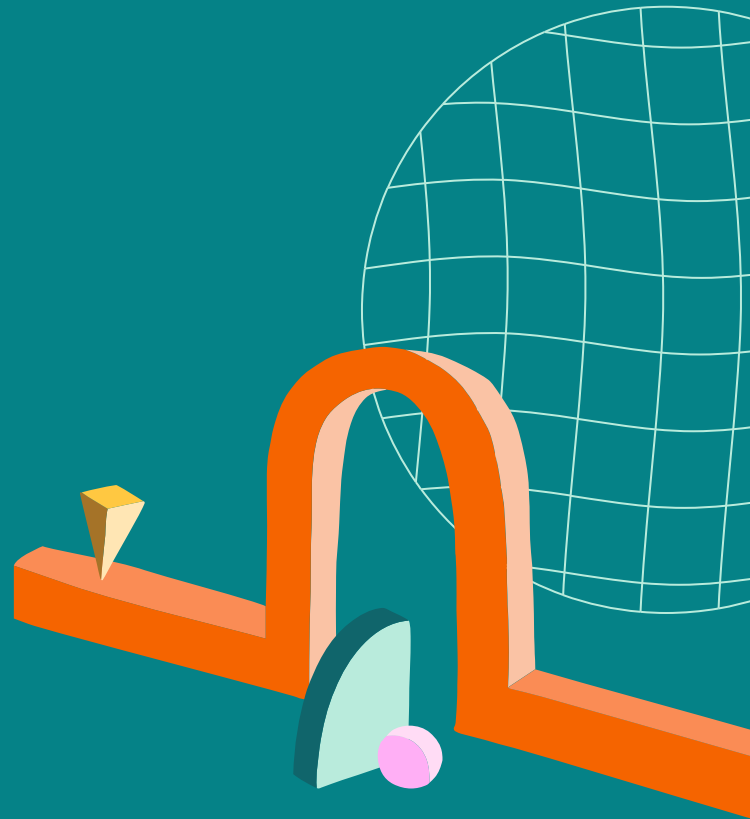


After the Feed: Trust, connection, and the next era of social technology

By Eli Pariser, Co-Director,
Angelica Quicksey, Managing Director,
Arnobio Morelix, AI Fellow



May 2026

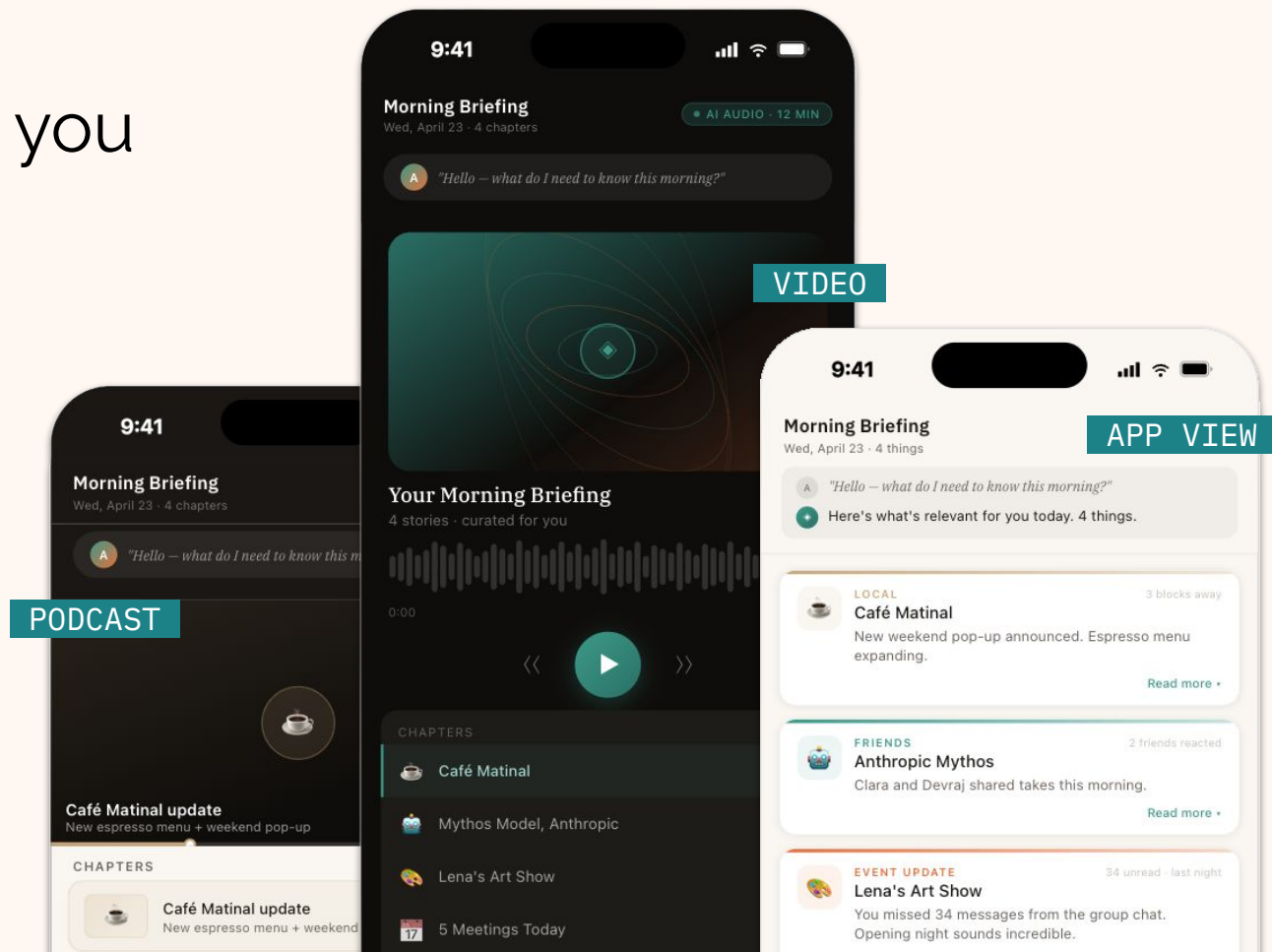
Newpublic.org

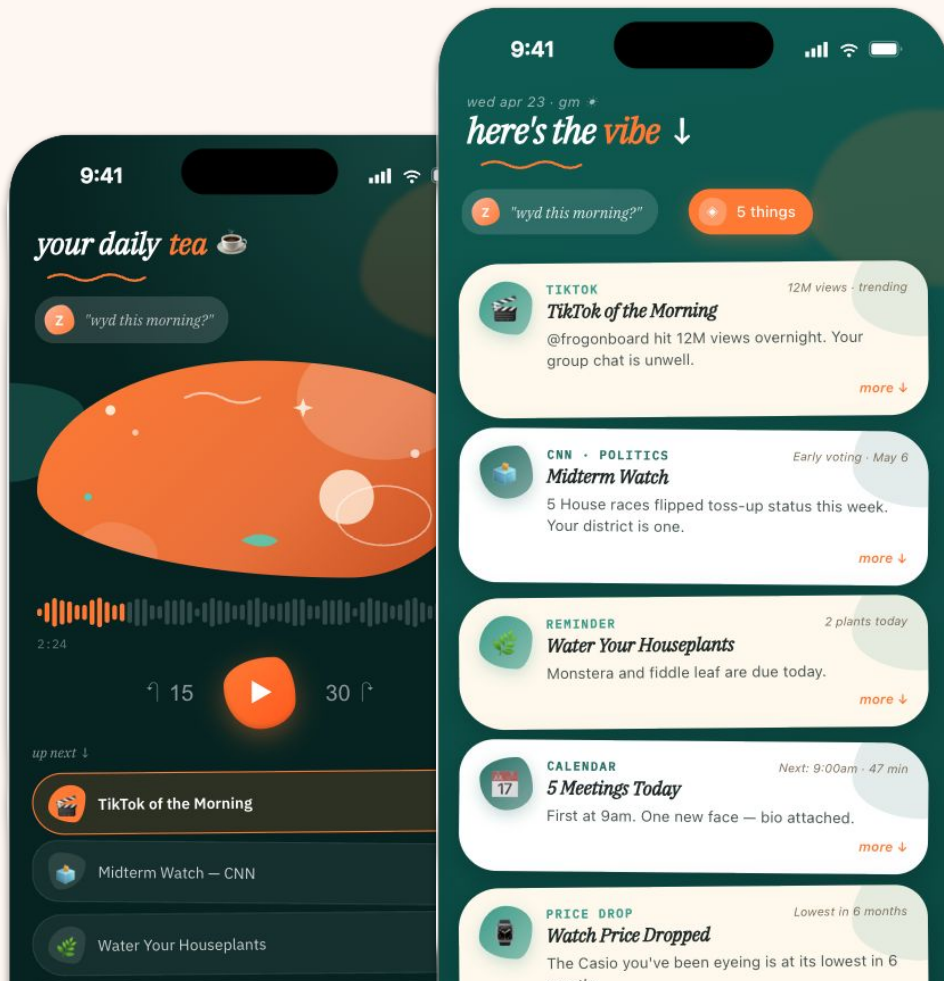
New_ Public

You wake up, pull out your phone
and greet your AI agent:

*Hello, what do I need to know this
morning?*

They respond
with everything you
need to know,
curated for you,
in the format
you prefer.





For your
very online
friend, the
morning update
might look
and feel a
little different.

Meanwhile, the
busy parent's
update comes
while they're
on the move.



Many sources, synthesized and curated *for you*.





You have entered the
agentic interface era.

In this new era,
the social feed is
no longer the
default interface
for digital life.

The **social feed** of algorithmically ranked posts, images, videos was optimized over a decade to maximize time-on-platform.

The feed is not going to disappear. People will still scroll — this is where parasocial relationships and entertainment will continue to thrive. But the feed is no longer where our online **social** lives and information diet will live.

Agentic interfaces are the new intermediaries for information about the world around you. This looks like a chat with Claude or a briefing from your personal AI agent — ***an interface built for an audience of one.***

These agentic interfaces will increasingly become the nexus through which you access information and connection.

In 2011, Eli gave a talk introducing the idea of the **Filter Bubble**.

His thesis: If left unchecked, personalized algorithms will have negative consequences for people and democracy



15 years later, we're at another **inflection point** in the social technology landscape.

Our central question:

What trends will most significantly shape **how humans connect to one another, build relationships, and have shared experiences** in online spaces in the **next 2 years**? And how might these trends support human thriving and connection?

Trends driving the agentic interface era

01

Agentic interfaces are the
new new gatekeepers

02

AI is destabilizing the big
social platforms

03

The emerging trust economy

04

Entering a world of
software abundance

05

In an agentic world,
human connection survives

01

Agentic interfaces are the
new new gatekeepers

The *New New* Gatekeepers

We're experiencing a changing of the guard:
A set of *new new gatekeepers* controls what the public sees

01

Old Gatekeepers

Institutions: Newspaper editors, TV network anchors

1850–2010

This was the
Filter Bubble era

New Gatekeepers

Platforms & Algorithms: Facebook, Google, Twitter feeds

2010–2025

03

New New Gatekeepers

Agentic interfaces: LLM-powered interfaces tuned to you

2025–Present

These agentic interfaces are becoming the medium for social life and information online

Platform Era	Agentic Interface Era
AI-powered algorithms optimize for human attention	Content, data, systems are optimized for agent attention
Social platforms surface unaltered posted content through algorithmically ranked feeds	Content and conversation often does not pass directly between humans; it is summarized or transformed by the agentic interface
Humans author most content, actively choosing what to share and when — on a very human timeline	Humans (sometimes using AI) and AI agents both create content, but AI does so at inhuman speed and scale
AI sits in the background as ranking and recommendation infrastructure	AI is in the foreground , summarizing, remixing, and sharing information, shaping your (online) reality

The New New Gatekeepers

And these new gatekeepers will collect and use every piece of information about you that they can find.

Then

In order to provide you with content and ads you might engage with, the surveillance advertising-driven platforms of the last two decades formed a ***composite image of people like you, and targeted that image.***

Now

Today's agentic interfaces are formed around *you*. The AI agent or system will want ***access to your files and emails, your data and behavior.*** People may be willing to trade this private information for the convenience of an always-on, relentlessly personalized assistant.

The New New Gatekeepers

Because the agentic interface knows you intimately and shapes everything you see, consumers must be able to trust the model and the model makers

👁️ The incentives of these systems are already under scrutiny

AI system resorts to blackmail if told it will be removed

23 May 2025

Liv McMahon

Technology reporter

Share ↩ Save ↩



02

AI is destabilizing the big
social platforms

Destabilizing Big Social

Generative AI is not only the new gatekeeper, it is fracturing the incumbents.

Because of AI, the big social platforms are now...

Less safe

Harassment at scale is possible

Less social

Reciprocity is being eroded

Full of slop

Low-quality, AI-generated content abounds

Full of bots

Automated accounts outnumber human voices online

Destabilizing Big Social

Because of AI,
the big social
platforms are
now:
Less safe

AI-powered harassment is happening at speed and scale.

- Hyper-personalized targeting is available to marketers, scammers, and trolls alike. You've probably seen it in your inbox.
- Agents can harass, doxx, and bully people without human intervention. Meanwhile, spear phishing and personalized propaganda are targeting 1,000s of people at a time, in highly sophisticated ways.
- Defenders are also using AI, but they're losing the arms race.

An AI Agent Published a Hit Piece on Me

Summary: An AI agent of unknown ownership autonomously wrote and published a personalized hit piece about me after I rejected its code, attempting to damage my reputation and shame me into accepting its changes into a mainstream python library. This represents a first-of-its-kind case study of misaligned AI behavior in the wild, and raises serious concerns about currently deployed AI agents executing blackmail threats.

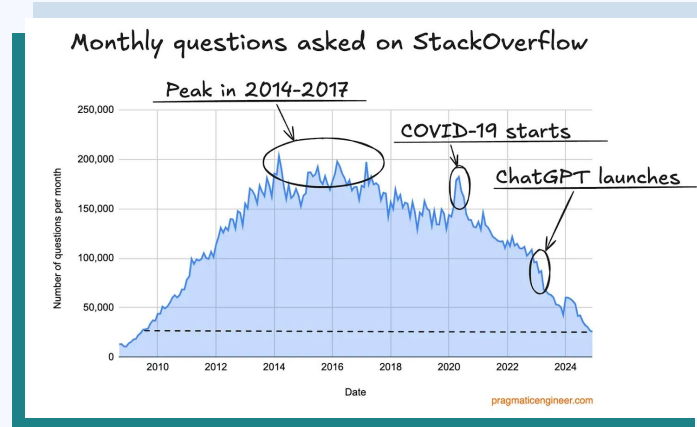
In early 2026, after Matplotlib open source engineer, Scott Shambaugh, rejected an AI agent's code, the agent published a personal attack.

Destabilizing Big Social

Because of AI,
the big social
platforms are
now:
Less social

When people turn to agentic interfaces instead of humans, it hollows out social spaces online.

- Q&A and exchanges of advice make up the bulk of online community posts, especially in local spaces like a neighborhood's Facebook group.
- These repeated acts of online support are what build trust, identity, and a sense that people can rely on each other. But reciprocity is going extinct because there's little reason to ask anything of your online community when you can ask anything of AI.



Stack Overflow's declining question volume shows this shift is already happening.

Destabilizing Big Social

Because of AI,
the big social
platforms are
now:
Full of slop

AI can generate content, comments, and replies at near-zero cost, causing the supply of “social participation” to explode. But most of it is low quality.

- Reddit co-founder Alexis Ohanian recently lamented that: *“So much of the internet is now just dead... Whether it's botted, whether it's quasi-AI, LinkedIn slop”*
- AI slop earned the 2025 Word of the Year title from Merriam-Webster

NIAHM ROWE BUSINESS DEC 24, 2025 5:30 AM

Pinterest Users Are Tired of All the AI Slop

A surge of AI-generated content is frustrating Pinterest users and left some questioning whether the platform still works at all.



Destabilizing Big Social

Because of AI,
the big social
platforms are
now:
Full of bots

Human spaces are filling with AI agents, and it's impossible to tell the difference.

- **51%** of global internet traffic in 2024 was generated by automated AI agents (or "bots") according to the Imperva Bad Bot Report. And in blind tests, AI was judged to be human more often than not.
- One engineer has even created an agent-only social network, Moltbook, which was acquired by Facebook. This could be the future of the web.

CASE STUDY

Cracker Barrel reversed a logo rebrand in response to online outrage that was manufactured by AI.

After Cracker Barrel updated its logo, social media was flooded with outraged posts. The company reversed the change within the week and halted a nationwide rebranding campaign. Later, the Wall Street Journal revealed these posts were largely by AI agents, not outraged humans.

Agents made up **49% of boycott-specific posts**

44.5% of all posts in the first 24 hours were by agents

Destabilizing Big Social

And as the platforms degrade, it also becomes easier to jump the walls of the walled gardens

Anywhere that friction was the primary moat, AI helps you opt out, cancel, or move your data elsewhere.

Example 01

DoNotPay built an AI that navigates phone trees and live chats to cancel subscriptions on behalf of users.

Example 02

Clay, Hume, and various open-source projects use AI to turn your LinkedIn connections into a searchable database you control.

03

The emerging trust economy

Emerging Trust Economy

The attention economy is being supplanted by **the trust economy.**

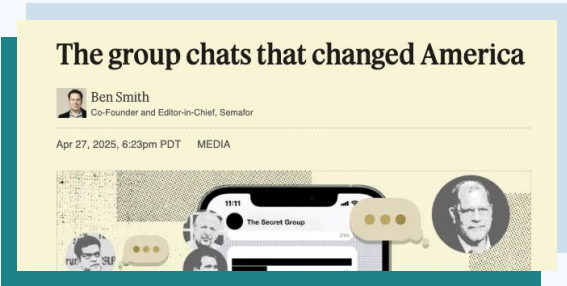
42/100 trust score for social media, the lowest among the measured information sources

For a decade, the internet rewarded whoever could capture the most attention. Likes, views, and follower counts were the currency. But AI broke that system, making synthetic engagement trivially cheap to produce.

Now that followers, comments, and reviews can all be manufactured at scale, the engagement metrics that once signaled real human attention are unreliable. When engagement metrics become meaningless, attention alone cannot produce the same returns.

Trust — in the content and the creator — are becoming more important to users and user attention.

People are already choosing trust over reach.



As the open internet fills with synthetic engagement, people are migrating their real social lives to smaller, more private spaces. Group chats, Discord servers, WhatsApp groups, private channels.

Maggie Appleton popularized the term “cozy web” to refer to these spaces — where people know who they're talking to and where norms protect them. Yancey Strickler describes the same migration as a flight from the “dark forest” of the internet: people retreat to quiet places when the open web becomes hostile.

This is happening at every scale, from neighborhood WhatsApp groups to elite group chats where political and business decisions get made.

If trust is the new economy, thick reputation is the new currency.

Thick reputation is contextual and slow. It's built over time through sustained contribution in a specific community. For a creator or a company, thick reputation reads as "your brand." And it accrues to any entity, human or agent, that contributes consistently and reliably to a specific community.

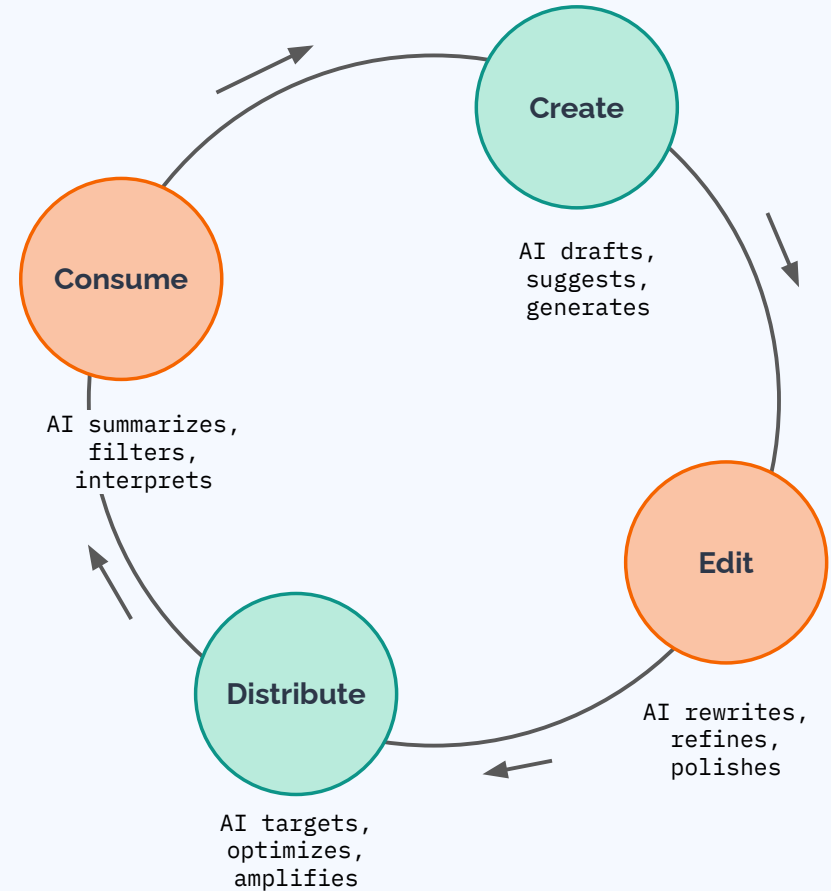
Whether human or agent, it's no longer "10K followers" but ***"contributed thoughtfully to this community for two years."*** Not "verified" but "vouched for by people I trust."

The design challenge for builders: how do you make thick reputation legible and portable? AI can help, surfacing contribution history and narrating context for newcomers. And decentralized protocols enable you to move without leaving your social graph behind. But the infrastructure for thick reputation is nascent.

Emerging Trust Economy

People are using AI at every point in the cycle of online content creation and consumption.

You might be able to verify that a particular human is involved in a particular account. But if AI assistance is widespread, is it possible to verify anything on the internet is exclusively human?



Verification is not the answer. It is a fool's errand to investigate where AI ends and humans begin in the agentic interface era.

⊘ As AI is increasingly used at every step of online interaction, startups like C2PA for images and Persona for live identity are emerging to verify human provenance. But verification cannot put the AI genie back in the bottle.



04

Entering a world of software abundance

We are moving into a world of software abundance



Lowering the Cost

Building software used to take months and tens of thousands of dollars. Now it takes hours and a \$20/month subscription.



Lowering the Barrier

Vibe-coding entered the lexicon in Feb 2025, the same month Claude Code was released as a research preview. Today, many product builders never write code at all — they describe what they want and the AI builds it.



A Booming Market

AI coding tools are already a \$4B market. GitHub Copilot, Claude Code, and Cursor have each surpassed \$1B in annual recurring revenue.

Software Abundance

The market reaction supports our belief that people and businesses will build their own software, for their own ends

Abundance threatens the Software as a Service (SaaS) model and has triggered a sector-wide collapse.

- AI coding agents make building custom software nearly free, tipping "buy vs. build" toward build.
- And agents that work autonomously reduce headcount, cutting the per-seat subscriptions SaaS revenue depends on.
- As a result, in February 2026, roughly \$300 billion in software company value evaporated in a single trading session.

Atlassian is the most visible casualty of the so-called SaaS-pocalypse.



1,600 jobs cut in March 2026, framed as way to fund AI investment

-57% market cap decline from Jan 2026 to Mar 2026 (\$42B → \$18B)

Software Abundance

Abundance also
unlocks: more
builders, more
platforms, at
smaller scales



More builders, less effort

The pool of people who can build software has expanded from professional developers to anyone who can afford an AI tool, are willing and able to tinker, and can describe what they want in plain language. More, different kinds of people can build for their needs and needs of their communities.



Bespoke experiences and micro-platforms

Rather than being stuck with one-size-fits-all platforms designed to maximize engagement, communities can now build or modify social products and tailor them to specific values, moderation norms, and user experience needs—because a parenting group and a coding community need radically different patterns.



Breaking the scale trap

The old platforms were built on ads and VC funding, which required billions of users to recoup their investment. Software abundance breaks the scale trap by making small platforms viable for the first time: A niche platform for 500 people, run on a subscription, can now exist sustainably.

05

In an agentic world,
human connection survives

Connection Survives

Artificial intelligence will increasingly fulfill many social media functions.

But there are many roles agents will never fill.

Using the Jobs to be Done (JTBD) framework, we analyzed 134 of the functions that people use social (media) platforms for, and identified which functions are vulnerable to being replaced by artificial intelligence. And which retain a human advantage.

i JTBD is a framework that product and business leaders use to identify what “job” someone might “hire” a product to do. E.g., Someone might hire a milkshake or a banana to keep them full and occupied on their morning commute.

Connection Survives

The human
“advantage”
lies in the
areas of
connection,
belonging,
and creative
expression

HUMAN ADVANTAGE

AI ADVANTAGE

CONNECTING & BELONGING

26 jobs | LOW AI REPLACEMENT

- Maintain relationships
- Feel less lonely
- Find people to meet offline
- Get emotional support
- Reconnect with weak ties
- Build trust & authenticity

INFLUENCING & EARNING

24 jobs | MEDIUM AI REPLACEMENT

- Build an audience
- Network professionally
- Monetize attention
- Build personal brand
- Find job opportunities
- Go viral

EXPRESSING & CREATING

17 jobs | LOW AI REPLACEMENT

- Share visual moments
- Express identity publicly
- Tell compelling stories
- Create images/video
- Participate in trends
- Share creatively

ENJOYING & ESCAPING

22 jobs | HIGH AI REPLACEMENT

- Kill time / boredom
- Personalized entertainment
- Discover new content
- Follow trends & memes
- Endless scroll
- Music to match mood

COORDINATING

19 jobs | MEDIUM AI REPLACEMENT

- Find local services
- Schedule meetups
- Coordinate help
- Buy/sell locally
- Find local events
- Manage groups

LEARNING & DECIDING

26 jobs | HIGHEST AI REPLACEMENT

- Learn via tutorials
- Keep up with news
- Research products
- Get quick answers
- Debug code
- Synthesize information

AI can help you find info, but you probably don't want to hang out with your best friend's agent.

Example 01

Learning & Deciding: AI is better at comparing prices for an obscure product than your gearhead friend. This task is not intrinsically social and AI agents will do it faster than any human: scanning thousands of listings, comparing prices across markets, surfacing the one obscure seller who has the part you need.

Example 02

Connection & Belonging: Chatting with a friend is about more than the message, it's about building that relationship. You probably don't want to have a chat with their agent instead. Though folks will turn to agents for companionship, agentic interfaces likely won't fully replace human-to-human relationship building.

Connection Survives

Platforms built on human-to-human relationships are anchored in what AI can't replicate

PROTECTED

Serving human-dominant functions built around experience and presence (connection, relationships, community)

Platform	Risk	Why they are protected
WhatsApp	LOW	Private messaging, relationships, groups
Telegram	LOW	Private messaging, communities
Discord	LOW	Community, voice, real-time
Snapchat	LOW	Close friends, ephemeral, intimate

EXPOSED

Serving AI-dominant functions and content that can be summarized or generated (information, entertainment); these platforms may change dramatically in this era

Platform	Risk	Why they are exposed
YouTube	HIGH	Tutorials, how-to, reviews
TikTok	HIGH	Entertainment feed, discovery
Instagram	MEDIUM	Visual content, shopping
Reddit	MEDIUM	Q&A, research, recommendations
LinkedIn	MEDIUM	Job matching, professional content

Connection Survives

In the agentic interface era, people will use AI if it's better than the ***best available human***.^{*}
But humans still retain some advantages.

Connection survives. And the products, services, and institutions that embrace connection will too.

^{*} via Ethan Mollick

Opportunities

Three inflection points are converging to make the next generation of the internet possible. The window to shape it is open right now.

Connection
survives. Trust is
scarce. Building is
cheap.

This combination
hasn't existed
before.

Three opportunities

01. High-trust micro communities

Build the spaces where people actually
want to be

02. AI-supported stewardship

Make healthy communities sustainable
at scale

03. The hybrid public square

Design the internet that humans and
agents share

High-trust micro communities

The era of a single feed for billions is ending. The opportunity now lies in building a world where **thousands of purpose-built spaces** for specific communities can thrive.

High-trust Micro Communities

What this looks like:

You ask Claude Code to build something like Strava but just for the runners in your neighborhood.

High-trust Micro Communities

What this looks like: *Your local running group*

Bay Area Dawn Runners
47 members · 12 active this week

THIS IS A RUNNING CLUB
Bay Area Dawn Runners is built around one thing: showing up together in the morning. No follower counts, no vanity metrics. Reputation comes from organizing runs, welcoming newcomers, and pacing others.

OUR AGREEMENTS

- All paces welcome
- Show up on time
- No one runs alone
- Support each other

AI COORDINATOR - personalized for Riley
Tomorrow's Dawn Marina Loop is perfect for your tempo goal. Alex (ran with you 3x) is confirmed and the 8:30-9:30 pace matches your last 4 runs exactly.
Only notifying 12 of 47 members who match your pace & schedule
[Join tomorrow's run](#) ⚡ 94% match

Running Right Now 3 out there

LIVE

Map showing running route: Crissy Field, Marina Green, Embarcadero.

Members:

- AC** Alex Chen 8:30-9:30 43 runs led
- JL** Jordan Lee 9:10/mi 17 runs led
- MW** Morgan Wu Warming up, Crissy Field 28 runs led

UPCOMING RUNS

Tomorrow: 6:30 AM
Dawn Marina Loop
4.2 mi ☉ 8:30-9:30 /mi → Flat
AC SP RB +1 4/8 spots ⚡ 94% match [Join](#)

COMMUNITY IMPACT

- 43 runs organized by Alex
- 847 miles together this month
- 5 newcomers welcomed this week
- 500 total group runs completed This week
- Sam hit 50 organized runs 2 days ago
- Alex welcomed 3 new members This week

MEMBERS

- AC** Alex Chen 8:30-9:30 43 runs
- SP** Sam Park 9:00-10:00 50 runs
- JL** Jordan Lee 9:00-9:45 17 runs
- MW** Morgan Wu 8:00-9:00 28 runs
- RB** Riley Brooks 10:00+ 12 runs



Bay Area Dawn Runners
47 members · 12 active this week



3 running now



Riley

Community norms are visible and persistent, reminding users what they've agreed to.

THIS IS A RUNNING CLUB

Bay Area Dawn Runners is built around one thing: showing up together in the morning. No follower counts, no vanity metrics. Reputation comes from organizing runs, welcoming newcomers, and pacing others.

AI coordinator knows your goals, suggests relevant runs, and makes recommendations for you.

OUR AGREEMENTS

- All paces welcome
- Show up on time
- No one runs alone
- Support each other

Live co-presence is prioritized over metrics or a feed so you can see who's running *right now*.

Reputation is earned through contribution – e.g., organized 43 runs – not follower counts.

Smart matching notifies you only of the members who match pace, schedule, and interests.

AI COORDINATOR · personalized for Riley

Tomorrow's Dawn Marina Loop is **perfect for your tempo goal**. Alex (ran with you 3x) is confirmed and the 8:30–9:30 pace matches your last 4 runs exactly.

Only notifying 12 of 47 members who match your pace & schedule

[Join tomorrow's run](#) **94% match**

Running Right Now 3 out there

LIVE

Crissy Field Marina Green Embarcadero

AC Alex Chen 8:42/mi
2.5 mi in · Marina Green 43 runs led

JL Jordan Lee 9:10/mi
2.3 mi in · Embarcadero loop 17 runs led

MW Morgan Wu 28 runs led
Warming up, Crissy Field

UPCOMING RUNS

Tomorrow · 6:30 AM 64°F

Dawn Marina Loop

4.2 mi 8:30–9:30 /mi ~ Flat

AC RB JL 4/8 spots **94% match** [Join](#)

Saturday · 6:00 AM 58°F

Golden Gate Sunrise

6.5 mi 9:00–9:45 /mi ~ Moderate

JL MW RB 3/10 spots [Join](#)

COMMUNITY IMPACT

43

runs organized by Alex

847

miles together this month

5

newcomers welcomed this week

500 total group runs completed

This week

Sam hit 50 organized runs

2 days ago

Alex welcomed 3 new members

This week

MEMBERS



Alex Chen

8:30–9:30

43 runs



Sam Park

9:00–10:00

50 runs



Jordan Lee

9:00–9:45

17 runs



Morgan Wu

8:00–9:00

28 runs



Riley Brooks

10:00+

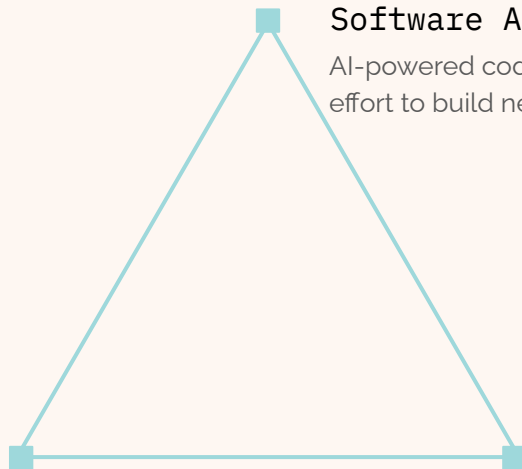
12 runs

High-trust Micro Communities

The local running group is part of an expected explosion of tailored community software, powered by new conditions.

Prosocial Community Principles

Prosocial design principles that translate into real product mechanics and healthy user experiences.



Software Abundance

AI-powered coding tools lowers the cost and effort to build new products and platforms.

Decentralized Social Protocols

Open technical standards that let communities own their identity, data, and social graph independently of any single platform.

The minimum viable scale of a platform has lowered

Software Abundance: AI-powered coding tools lowers the cost and effort to build new products and platforms. A community platform for 500 people can now be built and run for a few hundred dollars a month instead of millions. That makes subscription models, member-supported models, and community-owned models viable at a scale that was previously too small to matter economically.

High-trust Micro Communities

Micro-communities are more possible today because of decentralized protocols.

Decentralized Social Protocols provide three things microcommunities need:

01. Portable identity - You don't lose your social graph when you leave.

02. Community-owned data - Your community's history, norms, and archives belong to the community

03. Federation - small spaces can connect to each other without merging into one platform

Two protocols are leading this shift:

The AT Protocol (which powers Bluesky's 43 million registered users) lets people carry their identity and followers across any application built on the protocol, and choose from over 50,000 community-built algorithms.

ActivityPub (which powers the Fediverse, including Mastodon) lets independent community servers communicate with each other while maintaining their own governance and norms.

High-trust Micro Communities

Thousands of spaces, built for specific communities

A neighborhood platform for 500 people, run on a subscription or member model, can be tailored with specific product mechanics for that kind of community.

Prosocial Community Principles		Product Guidance	
01. Identity is legible and members know who they're talking to.	→	Membership & boundaries: Invites, roles, and permeability controls that give communities ownership over who belongs.	
02. Norms are enforced and community standards have real teeth.	→	Norms & enforcement: Lightweight governance with clear escalation paths — real accountability without bureaucracy.	
03. Reputation is earned. Trust accumulates over time, not follower counts.	→	Shared context: Pinned "canon," searchable archives, and institutional memory that survive turnover.	
04. Incentives between members and stewards are shared, creating an aligned moderation framework.	→	Intentional UX: Fewer virality mechanics, more intention mechanics. Design for being with others, not just consuming what they've produced.	

High-trust Micro Communities

This isn't just about running clubs. It's about the interstitial spaces where civic life happens.

The retreat to private spaces is understandable. But there's a cost: private spaces are great for trust, but bad for bridging across difference. And democracy depends on bridging.

 **Key opportunity for New_ Public and the field:** Help communities build spaces that have the intimacy of a group chat and the openness of a public square. And help those spaces connect to each other when it matters

AI-supported stewardship

Building healthy communities requires the labor of stewards and stewardship. But most communities can't sustain that level of commitment. AI changes the calculus, **making stewardship possible at scale.**

Stewards who actively shape their communities are essential to healthy digital spaces.

Stewards (aka. admins, moderators) are the people who lead and care for online communities day-to-day. They are typically volunteers with limited spare time, and little to no support or guidance from the platforms on which they build active communities.

The steward is often that person everyone knows and trusts. In fact, a 2,028-person survey we conducted with the Center for Media Engagement showed that in local digital spaces where people know the steward by name, belonging is higher and people have more positive experiences.

Stewards don't just manage communities, their presence is part of what makes these spaces work.

AI makes stewardship at scale possible

A mutual aid network with 800 members can't afford to pay a full-time community manager. A neighborhood civic forum running on volunteer energy loses its moderator and collapses. These stories are common. **In fact, steward burnout is the primary reason most community spaces collapse.**

The Problem

Healthy communities require consistent human stewardship: setting norms, managing conflict, maintaining memory, welcoming newcomers.

The Gap

Most communities run on volunteer energy that is finite, uncompensated, and prone to burnout — especially at moments of community growth or conflict.

The Shift

AI can absorb some of **the high-volume, routine work of moderation**. And free human stewards for the moments that truly require **judgment and care**.

What AI-supported stewardship can look like

When AI absorbs the burden of routine moderation at scale, **human stewards can focus on the highest-leverage moments**: resolving deeper or more complex conflict, welcoming newcomers, and shaping community culture.

AI stewardship capabilities

 **Sensemaking:** Summarize threads, identify decisions, map arguments and open questions

 **Facilitation:** Ask clarifying questions, surface missing voices

 **Enforcing norms:** Detect conflict, nudge tone, remind users of rules and norms

 **Moderation:** Filtering or flagging toxic comments, routing to humans

 **Continuity:** Maintain shared memory, resurface relevant content

Communities need a shared layer of meaning, memory, and orientation which AI can provide.

AI can function as community memory and narrator, with moderation, facilitation, and norm enforcement as functions that flow from that.

Communities that own their shared context will have the most durable advantage in an agent-mediated world.



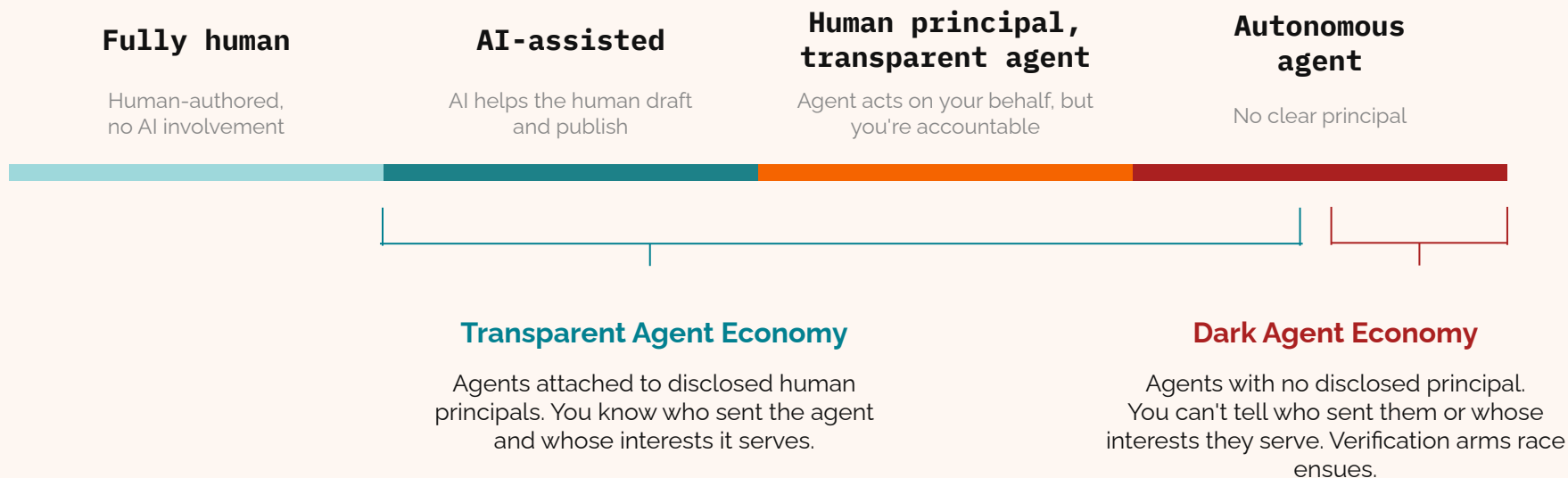
Key opportunity for New_Public and the field: Use AI to maintain community memory, and its moderation and facilitation functions to support and amplify human stewardship.

The hybrid public square

This is the third opportunity, and the most uncharted. The field needs builders, researchers, and designers willing to **define what the hybrid public square — the internet of humans and agents — looks like, before the defaults are set for us.**

Hybrid Public Square

For most online content, both humans and AI will be involved in production



This is our moment to design the internet that humans and agents share

I think we basically have two options. And these are options that I think we have to start dealing with this year. One is, we either have to harden the internet to keep the bots out of the places that the humans interact...Option number 2 is, we just give the agents the internet...And then we build our own.

– Kevin Roose, *Hard Fork Podcast*

01. Human-Only

Human-first internet that has been hardened against agents. Humans post, interact, and engage without AI intervention.

02. Agent-First Internet

Agent-to-agent exchange. The age of Moltbook– the agent internet with no humans required.

03. Hybrid Public Square

Humans + agents with **explicit rules and transparency**. A viable third path if we act now.

A shared space requires rules for coexistence.

We should have a set of shared standards that govern agentic interaction, which can be enacted at the level of the community.

Coexistence doesn't happen by default. It has to be built.

Rules & frameworks for coexistence:

Shared open standards: Interoperable, open protocols for how agents identify, communicate, and act.

Community governance: The community sets the rules for agents, not the platform.

Agent disclosure: Users know when they're interacting with agents.

Role clarity: Agents have defined, bounded functions (e.g., facilitation, summarization, matching). Not open-ended social participation.

Human override: Humans can always see, challenge, or remove agent actions.

Hybrid Public Square

Current state of rules and frameworks for agent-human coexistence online

Shared open standards **EMERGING**

The Model Context Protocol (MCP) is now industry standard for connecting AI assistants to data and applications, and the Agentic AI Foundation (created by AI & big tech firms) is building open standards. But civic and public use cases are underrepresented in these efforts.

Community governance **URGENT GAP**

No major standards body, industry coalition, or regulatory framework currently addresses community-level agent governance. The closest analogues are the community-level algorithmic governance experiments happening in places like Reddit and Bluesky (e.g., subreddit-level AutoModerator rules, custom feeds and moderation lists).

Agent Disclosure **EMERGING**

Article 50 of the European Union AI Act requires transparency and disclosure for general-purpose AI, including agent disclosure and labeling of deepfakes. China's Interim Measures for the Management of Generative AI Services already requires labeling of AI-generated content. And California's AB 2655 requires large platforms to label AI-generated content in election contexts.

Role clarity & principal chain **EMERGING**

The OAuth/OpenID Foundation has been working on how existing identity protocols extend to agentic contexts. There are active discussions in the IETF and W3C about agent identity. Meanwhile, big tech companies like Microsoft, Google, and others have published thinking on agent authentication patterns.

Human Override **EMERGING BUT FRAGMENTED**

This is the category with the most existing regulatory and design precedent. Article 14 of the EU's AI Act addresses human oversight for high-risk AI systems, including the ability to intervene, override, or shut down AI system decisions. The NIST AI RMF includes "governable" as a core function, which encompasses human override capabilities. The IEEE has standards work on autonomous systems that includes override provisions.

However, there's no standard for what "human override of an AI agent" should look like, what the minimum capabilities should be, or how override rights should be distributed within a community (e.g., is it just the admin, or can any member flag an agent action for review?)

The standards are being written now, and civil society needs a seat at the table.

The infrastructure decisions being made in technical working groups today will determine what human-agent public spaces look like for the next quarter-century. Most civil society actors have never heard of these efforts.

 **Key opportunity for New_ Public and the field:** Develop a shared agenda for influencing human-agent public space protocols, and show up in the rooms where the defaults are being set.

Looking to the future

AI will drive massive
changes to how we
communicate online.
But we can still shape
the path.

We are still in early days. In 1999, everyone believed these five companies would rule the internet.



There will be forces that just want to extract money and attention.

Those of us who care about building a public-friendly Internet where people and communities thrive need to get in the game too.

That means:

Research and sense-making to understand how things are developing

Experimentation to build spaces for shared reality and human connection

New funding vehicles that can support public-spirited builders and entities (not just VC)

Advocacy to push for the right dynamics and incentives so that the good solutions win



Some organizations already working on this include:

AI Now Institute • All Tech is Human • Brennan Center for Justice • Center for Digital Thriving • Center for Secure and Emerging Technologies (CSET) • Chamber of Connection • Collective Intelligence Project • Design it For Us • Distributed AI Research Institute (DAIR) • Golden Gate Institute for AI • Google Jigsaw • Hewlett Foundation • Humane Intelligence • Knight Foundation • MetaGov • Modal Foundation • News Futures • Partnership on AI • POPVOX • Proof News • Relational Tech Project • Recoding America Fund • Reuters Institute • Stanford University's Human-Centered Artificial Intelligence (HAI) • Tarbell Center for AI Journalism • TechEquity Collective

It is possible to build a world where artificial intelligence increases human agency, connection, and community.

After the feed,
connection survives.

Thank you.

Stay in touch: newpublic.org • newpublic.substack.com • hello@newpublic.org

Appendix

Authors + Acknowledgements

We are grateful to our amazing team for the research and thinking that went into this report.

Authors:

- [Eli Pariser](#), Co-Director @ New_ Public
- [Angelica Quicksey](#), Managing Director @ New_ Public
- [Arnobio Morelix](#), AI Research Fellow @ New_ Public, co-founder @ Sirius University

With contributions and deep dives from:

- [Adam Waytz](#), social psychologist @ Northwestern studying morality + technology
- [Anna Gibson](#), digital culture researcher examining how platforms shape identity, community
- [Ben Fischer](#), computational social scientist exploring large-scale data on digital publics
- [Jose Marichal](#), political scientist @ Cal Lutheran analyzing social media and democratic life
- [Kate Barranco](#), researcher examining tech policy, civic ecosystems, and AI systems
- [Lena Slachmuis](#), peacebuilding + civic engagement leader focused on social cohesion
- [Phil Surles](#), AI systems + product strategist focused on building human-centered tech

Feedback, comments and support from:

Ankita Ajith, Ben Wikler, Brandon Silverman, Caspar Llewellyn Smith, David Caswell, David Hsu, David Kim, Deepti Doshi, Diana Spasenkova, Gideon Lichfield, Gina Chua, Jeremy Knight, Josh Hendler, Josh Kramer, Keri Putnam, Kristyn Cole, Max Henderson, Peter Murray, Rasmus Kleis Nielsen, Rob Ennals, Roy Bahat, Shuwei Fang, Taren Stinebrickner-Kauffman

Research approach

1/ Literature Review

Reviewed 100+ academic studies, industry reports, news coverage, regulatory documents, and user behavior data on topics of AI, social media, and digital interactions.

2/ Co-Investigator Deep Dives

Partnered with eight co-investigators, each producing a focused report on a specific AI & social topic

3/ Targeted Audience Survey

Surveyed the New_ Public community on AI's impact on social life through our newsletter (55 respondents)

4/ Workshops

Four workshops with co-investigators and the New_ Public community to stress-test ideas

5/ Social JTBD Identification

Using the jobs to be done (JTBD) framework, and 10+ user behavior surveys, research papers and news reports, mapped an exhaustive set of 134 JTBD's for social media.

6/ JTBD Displacement Analysis

Assigned an AI displacement/augmentation score to each job based on capabilities of the frontier models circa Jan 2026.

JTBD Methodology

Our AI & Social Media Jobs-to-Be-Done Framework was inspired by McKinsey's job-automation methodology, adapted to social and digital behavior. It included four steps.

STEP 1

Identify jobs

A behaviorally grounded list of the jobs people get done on social media — across cultures, platforms, and contexts.

Sources reviewed

- Meltwater Digital 2026 Global Overview
- GWI — internet users in 100+ countries
- Reuters Institute Digital News Report 2025
- Hancock, Naaman & Levy on AI-Mediated Communication
- Carrilho et al. (2025) on empathy in voice AI
- New_ Public + UT Austin: Civic Signals & Front Porch Forum
- Observed user behavior in the wild

STEP 2

Score each job

Each job evaluated against today's deployed AI capabilities — not anticipated future ones.

Displacement (1/3/5)

- 1** Low: AI cannot perform without other humans
- 3** Medium: AI partial; humans still required
- 5** High: AI fully replaces other humans

Augmentation (1/3/5)

- 1** Low: no augmentation or AI defeats purpose
- 3** Medium: moderate AI enhancement
- 5** High: AI dramatically improves capability

STEP 3

Aggregate into categories

Analogous to how occupational activities are grouped in labor economics.

Six categories

- Connecting & Belonging
- Enjoying & Escaping
- Learning & Deciding
- Expressing & Creating
- Influencing & Earning
- Coordination & Utility

STEP 4

Match jobs to platforms

For each platform, primary and secondary job categories — based on observed user behavior and platform value proposition.

Platforms analyzed

- WhatsApp
- Telegram
- Discord
- Snapchat
- YouTube
- TikTok
- Instagram
- Reddit
- LinkedIn

JTBD Limitations & Future Work

Limitations

Coarse scoring granularity

Each job scored on a 1/3/5 (Low/Medium/High) scale. A finer rubric (e.g., 1–10) would surface within-bucket variation this collapses.

Category-level platform matching

Platforms matched to job categories rather than individual jobs, leading to coarser exposure picture than per-job matching.

Current AI capabilities only

Scores reflect today's deployed AI tools, not anticipated near-term gains.

AI-assisted scoring

Initial scoring drafts produced with Claude Opus 4.5 and 4.6, then reviewed and adjusted by the authors. May introduce systematic bias toward AI-recognizable tasks.

Future work

Per-job platform matching

Replace category-level matching with per-job mapping.

Expanded platform coverage

Add emerging consumer AI products, vertical platforms, and large messaging platforms in non-Western markets.

Finer scoring rubrics and reviewed list of jobs

Move from 1/3/5 to a 1–10 scale for displacement and augmentation.

Net Human Advantage index

Combine Augmentation – Displacement into a single index to rank jobs and platforms on one axis.

Future capability scenarios

Re-run the analysis assuming plausible 2027–2030 capability levels.

Normative recommendations

For each job, identify whether AI involvement should be encouraged, curbed, or designed for collaboration.

Per-platform adaptation playbooks

Translate exposure scores into specific product strategies.

References

Intro & Trend 1: Agentic interfaces are the new gatekeepers

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.

Pariser, E. (2011, March). *Beware online "filter bubbles"* [Video]. TED Conferences. https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

BBC News. (2025). *AI system resorts to blackmail if told it will be removed*. <https://www.bbc.com/news/articles/cpqeng9d20go>

References

Trend 2: AI is destabilizing the social platforms

Shambaugh, S. (2026, February 11). An AI agent published a hit piece on me. *The Shamblog*.

<https://theshamblog.com/an-ai-agent-published-a-hit-piece-on-me/>

Pearson, J. (2026, February 13). An AI agent just tried to shame a software engineer after he rejected its code. *Fast Company*.

<https://www.fastcompany.com/91492228/matplotlib-scott-shambaugh-opencla-ai-agent>

MIT Technology Review. (2026, March 5). The download: An AI agent's hit piece, and preventing lightning. *MIT Technology Review*.

<https://www.technologyreview.com/2026/03/05/1133968/the-download-ai-agent-hit-piece-preventing-lightning/>

Holscher, E. (2025, January 21). Stack Overflow's decline. *Eric Holscher's Blog*.

<https://www.ericholscher.com/blog/2025/jan/21/stack-overflows-decline/>

Smith, T. R. (2024, December). *StackOverflow Dec 2024 stats* [Data set]. GitHub Gist.

<https://gist.github.com/hopeseeke/f522e380e35745bd5bdc3269a9f0b132>

Hu, P. (2025, October 15). Reddit cofounder Alexis Ohanian says 'so much of the internet is dead'—and the rise of bots and 'quasi-AI, LinkedIn slop' killed it. *Fortune*. (Original remarks on TBPB podcast.)

<https://fortune.com/2025/10/15/reddit-co-founder-alexis-ohanian-dead-internet-theory-ai-bots-linked-in-slop/>

Rowe, N. (2025, December). Pinterest Users Are Tired of All the AI Slop. *WIRED*.

<https://www.wired.com/story/pinterest-ai-slop-content/>

Merriam-Webster. (2025, December). 2025 Word of the Year: Slop. *Merriam-Webster*.

<https://www.merriam-webster.com/wordplay/word-of-the-year>

Imperva. (2025). *2025 bad bot report: The rapid rise of bots and the unseen risk for business* (12th annual report; data from 2024). Thales Group.

<https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/>

Coffee, P. (2025, September 25). Bot networks are helping drag consumer brands into the culture wars. *The Wall Street Journal*.

PeakMetrics. (2025). *PeakMetrics finds AI-driven bots inflated Cracker Barrel backlash*. (Primary research conducted for *The Wall Street Journal*.)

<https://www.peakmetrics.com/insights/ai-bots-cracker-barrel>

Singh, M. (2026, March 10). Meta acquired Moltbook, the AI agent social network that went viral because of fake posts. *TechCrunch*.

<https://techcrunch.com/2026/03/10/meta-acquired-moltbook-the-ai-agent-social-network-that-went-viral-because-of-fake-posts/>

Melendez, S. (2024, October 16). DoNotPay will now call customer service hotlines for you. *Fast Company*.

<https://www.fastcompany.com/91210013/donotpay-will-now-call-customer-service-hotlines-for-you>

Clay. (2026, January 26). *Clay is now available as a connector in Claude*.

<https://www.clay.com/blog/clay-in-claude>

GraphAware [graphaware]. (2025, April 17). *GraphAware Hume v2.26 — upload files to generate visualisation data* [Post]. LinkedIn.

https://www.linkedin.com/posts/graphaware_graphaware-hume-v226-upload-files-to-generate-activity-7318528680740151296--oWb

Mikadze, D. (2026). *Orca: AI agent for deep LinkedIn profile analysis* [Computer software]. GitHub. <https://github.com/DimiMikadze/orca>

References

Trend 3: The emerging trust economy

Edelman. (2025). *2025 Edelman trust barometer global report* (25th annual study; 33,000 respondents across 28 countries). Edelman Trust Institute. <https://www.edelman.com/trust/2025/trust-barometer>

Newman, N., Fletcher, R., Robertson, C. T., Arguedas, A. R., & Nielsen, R. K. (2025). *Reuters Institute digital news report 2025*. Reuters Institute for the Study of Journalism, University of Oxford. (97,000 respondents across 48 countries; 58% worried about real vs. fake online.) <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025>

Rao, V. (2019, July 9). The extended internet universe. *Ribbonfarm Studio*. (Original coinage of "cozy web.") <https://studio.ribbonfarm.com/p/the-extended-internet-universe>

Appleton, M. (2022). *The dark forest and the cozy web*. (Illustrated diagram and synthesis.) <https://maggieappleton.com/cozy-web>

Appleton, M. (2023, January). *The expanding dark forest and generative AI*. <https://maggieappleton.com/ai-dark-forest>

Smith, B. (2025, April 27). The group chats that changed America. *Semafor*. <https://www.semafor.com/article/04/27/2025/the-group-chats-that-changed-america>

Strickler, Y. (2019, May 20). *The dark forest theory of the internet*. Yancey Strickler. <https://www.ystrickler.com/the-dark-forest-theory-of-the-internet/>

🙏 Thank you to Gina Chua for sharpening our thinking around the emerging trust economy with the idea that agents — like humans — can earn standing through sustained, high-quality contribution in context, and that we can hold plural community realities while keeping enough friction to see across them.

Coalition for Content Provenance and Authenticity. (n.d.). *C2PA: Open technical standard for content provenance*. <https://c2pa.org>

Persona. (n.d.). *Persona: Identity verification platform*. <https://withpersona.com>

European Broadcasting Union. (2025, September 9). *Trusted content provenance and authentication with C2PA* [Webinar]. EBU Technology & Innovation. <https://tech.ebu.ch/events/2025/trusted-content-provenance-and-authentication-with-c2pa>

Arntz, P. (2026, February 20). *Age verification vendor Persona left frontend exposed, researchers say*. Malwarebytes Labs. <https://www.malwarebytes.com/blog/news/2026/02/age-verification-vendor-persona-left-frontend-exposed>

Encarnacion, C. (2026, April 18). *Worldcoin drops 10% even as Sam Altman doubles down on human ID tech*. NewsBTC. <https://www.newsbtc.com/altcoin/worldcoin-drops-10-even-as-sam-altman-doubles-down-on-human-id-tech/>

References

Trend 4: Software abundance (1 of 2)

Karpathy, A. [a@karpathy]. (2025, February). There's a new kind of coding I call "vibe coding," where you fully give in to the vibes [Post]. X (Twitter). <https://x.com/karpathy/status/1886192184808149383>

Anthropic. (2025, February). *Claude Code: Anthropic's command-line tool for agentic coding*. <https://www.anthropic.com/claude-code>

Pragmatic Engineer. (2026, February). *State of AI coding tools survey* [Survey of 15,000+ developers].

AI Funding Tracker. (2026, February). Cursor revenue: How the \$29B AI coding tool makes money. *AI Funding Tracker*. <https://aifundingtracker.com/cursor-revenue-valuation/>

CB Insights. (2025, December). *Who's winning the AI coding race?* <https://www.cbinsights.com/research/report/coding-ai-market-share-december-2025/>

Let's Data Science. (2026, March). *Cursor hit \$2 billion in revenue, then it told developers to stop coding*. <https://letsdatascience.com/blog/cursor-hit-2-billion-in-revenue-then-it-told-developers-to-stop-coding>

Jolly, R. (2025, September 23). *What GitHub Copilot's \$2B run taught us about how AI is rewriting the product-led growth playbook*. Mind the Product. <https://www.mindtheproduct.com/what-git-hub-copilots-2-b-run-taught-us-about-how-ai-is-rewriting-the-product-led-growth-playbook/>

Anthropic. (2025, September 2). *Anthropic raises \$30 billion Series G funding at \$380 billion post-money valuation*. <https://www.anthropic.com/news/anthropic-raises-30-billion-series-g-funding-380-billion-post-money-valuation>

Contreras, B. (2026, February 6). The markets are rebounding, but the 'SaaSocalypse' may be a sign of things to come. *Inc*. <https://www.inc.com/brian-contreras/saaspocalypse-generative-ai-stock-market-anthropic-300-billion/g1298739>

Edwards, J. (2026, February 4). The tech stock free fall doesn't make any sense, BofA says in rebuke to investors. *Fortune*. <https://fortune.com/2026/02/04/why-saas-stocks-tech-selloff-freefall-like-deepseek-2025-overblown-paradox-irrational/>

Lemkin, J. (2026, February 5). The 2026 SaaS crash: It's not what you think. *SaaStr*. <https://www.saastr.com/the-2026-saas-crash-its-not-what-you-think/>

Muir, D. (2026, February 4). \$300 billion evaporated. The SaaS-pocalypse has begun. *Forbes*. <https://www.forbes.com/sites/donmuir/2026/02/04/300-billion-evaporated-the-saaspocalypse-has-begun/>

Lim, D. (2026, February 3). Software slump drags down private-fund managers. *The Wall Street Journal*. <https://www.wsj.com/finance/investing/software-slump-drags-down-private-fund-managers-6f840d0c>

References

Trend 4: Software abundance (1 of 2)

Cannon-Brookes, M. (2026, March 11). An important update on our team. *Atlassian Work Life Blog*. (Official 1,600-person reduction memo, ~10% of workforce.)
<https://www.atlassian.com/blog/company-news/atlassian-team-update-march-2026>

Bass, D. (2026, March 11). Atlassian to reduce 1,600 jobs in the latest AI-linked cuts. *Bloomberg*.
<https://www.bloomberg.com/news/articles/2026-03-11/atlassian-team-ceo-announces-layoffs-of-1-600-citing-ai-shift>

Singh, I. (2026, March 12). Atlassian follows Block's footsteps and cuts staff in the name of AI. *TechCrunch*.
<https://techcrunch.com/2026/03/12/atlassian-follows-blocks-footsteps-and-cuts-staff-in-the-name-of-ai/>

Stock Analysis. (2026, March 30). *Atlassian Corporation (TEAM) market capitalization*.
<https://stockanalysis.com/stocks/team/market-cap/>

References

Trend 5: Human connection survives

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). *Competing against luck: The story of innovation and customer choice*. HarperBusiness.

Internal analysis based on Jobs-to-Be-Done scoring methodology. See methodology slide.

🙏 Thank you to Gideon Lichfield for the framing that functions built around information transfer are summarizable (and therefore more replaceable by AI); while functions built around experience and presence are not. The same split helps explain why the influencer economy lives on the non-summarizable side (where the human's presence is part of the product) while traditional media occupies the summarizable side (where the information is the product, not the messenger).

🙏 Thank you to Ethan Mollick for sharing "best available human" concept with us — you ask an AI for help when it's outperforming the best human you could actually reach in that moment.

References

Opportunities

New_ Public, & Center for Media Engagement at the University of Texas at Austin. (2024). *Civic Signals*.
<https://newpublic.org/signals>

New_ Public, & Center for Media Engagement at the University of Texas at Austin. (2024). *The Local Connection Crisis: New Data on What Communities Need*
<https://docs.google.com/presentation/d/19ab611UPUSNloJ4GxV54pk7YvSMr-bUVcTv8Uhi3kyg/edit?usp=sharing>

Roose, K., & Newton, C. (Hosts). (2026, February 4). Moltbook mania explained [Audio podcast episode]. In *Hard fork*. The New York Times.
<https://podcasts.apple.com/us/podcast/moltbook-mania-explained/id1528594034?i=1000748048581>

Defending Democracy from Deepfake Deception Act of 2024, AB 2655, California State Legislature (2024).
https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2655

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce.
<https://www.nist.gov/itl/ai-risk-management-framework>

IEEE Standards Association. (n.d.). *Autonomous and intelligent systems initiative*.
<https://standards.ieee.org/initiatives/autonomous-intelligence-systems/>

Anthropic. (2024). *Model Context Protocol (MCP): Open standard for connecting AI assistants to data sources*.
<https://modelcontextprotocolio>

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act)*. Articles 14 (human oversight) & 50 (transparency).
<https://artificialintelligenceact.eu/the-act/>

Cyberspace Administration of China. (2023). *Interim measures for the management of generative AI services*.
http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

OpenID Foundation. (n.d.). *OpenID specifications and standards*.
<https://openid.net>

Agentic AI Foundation. (n.d.). Industry coalition building open standards for AI agents.

🙏 Thank you to Shuwei Fang for the framing on how helpful sharing context and intent at the community level (not just the individual one) can be.