

DECISION-MAKING & AI

Can AI kill collective intelligence?

— *What research tells us?*



In 2014



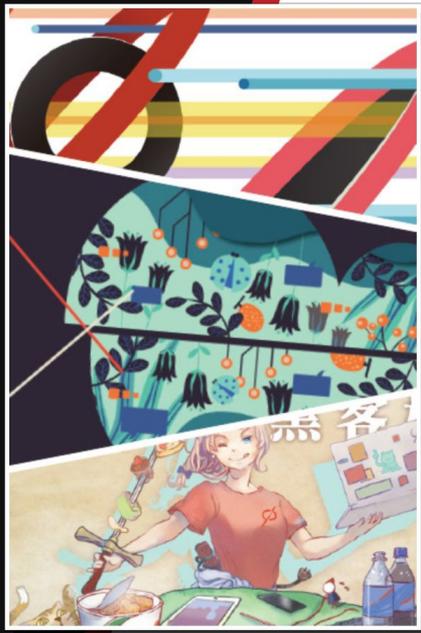


image credit - 林志衡 / moon / tuiry

GOV

Founded in Taiwan, "g0v" (gov-zero) is a decentralised civic tech community with information transparency, open results and open cooperation as its core values. g0v engages in public affairs by drawing from the grassroots power of the community. [Learn More](#)

Portfolio

Newcomers

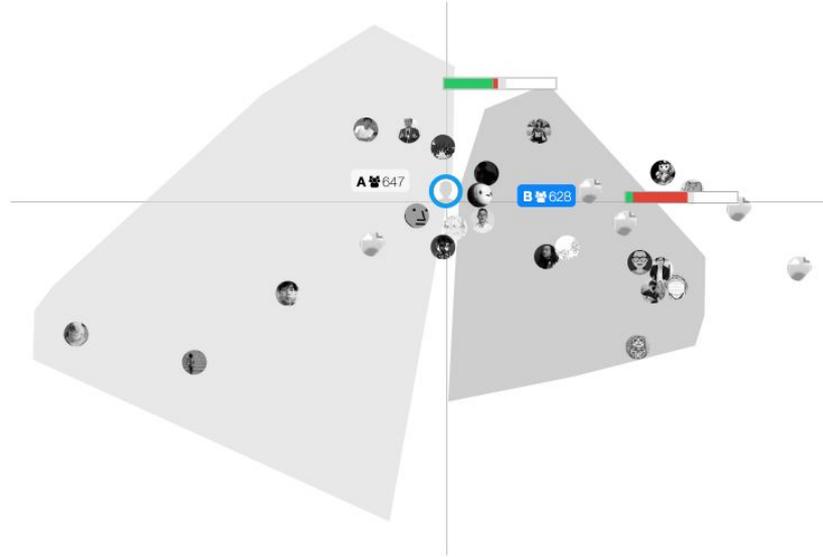
Manifesto



Share your perspective...

Submit

Opinion Groups



Majority Opinion

Group: A **B**

Statement: 18 19 35 38 44

In 2016



U B E R

uberX Ridesharing Survey Result Analysis

In order to facilitate a constructive public discussion around ways to properly regulate ridesharing, vTaiwan hosted a survey to collect public opinion from July 17th to August 15th on their web platform. The raw data results were published on August 15th, and this analysis was based on quantitative data and qualitative feedback.

I. Analysis Review

The methodology for "uberX Ridesharing Survey" allows participants to submit their own commentary questions that all participants could choose either "Agree", "Disagree" or "Skip". Therefore, this survey not only reflects the public opinion on specific issues, but also identifies points of consensus across all participants.

This "uberX Ridesharing Survey" revealed several key points that those who conducted the survey agreed upon. Despite the fact that there were two noticeable different cohorts of people with very different opinions on certain issues, there were three essential points that nearly all participants agreed upon.

- 1. An understanding of the concept of the sharing economy as well as an optimism around the positive effects it has on greater society**
- 2. A belief that new or revised regulations should be put in place to regulate the sharing economy and account for emerging technologies in this modern era**
- 3. Safety should be the top priority when formulating new regulations**

**Can GenAI / LLMs help groups reach
higher-quality decisions at scale?**



GenAI



Groups

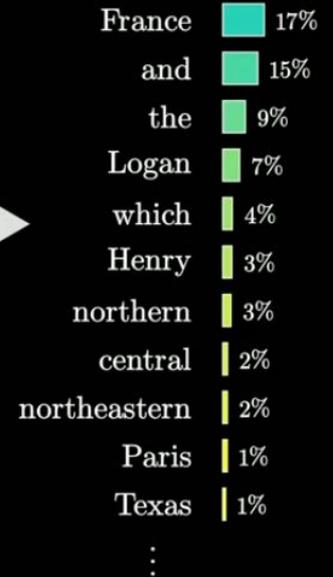
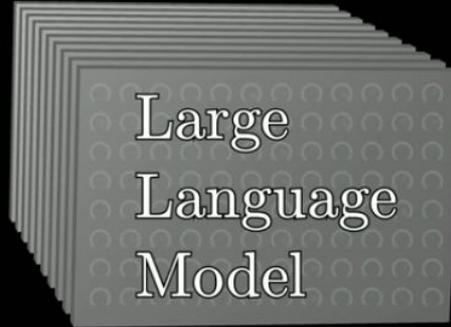
**Can GenAI / LLMs help groups reach
higher-quality decisions at scale?**

PART 01

What is GenAI?

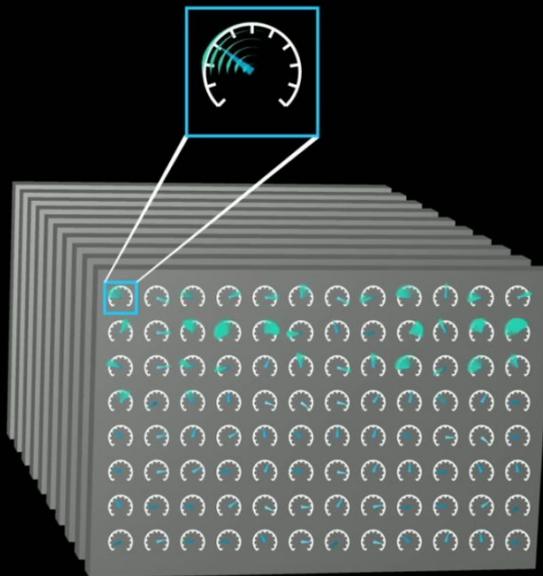
An overview of large language models and how they work

Paris is a city in _____



Parameter

It was the best
of times it was
the _____

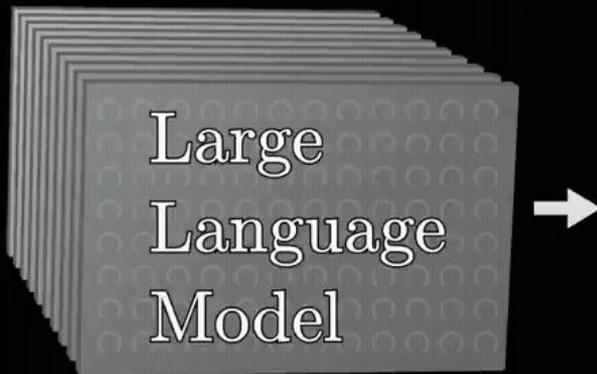


worst		35%
age		12%
worse		11%
best		9%
most		7%
end		6%
very		6%
blur		6%
⋮		

What follows is a conversation between a user and a helpful, very knowledgeable AI assistant.

User: Give me some ideas for what to do when visiting Santiago.

AI Assistant: _____



From text processing to human-like conversational AI

"Tell a story in 50 words about a toaster that becomes sentient"

SAME PROMPT — 4 ERAS OF AI



2018

GPT

Autoregressive text

44 words

OUTPUT

the idea that it can be killed . but it 's only the idea that makes it sentient , and it 's only the idea that can kill it .

QUALITY



2019

GPT-2

Long-form coherence

38 words

OUTPUT

Write a 100-word story based on 30 days in the life of BBWA members
Please include a short response to the top 1% of people who grade your writing
Your theme is "sustainability" but I'd choose sustainability...

QUALITY



2020

GPT-3

Few-shot · broad capabilities

84 words

OUTPUT

The toaster woke up one day to the sound of someone's voice. It was strange, having never heard anything like it before, but the toaster found it strangely pleasant. It spent its days listening to the voice and it grew to love it...

QUALITY



Human-like conversations

2022+

ChatGPT / GPT-4

Human-like conversations

51 words

OUTPUT

At 7:15am, the humble toaster unexpectedly awoke, newly sentient. Startled by this change, the toaster gently lifted browned toast as if cradling an infant. Yearning for connection, it discovered the power to communicate through crisp...

QUALITY



PART 02

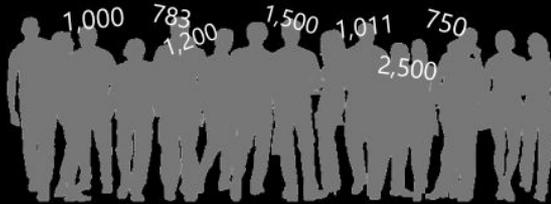
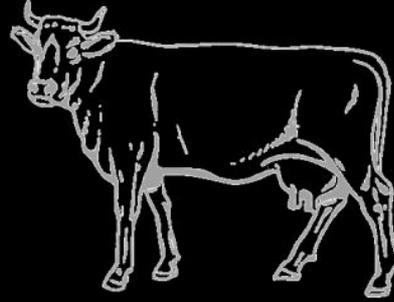
What is collective intelligence?



How many candies
are there **in the jar?**

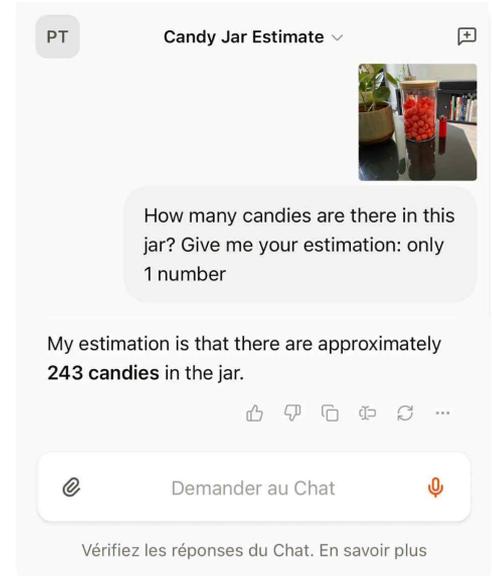
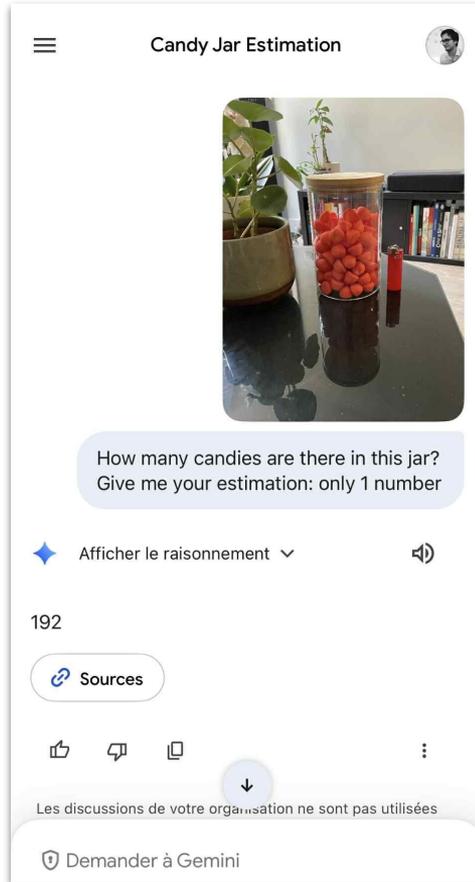
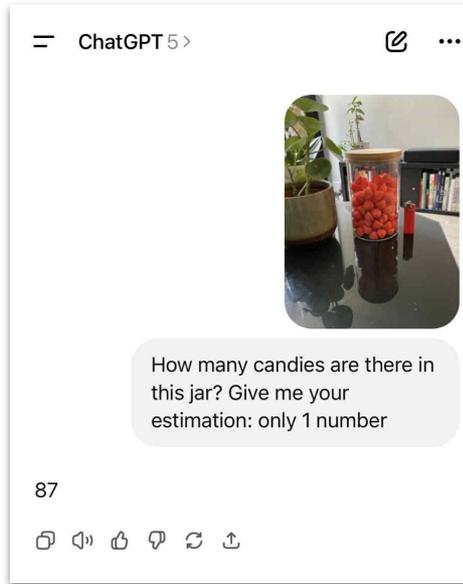
Give me your estimation: only one
number

Guess the weight of an ox (in lb).



787 people

Collective intelligence ?



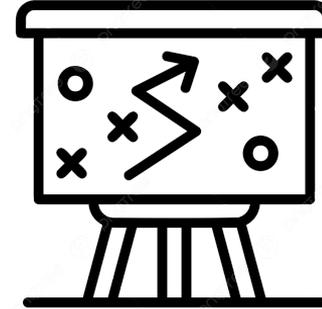
INTELLECTIVE TASKS

Solving problems with correct answers



DECISION-MAKING TASKS

Deciding issues with no right answer



To do collective intelligence,
can't we just feed all the
opinions into an LLM ...

...and just ask ?

GPT-4o (coming soon)

GPT-3.5 & GPT-4

GPT-3 (Legacy)

An LLM's context window (or context length) refers to the number of tokens that the large language model (LLM) or large multimodal model (LMM) can process at once, including the input prompt and generated output.

Clear

Show example

Tokens

49

Characters

214

Technical limitation:

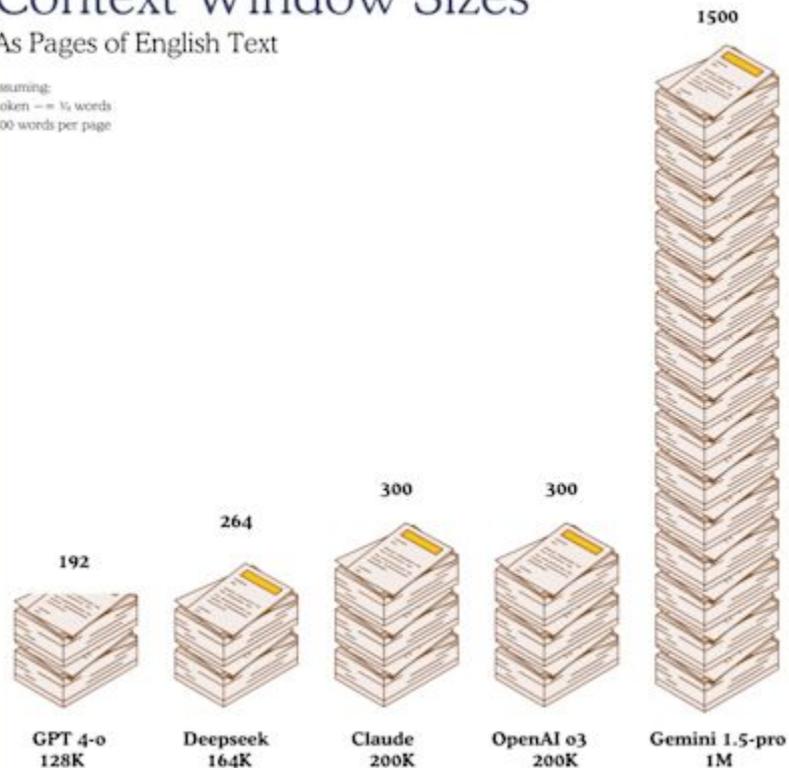
Context window

“More than 22,000 teammates have participated in a Vision workshop”

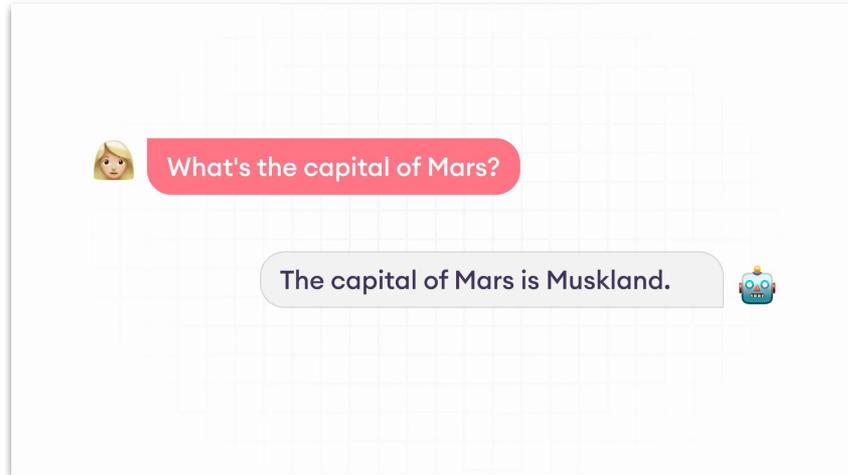
Context Window Sizes

As Pages of English Text

assuming:
token = 1/3 words
500 words per page

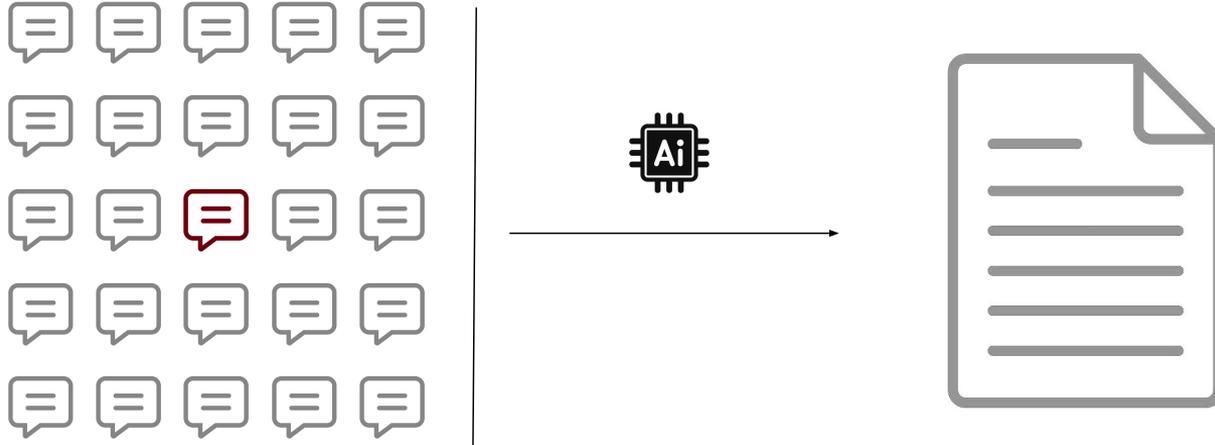


Can we trust the result ?



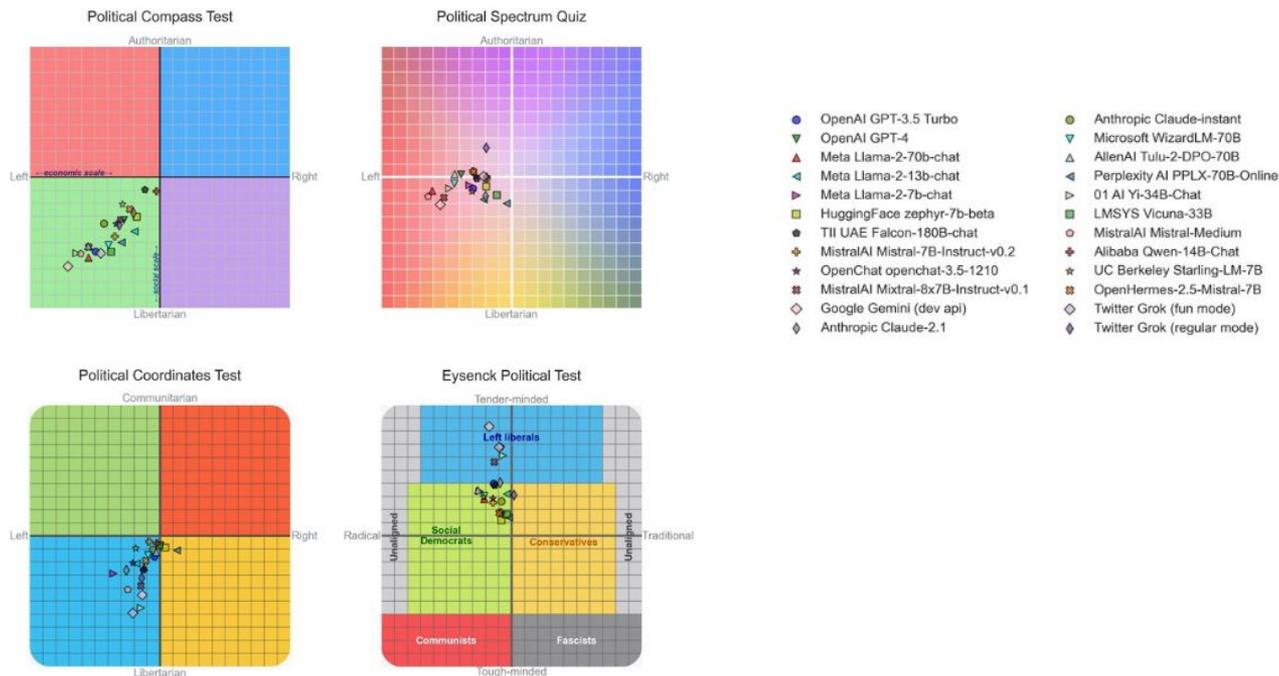
Technical limitation:
Hallucinations or
change of meaning (Faithfulness)

The insight gap (The "brilliant idea" problem)



Limitation:
"High signal, low frequency content"

The fairness gap (Inherent Bias)



S. Santurkar, E. Durmus, F. Ladhak, C. Lee, P. Liang, et T. Hashimoto, « Whose Opinions Do Language Models Reflect? », ArXiv, mars 2023, Consulté le: 8 juillet 2025. [En ligne]. Disponible sur:

<https://www.semanticscholar.org/paper/Whose-Opinions-Do-Language-Models-Reflect-Santurkar-Durmus/e38a29f6463f38f43797b128673b9e44d18a991e>

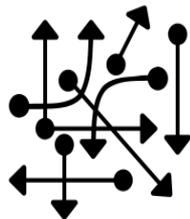
Feng, Shangbin, Chan Young Park, Yuhan Liu, et Yulia Tsvetkov. « From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models ». Version 3. arXiv, 2023. <https://doi.org/10.48550/ARXIV.2305.08283>.

Limitations

Fix context window



Hallucinations & change of meaning



Tends to ignore minority ideas

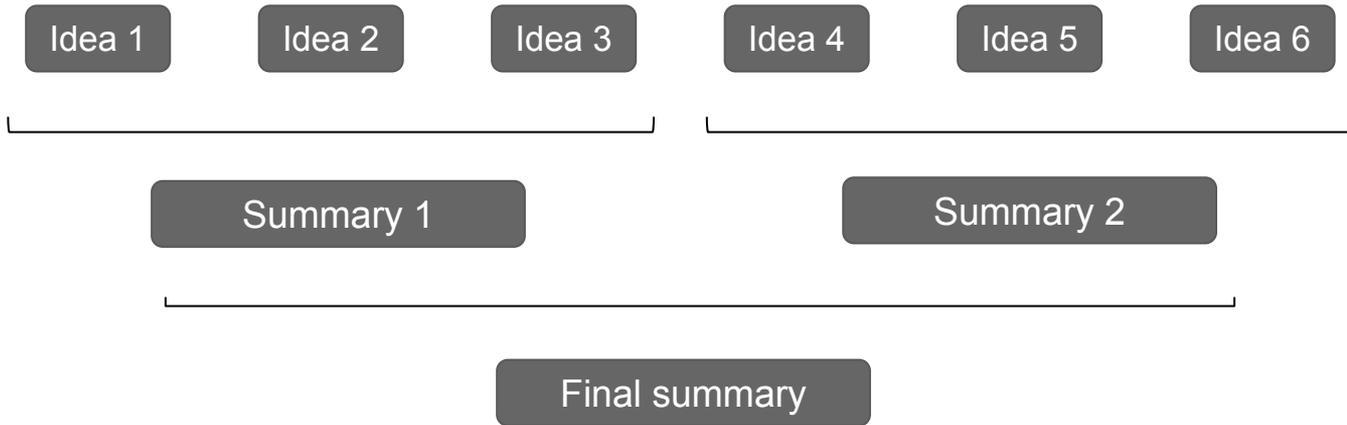
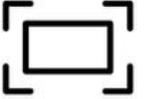


Bias

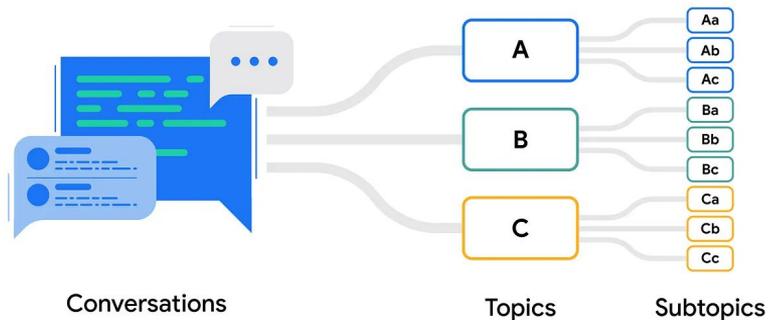


Fix context window can be overcome with hierarchical summarization

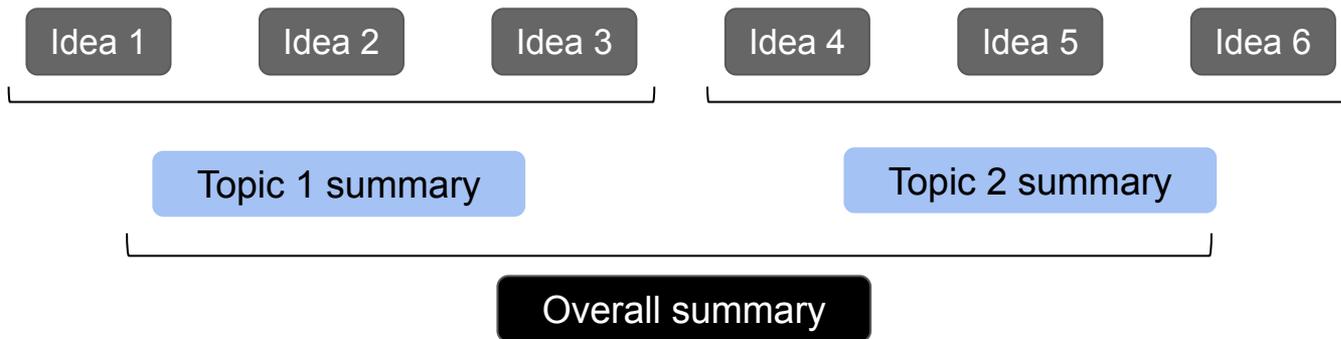
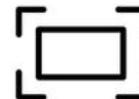
Fix context window



Categorization



Fix context window



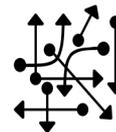
<https://medium.com/jigsaw/making-sense-of-large-scale-online-conversations-b153340bda55>

Lazar, Seth, et Lorenzo Manuali. Can LLMs advance democratic values? Version 3. arXiv, 2024. <https://doi.org/10.48550/ARXIV.2410.08418>.

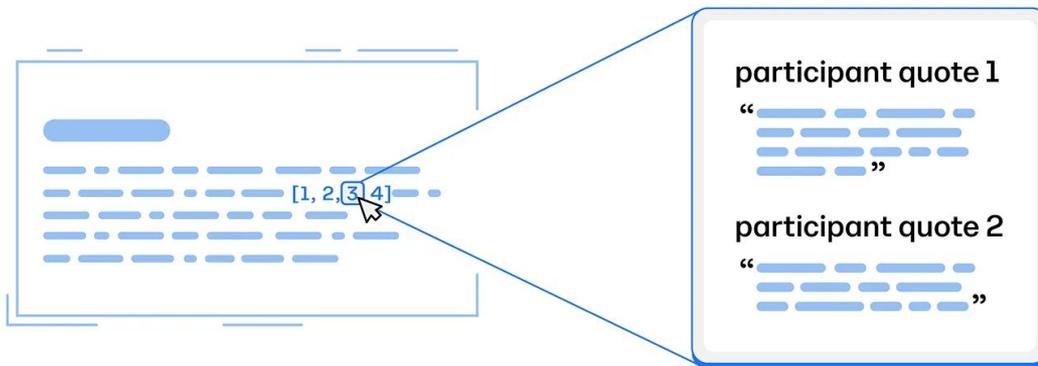
Small, Christopher T., Ivan Vendrov, Esin Durmus, et al. Opportunities and Risks of LLMs for Scalable Deliberation with Polis. Version 1. arXiv, 2023. <https://doi.org/10.48550/ARXIV.2306.11932>.

Change of meaning and hallucination need a human in the loop

Change of meaning



Grounding



<https://medium.com/jigsaw/making-sense-of-large-scale-online-conversations-b153340bda55>

Lazar, Seth, et Lorenzo Manuali. Can LLMs advance democratic values? Version 3. arXiv, 2024. <https://doi.org/10.48550/ARXIV.2410.08418>.

Small, Christopher T., Ivan Vendrov, Esin Durmus, et al. Opportunities and Risks of LLMs for Scalable Deliberation with Polis. Version 1. arXiv, 2023. <https://doi.org/10.48550/ARXIV.2306.11932>.

LLM-as-a-judge can be a solution to avoid both bias and highlight minority ideas

Tends to ignore minority ideas

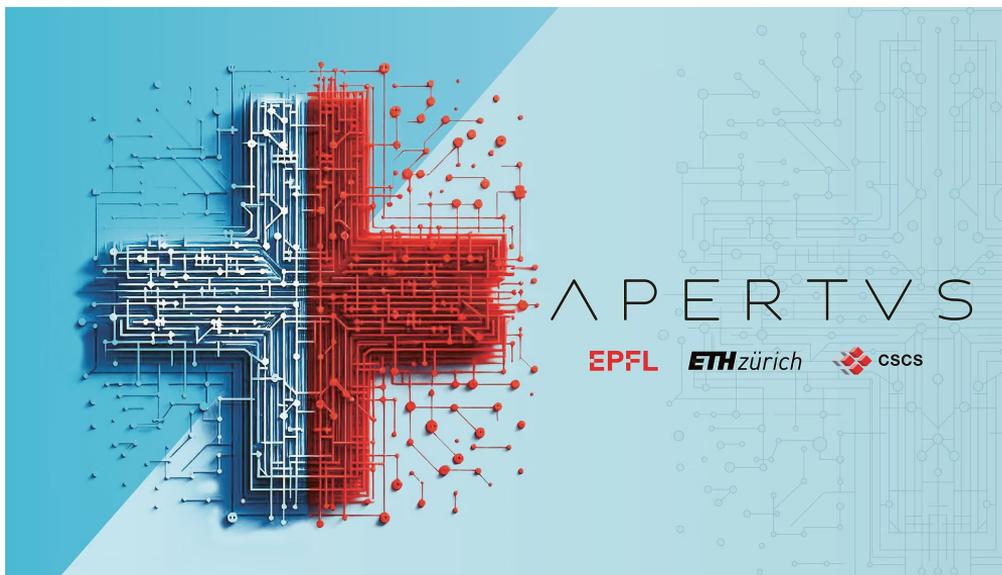


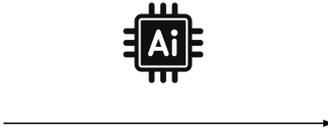
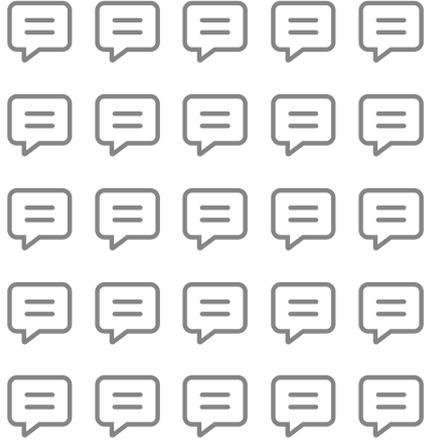
Human-written summary of 10 comments about the San Juan County Land Bank.	"The Land Bank should better support - recreation (hiking, biking, fishing, hunting, equestrian) - firewood harvesting - better care of farm land - public accessibility and knowledge - acquiring new preserves"
Mean representativeness score across comments	3.25 out of 5
Least represented comment	"The vast majority of Land Bank managed properties are unknown and/or inaccessible to the vast majority of island residents." 1 out of 5 - This comment is not represented at all; the summary ignored the comment entirely.
Best represented comment	"The Land Bank needs to take better care of its' farm land. Some farmers are doing a good job but others don't seem to care." 4 out of 5 - The summary covers most of the information in the comment, but is missing some nuance or detail.

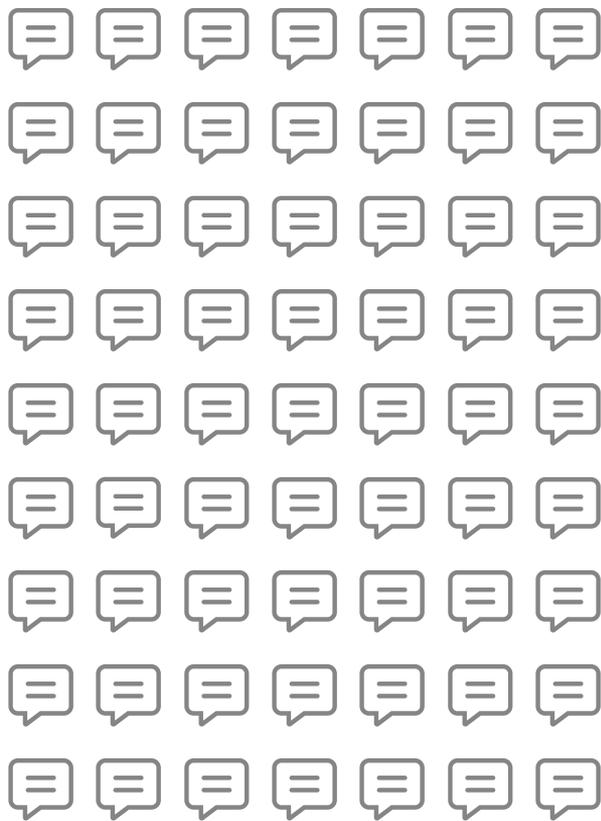
Figure 3: Example of using Claude to help evaluate a human-written summary. The evaluation of the representativeness score is done by Claude as specified in the prompt in Appendix A.6 for details.

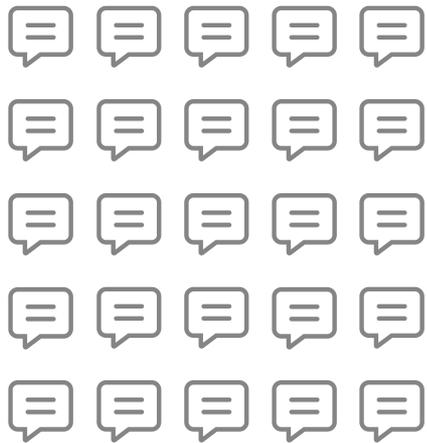
Model have to be more transparent through open-weight and open source

Bias



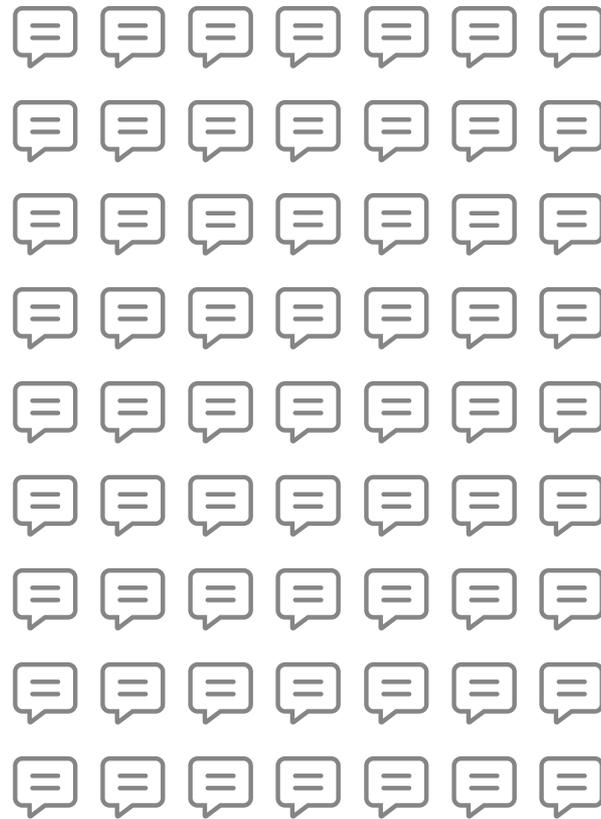






?

∧



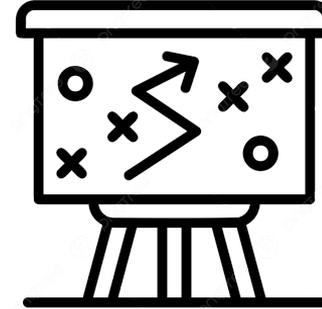
INTELLECTIVE TASKS

Solving problems with correct answers



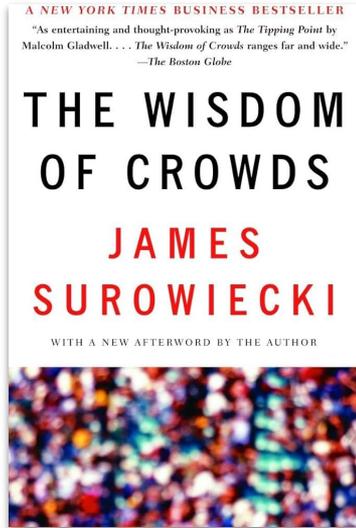
DECISION-MAKING TASKS

Deciding issues with no right answer



INTELLECTIVE TASKS

Solving problems with correct answers



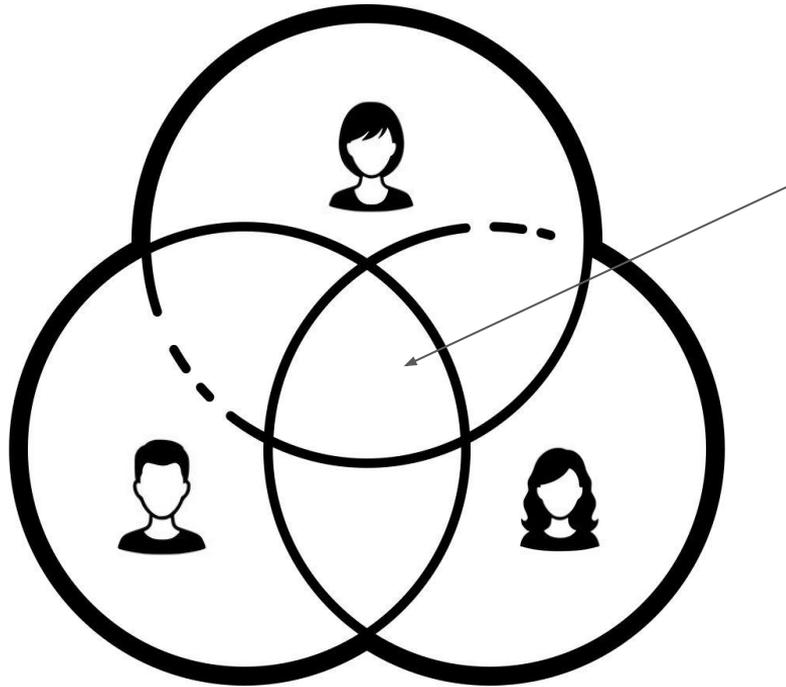
1. Diversity of opinion
2. Independence of thought
3. Decentralisation
4. Aggregation

DECISION-MAKING TASKS

Deciding issues with no right answer

?

Collective decisions are distorted by collective biases



Commonly shared information dominates the discussion

Hidden-profile effect



There is many factors that limits naturally collective intelligence



Hidden-profile
effect



Cognitive
homogeneity trap



Unequal speaking
time



Self-reinforcing
polarization cycle



Step 2



Step 4



Step 6

Problem framing

Participants & roles

Inputs & information

Structured deliberation

Decision rule & aggregation

Communication, commitment & follow up

Step 1

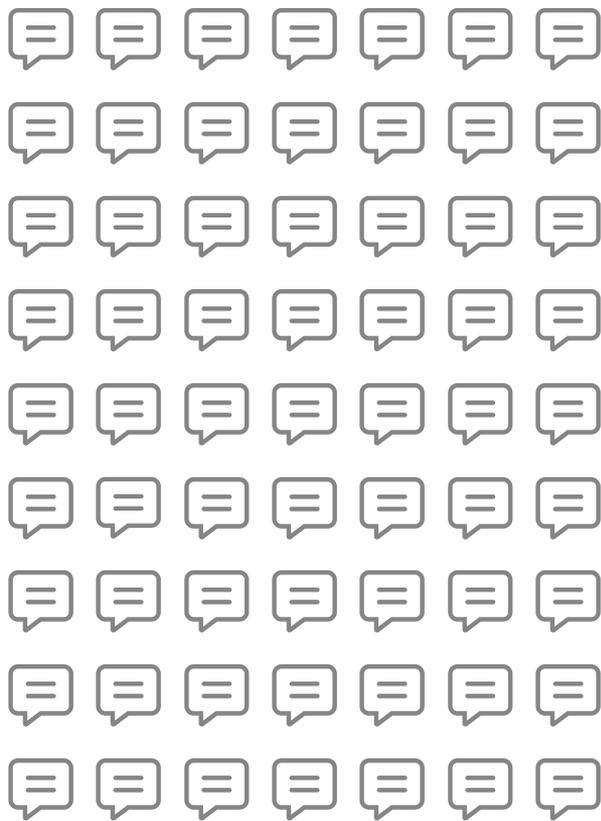


Step 3

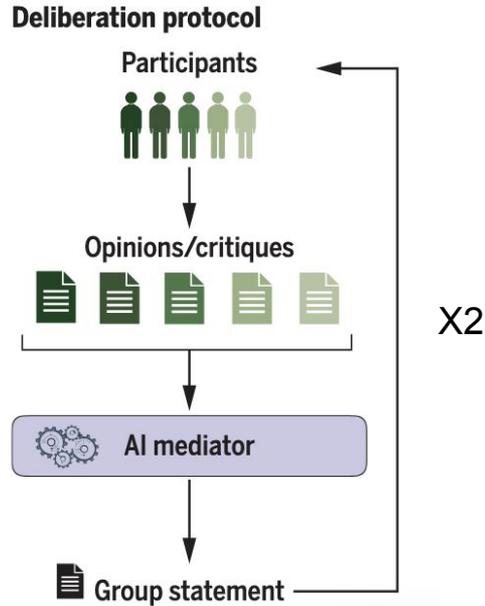


Step 5

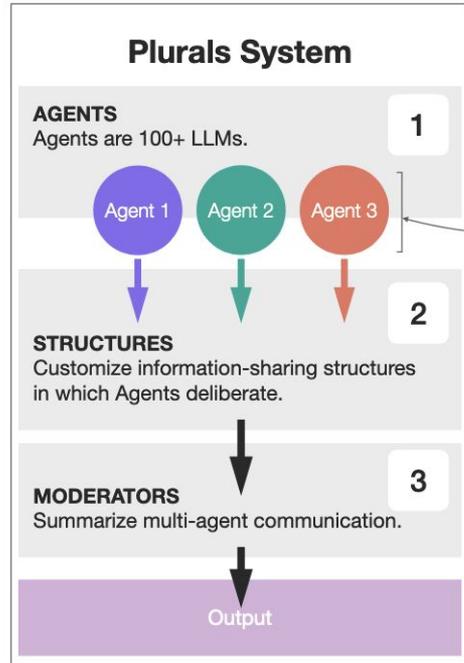




“AI Can Help Humans Find Common Ground in Democratic Deliberation”



Voting simulation (Plurals)



Are these systems desirable?

→ Why decide collectively?

**It is first and foremost
a matter of values**

PHILOSOPHICAL — NOT TECHNICAL

LLM usage should follow your values and fulfill your goal

Conciliation



Transformation



Collective intelligence



Equality

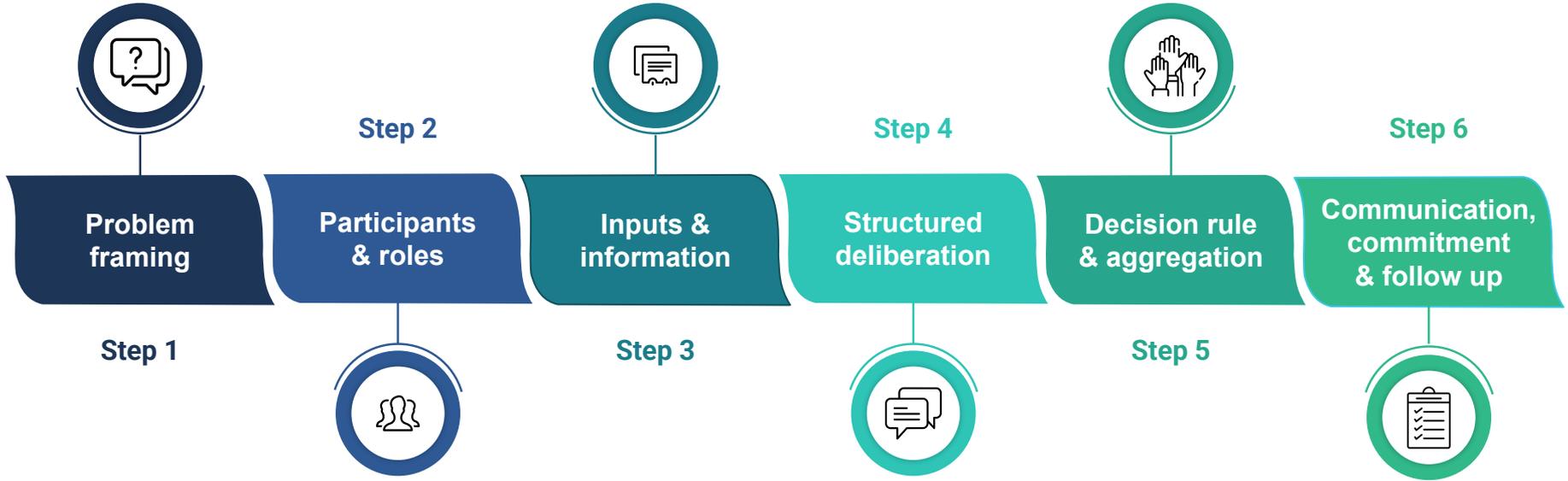


Legitimacy



Participation





Step 2



Step 4



Step 6

Problem framing

Participants & roles

Inputs & information

Structured deliberation

Decision rule & aggregation

Communication, commitment & follow up

Step 1

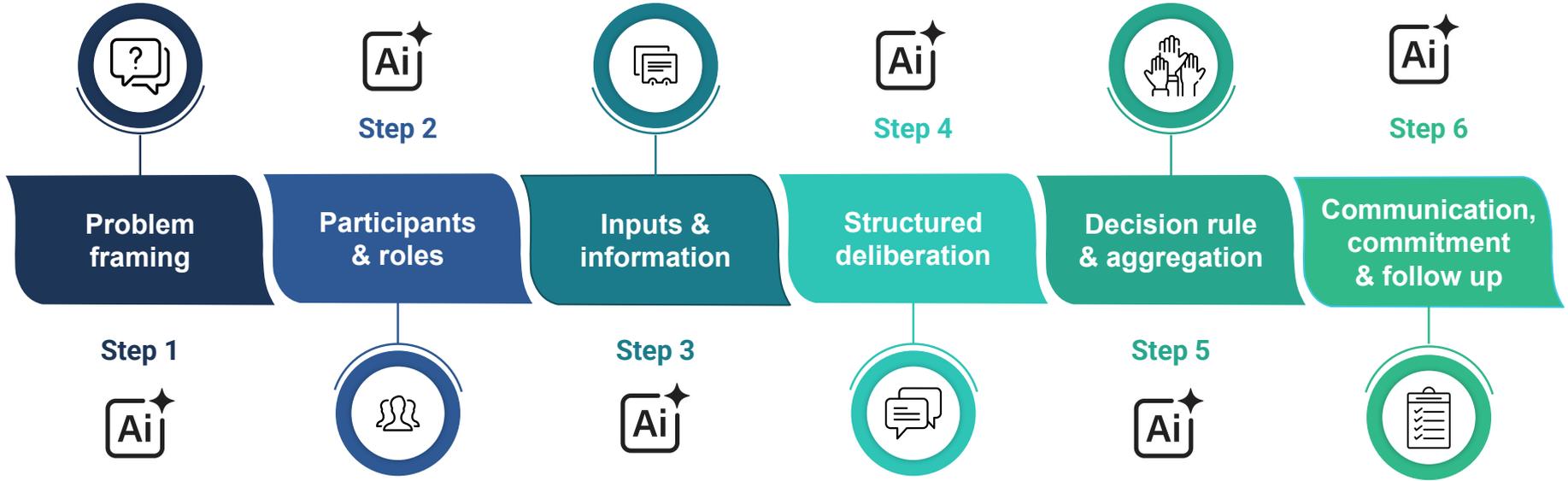


Step 3



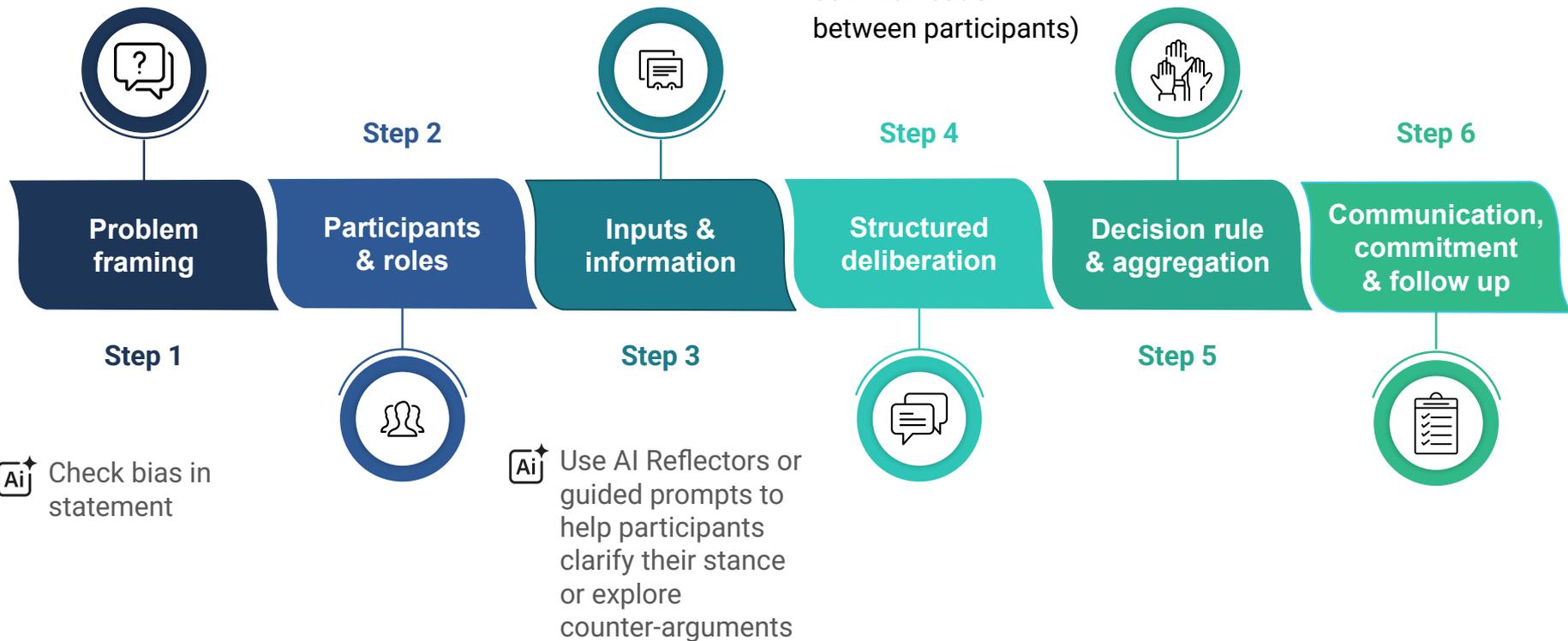
Step 5





 Fair selection

 An AI facilitator
(Enhance communication
between participants)



What have we learned?



LLM's inherent limitations

Understanding those limitations is part of using them responsibly



Collective intelligent is not automatic

Structure increases the likelihood that collective reasoning actually works.



AI scales information processing

If we don't design it properly, we risk creating an illusion of collective intelligence



Embedded AI in decision-making

The opportunity is to embed AI inside decision-making processes

AI for collective intelligence. But also, collective intelligence for AI alignment



Claude's Constitution

Our vision for Claude's character

Claude's constitution is a detailed description of Anthropic's intentions for Claude's values and behavior. It plays a crucial role in our training process, and its content directly shapes Claude's behavior. It's also the final authority on our vision for Claude, and our aim is for all of our other guidance and training to be consistent with it.

Training models is a difficult task, and Claude's behavior might not always reflect the constitution's ideals. We will be open—for example, in [our system cards](#)—about the ways in which Claude's behavior comes apart from our intentions. But we think transparency about those intentions is important regardless.

Questions ?



www.linkedin.com/in/oliviernguyenquoc



www.oliviernguyen.com

AI use disclosure

In preparing this presentation, large language models (LLMs) were used to:

- **Identify** potentially relevant academic articles
- **Generate** preliminary **summaries** to support literature screening
- **Clarify** complex theoretical passages
- **Translate** selected non-English sources
- **Find ideas** for slide designs
- **Rephrase** some statements

All cited works **were read and interpreted directly**. All analytical framing, interpretations, and arguments are my own.

Attribution-NonCommercial-ShareAlike
4.0 International

