MAY 2025

BLUEPRINT ON Prosocial Tech Design Governance

LISA SCHIRCH

Council on Technology and Social Cohesion

KROC INSTITUTE FOR INTERNATIONAL PEACE STUDIES
KEOUGH SCHOOL OF GLOBAL AFFAIRS

Toda Peace Institute

Table of Contents

Executive Summary 3

The State of Tech Design Governance and Regulation in 2025 7

Section A: Advance Prosocial Tech Design 9

A1. Defining Prosocial Tech Design Tiers 13

A2. Tier 1: Minimum Platform Building Standards of Prosocial Tech 15

Tier 2: Low-Barrier UX Designs to Support Prosocial Norms 17

Tier 3: Prosocial Algorithms and Recommender Systems 19

Tier 4: Advanced Prosocial Designs for Social Cohesion 21

A3. Tier 5: Middleware to Support Data Sovereignty, Portability, Interoperability 23

Section B: Provide Foundational Governance for Platform Research 24

B1. Require Democratic Platform Oversight, Transparency, and Audits 25

B2. Develop a Data Standard for Prosocial Tech Metrics 26

B3. Enforce Safe Harbor Protections for Accredited Researchers 28

Section C: Shift Market Forces to Support Prosocial Design 29

C1. Enforce Competition, Antitrust, Anti-monopoly Laws 31

C2. Codify Product Liability Adverse Impacts 32

C3. Incentivize and Invest in Prosocial Tech 33

Conclusions and Caveats 35

Research Methodology and Workshops 37

About 38

Recommended Citation: *Schirch, Lisa. "Blueprint on Prosocial Tech Design Governance." Council on Technology and Social Cohesion with University of Notre Dame and Toda Peace Institute. May 2025.*

Special thanks *for contributions from Renee Black, Devika Malik, Lena Slachmuijlder, Ravi Iyer, Michele Giovanardi, Paul Heidebrecht, Laure X Cast, Julia Kamin, Helena Puig Laurrari, and Jonathan Stray at the Council on Tech and Social Cohesion.*

Executive Summary

This *Blueprint on Prosocial Tech Design Governance* lays out actionable recommendations for governments, civil society, researchers, and industry to design digital platforms that reduce harm and increase benefit to society.

The *Blueprint* responds to the **crisis in the scale and impact of digital platform harms**. Digital platforms are fueling a systemic crisis by amplifying misinformation, harming mental health, eroding privacy, promoting polarization, exploiting children, and concentrating unaccountable power through manipulative design.

Digital platforms are not neutral. Their design influences human behavior. Relatively few abusive users spread most harmful content online. Platform designs often enable and incentivize harmful content, fraud, bots and fake accounts. Most "Trust and Safety" efforts inside large tech platforms downplay their design choices and steer regulators to focus on content moderation or "downstream" removal of harmful content.

Building on eight years of research, the Council on Technology and Social Cohesion conducted 12 workshops between 2023-2025 to conduct a systems analysis of the root causes of harmful content online with over 450 civil society experts. This Blueprint emerges from these collective insights and the foundational work of the Council's founding members including the Prosocial Design Network, Build Up, New_Public, Integrity Institute, Center for Humane Technology, Toda Peace Institute, University of Notre Dame, Search for Common Ground, Exygy, the Alliance for Peacebuilding, the Center for Human-Compatible AI, GoodBot, and the University of Southern California's Neely Center.

A systems approach to digital harms focuses on root causes rather than symptoms. Harms for digital platforms are not accidental but stem from deliberate design choices that prioritize profit at the expense of individual and societal well-being. The Integrity Institute's formulation of the problem illustrated below emphasizes the need to focus on how platform designs incentivize harmful content and behavior in online communities.[1]

Digital harms largely arise from **platform design** which can be **changed and regulated.**

Prosocial tech design governance is a framework for regulating digital platforms based on how their design choices—such as algorithms and interfaces—impact society. It shifts focus "upstream" to address the root causes of digital harms and the structural incentives influencing platform design.

A Systems Approach to Prosocial Tech Design

Digital harms stem from design. Platform design decisions are policy choices with social consequences.

Monetization, data extraction, recommendation systems, metrics and UX design are deeply interconnected and form a core system that

What the design **allows** you to do

What the design **prevents** you from doing

What the design **persuades** or **incentivizes** you to do

What the design **amplifies** and **highlights**

What the design **manipulates** or **deceives** you to do

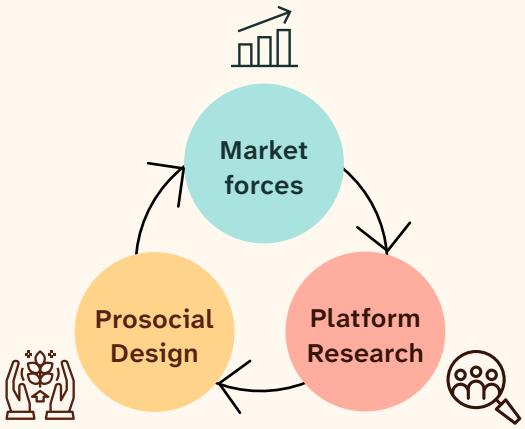
1 Integrity Institute. "*Focus on Features: Prevent Harm Through Design*". 2025. https://features.integrityinstitute.org/.

causes many digital harms.[2] Many platforms monetize through ad-based revenue, requiring maximizing user attention through **UX design** optimized to hook users, such as infinite scroll, and social feedback features like the Like button and emojis. Algorithms and **recommendation systems** predict and promote content most likely to keep users engaged to enable more **data extraction** that maximizes profit by targeting users with personalized content and ads. These design choices draw upon **metrics** for measuring success based on the number of daily users, time spent, and clicks instead of prosocial indicators.

Prosocial Tech Design Governance

Tech governance is a wider umbrella, including government regulation as well as a broader set of stakeholders who help to develop norms, processes, and institutions for thriving prosocial tech innovation.

Prosocial tech design governance reduces digital harms by 1) advancing prosocial design, 2) protecting independent platform research, and 3) shifting market forces to enable prosocial innovation. The Blueprint has three sections with three recommendations in each section.



Prosocial tech design governance advances understanding of prosocial tech design, improves independent research on platform impacts on society, and shifts market forces to enable prosocial innovation.

Policy Recommendations

SECTION A

Advance Prosocial Tech Design

Prosocial design aims to reduce digital harms and enable individuals to communicate with others in ways that uphold human dignity and enable public problem solving.

A1. Implement a Certification System for Prosocial Platform Tiers

Modeled after systems like LEED (for sustainable building), B Corp (for ethical business), and/or Environmental, Social, and Governance (ESG) standards, a tiered system creates a shared vocabulary and design benchmarks to align designers, investors, regulators, and civil society around responsible technology. It offers professional recognition and incentives—reducing legal and reputational risks, increasing user trust and retention, and linking compliance to

2 Black, Renee, Thinking in Systems Toward Responsible Tech Futures: How Designs, Algorithms & Business Models Impact Society and Democracy. GoodBot Society. Oct 2024. www.goodbot.ca/systems-thinking

public funding opportunities.

TIER 1: **Minimum Building Standards**

TIER 2: **Low Barrier Prosocial Designs**

TIER 3: **Prosocial Algorithms & Recommendation Systems**

TIER 4: **Advanced Prosocial Platforms for Social Cohesion**

TIER 5: **Middleware to Support Data Sovereignty, Portability, Interoperability**

A2. Require Minimum Tech Design Building Codes

Minimum building standards ensure that all companies adhere to the same rules, preventing unfair competition and protecting consumers from unsafe platforms. Minimum Tech Design Building Standards would reduce compliance risks and emphasize proactive safety measures, data protection, and user agency. The Neely Design Code offers a starting point for minimum design standards emphasizing user agency and control over content.

A3. Support Middleware for Data Sovereignty, Portability, and Interoperability Protocols

To ensure equitable, user-centered digital ecosystems, policymakers should enact legislation that requires dominant platforms to support certified middleware providers through open APIs, standardized data formats, and real-time data access. Middleware solutions—third-party services that mediate the user-platform relationship—advance data sovereignty, enable portability, and enforce interoperability across dominant digital platforms.

SECTION B

Provide Foundational Governance[3] for Platform Research

Research reveals how platforms are impacting individuals and society. Without research, governments and civil society are less able to hold platforms to account or reward them for their social impacts.

B1. Mandate Democratic Platform Oversight, Transparency, and Audits

Digital platforms operate in a black box, with complex algorithms and vast user data. Their decision-making on content moderation, ad targeting, and recommendation systems often remains hidden. Without transparency, verifying whether platforms uphold ethical standards, protect user rights, or comply with the laws is impossible. Requiring transparency would allow independent auditors and academic researchers to understand how technology design choices impact society.

B2. Develop a Data Standard for Prosocial Tech Metrics

3 Black, Renee "Foundational Principles for Advancing Prosocial Design Governance," GoodBot, May 2025. https://www.goodbot.ca/tech-policy/foundational-principles

Prosocial metrics are vital for reducing digital harm because they redefine what counts as “success” in the digital ecosystem. Current engagement metrics reward proxy metrics such as user base and revenue generation. Prosocial metrics act as alternative price signals, guiding innovation and investment in technologies that serve the public good rather than exploiting user vulnerabilities.

B3. Enforce Safe Harbor Protections for Accredited Researchers

Safe harbor protections create a legal shield for research in the public interest, enabling independent experts to investigate the societal impacts of digital platforms without facing legal threats, account suspension, or data barriers. Researchers report that lawsuit risk for ‘un-permissioned research’ significantly hinders independent assessment of platform behavior.

SECTION C

Shift Market Forces to Support Prosocial Design

Market forces incentivize design choices and metrics for measuring success. Shifting market forces is necessary to reduce digital harms and support prosocial design.

C1. Enforce Competition, Antitrust, & Anti-Monopoly Laws to Boost Prosocial Innovation

Enforcing competition, antitrust, and anti-monopoly laws is essential to reducing digital harms and advancing prosocial innovation. When a handful of tech giants dominate infrastructure, data, and user attention, smaller platforms with ethical, inclusive, or democratic designs struggle to gain visibility or viability. Governments can open space for prosocial tech alternatives to emerge and thrive by leveling the playing field. A more competitive digital environment empowers users with real choices and drives innovation that aligns with public interest rather than monopolistic control.

C2. Codify Product Liability for Adverse Impacts

Codifying product liability holds tech companies accountable for systemic risks like polarization, privacy violations, and disinformation. Venture capital often ignores these impacts, as traditional financial metrics exclude social impacts. Prosocial metrics would enable a price signal for these costs. By requiring large platforms to internalize these negative externalities through mechanisms like insurance, taxes, or design-based liability, regulators can shift incentives away from antisocial platform designs.

C3. Incentivize & Invest in Advanced Prosocial Technology Markets

Prosocial tech startups need guidance on sustainable monetization models and business advice. Public and private funds are crucial for supporting and scaling prosocial technologies across different prosocial tiers. Some use crowdfunding and tech cooperative ownership models to support prosocial innovation. Some prosocial tech companies can utilize monetization models, such as subscription-based and pay-per-use, to remove ad-based incentives for divisive content. Others look to token sales, XPRIZE competitions, and retroactive funding to provide innovative ways to fund prosocial tech. Governments can offer tax incentives, grants, and public funding to support prosocial tech. Private funding mechanisms include angel investors and philanthropic donations. Transparent government procurement can also help small, prosocial companies develop new technologies.[4]

4 This framework emerged from a May 1, 2024 workshop with VC and private funders for this Blueprint facilitated by Laure X Cast from the Integrity Institute, cosponsored with Lisa Schirch from Council on Tech and Social Cohesion.

The State of Tech Design Governance and Regulation in 2025

A substantial body of literature has focused on **government regulations** on digital platforms to address societal harms. In *Digital Empires: The Global Battle to Regulate Technology*, Anu Bradford outlines three major paradigms of tech governance: the US market-driven model prioritizing innovation and corporate freedom, the Chinese state-driven model focused on control and surveillance, and the EU rights-driven model emphasizing content moderation, privacy, and human dignity.[5] In *The Tech Coup: How to Save Democracy from Silicon Valley*, Marietje Schaake argues that unregulated tech companies have amassed unprecedented power, undermining democratic institutions. She critiques claims that regulation stifles innovation, arguing that tech monopolies are the real threat to innovation.[6]

In her review of tech regulation in the Global South, Devika Malik found some countries like India, Brazil, Nigeria, and Sri Lanka are advancing legal frameworks on antitrust, data protection, and interoperability policies to foster digital innovation and reduce dependency on monopolistic platforms. Some of these countries are also advancing state control on platforms to suppress dissent. Digital authoritarians can weaponize even well-intentioned tech regulations to curtail fundamental freedoms, enabling censorship and surveillance.

Tech regulation with legal rules and government enforcement is important but not enough. In "Fix the Internet, Not Big Tech," Cory Doctorow warns against focusing too much on efforts to change large digital platforms. Instead he advocates for systemic reforms to internet infrastructure.[7] These system changes are necessary for prosocial tech designs to thrive.

Tech governance is a wider umbrella, including government regulation as well as a broader set of stakeholders who help to develop norms, processes, and institutions for thriving prosocial tech innovation. Private foundations, universities and civil society groups are innovating norms, metrics, monetization pathways, and novel protocols to build a prosocial internet, emphasizing human rights and ethics in tech governance. For example, the "Design from the Margins" framework centers marginalized communities.[8] Global civil society promotes stringent antitrust laws and corporate responsibility measures to curb monopolistic practices and encourage more equitable, prosocial digital environments necessary for democracy and human rights.[9]

This Blueprint argues we need to move beyond regulation toward **tech design governance** of digital platforms.

Examples of Prosocial Tech Design Policies

These examples show the growing emphasis on integrating prosocial designs directly into technology development to address issues like data privacy and user protection.

- **California's Consumer Privacy Act (CCPA)** requires companies to offer clear and accessible ways for users to opt out of data collection.

---

5 Bradford, Anu. *Digital Empires: The Global Battle to Regulate Technology.* New York: Oxford University Press, 2023.

6 Schaake, Marietje. *The Tech Coup: How to Save Democracy from Silicon Valley.* Princeton, NJ: Princeton University Press, 2024.

7 Doctorow, Cory. "Fix the Internet, Not Big Tech." *Knight Foundation*, June 12, 2023.

8 Rigot, Afsaneh. *Design From the Margins.* Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, May 13, 2022.

9 Devika Malik, *Mapping Tech Design Regulation in the Global South,* Report No. 216. Toda Peace Institute, March 27, 2025.

- **Minnesota Algorithmic Transparency Act** requires social media platforms with over 10,000 monthly users to disclose information about their algorithms, user limits, notifications, and product experiments.
- **Utah Digital Choice Act** requires platforms to enable data portability and mandates interoperability, ensuring users can transfer data and communicate across platforms using open protocols.
- **US Federal Trade Commission's Dark Patterns Regulations** combats manipulative user interface designs that trick users into making unintended decisions, such as sharing personal data. Section 5 of the FTC Act prohibits "unfair or deceptive acts or practices in or affecting commerce."
- **US Antitrust Cases** include Meta's trial with the US Federal Trade Commission (FTC), focusing on Instagram and WhatsApp acquisitions, allegedly eliminating competition in personal social networking. In April 2025, Google faced two antitrust rulings, with a judge ruling that Google monopolized digital advertising and Department of Justice (DOJ) advocating for the divestiture of Google's assets to restore competition.
- **EU's General Data Protection Regulation (GDPR) - Privacy by Design** mandates that organizations must integrate data protection into the design and development of their systems from the outset.
- **EU's Digital Services Act (DSA)** includes design-focused regulations that require online platforms to assess and mitigate risks such as illegal content, ensuring transparency in algorithms and improving the design of user interfaces to prevent harmful content from spreading.
- **Brazil's General Data Protection Law (LGPD)** includes provisions on algorithmic transparency, granting individuals the right to request clear and adequate information about automated decision-making processes that significantly affect them.
- **Kenya's Data Protection Act 2019 (DPA)** emphasizes data protection by design and by default. It lays the groundwork for data interoperability by granting individuals the right to data portability.
- **Nigeria's Data Protection Act (NDPA)** enhances user control and promotes interoperability through the right to data portability.
- **Sri Lanka's Personal Data Protection Act, No. 9 of 2022 (PDPA)** includes safeguarding personal data.
- **India's Consumer Protection (E-Commerce) Rules** mandates transparency in automated platform product ranking, requiring disclosure of algorithms to ensure fairness.
- **UK's Online Safety Bill** requires platforms to build in age-appropriate design features that detect and remove illegal or harmful content, and mandates greater transparency on content moderation and algorithms.

Proposed Laws

- **The Algorithmic Accountability Act (US)** would require companies to conduct impact assessments of their algorithms, reducing bias in AI systems and improving transparency in tech design.
- **Sri Lanka's National Digital Economy Strategy 2030** emphasizes the importance of interoperability to drive digital transformation.
- **The Digital India Act** proposes that platforms reveal algorithm functionality.
- **India's Digital Competition Bill** aims to foster fairness in the digital ecosystem by addressing anti-competitive practices, like service bundling, particularly on large digital platforms.
- **The US Platform Accountability and Transparency Act** (PATA) aims to enhance transparency by requiring large social media platforms to share data with independent researchers and the public.

SECTION A

Advance Prosocial Tech Design

This section lays out a comprehensive guide to prosocial tech design, contrasting it with anti-social tech design and demonstrating how platform architecture impacts society.

Key Recommendations Recap:

A1. Implement a certification system for prosocial platform tiers
A2. Require minimum tech design building codes
A3. Support middleware for data sovereignty, portability, and interoperability protocols

Platform Design Steers Human Behavior and Impacts Society

Digital harms—such as misinformation, addiction, online radicalization, privacy violations, and mental health crises—are not just the result of bad actors but are deeply embedded in the design of digital platforms themselves. Prosocial tech design responds to the widespread use of deceptive and manipulative tech designs. **No technology has a "neutral" design.** Tech designers make important choices that influence human behavior, including the following:

- What the design **allows** you to do
- What the design **prevents** you from doing
- What the design **persuades** or **incentivizes** you to do
- What the design **amplifies** and **highlights**
- What the design **manipulates** or **deceives** you to do

Prosocial Design Supports Social Cohesion

Prosocial design in technology aims to cultivate healthy online interactions, safety, well-being, and dignity by integrating prosocial principles into digital platforms. Social cohesion is a widely used term for the glue that holds society together. This paper draws on Search for Common Ground's more detailed definition of social cohesion as 1) individual agency, horizontal social capital within and between groups, and vertical social capital or public trust between the public and institutions.[10] Prosocial platform designs can contribute to three dimensions of social cohesion: individual agency, social trust between groups, and public trust between institutions and society.[11]

10 Institutional Learning Team, "Building Social Cohesion in the Midst of Conflict: Identifying Challenges, Measuring Progress, and Maximizing Results," Search for Common Ground. November 2020.

11 Schirch, Lisa. "The Case for Designing Tech for Social Cohesion: The Limits of Content Moderation and Tech Regulation." *Yale Journal of Law & the Humanities* 34, 2023: 1–34.

Contrasting Prosocial and Antisocial Tech

Prosocial platforms are distinct from antisocial tech in at least five ways, illustrated in the table below. Platform design choices contribute significantly to personal, community, and societal harm. Antisocial tech is often driven by ad-based and data extraction-based monetization that expects rapid scaling and returns, as measured by daily active users, time spent, and clicks. To optimize user engagement, recommendation systems use algorithms that prioritize keeping users engaged with divisive and addictive content.

	Antisocial Tech	**Prosocial Tech**
Monetization	Operate as data brokers and advertising platforms	Operate as public services designed for the common good
Metrics	Use indicators such as daily active users, clicks, time on platform	Use indicators such as user-reported satisfaction and intergroup communication
Data	Minimum privacy, maximizing data collection	Maximize privacy, minimize data collection; Privacy by Design
Recommendation Systems	Amplify highly emotional content, incentivizing outrage and divisive content	Amplify trusted sources, incentivizing areas of agreement between groups
UX Design	Reduces human agency with manipulative design tricks to maximize user engagement; unintentionally undermines social cohesion	Maximizes human agency and control over user experience; designed to enhance social cohesion

Antisocial and Deceptive Tech Design

Antisocial tech designs can amplify highly emotional and divisive content and incentivize antisocial behaviors, including violating social norms, breaking rules, deception, theft, verbal attacks, threats of physical attacks, and harm to self or others. Examples include:

- **Infinite Scrolling** to keep people engaged as long as possible
- **Algorithmic Extremism** amplifies content that arouses strong emotions or false information
- **Confusion** from mixing ads and personal content with news, making it hard to distinguish journalism from propaganda.

Social media platforms often cue users to promote themselves through design patterns that mimic performance metrics—like hearts, likes, and comments—which position peer feedback as deeply personal. For example, the profile card advertises a user's identity and status. Voting systems like the Like button or upvote are designs that offer social validation and comparison. These interactions transform identity construction into a public spectacle, where self-worth is tied to visible approval. This environment fosters heightened psychological vulnerability, as users become conditioned to seek validation through curated personas and reactive content. In this system, attention itself becomes the reward—regardless of whether it is positive or negative.

As a result, users may increasingly engage in provocative, exaggerated, or emotionally charged behavior, since controversial or extreme posts often generate more feedback than prosocial or compassionate content. This dynamic not only distorts self-perception but also contributes to a culture where negative attention can be more gratifying—and socially rewarded—than genuine human connection.[12]

Deceptive tech designs, also called "dark patterns," manipulate or mislead people to behaviors in the interest of tech companies rather than those using the platform. These designs draw on psychological insights and cognitive biases to encourage certain behaviors.[13] The US Federal Trade Commission's 2022 report, *Bringing Dark Patterns to Light*, analyzes how manipulative design practices deceive consumers, undermine informed choice, and violate consumer protection laws.[14] The Advertising Standards Council of India (ASCI) published guidelines identifying deceptive design patterns such as drip pricing, bait-and-switch tactics, false urgency, and disguised advertisements as misleading and detrimental to consumer trust.[15]

- **Hiding Cancelation Options**
- **Default settings** on a new account are set for minimal security or privacy, helping the company maximize data collection
- **Complexify** privacy policies to discourage user privacy
- **Gamify** platforms to gather data. Facebook's "Ten Year Challenge" had users post photos from 10 years ago and today, likely training facial recognition software.[16]
- **Hidden Costs** of luring users to watch videos with the intent of deception

CASE STUDY OF ANTISOCIAL DESIGN:

Facebook's Algorithms of Trauma

A case study examined by the Panoptykon Foundation demonstrates how Facebook uses algorithms to deliver personalized ads that may exploit users' mental vulnerabilities. The study found Facebook's algorithms expose users to disturbing, trauma-related content despite their efforts to opt out, revealing that the platform does not give users meaningful control over what they see. The study profiles Joanna, a mother with health anxiety, whose Facebook feed showed health ads about cancer and disorders. When she selected "See fewer parenting ads," similar content soon returned.

Dorota Głowacka, Karolina Iwańska. *Algorithms of Trauma: New Case Study Shows That Facebook Doesn't Give Users Real Control over Disturbing Surveillance Ads.* Panoptykon Foundation. September 28, 2021.

Map of Digital Harms

Digital platforms affect individuals and society in various ways, in market, social, and political exchanges, as illustrated in the table below, categorizing digital harms. An externality is an indirect cost outside the intended goal. In the automobile

12 Cular, Circ. "9 Design Interactions Responsible for the Social Media Mental Health Crisis." *UX Magazine*, October 12, 2023.

13 Brignull, Harry. *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You.* UK: Testimonium Ltd, 2023.

14 Federal Trade Commission. *Bringing Dark Patterns to Light.* Washington, D.C.: Federal Trade Commission, September 2022.

15 Advertising Standards Council of India. *Conscious Patterns: Study of Deceptive Patterns in Top Indian Apps.* Mumbai: ASCI Academy, 2024.

16 O'Neill, Kate. "Facebook's '10 Year Challenge' Is Just a Harmless Meme—Right?" *WIRED,* January 15, 2019.

industry, externalities include accident deaths and carbon emissions. Externalities often lack specific pricing.[17] Toxic polarization online is one serious example of a *negative externality* of digital platforms.[18] Solving collective challenges like the climate crisis, healthcare, or migration is more difficult when a society faces toxic polarization.

Market Harms	**Social Harms**	**Political Harms**
Fraud and scams	Defamation	Polarizing political rhetoric
Human trafficking	Bullying and harassment	Political surveillance
Intellectual property theft	Enabling self-harm	Toxic speech
Sale of illegal, counterfeit, or regulated goods and services	Child Sexual Abuse Material	Calls, threats, or instructions to enable violence
Dynamic pricing charging different prices for the same thing to different audiences	Sexual harassment and exploitation	Dangerous misinformation and disinformation
Monopolies reduce consumer protections and choice	Graphic, Violent, or Sexual Content	Violent extremist recruitment
Algorithmic bias against marginalized groups	Algorithms that amplify and discriminate	Election interference
Selling personal data and exposing it to hacking	Toxic polarization	Cognitive warfare to confuse and divide societies

17 Puig Larrauri, Helena. *Societal Divides as a Taxable Negative Externality of Digital Platforms: An Exploration of the Rationale for Regulating Algorithmically Mediated Platforms Differently.* Ashoka Tech & Humanity, 2023.

18 See Barrett, Paul, Justin Hendrix, and Grant Sims. 2021. "How Tech Platforms Fuel U.S. Political Polarization—and What Government Can Do About It." *Brookings Institution.* September 27, 2021. Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "How Social Media Shapes Polarization." *Trends in Cognitive Sciences* 25, no. 11 (2021): 913–16; González-Bailón, Sandra, and Yphtach Lelkes. "Do Social Media Undermine Social Cohesion? A Critical Review." *Social Issues and Policy Review* 17, (1) 2023: 1–26.

A1. Defining Prosocial Tech Design Tiers

Without codes, builders might cut corners, causing dangerous structures. Building codes establish minimum safety standards for architects, engineers, and builders. They ensure standards across regions, making construction predictable and fair. Building codes protect lives, property, and public trust by embedding safety into construction.

Just as cities enforce building codes for housing to ensure public safety and urban planning, investors, regulators, and companies could also develop tech platform building codes to reduce harm and contribute to public goods. Architects and civil engineers rely on tiered building codes based on safety and performance requirements. Mandatory codes like the International Building Code (IBC) set minimum structural safety and health standards. Basic homes must meet minimum requirements, while hospitals and skyscrapers need stricter codes for earthquake resilience. Green Building Standards, like LEED or the WELL Building Standard, offer optional tiers focusing on energy efficiency, carbon reduction, and resilience, promoting environmental stewardship and social well-being. While some governments now incentivize these higher standards, certification remains largely voluntary.

A tiered system for prosocial tech design could facilitate the gradual and scalable integration of ethical principles into digital platform designs. Tiers might signal different entry points. with some adopting Tiers 1 and 5, for example, but not other tiers. Lower tiers offer easier and smaller design shifts, where higher level tiers require more fundamental shifts to monetization and design. This is a beginning prototype of what a tier system could look like. A tier system could define tiers differently than this report.

Tier 1 addresses UX design drivers of digital harm. Tier 2 offers prosocial UX designs that are “low barrier,” meaning they do not interrupt ad-based monetization. Tier 3, 4, and 5 offer prosocial designs that address more fundamental system drivers of harm. Tier 3 addresses algorithmic recommendation systems. Tier 4 offers types of platforms designed explicitly for public decisionmaking and intergroup understanding. Tier 5 offers more fundamental system solutions through middleware protocols.

Tier 1: Minimum Building Standards offers a baseline of widely agreed upon, research-based prosocial tech designs to prevent online harms. It includes Safety by Design and Privacy by Design frameworks without altering the underlying profit model tied to ads and data collection.

Tier 2: Low Barrier Building Standards go beyond minimum standards of harm reduction to foster healthy

interactions between users with tags, buttons, nudges, boosts, and friction designs without altering the underlying profit model tied to ads and data collection.

Tier 3: Prosocial Algorithms and Recommender Systems offer a wide range of AI-driven algorithmic tools to guide what information platforms show and recommend. Tier 3 may reduce user-engagement markers used by ad-based marketing.

Tier 4: Advanced Prosocial Designs focuses on platforms facilitating respectful dialogue, community governance, mutual aid, informational integrity, and collective problem-solving. This model relies on public and private funding over ad-based monetization.

Tier 5: Middleware to Support Data Sovereignty, Portability, and Interoperability addresses the underlying tech stack to find solutions.

A Tier System Benefits Users, Tech Companies, Investors, and Governments

A Professional Certification System for Prosocial Technology Tiers could recognize platforms, products, and practitioners achieving distinct tiers of prosocial technology design. Modeled after systems like LEED (for sustainable building), B Corp (for ethical business), and/or Environmental, Social, and Governance (ESG) standards, this framework would require an independent, nonprofit entity—supported by public funding or industry fees—to administer audits, verify compliance, and issue designations.

A tier system could benefit UX designers, investors, and regulators with a roadmap for balancing profitability with social responsibility in the following ways.

- **Establishes a shared vocabulary** across the diverse stakeholders involved in digital platform development—designers, investors, regulators, and civil society actors, reducing ambiguity around what constitutes "responsible" or "prosocial" technology.
- **Defines specific, actionable design benchmarks** to foster more effective collaboration, regulation, and investment to translate abstract ethical commitments into real interface choices and algorithms.
- **Help consumers make decisions** on platform design standards for safety and privacy.
- **Offers professional recognition**[19] **for investors, companies, and UX designers**, creating a new ecosystem of recognition around ethical tech. A higher-tier designation may signal to regulators, insurers, and legal actors that the platform meets or exceeds industry standards for digital safety, reducing the likelihood or severity of regulatory penalties, civil suits, or public backlash.
- **Reduces risks and liabilities to brand safety, helping investors and companies** limit exposure to legal and reputational risks and evaluate long-term growth potential
- **Boosts user satisfaction and loyalty** with measurable improvements in retention, well-being, and meaningful participation, offering a compelling alternative to extractive engagement models.
- **Links to eligibility for public funding.** Regulators can use this framework to promote and reward socially beneficial technologies or incentivize ethical innovation such as with a tax relief system.

19 Black, Renee, "Imagining Professional ProSocial Design Governance" GoodBot Society. May 10 2025. https://www.goodbot.ca/tech-policy/prosocial-professionals

A2. TIER 1: Minimum Platform Building Standards of Prosocial Tech

Without codes, builders might cut corners, causing dangerous structures. Building codes establish minimum safety standards for architects, engineers, and builders. They ensure standards across regions, making construction predictable and fair. Building codes protect lives, property, and public trust by embedding safety into construction. Minimum tech design building standards

A Lack of Minimum Standards Hurts Ethical Design

1. **Uneven burdens:** Platforms that voluntarily invest in trust and safety or transparent moderation face higher costs, while exploitative competitors can grow rapidly by neglecting these responsibilities.
2. **Lack of market signals for trustworthiness:** Without standard metrics or disclosures, users and regulators struggle to compare platforms' ethical claims, letting deceptive or harmful actors blend in with responsible ones.
3. **Fragmentation and duplication:** Each ethical platform must build custom safeguards (e.g., for consent, safety alerts, or reporting mechanisms), which drains resources that could go toward meaningful innovation.
4. **Discourages innovation in governance:** In a deregulated environment, platforms often see responsible design as risky or costly. A building code would normalize baseline safety and open the door to creative competition above that baseline.

How a Building Code Helps Level the Playing Field

A basic building code would make it easier for ethical and prosocial platforms to compete by setting clear, enforceable baseline obligations for all players.

1. **Prevents a race to the bottom:** Without shared standards, dominant platforms can cut corners (e.g., maximizing engagement through addictive or harmful design) while smaller prosocial platforms absorb higher costs. Standards would raise the floor so all platforms must internalize certain social responsibilities.
2. **Reduces compliance uncertainty:** Clear standards lower the legal and operational risk for startups trying to build responsibly. Rather than guessing what is "ethical enough," they can focus on innovation within a known compliance framework.
3. **Supports interoperability and portability:** A building code could mandate open APIs, data portability, or standard labeling practices—empowering users to switch services and reducing the network effects that entrench dominant platforms.
4. **Encourages fair competition:** When foundational responsibilities (like child safety, data protection, or content labeling) are required of all platforms, large incumbents can't undercut values-driven competitors by externalizing societal harms.

Tier 1 offers a baseline of designs, policies, and processes that respect individual users' safety, privacy, and agency.

- The **Safety by Design** initiative, launched by Australia's eSafety Commissioner in 2018, integrates user safety into online products' core design. As defined by the Trust and Safety Professional Association, this proactive approach focuses on reducing harm through user controls and preventive technologies to improve cybersecurity

and prevent identity theft.[20] For example, requiring platforms to be transparent in ad placement and funding and offering tools for users to understand and control ads are part of Safety by Design.

- The **Privacy by Design** framework follows seven principles, including privacy as default in platform design. The GDPR requires organizations to implement data protection through measures like data minimization. Platforms must maintain transparency in monetizing data.[21]
- **User Agency by Design** ensures platforms empower users to indicate preferred content, prioritize quality over engagement, provide opt-out options for revenue features, and reject deceptive designs.
- **Procedural Justice by Design** involves transparent decision-making that empowers users to shape platform rules and ensure fair treatment and outcomes.

The Neely Design Code from the University of Southern California Marshall's Neely Center for Ethical Leadership and Decision-Making consulted experts and practitioners to identify the following nine evidence-based *minimum standards* for companies that host online social interactions.[22]

Neely Design Code

1. Allow users to easily and explicitly **indicate content they do or do not want**, and respect users' explicit preferences even if contradicted by users' engagement.
2. **Replace engagement optimizations** (e.g., view time, comments, shares, more ad distribution) with optimizations for user-perceived quality, especially on sensitive topics.
3. **Prioritize public amplification** to actors that varied users explicitly know and trust.
4. **Allow users to accessibly opt-out of revenue-maximizing design features** (e.g., optimizing for time spent, infinite scroll, auto-play) that encourage greater usage. Make this the default for minors. Offer all users tools to limit their platform usage.
5. **Provide transparent, sensible rate limits** for new, untrusted users who access functionality that can be used to target or influence others.
6. **Enable privacy by default** for any situation involving minors or when a significant number of users expect their information to be inaccessible.
7. **Prohibit the public distribution of sexually explicit content** depicting individuals who do not explicitly provide permission.
8. **Enable device-based parental controls** with appropriate defaults for minors for any online activity that a broad group of parents considers risky.
9. **Publicly provide product experimentation results** on outcomes of societal interest for any meaningful product design decision.

---

20 Trust & Safety Professional Association. "Implementing Safety by Design." In *Trust & Safety Curriculum*. https://www.tspa.org/curriculum/ts-curriculum/safety-by-design/implementing-safety-by-design/

21 Cavoukian, Ann. 2011. *Privacy by Design: The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario, Canada.

22 "Neely Design Code." Neely Center for Ethical Leadership and Decision Making. *USC Marshall School of Business*. Accessed May 7, 2025. https://www.marshall.usc.edu/institutes-and-centers/neely-center-for-ethical-leadership-and-decision-making.

TIER 2: Low-Barrier UX Designs to Support Prosocial Norms

Tier 2 design promotes healthy user interactions through low-barrier elements that benefit both users and tech companies. It uses design features to support prosocial norms, understanding cues, and friction for healthier engagement. Tier 2 designs are feasible for Big Tech and startups without disrupting profit models. The Prosocial Design Network and Integrity Institute offer libraries of low-barrier UX designs. A report on tech design to prevent Technology-Facilitated Gender-Based Violence (TFGBV) also reviews many Tier 2 platform designs.[23]

Tags and Signaling Systems

Tags and signaling systems on digital platforms enable users to communicate identity and trustworthiness, shaping social norms. eBay's reputation system allows buyers and sellers to leave ratings and reviews, serving as reputation capital that encourages cooperative behavior. Public feedback creates accountability, promoting civility and responsiveness.[24] Reddit's karma system reflects user contributions through upvotes. While karma rarely affects posting, it indicates credibility, with high-karma users viewed as more established. Some subreddits use karma thresholds to prevent spam.

Reaction Buttons

Reaction buttons like "like," "respect," and emoji responses can guide user behavior and foster empathy in online spaces when intentionally designed. These features shape social norms and promote healthier interactions. LinkedIn's "Curious," "Celebrate," "Insightful," and "Support" reactions enable users to appreciate different viewpoints and express solidarity. One study found participants were more inclined to engage with opposing political views when a "Respect" button was available, compared to "Like" or "Recommend" options. The term "Respect" may encourage users to acknowledge different viewpoints without agreeing, fostering civil online discourse.[25]

Prompts, Nudges, Boosts, Coaches, and Pop-Ups

Researchers document many ways of encouraging people to increase prosocial behavior. While designers can "nudge" or prompt people toward prosocial behavior, "boosting" and coaching is a more transparent approach that builds an individual's capacity and respects individual agency.

In the early 2000s, the online auction company eBay hired conflict resolution expert Colin Rule to design a dispute resolution system using scripts to prompt buyers and sellers in effective communication to solve their disputes. eBay. The automated system helped to solve millions of disputes.[26]

Go Vocal (formerly CitizenLab) conducted experiments on Reddit, nudging users to "Remember the human" before commenting. Research demonstrated these nudges improved civil discourse and community health.[27] Several

---

23 Slachmuijlder, Lena, and Sofia Bonilla. Prevention by Design: A Roadmap for Tackling TFGBV at the Source. Integrity Institute, Council on Technology & Social Cohesion, and Search for Common Ground, March 2025.

24 Tadelis, Steven. "Reputation and Feedback Systems in Online Platform Markets." *Annual Review of Economics* 8. 2016: 321–340.

25 Stroud, Natalie Jomini, Ashley Muddiman, and Josh Scacco. "Like, Recommend, or Respect? Altering Political Behavior in News Comment Sections." *New Media & Society* 19 (11). 2017: 478–497.

26 Rule, Colin. "Making Peace on eBay: Resolving Disputes in the World's Largest Marketplace." ACResolution Magazine, Fall 2008, 8–11.

27 Matias, J. Nathan. Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences*, 116(20), 2019: 9785–9789.

large-scale experiments verify that posting reminders of guidelines improves norm coherence.[28] Pennycook and Rand's meta-analysis of 20 experiments with 26,000 participants found that accuracy prompts significantly reduce misinformation sharing across various contexts.[29] Capraro and Celadin (2022) conducted online experiments where participants viewed true and false news items with share buttons. Adding "think if this news is accurate" to share buttons reduced participants' likelihood of sharing fake news by over 25% compared to controls without prompts.[30] Jesse and Jannach (2020) conducted a comprehensive survey identifying 87 digital nudging mechanisms within recommender systems. They propose integrating these mechanisms to influence user decision-making, especially to reflect on their content consumption patterns.[31]

Using Jigsaw's Perspective API to rate comment toxicity, LLMs can identify toxic social media conversations and prompt users with popups that prompt users to revise comments that receive high toxicity scores. OpenWeb's study of 50,000 users showed how real-time feedback reduces toxicity. Users receive warnings for offensive content and can edit or proceed with posting. 34% of warned users edited comments, with 54% making acceptable changes. This feedback increased civil comments by 12.5%.[32]

Friction Slow Downs

Friction deliberately adds barriers to prevent antisocial behaviors on platforms. While most designs aim to reduce friction for engagement, purposeful friction in ethical technology design can promote reflection and minimize harm. For example, WhatsApp implemented limits on message forwarding to reduce viral falsehoods. Platforms like Medium and Twitter tested prompts to "read the article before sharing" to boost content literacy.[33] A study by legal and technical experts analyzed WhatsApp's friction on message forwarding, finding it reduced the spread of false news without infringing on speech rights. They advocate for "design friction" as a concept that deserves more attention in regulatory frameworks, particularly for encrypted platforms.[34] However, friction slow-downs do not always have prosocial outcomes. Adding friction to news articles, for example, can slow the spread of important public information.

28 Horta Ribeiro, Manoel, Robert West, Ryan Lewis, and Sanjay Kairam. "Post Guidance for Online Communities." *Proceedings of the ACM on Human-Computer Interaction,* CSCW 2025; Kim, Jisu, Curtis McDonald, Paul Meosky, Matthew Katsaros, and Tom Tyler. "Promoting Online Civility Through Platform Architecture." *Journal of Online Trust and Safety*, 1(4), 2022.

29 Pennycook, Gordon, and David G. Rand. "Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation." *Nature Communications* 13, no. 1. 2022: 2333.

30 Capraro, Valerio, and Tatiana Celadin. "I Think This News Is Accurate: Endorsing Accuracy Decreases the Sharing of Fake News." *Personality and Social Psychology* Bulletin 49 (12). 2023: 1635–1645.

31 Jesse, Mathias, and Dietmar Jannach. "Digital Nudging with Recommender Systems: Survey and Future Directions." *Computers in Human Behavior Reports* 3. 2021: 100052.

32 OpenWeb. "Can Machines Change Human Behavior? OpenWeb, Using Jigsaw's Perspective API, Releases Case Study Measuring the Effects of Real-Time Feedback and 'Nudges' in Decreasing Toxicity." *OpenWeb*, September 23, 2020.

33 Hutchinson, Andrew. "Twitter Shares Insights Into the Effectiveness of Its New Prompts to Encourage More Considerate Tweeting." *Social Media Today*, September 24, 2020.

34 Gordon-Tapiero, Ayelet, Paul Ohm, and Ashwin Ramaswami. "Fact and Friction: A Case Study in the Fight Against False News." *UC Davis Law Review* 57 (1) 2023: 171–242.

TIER 3: Prosocial Algorithms and Recommender Systems

Algorithms and recommender systems decide what information platforms show and recommend. These AI-driven systems significantly impact society and should reflect democratic values and public interest.[35] In *Invisible Rulers*, Renée DiResta examines how algorithms, influencer networks, and crowdsourcing dynamics interact to shape public perception, turning lies into viral realities that distort democratic discourse.[36]

Prosocial algorithms and recommender systems could reorient social media to enhance the social fabric by helping people experience a wider diversity of viewpoints and discover more context.[37] There are a wide range of prosocial algorithms and recommender systems.

Prosocial Content

Civility and Toxicity Reduction Algorithms aim to improve the quality of dialogue by promoting respectful and constructive speech. Toxicity Filters like Detoxify and Perspective API downgrade hateful content. Civility Ranking algorithms rank content by politeness, empathy, or reasoning.[38]

Bridging Algorithms rank areas of common ground between diverse groups higher to surface agreement.[39] The platform Pol.is, for example, collects open-ended input and uses machine learning to group people by shared opinions. It highlights areas of consensus and constructive disagreement, enabling inclusive democratic decision-making.[40]

Fairness Algorithms identify bias in machine learning models, preventing discrimination based on protected attributes. These tools are vital in hiring, credit scoring, healthcare, and content recommendation. For example, IBM's AI Fairness 360 provides metrics and algorithms to assess discrimination and support adjustments, enabling transparent AI development.

Early Warning Algorithms predict potential incitement to violence in digital spaces using pattern recognition and natural language processing. The University of Pennsylvania's Machine Learning for Peace, for example, tracks political rhetoric and online harms to forecast instability by analyzing vast amounts of data and identifying patterns and trends that may indicate rising tensions or the spread of harmful content.

Content Diversity

Exposure Diversity Algorithms aim to reduce echo chambers and promote cognitive empathy by diversifying what users see. The "Designing for Serendipity" concept incorporates randomness into digital recommendations, enabling

35 Stray, Jonathan. "Designing Recommender Systems to Depolarize." *First Monday* 27 (5). 2022.

36 DiResta, Renée. *Invisible Rulers: The People Who Turn Lies into Reality. New York: PublicAffairs, 2024*

37 Weyl, E. Glen, Luke Thorburn, Emillie de Keulenaar, Jacob Mchangama, Divya Siddarth, and Audrey Tang. "Prosocial Media." *arXiv preprint* arXiv:2502.10834 February, 2025.

38 Bao, Jiajun, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. "Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations." In *Proceedings of the Web Conference* 2021, 312–23. New York: Association for Computing Machinery, 2021.

39 Ovadya, Aviv, and Luke Thorburn. *Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance.* arXiv preprint, January 2023.

40 Small, Christopher T., Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. "Polis: Scaling Deliberation by Mapping High Dimensional Opinion Spaces." *Recerca. Revista de Pensament i Anàlisi* 26 (2). 2021: 1–26.

users to make unexpected yet valuable discoveries that spark creativity. Flipfeed (MIT Media Lab prototype) reorders Twitter feeds to increase ideological diversity by showing posts from users with opposing political views. It filters for civil, constructive content to gently introduce alternative perspectives without triggering defensive responses.[41] Some research finds content diversity can backfire. Any content diversity process should therefore be tested.[42]

Locality-aware algorithms prioritize content contextually relevant to a user's geographic location. These algorithms surface local journalism, community discussions, and content shared by mutual connections or neighbors. Civic tech tools tailor information to voter addresses, highlighting local races and organizing efforts overlooked by national apps.

Content Quality

Credibility Score Algorithms boost sources rated for accuracy and transparency. X's Community Notes enables users to label misleading tweets, affecting visibility. When tweets are flagged with "helpful" notes, the algorithm adjusts their visibility based on trust scores. Tweets can be labeled "misleading" and deboosted.[43]

Rational Discourse Optimizers use algorithms trained to detect argument structure, not emotional salience or popularity. The ArgumenText project detects arguments in text, scores their logical structure, and identifies evidence and counterarguments.[44] Kialo uses algorithms to organize discussions hierarchically, creating tree-like structures of pro and con arguments.

Healthy Content and Consumption

Digital Nutrition Scoring Algorithms optimize for user health and mindfulness by recommending content that improves psychological health, empathy, and civic knowledge—like a "healthy diet for the mind." The Center for Humane Technology's Digital Nutrition Project shifts tech design toward "well-being by design" and "time well spent."

Time-aware Nudging Algorithms detect prolonged or repetitive use patterns and prompt users to take breaks or engage in alternative activities. For example, YouTube and TikTok have introduced reminders like "Take a break" or "You've been scrolling for a while" to help reduce compulsive behavior and support digital well-being.

Mental Health Algorithms detect emotional distress signals through search queries, posting patterns, and behaviors, then surface supportive content. Platforms like Instagram and TikTok offer crisis resources when users search terms related to self-harm or depression. Systems prioritize uplifting content during vulnerable periods to create a safer digital environment.

---

41 Anand, Arvind Narayanan, and Neha Narula. *FlipFeed: A Platform for Exploring the Effects of Algorithmic Oppositional Exposure on Social Media*. MIT Media Lab, 2018.

42 Prosocial Design Network. Digital Intervention Library. 2025.. https://doi.org/10.17605/OSF.IO/Q4RMB

43 Wojcik, Stefan, et al. 2022. "Birdwatch: Crowd Wisdom and Bridging Algorithms Can Inform Understanding and Reduce the Spread of Misinformation." *arXiv preprint* arXiv:2210.15723. https://doi.org/10.48550/arXiv.2210.15723.

44 Daxenberger, Johannes, *et al*. ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario. *Datenbank Spektrum* 20, 115–121. 2020.

TIER 4: Advanced Prosocial Designs for Social Cohesion

Advanced levels of prosocial platform design aim to build social cohesion, which is defined here as individual agency, intergroup social trust, and public trust in institutions. **Democracy tech and Civic tech** refers to technologies developed to empower citizens, enhance civic engagement, and strengthen democratic participation. **Deliberative tech** refers to a class of civic tech that enables a large-scale exchange of views between the public in an iterative discussion, allowing participants to evolve in their understanding. Unlike traditional polls or surveys, deliberative technologies enable participants to share their unique perspectives, listen at scale to others' experiences, values, and ideas, identify common ground, and vote on others' proposals or statements.[45]

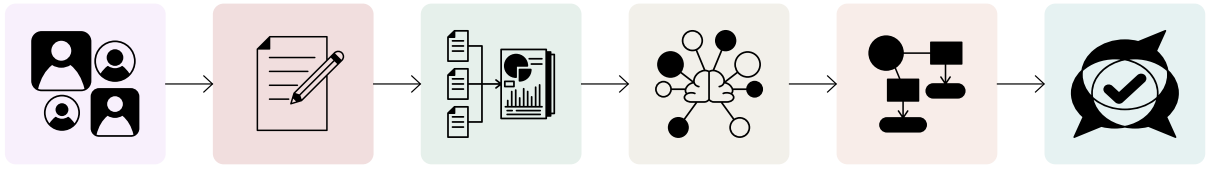
Principles of Democracy and Civic Technologies

Democracy technologies foster inclusive, informed, and values-aware collective decision-making. The following principles guide the ethical and practical development of such tools:

1. **Participation:** Users can propose, deliberate, vote on, and co-create solutions, policies, or shared knowledge.
2. **Pluralism:** Deliberative tools must welcome diverse voices and intentionally cultivate a sense of belonging, ensuring that disagreement does not result in exclusion but fosters deeper mutual understanding.
3. **Design from the Margin:** A justice-oriented approach to design begins not with the average user but with those most excluded or harmed by dominant systems. By centering the needs and insights of marginalized communities, deliberative tech can surface insights that improve equity and strengthen outcomes for everyone.
4. **Open and Accountability:** Some deliberative technologies are open-source by default, allowing for transparency in how decisions are facilitated, data is handled, and algorithms function. Open code enables public scrutiny, ethical accountability, and safeguards against covert manipulation or bias.
5. **Interoperability and Shared Infrastructure:** Some deliberative systems allow users to integrate, fork, or remix tools across platforms and communities.
6. **Adaptability and Localization:** To be inclusive and effective, deliberative tools must be adaptable to different cultural, linguistic, and ethical contexts. Communities should be able to modify or translate systems to reflect local norms, values, and governance needs—ensuring relevance across settings.

The Digital Anatomy of Deliberative Tech

Deliberative technologies go beyond information sharing or polling to create structured opportunities for users to listen, reflect, and co-create. Deliberative platforms facilitate different types of intergroup work, illustrated below, that contribute to dialogue and an exchange of views. Instead of thinking of "which platform is best," it is helpful to think more about a group's goals and which platforms are designed to achieve them.



45 Schirch, Lisa. *Defending Democracy with Deliberative Technologies.* Keough School Policy Brief Series. Notre Dame, IN: Keough School of Global Affairs, 2024.

A public problem-solving process may want to "stack" or sequence different civictech platforms; depending on the goals. LLMs can be used in some steps of the process to synthesize public input or generate options addressing diverse stakeholder interests.

The design of deliberative technologies supports deeper, more inclusive, and values-driven public engagement by guiding participants through a structured decision-making process. These platforms offer participants space to name, define, and frame the issue in their own words. Some platforms provide access to relevant information or educational materials, clearly distinguishing between facts, expert opinions, and personal perspectives.

Some deliberative platforms identify areas of consensus and disagreement while illuminating the underlying values and moral frameworks behind each perspective. Some enable users to discover misperceptions about others' views and encourage reflection on how those misunderstandings influence division or polarization. Through features like structured proposal contribution, weighted online voting, and even participatory budgeting, deliberative technologies give users a concrete role in shaping outcomes. In doing so, they shift public discourse from reactive commentary to co-creative governance.

Uses of Deliberative Technology

There are a wide variety of uses of deliberative technology.

- **Citizens' Assemblies:** Demographically-representative individuals represent the broader public in an assembly to deliberate on public issues.
- **Civil Society Consultations:** Representatives of civil society organizations and groups meet to develop civil society platforms.
- **Participatory Budgeting:** Community members allot budgets for their priorities.
- **Public Input in Peace Processes:** People discuss the pros and cons of ceasefire or political agreements and identify core values and redlines for what they could accept.
- **Platform and AI Governance:** People participate in how digital platforms should work, including for example content moderation and AI-training systems.
- **Social Movement Organizing:** Members deliberate on movement goals and tactics.
- **Future Planning:** Participants can envision an ideal community in 20 years, collaboratively outlining what it would take to achieve that vision.

A3. TIER 5: Middleware to Support Data Sovereignty, Portability, Interoperability

Data sovereignty, portability, and interoperability are interrelated principles that define who controls digital data, how it moves, and how systems communicate. Together, they form the foundation for a more democratic, user-centered, and prosocial digital ecosystem.

Data sovereignty determines legal control over data and ensures individuals and nations maintain rights over data within their jurisdictions. The EU's GDPR enforces this by protecting EU citizens' data that is processed abroad.

Data portability empowers users to move data between platforms, reinforcing autonomy and enabling exit from exploitative systems. **Interoperability** enables systems to work together through data exchange via open protocols.[46]

Governments and advocacy groups such as the Electronic Frontier Foundation and OpenForum Europe have pushed for laws and standards that guarantee these principles. By mandating open protocols and data portability, policymakers can help prevent monopolistic "walled gardens" and support smaller, prosocial platforms prioritizing ethical design and user well-being. For example, the Beckn Protocol in India enables seamless interoperability across digital commerce platforms by providing a universal set of APIs for discovering, ordering, and fulfilling services.[47]

Middleware—third-party tools between users and platforms—lets people choose how they curate content. Instead of relying on centralized platform decisions, middleware enables users to select filters for fact-checking and hate speech removal, reflecting their values. Users could opt for news feeds prioritizing verified content over engagement. Middleware introduces competition while reducing large platforms' power to shape digital information.[48]

Middleware depends on interoperability to access real-time content and services from major platforms. It requires data portability to transfer user profiles and preferences. It enhances data sovereignty by giving users control over who mediates their online experiences without requiring those users to abandon existing platforms entirely. Laws like the EU's Digital Markets Act support middleware ecosystems by mandating platform openness and reducing data lock-in. Encouraging a competitive market for middleware providers could introduce innovation and diversity in recommendation systems.

A user who wants their feed curated by democratic quality rather than algorithmic engagement can port their data (portability), middleware can access content streams (interoperability), and the user can legally choose who governs their information experience (sovereignty).[49] By integrating these principles, middleware offers a pluralistic, user-empowering approach to tech governance that decentralizes control and fosters ethical, inclusive, and civic-minded digital spaces.

---

46 Herr, Trey, Will Loomis, and Stewart Scott. "User in the Middle: An Interoperability and Security Guide for Policymakers." Atlantic Council, June 28, 2023.

47 Foundation for Interoperability in Digital Economy (FIDE). *Beckn Protoco*l. Bengaluru: FIDE, 2025.

48 Fukuyama, Francis, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed, Marietje Schaake, *Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy*. Stanford, CA: Stanford Cyber Policy Center, 2020.

49 Brkan, Maja. "Do Algorithms Rule the World? Algorithmic Decision-Making and Data Portability in the GDPR and Beyond." *International Journal of Law and Information Technology* 27(2) 2019: 91–121.

SECTION B

Provide Foundational Governance for Platform Research

This section offers more details to explain the Blueprint's three recommendations to provide foundational governance for platform research.

Key Recommendations Recap:

B1. Mandate democratic platform oversight, transparency, and audits
B2. Develop a data standard for prosocial tech metrics
B3. Enforce safe harbor protections for accredited researchers

Research provides insights into how design changes affect user behavior. Research helps identify harmful design features. Dozens of independent researchers and whistleblowers inside tech companies document how platform design translates into antisocial impacts and deceptive intentions of platform design.[50] A robust community of interdisciplinary researchers also explores how prosocial designs translate into positive social impacts.

Research on digital platform design and AI systems are similar. Due to commercial incentives and opaque architectures, these issues require independent researchers to uncover risks. Both affect vast user bases and demand ethical scrutiny, yet operate with minimal accountability in regulatory gray areas.

Methods for Collecting Data on Tech's Societal Impacts

There are four broad approaches to gathering data on these types of metrics.

1. **Public Surveys** ask users to rate their perceptions of and experiences on digital platforms. The University of Southern California uses longitudinal tracking surveys called the Neely Social Media Index and the Neely-UAS AI Index.[51]
2. **Online-Offline Correlations** compare online and offline activity to measure digital impacts on social cohesion. For example, researchers can compare online disinformation levels about elections with offline voting patterns during elections.
3. **Platform Democracy** enables public participation in platform governance through transparent decision-making processes, user councils, and mechanisms like voting or feedback systems to shape policies on content moderation, algorithms, and data usage.[52]
4. **Embedded Digital Metrics**: Embedded digital metrics, addressed in the next section, could enable cost-effective, scalable, and regular feedback about platform impacts. Because of the expense and slowness of survey feedback, embedded digital signals are necessary to measure social cohesion indicators.

50 See for example, Vaidhyanathan, Siva. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy.* New York: Oxford University Press, 2018; Alberts, Lize, Ulrik Lyngs, and Max Van Kleek. "Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces That Act Socially." *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024): Article 202; Haugen, Frances. *The Power of One: How I Found the Strength to Tell the Truth and Why I Blew the Whistle on Facebook.* New York: Little, Brown and Company, 2023.

51 Neely Center for Ethical Leadership and Decision Making. *USC Neely Social Media Index*. Los Angeles: University of Southern California Marshall School of Business, 2023.

52 Ovadya, Aviv. *Towards Platform Democracy: Policymaking Beyond Corporate CEOs and Venture Capital*. Berlin: Stiftung Neue Verantwortung, 2021.

B1. Require Democratic Platform Oversight, Transparency, and Audits

Just as financial auditors assess the integrity of corporate accounting, digital auditors can independently evaluate algorithms, manipulative UX designs, and data privacy protection. By providing regular, standardized assessments, professional auditors can help identify systemic risks, expose regulatory blind spots, and offer recommendations that protect users and democratic institutions from digital harm. Digital platforms operate in a black box, with complex algorithms and vast user data. Yet, their decision-making on content moderation, ad targeting, and recommendation systems often remains hidden. Without transparency, understanding how platform designs are impacting society is impossible.

Requiring transparency[53] would allow independent auditors and academic researchers to understand how technology design choices impact society. Researchers formed the Coalition for Independent Technology Research to advance the public interest in understanding how technology platforms affect individuals and societies.[54]

Tech companies discourage research on their platforms' social impacts since scrutiny can expose harms and threaten profits. Their resistance stems from protecting profits and preventing reputational damage, often at the expense of transparency. Companies often cite competition, user privacy, and legal concerns as reasons for restricting data sharing, which in turn impedes independent research efforts.[55] This asymmetry in access to information means that independent researchers are often unable to verify claims made by the platforms or explore harms at the scale at which they occur.[56]

Researching tech platforms' social impacts is difficult due to limited transparency. Companies do not release datasets, algorithms, or moderation decisions that shape user behavior, forcing researchers to rely on incomplete methods like scraped data or user surveys. Independent researchers lack access to study the impact of digital platforms.

Big Tech companies may suppress independent research while promoting their studies, widening the knowledge gap. While social media companies earn billions in profit, independent researchers often lack the resources to study online impacts on society. In this "David and Goliath" battle against well-funded tech giants, comprehensive analyses of digital technologies' effects on mental health and societal well-being are missing.[57] Meta's Discontinuation of CrowdTangle, a tool instrumental for researchers and journalists in tracking misinformation, raises concerns about the implications of reduced transparency tools for studying social media's societal impacts. [58]

Studies involving digital harms raise ethical concerns about user consent, surveillance, and potential platform reputational damage—leading to resistance. Researchers can face legal threats (e.g., under the US Computer Fraud and Abuse Act), defamation lawsuits, or non-cooperation from tech firms.

---

53 Black, Renee "Foundational Principles for Advancing Prosocial Design Governance," GoodBot, May 2025. https://www.goodbot.ca/tech-policy/foundational-principles

54 Coalition for Independent Technology Research. "A Case for Ethical and Transparent Research Experiments in the Public Interest." *Independent Tech Research*, 2023.

55 Heath, Ryan, and Sara Fischer. "Researchers Fight to Access Big Tech Data." *Axios*, 1 Aug. 2023,

56 Barakat, Hanna. "How Information Asymmetry Inhibits Efforts for Big Tech Accountability." *TechPolicy.Press*, APRIL 16 2025.

57 Orben, Amy, and Nathan Matias. "Fixing the Science of Digital Technology Harms." *Science* 388, no. 6743. April 10, 2025: 152–155.

58 Elliot, Vittoria. "Meta Kills a Crucial Transparency Tool at the Worst Possible Time." *Wired*, March 25, 2024.

B2. Develop a Data Standard for Prosocial Tech Metrics

"If you can't measure it, you can't grow it." Nancy McMahon's insight, published in the *Stanford Social Innovation Review* in 2013, captures a foundational truth.[59] In today's digital world, metrics are the invisible currency that guides design choices, business models, and social norms. They function as price signals—proxies for value that determine where attention, investment, and innovation flow. Markets reward what they measure.

Unfortunately, most digital metrics reward attention, emotion, and scale, not truth, well-being, privacy, safety, or social cohesion. Time-on-site, clicks, shares, and daily active users dominate as key indicators of success. But these metrics favor content that inflames rather than informs—leading platforms to ignore metrics related to polarization, manipulation, or distrust.[60]

Metrics as Price Signals

A **price signal** conveys the perceived value of a good or behavior. In the digital economy, engagement metrics act as price signals that reward what gets attention—not necessarily what builds trust or strengthens democracy. Traditional platform metrics include

- Time spent on the platform
- Number of likes, shares, or comments
- Content virality
- Daily/monthly active users

Not all high-metric content is valuable. Outrageous or false content might get more engagement, but that doesn't mean it's good. Platforms fine-tune their systems using A/B testing and machine learning loops to maximize these metrics, reinforcing designs that exploit vulnerabilities rather than serve the public good. Platforms chasing metric signals can incentivize harmful content because it looks "valuable" in the data. When platforms use engagement metrics as price signals, they create invisible markets for attention, emotion, and influence.

A Price Signal for Social Cohesion

Companies, investors, governments, and the public need ways to measure both antisocial and prosocial impacts alongside growth and profit. A price signal for social cohesion could measure and reward prosocial outcomes like empathy, trust, and civic participation. That means building new embedded metrics that capture the three main dimensions of social cohesion: individual agency, intergroup pluralism and trust, and vertical cohesion (trust between people and institutions).[61]

Margarita Quihuis and Mark Nelson, co-directors of the Stanford Peace Innovation Lab, have pioneered the development of the Peace Data Standard. This framework reimagines peace as a measurable, data-driven outcome of technology-mediated human interactions. Their work defines peace not merely as the absence of conflict but as the presence of positive, prosocial behaviors that foster mutual benefit across social divides.[62] Identifying four key

59 Mahon, Nancy. "Metrics Matter." *Stanford Social Innovation Review*. 2013.

60 Brady, William J., Steve Rathje, Laura Globig, and Jay J. Van Bavel. "Estimating the Effect Size of Moral Contagion in Online Networks: A Pre-registered Replication and Meta-analysis." OSF Preprints. April 2, 2025.

61 Schirch, Lisa, Mark Nelson, Margarita Quihuis, "Prosocial Tech Metrics". *Toda Peace Institute*, forthcoming 2025

62 Quihuis, Margarita. "Quantifying Positive Peace: The Imperative for a Peace Data Standard." *Medium*, November 2, 2023.

components—group identity information, behavior data, longitudinal data, and metadata—provides a structure for collecting and analyzing "peace data" to assess and enhance intergroup engagement.[63]

Examples of Prosocial Metrics and Indicators

A **metric** is a numerical value used to track performance or behavior. It's often a **raw data point** or a directly measured outcome. An **indicator** is a signal or proxy for a broader concept. A composite signal might combine multiple metrics to express a more abstract idea like "trust" or "well-being."

Individual agency and well-being, intergroup social trust, and state-society public trust are indicators of social cohesion. The table below offers examples of embedded metrics for social cohesion.

Individual Agency & Well-Being The ability of users to make informed choices and participate meaningfully	**Active Participation Rate:** % of users posting, commenting, or voting in platform governance
	Well-Being Check-Ins: User-reported satisfaction and trust via brief in-platform surveys
	Belonging Score: % of users who agree with "I feel like I'm part of a community on this platform."
Social Trust and Intergroup Pluralism Quality of interaction across diverse social, cultural, or political groups	**Cross-Group Interaction Rate:** Frequency of engagement across identity lines
	Participant Diversity: Demographic spread (age, gender, geography, ideology, language) of active users
	Inclusivity Score: Participation rates among marginalized or underrepresented groups
	Supportive Behavior Frequency: How often users offer help, praise, or positive feedback
Public Trust Strength of the relationship between users and public institutions	**Civic Engagement Rate:** Frequency of participation in digital public forums, petitions, or surveys
	Institutional Sentiment: AI analysis of public discourse about government and institutions
	Trust Signal Keywords: Frequency of words like "fair," "transparent," or "corrupt" in user feedback

63 Guadagno, Rosanna E., Mark Nelson, and Laurence Lock Lee. "Peace Data Standard: A Practical and Theoretical Framework for Using Technology to Examine Intergroup Interactions." *Frontiers in Psychology* 9. 2018: 734.

B3. Enforce Safe Harbor Protections for Accredited Researchers

Despite digital platforms and AI companies' commitments to encourage third-party discovery of issues, many impose barriers that deter safety research. Terms of Service prohibit probing for vulnerabilities, and researchers risk legal action for such evaluations. This lack of transparency hampers identifying digital platform risks, including AI systems, which could affect national security and public trust. Researchers report that the possibility of a lawsuit for 'un-permissioned research' is a significant obstacle to researchers independently assessing platform claims and behavior. New reports from the Knight First Amendment Institute[64] and Federation of American Scientists[65] identify a range of related protections for AI researchers and tech company whistleblowers.

Create Safe Harbor Provisions as legal protections for independent researchers who evaluate platforms and AI systems in good faith, shielding them from liability or punitive action when acting in the public interest. Safe harbor policies aim to prevent tech companies from suing or criminally accusing public-interest researchers who use automated means to collect public-facing platform information in cases where researchers take appropriate steps to safeguard data and user privacy. Companies could not hold researchers liable for contract violations or threaten criminal liability if the research met the prescribed conditions.[66]

Standardize and Develop Norms and Mechanisms for Flaw and Design Disclosure to ensure developers receive timely, constructive feedback while protecting researchers.

Clarify Terms of Service to Permit Research so tech companies don't penalize good-faith testing of models, especially through publicly accessible interfaces.

Support Infrastructure for Independent Research and encourage the creation of shared tools, datasets, and secure sandboxes to facilitate rigorous, transparent AI safety evaluations by independent experts. It is important to recognize the challenges nonprofits (especially from the Global South) face in accreditation systems. Existing accreditation systems to access data are more accessible for academic institutions from the Global North.

Expand Bug Bounty Programs Companies like Google, Meta, and Microsoft run bug bounty programs to strengthen cybersecurity. Antisocial design issues like manipulative UX patterns and harmful content amplification are design-level rather than code-level problems. Expanding bug bounty programs to include design harms could reward researchers for flagging deceptive patterns and discriminatory systems. These programs could also incentivize AI researchers to identify safety flaws like misinformation or bias in AI models.

64 Johnson, Nadine Farid. "Responsible AI Regulation: Supporting Independent Researchers." Knight First Amendment Institute, March 14, 2025

65 Klyman, Kevin, Sayash Kapoor, and Shayne Longpre. *A Safe Harbor for AI Researchers: Promoting Safety and Trustworthiness Through Good-Faith Research*. Federation of American Scientists, June 28, 2024.

66 Black, Renee "Foundational Principles for Advancing Prosocial Design Governance," GoodBot, May 2025. https://www.goodbot.ca/tech-policy/foundational-principles

SECTION C

Shift Market Forces to Support Prosocial Design

This section addresses how market forces incentivize design choices and metrics for measuring success. Shifting market forces is necessary to reduce digital harms and support prosocial design.

Key Recommendations Recap:

C1. Enforce competition, antitrust, and anti-monopoly laws to boost prosocial innovation
C2. Codify product liability for adverse impacts
C3. Incentivize and invest in advanced prosocial technology markets

Market forces drive antisocial and deceptive tech design.[67] Venture capital (VC) funding is the predominant source of financing for many major tech platforms, especially in their early and growth stages. It plays a significant role in entrenching antisocial tech design, expecting rapid scaling, high returns, and market dominance—often at the expense of ethical development. Here's how:

- **Data Brokers and Ad-Based Monetization:** Pressures from VC investors push many tech companies to monetize as data brokers and ad platforms. More data extraction increases platform value. This incentivizes invasive tracking and deceptive interface designs that obscure consent. Platforms monetize user behavioral data through targeted advertising.
- **Growth-at-All-Costs Culture:** VCs expect startups to achieve hypergrowth to deliver 10x or 100x returns with "growth hacking" or rapid design experimentation to optimize user behavior.
- **Engagement Maximization Metrics:** Platforms gain financially when users spend more time scrolling and reacting. Emotionally charged content—especially outrage, fear, or tribalism—drives more engagement, leading algorithms to amplify polarizing content. Engagement-based success metrics (time spent, clicks) misalign with the public interest.
- **Exploiting Human Psychology:** The focus on growth commonly involves exploiting psychological vulnerabilities like social comparison, which fosters compulsive use rather than reflective, healthy interaction. Platforms adopt addictive features (e.g., infinite scroll) in early design.
- **Deceptive Interface Design:** Manipulative patterns such as hiding unsubscribe buttons or auto-enrolling users) exploit cognitive biases to maximize clicks, subscriptions, or data consent.
- **Ignore Externalities and Delay Accountability:** Most VC-backed platforms externalize harms (e.g., polarization, harassment, disinformation) because they are not financially liable for social consequences. VCs often encourage a "scale now, fix later" mentality that defers questions of harm, bias, or regulation until after market domination—when accountability becomes harder to enforce.
- **Exit Pressures:** Startups are often pushed toward IPO or acquisition, encouraging designs that increase market valuation through scale and data extraction—not social responsibility.

67 See Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019; Barford, Paul and James Mickens, et al., *The Drivers of Platform Harm: Toward an Evidence-Based Framework for Measuring and Mitigating Platform Harms*. Cambridge, MA: Belfer Center for Science and International Affairs, 2023.

- **Market Consolidation and Lock-In:** VCs favor “winner-take-all” platforms that monopolize user networks and data. This discourages interoperability, user control, and democratic governance—since these features reduce platform lock-in.

Ethical design becomes a liability under these financial logics. To better incentivize fund managers to invest in prosocial tech, policymakers and philanthropic actors should establish blended finance mechanisms—such as public-private co-investment funds, risk guarantees, or outcome-based grants—that reduce the financial risk of investing in socially beneficial technologies. Prosocial investments could be more competitive by aligning financial incentives with long-term social impact.

C1. Enforce Competition, Antitrust, Anti-monopoly Laws

Market concentration stifles innovation and locks users into systems optimized for profit, not well-being. Many big tech companies operate as monopolies using opaque tactics and value chain domination. Technology monopolies make it difficult for smaller prosocial platforms to grow. When a handful of tech giants dominate infrastructure, data, and user attention, smaller platforms with ethical, inclusive, or democratic designs struggle to gain visibility or viability.

By leveling the playing field—through fair API access, anti-bundling rules, and structural separation—governments can open space for prosocial alternatives to emerge and thrive. A more competitive digital environment empowers users with real choices and drives innovation that aligns with public interest rather than monopolistic control.

Enforcing competition, antitrust, and anti-monopoly laws can open space for prosocial alternatives to emerge and thrive by leveling the playing field.[68] Governments can adopt, amend, or remove regulations to enhance competition among platforms, enabling new platforms incorporating prosocial designs to compete with big-tech platforms. Governments play a crucial role in preventing large technology corporations from exploiting dominant market positions to the detriment of consumers and undermining markets,[69] including those involved in prosocial technological innovation. Governments can reduce entry barriers, facilitate market access, mitigate incumbency advantages, and lower compliance costs for smaller companies, benefiting from prosocial innovation.

Ensuring equitable access to APIs and infrastructure, maintaining platform neutrality, and implementing regulations, such as anti-bundling measures, are crucial for promoting competition and innovation. Authorities can also break up corporations that dominate multiple layers or substantial portions of the technology stack, including hardware, software, cloud services, and app ecosystems, to limit their capacity to monopolize various market sectors. This approach empowers users to opt for prosocial services rather than profit-driven alternatives. It also makes it easier for startups to compete without being overshadowed by pre-installed or bundled offerings from major platforms.[70]

However, anti-monopoly regulations can introduce new risks. Fragmentation could lead to inconsistent safety standards, weaker moderation, and reduced accountability, especially if new entities lack their predecessors' capacity or regulatory oversight. Users may face a fragmented digital experience and diminished protections without interoperability mandates. Dividing complex infrastructure and data systems could also increase security vulnerabilities and create governance blind spots. To ensure that antitrust action reduces harm rather than exacerbating it, prosocial design standards and enforceable transparency requirements are necessary to complement these laws.

68 Khan, Lina M. "Amazon's Antitrust Paradox." *Yale Law Journal* 126(3) 2017: 710–805.

69 Black, Renee, "Addressing the Enabling Conditions for ProSocial Technology Markets" GoodBot Society. May 5, 2025. https://www.goodbot.ca/tech-policy/enabling-conditions

70 Hovenkamp, Herbert. *Tech Monopoly.* Cambridge, MA: MIT Press, 2024.

C2. Codify Product Liability Adverse Impacts

Taxing the polarizing externalities of digital platforms—such as algorithmic amplification of outrage, misinformation, or societal division—offers a promising way to internalize the social costs currently offloaded onto the public. Like carbon taxes designed to incentivize environmental responsibility, a "polarization tax" could push platforms to redesign their systems to promote empathy, pluralism, and civic trust rather than engagement at any cost. By turning harmful social outcomes into measurable liabilities, this approach could align platform incentives with democratic resilience. The revenue collected could also fund public-interest initiatives such as media literacy, civic tech, or mental health programs—reinforcing a broader ecosystem of digital well-being.

Codifying product liability for the adverse social impacts of platform design is essential to reducing digital harm because it shifts accountability from individual users and content moderation to the systemic design choices that drive large-scale societal risk. Many dominant platforms externalize harms like polarization, disinformation, and diminished civic trust—costs borne by society but absent from company balance sheets. By imposing financial disincentives such as liability, insurance requirements, or taxation based on metrics like a "polarization footprint," regulators can align corporate incentives with public well-being.

This approach also advances prosocial design by creating a regulatory environment where ethical innovation is not a competitive disadvantage. If platforms are held accountable for the negative externalities of their design, stakeholders—from investors to engineers—have a tangible reason to prioritize features that promote safety, cohesion, and democratic integrity. Liability thus becomes a market-correcting mechanism, encouraging transparency, risk reduction, and the development of technologies that serve the common good.

Traditional VC-funded tech platforms often oppose research or regulations that link their platform design to societal harm, instead focusing on content policies focused on individual harm. A system-wide analysis recognizes negative externalities, such as data breaches, misinformation, polarization, and community harms, which impose significant societal costs not reflected in financial metrics. Regulations should impose distinct obligations on large platforms, whose models create societal-level risk at the scale, including financial disincentives.

These include requiring insurance, applying fees, or taxing platforms using designs that pollute information ecosystems and algorithms that promote divisive content. Success metrics must include these externalities, allowing stakeholders, such as investors, regulators, and users, to assess long-term sustainability and prosocial impact.

For example, Build Up researched the potential of taxing platforms, data centers, and/or data brokers based on their "polarization footprint"—a measure of how content recommendation systems contribute to polarization.[71] Enforcement would require prosocial metrics outlined earlier in this Blueprint to meaningfully measure platform impacts on social cohesion.

71 Puig Larrauri, Helena. *Societal Divides as a Taxable Negative Externality of Digital Platforms*. Build Up, March 2023.

C3. Incentivize and Invest in Prosocial Tech

Prosocial funding and monetization models for technology already exist. Funding models refer to how a platform raises capital, especially in early stages (e.g., grants, venture capital, crowdfunding), including covering startup and growth costs. Monetization models refer to how the platform generates ongoing revenue (e.g., subscriptions, ads, licensing) to sustain operations over time.

Prosocial Funding Models

Prosocial Venture Capital funders prioritize social impact, ethical technology, and public interest over purely profit-driven motives. Funds like Purpose Ventures promote alternative ownership models and non-extractive financing. Organizations like Village Capital and the Next Billion Capital seek out investments to create positive social impacts in marginalized communities.

Public Funding, such as government grants and procurement programs support new civic tech innovations (see Prosocial Tier 4). Public funding, such as the European Commission's Next Generation Internet (NGI) of Trust, align technology with ethical guides on Public Digital Infrastructure and the UN's Global Digital Act that advance digital rights, public participation, and social trust. Transparent and fair government procurement can also help small, prosocial companies.

Philanthropic Funding from foundations, donor-advised funds, or mission-aligned investors provides early-stage capital to ventures that prioritize social impact, equity, and public interest over rapid profit. Funders such as the Mozilla Foundation back startups developing technologies that serve marginalized communities, protect privacy, or strengthen democracy. Challenge-driven philanthropic funding incentivizes innovation for the public good. For example, XPRIZE offers large, open prizes to teams that can solve pressing global challenges—ranging from climate change and pandemic preparedness to education and clean water.

Community and cooperative funding models offer prosocial alternatives to traditional venture capital by emphasizing collective ownership. Through crowdfunding platforms like Kickstarter, users contribute directly to projects they believe in, fostering a sense of shared responsibility. Platform cooperatives take this further by making users or workers co-owners of the tech they use, distributing revenue among members. Mutual aid and commons funding supports community-managed digital infrastructure, such as mesh networks or data trusts. Quadratic funding, used in open-source and civic tech ecosystems like Gitcoin, matches community donations in a way that amplifies broad-based support, ensuring that funding reflects collective priorities rather than wealthy backers. An example is Zebras Unite, which advocates for cooperative and mission-aligned tech ventures that balance purpose with profit while avoiding exploitative growth models.

Universities, NGOs and multilateral contracts provide a prosocial funding pathway for technology by directly supporting tools that advance peacebuilding, election integrity, and humanitarian response. Organizations like the United Nations or the International Committee of the Red Cross often fund the development of tech platforms for conflict monitoring, digital ID systems for displaced populations, or secure communication tools for civil society actors. These contracts prioritize mission alignment over profit, ensuring that technology serves urgent societal needs and is shaped by the field realities of those it aims to help.

For example, the University of Waterloo's Grebel Peace Incubator supports tech innovations that unpack complex global challenges. The University of Notre Dame's Peacetech and Polarization Lab is developing opensource tools

for using LLMs to support civic discourse and deliberative technologies. George Mason University's Center for Peace Tech incentivize entrepreneurship to support conflict-affected communities. The Civic Health Project is building an LLM detoxifier bot for social media.

Prosocial Monetization Models

Prosocial monetization models avoid data extraction and advertising, as these are root drivers of antisocial and deceptive tech design.[72]

Models with high potential for supporting prosocial outcomes—such as subscription-based, software-as-a-service (SaaS), licensing, and pay-per-use—focus on value delivery without exploiting user data or manipulating engagement. Freemium with ethical paid tiers allows users to access core features for free while offering advanced tools behind a paywall. Sliding scale pricing adjusts costs based on a user's income or capacity, often used in education and nonprofit tech. For instance, companies like Microsoft, Netflix, and Salesforce exemplify approaches that avoid invasive advertising or coercive engagement tactics.

Prosocial Licenses

Licenses determine who can use the software, how they can use it, and whether they must share their changes. GPL (General Public License) and AGPL (Affero General Public License) are critical for prosocial tech startups because they ensure the software remains open, transparent, and resistant to corporate enclosure, aligning with democratic ownership, accountability, and shared benefit. GPL requires modified versions to stay freely available under the same license. At the same time, AGPL extends these protections to software-as-a-service platforms by requiring source code disclosure when servers host software. Together, they establish a legal commons that prevents proprietary appropriation and fosters community reciprocity.

A prosocial tech startup could adopt these licenses to build trust, attract public-interest collaborators, and ensure openness. To sustain this model, the startup could combine public sector support (grants, procurement), mission-aligned venture capital (with capped returns), and community funding like quadratic funding or retroactive public goods rewards. Equity-free accelerators, co-op ownership, and subscription models can provide financial stability without compromising ethical commitments or subjecting the platform to extractive monetization pressures.

72 See Cast, LX. "Aligning Incentives to Develop Prosocial Technology: Funding Infrastructure." *Tech Policy Press*, September 17, 2024.

Conclusions and Caveats

All forms of tech governance face significant challenges. These include the following.

This Blueprint advances a vision for **prosocial tech design**, where platform designs and systems support public interest outcomes like democratic resilience, user agency, and social cohesion rather than engagement maximization or profit at any cost. It proposes a tiered framework that classifies platforms based on design risk and incentivizes safer, more ethical practices through tools like certification, liability, and building codes. A second central theme is a need for **research transparency and protections**; the paper calls for mandated platform audits, researcher safe harbors, and public infrastructure to enable independent scrutiny of algorithmic systems and user experiences. Without these safeguards, crucial insight into systemic harms—such as manipulation, bias, and disinformation—remains inaccessible. Finally, the paper addresses **shifting market forces**, emphasizing the importance of antitrust enforcement, interoperability, and public investment to open space for ethical alternatives. Together, this Blueprint outlined a path toward governance that aligns platform incentives with democratic values and collective well-being.

Implementing prosocial tech governance is not without challenges, dangers, and cautions.

Bureaucratic Challenges

Government efforts to regulate technology face bureaucratic challenges due to siloed agencies that separately address privacy, cybersecurity, data, or infrastructure, leading to fragmented and sometimes conflicting policies. Compounding this, the global nature of digital platforms creates jurisdictional conflicts, as national laws often clash or fail to align, making consistent and effective governance difficult across borders.

Ethical Challenges

Ethical challenges in technology governance arise from the complexity of defining standards for emerging tools and platforms. Competing priorities create persistent tensions between individual rights and collective well-being.

Technical Challenges

Technology evolves faster than governance or regulatory frameworks, making it challenging for governance to keep up with emerging issues like AI, blockchain, and quantum computing. The opacity of complex algorithms, AI, and machine learning makes it difficult to ensure accountability and fairness.

Trust and Safety Challenges

Tensions also arise between protecting intellectual property and promoting open-source innovation. Growing public distrust in government surveillance and corporate data misuse undermines regulatory efforts and complicates the enforcement of policies intended to protect users and foster innovation.

Economic Challenges

Large tech companies have significant power, lobbying influence, and control over markets, making it hard to regulate their practices effectively.

While prosocial tech design governance aims to reduce harm, it carries several potential dangers if not carefully implemented.

Eurocentric and Ideological Design Bias

Technology platforms from Western countries often prioritize Western norms and values, which can marginalize local contexts and create products unsuited for the Global South. This bias, along with Western-centric datasets for AI systems, can lead to systemic bias and a lack of transparency in algorithmic decision-making. The Global South's linguistic diversity challenges the large-scale use of LLMs to detect toxic content or help groups talk to each other. Platforms prioritize dominant languages, ignoring diverse communities due to limited profitability. Defining what constitutes "prosocial" may reflect dominant cultural or political values, potentially marginalizing dissenting views or reinforcing existing power structures. This becomes particularly problematic in polarized or authoritarian contexts.

Individual Agency

Systems designed to be "good" may still centralize power and reduce user autonomy or informed consent. For example, nudges designed to encourage thoughtful interaction may frustrate users, reducing their sense of control.

Censorship risks

Without clear standards, autocratic governments could use design to achieve political ends. Tools meant to nudge users or demote harmful content can easily slip into soft censorship, where platforms or governments suppress legitimate speech under the guise of harm reduction. In countries with weaker democratic institutions, governments may use technology regulations to suppress dissent, enhance surveillance, and enforce censorship.

Ethicswashing

Without clear standards, corporations might adopt prosocial design labels (like ESG or ethical tech seals) without a certification process for branding or to preempt stronger regulation without making meaningful changes.

Unintended consequences

Even well-meaning design interventions can backfire in complex systems, increasing mistrust or reducing engagement, especially from marginalized groups. Ultimately, prosocial tech design requires careful navigation of these competing ethical priorities, relying on transparent deliberation, pluralistic governance, and context-specific trade-offs rather than one-size-fits-all solutions.

Research Methodology and Workshops

From December 2023 to April 2025, the Council on Tech and Social Cohesion (hereafter "Council") members organized 8 workshops with a total of over 450 people to focus on a variety of topics related to prosocial tech design and policy. These workshops brought together venture capital, private funders, technologists, civil society leaders, bridgebuilding and peacebuilding practitioners, and policy experts. Each workshop revised drafts of the Blueprint outlining a systemic approach to prosocial tech design, including economic and policy incentives as well as the need for research and transparency to understand platform design impacts on society.

- **Nairobi, Kenya:** Eight members of the Council presented at 3-day Build Peace conference in Nairobi, Kenya from December 1-4, 2024, attended by approximately 250 civil society groups.
- **Designs Solutions Summit, Washington, DC:** Ravi Iyer of the Council organized and facilitated a 1 day summit with 30 people to explore tech design options to address threats to elections on March 12, 2024.
- **US Policymaker Meetings:** On March 12, Ravi Iyer and Lisa Schirch met with individual Congressional offices and the US Federal Trade Commission to discuss the Neely Design Code on March 13, 2024.
- **Defending Democracy with Deliberative Technologies Symposium, Washington DC:** The University of Notre Dame Keough School's Washington DC office space on March 14, 2024, with Nobel Peace Prize winner Maria Ressa. The Symposium included 150 Washington DC think tanks and policy advocates.
- **Venture Capital and Prosocial Design, San Francisco:** Laure X Cast and Lisa Schirch of the Council facilitated a 1-day conference for 50 Venture Capitalists and private funders on "Incentives for Prosocial Technology" at the Internet Archive in San Francisco on May 1, 2024.
- **Prosocial Designs for Comment Sections, San Francisco:** The Plurality Institute and the Council co-facilitated a 1-day workshop for 45 researchers, technologists and journalists on May 2, 2024.
- **Deliberative Tech in Polarized Contexts, University of Notre Dame:** The Toda Peace Institute, University of Notre Dame's Kroc Institute, and Council co-organized a 4-day workshop with six technologists leading sessions for 47 civil society members from 19 countries from June 24-27, 2024.
- **Brussels, Belgium:** Lena Slachmuijlder of Search for Common Ground and the Council organized a 1-day workshop for 24 people on October 8, 2024 to review the first draft of the Blueprint.
- **European University Institute in Florence, Italy:** Michele Giovanardi of the EUI and Lisa Schirch, Ravi Iyer, and Lena Slachmuijlder of the Council organized a 2-day workshop for 17 people on October 9-10, 2024 to discuss and expand the draft Blueprint.
- **Center for International Governance Innovation (CIGI) in Waterloo, Ontario:** Paul Heidebrecht and Renee Black from the Council organized a 1-day workshop for 28 people on December 3, 2024 to review a revised Blueprint draft.
- **Notre Dame Law School:** Lisa Schirch participated on a panel on "The Regulation of Social Media" with members of the Meta Oversight Board on March 28, 2025, discussing ideas on tech design regulation.
- **Democracy Exchange, Toronto:** Renee Black and Lisa Schirch from the Council led a session on the Blueprint for at the Democracy Exchange on April 4-5, 2025.

About

This report is co-published by three organizations.

The Council on Technology and Social Cohesion

The idea for a Council emerged from a podcast by the Center for Humane Technology and Search for Common Ground in 2021. However, the groundwork for the Council emerged from the foundational work of organizations like Build Up, Toda Peace Institute, New_Public, Psychology of Technology Institute, Integrity Institute, and the University of Notre Dame's Peacetech and Polarization Lab. The Council on Technology and Social Cohesion launched in February 2023 with a global conference on "Designing Technology for Social Cohesion." The conference attracted 350 people and 36 speakers on prosocial tech design were present. In 2024, new partnerships formed with other key organizations, including the Exygy, GoodBot, the Alliance for Peacebuilding's Community of Practice on Digital Peacebuilding, the University of Waterloo's Grebel Peace Incubator, the European University Institute's Global Peacetech Hub, the University of Berkeley's Center for Humane AI, the Civic Health Project, Plurality Institute, and many more organizations.

The University of Notre Dame's Peacetech and Polarization Lab

The PeaceTech and Polarization (PTAP) Lab, part of the Kroc Institute for International Peace Studies within the Keough School of Global Affairs, offers students the opportunity to study and research the design and functions of technology and its intersection with peacebuilding and democracy. The Lab aims to contribute to social cohesion through technology that supports public problem-solving.

The Toda Peace Institute

The Toda Peace Institute is an independent, nonpartisan institute committed to advancing a more just and peaceful world through policy-oriented peace research and practice. The Institute commissions evidence-based research, convenes multi-track and multi-disciplinary problem-solving workshops and seminars, and promotes dialogue across ethnic, cultural, religious, and political divides. It catalyzes practical, policy-oriented conversations between theoretical experts, practitioners, policymakers, and civil society leaders to discern innovative and creative solutions to the significant problems confronting the world in the twenty-first century (see www.toda.org for more information).