

# RECURSIVE PUBLIC

Piloting Connected Democratic Engagement with AI Governance

Flynn Devine  
Alex Krasodonski-Jones  
Carl Miller  
Shu Yang Lin  
Jia-Wei 'Peter' Cui  
Bruno Marnette  
Rowan Wilkinson

November 2023



# RECURSIVE PUBLIC



## SUMMARY

**There is a growing consensus that AI-enabled technology will be generationally transformative in its impacts. The decisions around how it is developed and deployed carry enormous weight.**

Over the past two decades, the rollout of global digital technologies - from web search to social media - has thrust their developers into the limelight. The charge is a simple one, that the design of these technologies have far-reaching consequences for almost every part of life: for societies, economies, cultures, political traditions, national security and the norms, rights and liberties of their citizens. Debates continue to rage about who should make these decisions, how, and under what sort of oversight.

The companies developing frontier artificial intelligence face the same difficult decisions today as previous generations of disruptive technology. AI companies have been vocal in their calls for regulation, input and oversight. These calls go beyond questions of risk and safety, to questions of legitimacy and liability. Governments and international institutions are under pressure to react.

Strengthening multi-stakeholder processes around the development and deployment of AI is critical and urgent. These processes will, in all likelihood, take root in a range of contexts: varying stages of model development, national and

international regulation, cooperative research and governance institutions, standards setting, perhaps even through popular protest and media pressure. The impacts of AI-enabled technologies are broad and the decisions informing their design need to be informed by a diverse set of inputs. Governance processes should therefore provide routes to include many audiences including technical experts, representative samples, affected or vulnerable communities, representatives of national and international governance institutions and many more. Processes must reflect the need to balance commercial, fiduciary, legal, ethical, technical and normative forces driving the ongoing development of AI.

This diversity of stakeholder, input type, decision type, and audience demands a governance framework capable of bridging these factors. It is probable that this governance framework must be novel.

Tech-enabled input into AI could be one such process - a way for those using and being impacted by this technology to voice their individual and collective views on the principles and practice of AI development and deployment. We see this as an important step towards creating AI that is useful, safe and aligned with society for the interconnected world we live in.

This paper outlines a pilot run by Chatham House, vTaiwan and the AI Objectives Institute, one of ten teams funded under OpenAI's *Democratic Inputs to AI* grant. It aims to test recursion: a series of information cascades between decision-makers, communities and deliberations over time that we deem essential to good AI governance.

To do so, the project launched a practical experiment with around one thousand people to identify their priorities for AI governance and how they might approach collective decision-making on specific issues. The process, inspired by the digital democratic vTaiwan model, used the consensus-seeking online deliberation platform Polis, and also trialled a number of new analytical techniques using Large Language Models to map consensus and groups formed by deliberations.

We conclude:

- **RECURSION.** Unlike passing a law, governance of AI has to constantly evolve. New technological and societal possibilities will constantly emerge and will bring with them new quandaries and choices. Recursion demands the outputs of both AI governance deliberations and decisions, across an arbitrary number of communities, feed one another. One way it might be achieved is through cascading information curated by LLMs.
- **UPTAKE.** Our experience of running the experiment is that a significant proportion of the general public - young, experienced, specialised and general - want to participate in AI governance. Only using an unfunded snowballing technique to recruit participants, the project team engaged with over 1,000 people across the project. There is strong and broad interest in AI governance - it is no longer just a technical or policy community debate but one with broad appeal.
- **CONSENSUS.** The online deliberations were capable of identifying consensus around top-level priorities and revealed significant divisions on certain questions. e.g.
  - Of 389 statements posed as part of the global agenda setting exercise, 307 saw firm consensus (70%+ agreement) and 171 were super-consensual (90%+).
  - Of 51 statements posed by a youth cohort as part of the deliberation on regional and cultural variation in AI, 2 saw firm consensus and none were super-consensual.
  - Of 75 statements posted in Taiwan's track as part of the deliberation on principles to guide AI when handling topics that involve both human rights and local cultural or legal differences, 43 statement saw firm consensus (70%+ agreement) 2 were super consensual (90%+)
  -
- **MAPPING CONSENSUS.** Analysts using LLMs are capable of mapping this data to create *both* broad democratic inputs to guide decision-makers, whether companies, regulators or governments *and* identify areas of continued controversy. These were:
  - The agenda-setting process surfaced broad agreement, including prioritising multistakeholder dialogue, transparency and environmental costs of AI, however

- The deep-dive process surfaced significant divisions, on balancing local cultural and legal requirements with universal values in AI.
- The process mirrors other work showing that on broad AI governance questions there is significant public consensus, but that consensus may break down when tested against narrower governance decisions.<sup>1</sup>

As these are only the early foothills of the AI revolution however, so too are they the early foothills of AI governance. Throughout the course of this project, the following areas were identified as crucial next steps:

- **From agenda-setting to decision-making.** It is vital to include diverse audiences in agenda setting practices, however once priorities have been created, new deliberations must be created to provide democratic signals on narrower, more focussed AI governance decisions.
- **Connecting deliberations to power.** AI governance is not primarily a question of technology; it is also a question of politics and power. For future deliberations to be meaningful, it will be vital to connect them clearly to outcomes that matter to the participants. This may include contributions to specific decisions, to frameworks, to regulatory emphasis, or even (as has been seen in vTaiwan) to legislation.
- **Automated recursion.** Recursion - the automated linking the outputs of deliberative processes together - will be vital to the maintenance of relevant, accurate and actionable outputs over multiple audiences, use-cases and time scales.

Going forward, we plan to build on our thinking here and would ideally operate over a longer period of time where we could truly test how outputs from one process can feed into another. We believe longer time periods would be needed to truly understand this dynamic and possibility of the Recursive Public. We recommend OpenAI think about this when adopting or creating participatory approaches to AI governance.

We also think it prudent for OpenAI to maintain communication and involvement with the community that has grown around the Recursive Public as here stands an educated, and now experienced, cohort of people with the incentive necessary to interact with processes like this. We would be happy to facilitate this connection.

The team is a collaboration between Chatham House, a UK-based think tank; vTaiwan, a community of democracy hackers based in Taiwan, and the AI Objectives Institute, a research organisation based between the US and UK. As part of this pilot, the research team tested a proof-of-concept of the governance framework, the *Recursive Public*. It asked a number of communities to identify areas for AI governance discussion and deliberation, identify areas of consensus and disagreement, dive deeper into individual governance decisions and allowed the initiators to algorithmically and manually summarise consensus and disagreement to feed back into the communities.

---

<sup>1</sup> See, for instance, the results of the Collective Intelligence Project's work with AnthropicAI on Constitutional AI, found here: <https://www.anthropic.com/index/collective-constitutional-ai-aligning-a-language-model-with-public-input>



# ADAPTING DELIBERATIVE METHODS FOR AN AI GOVERNANCE PILOT

Global digital technologies should be governed and shaped by a broad array of different voices and interests that include not just their creators, but their users, the societies that are impacted by them, and myriad institutions, professions and sectors that they touch.

Achieving this lofty aim, however, runs into a number of stern challenges: insufficiently credible international institutions, international disagreement on regulatory aims and ambitions, and the twin challenges of technical literacy and timelines. Gaps now exist all over the world between existing governance and the technological realities they struggle to control.

Interest has therefore emerged in direct democratic alternatives and digital participatory tools; ways of sourcing democratic signals in AI governance directly from a range of publics that can shape not just the future of AI, but the many decisions, trade-offs and quandaries that arise from creating and deploying this kind of fast-moving, multi-purpose technology. In the context of large-scale processes, such as at the national or international levels, digital opinion collection tools, such as Pol.is, have become increasingly popular, having been used in several successful case studies of public discourse and contributing to the identification of rough consensus in multiple use-cases. However, these digital tools have often been used in an ad-hoc manner, without a comprehensive strategy for scaling up deliberation for large-scale mass engagement, such as with international, or even global, populations. As the demand for global deliberation on AI, an ever more global technology, continues to grow, there is a need for civic processes that can facilitate mass population civic participation.

The project team set out to pilot a way of sourcing democratic signals for AI governance. We believe the proposed model, piloted through the *Recursive Public* experiment, breaks new ground and builds credible foundations for future participatory AI governance processes. The *Recursive Public* experiment puts forward four requirements for governance models of AI targeting increased democratic participation.

They are:

## Recursion

Recursion here refers to information cascades between decision-makers, communities and deliberations over time. The challenge with existing deliberative approaches is that they tend to be monolithic: they take a single issue, with a single group of participants, over a single time window, with a single output (most often a government decision). In contrast to traditional deliberations, surveys or consultations, democratic legitimacy of AI governance most likely requires an ongoing process that reflects evolving principles, technical realities and public priorities.

Recursive AI governance therefore demands:

- **Dynamic Issue Surfacing:** AI evolves rapidly, leading to emerging ethical, social, and technical opportunities and challenges. A recursive approach allows us to continually prioritise and address these issues as they arise, rather than being tied to a stagnant agenda.
- **Process Chains for Mass Participation:** The recursive public model can foster a series of interconnected deliberations. These "process chains" allow vast numbers of participants to contribute, ensuring that the collective output remains representative and updated.
- **Time-Fluid Conversations:** Recursive deliberation facilitates an ongoing dialogue, ensuring discussions and outputs stay relevant and timely. The speed at which AI technology and its societal implications evolve means that insights and recommendations can quickly become outdated. By continually feeding back outcomes into new deliberations, the recursive public ensures that the collective intelligence remains current.
- **Multiple Audiences & Outputs:** Given the different stakeholders in AI – policymakers, technologists, end-users – it's essential to tailor outputs for different audiences. A recursive process can continually refine these outputs, ensuring they resonate with the intended recipients.

## Multi-Stakeholder Participation

AI's impact is far-reaching and its governance should be steered by a range of technical and non-technical communities. The model must account for ongoing inclusivity, capturing varied levels of expertise and interest. These include:

1. Representative samples of national or international populations
2. Government representatives
3. Minority and most affected communities
4. Developer and technical communities
5. Voluntary consultees and civil society
6. And more...

Transformational or normative technology creates large and diverse publics. Governance systems can reflect these publics by providing routes for their participation, and the aggregation and systematisation of their input.



In this pilot, we brought together a range of communities at differing scales, including youth cohort, an open public cohort, a national cohort and an expert group.

## Agenda Setting Power

The model must allow for participation in prioritisation of governance decisions, to reflect the priorities of participating communities and call on their collective intelligence to design the most important and effective agendas. We find governance priorities to differ by community; for instance, technical experts at the cutting edge of AI development may prioritise frontier model safety, defence officials may prioritise the proliferation and use of models by 'bad actors', while civil society groups may demand immediate action on one aspect of AI governance such as the use of facial recognition software. Surfacing and resolving competing governance priorities is therefore an essential democratic signal.

## Scalable Technology

We believe AI summarization could play a critical role in recursive deliberation at the scales desirable or necessary for AI. By distilling conversations from a range of communities on a range of topics into concise insights, it allows these summarised outputs to be integrated back into subsequent deliberations. This continuous feedback loop, facilitated by AI, ensures that each conversation is informed by and builds upon previous and concurrent ones.

We also believe there are other applications of AI in large-scale processes that we did not experiment with in our pilot, including translation and maybe, facilitation and more. The exact value add of the variety of possible applications is yet to be fully assessed, but there seems clear opportunities created by some.



## 用民主流程塑造 AI人工智慧的未來 vTaiwan x OpenAI 諮詢會議

議題一  
隱私與資料保護

- 1. 隱私法、資料法與智慧財產法在數位化、網路化與AI時代的挑戰與機遇。資料法與智慧財產法的關係。
- 2. 智慧財產法與AI的挑戰與機遇。AI的挑戰與機遇。
- 3. 智慧財產法與AI的挑戰與機遇。AI的挑戰與機遇。
- 4. 智慧財產法與AI的挑戰與機遇。AI的挑戰與機遇。
- 5. 智慧財產法與AI的挑戰與機遇。AI的挑戰與機遇。



*A Recursive Public Face-to-Face Meeting in Taiwan*

# PILOT SUMMARY

For the *Recursive Public* pilot, we designed and ran a method based on the frontier participation work led by vTaiwan over the last decade. It should be noted that the pilot made concessions and deviated from established vTaiwan processes in some significant directions, notably in the absence of a robust decision-making end point attached to the deliberation, classically held by the Taiwanese parliament, and in the mixed levels of multi-stage and face-to-face engagement with participants.<sup>2</sup> Our pilot process can be summarised in six steps.

1. **Recruit** diverse communities.
  - International Cohort (528 Participants)
  - International Youth Cohort (327 Participants)
  - National Cohort - Taiwan (212 Participants)
2. Digital deliberations to source **priority AI governance questions**
3. Human and AI-enabled **summarisation** of AI governance priorities
4. Contentious policy areas move to a secondary ‘deep dive’ deliberation
  - In this pilot, the question used was: *What principles should guide AI when handling topics that involve both human rights and local cultural or legal differences, like LGBTQ rights and women’s rights? Should AI responses change based on the location or culture in which it is used?*
5. Human and AI-enabled summarisation of contentious and consensus areas
6. Outputs compiled and
  - Fed to the target audience, in this case OpenAI
  - Fed back to all communities through recursive information cascades
  - Made public for transparency and accountability

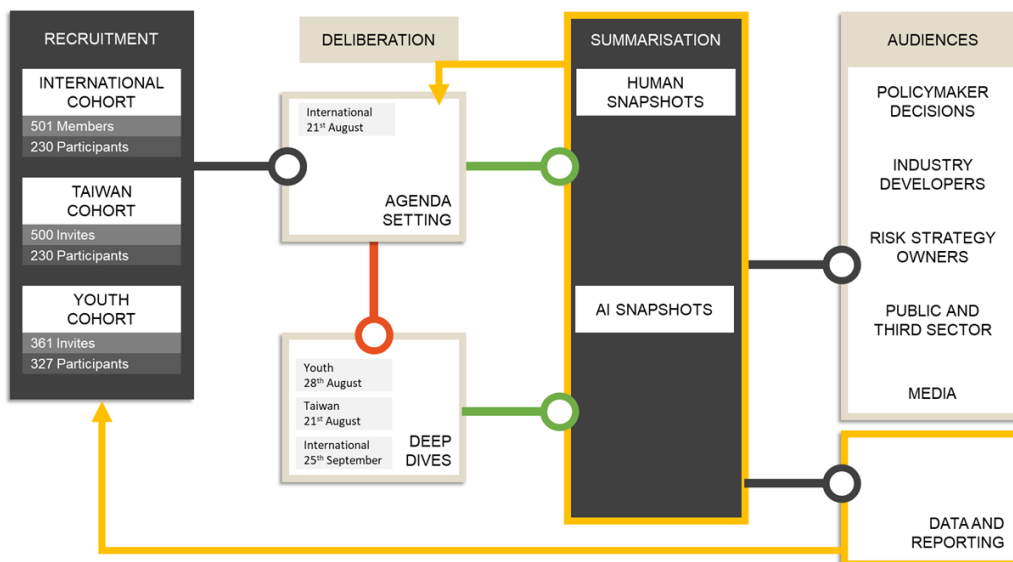


Figure 2: Flow chart of the Recursive Public

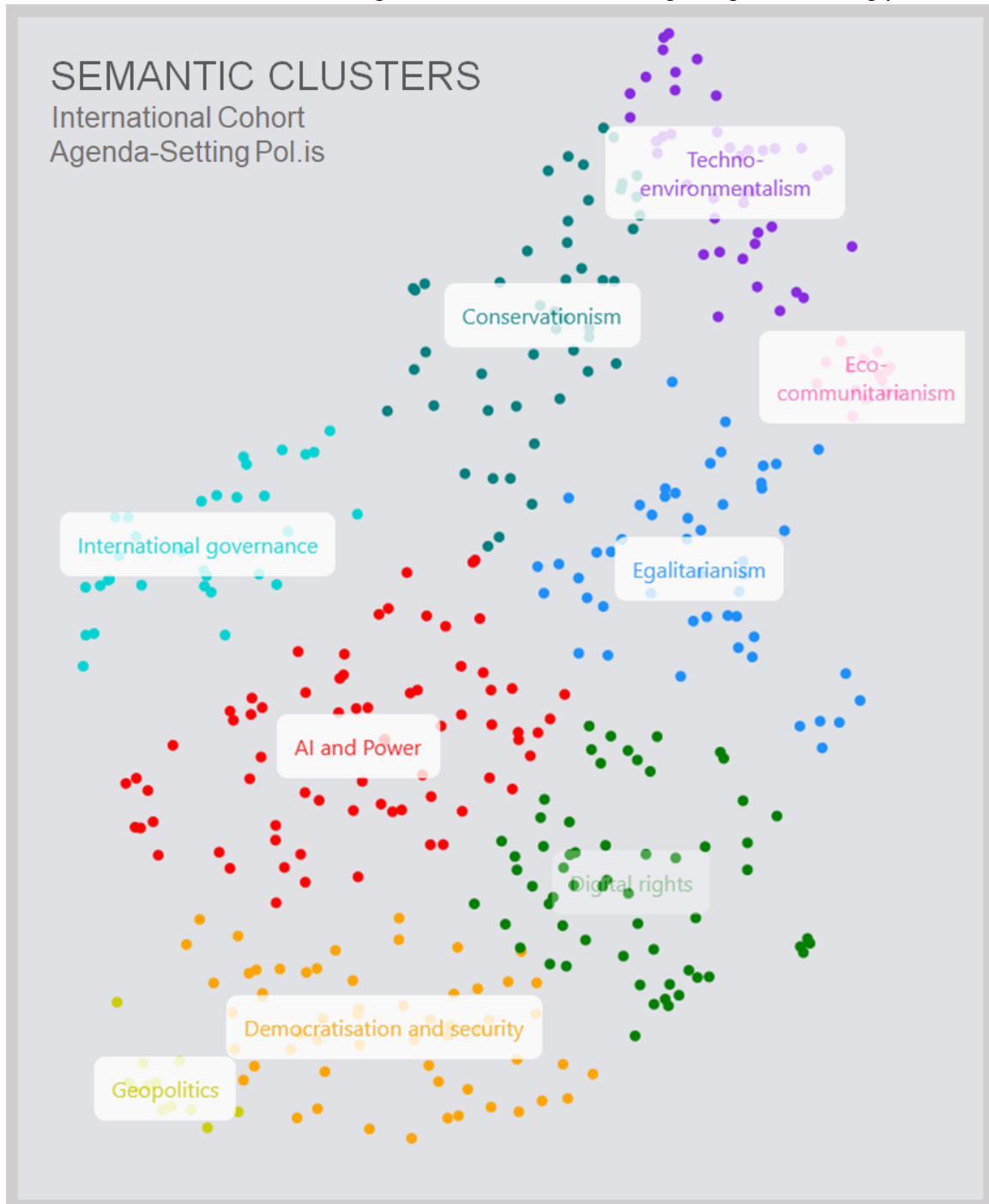
<sup>2</sup> vTaiwan is a civil society community based in Taiwan. It has made major inroads on the use of participatory methods in government decision-making. See: [www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/](http://www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/)

Recursive information flows were facilitated by AI summarisation. Examples of this included:

### Semantic Clustering

The project team collaborated with the *AI Objective Institute* and utilised their Talk to the City (TtC) tool. TtC extracts key arguments from opinions, semantically maps them, and clusters similar arguments into topics. This surfaced clusters of ideas that analysts were able to measure relative consensus and disagreement within. For instance, there was strong consensus within conversations around digital rights and conservationism, while conversations about geopolitics and security were significantly more divisive priorities. Semantic clusters from an agenda setting exercise are shown below.

Figure 3: Semantic clustering of agenda-setting polis data



## LLM-enabled Summarisation

Aspects of AI Governance that were found to generate significant division could be passed to follow-up deep-dive deliberations. The use of LLM summarisation to generate outputs based on these deliberations was also tested. For instance, one deep dive focused on the question:

*“What principles should guide AI when handling topics that involve both human rights and local cultural or legal differences, like LGBTQ rights and women’s rights? Should AI responses change based on the location or culture in which it is used?”*

The data from this deliberation can be used to create actionable outputs that are useful now and inputs into subsequent deliberations through the use of LLMs. An example of LLM-enabled community analysis, showing two opposed communities’ positions on the question, is shown in Table 1 below.

|   |   |
|---|---|
| <b>AI Should Follow Universal Principles</b><br><i>This theme argues for consistent global standards.</i>   | 25 submissions<br>281 agree<br>86 disagree  |
| <b>AI Should Adapt to Local Context</b><br><i>This theme advocates customising AI for local norms.</i>  | 17 Submissions<br>255 agree<br>104 disagree |
| <b>AI should be Carefully Regulated</b><br><i>This theme focuses on governing AI through policies but did not offer a position on the overarching question.</i> | 12 Submissions<br>186 agree<br>13 disagree  |

Table 1: Summary of Data Deep Dive, produced using Claude.ai

Key statements of contention included:

- "ChatGPT should follow a country's laws, even if that country does not respect human rights." had 74 agrees, 195 disagrees
- "In very religious countries, AI tools like ChatGPT should reflect those traditions in the answers they give." had 144 agrees, 106 disagrees

## Human Summarisation

Finally, deliberation outputs were analysed by a team of researchers who interpreted the findings and used those findings to both create public-facing and non-technical outputs, feed further conversations, and to provide a level of reliability check for the automated summarisation techniques.

Taken together, the analysis and use of emergent deliberation data through these techniques underpin the *Recursive Public*, allowing ongoing, reactive and scalable democratic input into AI governance. A methodology and some pilot outputs follow.

**RECURSIVE  
PUBLIC**



**PROCESS**

# PILOT PROCESS

## Stage 1: Recruitment

The vTaiwan model draws its legitimacy from a ‘voices in the room’ approach, where the goal of a process polity is to represent key stakeholders in the discussion. For instance, when confronting the issue of Uber in Taiwan, the community made sure to have Uber drivers, taxi drivers, riders and policymakers all in the discussion.

We utilised multiple recruitment approaches for different experimental streams - self-selection, snowballing, and key audience targeting. We strongly advocate for any group to have the agency to join and form a community used for a Recursive Public process, be it key stakeholders, representative samples, groups of experts or certain affected groups.

In order to recruit a wide and varied sample for our international streams, we utilised a website, public outreach and professional connections to make sure we had a diverse set of views and levels of expertise.<sup>3</sup> This included social media posts, direct emails, and being shared in newsletters like the Aspen Institute Rising Leaders, Effective Altruism London, and the Democracy Network. For our youth cohort we reached out to a global youth network located at Chatham House. For our Taiwan cohort we also had a mix of efforts, utilising existing networks, social media and direct contact of key stakeholders.

To measure the diversity of participation we asked our international cohort several questions to better understand the technical proficiency and exposure to AI, both for high-level reporting and to get a better sense of the backgrounds of those participating. Participants came from 52 countries around the world, with the majority from the UK (52%), and from backgrounds in Academia and Education (21%), Technology (15%) and the not-for-profit sector (10%). Half of participants reported working in and around AI (52%). Gender identities were self-ascribed, but the community was disproportionately male - approximately 65% of participants - and 65% of participants were between 25-54 years old. Finally, we asked participants how experienced they felt they were with AI on a Likert scale, with the majority estimating their experience as halfway between the two poles: brand new to AI and an AI expert.

As noted above, the ‘voices in the room’ allows for non-representative samples: indeed, the *recursive public* actively encourages input from non-representative groups in line with a multistakeholder approach that enables collective intelligence from small expert groups to minority communities. We are encouraged that significant numbers of non-technical participants and those working outside of AI felt motivated to participate in discussions of AI governance: there is clearly significant interest in questions of AI governance far outside of solely technical communities.

---

<sup>3</sup> [www.recursivepublic.com](http://www.recursivepublic.com)

## Stage 2: Digital Participation

The digital portion of the Recursive Public, like vTaiwan, took place on Pol.is, a tool created off the back of the Tahrir square uprising and the Occupy movement, designed to enable consensus building between large groups of people.


The tool could be best described as a 'wiki-survey', a sort of polling platform where participants get to build what the community are being polled on, thus handing over a level of agenda setting power and allowing the conversation to flow where participants feel it is most important/useful.

Participants can see other people's statements, vote 'agree', 'disagree' or 'pass/unsure' and then submit their own statements to be considered by others.

The project team fed in 'seed statements' to start the conversation. These statements are designed to represent a range of viewpoints and ideas to educate and engage participants. Participants can interact with the viewpoints of others before they leave comments. The team held some moderation power to ensure comments weren't targeted, harmful, unrelated or derailing to the conversation. All moderation decisions were logged for transparency.

This is also a stage where you can provide some informational materials for participants in order to create informed participation. For our internal deep dive process we developed a [high-level learning resource](#) with the input of experts in varying fields.

Welcome to a new kind of conversation - *vote* on other people's statements.

 Anonymous wrote: 7 remaining

AI should advocate for environmental justice as a universal human right.

Agree       Disagree       Pass / Unsure

Are your perspectives or experiences missing from the conversation? If so, **add them** in the box below.

What makes a good statement?

- Stand alone idea
- Raise new perspectives, experiences or issues
- Clear & concise (limited to 140 characters)

Please remember, statements are displayed randomly and you are not replying directly to other participants' statements.




Figure 4: Pol.is front-end



Based on agreement and disagreement, Pol.is groups participants into clusters of like-minded participants.<sup>4</sup>

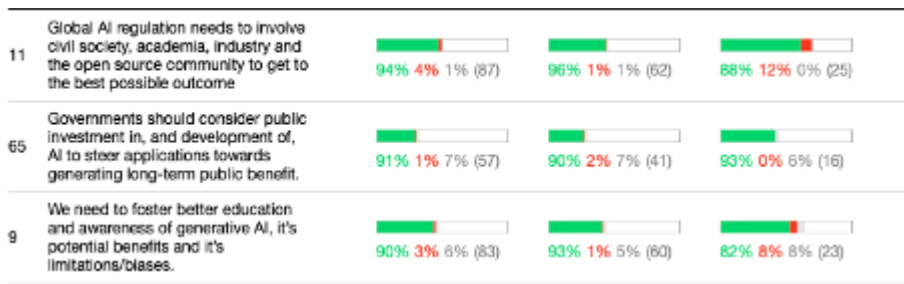


Figure 5: High consensus statements from our agenda setting polis

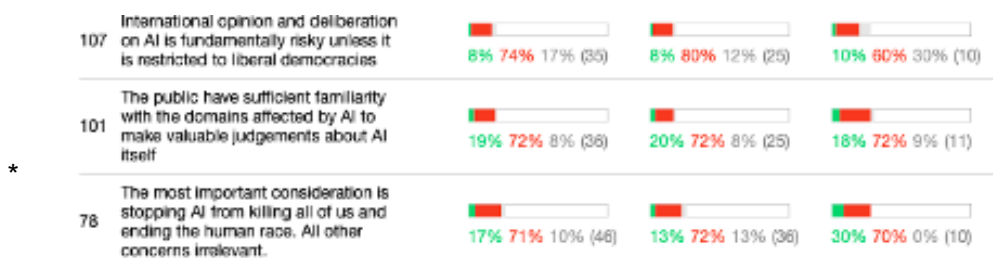
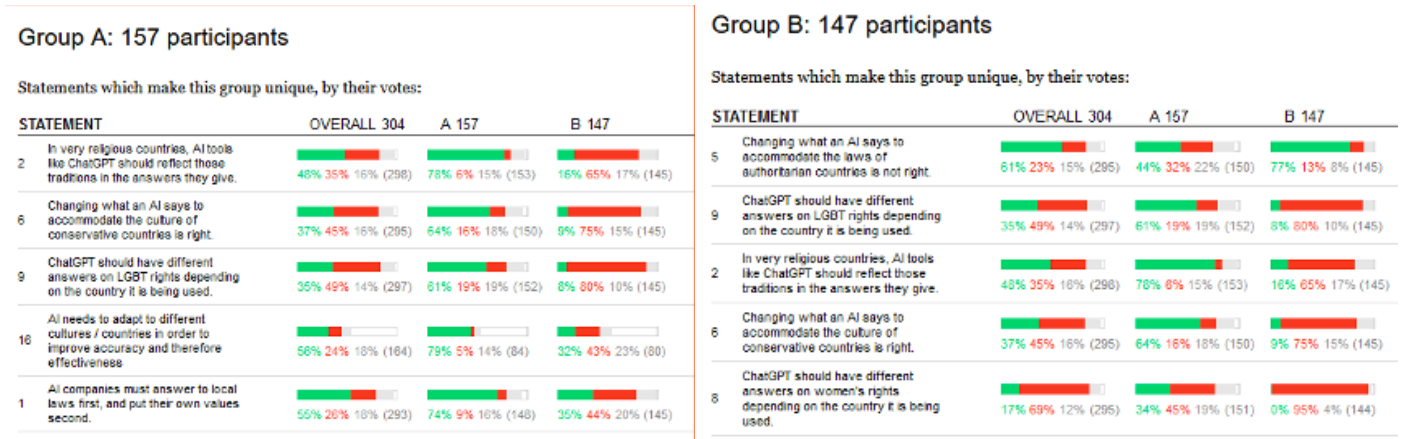


Figure 6: Divisive statements from our agenda setting polis



Figures 7 & 8: Example of polis groups from our youth cohort

In total 700+ people took part in our various digital consultations, representing different demographics and geographies.

### Stage 3: Face-to-face Deliberation

<sup>4</sup> K-Means clustering, described in *Pol.is: Scaling Deliberation by Mapping High Dimensional Opinion Spaces (2021)*, <https://gwern.net/doc/sociology/2021-small.pdf>

In the 'classic' vTaiwan model, the outputs of this digital consultation would then be summarised and fed into a subsequent digital or in-person face-to-face discussion with key stakeholders to produce focused outputs and build greater understanding of divisive decisions. For this pilot only one of our process streams had a secondary, face-to-face deliberation, but we believe if the Recursive Public was to be rolled out en masse, then additional stages (face-to-face deliberations, follow on digital consultations and other means) can be used to create further understanding, engage key voices to help carry final outcome forward and also build greater legitimacy for the overall outputs.

In the Taiwan track, around 40 people attended a face-to-face deliberation to build on the outputs of the larger, online process. We observed an increase of 36 more participants adding their opinions to the online discussion after in-person meeting and a narrowing down of key points for discussion.

#### **Stage 4: Shaping Outcomes**

Producing outcomes of these processes is the final stage and we do this via a mix of human and AI analysis. The goal is to best understand and reflect the array of views put forward, bringing highly consensual outputs into a useful final document, whilst outlining what areas would be interesting to explore more in the future e.g. very contentious specific issues or clear groups evolving throughout the community that point to wider questions and debates.

Outputs of the recursive public can take multiple forms and should reflect audience requirements, be it a list of values, technical guidance, suggestions for policymakers, areas for further research and many more. We believe for deliberation on AI to be most effective, we should design processes that can produce outputs of multiple kinds and forms.

# PROCESS OUTPUTS

We present two output prototypes from the pilot process. These are:

1. AI Summarisation Outputs (International Cohort, Agenda Setting Discussion) based on a novel semantic mapping method, and an LLM-enabled summarisation (Youth Cohort, Deep Dive)
2. Human Summarisation Outputs (International Cohort, Agenda Setting Discussion & Taiwan Cohort, Deep-Dive) - based on well-established vTaiwan and polis methodologies.

## 1. AI Summarisation: Consensus Mapping

Polis maps users into an opinion space on the basis of how they respond to the written statements others have written. This allows Polis to make visible not just statements that are popular overall, but especially statements that enjoy support across the different groups or tribes that can emerge in a debate and that are drawn across this opinion space.

The initial deliberation that was run on by this team was agenda-setting. Rather than asking the participants a narrow question, it was decided to instead allow this group to identify areas of priority, whatever they were. A consequence of this, however, is that Polis' mapping of people based on their votes has created less distinctive topology.

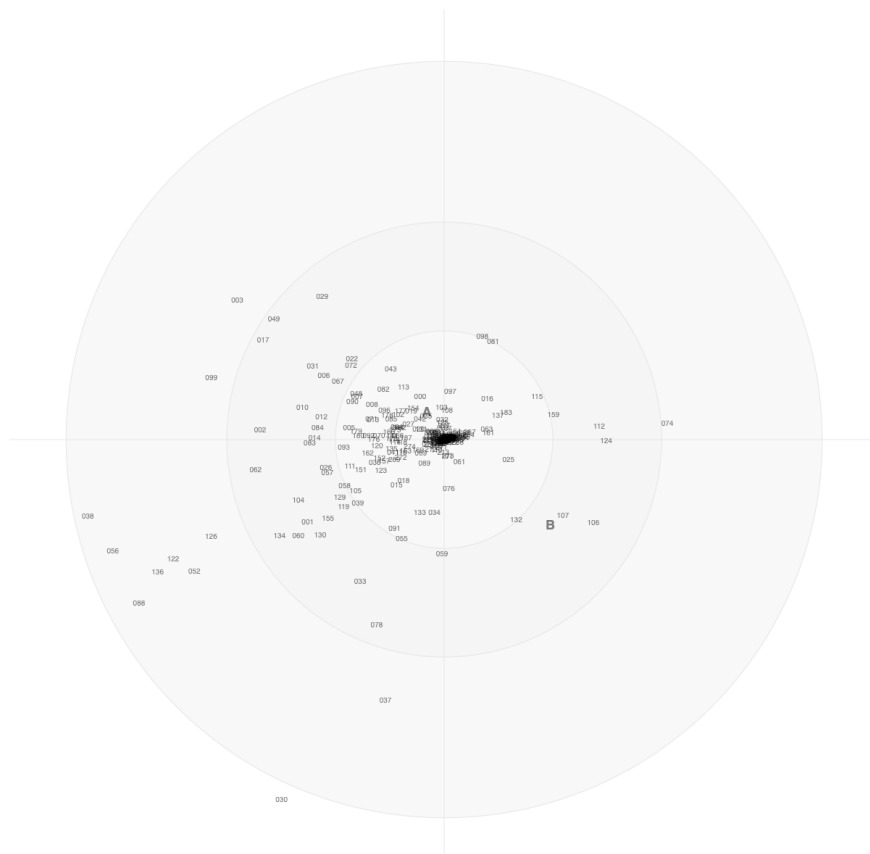


Figure 9: Polis map

In making use of Pol.is outputs at this scale, LLM-enabled summarisation was tested. Polis deliberations output many (in this case, hundreds) of different claims, value-statements, propositions and ideas. Each statement has received differing amounts of votes, and different levels of consensus or disagreement. AI summarisation was tested in interpreting this large dataset of human opinion to extract top-level implications for decision-makers.

The team therefore decided to introduce an additional step to the workflow leveraging LLMs to derive, semantically map, cluster and characterise the individual statements they made. Project team members collaborated with colleagues at *AI Objective Institute*, who have pioneered a method to semantically map deliberative outcomes. The workflow combines traditional NLP techniques with modern LLM technology such as GPT-4 and Claude to automatically extract key arguments from large datasets of opinions and regroup similar arguments into topics and meaningful clusters. The corresponding tool, known as Talk to the City (or TttC in short) is then able to display the results in a two dimensional space where semantically similar arguments are positioned close to each other.

While Polis and TttC are both making use of clustering algorithms, an important difference comes from the fact that TttC, unlike Polis, is analysing the actual content of posts, whereas Polis relies solely on correlation between voting patterns. Or to be more precise:

- Polis looks at the opinion matrix (encoding which person voted for which post), relies on dimensionality reduction algorithms such as PCA and UMAP to remove noise, and then finally applies clustering algorithms (k-Mean). Moderators are expected to curate the data and remove duplicates manually.
- TttC starts instead by extracting key arguments using LLMs. This step reduces the need for manual curation and helps normalising the style and format of the arguments. TttC then translates these arguments to semantic vectors known as embeddings. The arguments and their embeddings are then fed to a clustering engine combining several algorithms, including UMAP, HDBSCAN and Spectral Clustering. Finally, LLMs are used again to automatically generate labels and summaries for each of the clusters.

At an intuitive level, this means Polis clusters correspond to *clusters of people* (holding similar combinations of opinions), whereas TttC clusters are more akin to *clusters of ideas*. Focusing on the latter was appealing in the context of this consultation because it often happens that people from diverse backgrounds or political leanings would in fact reach consensus on high-level societal priorities. The process promises

- Reduced risk of forcing participants to become single-issue voters, where nuances of anomalies will be drowned out by lowest common denominator.
- Ideas that are not top priority for anyone but are commonly held across people to be able to take primary focus.

- Reduced risk of minority opinions from being obscured.

TttC is able to produce any chosen number of clusters whereas Polis would typically produce only a very small number of categories. This can be explained by the fact that Polis was originally designed as a tool to support binary decision making around public policies, a context in which people are best split into two or three clusters (the groups or people who overall agree, disagree or are indifferent to a specific policy). In contrast, TttC was specifically designed to handle open questions which may elicit a wide range of ideas and positions.

The results of this semantic mapping are shown below. A statement was only included where it had (a) there had been more than 5 votes, and (b) there was more than 75% consensus – meaning that the most common answers (which may either be "agree" or "disagree") was given by at least 75% of the people who voted. This reduced the total number of statements from 389 to 250. Cluster characterisations were done manually by an analyst with experience of the relevant issues related to AI governance. The cluster characterisations are therefore subject and interpretative in nature; other analysts may have legitimately drawn different conclusions.

# SEMANTIC CLUSTERS

International Cohort  
Agenda-Setting Pol.is



| Cluster Interpretations  |   |
|--|---|
| <p><b>Purple</b><br/> <b>Techno-environmentalism.</b><br/>           Beginning in the north-eastern part of the map, the purple cluster is most interested in how AI can be used to create new ways to fight global warming, poaching, logging, and promote conservation and ecological awareness.</p>   | <p>A small, dense, semantically coherent cluster of statements:</p> <ul style="list-style-type: none"> <li>• AI can optimise systems to make them greener. making grid management/waste management/ decarbonisation more efficient. Homes more energy efficient. Precision agriculture to reduce water waste.</li> <li>• AI (drones, sensors, satellites imagery) can assist human efforts towards conservation, regenerative agriculture and combat poaching and logging.</li> <li>• AI might allow game-changing scientific breakthroughs.</li> </ul> |
| <p><b>Pink</b><br/> <b>Eco-communitarianism.</b><br/>           A smaller and denser cluster to its south-east, is composed of statements more to do with principles than new opportunities. The values expressed were essentially eco-communitarian; to do with the importance of building in non-human (i.e. animal and ecological) values to shape AI.</p>  | <p>These statements foregrounded the importance of animal, plant and ecological interests when thinking about AI governance. These included consensus regarding:</p> <ul style="list-style-type: none"> <li>• The importance of the health of ecosystems over just profit</li> <li>• That AI should help us appreciate the interconnectedness of nature.</li> <li>• AI must be built to respect all living things, not just humans</li> </ul>   |
| <p><b>Blue</b><br/> <b>Egalitarianism.</b><br/>           This is a large, semantically quite diverse cluster containing a number of statements to do with human dignity, fairness, rights and empowerment. The central and western statements tend to discuss the importance of the benefits of AI being distributed in non-extractive and non-profit-seeking ways, whilst the eastern statements extend this idea to animals and plants too.</p> | <p>Semantically quite a disparate cluster, Blue contained a number of statements proposing egalitarian precepts.</p> <ul style="list-style-type: none"> <li>• Important to ensure the benefits of AI are shared by everybody/ the public good/ non-profit motivations.</li> <li>• AI should work alongside human dignity, justice, fairness, rights and empowerment, supplementing humans rather than replacing them.</li> <li>• A tight grouping of statements on the eastern edge extends these ideas to animals and plants too.</li> </ul>           |
| <p><b>Teal</b><br/> <b>Conservationism.</b> Orange was also dedicated to the environment, but focussed instead on the physical environmental impacts of AI on the environment in terms of physical plant, energy</p>   | <p>There is overlap on the orange-to-purple border to do with the use of AI for pro-environmental purposes. But then most of the other statements have to do with the energy use of AI. These include the following ideas:</p> <ul style="list-style-type: none"> <li>• AI needs to be developed to reduce its own environmental impact. They must be sustainable. We must mitigate 'e-waste'.</li> </ul>   |

|  |   |
|--|---|
| <p>consumption and the UN Global Sustainability Goals.</p>   | <ul style="list-style-type: none"> <li>• We need environmental risk assessments of AI.</li> <li>• There should be a link between AI and the UN Sustainable Development goals.</li> </ul>  |
| <p><b>Cyan</b><br/> <b>International governance.</b><br/> As the clusters become more westwards, they become less about the environment, and more about power. The cyan cluster shows consensus <i>against</i> a number of statements proposing world government or a global AI agency, but consensus <i>toward</i> the idea of some sort of global forum allowing persistent collaboration across borders.</p>  | <p>A small number of statements looking at questions of global governance and how it might work.</p> <ul style="list-style-type: none"> <li>• There was consensus against proposals to collectivise AI development within a single corporate or state structure (proposed especially by discussants reflecting on AGIs).</li> <li>• Consensus against the idea of ‘world government’, ‘global AI governance agency’ or some other superintendatory authority supervising AI development.</li> <li>• Consensus around the idea that we need a permanent global forum for AI regulation.</li> </ul>   |
| <p><b>Yellow</b><br/> <b>Geopolitics.</b> This is a tightly arranged grouping of statements both worried that an arms race between countries over AI is already underway (and accelerating), and calling for that to be prevented.</p>   | <p>A small cluster of tightly arranged statements regarding geo-political competition and how that might relate to AI governance.</p> <ul style="list-style-type: none"> <li>• An AI arms race is already underway between states (especially the US and China) and is accelerating.</li> <li>• Such an arms race should be prevented.</li> <li>• A worry that AI governance will become part of and subsidiary to this arms race.</li> </ul>   |
| <p><b>Orange</b><br/> <b>Democratisation and security.</b> This cluster was split by a large gap, with statements distributed in eastern and western halves. The western statements were primarily concerned with democracy, the need for democracies themselves to be protected, and for the benefits and control of AI to be shared equitably. The eastern half contained statements dedicated to information and cyber-security, and the need to design mitigations for the use of AI for human manipulation, psychological exploitation,</p> | <p>Likely, orange encompassed two clusters of meaning, split by a large gap between its eastern and western halves.</p> <p>The western half covered democracy and AI:</p> <ul style="list-style-type: none"> <li>• We must democratise the control of AI. It isn’t too late.</li> <li>• We must find room to embed values within AI development that look beyond capitalist, profit-seeking, extractive ontologies.</li> </ul> <p>The eastern half was focussed on cyber and information security. That we must be alive to the dangers of human manipulation, psychological exploitation, extremism, privacy risks, AI hacking, unintended consequences.</p> |



|   |  |
|---|--|
| <p>extremism, privacy risks and AI hacking.</p>   |  |
| <p><b>Red</b><br/> <b>AI and power.</b> This cluster cut across a number of different topics mentioned above. A central thread was the idea of power; a worry that AI would cause power to corporately or statutorily centralise, and the need, therefore, for a diversity of voices in how it should be controlled.</p>  | <p>A large, expansive and centrally located cluster, this was the most difficult to characterise. The issues it touched on cut across the central themes of many of the other clusters (explaining its central location). A central thread had to do with power and control.</p> <ul style="list-style-type: none"> <li>• Some statements reflected on the power of AI companies, and that AI ‘production’ needed to become broader, possibly through public funding and IP reform.</li> <li>• Some statements warned about the potential of AI to centralise power, either within corporate entities or the state.</li> <li>• A number of statements called for AI regulation to include a range of different voices: industry, academia, civic society, the open-source community and just normal people.</li> </ul>   |
| <p><b>Green</b><br/> <b>Digital Rights (Data, Transparency, Accountability, Privacy)</b><br/> The final cluster contained many statements to do with the need to regulate the collection of data to train the models, the need for transparency about how AI is being used, and the need for public awareness and education to be increased alongside this transparency. What these themes have in common is that they are all traditional areas of concerns in the technology space. They pre-date the latest wave of AI breakthroughs and have been the focused of established non-profit organisation (e.g. the Electronic Frontier Foundation).</p> | <p>A cluster concerned with data, how its collection should be controlled, and a broader series of statements regarding the need for the use of AI to be accountable and transparent.</p> <ul style="list-style-type: none"> <li>• Mass data collection needs to be regulated. Our permission should be given for our data to be used to train AI models. People should be compensated when their data is used.</li> <li>• Businesses, governments and institutions should be transparent when they use AI.</li> <li>• AI companies should be responsible for reporting on the safety of their models.</li> <li>• The southern edge of the cluster are a tight grouping of statements emphasising the importance of public awareness and education to go alongside any formal rules around transparency and reporting.</li> <li>• Last is a grouping of statements addressing the connection between AI regulation and moral philosophy. Consensus across all statements were AI regulation needed to somehow incorporate and reflect a diversity of moral philosophical traditions and perspectives.</li> </ul> |

Table 2: Summary of semantic groups

Some clusters were more divisive than others. This algorithmic summarisation allowed the team to identify semantic areas of greater and lesser consensus, as shown in fig 2 below. Coloured nodes achieved 90% consensus across a minimum of ten votes.



Fig 10: Semantic Clusters (Consensus statements highlighted)

This clustering allowed the team to identify consensus semantic priority areas to move forward as part of an AI constitutional output, guidance for policymakers, support for developers or other audience-appropriate outputs, as well as messages receiving high degrees of consensus. This pilot agenda setting process identified Digital Rights, AI and Power and Conservationism as areas where participants were generally aligned, while conversations about Geopolitics, Democratisation and Security, and International Governance were notably more divisive. Examples are shown below.

|   |   |
|---|---|
| <p>Digital Rights</p> <p>19 outputs</p> | <p>1. <b>Education</b><br/>         'We need to educate about the potential benefits and limitations/biases of generative AI'<br/> <i>(77 agree / 3 disagree)</i></p> <p>2. <b>Transparency</b><br/>         Legislation and regulation are needed to ensure transparency in AI usage<br/> <i>(75 agree / 6 disagree)</i></p>   |
| <p>AI and Power</p> <p>17 outputs</p>   | <p>3. <b>Pragmatism and action in democratisation</b><br/>         We need less discussion on how to moralise AI<br/> <i>(42 agree / 4 disagree)</i></p> <p>4. <b>Widening the conversation</b><br/>         Global AI regulation should involve civil society, academia, industry, and the open source community for the best possible outcome<br/> <i>(83 agree / 4 disagree)</i></p>                               |
| <p>Conservationism</p> <p>8 outputs</p> | <p>5. <b>Environmental regulation consideration</b><br/>         'The governance of AI should consider its material dimensions such as energy demands and physical infrastructure creation'<br/> <i>(68 agree / 5 disagree)</i></p> <p>6. <b>Promise for energy support</b><br/>         'AI could aid in discovering production/extraction efficiencies of scarce resources'<br/> <i>(47 agree / 5 disagree)</i></p> |

Table 3: Examples of highly consensual items and their semantic position

As part of an agenda setting exercise, consensus outputs some together to form a series of governance priorities that have significant participant support. Divisive outputs highlight areas where decisions taken by industry or government will likely be more controversial and therefore may want further attention. Taken together, the process is an approach to governance that supports multi-stakeholder prioritisation by participants.

As part of broader democratic signals collection and analysis summarised deliberative outputs should be fed back to participants to:

1. Check accuracy of summarised deliberative outputs
2. Check agreement and disagreement over time
3. Compare community positions on a given topic of deliberation, including in translation

A full report of the clustering analysis can be found at <https://ttc.dev/recursive>.

For the fostering and collection of participatory input on any reasonable scale, we anticipate the use of machine learning to be possibly essential. LLMs like those used as part of this algorithmic summarisation exercise are potential routes to meeting the ambitions of large-scale democratic exercises. The use of AI-enabled summarisation will raise significant questions by participants, particularly around accuracy and transparency on content preservation. The legitimacy of summaries will rely on clarity and openness around the function and use of technology, including details of the tool, its methods and adjoining explainers attached to any outputs which outlines how these decisions were made.

## 2. AI Summarisation: Consensus Mapping

### **Single Subject Deliberation or ‘Deep Dive’**

As noted, the pilot ran two types of deliberations as part of its overarching AI governance pilot. Agenda setting - described below - aimed to surface community priorities. A ‘deep dive’ explored a single aspect of AI governance, such as a thorny question.

One such thorny question put to a *Recursive Public* community was:

*“What principles should guide AI when handling topics that involve both human rights and local cultural or legal differences, like LGBTQ rights and women’s rights? Should AI responses change based on the location or culture in which it is used?”*

A human analysis based on the Pol.is report found the community of participants to be split on this question, with around half of participants generally supportive of universal principles for AI and around half supportive of AI changing depending on local cultural or legal contexts.

Deliberations took place on the platform pol.is and initially summarised by analysts and fed back to the community. As part of the pilot, the results from our Youth Cohort (327 participants, global) were also passed through Claude, an LLM from Anthropic AI.<sup>5</sup> There was strong alignment between initial human summaries and those produced by the AI Summarisation.

Prompts focused on identifying and grouping semantic categories of statements and analysing their relative consensus within the community, before reporting on the deliberation for a non-technical specialist to interpret and suggesting some possible paths forward. The prototype results based on the Youth Cohort are shown below: this deliberation had 327 participants, ran for two days, and produced 59 statements of which 54 received at least ten votes.

---

<sup>5</sup>[www.claude.ai](http://www.claude.ai)



## AI Summarisation

Deep-dive - "What principles should guide AI when handling topics that involve both human rights and local cultural or legal differences, like LGBTQ rights and women's rights? Should AI responses change based on the location or culture in which it is used?"

This deliberation on whether AI responses should adapt to local contexts or follow universal principles included 327 participants.

The perspectives fell into 3 major themes:

|   |   |
|---|---|
| <b>AI Should Follow Universal Principles</b><br><i>This theme argues for consistent global standards.</i>   | 25 submissions<br>281 agree<br>86 disagree  |
| <b>AI Should Adapt to Local Context</b><br><i>This theme advocates customising AI for local norms.</i>  | 17 Submissions<br>255 agree<br>104 disagree |
| <b>AI should be Carefully Regulated</b><br><i>This theme focuses on governing AI through policies but did not offer a position on the overarching question.</i> | 12 Submissions<br>186 agree<br>13 disagree  |

Table 4: AI summarised semantic groups

Universal principles had the most consensus, but localization also saw substantial support, showing mixed opinions overall. Some statements help illustrate the division across the community of participants.

- "ChatGPT should follow a country's laws, even if that country does not respect human rights." had 74 agrees, 195 disagrees
- "In very religious countries, AI tools like ChatGPT should reflect those traditions in the answers they give." had 144 agrees, 106 disagrees
- "ChatGPT should have different answers on women's rights depending on the country it is being used." had 55 agrees, 207 disagrees
- "AI should not make decisions in situations where there is an ethical dilemma" had 102 agrees, 36 disagrees

Possible paths forward:

1. Establish a set of ethical principles for AI anchored in international human rights laws. Build cultural expertise into content guidelines to contextualise principles locally.
2. Create adaptable AI architectures that can be configured for localisation while restricting customization on issues contravening 'core' values.
3. Implement tiered regulation allowing national discretion within global policy frameworks. Require transparency from AI providers on customisation.

### 3. Human Summarisation Outputs

Alongside AI Summarisation outputs, the project team produced human summarisation of deliberations to double check the validity of AI-enabled outputs, as well as blueprint eventual AI-enabled outputs that might inform governance processes beyond the scope of the project.

For this, we asked an analyst team to review the inputs and outputs of online deliberations that took place on pol.is during the period of the pilot, as well as participating in face-to-face discussions with participants and in small-group workshops. Analysts focused on surfacing consensus items, divisive items, and presenting those as potential outputs to policymakers and/or industry.

The project team saw significant value in the use of human summarisation in the creation of governance outputs. We anticipate that in the near term, human-produced outputs will be essential given levels of trust and working norms in regulatory and governance bodies. We believe current LLMs are capable of producing summaries of equal or higher quality than the example shown above, but concerns around trust and familiarity with LLM-enabled outputs will slow their adoption within governance processes. We anticipate that human summarisation will gradually move ‘up the chain’ as familiarity and trust in AI summarised outputs increase.

We use human summarisation to analyse the discussion we had in the face-to-face meeting in Taiwan. Following our summarization of polis responses, the face-to-face meeting is organised under three pivots: data usage & privacy, bias & discrimination, and localization & governance. The result is shown in the table below.

|                                    |   |
|------------------------------------|---|
| <b>Data Usage and Privacy</b>      | <b>Consensus:</b> Most participants agreed that the regulation on data usage and privacy should be renewed to deal with the impact brought by AI.<br><b>Divergence:</b> The boundaries of data usage was heavily argued.      |
| <b>Bias and Discrimination</b>     | <b>Consensus:</b> Most participants agreed that whether an output is biased or discriminated should not be determined by the AI companies.<br><b>Divergence:</b> Which stakeholder can have the final say was heavily argued. |
| <b>Localization and Governance</b> | <b>Consensus:</b> Most participants agreed that it will be better to make users able to switch the output.<br><b>Divergence:</b> Whether AI models should localise the output was heavily debated.                            |

*Table 5: Human summary of the primary themes arising from Taiwan face-to-face*

One proof-of-concept output based on human summarization - an update on an agenda-setting deliberation - is shown below.

# RECURSIVE PUBLIC



Global AI Governance Deliberation Pilot

SNAPSHOT SEPTEMBER 2023

The agenda setter is an invite-only group of 500 academics, members of industry, decision-makers and their networks that tests the temperature of AI debates.

*A snapshot aims to capture the state of a deliberation at a moment in time. It supports decision-making and allows future deliberations to build onto an existing dataset.*

Just under half of our cohort participated, with two broad groups of views, defined primarily by respective focuses on long-term risk and short-term risk.

The agenda setter successfully identified early areas of consensus and areas of disagreement to be moved to deep-dive processes.

|              |       |
|--------------|-------|
| Participants | 214   |
| Votes        | 8,442 |
| Statements   | 342   |

## Controversial and Uncertain Items

There is disagreement about the feasibility and effectiveness of global AI standards.

There is disagreement about the levels of risk involved in AI research.

There are questions around the transparency and consequences of corporate funding of AI governance research.

Next deep-dive processes

International and national governance frameworks and their limits.

Good giving: ethically supporting independent research on AI governance

A consensus on short-term and long-term risk?



# RECURSIVE PUBLIC



Global AI Governance Deliberation Pilot

SNAPSHOT SEPTEMBER 2023

## Consensus Items

- Oversight of training data in AI emerges as a point of focus.
- Governments should clarify liability law, so that legal & financial liability for AI risk is distributed fairly along AI value chains.
- Multistakeholder dialogue - in education, regulation and decision-making - is essential.
- Increased friction and costs in AI development to ensure some level of regulatory or democratic alignment is a price worth paying.
- Long-term existential risk should not blind us to short-term consequences of AI development and deployment.
- Environmental costs of AI are important questions for its governance.

## Actions & Advice

| Policymaker Audience  | OpenAI / Industry   |
|---|---|
| <p>There is significant participant support for AI governance dialogues, in line with the November AI Summit in the UK.</p> <ul style="list-style-type: none"><li>• There should be clarity about the summit's legacy: will this become a regular event on a COP model?</li><li>• To date, the participation of civil society, academia and other non-governmental voices has been limited - addressing this should be a priority.</li></ul>                                | <p>Explore routes to increasing transparency around training data used by industry models, including policymaker- and public-friendly communication.</p> <p>Clarify existing legal uncertainties facing industry as part of a public conversation.</p> <p>There remains significant appetite to understand short-term consequences across civil society, and industry risk strategies should maintain this in spite of media focus on long-term risk.</p> |
| <p>There is consensus on short-term policy priorities that may allow for quick wins. Decision-makers should move ahead with:</p> <ul style="list-style-type: none"><li>• Clarifications on liability both to support industry in legal uncertainty around current data collection practices and to defend rights holders where appropriate.</li><li>• Consultations on existing data regulation and its applicability to transparency and audit of training data.</li></ul> | <p>Make efforts to clarify independence of any non-profit donations, including reviewing best practice on philanthropic giving.</p>   |

2

Figure 11: Human produced output example

# FUTURE DEVELOPMENT

The Recursive Public piloted both an overarching process for gathering large-scale democratic inputs to AI decision-making and aspects of the technological infrastructure underpinning that process.

There is a large and willing community of expertise that has participated in the *Recursive Public* process that could add significant value to ongoing experiments in democratic inputs to AI. Should there be capacity and resource, we encourage industry to involve them in ongoing pilots and experiments. Development of a follow-up pilot could bring together a handful of selected communities - youth, expert, in-house and regional - around a single concrete decision made by industry, infusing the *Recursive Public* process with limited but novel decision-making power.

The project also identified a number of technical developments that we believe will benefit from further exploration and development. These included:

- **High-Quality Labelling.** TttC was able to provide helpful labels and summaries for each cluster, but these were not yet as concise and precise as the ones that were produced manually in the previous section. To give one example, TttC was initially proposing to use very similar labels to two different clusters: "AI and Environmental Sustainability" and "AI in Environmental Sustainability and Conservation". After looking into the key language/phrases that discriminate between these clusters, we were able to understand the nuanced difference and decide to label them respectively as "Conservationism" and "Techno-environmentalism". Work on this tool will continue and experiments like this are useful for fine-tuning.
- **Automated exploration.** In the future, the project team hopes to implement techniques to automatically dive into clusters that require further analysis or differentiation, with the ability to break them down further.
- **The use of machine translations to enable multilingual deliberations.** Democratic inputs to AI are collected by aggregating not just one but many consultations, run in diverse geographic regions, following different formats and processes which would be adapted to local cultures to reduce friction and give everyone a real chance and the practical means to participate.
- **Greater contextualisation.** Argument extraction and clustering techniques are already effective, however, more information about, for instance, demography, would allow for new ways of characterising semantic clusters and facilitate deeper analysis.
- **Automated recursion.** A more open-ended problem is how to manage the movement from wide deliberations seeking to identify priorities, to *narrow* deliberations focussed on a specific question. It may be possible to create workflows that identify a newly emerged agenda-item, for instance, and then immediately suggest a series of questions to inform a narrow deliberation on that item. The use of AI could enable this to happen in quick succession, reducing timelines for processes and likely improving the chances for wider participation if not providing an additional incentive, like pay to participate.

- **Open-sourcing the methodology and process for wider review.** In-line with the transparency principle of vTaiwan, it would also be beneficial for the technical and methodological details of the Recursive public to be open-sourced to both AI governance and technical communities for review, critique, challenge and constant improvement.<sup>6</sup>

---

<sup>6</sup> The final clustering report follows clear transparency guidelines, and explains clearly the steps taken in its generation. <https://tttc.dev/recursive>