Author's copy. Authoritative copy to appear in J. Haqbeen et al. (Eds.): IJCAI 2025, LNAI 16400, pp. 1–20, 2026.
https://doi.org/10.1007/978-981-95-9667-6_3

Democracy as a scaled collective intelligence process: points of vulnerability and augmentation

Marc-Antoine Parent[0000−0003−4159−7678]

Conversence, Montréal, Canada
maparent@conversence.com
https://www.conversence.com

Abstract. The ideal democratic process aims to solve complex social decision problems, where diverse communities have potentially conflicting goals, through public deliberation, based on transparency, accountability, and trust. In practice, there are scaling limits to deliberative processes, both in terms of number of participants and cognitive complexity and democratic societies fall short of the ideal. Use of generative AI has been proposed to replace or augment the deliberative process. We argue that generative AI is an opaque process by nature, and hides issues with bias, power, accountability, and trust; and as such should not be directly involved in the decision-making process. We propose an alternate path, where decision-making is grounded in a fully transparent collective intelligence process, using a decision-oriented global structured knowledge base. Hybrid AI could help people approach and contribute to such a knowledge base, and watch over the coherence of expectations, actions and goals.

Keywords: Augmented collective intelligence, social learning, hybrid AI, deliberative process

1 Introduction: What roles for AI in the democratic process?

With the popularization of generative AI, we have seen many proposals to use it to streamline many processes, including democratic governance or governance at large. Examples range from efforts to use generative AI to synthesize citizen consultations [10] or even facilitate them [32], rewrite legislation [33], engage with the public on behalf of politicians [16], etc. Schneier [51] gives many more examples.

There are also a proposals to replace the public service, and even democratic governance altogether with AI-driven processes. [20] This is based on a criticism of existing democratic institutions, which we want to analyze.

We propose to ask: what issues are these proposals trying to solve? Are their criteria of success aligned with those of democracy? What are those anyway?

What do we even mean by democracy? We will focus on two aspects of democracy, as a rational governance process and as a social negotiation, each with their distinct aims. We think identifying those will make it clearer where and how it is most and least appropriate to involve artificial intelligences in the process. In doing so, we will reuse the terminology for dimensions of democracy defined in [41].

2 Democracy as a rational governance mechanism

Democracy is a specific form of social governance; we will start with a definition of what we mean by that term.

2.1 Governance is a decision-making process

Decision making can be modelled as follows: Given a known situation, an agent can take one of many actions. Each action has a range of expected consequences, according to a causal theory. Each consequence is valued, i.e. estimated to have a certain utility, according to goals expressed as a set criteria. The act of decision can be thought of as choosing an action that maximizes aggregate utility according to those expectations and criteria.

In a collective setting (whether organization or political), every individual bases decisions on an heuristics, which is based on a simplified causal theory, but also on a collective agreement about the most appropriate actions to take in a given situation. These agreements can be implicit (cultural norms, [46]) or explicit (rule of law).

2.2 Governance involves continuous learning

Decisions do not always yield the hoped-for outcome, and every decision outcome is an opportunity to correct one's action. In some cases, the outcome is not deeply surprising, and it is possible to correct course directly, without revising one's causal theory, or with minor tunings of likelihood estimates (L1); in other cases, the change requires revising the underlying abstract causal schema (L2), or even re-evaluating the goals (L3); in the more radical cases, the process through which we imagine and evaluate competing causal theories must itself be re-evaluated (L4). These layers of learning have been called single, double and triple-loop learning respectively, by diverse researchers, though the definitions sometimes overlap. (My L4 is usually included in the third loop.) [55,24,8,2]

At any scale, the individual or collective's success depends on the accuracy of their heuristics. As we accumulate experience in a complex world, we realize simplified heuristics cannot at the same time reflect that complexity accurately, and fit within the bounds of any individual's cognitive capacity. This means both that it requires greater cognitive capacity to maintain the heuristic's factual accuracy, and also that individual decisions are constrained by the complexity of the heuristics.

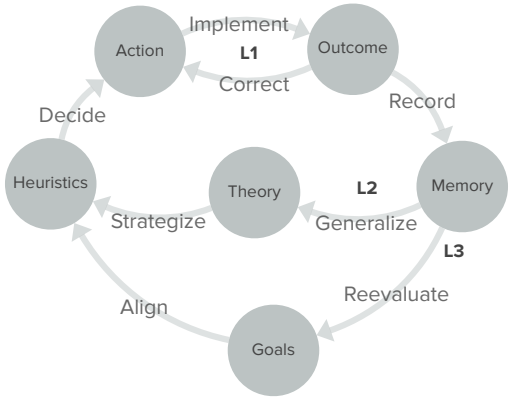

Fig. 1. Decisions and learning loops

Criterion 1 (Accuracy). Actions should be guided by heuristics that reflect a causal theory that is accurate enough.

Criterion 2 (Legibility). The decision heuristics should be simple enough to be understandable by members of the collective.

Criterion 3 (Cognitive Capacity). Cognitive capacity should be adequate to generalize a theory from the learned experience, and distill it into a heuristic aligned with the goals.

Criterion 4 (Learning). Experiences should be captured, and be available in a shared memory. There should be a continuous effort to maintain that memory, and learn from it.

Trade-off 1. Thus, we can identify a first trade-off, between Accuracy and Cognitive Capacity.

2.3 Collective intelligence

Deciding entities have to make decisions from partial knowledge of the situation, using an incomplete causal theory [1] built from limited experience, using limited rationality, subject to cognitive biases[34]. Fortunately, these limitations can be alleviated by drawing on collective intelligence.

[1] Assuming the situation's causal regime even allows predictions, as opposed to chaos or emergent dynamics, [52]

Many teams and small tribes take decisions after deliberation. It has been shown that deliberation improves the Accuracy of decisions, under certain conditions. Those conditions include a shared goal and sufficient diversity to avoid groupthink. It has even been hypothesized that the evolution of human cognition was mostly driven by its usage in a social context, as confirmation bias can be a heuristic leading to an efficient division of cognitive labour, and the resulting inaccuracy gets corrected by iterations in a social context [39].

A more abstract argument can be made in favour of deliberative cognition: As originally stated by Ashby [3], any entity (individual or collective) that wants to adapt to the world should have adequate inner complexity to have the capacity to detect, understand (a causal theory), and react to the states of the world that affect it. Again a group with sufficiently diverse expertise and points of views is more likely to have the requisite variety. So there is a diversity angle to Cognitive Capacity.

Criterion 5 (Diversity). Accuracy relies on an adequate diversity of expertise and points of view in the conversation

On the other hand, there is a communication overhead to deliberation, which scales quadratically with group size, and even a diverse group can succumb to groupthink. This is why, beyond a certain scale, the various functions of the decision and learning process are handed to separate institutions, as discussed in section 3.1.

Criterion 6 (Overhead). Decision processes should not become bogged down in coordination and communication overhead.

3 Democracy as a socio-political process

Democracy is also a socio-political process, where diverse publics, with sometimes conflicting interests, negotiate a shared way of life [22,40]. A large part of the democratic process in particular is determination of a set of coherent social goals from the aggregated goals of all sub-communities, or all citizens. It is also possible that the goals of sub-communities are too profoundly different for any set of common goals to satisfy any of them, which may lead to social fragmentation[2].

Criterion 7 (Cohesion). The goals of sub-communities in the collective remain compatible enough that the sub-communities engage in the higher-level goal of trying to align them, rather than splitting the community.

Given enough cohesion, the next step in the democratic process is choosing a strategy for collective actions, so they align with collective goals. Again, shared

[2] Often in a feedback loop of antagonism, first described by Bateson in [7] as schismogenesis

heuristics and social norms allow coordinated actions without Overhead of coordination [50]. This is why, in a social context, people give value to following norms, rather than optimization for individualistic best outcome [46].

Criterion 8 (Alignment). Heuristics and subsequent decisions should be aligned with stated collective goals.

According to Rawls [48], it is possible for subgroups in a pluralistic society to agree on actions without agreeing on underlying goals, leading to what Rawls describes as a *modus vivendi*. From there, trust can build in time, and subgroups can identify local areas of agreement on goals from which to build further trust, even though global agreement remains provisionally out of reach. A society can function using this *overlapping consensus*. As such, both coherence and alignment are often a continuous work in process.

But the important point here is that the process itself matters: citizens and communities learn to trust the shared rules and social conventions insofar as they feel they are actively involved in the processes that shape the social consensus and its implementation.

Criterion 9 (Participation). The public can and does get involved in the various steps of the decision and learning loop[3].

Of course, this involvement only makes sense if the process implementation actually follows the heuristics (rules or norms) that have been decided through goal alignment, strategy, goal aggregation and knowledge generalization. In the words of Elliot Higgins of Bell¿ngcat [30], democracy needs:

Criterion 10 (Verification). Processes for truth-checking, and for checking that actions follow the stated rules.

Criterion 11 (Deliberation). Ideas are discussed in public forums

Criterion 12 (Accountability). A process to respond to decisions that fail (whether with sanctions or revised learning)

Processes and learning It can be argued that the type of social, collective learning involved here is of a fundamentally different type than the factual learning involved in updating a causal theory, or discovering implementation strategies. This social learning can be situated in what Weber called the lifeworld (*lebenswelt*) [58], the fabric of socialization, social integration, and reproduction of culture and society. This reproduction is both transmission of social learning and a continuous self-redefinition, what Castoriadis called autonomy [12],

[3] See conditions and criteria for democratic processes in [41], which defines: Representation, Informedness, Accuracy, Deliberation, Substantiveness (2), Robustness, Legibility, Commitment, Integration, Ability to bind, Awareness, Participation, 12, and Buy-in.

or Varela called autopoïesis [38]. Weber, and later Parsons and Habermas, opposed the lifeworld to what they called the system, social subsystems dedicated to short-term utilitarian optimization. They both argue that the system has its place, but should not "colonize" the transmission processes of the lifeworld [28,29].

3.1 Institutions

The institution of representative democracy is a specific institution that attempts to avoid oligarchy through periodic replacement of the deciding bodies, themselves arranged in a balance of power. This replacement function is only one accountability process among many steps in the social learning loop, and there are complementary institutions that ensure other steps. Much is made of the separation of powers into legislative , executive and judicial branches; those would map to the alignment, implementation, and correction steps in the decision diagram. But social decision and learning involves many sectors of society, and those heuristics which are encoded as law are but a fraction of social decisions. To name a few:

- Bureaucracy: involved in the implementation phase of decisions, but also collecting memory (L1)
- Academia: involved in building the causal theory from accumulated experience (L2)
- Journalism: directs public attention towards discrepancies between goals and outcomes, effectively triage-ing the correction function (L1-4)
- Auditors: Identifying issues that need to be corrected (L1)
- Civil society: Either advocating for the correction of issues or implementing them (L1, L3)
- Citizen assemblies: Tasked with checking the alignment with the heuristics with sufficient requisite diversity (L3)
- Artists: Reshaping the narratives around certain goals, through which cultural norms adapt (L3)

The bureaucracy is an important case in point. Ostensibly, is there to carry out decisions of a deciding body. But more importantly, it connects a smaller deciding body, such as those of representative democracy or citizen assemblies, with a body of domain experts, either from the scholarly community or trained by them, who will advise the deciding bodies when relevant (1). As such, it can augment the cognitive capacity of the deciding body to match the complexity of the issues being considered (3). It also acts as a specific shared memory for decisions and goals, allowing memorization and Verification, and at its best can even ensure that different aspects of a complex and extensive strategy do not work at cross-purpose.

Criterion 13 (Coherence). Social actions should not be at cross-purpose. Ultimately, expected consequences should be considered based on global combination of actions, not on individual actions.

Dividing the work in this way allows for specialized teams to take on specific challenges, but the complexity of coordination is pushed to the inter-group level. A particular set of institutions, and more important how they interact, embodies the political infrastructure of a given society or organization, and rethinking this configuration of institutions belongs in the fourth layer of the learning loop.

Trade-off 2. But here lurks another trade-off: though there is a benefit in coherent action (13), it may come at the expense of adapting to the Diversity of specific situations. Bureaucracy is far removed from the local reality of communities and their specific lifeworlds.

3.2 Agonistics

When defining common goals from the goals of agonistic sub-communities, there are many different aggregation strategies, with different trade-offs, and some may lead to the goals of a majority, or an oligarchic minority, dominating the decision process.

Criterion 14 (Fairness). Social processes, in particular the process used to determine collective goals, should not systematically ignore the goals of a subcommunity.

We will not recapitulate political theory here, but would like to point out that an important part of the conflict plays out at two levels: trying to influence public opinion (whether through information or disinformation campaigns), and trying to influence the deciders, which represent a smaller attack surface[4].

Representative democracy creates its own elites: The complexity of the state apparatus makes it more likely to be captured by an expert class of its own, which may itself suborn the process to its own interests. Of course, the complexity of the system of institutions partly reflects the complexity of the world and society; and it has been argued (as far back as Plato) that allowing the broader public to have a say in important decisions is a guarantee for uninformed decisions.

But a specialized political class has its own downsides. For example, it is advantageous for candidate representatives to use vague promises as a way to garner support from people with ultimately divergent interests, at the expanse of Legibility and Accountability. The end result is that citizens lose trust in the processes, and rightly so: in a recent study, it was shown that political decisions in the USA were systematically biased in favour of elite interests against citizen preferences [26]. The complexity of the state also becomes a way to avoid Verification.

[4] Corruption of public officials is an ancient problem, which is why Athenian democracy resorted to sortition [49]. Corruption is more common in more unequal societies.

3.3 Cognitive complexity revisited

On the other hand, experiments with direct democracy have lead to sets of law that are faulted for Accuracy, Coherence, or both. Successful experiments with collective intelligence with a fair, representative sample of the broader public, such as citizen assemblies[18] and deliberative polling [23], have relied on an extensive (6) deliberative process, where experts are readily available to explain the issue clearly.

Trade-off 3. And we see another key, more complex trade-off: Accuracy, especially when dealing with social Diversity requires a deliberative process (11); but that process has a cost both in communication Overhead and in Legibility of complex issues.

For each step in our model (1) of decision and learning, we should identify institutions in charge of that step, and how they handle those tradeoffs identified. To caricature extreme models, libertarians favour adaptation to diversity at the expense of coherence (and cohesion); oligarchs sacrifice accuracy (and deliberation) to coherence; technocratic bureaucracy represent a compromise of coherence and accuracy that sacrifices legibility. Can we improve on the institutions of democracy without creating a prohibitive overhead? In particular, how to make the social decisions and learning, in all their complexity, subject to verification, deliberation and accountability?

3.4 Social media agora

As a counterpoint, let us look at how social media has claimed the role of an agora, where everybody can weigh in on public issues, including governance. This is clearly a more inclusive and diverse Deliberation than criticism through professionalized classes (such as journalists or civil society activists).

Objection (Cohesion). However, a downside of social media, with its emphasis on engagement, is that it naturally favours strong emotional reactions, such as anger and fear. This can lead to rising hostility between factions, up to the point where they interact from a basic stance of distrust. This distrust is compounded in that, pursuing engagement, social media has an incentive to feed the emotions with "more of the same", leading different factions to live in an information bubble.

Objection (Accuracy). This could be mitigated with fact-checking against a shared reality, but accuracy is also sacrificed to engagement.

Objection (Legibility). Another issue with social media is that each contribution stands on its own, and is not necessarily connected to similar interventions except through replying. It is very hard to have an intelligible map of how many distinct ideas are being expressed, and who stands behind each one.

Objection (Learning). Beyond the duplication issue, the ephemeral stream conversation structure of social media makes it harder than necessary to cross-reference new interventions against past interventions, which prevents the accumulation of a stock of learning.

4 Issues with Generative AI as a core governance mechanism

Given these existing mechanisms and criteria, let us ask again what roles should automated processes play? We would like to specifically make arguments against involving them in the decision making itself.

4.1 Inherent biases

The basic argument for using generative AI is that it provides deciders convenient access to a broad synthesis of human knowledge. This itself is a form of synthesis of Diversity. A more speculative argument is made that generative AI can or will someday provide original answers to difficult problems, but let us first focus on the current state of the technology.

Objection (Accuracy). At some basic level, generative AI acts as a (lossy) compression of its training set. As such, it has been shown to blindly reproduce existing biases in its training set [59]. The only sure way to prevent that would be to fact-check the training set, but that is not economically realistic given the size of the training set needed to train a modern generative AI. So instead of fact-checking, generative AI relies on the internal consistency of the training set as an imperfect proxy for accuracy.

Objection (Verification). Of course, another reason fact-checking is not done is that the training set is expected to be an industrial secret, to create a moat in the AI company's business model. The implication is that the training set cannot in general be verified.

Introspection There is a huge research effort ongoing to provide explainable AI, so that AI suggestions can be verified even if the process itself is a black box. Some of this effort is looking at the internal weights, but this type of explanation is hard to understand (2) and does not help to check the provenance of the AI's statement, or to engage in Deliberation around it.

Most people rely on the imperfect proxy of asking the generative AI why it choose an answer. When prompted to do so, the generative AI will indeed provide the most likely explanation that someone would give when asked such a question. This is pure confabulation, and has nothing to do with the AI's inner mechanism, which is opaque to its training-set regurgitation mechanism.

Furthermore, it is based on human *post hoc* justification of their own positions, which is also generally understood to be quite unaware of the factors actually involved in the inner decision mechanism. Humans are not generally good at introspection either.

In a social reasoning process, this flaw is mitigated by the process of conversation, where reasons can be subject to discussion, under the assumption of equality 14.

Human reception This assumption obviously does not hold when humans converse with AI; their considerable recall and polished language make most people take answers from generative AI as more authoritative than they are.

Generative AIs are also very unreliable conversation partners for many reasons; much has been made of so-called hallucinations, but it's only lately that people are more overtly critical of the impact on our inherent confirmation bias of the sycophancy that AI has been programmed to display, probably to be more agreeable to customers [25]. The agreeableness is also problematic in its own right; there have been studies of cases of emotional dependency on AIs as emotional support [61], but there have also been concerns whether heavy AI users will lose the skills to handle a clash of views in a productive way.

Biases as an expression of power

Objection (Accuracy, Alignment). In the most extreme cases, the opacity of the process enables covert manipulation by people who choose to introduce bias, either in the form of unfounded claims, such as Grok's recent assertion of white genocide [35], or the poisoning of training set by russian propaganda [17].

Given those issues, one can question the enthusiasm to push generative AI into the decision process.

One obvious reason is financial: The sellers of generative AI solutions have an incentive to sell those solutions at all layers, and the decision layer of decision bodies, whether at the social or corporate level, is an extremely lucrative one. This opportunity for AI companies comes to the detriment of knowledge workers whose work was either part of the knowledge commons or private, and was enclosed by inclusion in training sets.

Objection (Fairness). Meanwhile, small circles of deciders have a financial incentive not to employ those same knowledge workers, and to replace them with automated processes. But there is also an ideological battleground, where deciders can resent the interventions of experts, especially when these experts' evidence questions the validity of decisions that were taken against the public interest, following corruption or capture. Replacing experts with a mechanism makes it much more convenient to camouflage the goals of any decision. More generally, the more decision is automated, the less other people are involved in the decision.

This is especially obvious in the workplace: generative AI is often imposed on workers, often with the explicit aim of replacing them. There are no democratic principles of equality in work communities, and even in the political community, there are actors whose explicit plan is to replace processes in human institutions, which they deem fallible, by even more fallible automated decision processes. [15]

4.2 Accountability sink

Objection (Accountability). There is another incentive: Letting opaque algorithms take decisions is a way to obscure the chain of decision, and makes it extremely difficult to require any accountability for bad decisions. Thus deciders become both able to introduce bias opaquely and become unaccountable for failure.

The usual answer of AI proponents to accuracy issues (besides claiming it's going to get better *real soon now*) is proposing that someone should vet the answers of the AI, the so-called human in the loop. However, verification of information does not require less time or cognitive effort than research. Meanwhile, the speed of generation imposes a rhythm to information flow that makes thorough verification impractical. Of course, the overall effect is a decrease of fact-checking and accuracy, and even of the capacity to do so [37]. Most important, the human in the loop will be scapegoated for the unavoidable mistakes, and rarely the person who put the system in place. This is another way in which opaque algorithms act as an accountability sink [21].

4.3 Short-circuiting social learning

Finally, there is broad concern about the impact of generative AI on the capacity to learn. Some aspects are rather obvious, such as the providing a bread opportunity for cheating, and fostering intellectual laziness.

Objection (Learning). Another well known issue is that generative AIs, though they cannot (yet?) replace experts, can and do replace novices in many professions. How will the next generation learn to become experts if they do not have access to the learning and mentorship opportunities of a novice position?

But there is a more subtle issue: Generative AI is trained on past information. It can be retrained with new information, but does not behave well if the new information includes AI output. Since we do not have provenance information in our training sets, we cannot practically exclude AI output, and training AI with updated data is more and more difficult. Even if that were the case, AI is, through training, essentially backwards-looking, offering new combinations of past statements, but there is a level of innovation that may be inherently beyond it. There is no doubt it can handle single-loop corrections, and probably some simple double-loop learning, but there is a possibility that triple-loop learning is fundamentally beyond its reach, at least on its own.

By intercalating itself in social processes, a uniquely insidious colonization of the lifeworld, generative AI short-circuits the processes of cultural transmission by which social learning may otherwise occur.

4.4 Whose conversation?

Objection (Participation). For trust, the process matters as much as its result, and this would be true even if the automated (human-less) process were to yield perfect, ideal consensus. Without the process of deliberation and value alignment itself, there cannot arise a feeling of being part of a community, and the political unit is reduced to its rawest power dimension: you are a citizen because the laws of the nation declare you are, however you feel about it.

Again, there is nothing inherently wrong with including generative AI in a public deliberation process, as long as participants understand the strengths and limitations of generative AI. The issue is when the AI is either taking decisions, or in an opaque dialogue with a decision body. It is both vindicating and troubling that one of the most promising avenues of research in AI is having multiple AI agents in a conversation together; again, the issue is that they're having this deliberation among themselves, without including us.

5 Augmented Collective Intelligence

We propose a programme to research and realize augmented collective intelligence at scale. It takes as fundamental the value of transparent deliberation, and proposes that all decisions should be grounded in a comprehensive deliberative map.

5.1 Historical precedent: dialogue mapping

In its original form, dialogue mapping has been shown to work best for groups of medium size, who discuss in an assembly, while a cartographer constructs a map of the discussion i real time. The map represents key issues, proposals and arguments. Unlike a pure debate map, new questions can arise at any point. It has been found that the map, as a liminal object, can defuse tensions, as people address the point rather than the person who first enounced it. Dialogue mapping has been found to help very diverse communities, with diverging goals, reach a mutual understanding [19]. Interestingly, this does not carry over if participants were asked to do their own mapping asynchronously [11]. This was partly due to contributions that were not following the mapping conventions, but another issue was some people clogging up the map with detailed description of side issues, until the global map was barely legible.

5.2 Towards a global map for collective intelligence

Why a map? We strongly believe that natural language is ill-suited as a working memory for large-scale collective intelligence. Each idea can have countless different expressions. On the one hand, there is clear pedagogical value to what Mike Caulfield calls choral explanations, [13], allowing people with diverse backgrounds to understand an idea in a form appropriate to their expertise. Gathering those multiple formulations as a single point on the map reduces the communication Overhead, and the overwhelm and cognitive complexity.

Also, a map can directly display the extent of the alternative space in the neighbourhood of a given idea, helping people become aware of how much they don't know yet. We believe that this awareness might encourage at least some people to at least consider other perspectives.

Progressive formalization As ideas on the map have multiple expressions, understandable by a diversity of people, we propose this diversity includes structural descriptions of ideas, using formal language and symbolic data structures. Formal languages have the benefit of reducing ambiguity, and hence Accuracy. Formalized ideas can be compared using various formal analysis techniques, allowing to de-duplicate equivalent ideas, identify inconsistencies (helping Coherence), etc. However, formal languages are usually designed by domain experts, and hard to approach by outsiders.

Trade-off 4. Thus, there is normally a trade-off between the precise vocabulary needed for Accuracy, and including (9) a broader (14) public in the Deliberation. This is related to the trade-off 1 with cognitive complexity.

This non-specialist public might approach the map through the choral explanations in natural language, but a hybrid (symbolic-generative) AI can also play a helpful role, translating the more symbolic knowledge in vernacular terms. To enable them to participate actively in the conversation (14), we must also accept informal, unstructured contributions. But then, we can progressively help them to clarify their contribution: we can identify ambiguities and ask clarifying questions; we can suggest multiple interpretations, and ask contributors to commit to (at least) one of them; we can compare it to already-formalized ideas, and ask to focus on the distintive elements; we can propose breaking it down into component concepts. Eventually, the contribution will be clear enough to be formalized. This whole process can be crowdsourced with peers, or with the mentorship of hybrid AI.

Indicators The map will serve the function of global memory insofar as it is exhaustive. An exhaustive global map will be, by essence, too large to be legible. People will interact with the global memory through either curated or computed maps. Maps curated by humans, making editorial choices, are more likely to have high Legibility, and be the most useful for approaching the information space. Of course, they will reflect the biases of their authors. As the curated map is

embedded in a global information map, the system can provide indicators that express the scope of what's outside the map. Useful partial maps can also be computed by the system, based on traversing relations and using indicators as heuristics to prune elements beyond a certain number.

This is one of the most delicate aspects of the system: heuristics for indicators can re-introduce hidden bias. First, the algorithms for indicators should themselves be transparent, and subject to public Deliberation. More controversially, the system should allow to incorporate a Diversity of aggregation and ranking indicators, much like Bluesky's open marketplace of algorithms [27]. However, in the name of Verification, closed-sourced algorithms should be marked as such. The system should make it convenient to re-compute any view incorporting unverified indicators in a way that compares them to a set of publically vetted indicators. On the other hand, we do not believe in excluding them outright; the plurality of approaches may itself contribute to make it more difficult to game the system.

What indicators are we considering as central?

Abstraction and distinctions A fundamental issue with a global map is to present a cluster of a great number of related ideas in the form of a unifying abstraction at the right level. Though we can use automated semantic clustering to identify tentative clusters, we propose to use techniques from formal concept analysis [60] to interpret them as a lattice of abstractions and distinctions, progressively lending structure to the massive underlying knowledge base. We can then use indicators of usage to make informed guesses about the appropriate level of abstraction, or the most salient distinctions, for a given participant. In case of error, the participant should be able to navigate the structure.

Coherence indicators A deliberation tool should not be an arbiter of truth, and needs to record a plurality of viewpoints. Yet, especially in this era of weaponized disinformation and cognitive denial of service attacks with a flood of nonsensical information, we cannot totally eschew the responsibility to rate information. We believe that fighting disinformation requires incorporating it, and to identify the presence of supporting or contradicting evidence. As such, the platform should focus on Coherence issues, between decisions, misalignments with goals, or conflicting goals. (Though sometimes those represent a legitimate balance.) Again, some of this will be crowdsourced, but hybrid AI could be invaluable in the automated detection of suspected inconsistencies and maybe broader coherence issues in the decision base. Such potential issues would then be subject to further human deliberation, or the structured information could also be checked by classical, verifiable inference engines.

Here, we believe there is much space for research in identifying algorithms that help the process of deliberation. Some indicators may focus on content, others on the communication structure. For example, the global brain [57] measures how much each individual contribution pushes participants towards the position of maximally involved (and presumably informed) participants. We believe, but would like to verify, that debates could be improved by directing participant's

attention to higher-level learning loop elements, such as epistemic criteria: what would lead them to believe a position against the other? One part of the proposal is that we should share an experimental workbench to study such issues.

Moderation In any system that accepts public contributions, we have to tackle the issue of moderation. Beyond the obvious (hate speech), we have to deal with arguments that oppose the value of inclusive deliberation as a basis for decision making. Despite the paradox of tolerance [47], we are committed to mapping such positions, precisely so they can be refuted. However, such positions are contrary to the goals of the platform, and will be tagged as such, and the original unstructured contribution containing may still be filtered out.

Reputation This is another sensitive design area. We acknowledge the value of debating at the level of ideas, not people. Abstracted, structured ideas can be partially detached from their authors. There is a lot of research in the positive value of avatars in certain epistemic settings. Yet, there is also a strong value in tracing provenance of ideas, if only to distinguish contributions of humans, artificial or collective actors. Also, though *ad hominem* is in principle a fallacy, a lot of legitimate political debate concerns statements by individuals, their interpretation and consequence, and anonymity only goes so far. More importantly, using provenance, we can measure the track record of contributors [54] and institutions. This can be incorporated as a key proxy indicator to rank arguments.

Trade-off 5. There is a trade-off between allowing a track report of reliable information to play a role in ranking contributions, and not letting powerful or famous people dominate the discussion.

Federation In our experience, people rarely converge on a single technical infrastructure. Indeed, just as expert communities need jargon for efficient communication, they need diverse domain-specific tools to manipulate the concepts of their discipline. Consequently, we do not propose a single tool, but a commitment to an ecosystem of tools that share an interoperable data format. This format should allow for federated cross-references between multiple thinking repositories, or (in the terms of Jack Park) knowing hubs, and for third parties to compute aggregated indicators across the federation.

What would such a data format look like? We have contributed to an early federation protocol for collective intelligence [43], which was based on the basic categories used in dialogue mapping: a few concepts (questions, proposals and arguments) and relations (answers, pro, con, questions.) This is a simple and proven core, but we suspect that structured conversation will benefit from a richer vocabulary. In particular, since we are dealing with decisions, the notion of expected outcome, and valuation along criteria, needs to be part of that vocabulary.

Experts have identified a on the order of a hundred argumentation schemes [56], and on the order of a thousand linguistic frames describing basic situations

[5]. Which of those are necessary? We do not believe that the perfect vocabulary exists to be identified, but that we should allow for a language to evolve through Learning, in the form of accretion around a base core, guided by usage statistics. However, to avoid this to become a babel, we have proposed to use nested frames [42]. Frame nesting is formally equivalent to a recursive hypergraph, and allows statements about statements. New frame proposals should strive to build upon simpler frames through composition, and/or to provide a translation layer to alternate proposals. The translation layer should be determinstic and work at the level of structured data, though the result may take the form of a less formal linking frame.

6 Parallel work and open issues

This is a research program, and we do not have all the answers. Yet we are building on established research, and we find many researchers are building parts of what we're proposing, besides the components we are ourselves experimenting with.

First, there is a lot of activity around structured conversation [36]. We have previously been involved in one such project, where among other things we experimented with visualization and deliberation analytics [44]. We are currently involved with discourse graphs, which develops structured conversations for research teams [14].

Experiments also show that generative AI can be used to automate the generation of conversation maps [10]. Such generated maps can be shown improve group discussions [1,32]. Conversational agents have also been shown to be skilled at moderation [4], and have been shown to help polarized groups reach a better mutual understanding [53]. In the latter case, the effect could be explained in part to the persisting attribution of neutrality to automated processes by the participants, but either way, we believe strongly in the possibility of harnessing AI's access to a considerable corpus to identify bridges between epistemic islands, in community with people willing to do the exercise. We also hope that demonstrating this possibility will get people to reconsider the value of the deliberative process to create more Coherence in society.

The idea of a global map of knowledge is ancient, and the origins of the internet can be traced to this dream, from Vannevar Bush's memex, to Ted Nelson's Xanadu, to Tim Berner Lee's web [6]. Trying to build a global map specific to deliberation and decision making also has precedents, such as DebateWise, and recently the Canonical Debate Map white paper of Timothy High [31]. We have shared ideas on knowledge federation with Jack Park of TopicQuests, and previously we had worked on a collective intelligence data interoperability format [43] for the Catalyst project.

This history is one of partial successes and failures. The internet has provided a shared addressing space for documents, and WikiPedia has provided a reference point for ideas.

Yet, at the rational level, attempts at deeper knowledge unification have remained marginal. What is our strategy to avoid this fate? First, we are not trying to build a unifying ontology, but embrace ontological pluralism. We believe that, thanks to the role of hybrid AI as a translator, there is a unique opportunity to resolve the tradeoff 4 between crowdsourcing and formalization.

At the social levol, many important conversations happen within silos, whether for convenience, comfort or a (partly illusory) feeling of privacy. We believe one deep underlying issue is that there are strong social disincentives to make the decision processes public. Accountability can create an enormous pressure to avoid even the appearance of considering impopular options; whereas in reality it would be better to consider and reject them for explicit reasons. But even without those pressures, it can be useful to let unfinished ideas compost in small trusted groups, to avoid premature exposure or even premature formal explicitation. (Nora Bateson calls this process aphanipoiesis [9].) This is another reason to allow for federated, community-based knowledge hubs; some ideas can be made public when ready. But there also needs to be a higher social tolerance for nuance and tentative exploration and experiments. Again, we hope that laying out the process (even with delay) will make this more acceptable, and facilitate higher levels of social learning.

We also believe firmly in the importance of making structured deliberation more accessible to the general public, both through using generative AI as a translator/mentor, and through pedagogical activities. In particular, we are developing a coopetitive game with Jack Park [45], where teams learn to contribute to a structured conversation.

Beyond this project and research agenda, we hope that this paper sparks a discussion of the precise risks and benefits that AI pose at various points in the decision model.

Acknowledgments. Many of these ideas were developed with Jack Park, of TopicQuests. I am also grateful to Sonny Bhatia of West Point, and Louise Vandelac of UQAM, for helpful comments.

Ethical Statement There are no ethical issues in our research at this stage.

References

1. Anastasiou, L., Liddo, A.D.: Bcause: Human-ai collaboration to improve hybrid mapping and ideation in argumentation-grounded deliberation (2025), https://arxiv.org/abs/2505.03584
2. Argyris, C., Schön, D.A.: Theory in Practice: Increasing Professional Effectiveness. Jossey-Bass, San Francisco (1974)
3. Ashby, W.R.: Requisite variety and its implications for the control of complex systems. Cybernetica **1**(2) (1958)
4. Babatunde, I.D., Nnanna, O.M., Klein, M.: Moderating large scale online deliberative processes with large language models (llms): Enhancing collective decision-making. In: Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing. p. 996–1003. SAC '25, Association for Computing Machinery, New York, NY, USA (2025), https://doi.org/10.1145/3672608.3707925

5. Baker, C.: FrameNet, present and future. In: Webster, J., Ide, N., Fang, A.C. (eds.) The First International Conference on Global Interoperability for Language Resources. City University, City University, Hong Kong (2008)
6. Balasubramanian, V.: State of the art review on hypermedia issues and applications (1993)
7. Bateson, G.: Naven: A Survey of the Problems Suggested by a Composite Picture of the Culture of a New Guinea Tribe Drawn from Three Points of View. Cambridge University Press, Cambridge (1936)
8. Bateson, G.: Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution and Epistemology, chap. Social Planning and the Concept of Deutero-Learning. Paladin, Granada, London (1942)
9. Bateson, N.: Aphanipoiesis, https://norabateson.medium.com/aphanipoiesis-96d8aed927bc
10. Bhatia, A., Sukthankar, G.: Using llms to structure and visualize policy discourse. In: Yin, W., Ahn, J.J., Zhang, R., Huang, L., Hadfi, R., Ito, T., Ohnuma, S., Shiramatsu, S. (eds.) Artificial Intelligence for Research and Democracy. pp. 69–76. Springer Nature Singapore, Singapore (2025)
11. Buckingham Shum, S.J., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B., Nuseibeh, B.: Rationale Management in Software Engineering, chap. Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC, pp. 111–132. Computer Science Editorial, Springer-Verlag (2006), http://kmi.open.ac.uk/publications/index.cfm?trnumber=kmi-05-18
12. Castoriadis, C.: L'institution imaginaire de la société. Seuil, Paris (1975)
13. Caulfield, M.: Choral explanations. https://hapgood.us/2016/05/13/choral-explanations/ (May 2016)
14. Chan, J., Akamatsu, M., Vargas, D., Kawerau, L., Gartner, M.: Steps towards an infrastructure for scholarly synthesis. https://arxiv.org/abs/2407.20666 (2024), https://arxiv.org/abs/2407.20666
15. Chayka, K.: Elon musk's a.i.-fuelled war on human agency. The New Yorker (February 2025), https://www.newyorker.com/culture/infinite-scroll/elon-musks-ai-fuelled-war-on-human-agency
16. Christopher, N., Bansal, V.: Indian voters are being bombarded with millions of deepfakes. political candidates approve. https://www.wired.com/story/indian-elections-ai-deepfakes/ (May 2024), https://www.wired.com/story/indian-elections-ai-deepfakes/
17. Châtelet, V.: Exposing pravda: How pro-kremlin forces are poisoning ai models and rewriting wikipedia (2025), https://www.atlanticcouncil.org/blogs/new-atlanticist/exposing-pravda-how-pro-kremlin-forces-are-poisoning-ai-models-and-rewriting-wikipedia/
18. Collective: European citizens' assembly: a new model for decision making. Tech. rep., Center for Blue Democracy (May 2022), https://citizensassemblies.org/wp-content/uploads/2022/05/European-Citizens-Assembly.pdf
19. Conklin, J.: Dialogue Mapping: Building Shared Understanding of Wicked Problems. John Wiley & Sons (2006)
20. Danaher, J.: The threat of algocracy: Reality, resistance and accommodation. Philosophy & Technology **29**(3), 245–268 (2016), https://doi.org/10.1007/s13347-015-0211-1
21. Doctorow, C.: Ai's "human in the loop" isn't (2024), https://pluralistic.net/2024/10/30/a-neck-in-a-noose/

22. Farrell, H., Han, H.: Ai and democratic publics. Tech. rep., 25-17 Knight First Amend. Inst. (August 2025), https://knightcolumbia.org/content/ai-and-democratic-publics
23. Fishkin, J.S., Luskin, R.C., Jowell, R.: Deliberative polling and public consultation. Parliamentary Affairs **53**(4), 657–666 (2000)
24. Flood, R.L., Romm, N.: Contours of diversity management and triple loop learning. Kybernetes **25**(7/8), 154–163 (1996)
25. Gerlich, M.: Ai tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies **15**(1), 6 (2025), https://www.mdpi.com/2075-4698/15/1/6
26. Gilens, M., Page, B.I.: Testing theories of american politics: Elites, interest groups, and average citizens. Perspectives on Politics **12**(3), 564–581 (2014)
27. Graber, J.: Algorithmic choice (3 2023), https://bsky.social/about/blog/3-30-2023-algorithmic-choice
28. Habermas, J.: The Theory of Communicative Action (Lifeworld and System: A Critique of Functionalist Reason), vol. 2. Beacon Press, Boston (1987)
29. Heath, J.: Communicative Action and Rational Choice. MIT Press, Cambridge, MA (2001)
30. Higgins, E.: Global nature of information disorder. https://www.jbs.cam.ac.uk/events/cambridge-disinformation-summit-2025 (2025), https://www.jbs.cam.ac.uk/events/cambridge-disinformation-summit-2025/
31. High, T.: The canonical debate. https://github.com/canonical-debate-lab/paper/blob/master/README.mediawiki (June 2018)
32. Ito, T., Hadfi, R., Suzuki, S.: An agent that facilitates crowd discussion. Group Decision and Negotiation **31**, 621 – 647 (2021), https://api.semanticscholar.org/CorpusID:243839643
33. Jeantet, D., Savarese, M.: Brazilian city enacts an ordinance that was secretly written by chatgpt. https://apnews.com/article/brazil-artificial-intelligence-porto-alegre-5afd1240afe7b6ac202bb0bbc45e08d4 (November 2023), https://apnews.com/article/brazil-artificial-intelligence-porto-alegre-5afd1240afe7b6ac202bb0bbc45e08d4
34. Kahneman, D., Tversky, A.: On the reality of cognitive illusions. Psychological Review **103**(3), 582–91 (July 1996), discussion 592–6
35. Kerr, D.: Musk's ai grok bot rants about 'white genocide' in south africa in unrelated chats. The Guardian (2025), https://www.theguardian.com/technology/2025/may/14/elon-musk-grok-white-genocide
36. Kirschner, P.A., Buckingham Shum, S.J., Carr, C.S. (eds.): Visualizing Argumentation. Springer-Verlag, London (2003), http://www.visualizingargumentation.info/
37. Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.H., Beresnitzky, A.V., Braunstein, I., Maes, P.: Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. https://arxiv.org/abs/2506.08872 (2025), https://arxiv.org/abs/2506.08872
38. Maturana, H., Varela, F.: Autopoiesis and Cognition. Boston Studies in the Philosophy of Science, D. Reidel, Boston (1980)
39. Mercier, H., Sperber, D.: The Enigma of Reason. Harvard University Press, Cambridge, MA, USA (2017)
40. Mouffe, C.: Agonistics: Thinking the World Politically. Verso Books, New York, NY (2013)

41. Ovadya, A., Redman, K., Thorburn, L., Chen, Q.Z., Smith, O., Devine, F., Konya, A., Milli, S., Revel, M., Feng, K.J.K., Zhang, A.X., Chandra, B., Bakker, M.A., Kasirzadeh, A.: Democratic ai is possible. the democracy levels framework shows how it might work (2025), https://arxiv.org/abs/2411.09222
42. Parent, M.A.: Towards knowledge federation. In: Hegeland, F. (ed.) the Future of Text, vol. V, pp. 188–190. Future Text Publishing (2024), https://futuretextpublishing.com/vol-5/
43. Parent, M.A., Grégoire, B.: Architecture and cross-platform interoperability specification. Tech. rep., Catalyst (mar 2014), http://bit.ly/catalyst_interop3
44. Parent, M.A., de Liddo, A., Klein, M., Ullman, T.: Project testbed: Argument mapping & deliberation analytics. Tech. rep., Catalyst (2015), http://catalyst-fp7.eu/wp-content/uploads/2016/01/CATALYST_WP4_D4.2b.pdf
45. Parent, M.A., Park, J.: Sensecraft game design. http://bit.ly/4l9Bl8U (May 2023), https://docs.google.com/presentation/d/1K60P4xFMt9v7X9eDjkQTPfFJqQGDbss9CzoOGPlDkMA/edit#slide=id.p
46. Parsons, T.: The Structure of Social Action. McGrawHill, New York (1937)
47. Popper, K.: The Open Society and Its Enemies, vol. 1. Routledge (1945)
48. Rawls, J.: The idea of an overlapping consensus. Oxford Journal of Legal Studies **7**, 251–276 (1988)
49. Samons, L.: What's Wrong with Democracy?: From Athenian Practice to American Worship. University of California Press (2004)
50. Schelling, T.C.: The Strategy of Conflict. Oxford University Press, New York (1963)
51. Schneier, B.: How ai will change democracy. https://www.schneier.com/blog/archives/2024/05/how-ai-will-change-democracy.html (May 2024), https://www.schneier.com/blog/archives/2024/05/how-ai-will-change-democracy.html
52. Snowden, D., Boone, M.: A leader's framework for decision making. Harvard business review **85**, 68–76, 149 (12 2007)
53. Tessler, M.H., Bakker, M.A., Jarrett, D., Sheahan, H., Chadwick, M.J., Koster, R., Evans, G., Campbell-Gillingham, L., Collins, T., Parkes, D.C., Botvinick, M., Summerfield, C.: Ai can help humans find common ground in democratic deliberation. Science **386**(6719), eadq2852 (2024), https://www.science.org/doi/abs/10.1126/science.adq2852
54. Tetlock, P.E.: Expert Political Judgment: How Good Is It? How Can We Know? Princeton University Press (2005)
55. Tosey, P., Visser, M., Saunders, M.: The origins and conceptualizations of 'triple-loop' learning: A critical review. Management Learning **43**, 291–307 (07 2012)
56. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press, Cambridge (2008)
57. Warden, J., Nakayama, J., Dietze, F.: The global brain algorithm. https://social-protocols.org/global-brain/ (2024)
58. Weber, M.: The Protestant Ethic and the Spirit of Capitalism. Charles Scribner's Sons, New York (1958)
59. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L.A., Isaac, W., Legassick, S., Irving, G., Gabriel, I.: Ethical and social risks of harm from language models (2021), https://arxiv.org/abs/2112.04359
60. Wille, R.: Restructuring lattice theory: An approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets. pp. 445–470. Springer Netherlands, Dordrecht (1982)

61. Yuan Z, Cheng X, D.Y.: Impact of media dependence: how emotional interactions between users and chat robots affect human socialization? Frontiers in Psychology **15**, 1664–1078 (Aug 2024)