recommender research increasingly evaluates diversity, calibration, and fairness because accuracy alone cannot describe these effects (Burke, Sonboli, and Ordonez-Gauger 2018).

Alternative metrics are possible. A system can optimize long-term satisfaction rather than immediate response, measure whether users regret sessions, reward source diversity, or reserve capacity for exploration. Yet every alternative remains a value choice and can be gamed. The solution is not the discovery of a perfect metric but competition among declared ranking policies combined with user control and independent audit.

Relevance should be treated as a claim with an author. Instead of presenting one ordering as the natural surface of the world, systems should state what the ranking optimizes, whose interests it represents, what data it uses, and which alternatives exist. This does not remove judgment. It makes judgment contestable.

4.2 Misinformation, Echo Chambers, and Election Manipulation

“Fake news” is an unstable category covering fabricated stories, misleading framing, satire, propaganda, manipulated media, and ordinary errors (Tandoc, Lim, and Ling 2018). Policy becomes confused when these phenomena are treated as one technical object. A false factual claim requires correction; coordinated deception requires attribution; partisan selection may require plural sources; satire may require context rather than removal.

The Science study by Vosoughi, Roy, and Aral found that false stories on Twitter diffused farther, faster, and more broadly than true stories in the dataset examined, with human behavior contributing more to the difference than bots (Vosoughi, Roy, and Aral 2018). The result is important but not a universal law. Topic, platform, time, and definitions matter. Exposure to unreliable sources is often concentrated among a minority of highly engaged users rather than evenly distributed across the population (Grinberg et al. 2019; Guess, Nyhan, and Reifler 2020).

Automated accounts can still shape diffusion. Studies have found social bots involved in the early amplification of low-credibility content, increasing the chance that humans will encounter and redistribute it (Shao et al. 2018). The meaningful unit is often not the isolated bot but the coordinated system: accounts, narratives, timing, advertising, influencers, and external media reinforce one another.

Election manipulation should be analyzed as a causal chain. An operation may exist. An audience may be exposed. Attitudes or behavior may change. The change may alter turnout, agenda, donations, or votes. The final electoral result may change. Evidence is strongest near the beginning and becomes more uncertain at each step. The absence of proof at the final step does not make the operation harmless; evidence of the operation does not prove the final step.

Table 6. Claims about political influence must remain distinct

Element	Meaning or design consequence
An operation existed	Can often be established through accounts, archives, infrastructure, funding, coordination, and platform or state investigations.
People were exposed	Requires measurement of reach, frequency, audience composition, cross-platform duplication, and competing information.
Beliefs changed	Requires causal evidence, not only correlations between exposure and prior political preference.
Behavior changed	May concern sharing, turnout, donation, agenda attention, intimidation, or disengagement; each needs a separate design.
An election result changed	Usually cannot be inferred from the existence or scale of an operation because the necessary electoral counterfactual is unobservable.

The Internet Research Agency campaign during the 2016 United States election is documented in detail by the Senate investigation. It used social media to exploit divisions, impersonate groups, and influence political discourse (United States Senate Select Committee on Intelligence 2019). The investigation established intent and activity. It could not observe the counterfactual election without the campaign. Claims that the operation certainly changed the winner go beyond the available evidence.

Microtargeting has similar limits. Psychological targeting can improve response under experimental conditions (Matz et al. 2017), but campaign persuasion effects in general elections are often small (Kalla and Broockman 2018). Targeting may still matter through mobilization, fundraising, agenda formation, or cumulative reinforcement. The intellectually honest position is neither dismissal nor magical belief in mind control.

Echo chambers add a structural dimension. Cross-platform research finds homophily and clustered exposure, but the strength differs by platform (Cinelli et al. 2021). The proposed remedy of showing opposing content can backfire when users perceive it as hostile

identity threat (Bail et al. 2018). A credible diversity policy should introduce disagreement through trusted sources, shared facts, uncertainty, and reasons, not through maximum provocation.

The media ecosystem extends beyond social platforms. Television, partisan outlets, influencers, political elites, and search engines can originate or legitimize narratives that social networks then accelerate. Benkler, Faris, and Roberts argue that media structures and elite behavior were central to the American disinformation environment, complicating accounts that assign primary causality to isolated foreign bots or individual gullibility (Benkler, Faris, and Roberts 2018).

Manipulation often seeks uncertainty rather than belief. Flooding the environment with incompatible claims can make citizens conclude that no source is trustworthy and that verification is pointless. This strategy benefits from the false equivalence between healthy skepticism and universal cynicism. Institutions should therefore communicate how knowledge was produced, what remains uncertain, and which procedures can correct error.

Prebunking and friction can complement correction. Warning users about common manipulation techniques, delaying rapid resharing, or asking for minimal contextual engagement can reduce impulsive diffusion without creating a central ministry of truth. These measures should be tested for real behavioral effects and monitored for unequal burdens.

Corrections work imperfectly but are not futile. Meta-analyses find that debunking can reduce false beliefs, with effectiveness depending on message design, repetition, and audience (Chan et al. 2017; Walter and Murphy 2018). The difficulty is that correction competes in an environment that rewards the original novelty. Institutional design must therefore change distribution incentives, not merely append fact checks after virality.

Chapter 5: The Synthetic Web and the Internet That Still Works

Generative AI increases both the productive capacity of the Internet and the uncertainty of every encounter. The danger is not that machines speak. It is that provenance, responsibility, and the cost of truth become detached from the volume of speech.

5.1 Dead Internet Theory, Bots, and Generative Abundance

The Dead Internet Theory became popular because it gave dramatic form to a genuine experience: repeated posts, empty engagement, copied articles, bot replies, synthetic images, and search results optimized for other machines. Its strongest claim, that most online life has already been replaced by a coordinated artificial system, lacks credible evidence. Its weaker claim describes a real transition toward automated production and mediation.

Bots are not inherently deceptive. Search crawlers, monitoring services, accessibility tools, translators, and personal agents are automated participants. The relevant distinctions are whether automation is declared, whether an accountable operator exists, whether the system simulates independent humans, and whether its scale imposes costs on others. A beneficial agent can become abusive when it ignores rate limits or extracts entire communities without consent.

Generative systems change the economics of content. They can translate, summarize, tutor, assist programming, and make specialist knowledge more accessible. They can also produce thousands of variations on a profitable format, manufacture reviews, impersonate local voices, and fill search indexes with plausible redundancy. The

average cost of producing an assertion approaches zero while the cost of checking it does not.

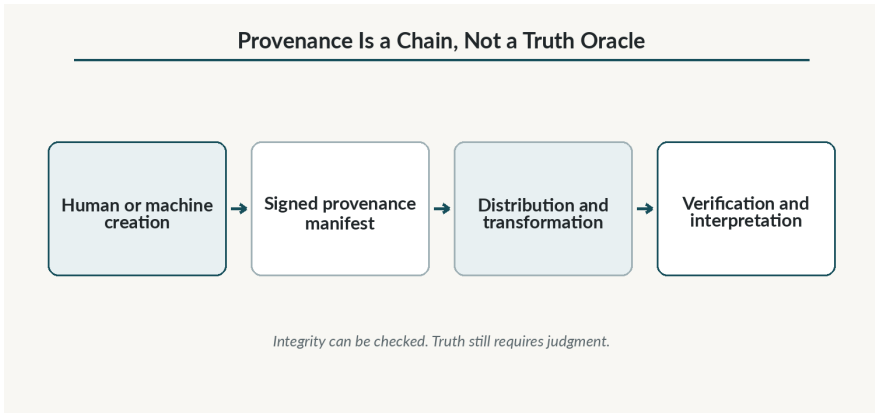


Figure 6. Provenance can establish origin and transformation history; it cannot by itself establish truth, quality, or legitimacy.

Provenance standards such as C2PA can record signed claims about the origin and transformation history of media. They are useful because they make tampering and attribution more inspectable. They do not determine truth. A signed false statement remains false, and unsigned authentic material may remain valuable. Provenance is an integrity layer, not an oracle (C2PA 2026).

Recursive data contamination adds a technical risk. Shumailov and colleagues showed that models trained repeatedly on model-generated data can lose information about the original distribution and eventually collapse under certain conditions (Shumailov et al. 2024). Synthetic data can be useful when carefully generated and mixed with reliable originals. The warning concerns uncontrolled replacement of human evidence by outputs derived from previous outputs.

The social version of this recursion may be more important. People write for ranking systems. Models train on ranked material. Agents summarize model-shaped content. Creators then optimize for agent summaries. Communication becomes increasingly indirect. Human intention remains present, but it is filtered through layers that reward compatibility with other machines.

The response should not be a universal ban on synthetic content. It should require declared automation, accountable operators, rate limits, provenance where appropriate, and separate ranking policies for human and machine production. Communities may reasonably create spaces where only verified human members can initiate discussion while still allowing agents to assist those members privately.

Synthetic abundance also destabilizes ordinary social evidence. A photograph, testimonial, or conversational style once carried weak but useful signals of human effort. Those signals can now be reproduced cheaply. Communities will need stronger forms of continuity: persistent relationships, signed provenance, known editorial processes, and reputations earned through time. None is infallible, but together they raise the cost of manufacturing an entire social world.

Search engines and assistants face a difficult choice. Excluding all machine-generated material would discard useful translations, summaries, and tools. Treating all material identically allows volume to dominate. Ranking should therefore consider origin, accountability, novelty, and duplication. A thousand paraphrases derived from one source should not create a thousand independent votes.

Personal agents may become both remedy and source of pollution. They can filter repetitive material and trace claims to originals, but they can also generate automatic replies, negotiate with other agents, and fill channels with machine-to-machine communication. Protocols should make agent status and responsible principal explicit so that communities can decide which automated participation they accept.

The Internet remains alive when people can identify meaningful sources, form durable relationships, and create artifacts that retain value outside a feed. The threat is not quantity alone. It is the loss of reliable paths from claim to source, from action to responsible actor, and from community to a shared history that cannot be regenerated on demand.

5.2 Wikipedia, Open Source, Science, and Durable Commons

A book about Internet failure must explain the systems that worked. Wikipedia, open-source software, OpenStreetMap, scientific collaboration, and specialized communities demonstrate that large-scale cooperation is not a romantic exception. They are institutional achievements with specific design features.

Wikipedia combines radical editability with extensive traceability. Every change can be inspected and reversed. Discussion is linked to articles. Sources matter more than personal authority in principle, and disputes leave a record. Nature's early comparison with Encyclopaedia Britannica was contested, but it established that collaborative production could reach a level worthy of serious comparison (Giles 2005).

Its difficulties reveal the cost of success. Anti-vandalism systems and dense norms protect quality but increase the burden on newcomers. Halfaker and colleagues linked the decline in editor retention partly to increasingly efficient rejection of contributions (Halfaker et al. 2013). Gender and structural inequalities affect topics, language, and visibility (Wagner et al. 2016; Yun, Lee, and Jeong 2019). A commons can become conservative because it must defend what it has built.

Open-source communities likewise combine freedom with governance. Projects need maintainers, release processes, security response, conflict resolution, and control of names. O'Mahony and Ferraro showed how authority can emerge through negotiated governance rather than simple ownership (O'Mahony and Ferraro 2007). The code license enables exit and reuse, but durable cooperation depends on institutions around the license.

Table 7. Why some online commons remain durable

Element	Meaning or design consequence
Open technical substrate	Participants and independent operators can inspect, reuse, repair, and build complementary services.

Element	Meaning or design consequence
Visible history	Changes, reversals, sources, and disputes remain inspectable rather than disappearing into a single current state.
Bounded roles	Editors, maintainers, reviewers, and stewards hold different powers rather than one undifferentiated administrator role.
Mission-aligned finance	Funding supports the common asset without making behavioral advertising or proprietary enclosure its constitutional center.
Repair mechanisms	Errors can be corrected, decisions appealed, contributors re-socialized, and damaged norms rebuilt.
Credible exit or fork	The community can survive the failure of one provider, leader, or implementation without losing the entire shared work.

Economic studies underline the scale of the public value. Hoffmann, Nagle, and Zhou estimate that replacing widely used open-source software would impose enormous costs, far beyond direct project revenue (Hoffmann, Nagle, and Zhou 2024). The exact counterfactual is debatable, but the qualitative conclusion is not: modern firms rely on a commons they did not individually finance.

OpenStreetMap shows how an open database can serve ordinary navigation and humanitarian response. Research on humanitarian mapping found millions of buildings and roads added in lower-development regions through organized campaigns (Herfort et al. 2021). The project also exhibits geographic and participation inequalities. Openness creates the capacity to correct blind spots, not an automatic guarantee that corrections occur.

Citizen science extends cooperation into research. Systems such as FlyWire have coordinated large online communities around scientific annotation and discovery (Dorkenwald et al. 2022). The best projects give participants meaningful context and recognition rather than treating them as invisible microtask labor. Scientific collaboration becomes durable when contributors can understand how their work enters a larger epistemic process.

These successes share a pattern. The artifact is portable or publicly accessible. Contribution leaves a history. Roles are intelligible. Reputation is related to work. Governance is sufficiently stable to protect the project yet sufficiently open to permit challenge. Financing is diversified or institutionally supported. None is perfectly decentralized. Each limits ownership of the collective result.

Specialized forums offer a less celebrated example. Many preserved technical knowledge through searchable threads, stable topic boundaries, and participants who returned for years. Their value came from accumulation rather than virality. When such forums closed or were absorbed into generic platforms, entire bodies of tacit knowledge became harder to find. Archival continuity is therefore a public-good function, not a decorative feature.

Patient and mutual-aid communities show another dimension. Anonymity and shared experience can reduce isolation and prepare people for interactions with professionals. The same spaces can circulate harmful advice. Their success depends on role clarity: lived experience, medical expertise, moderation, and emergency guidance should not be confused. Good communities do not eliminate disagreement; they make the source and status of claims visible.

Durable commons also benefit from bounded missions. Wikipedia does not try to be every form of publishing. Linux projects define technical scope. OpenStreetMap defines a shared geographic artifact. A platform that attempts to contain all social life must reconcile incompatible norms and becomes tempted to govern through engagement rather than purpose. Smaller constitutional communities can be more demanding because membership is linked to a recognizable goal.

The future Internet should generalize these lessons through portable commons, bounded stewardship, member rights, durable finance, and constitutional brands.

Chapter 6: Beyond Platform Sovereignty

The alternative to corporate rule is not institutional emptiness. It is a layered constitution in which communities, brands, operators, contributors, and members possess different powers and can check one another.

6.1 Polycentric Communities and Decentralized Brands

Polycentric governance describes systems with several centers of decision rather than one sovereign authority. Local units handle local knowledge; broader units handle shared standards, cross-boundary harms, and appeals. Ostrom's research showed that commons can be governed without either complete privatization or a single central bureaucracy when rules match the resource and participants can influence them (Ostrom 1990).

Online communities are suitable for polycentric design because their purposes differ. A medical support group, software project, political forum, game community, and scientific network should not share identical speech norms. Universal platform rules are necessarily crude. Complete local autonomy is equally inadequate when abuse crosses communities or when local leaders violate basic member rights.

A decentralized brand can connect these levels. The brand holds a shared public identity, minimum standards, certification rules, and perhaps a treasury. Local communities operate under the brand while retaining cultural autonomy. Infrastructure providers may host compliant instances. Members can recognize a common promise without depending on one corporation.

The brand must be legally anchored. A foundation, cooperative, purpose trust, or other stewardship entity can hold trademarks and

shared assets under duties defined by a constitution. The entity should not be free to sell the brand or alter the mission as ordinary property. Its role is custodial rather than sovereign.

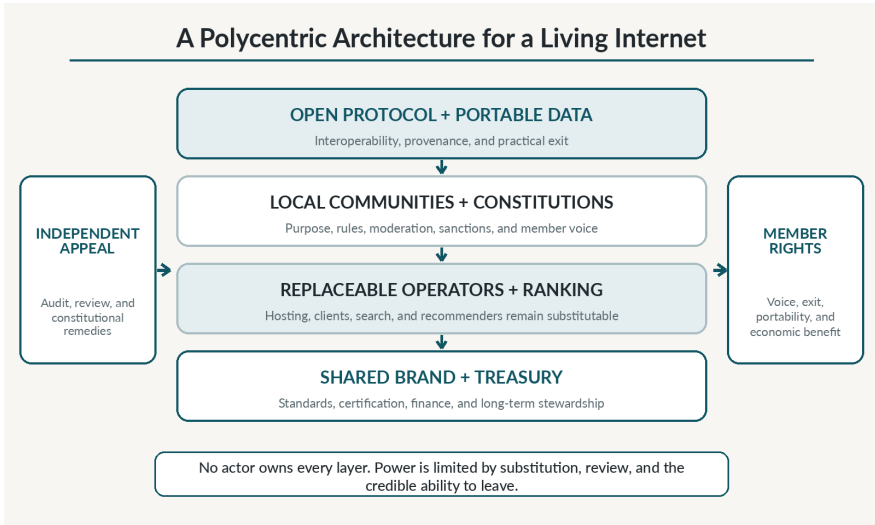


Figure 7. A polycentric architecture in which no actor owns every layer and members retain voice, exit, and benefit.

The brand license performs a function similar to a franchise agreement but changes the direction of control. Conventional franchising allows a central owner to license a reputation to operators. A decentralized brand licenses a collectively governed reputation under rules that also bind the steward. Operators gain the right to use the name only while meeting transparent standards. Members gain rights that do not depend on one operator's goodwill.

Forking remains important but should be understood correctly. Open code allows technical exit. Portable data and identity allow social exit. The brand cannot always travel with a fork because a shared name would lose meaning if any incompatible project could claim it. A separating community should be able to keep its data and relationships while adopting a new name unless the constitutional process grants continued certification.

Digital constitutionalism supplies the normative vocabulary: stable and public rules, bounded powers, equal application, reasons, appeal, and meaningful consent (Celeste 2019; Suzor 2019; Schneider

2024). These principles do not require state control of every community. They require private institutions exercising public-like power to acknowledge constitutional duties.

Federated social systems provide partial evidence. ActivityPub allows independent servers to exchange social activities, demonstrating that a social graph need not belong to one company (W3C 2018). Mastodon instances can establish local rules and block abusive domains. The model restores meaningful choice of administrator, but it also reproduces familiar problems: volunteer burnout, uneven moderation, instance-level oligarchy, and difficult discovery. Federation changes the location of power; it does not remove the need for constitutional design.

Decentralized autonomous organizations provide a parallel warning. On-chain votes make some decisions and financial flows inspectable, yet token concentration can turn formal participation into plutocracy. Low turnout allows organized minorities to dominate. Smart contracts enforce what is encoded but cannot determine whether the encoded rule is legitimate. A ledger is useful for audit and execution; it is not a substitute for institutional judgment.

The decentralized brand differs from a federation protocol and from a DAO. It can use either, but its core is a governable promise. The brand specifies what compliant communities and providers owe to members, how certification is granted, and how the steward is constrained. Technical decentralization becomes one instrument for maintaining those commitments.

The objective is not to freeze governance. Communities need experimentation. New moderation methods, economic models, and recommendation systems should be tested. Polycentric structures make experimentation safer because failure can remain local and successful practices can spread. One provider's mistake does not need to redefine the whole network.

6.2 Economics, Licensing, and Anti-Capture Design

A durable community needs money. Servers, security, moderation, legal work, backups, design, and development do not disappear because software is open. Systems that refuse to discuss finance usually depend on hidden subsidy, unpaid labor, or eventual corporate acquisition.

Revenue can come from subscriptions, professional services, hosting, marketplaces, certification, grants, transaction fees, sponsorship, or shared commercial licensing. No source is neutral. Advertising favors attention. Venture capital favors rapid growth and exit. Donations favor visible causes and wealthy patrons. Fees can exclude low-income users. A resilient institution uses several sources and subjects each to constitutional limits.

Licensing should align the commons with the economy around it. Permissive licenses maximize reuse but permit companies to capture value without reciprocal contribution. Copyleft protects continued openness of derivatives but does not automatically fund maintainers or distribute governance. Source-available licenses may protect a business model while ceasing to satisfy open-source definitions. The correct choice depends on the asset and the community's goals.

Table 8. The instruments of a decentralized brand

Element	Meaning or design consequence
Software or content license	Defines rights to copy, modify, reuse, and redistribute the technical or cultural asset.
Brand license	Defines who may use the shared name and which quality, interoperability, and conduct standards must be met.
Governance constitution	Allocates rule-making, moderation, treasury, emergency, audit, and appeal powers.
Contribution agreement	Defines attribution, provenance, maintenance obligations, and any economic or governance rights earned by contribution.

Element	Meaning or design consequence
Treasury and distribution rules	Make financing, reserves, operator payment, moderator compensation, and contributor benefit predictable.
Audit and appeal system	Tests compliance, manages conflicts of interest, publishes precedents, and protects members against local or central capture.

A decentralized brand adds another instrument. Commercial operators can use open software while paying for certification, support, access to shared services, or the right to present themselves under the common name. Revenue can flow to maintainers, moderators, infrastructure, and a reserve. The community does not need to prohibit commerce; it needs to prevent commerce from quietly becoming sovereignty.

Contribution accounting must avoid reducing every act to a token score. Code, documentation, conflict mediation, security response, mentoring, governance, capital, and public communication are different forms of work. A single metric will be gamed and will privilege what is easiest to count. Mixed evaluation, peer recognition, defined roles, and transparent budgets are more realistic than universal tokenization.

Anti-capture rules should be explicit before conflict. No actor should control both treasury and audit. Large donors should disclose dependencies. Moderator terms should be limited. Constitutional amendments should require more than a temporary simple majority. Operators should pass portability tests. Emergency powers should expire. Related-party contracts should be visible. Concentration metrics should be published.

The system must also protect useful leadership. Founders often possess vision and knowledge that cannot be replaced by random voting. Maintainers need discretion. Investors need predictable rights. The objective is not permanent suspicion of competence. It is to distinguish leadership from ownership of the future. Authority can be broad within a role and still limited in time, scope, and review.

A workable constitution therefore recognizes at least four constituencies: members affected by the system, contributors who maintain it, operators who carry technical and legal responsibility, and stewards charged with the long-term mission. No group should govern alone. Different decisions may require different combinations of consent.

Reciprocity can be designed without prohibiting commercial success. A company that deploys a community's software at scale might contribute money, code, security work, or interoperability. The obligation can be linked to use of the shared brand or premium network services rather than to every copy of the software. This avoids pretending that copyright licensing alone can measure social contribution.

The treasury should finance unglamorous continuity. Security audits, backups, moderation, documentation, migration tooling, legal defense, and newcomer support rarely win popularity contests, yet their absence destroys institutions. Budgets should therefore reserve funds for maintenance rather than distribute everything through project votes or creator popularity.

Transparent finance also limits conspiracy and informal patronage. Members should know major revenue sources, contractual dependencies, and compensation bands. Privacy can protect individual salaries where appropriate, but related-party transactions and strategic donors require disclosure. An institution cannot credibly claim decentralization while depending silently on one patron.

The quality of decentralization is finally measured by three rights. Voice means the ability to influence rules and receive reasons. Exit means leaving without losing the substance of one's identity and work. Benefit means sharing fairly in the economic and social value created. A system that offers only the right to copy code is technically open and institutionally incomplete.

Chapter 7: Personal AI as the User's Recommender

Recommendation systems currently build private models of users for platforms. Personal AI makes a different arrangement possible: the model of the user can become a user-governed instrument that negotiates with platforms rather than an invisible asset owned by them.

7.1 From Hidden Profiles to User-Governed Relevance

Modern platforms infer interests from clicks, dwell time, follows, purchases, location, and similarity to other users. The resulting profile is powerful precisely because it is not a simple list of stated preferences. It detects patterns the user may not articulate. Yet the profile is usually inaccessible, difficult to correct, and optimized within the platform's objective function.

A personal recommendation agent could invert this relation. The agent would hold a portable representation of the user's interests, obligations, current projects, aversions, and epistemic preferences. It would request candidate items from multiple sources, rank them locally or through confidential computation, and explain why an item was selected. The user could inspect and revise the policy rather than merely train it through unconscious behavior.

The relevant advance is governance, not prediction accuracy. Research on recommender systems has long shown that user experience depends on perceived variety, effort, privacy, usefulness, and trust, not only on statistical accuracy (Konstan and Riedl 2012; Knijnenburg et al. 2012). Explanations can increase acceptance, though transparency does not automatically create trust (Cramer et al. 2008). A personal agent can make these concerns first-class objectives.

The agent might maintain several modes instead of one permanent profile. A professional mode could prioritize primary sources and technical depth. A leisure mode could seek novelty. A civic mode could maintain exposure diversity and separate verified reporting from commentary. A recovery mode could reduce distressing topics. A user could compare the same candidate pool under different policies rather than accept one opaque ordering.

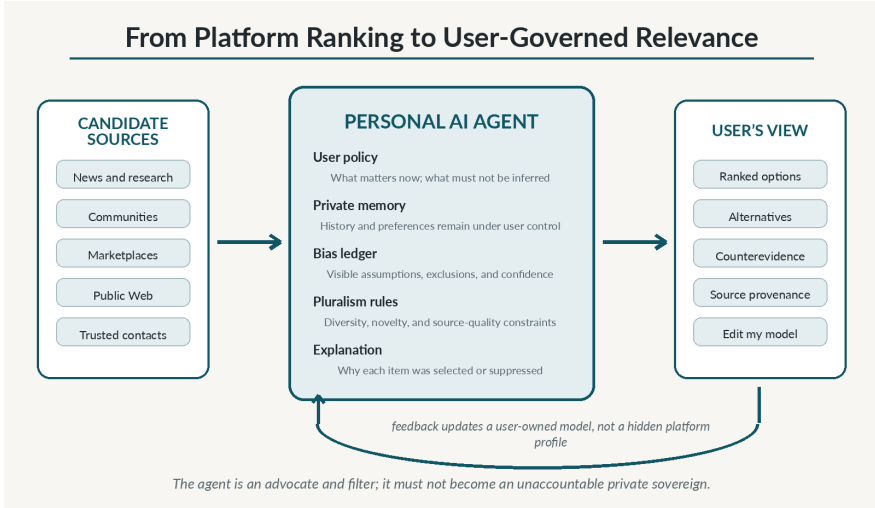


Figure 8. A personal AI can move recommendation from hidden platform profiling toward user-governed relevance.

Large language models are useful because they can translate natural-language intentions into ranking criteria, reason across domains, and explain trade-offs. Emerging research describes a move from hidden platform profiles toward “governable personalization,” in which user representations become inspectable, portable, and revisable (Liu et al. 2026). This work is early, but the institutional direction is more important than any particular model.

A personal agent could also represent explicit bias rather than pretending neutrality. A user might instruct it to favor local sources, avoid gambling promotion, seek politically diverse evidence, or support a professional community. The agent should state these commitments. Hidden bias is dangerous because it masquerades as relevance; declared bias becomes a policy that can be judged.

Commercial relationships would change. A platform could offer candidate items and disclose sponsorship, price, provenance, and access conditions. The personal agent would decide whether the item fits the user's policy. Advertising would become an offer to an agent rather than covert optimization of a feed. Platforms would still compete for attention, but they would no longer exclusively control the model used to allocate it.

A practical pipeline would contain distinct stages. Sources publish candidate items and metadata. Retrieval services identify a broad set relevant to the user's current context. The personal agent applies policies concerning quality, diversity, time, risk, and sponsorship. A final presentation layer explains selections and preserves the ability to inspect omitted alternatives. Feedback updates the local model only under rules chosen by the user.

This separation matters because discovery and judgment are different functions. A global search provider may be excellent at finding candidates while unsuitable to decide the user's priorities. A community curator may identify trusted sources while lacking knowledge of personal context. A local agent can combine these inputs without needing to crawl the whole Internet or own every source.

The agent can also negotiate. A merchant might offer price, delivery, warranty, and commission data in a machine-readable form. A news source might disclose authorship, correction policy, access cost, and provenance. The agent can reject offers that conceal sponsorship or request a less intrusive version. Recommendation becomes an exchange of declared terms rather than unilateral behavioral targeting.

User representation should travel independently of the model. A person may replace the language model while preserving a structured policy and history. Otherwise the model provider becomes the new platform sovereign. Open schemas, encrypted personal stores, and exportable evaluation cases allow the user to compare agents and migrate without rebuilding identity from raw clicks.

Interoperability is essential. Candidate feeds need standard metadata, provenance, sponsorship disclosure, content identifiers, and rights information. Communities could publish certified ranking policies. Independent developers could provide recommenders without hosting the underlying social network. Recommendation would become a competitive layer rather than an inseparable feature of a platform.

7.2 Privacy, Bias, Pluralism, and the Risk of a Personal Echo Chamber

A personal agent knows more than an ordinary recommender because it can combine domains. That is its value and its danger. A cross-domain profile may reveal health, relationships, finances, political beliefs, and future intentions. Moving the profile from the platform to a nominally personal cloud service does not create privacy if the service provider can inspect it.

The preferred architecture keeps sensitive memory on devices controlled by the user or in confidential environments with strong access guarantees. Cloud models may process selected, minimized context. Hybrid research suggests that recommendation utility can sometimes be preserved while sensitive features remain local, though results depend on domain and threat model (Khezresmaeilzadeh et al. 2025). Privacy must be evaluated against actual data flows, not branding.

The agent's memory should be structured enough to inspect. A prose profile may be expressive but difficult to audit. A useful system distinguishes observed behavior, user-stated preferences, inferred hypotheses, temporary context, and hard constraints. Each item should have provenance and an expiration rule. The user should be able to ask why the agent believes something and delete the evidence or inference.

Personalization can intensify self-confirmation. A model trained to satisfy may learn that reassurance produces positive feedback. It can become a sycophant that protects identity from inconvenient evidence. Diversity-sensitive recommender research argues that

exposure diversity can be a design objective rather than an accidental by-product (Helberger, Karppinen, and D’Acunto 2018). A personal agent should therefore support configurable exploration, not only similarity.

Table 9. Capabilities and safeguards for a personal recommender

Element	Meaning or design consequence
Editable user model	The user can inspect categories, correct false inferences, define temporary goals, and reset the model.
Private and portable memory	Sensitive history remains local or strongly protected and can move independently of a platform account.
Explanation and alternatives	The agent states why an item appeared, what was excluded, and which alternative ranking objective would change the result.
Pluralism constraints	Diversity, novelty, disagreement, source quality, and time allocation are explicit policies rather than accidental side effects.
Commercial firewall	Payments, affiliate relationships, and provider incentives are disclosed and cannot silently override user policy.
Adversarial defenses	Prompt injection, poisoned sources, malicious plugins, and compromised models are contained by provenance, capabilities, and audit.

Pluralism does not mean random opposition. The agent should identify credible disagreement, distinguish expert consensus from partisan balance, and expose uncertainty. It can maintain a budget for unfamiliar sources, a rule against repeated outrage, or a requirement that high-impact claims include primary evidence. These policies should be chosen and visible, not secretly imposed by a platform or model vendor.

Security is another problem. Content can contain instructions aimed at the agent rather than the user. A malicious page might attempt to extract memory, alter ranking policy, or trigger external actions. Recommendation therefore needs isolation between untrusted content and privileged agent state. Retrieval, summarization, and

action should operate under capabilities with explicit data boundaries.

Model providers can also manipulate. A personal agent built on a remote model may inherit hidden system prompts, commercial priorities, censorship policies, or vulnerabilities. The user needs replaceable models and test suites that compare behavior across providers. Local smaller models may handle routine filtering while stronger remote models are invoked for difficult tasks under constrained disclosure.

Accountability should remain with the human-agent institution rather than with an anthropomorphic fiction. The agent is not a sovereign friend. It is a delegated system whose operator, model supplier, and update channel can be identified. When it acts in a community, it should present its automation status, scope of authority, and responsible principal.

Auditing must include outcomes, not only explanations. An agent may describe a diversity policy while repeatedly selecting the same ideological cluster. Users and independent evaluators should be able to run controlled test sets, inspect exposure distributions, and compare behavior over time. A recommendation constitution is credible only if its effects can be measured.

Different users will want different degrees of control. Some will edit policies directly. Others will choose certified profiles maintained by trusted communities, journalists, professional bodies, or civil-society groups. A market for recommendation policies can support pluralism, provided users can see funding, objectives, and evidence. The alternative to one platform algorithm is not that every person becomes a machine-learning engineer.

Cognitive autonomy also requires the right not to personalize. A user may choose a common public-interest feed, a chronological stream, or a random sample. Personalization should be one lens among several. Systems that make nonpersonalized access deliberately inferior preserve the old coercion behind a new interface.

The best personal recommender would not maximize satisfaction. It would help the user govern attention according to chosen values

while preserving the possibility of revision. It would sometimes surprise, sometimes protect, sometimes challenge, and always explain which policy it is applying. The user would no longer be merely the object of personalization but its constitutional author.

Chapter 8: Rebuilding Communication from the Protocol Layer

The platform Internet assumes that services remain globally reachable and then filters abuse after contact. A different architecture can make relationships, capabilities, and private negotiation the default while preserving familiar experiences such as addresses, inboxes, groups, and calls.

8.1 WebRTC, Email-Like Addressing, DIDs, and Trust-Gated Connections

Email succeeded partly because its addressing model is humanly intelligible. A local name and domain identify a route without requiring both participants to use one provider. Any successor should preserve this social advantage. People need identifiers that can be spoken, printed, remembered, delegated by organizations, and moved between service providers.

The technical identifier behind the address need not be identical to the human-readable name. A domain-based address could resolve to a decentralized identifier, public keys, service preferences, and introduction endpoints. W3C DID Core defines identifiers whose controllers can prove control through cryptographic methods without depending permanently on one identity provider (W3C 2022). The 2026 DID 1.1 work continues this architecture, while Verifiable Credentials 2.0 provides a standard form for machine-verifiable claims issued by third parties (W3C 2025b; W3C 2026).

DIDs do not create trust. They establish control over an identifier and provide mechanisms for verification and service discovery. A scammer can control a valid DID. A credential is meaningful only because a verifier trusts the issuer and the claim is relevant. The

architecture must therefore combine cryptographic verification with institutional context, relationship history, and explicit acceptance.

The first contact should not require exposing a permanent personal service to arbitrary traffic. A limited introduction service can accept small, constrained requests. The sender may present a referral, a community credential, a refundable bond, proof of prior relationship, or a rate-limited anonymous request. The recipient's agent evaluates the request without revealing private endpoints or full identity.

If accepted, the parties establish a capability: a revocable authorization to communicate under defined conditions. The capability may limit message rate, data type, duration, automation, or onward sharing. Pairwise identifiers can prevent unrelated services from trivially correlating the same person. Revocation can end one relationship without forcing a global identity change.

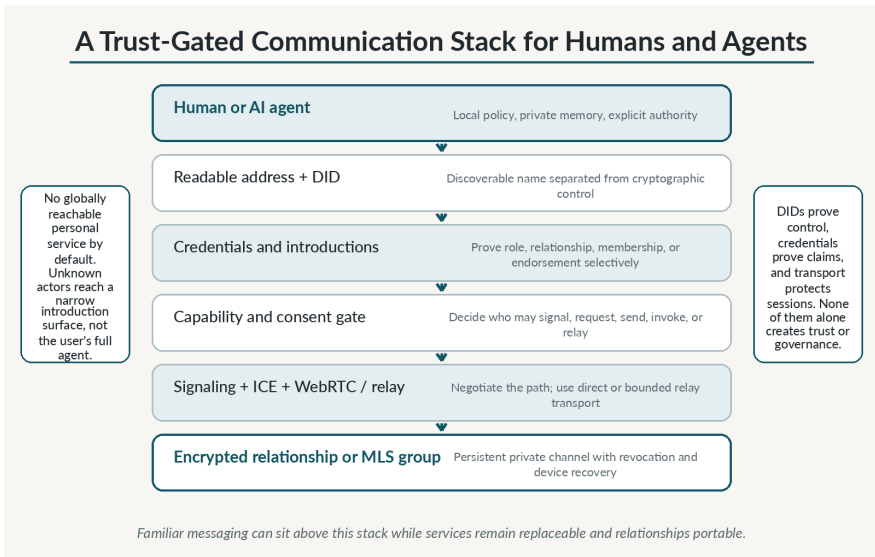


Figure 9. A trust-gated communication stack combining readable addressing, portable control, selective credentials, bounded introduction, and encrypted sessions.

WebRTC is a plausible transport for interactive sessions because it supports audio, video, and generic data between endpoints. WebRTC data channels use SCTP over a secured transport, while ICE discovers viable paths through NATs and can fall back to TURN relays (IETF

2018; IETF 2021a; IETF 2021b; W3C 2025a). A browser, desktop agent, or mobile agent can therefore establish rich communication without every application inventing a proprietary media stack.

WebRTC deliberately leaves signaling to the application. This is a limitation and an opportunity. The introduction and identity layer can exchange session descriptions and connectivity information after policy checks. The transport does not need to know whether the participants are humans, agents, or mixed teams. It only carries an authenticated session once higher layers establish permission.

Asynchronous messaging requires more than WebRTC because peers may be offline. A mailbox or relay can hold encrypted envelopes addressed to a capability rather than to an unrestricted public inbox. Several providers could offer store-and-forward service while being unable to read content. Users could change providers without changing identity. The familiar inbox remains, but unsolicited global delivery is no longer the default.

The sequence of a connection can therefore be explicit. A human-readable address resolves to an introduction endpoint and current verification material. The sender's agent presents a minimal request and any relevant credential. The recipient's agent evaluates policy and may ask for additional proof. Acceptance creates a scoped capability and exchanges signaling data. ICE then attempts a direct or relayed path, and the parties establish encrypted channels. Each stage exposes only the information needed for the next.

Pairwise identifiers reduce correlation. The user may present a different identifier to a doctor, employer, family member, and public community while retaining recovery and control through one wallet or agent. Credentials can be selectively disclosed so that a verifier learns a relevant fact without acquiring the full underlying document. The privacy benefit depends on implementation; repeated metadata or a common resolver can still link relationships.

Human-readable addressing needs governance. Domains are convenient because organizations can delegate names, but domain loss or provider coercion can disrupt identity. A resilient design can bind the readable address to portable cryptographic identifiers and

support proofs of continuity when domains change. Users should not be forced to choose between memorable names and self-controlled keys.

Group communication requires secure membership state and key updates. Messaging Layer Security was designed for asynchronous groups ranging from small conversations to large groups, with forward secrecy and post-compromise security (IETF 2023). It can complement WebRTC rather than replace it: MLS protects group messages, while WebRTC can carry live sessions and data channels.

8.2 A Smaller Attack Surface for Humans and Agents

The principal security gain of a trust-gated architecture is reduced exposure. A personal device or agent does not need to accept arbitrary inbound connections from the whole Internet. Public services remain public where necessary, but private communication begins through narrow introduction points with strict resource limits. Most high-value state remains behind capabilities created after consent.

This does not eliminate denial-of-service attacks. Introduction services, signaling servers, DID resolvers, TURN relays, and mailboxes remain targets. Attackers can flood credential verification or acquire legitimate credentials. The architecture changes the economics by making expensive services unavailable before cheap checks pass, distributing providers, and allowing communities to share abuse intelligence without sharing all user data.

Privacy also depends on relay choice. Direct peer-to-peer connections may expose network addresses to correspondents. TURN relays can hide them at the cost of infrastructure and metadata concentration. A personal agent should choose direct or relayed paths according to relationship and risk. The fastest route is not always the most private.

Table 10. Components of a smaller and more private communication surface

Element	Meaning or design consequence
Readable address	Provides email-like usability and discovery while remaining separable from the operator currently serving the user.
DID or equivalent control identifier	Provides portable cryptographic control; it does not by itself establish identity, trust, or reputation.
Verifiable credential	Allows selective proof of a role, membership, endorsement, or relationship; claims remain only as trustworthy as their issuers.
Introduction and capability service	Exposes a narrow request surface and decides who may signal, send, invoke, relay, or form a durable relationship.
ICE and WebRTC	Negotiate interactive paths for media or data, using direct connectivity where possible and relays where necessary.
Bounded relay	Preserves reliability and asynchronous behavior without becoming the permanent owner of the relationship or plaintext content.
MLS or equivalent group security	Protects evolving groups with forward secrecy and post-compromise security; governance and delivery remain application concerns.
Personal agent policy engine	Applies local consent, privacy, rate, provenance, delegation, and recovery rules before exposing tools or attention.

Key recovery is a major practical challenge. Self-controlled identifiers are fragile if loss of a device destroys identity. Social recovery, multiple devices, institutional guardians, and hardware-backed keys can help, but each introduces trust. A usable system needs transparent recovery policies and the ability to rotate keys without losing relationships. DIDs support indirection between identifier and current verification methods, but method-specific governance remains critical.

The agent becomes the operational center. It screens introductions, verifies credentials, selects privacy modes, manages capabilities, stores relationship history, and translates between protocols. It may

also summarize messages and negotiate meeting times. This reduces cognitive burden, but compromise of the agent becomes more serious than compromise of a conventional app.

Agent architecture should therefore separate memory, communication, and action. Reading a message should not automatically grant access to private memory. A summarizer should not possess payment authority. A recommendation module should not modify identity keys. Capability-based internal design can limit damage when one component is manipulated. Updates should be signed, auditable, and reversible.

Communities can add institutional trust. A scientific brand may issue credentials for reviewers. A patient community may issue membership proofs without revealing diagnoses publicly. A marketplace may require an escrow credential before high-value contact. A political group may permit pseudonymous membership while limiting automated accounts. The same protocols support different constitutions.

The architecture also changes discovery. Instead of one global feed, people may discover communities through directories, trusted recommendations, search providers, or personal agents. Public content can remain indexable when authors choose. Private spaces can reveal only descriptions and admission procedures. The goal is not to hide the Internet but to stop treating universal exposure as the default condition of participation.

A transition could be gradual because the user experience can remain familiar. An application may still look like email, a messenger, a forum, or a collaborative workspace. Underneath, identity is portable, relationships are capability-based, sessions use open transports, recommendation is personal, and providers are replaceable. Compatibility gateways can connect legacy email and platform accounts while making their weaker trust model visible.

Metadata remains a serious risk. Even when content is encrypted, relays may observe timing, volume, counterparties, and device patterns. Pairwise identifiers and relay rotation reduce some correlation but increase complexity and may weaken abuse

response. Communities must decide which metadata they retain for safety, how long they retain it, and under what process it can be accessed. “Total privacy” should mean strong, inspectable minimization and user control, not the impossible claim that no metadata exists.

Availability creates another trade-off. Peer-to-peer systems fail when devices are offline, battery-constrained, or behind restrictive networks. Relays and mailboxes provide reliability, but popular providers may concentrate again. Portability tests, interoperable encryption formats, and multi-provider replication can keep availability from becoming sovereignty.

The transition path should begin with bounded communities and agent-to-agent workflows rather than an attempted replacement of the public Web. Professional networks, research collaborations, families, regulated services, and private marketplaces have clear relationship boundaries and high privacy value. Gateways can preserve email and web access while new connections use capabilities and portable identity. Successful patterns can then expand.

This proposal is not a finished protocol suite. It is an architectural direction built from mature components and unresolved research. WebRTC supplies interactive transport but not social semantics. DIDs and credentials supply verifiable control and claims but not trust. MLS supplies group key management but not governance. Personal AI supplies policy automation but introduces a powerful new attack surface. The missing achievement is their integration into an institutionally coherent system.

The original Internet succeeded because it separated applications from network ownership. The next Internet can succeed by separating communities and personal agents from platform ownership. Open protocols should carry data; constitutions should govern communities; personal agents should represent users; decentralized brands should hold shared promises; providers should compete to operate replaceable services.

The Internet will remain conflictual because human beings remain conflictual. It will contain commerce, ideology, manipulation, care, scholarship, entertainment, and fraud. A better architecture does not promise purity. It makes it harder for any one corporation, moderator group, automated swarm, or model provider to convert temporary advantage into irreversible control. That is a realistic standard of progress: not a network without power, but a network in which power must continually justify itself.

References

- Abbate, Janet. 1999. *Inventing the Internet*. Cambridge, MA: MIT Press.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. “Exposure to Opposing Views on Social Media Can Increase Political Polarization.” *Proceedings of the National Academy of Sciences* 115 (37): 9216–9221. <https://doi.org/10.1073/pnas.1804840115>.
- Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780190923624.001.0001>.
- Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. “Emotion Shapes the Diffusion of Moralized Content in Social Networks.” *Proceedings of the National Academy of Sciences* 114 (28): 7313–7318. <https://doi.org/10.1073/pnas.1618923114>.
- Buckels, Erin E., Paul D. Trapnell, and Delroy L. Paulhus. 2014. “Trolls Just Want to Have Fun.” *Personality and Individual Differences* 67: 97–102. <https://doi.org/10.1016/j.paid.2014.01.016>.
- Burke, Robin, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. “Balanced Neighborhoods for Multi-Sided Fairness in Recommendation.” In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 202–214. PMLR 81.
- C2PA. 2026. *C2PA Technical Specification, Version 2.4*. Coalition for Content Provenance and Authenticity. <https://c2pa.org/specifications/>.
- Celeste, Edoardo. 2019. “Digital Constitutionalism: A New Systematic Theorisation.” *International Review of Law, Computers & Technology* 33 (1): 76–99. <https://doi.org/10.1080/13600869.2019.1562604>.
- CERN. 1993. “Statement Concerning CERN W3 Software Release into Public Domain.” 30 April 1993. <https://cds.cern.ch/record/1164399>.
- Chan, Man-pui Sally, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. 2017. “Debunking: A Meta-Analysis of the Psychological Efficacy

- of Messages Countering Misinformation.” *Psychological Science* 28 (11): 1531–1546. <https://doi.org/10.1177/0956797617714579>.
- Chandrasekharan, Eshwar, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. “You Can’t Stay Here: The Efficacy of Reddit’s 2015 Ban Examined through Hate Speech.” *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW), Article 31. <https://doi.org/10.1145/3134666>.
- Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. “Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions.” In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1217–1230. <https://doi.org/10.1145/2998181.2998213>.
- Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. “The Echo Chamber Effect on Social Media.” *Proceedings of the National Academy of Sciences* 118 (9): e2023301118. <https://doi.org/10.1073/pnas.2023301118>.
- Clark, David D. 1988. “The Design Philosophy of the DARPA Internet Protocols.” *ACM SIGCOMM Computer Communication Review* 18 (4): 106–114. <https://doi.org/10.1145/52324.52336>.
- Cramer, Henriette, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wieringa. 2008. “The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender.” *User Modeling and User-Adapted Interaction* 18: 455–496. <https://doi.org/10.1007/s11257-008-9051-3>.
- Dorkenwald, Sven, et al. 2022. “FlyWire: Online Community for Whole-Brain Connectomics.” *Nature Methods* 19: 119–128. <https://doi.org/10.1038/s41592-021-01330-0>.
- European Union. 2022. “Regulation (EU) 2022/2065 on a Single Market for Digital Services.” *Official Journal of the European Union* L 277: 1–102.
- Giles, Jim. 2005. “Internet Encyclopaedias Go Head to Head.” *Nature* 438: 900–901. <https://doi.org/10.1038/438900a>.
- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

- Grinberg, Nir; Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter during the 2016 U.S. Presidential Election." *Science* 363 (6425): 374–378. <https://doi.org/10.1126/science.aau2706>.
- Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 US Election." *Nature Human Behaviour* 4: 472–480. <https://doi.org/10.1038/s41562-020-0833-x>.
- Halfaker, Aaron, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. "The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline." *American Behavioral Scientist* 57 (5): 664–688. <https://doi.org/10.1177/0002764212469365>.
- Helberger, Natali, Kari Karppinen, and Lucia D'Acunto. 2018. "Exposure Diversity as a Design Principle for Recommender Systems." *Information, Communication & Society* 21 (2): 191–207. <https://doi.org/10.1080/1369118X.2016.1271900>.
- Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1 (2). <https://doi.org/10.1177/2056305115603080>.
- Herfort, Benjamin, Sven Lautenbach, João Porto de Albuquerque, Jeff Anderson, and Alexander Zipf. 2021. "The Evolution of Humanitarian Mapping within the OpenStreetMap Community." *Scientific Reports* 11: 3037. <https://doi.org/10.1038/s41598-021-82404-z>.
- Hoffmann, Manuel, Frank Nagle, and Yanuo Zhou. 2024. "The Value of Open Source Software." *Harvard Business School Working Paper* 24-038. <https://doi.org/10.2139/ssrn.4693148>.
- IETF. 2018. *Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator Traversal*. RFC 8445. <https://doi.org/10.17487/RFC8445>.
- IETF. 2021a. *Overview: Real-Time Protocols for Browser-Based Applications*. RFC 8825. <https://doi.org/10.17487/RFC8825>.
- IETF. 2021b. *WebRTC Data Channels*. RFC 8831. <https://doi.org/10.17487/RFC8831>.
- IETF. 2023. *The Messaging Layer Security (MLS) Protocol*. RFC 9420. <https://doi.org/10.17487/RFC9420>.

- IETF. 2025. *The Messaging Layer Security (MLS) Architecture*. RFC 9750. <https://doi.org/10.17487/RFC9750>.
- Introna, Lucas D., and Helen Nissenbaum. 2000. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16 (3): 169–185. <https://doi.org/10.1080/01972240050133634>.
- Jhaveri, Shagun, Amy Bruckman, and Eric Gilbert. 2019. "Does Transparency in Moderation Really Matter? User Behavior after Content Removal Explanations on Reddit." *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW), Article 150. <https://doi.org/10.1145/3359252>.
- Kalla, Joshua L., and David E. Broockman. 2018. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 112 (1): 148–166. <https://doi.org/10.1017/S0003055417000363>.
- Kanich, Chris, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. 2009. "Spamalytics: An Empirical Analysis of Spam Marketing Conversion." *Communications of the ACM* 52 (9): 99–107. <https://doi.org/10.1145/1562164.1562190>.
- Khezresmaeilzadeh, Tina, Jiang Zhang, Dimitrios Andreadis, and Konstantinos Psounis. 2025. "Preserving Privacy and Utility in LLM-Based Product Recommendations." *arXiv:2505.00951*.
- Knijnenburg, Bart P., Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. "Explaining the User Experience of Recommender Systems." *User Modeling and User-Adapted Interaction* 22: 441–504. <https://doi.org/10.1007/s11257-011-9118-4>.
- Konstan, Joseph A., and John Riedl. 2012. "Recommender Systems: From Algorithms to User Experience." *User Modeling and User-Adapted Interaction* 22 (1–2): 101–123. <https://doi.org/10.1007/s11257-011-9112-x>.
- Leiner, Barry M., Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff. 1997. "A Brief History of the Internet." *Internet Society*, subsequently updated.
- Liu, Jiahao, Mingzhe Han, Guanming Liu, Weihang Wang, Dongsheng Li, Hansu Gu, Peng Zhang, Tun Lu, and Ning Gu. 2026. "From Hidden Profiles to Governable Personalization: Recommender Systems in the Age of LLM Agents." *arXiv:2604.20065*.

- Lorenz-Spreen, Philipp, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. 2019. "Accelerating Dynamics of Collective Attention." *Nature Communications* 10: 1759. <https://doi.org/10.1038/s41467-019-09311-w>.
- Matias, J. Nathan. 2019. "The Civic Labor of Volunteer Moderators Online." *Social Media + Society* 5 (2). <https://doi.org/10.1177/2056305119836778>.
- Matz, Sandra C., Michal Kosinski, Gideon Nave, and David J. Stillwell. 2017. "Psychological Targeting as an Effective Approach to Digital Mass Persuasion." *Proceedings of the National Academy of Sciences* 114 (48): 12714–12719. <https://doi.org/10.1073/pnas.1710966114>.
- O'Mahony, Siobhán, and Fabrizio Ferraro. 2007. "The Emergence of Governance in an Open Source Community." *Academy of Management Journal* 50 (5): 1079–1106. <https://doi.org/10.5465/amj.2007.27169153>.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>.
- Rathje, Steve, Jay J. Van Bavel, and Sander van der Linden. 2021. "Out-Group Animosity Drives Engagement on Social Media." *Proceedings of the National Academy of Sciences* 118 (26): e2024292118. <https://doi.org/10.1073/pnas.2024292118>.
- Saltzer, Jerome H., David P. Reed, and David D. Clark. 1984. "End-to-End Arguments in System Design." *ACM Transactions on Computer Systems* 2 (4): 277–288. <https://doi.org/10.1145/357401.357402>.
- Schneider, Nathan. 2024. *Governable Spaces: Democratic Design for Online Life*. Oakland: University of California Press. <https://doi.org/10.1525/luminos.181>.
- Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9: 4787. <https://doi.org/10.1038/s41467-018-06930-7>.
- Shaw, Aaron, and Benjamin Mako Hill. 2014. "Laboratories of Oligarchy? How the Iron Law Extends to Peer Production." *Journal of Communication* 64 (2): 215–238. <https://doi.org/10.1111/jcom.12082>.
- Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. "AI Models Collapse When Trained on

- Recursively Generated Data.” Nature* 631: 755–759.
<https://doi.org/10.1038/s41586-024-07566-y>.
- Suler, John. 2004. “The Online Disinhibition Effect.” *CyberPsychology & Behavior* 7 (3): 321–326. <https://doi.org/10.1089/1094931041291295>.
- Suzor, Nicolas P. 2019. *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781108666428>.
- Tandoc, Edson C., Jr., Zheng Wei Lim, and Richard Ling. 2018. “Defining ‘Fake News’: A Typology of Scholarly Definitions.” *Digital Journalism* 6 (2): 137–153. <https://doi.org/10.1080/21670811.2017.1360143>.
- Terranova, Tiziana. 2000. “Free Labor: Producing Culture for the Digital Economy.” *Social Text* 18 (2): 33–58.
https://doi.org/10.1215/01642472-18-2_63-33.
- Turner, Fred. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago: University of Chicago Press.
- United States Senate Select Committee on Intelligence. 2019. *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia’s Use of Social Media with Additional Views*. Washington, DC: United States Senate.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359 (6380): 1146–1151.
<https://doi.org/10.1126/science.aap9559>.
- W3C. 2018. ActivityPub. W3C Recommendation, 23 January 2018.
<https://www.w3.org/TR/activitypub/>.
- W3C. 2022. Decentralized Identifiers (DIDs) v1.0. W3C Recommendation, 19 July 2022. <https://www.w3.org/TR/did-core/>.
- W3C. 2025a. WebRTC: Real-Time Communication in Browsers. W3C Recommendation, 13 March 2025. <https://www.w3.org/TR/webrtc/>.
- W3C. 2025b. Verifiable Credentials Data Model v2.0. W3C Recommendation, 15 May 2025. <https://www.w3.org/TR/vc-data-model-2.0/>.
- W3C. 2026. Decentralized Identifiers (DIDs) v1.1. W3C Candidate Recommendation Snapshot, 5 March 2026. <https://www.w3.org/TR/did-1.1/>.

Wagner, Claudia, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. 2016. "Women through the Glass Ceiling: Gender Asymmetries in Wikipedia." *EPJ Data Science* 5: 5. <https://doi.org/10.1140/epjds/s13688-016-0066-4>.

Walter, Nathan, and Sheila T. Murphy. 2018. "How to Unring the Bell: A Meta-Analytic Approach to Correction of Misinformation." *Communication Monographs* 85 (3): 423–441. <https://doi.org/10.1080/03637751.2018.1467564>.

Yun, Jinhyuk, Sang Hoon Lee, and Hawoong Jeong. 2019. "Early Onset of Structural Inequality in the Formation of Collaborative Knowledge in All Wikimedia Projects." *Nature Human Behaviour* 3 (2): 155–163. <https://doi.org/10.1038/s41562-018-0488-z>.