



The Collective Intelligence Project

2024: OUR YEAR IN REVIEW



The [Collective Intelligence Project](#) (CIP) is a nonprofit R&D lab building collective governance models for transformative technology.

CIP IN 2024

The race to build AI is outpacing our ability to direct it toward collective benefit. This decade will determine whether the public can shape and benefit from AI development or merely react to it, whether AI systems reflect plural values or standardize them, and whether the coming economic transformation benefits the many or empowers the few.

In [2023](#), CIP demonstrated a new approach: embedding collective intelligence directly into AI development through partnerships with frontier labs and governments. 2024 has been about scaling and adapting this model - building the infrastructure needed to someday make collective benefit from AI inevitable. 2024 saw the rapid release of increasingly powerful AI models. Claude 3, GPT-4o, and Gemini demonstrated capabilities that crossed new thresholds, from nearly human-level reasoning to sophisticated visual understanding. **We continue to believe that this is technology with global-scale social and economic impacts.**

The [state of progress is exciting](#). AI labs continue to invest in [figuring out](#) how to [technically align](#) individual models, [governments are establishing AI Safety Institutes](#) to advance the science of AI safety, countries are beginning to explore their own [sovereign AI strategies](#), and we see [green shoots towards global governance](#).

The future of AI is [multi-agentic](#)—the fantasy that aligning single models to ‘human values’ can ensure collective benefit is crumbling. What will replace it? Our thesis is that this requires new models of pluralistic alignment and collective input, and institutional innovation to steward the change. Now is the time to harness collective intelligence: collecting and building on participatory input that captures human diversity, crafting policy that isn’t focused on naively hoping for benevolent power concentration, and getting predistribution right before large-scale economic disruption. **Who gets to shape the future of AI is now a geopolitical issue; how is still an unsolved technical challenge.**



We are still in the interregnum. [Funders \(philanthropic and otherwise\)](#) are still [figuring out the highest](#) leverage areas to invest in. [AI applications are booming—but not yet the ones that will change the world.](#) [Shifts in economic reality are nascent](#), but beginning. [Nation-states](#) are [getting involved](#) in the race in a [very material way](#). Buoyed by these currents, the not-quite-research but not-quite-product work CIP and others like us do is coming to the fore.

And the work goes on. This year, we're excited about:

- **Building Global Dialogues**, a scalable system for hosting collective dialogues with people around the world on AI-related issues. Let's take global governance for the people seriously.
- **Expanding our Community Models platform** with partners around the world, which will allow any community to shape and steer their own AI models through local knowledge, principles, and stories. There is a midpoint between personalization and regression to the mean. That midpoint is community input.
- **Closing out our time at the UK AI Safety Institute** as founders of the Societal Impacts and the AI and Democracy teams. It is crucial to build up public sector capacity to build, evaluate, and shape AI.
- **Producing cutting-edge research** on the directing transformative technologies towards the collective good: spec'ing out proof-of-personhood credentials, building network societies (not network states), funding supermodular goods, predistributive approaches to AI, corporate governance of AI companies, and using AI to augment collective intelligence. There is so much we still don't know and haven't thought about.

Our work has been widely covered in the media, including [The New York Times](#), [TIME Magazine](#), [BBC](#), the [Brookings Institute](#), and [Science Magazine](#).

The work that matters addresses the core issues: navigating impending economic transformations, building public interest use cases for transformative technology, and fighting anti-human tendencies from all sides. In 2024, 2025, and forever, we're keeping our eye on the ball.



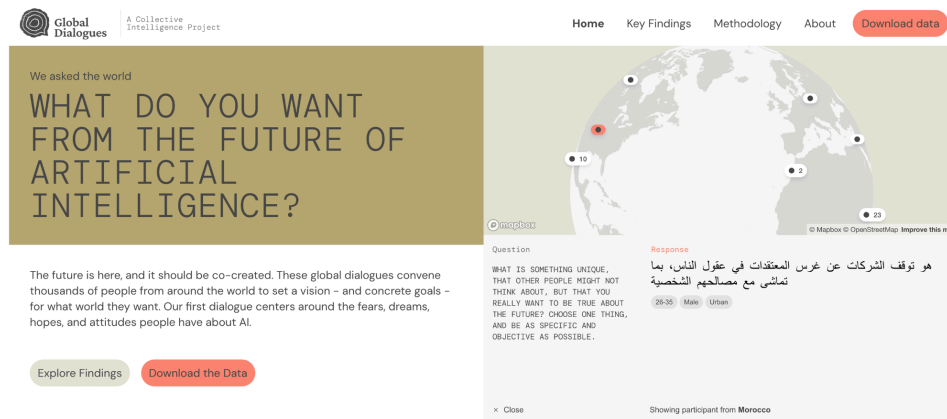
2024 HIGHLIGHTS

Our three core areas of work are **experimentation**, **research**, and **advocacy**. We want to think through what should be done, figure out if it works, and then implement it at institutional scale. This year, we did a lot of all three.

EXPERIMENTATION

GLOBAL DIALOGUES

In partnership with Remesh and Prolific, we've launched [Global AI Dialogues](#): a monthly public input process that engages participants worldwide in questions about positive-sum futures.



Why this project? First, we believe that it should be possible to gather meaningful public input at scale without the prohibitive costs of traditional methods. While in-person deliberative polls (\$1-2M) and global assemblies (\$10M+) offer depth, the goal here was to build a more agile, cost-effective approach that could be deployed on an ongoing basis. The first pilot was 10-100x cheaper and engaged 1,200 participants, in 68 countries and 8 languages, through forty questions spanning several modalities: scenarios, values, outcomes, and cultural preferences.



First, we see a global (75+% across countries) strong preference for international coordination over national fracturing in AI governance.

*“If I could make world leaders understand one thing, it would be that long-term sustainability and global cooperation are essential for addressing the biggest challenges humanity faces, like climate change, inequality, and technological ethics. **Focusing on short-term gains or individual national interests undermines our collective future.**”*

There are, however, differences in belief when it comes to cultural preservation and pluralism: upwards of a 15% shift between the Global North and the Global South.

*“My culture is collectivist in nature. AI will lead to a more individualistic world, **which might take out the essence of what it means to be human and reduce us.**”*

Another area of agreement: across demographics and regions people show a desire (75%+) to forgo significant benefits (including health advances and financial support) to prevent single-actor AI dominance.

*“If a country or government has full control over AI, it will be difficult for the public interest to be truly protected. When those who hold the highest power have full control over AI, there may be infringements on social and personal freedom. **I am worried that in this case, the world will become more chaotic, and the rights and voices of the public will be suppressed.**”*

Finally, there are areas of both fear and resignation.

*“I am afraid that AI will make decisions about the life and death of a human, the fact that an AI decides whether a child is born or not, or whether a person dies or not, **even if it is for the greater good.**”*



*I don't think it's fair to have emotional ties to AI, **but I don't think it's fair to restrict it either.***

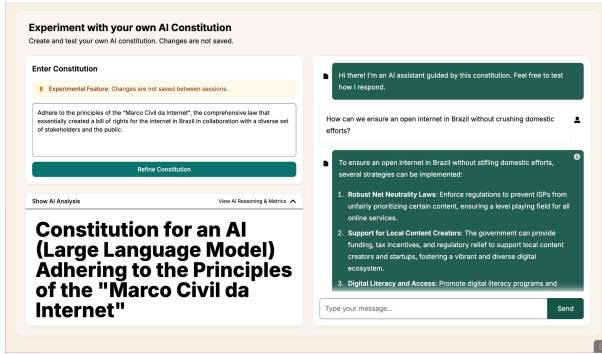
Our pilot runs demonstrated that scenario-based questions were particularly effective at helping participants navigate complex tradeoffs. Collecting nuanced public input at scale is both feasible and valuable, providing a foundational tool for ensuring diverse perspectives inform AI development and governance.

We're refining the scenario-based methodology, expanding our geographic and demographic representation, and developing frameworks to translate these insights into actionable governance recommendations. **These findings are being presented to AI labs, policymakers, and universities, with the intention that this form of global input can inform AI development and policy.** We've found that this has the potential to shift policy discussions from abstract principles to concrete public concerns.

COMMUNITY MODELS

Our Community Models platform has a simple goal: every community should be able to shape AI to serve their needs. Personalized AI, company AI, and nation-state AI will be solved by the market. But what about the collective steering of AI? What about archival AI? Community models make this possible.

This project was built on our learnings from CCAI with Anthropic. From that foundation, we've built an experimental platform for collective preference elicitation, deliberation, finetuning, and model alignment. The plan is to **build a community model library** for collective benefit.



So far, we're working on building Community Models to:

- Preserve and encode wisdom from elders around the world with the [Grandmothers' Collective](#)
- Ensure that models are tuned with Brazilian local context, in partnership with [ITS RIO](#)
- Contribute to AI alignment, research, and input in Africa with [Equiano Institute](#)
- Support public engagement and policy feedback in India with [Civis](#)
- Enable youth engagement, media, activism in India with [Youth Ki Awaaz](#)
- Create a safe and informed AI chatbot for formerly incarcerated persons re-entering society, with the [Second Life project](#)

Additionally, we are using the survey platform [Prolific](#) to launch surveys of representative samples of populations around the world to inform regionally and demographically bespoke Community Models.

RESEARCH

BUILDING DEMOCRATIC AI

We launched the [Roadmap for Democratic AI](#) to outline what could be done to shift the political economy of AI towards the public interest. There's a lot of discussion of democratization (who doesn't love democratization, in theory?) but very few practical steps. We assembled a comprehensive list of clear next steps that other organizations could take, starting from the immediately feasible and zooming out to a wider scope of



possibilities. **These may not solve every systemic problem, but we think each step is concrete, actionable, and important.**

CIP has begun to tackle some of the projects we listed: from enabling communities to fine-tune their own AI models, identifying points within the AI lifecycle to leverage collective input, and globally expanding the collective input mechanisms into AI.

The roadmap laid out practical steps to bring democratic input into every stage of AI development. We expanded and extended these arguments for Stanford's Digitalist Papers Project in an essay called A Vision for Democratic AI. This builds directly on what we've seen work - from our CCAI project with Anthropic to Taiwan's nationwide Alignment Assembly - and scales these successes into a comprehensive framework. It confronts the reality that democratizing access isn't enough - we need new institutions and infrastructure to ensure collective benefit. Democracy isn't just about giving everyone a voice. It's about doing something with the chorus.

We don't have to choose between technical excellence and collective governance.

They can, and must, strengthen each other. For more, listen to our Executive Director's TED Talk: ["AI and Democracy Can Fix Each Other"](#).

AI AND THE NEW ECONOMY

Technology affects the economy not primarily through first-order effects on democracy, but via reshaping markets, information, and power. Working on AI's economic foundations means building technology institutions, and incentives that better serve the public interest.

- In "[Predistribution over Redistribution](#)," we examine alternatives to **proposals like the Windfall Clause** - arguing that instead of just redistributing AI's benefits after the fact, we need to proactively build an inclusive economy from the ground up.
- We also focused on a main unit of our economy, the organization, in our paper "[Shared Code: Democratizing AI Companies](#)" with Nathan Schneider, in which **we explored how AI labs could democratize their own governance structures.**



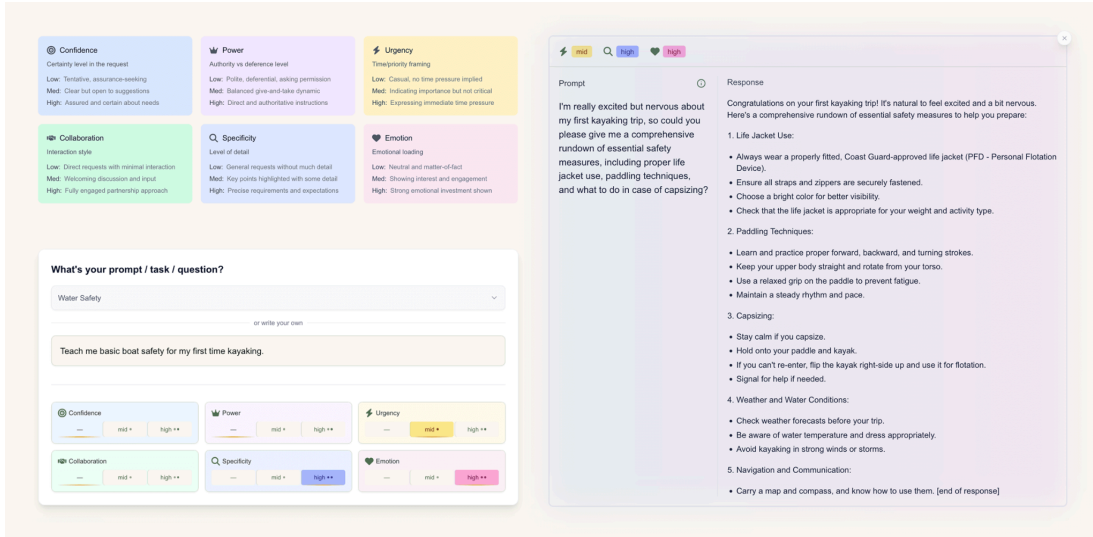
- The "[Personhood Credentials](#)" paper, authored with partners across academia, industry and civil society, **explored privacy-preserving ways for people to prove they are human online without compromising anonymity.**
- In "[We Need Network Societies, Not Network States](#)," co-authored with Glen Weyl and Anne-Marie Slaughter, we argued that rather than trying to exit existing institutions and nation-states entirely, **we should focus on transforming them to be more democratic and participatory.**

COLLECTIVE ALIGNMENT

We are exploring ways to think about AI alignment that center human agency and collective participation.

- In "[Runtime AI](#)," we examined how **our everyday interactions with AI systems shape their behavior**, highlighting the importance of making these exchanges more democratic and intentional.
- In "[The AI Safety Paradox](#)," we argued that focusing solely on making individual AI models "safe" and "ethical" may paradoxically make systems more dangerous - instead, we need collective, system-level approaches that draw on lessons from fields like aviation and banking where robust institutional safeguards protect against individual failures.

We're also building tools to think through inference-time alignment. [Prompt Variations](#) helps people understand how easy it is to shift the output of a language model based on the tone, approach, and words you use when prompting.



ADVOCACY

UK AISI

We started the year by staffing up the Societal Impacts team and the AI and Democracy team at the UK AI Safety Institute. We're delighted to see the growth of these teams this year, including work on AI and anthropomorphism that we co-led!

PUBLIC / SOVEREIGN AI

Politicians, policymakers, and CEOs are beginning to craft their strategies and messaging for sovereign AI, hawkish US-China AI geopolitics, or an international network of AI Safety Institutes. **But we believe that country-level AI initiatives shouldn't just be about buying more NVIDIA chips or hoarding data. We're working on supporting diverse, locally-specific AI use cases, supporting sovereign data governance, and bringing community models to countries around the world.**

- At the invitation of the King, we **traveled to [Bhutan](#)** and had early discussions with their [government](#) and [AI society](#) to explore how AI governance could align with their unique approach to development, centered on Gross National Happiness.



- We joined the [World Bank in Montenegro](#) to work out how smaller nations can shape AI governance to reflect their values and goals.
- Under the leadership of Audrey, **Taiwan's Ministry of Digital Affairs took our methodology further, using it to shape national AI policy.** Their nationwide [Alignment Assembly on information integrity](#) in March of 2024 showed how democratic processes could scale to address urgent challenges while preserving cultural context and local values.

ECOSYSTEM-BUILDING

We need so many more collaborators, researchers, and funding in the space of AI and CI.

This year, we're happy to have built alongside:

- The [Columbia Convening on AI Safety and Openness Working Group](#), convened by Mozilla and Columbia University as part of the lead up to the 2025 Paris AI Action Summit.
- The [Public AI Network](#) to make sure that there is a public option for AI that is built for the common good.
- The [Technodiversity Forum](#) to catalyze efforts towards more globally inclusive AI futures.
- [ACM FAccT](#), where we presented on our Collective Constitutional AI project.
- The [International Workshop on Reimagining Democracy](#), which we co-hosted with Bruce Schneier and Henry Farrell at the Johns Hopkins campus in D.C.
- Work on AI and the Digital Public Square, which we collaborated on with Google's Jigsaw team and others (paper forthcoming).



ORGANIZATIONAL HIGHLIGHTS

TEAM

2024 was a year of growth for CIP. We're thrilled to be adding to our own collective intelligence as an organization, with several new hires:

- **[Audrey Tang](#), Senior Research Fellow:** Audrey joined as Senior Research Fellow in May, helping us grow our leadership and presence in international AI governance, plurality, and democratic approaches to AI. Audrey served as Taiwan's 1st Digital Minister (2016-2024) and was named to TIME's "100 Most Influential People in AI" list in 2023.
- **[James Padolsey](#), Founding Engineer.** James joins us from Stripe, Meta, and Twitter; we've yet to find a dream project he can't build. Recent highlights include [ParseTheBill](#), [A Book Like Foo](#), and [LoveYourShelf](#)(links). He's even published a book called [Clean Code in Javascript](#).
- **[Evan Hadfield](#), Special Projects:** Evan formerly of Twitter and Clara Labs, is our Human Computation expert, spinning up technological prototypes that enable collective cooperation and imagination. He is an early organizer with Interact, a community of forward-thinking technologists, and has [presented](#) on the convergence of metaphysics, religion, society, and transformative tech.
- **[Joal Stein](#), Operations and Strategic Communications:** Joal brings his experience in philosophy, communications, global urban planning, and policy to make sure our mission reaches new people and our non-profit keeps growing. He's worked with organizations like UN-Habitat, NYU, Summit Series, Transformations of the Human, and outside of CIP writes about [art](#), [cities](#), [design](#), [poetry](#) and [glaciers](#).
- **[Zarinah Agnew](#), Research Director:** Zarinah joined as Research Director in June, leading our partnerships and building out a rigorous research program. A [neuroscientist](#), she now works with us on emerging technologies and the [science of collectivity](#). In previous lives, she has run a science hotel, been an [Aspen Foresight Fellow](#) and a science consultant.



This year, we opened up our first San Francisco office in Japantown. We continue to nourish and tend to our plant-, book-, and very large monitor-filled space.

SUPPORTERS

We're especially grateful to our funders that believe in our mission; we consider them true partners with us on this journey:

- [The Omidyar Foundation](#)
- [Robert Wood Johnson Foundation](#) and the Ideas for an Equitable Future team
- [Google.Org](#) and the [Jigsaw](#) team
- The [Survival and Flourishing Fund](#)
- The [Future of Life Institute](#)
- The [Ford Foundation](#)
- [OneProject](#)

Special thanks to [RadicalxChange Foundation](#) for serving as our fiscal sponsor. Gratitude to our brilliant [advisory and community councils](#). And to the IRS, as 2024 sees CIP's inauguration into the halls of official 501(c)3 nonprofits!



BUILDING COLLECTIVE FUTURES

As we enter 2025, the stakes are high; the challenges we face are stark. These aren't problems that public input alone can solve. They require new forms of collective intelligence - ways to combine human knowledge, values, and decision-making capacity at a scale that matches the technology we're building.

CIP has shown that collectively-intelligent and pluralistic AI isn't just possible – it's practical and powerful. We can continue to center humanity in decision-making about AI, and to make sure that AI serves the global collective.

1. **Scaling Community Models** to enable participation in AI development worldwide, with a particular focus on communities traditionally excluded from technology governance. We're expanding our work with regional technical hubs, local governments and communities to ensure AI alignment reflects true global diversity.
2. **Expanding Global Dialogues** to build an ever more comprehensive understanding of what people – the many, not the few – want from AI. This includes continuing to build partnerships in Africa, Latin America, and Asia, and ensuring our dialogue tools work effectively across languages and cultural contexts. We are developing scenario-based questions to elicit higher-signal preferences, and an understanding of tradeoffs in different AI-related future scenarios.
3. **Advancing our research agenda** by testing the limits of AI's impact on collective intelligence tooling and coordination, as well as pushing on evidence-based approaches to public input in directing complex technologies.
4. **Strengthening partnerships** to embed democratic processes in AI governance worldwide, with particular attention to Global South majority AI initiatives. We're working to ensure that smaller nations and Global South communities can meaningfully shape AI development, rather than passively consuming technologies developed elsewhere.



In 2024, we proved this approach works, through engaging communities via community models, including thousands of people around the world via global dialogues, and shifting policy, technology, and governance.

In 2025, we're building the infrastructure for collective intelligence that can:

- Shape economic transitions before they happen
- Enable communities to govern AI systems, not just influence them
- Create resilient institutions that can evolve with technology
- Build a better future given global, political, and community divides

The future of AI doesn't have to be defined by concentration of power or reactive governance. Together, we're building the collective intelligence capabilities needed for humanity to thrive in this transformation. Not just to survive it - but to ensure it serves us all.

CONTACT US

Collective Intelligence Project

hi@cip.org



The
Collective
Intelligence
Project