nature human behaviour

Perspective

https://doi.org/10.1038/s41562-024-01959-9

How large language models can reshape collective intelligence

Received: 6 November 2023

Accepted: 17 July 2024

Published online: 20 September 2024

Check for updates

Jason W. Burton[1,2] ✉, Ezequiel Lopez-Lopez[2], Shahar Hechtlinger[2,3], Zoe Rahwan[2], Samuel Aeschbach[2,4], Michiel A. Bakker[5], Joshua A. Becker[6], Aleks Berditchevskaia[7], Julian Berger[2,3], Levin Brinkmann[8], Lucie Flek[9,10], Stefan M. Herzog[2], Saffron Huang[11], Sayash Kapoor[12,13], Arvind Narayanan[12,13], Anne-Marie Nussberger[8], Taha Yasseri[14,15], Pietro Nickl[2,3], Abdullah Almaatouq[16], Ulrike Hahn[17], Ralf H. J. M. Kurvers[2,18], Susan Leavy[19], Iyad Rahwan[8], Divya Siddarth[11,20], Alice Siu[21], Anita W. Woolley[22], Dirk U. Wulff[2,4] & Ralph Hertwig[2]

Collective intelligence underpins the success of groups, organizations, markets and societies. Through distributed cognition and coordination, collectives can achieve outcomes that exceed the capabilities of individuals—even experts—resulting in improved accuracy and novel capabilities. Often, collective intelligence is supported by information technology, such as online prediction markets that elicit the 'wisdom of crowds', online forums that structure collective deliberation or digital platforms that crowdsource knowledge from the public. Large language models, however, are transforming how information is aggregated, accessed and transmitted online. Here we focus on the unique opportunities and challenges this transformation poses for collective intelligence. We bring together interdisciplinary perspectives from industry and academia to identify potential benefits, risks, policy-relevant considerations and open research questions, culminating in a call for a closer examination of how large language models affect humans' ability to collectively tackle complex problems.

In January 2023, ChatGPT gained 100 million users just two months after its launch[1], making it the fastest-growing web application ever and signalling both a striking advancement of the underlying large language model (LLM) technology and a new era for the online information environment (Fig. 1). Recent developments of LLMs have spurred high-profile debates on epistemological (for example, do LLMs 'understand' language?[2]), ethical (for example, how might LLMs propagate harmful stereotypes and social biases?[3,4]) and metaphysical aspects of LLMs (for example, are there "sparks of artificial general intelligence" in GPT-4 (ref. 5), one of 2024's most advanced LLMs?), but there is limited understanding of how they will affect the collective intelligence (CI) that underpins the success of groups, organizations, markets and societies.

LLMs are artificial intelligence (AI) systems that use massive amounts of input data and deep learning techniques to analyse and generate text (for example, BERT, LLaMA and the prominent generative pre-trained transformer (GPT) series). As LLMs become increasingly accessible to the public, their general-purpose ability to process vast amounts of information and output human-like text poses unique, pressing questions to CI at large. Text is the primary medium of communication in the digital age[6], and LLMs are already being integrated into online environments and adopted as tools for communication and information search. Of course, CI is not only based on language, and developments in other forms of generative AI such as image, video and audio generation may also have important effects on CI in the future—a point we return to later in the Perspective. However, it is LLMs that are

A full list of affiliations appears at the end of the paper. ✉e-mail: jb.digi@cbs.dk

https://doi.org/10.1038/s41562-024-01959-9

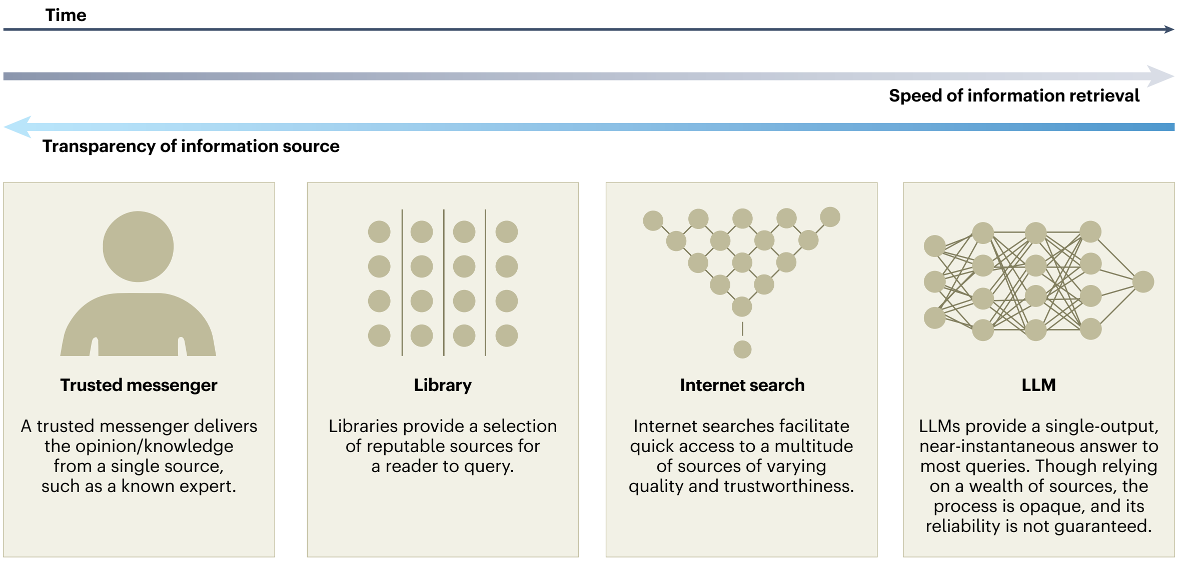

Fig. 1 | Development of information environments over time. A general trend is observed whereby new technologies increase the speed at which information can be retrieved but decrease transparency with respect to the information source.

positioned to most imminently affect key collective processes such as civic deliberation and elections, as well as how people interact with and relate to each other in everyday life.

Given these wide-reaching implications, we synthesize interdisciplinary perspectives from industry and academia to identify ways in which LLMs can reshape CI, for better and worse, considering LLMs' current and potential capabilities. In doing so, we provide an overview of priority areas for researchers, policymakers and technologists alike to consider.

CI and its importance

CI refers to the ability of individuals to collectively act in ways that seem intelligent, often displaying intelligence surpassing that of individuals acting alone on tasks such as idea generation, problem-solving, estimation, inference and decision-making[7,8]. CI manifests itself across society in myriad ways and varying scales in governing the collective memory, attention and reasoning processes essential to any intelligent system (for frameworks, see refs. 7,9–13). For instance, large-scale, macro-level CI can be observed in markets, where self-organized competition among individual buyers and sellers can efficiently set prices[14,15], and in the 'wisdom of crowds', where aggregating individuals' judgements can be used to identify correct alternatives and boost the accuracy of estimations[16–19]. On smaller scales, micro-level CI can be observed in teams and organizations, which overcome individuals' limitations in time, knowledge and computational capacities by specifying roles and workflows to manage collective attention and facilitate collaboration[20–23]. Existing literature on CI suggests there are several basic components that underlie its emergence. Here, we outline three to guide our discussion of the intersection of LLMs and CI: diversity, individual competence and aggregation.

First, in many contexts, diversity among individuals can promote CI. Diversity may be attributable to demographic and cultural differences, referred to as identity diversity, or to differences in how people represent and solve problems, referred to as functional diversity[24,25] (alternatively, see the distinction between surface-level and deep-level diversity[26–28]). Functional diversity, which may be promoted by identity diversity, ensures that a collective thoroughly searches a solution space when solving problems; it can also lead to error correction, as the mistakes or oversights of one individual can be cancelled out by others with a functionally different approach to the task at hand[24,25,29–31]. Although no individual may possess all relevant information or the correct model to address a given task, the diverse, distributed cognition of the many can coalesce into collectively intelligent outcomes.

Second, CI is facilitated by individual competence being calibrated for a given task. The individuals in a collectively intelligent group should not be completely naive, but they need not be experts. The need for individual competence depends on the task and the degree of diversity present; increasingly homogenous groups must be increasingly competent, and vice versa[31–33].

Third, CI requires an appropriate mechanism of aggregation to translate individual beliefs and behaviours into a collective outcome[34–36]. In some contexts, an explicit, formal aggregation rule may be applied (for example, majority voting)[16,18,30,37,38]. In others, aggregation is achieved implicitly through individuals' interactions (for example, traders' buying and selling behaviours aggregate to form market prices)[34,39]. In these contexts, interaction must follow a network structure that manages the competence–diversity trade-off as needed for a given task[39]. For example, high centralization may facilitate the rapid exchange of information needed for a quick consensus. However, that same structure may render a group susceptible to excessive social influence and 'groupthink'[40] that can undermine thorough exploration of a solution space in cases where quality is more important than speed[41–48].

Together, these components help to explain when CI emerges as well as when it may fail. Human history is riddled with market crashes, organizational failures and collective decisions gone awry. CI requires stewardship to create conditions that allow individuals to interact meaningfully and productively[49].

LLMs for and of CI

Recent technological advancements have opened up new dimensions to harness CI at scale[50]. Digitalization has increased capacities to store, communicate and compute information[45], as well as capacities for individuals to query information (Fig. 1) and connect with one another. In the current online information environment, complex systems of humans and machines have given rise to new forms of large-scale CI and public goods[51], such as crowdsourced knowledge commons (for example, Wikipedia and Stack Overflow), prediction markets (for example, Metaculus and PredictIt) and deliberation forums (for example, Reddit and Polis)[52]. Crucially, there are synergies: technology supports CI, and CI supports technology. The quality and utility of crowdsourced knowledge commons, prediction markets, BOX 1

Harnessing CI for developing LLMs

The most advanced LLMs are typically trained in a two-step process. First, in pretraining, LLMs construct a generative statistical model of billions of human-written texts[171,172]. Then, human annotators provide feedback on what model outputs they perceive as most helpful and non-harmful, and the model is updated accordingly—a process called reinforcement learning from human feedback[173].

Selecting data for pretraining LLMs. The vast intake of information during pretraining highlights LLMs' capacity to process and integrate a wealth of diverse perspectives, which is crucial for mitigating biases and imbuing models with contextual awareness[87]. Importantly, LLMs do not—and should not[3]—accommodate diversity through sheer size alone; they also accommodate it through the quality and coverage of training data[174]. Building on insights from CI, errors within the training data should be random rather than systematic. A smaller dataset from diverse sources might therefore be more suitable in some cases than a larger dataset from a single source.

Aggregating human feedback. In reinforcement learning from human feedback, it is important not to sideline groups with underrepresented cultural, linguistic or personal backgrounds[3,173]. At the same time, unconstrained maximization of diversity can come at the expense of the wisdom of the crowd. This tension points to a still broader challenge: gearing LLMs to accurately reflect and integrate the social beliefs, values and norms of different populations. Existing attempts to address this have leveraged LLMs' superior summarizing[112], dialoguing[53] and consensus-building capacities[119]. Grant programmes that fund research on democratic processes for AI governance may advance these efforts[54].

Optimizing LLM usage. Insights from CI can also inform the mechanistic setup of LLMs to attune their outputs to predefined goals, such as diversity. For instance, the temperature parameter can be used to calibrate the diversity of an LLM's output, with higher values overrepresenting rare instances from the training data (for example, minority views) and lower values overrepresenting the most frequent instances (for example, majority views). Appropriate prompt engineering (for example, to represent diverse demographics) can also, in principle, foster output diversity in LLMs[87]. However, some Western LLMs exhibit biases towards Western, educated, industrialized, rich and democratic values[175] and persistently struggle to adapt to other cultures[176]. Employing a variety of LLMs, each trained with data from distinct cultural contexts, can further enhance the diversity of LLM outputs[177–179].

Ensembles of LLMs. Beyond any single LLM, CI insights can also inform the development of ensemble methods that increase performance by combining multiple models. Just as ensemble methods have seen success in traditional machine learning domains (for example, aggregating the outputs of many different classifiers often leads to more accurate predictions than any single classifier)[180], emerging evidence suggests that combining several LLMs can give rise to systems that exhibit CI on their own. For example, multi-agent systems in which multiple LLMs are allowed to converse with one another display enhanced mathematical and strategic reasoning capabilities[56] and can complete tasks more efficiently than any single LLM[57].

Iterations of training, fine-tuning, usage optimization and ensembles of LLMs are set to intertwine collective knowledge from humans and machines ever more closely, giving rise to a collective cultural process[145]. Deliberately and carefully leveraging components of CI makes it possible to steer this process towards LLMs that cater to societal goals.

deliberation forums and other technology-enabled CIs are tied to the active involvement of the individuals who contribute to and use them.

Analogously, LLMs can support CI, but they can also be viewed as a product of CI. LLMs are trained on collective data that encapsulate the contributions of countless individuals, and LLMs are often fine-tuned with collective human feedback. Prompting an LLM with a question is like a distilled form of crowdsourcing. The responses LLMs generate are shaped by how masses of other people have tended to respond to similar questions and align with the collective preferences reflected in the fine-tuning process (Box 1). Despite growing excitement about what LLMs can do for CI and vice versa[53,54]—such as leveraging research on CI to inform the design of LLMs (Box 1) or using LLMs to simulate human CI (Box 2)—explicit links between LLMs and the CI literature remain understudied.

LLMs can simultaneously enable new, heightened CI and threaten society's ability to solve problems. As with the Internet and social media[55], the consequences of LLMs will probably vary across use cases and populations. In the following sections, we provide a scoping pass at anticipating such consequences and propose recommendations for handling them (see Table 1 for an overview) with a primary focus on individual LLMs. Still, we acknowledge that combining multiple LLMs can give rise to systems that exhibit CI on their own with further associated benefits[56,57] and risks[58].

How LLMs can help CI

LLMs are trained on broad data and can be adapted or fine-tuned to a wide range of downstream tasks[59], making them versatile and capable of being integrated into a variety of collective processes, including idea generation, deliberation and preference aggregation. In some instances, LLMs offer functionalities that previous technologies could not provide (for example, prompted idea generation). In others, LLMs improve on functionalities provided by previous technologies (for example, traditional supervised machine learning models can label information, but LLMs can perform as well as or better than such models with zero- or few-shot learning[60,61]). Here we outline ways in which LLMs can be used as tools for CI, including both demonstrated use cases and possible future applications.

Increasing accessibility and inclusion in online collaborations

Increasing accessibility and inclusion in online collaborations means ensuring that all stakeholders have opportunities to participate, regardless of their differences. This can facilitate the collaboration of larger, more engaged collectives, which is desirable on two counts. First, increasing group size typically enhances the wisdom of crowds (but see refs. 62,63 on the wisdom of small, select crowds and ref. 64 on the disruptive capabilities of small teams). In binary choice, this is true as long as the average individual accuracy is above chance[29,30]. For continuous estimation tasks, following the central

BOX 2

Using LLMs to simulate human CI

Empirical research on CI in human groups is often restricted by logistical issues. The resources required to organize, recruit and pay high numbers of individuals to participate in an experiment can be prohibitive and can force researchers to limit the scope of their studies (for example, by focusing only on dyadic or small-group dynamics, or considering only a select few experimental conditions). Even CI experiments with large participant samples can result in comparatively few collective-level observations and limited statistical power, depending on the study design. Already deployed in general social science research (for example, refs. 181,182), LLMs are uniquely positioned to address these logistical issues and support CI research.

Simulating human participants. LLMs have emulated human judgements in voter preferences[87], numeric estimation tasks[183] and moral assessments[184]. With basic prompts, virtual populations can be generated with accompanying demographic representativeness in a process called "silicon sampling"[87]. LLMs can thus be used to conduct computational experiments—a popular tool in CI research[185,186]—for richer simulations of collective behaviours that could not otherwise be experimented on[187,188]. Unlike human participants, silicon samples and simulations can be reused without the effects of distraction, exhaustion, learning and memory over repeated trials. Comparatively cost-effective silicon samples could be used to explore ideas, hone experimental design and pilot studies with artificial groups before recruiting human participants.

Risks of LLM-based simulations. Any LLM is constrained by its training[189]. This process is commonly opaque, and the nature of training sets limits the fidelity of LLM output for underrepresented groups, including by nationality, political preferences and digital presence[175]. In response to this challenge, the Turing Experiment[183] has been developed to assess the replication of human participant responses at a group level. Another training-driven constraint relates to the period of training data. Although knowledge gaps may in part be mitigated by online query capabilities, the underlying model that is not trained on up-to-date data may be deficient in reflecting changing language patterns and new states of the world.

Moreover, task performance across domains varies not just across LLM editions but also within the same version[190]. This inconsistency underscores the need for researcher documentation of model usage and the measurement properties of LLMs (for example, test-retest reliability), and preferably greater disclosure by LLM creators. More broadly, as LLMs reshape research, ethical codes must be revised to ensure responsible use.

limit theorem, aggregating larger samples of independent judges typically returns more accurate estimates[17,65]. Second, active participation gives legitimacy to collective outcomes by allowing stakeholders to voice their beliefs and assume shared responsibility, which in turn leads them to view the outcome as more just[66–70]. However, there is a well-documented dilemma between group size (or participation) and performance in the group decision-making and deliberative democracy literature[66,71–74]: Increasing the size of a group may introduce more individual competence, diversity and shared responsibility, but it also imposes administrative coordination costs that, for some tasks, undercut productivity[72–74], deliberative quality[67,71] and collective performance[75,76].

LLMs may be a valuable tool for reducing barriers to participation so that the benefits of larger, more engaged groups can be reaped without exorbitant coordination costs. LLMs can translate multilingually, enabling rapid communication across language barriers[77,78]. They can also provide writing assistance, which could be particularly helpful to non-native English speakers who are frequently discriminated against in Anglocentric domains such as academic publishing[79–82]. LLMs can summarize masses of text so that, for example, late joiners to a project or discussion can review what has already been said without being faced with an overload of information and without slowing or derailing incumbents. In the longer term, personal LLMs might even act as delegates that engage in deliberative discussions on behalf of their human owners, thereby reducing (or, in the extreme case, entirely removing) the cognitive burden of deliberation and accelerating discussions that would take years for humans alone. In these ways, LLMs offer potential new routes towards online collaborations that are larger, more diverse and more equitable.

Accelerating idea generation

Idea generation is typically the first step of any problem-solving or innovation process. Increasingly, CI approaches such as crowdsourcing and open innovation tournaments have leveraged the scale and diversity of crowds to generate high-quality ideas[83,84]. LLMs can contribute to these processes by enhancing the efficiency of generating ideas.

Most straightforwardly, LLMs can serve as a crowd that can be near-instantaneously queried. Two recent studies comparing LLM-generated ideas with those generated by groups of individuals showed that it took people days or months to generate the same number of ideas that LLM tools produced in a few hours[85,86]. However, evaluations of human-generated versus LLM-generated ideas are mixed. For example, Girotra et al. observed that GPT-4-generated ideas are, on average, of higher quality than human-generated ideas, but they also exhibit greater variance in quality[86]. Boussioux et al. observed that humans' ideas were more novel, and found no significant differences between the best GPT-4-generated ideas and the best human-generated ideas in terms of the perceived quality, value and feasibility[85]. However, the quality of LLM-generated ideas may improve in future models. Moreover, the diversity of generated responses can be enhanced through techniques such as in-context impersonalization, where the LLM is guided to represent various demographics[87–89], and other evidence suggests that passive exposure to LLMs can increase the diversity of ideas generated by humans[90].

Another way LLMs could enhance collective idea generation is by augmenting individual humans by, for example, providing starting points or 'icebreakers'. Indeed, GPT-4 makes individuals about 40 times more productive at generating ideas[86]. LLMs could also serve as sounding boards for ideas. As exposure to ideas enhances individuals' creativity[91] (but see ref. 90 for counter-evidence under passive exposure to LLMs), using LLMs this way could be particularly beneficial for less experienced or capable individuals[92], further promoting diversity and opportunity in open innovation processes. Relatedly, LLMs can provide individuals with an 'outside view' when prompted accordingly, which can facilitate a kind of dialectical bootstrapping where an individual assumes several varying perspectives to repeatedly generate ideas[93,94]. Moreover, LLMs' ability to search and summarize vast amounts of information could help to surface overlooked but relevant inputs in

Table 1 | Paper overview

How LLMs can help CI	
Increasing accessibility and inclusion in online collaborations	LLMs can reduce barriers to participation and coordination costs by providing translation, writing assistance or summarizations, or even acting on behalf of human individuals, leading to new forms of diverse, equitable collaboration.
Accelerating idea generation	LLMs can enhance the efficiency of generating ideas by posing as a crowd to be queried, augmenting human individuals by providing starting points and (re)combining seemingly disparate ideas.
Mediating deliberative processes	LLMs can provide deliberation support to human individuals by prompting them to consider specific information or rephrase arguments, and/or serve as a facilitator to oversee speaker queues and request elaborations on newly risen topics.
Aggregating information across a group	LLMs can generate summary statements that synthesize disparate views, clarify shared objectives and identify areas of agreement.
How LLMs can hurt CI	
Disincentivizing individuals from contributing to collective knowledge commons	Widespread use of LLMs as substitutes for open knowledge commons (for example, wikis) can threaten the health of such commons by deterring individuals from engaging with original source material and making new contributions.
Propagating illusions of consensus and pluralistic ignorance	If certain viewpoints are underrepresented or excluded entirely from an LLM's training data, interactions with an LLM may lead people to believe there is a consensus on an issue even if none exists.
Reducing functional diversity among individuals	Reliance on one or few LLMs can homogenize individuals' privately held beliefs and lead to premature convergence by limiting opportunities for diverse social learning strategies.
Removing friction in the production of false or misleading information	LLMs can deliver erroneous information privately to users en masse and induce collective biases, and LLMs can be used to aid deliberate disinformation campaigns.
Recommendations	
Truly open LLMs	Open access to model weights, code, data sources and model checkpoints would help prevent a monolithic model landscape.
Greater computational resources for researchers	Government-subsidized computational resources should be made available to enable new, diverse, independent research on LLMs.
Third-party oversight of LLM use	LLM developers must be open to external audits, content detection mechanisms and other measures to increase understanding of how LLMs are used in the real world.

groups' ideation processes to facilitate breakthrough ideas, which often come from recombining existing knowledge, particularly from seemingly disconnected fields[95–97]. These complementary strengths point to the potential for future LLM–human teams optimized for CI.

Mediating deliberative processes

A central challenge for efforts to elicit CI, such as deliberative democracy, is that people may not engage in sufficiently informed, meaningful ways. It is argued that not only are many people unwilling to participate in collective, deliberative processes[98], but also they are not competent enough to do so, due to limited cognitive bandwidth and motivated reasoning[99–101]. Yet it is unclear whether collective, deliberative processes sometimes fail due to a lack of informed, meaningful engagement, or whether people do not engage in such processes due to disillusionment with the processes themselves[102], or whether people may even be engaging in rational inattention[103]. It seems plausible that LLMs could be used to increase the attractiveness and decrease the cognitive load of engaging in deliberative processes. For instance, an LLM could take the role of an interlocutor who actively engages with contributors, asking guiding questions to help them to clearly express their opinions. Just as decision support software can augment a human analyst by automating quantitative reports[104], an LLM could provide user-friendly deliberation support, potentially by prompting participants to consider specific information or refine their arguments on the basis of the LLM's evaluation of previously contributed content[60,61,105–108]. The value of LLMs as a cognitive aid has already been demonstrated in education, where LLMs have guided self-learning in adults[109], and in divisive political debates, where LLMs have suggested ways to rephrase arguments to increase the perceived quality of debate without changing the core content[110].

Alternatively, an LLM could act as a facilitator, overseeing the deliberative process as a whole. For example, an LLM could manage speaker queues (that is, who should speak to whom, when and on what topic) or request elaboration if a statement cannot be confidently classified with a previously seen topic label[111,112]. Further research is needed to develop, deploy and evaluate such capabilities at scale, but in recent demonstrations of AI-facilitated deliberation, such as the Stanford Online Deliberation Platform, participants have reported satisfaction with the process and felt that it proceeded with trust and empathy[113,114].

Aggregating information across a group

When diverse individuals collaborate, differences in language, culture, education or expertise can pose challenges to effective communication and coordination[115]. In such cases, LLMs could help to bridge divides by generating summary statements that synthesize disparate views, clarify shared objectives and identify areas of agreement.

The summarization of opinions has long been of interest to the natural language processing community, and recent advances in LLM development allow for fine-grained sentence selection and the generation of meta-reviews summarizing multiple opinions[116–118], which broadens the scope of potential applications. These capabilities suggest that LLMs can identify subtleties of disagreement or conditional agreement and rephrase ideas in ways that enable others to relate to them. For example, Bakker et al. fine-tuned LLMs to generate "consensus statements" that are designed to maximize group-level agreement on the basis of a set of input opinions[119]. Relatedly, Small et al. developed an LLM that can aggregate a much larger set of opinions, showing that LLMs can be used to process large amounts of written opinions or comments[112]. Continuing in this direction, LLM-powered collective decision-making systems could promote efficient coordination by formulating consensus-based judgements tailored to each participant's perspective on the basis of massive quantities of often-vague stances expressed in natural language.

How LLMs can harm CI

In some instances, the risks of LLMs harming CI go hand-in-hand with potential benefits to CI. For example, LLMs can promote coordination by generating consensus statements, but their opaqueness can create illusions of consensus or obscure important differences in opinion between groups. In other cases, risks imposed by LLMs relate to their position within the broader information environment (for example, disincentivizing individuals from contributing to transparent, collective knowledge commons). In this section, we outline risks considering both current and potential near-term developments.

Disincentivizing individuals from contributing to collective knowledge commons

A prime example of CI is the crowdsourced development and maintenance of online collective knowledge commons. These shared, open resources—such as wikis, Internet archives, open-source software repositories and discussion boards—promote interaction among large groups of individuals that can produce outcomes that would be impossible for any one individual. Typically, these resources involve some degree of shared ownership and transparency on how individual contributions are handled, which naturally incentivizes individuals to contribute because it is clear whether and how contributions will be recognized. However, the widespread use of LLMs as substitutes for these collective knowledge commons could create an information environment that undermines this form of CI[120]. This logic applies not only to large-scale, Internet-based commons but also to smaller-scale commons such as organizational teams' wikis, and, if human individuals turn to LLMs rather than one another for collaboration, also analogue commons in workplaces (for example, 'lunch and learn' meetings).

The efficiency and availability of LLMs for content generation can lead people to rely on them rather than engaging with and contributing to other open collective knowledge commons. This may subsequently decrease the production rate of new human-generated material (as opposed to LLM-recycled material) and, in turn, decrease the quantity and quality of collective knowledge that can be shared, remixed and learned from[121,122]. This also threatens the health of the platforms that both feed these algorithms their training data and serve as essential pillars of the online information environment. The plentiful and high-quality user-generated content on platforms such as Reddit, Wikipedia and Stack Overflow is often used to train LLMs, potentially resulting in a paradox of reuse: as people increasingly rely on LLMs for information search, they cease to engage with the original source material. This trend could decrease audience size and engagement on the platforms, with implications for their long-term vibrancy, as people may be less inclined to contribute due to a reduced audience and a lack of credit when their content reaches people anonymously through an LLM[120,123–126].

Furthermore, the rising prominence of LLMs (and generative models in general) may disincentivize individuals from releasing their creative work into the public domain at all (for example, due to copyright issues[127]). Concerns over labour replacement or the misuse of their contributions as mere data for training LLMs could deter people from open-sourcing code or content. For example, the 2023 Writers' Guild of America strike reflected serious concerns around AI and included demands to ban studios from using writers' creative materials for training LLMs. If more content moves towards private, non-scrapable platforms, it not only puts substantial control in the hands of those platforms but also further constrains the diverse, open nature of the knowledge commons that promotes CI.

Propagating illusions of consensus and pluralistic ignorance

Although LLMs can assist CI by aggregating information across a group to, for example, identify or generate consensus statements, reliance on LLMs as aggregators also introduces novel risks to CI.

When an LLM is used in a collective process to aggregate information, it is most likely to generate responses that reflect the opinions or beliefs that appear most frequently in the training data. Yet certain viewpoints may be underrepresented or excluded entirely from the training data, leading LLMs to provide responses that neglect alternative opinions or less prevalent facts. As people interact with the model and treat it as an authority despite its opaqueness (Fig. 1), they may see responses leaning towards a specific perspective, leading them to believe there is a consensus on that issue even if none exists. In turn, this may lead to the propagation of illusions of consensus, whereby the repeated claim of a single source is misinterpreted as a true consensus supported by multiple independent sources[128]. Combined with the spiral-of-silence mechanism—where individuals become less likely to voice their opinions publicly as they perceive them to be in the minority[129]—this could eventually lead to groups and societies with non-pluralistic views on multifaceted matters.

These issues are largely absent from successful examples of traditional CI systems such as Wikipedia, which is based on principles such as neutrality and pluralism[130]. The transparency of the editorial process on Wikipedia, coupled with well-documented revisions of articles and discussions towards consensus-building, is central to upholding these principles[131]. Additionally, Wikipedia's multilingualism enables different narratives and viewpoints to coexist and coevolve in various language editions simultaneously[132]. Previous efforts aimed at unifying facts in a shared database, such as Wikidata, have faced criticism as this would require uniformity of narratives across languages and communities[133]. This criticism seems analogous to the ongoing development of authoritative LLMs with opaque, proprietary training data with little regard for what it means for the representation of marginalized opinions.

Reducing functional diversity among individuals

Widespread individual reliance on LLMs as an information source has the potential to undermine CI by dissolving one of its key components: functional diversity[24,25]. Generally, the accuracy of collective, aggregate judgements is maximized when individual group members are independent of each other, such that they retain their diverse approaches to the task at hand and correct for one another's errors[134–138]. If group members consult the same LLMs, they might introduce a correlation between their sources of information[139]. Shared information—regardless of its quality—may limit the benefits of aggregation due to higher similarity between individual responses[140,141]. In a worst-case scenario, the frequency of low-quality information may even increase if LLMs provide bad advice.

Beyond potentially homogenizing individuals' privately held beliefs, LLMs may further reduce functional diversity when embedded into interactive group processes—for example, when used in open-ended tasks such as serving as sounding boards in brainstorming activities (see section on 'Accelerating idea generation'). In these tasks, it is often advantageous for groups to foster individuals' diverse search strategies and divide attentional resources to cover larger grounds of a solution space[48]. However, an LLM providing suggestions on how to start a problem-solving process may suggest multiple similar strategies, leading to premature convergence on a path or solution—although this could be partially mitigated by increasing the LLM's 'temperature' (that is, randomness) to elicit more diverse suggestions. LLMs used to mediate a deliberative process (see section on 'Mediating deliberative processes') could improve the efficiency of communication between individuals but may also lead to premature convergence by limiting opportunities for diverse social learning strategies, which can be beneficial for problem-solving[142,143] (but see ref. 144 for a discussion of when such social influence may be detrimental to CI for estimation tasks). What may be perceived as inefficiency or conflict in the moment could instead be necessary to cultivate the transient diversity needed to reach a high-quality solution[48].

Furthermore, as LLMs become progressively integrated into cultural processes (for example, education) in the future, they might threaten cultural and functional diversity at the population level[145]. Consequently, LLMs could align not just the information that is readily available in the present but also the thought processes that govern the acquisition, spread and aggregation of new information.

Removing friction in the production of false or misleading information

The speed and ease with which LLMs can be used to generate coherent content is key to many ways they can promote CI. Yet this capability could also jeopardize CI in two ways. First, LLMs are currently prone to 'hallucinating': generating incorrect information in response to requests[146]. False or misleading information is not a new feature of the information environment, but the way it is delivered by LLMs may pose a novel risk (Fig. 1). Whereas false or misleading information in the public domain can be fact-checked, evaluated with veracity cues (for example, tracing the original source of a claim) or otherwise corrected, private LLM-to-user dialogue is largely untraceable. A coherent, authoritative-sounding hallucination could be delivered en masse without any oversight, consequently inducing collective bias towards erroneous information. This problem may worsen if, as the issue of hallucinations is ameliorated (for example, by augmenting LLMs with information retrieval from accepted knowledge sources[147]), the authority of LLMs' output becomes harder to challenge, and people use them to verify information from other channels.

Second, LLMs' ability to rapidly generate content could undermine CI by aiding deliberate disinformation campaigns[4,148] (but see ref. 149 for counter-arguments). Just as LLMs can lower barriers to entry for benevolent online collaborations, they could enable propagandists to target audiences they would not be able to communicate with otherwise[148]. LLMs could also be used to produce many subtly distinct messages that avoid automated detection[148] and generally drive down the costs of creating such content by as much as 70%[150] (but see ref. 151, which argues that the costs of content distribution may be more crucial than the costs of content creation). Moreover, if LLMs and their training data are controlled or co-opted by ill-intentioned actors, the private, LLM-to-user delivery of information could be leveraged to steer civic deliberation, even without the use of disinformation per se. For example, a state-run LLM could be trained on data with anti-state content deliberately excluded without citizens' knowledge.

Striking a balance

What can be done to curb the challenges LLMs pose to CI without undercutting the opportunities? We propose three recommendations: truly open LLMs, better access to computational resources and LLMs for researchers, and greater oversight of how LLMs are used and what harms they cause in the real world.

Unlike accessing LLMs through a gated API, having truly open LLMs means the model weights, codebase and details on the training procedure and data sources are available publicly[152], thereby facilitating research and development[153,154]. Although pushing for truly open LLMs alone will not prevent the centralization effects that accompany LLM development[155], it steers away from a monolithic model landscape dominated by just a few developers. In doing so, this ensures that the bounds of acceptable speech are not defined by a handful of private companies and avoids illusions of consensus, as downstream users can fine-tune their own models. Although making LLMs openly available runs a risk of facilitating model misuse (for example, for cyberattacks), this risk is unlikely to be beyond what is already experienced with closed LLMs and existing technologies such as web search[154].

Truly open access is useful for letting users build on and fine-tune LLMs, but it does not address the vast computational costs of research and development of LLMs for CI. The high cost of developing and deploying LLMs has put such research beyond the capabilities of all but the best-resourced companies and research groups. Government efforts to provide computational resources to academics and researchers, such as the US National Artificial Intelligence Research Resource[156], can increase access to the computational resources needed to build, fine-tune and research LLMs, including examining the risks and benefits they hold for CI.

A prerequisite for addressing risks effectively is knowing not only how LLMs were developed and the data sources they were trained on (as called for by the European Union AI Act[157]) but also what they are actually being used for and what harms they are causing. Past research has shown the potential risks of LLMs in the laboratory, and AI developers must be forthcoming about how their models are used in the real world[158] and cooperate with reasonable third-party oversight (for example, through external audits[159,160] and content detection mechanisms[161]). This can help developers, researchers and policymakers to better understand what measures are needed to reduce potential harms to users. Developers of LLMs often impose use restrictions, specifying what their models can and cannot be used for. Providers of products and services that use LLMs have important insight into how users interact with them and therefore must enforce use restrictions appropriately. Going further, the CI of LLM users themselves could be leveraged for participatory, post-deployment oversight (for example, as in Wikipedia's Objective Revision Evaluation Service, where oversight and governance of content moderation algorithms are delegated to Wikipedians)[162]. In either case, transparency about how LLMs are used could also translate into better interventions or standardized protocols for contexts where users may be unaware of LLMs' limitations—for example, to prevent LLM misuse in organizational processes. Protecting the platforms where risks to CI materialize may also be an effective approach to attenuating those risks. For example, disinformation typically spreads through social media platforms. Here, shoring up existing defences, such as third-party fact-checking[163], is likely to be an effective near-term solution[164–167]. However, the effectiveness of such defences may dwindle if the reach and sophistication of LLM-powered disinformation campaigns increase, making our call for greater oversight of LLM usage all the more pertinent.

Other forms of generative AI

Alongside the recent development of LLMs, there have been notable advances in other forms of generative AI. These include AI systems that take text or other multimodal inputs to generate images (for example, Midjourney), videos (for example, Sora) or audio (for example, SeamlessM4T). Although many of these systems are not yet developed enough to reshape CI in an immediately meaningful way, it is very likely that these other emerging forms of generative AI could affect CI in the future.

Do the benefits, risks and recommendations related to the LLM–CI connection that we have highlighted extend to these other forms of generative AI? At present, AI applications for image, video and audio generation are used primarily for entertainment purposes, and there is currently insufficient empirical evidence to offer broad conjectures. However, consider the most developed and accessible of these other forms: image-generating AI applications such as DALL·E, Midjourney and Stable Diffusion. Just as LLMs can help to accelerate the generation of ideas expressed in natural language or computer code (see 'Accelerating idea generation'), image-generating AI applications can help to accelerate the generation of candidate visual designs and prototypes as well as encourage divergent thinking[168,169]. Image-generating AI can also be trained on visual designs from different designers to aggregate their non-linguistic styles, mimicking the kind of information aggregation discussed in the section on 'Aggregating information across a group'. Yet, unsurprisingly, the implications of image-generating AI are not all positive. If widely adopted, image-generating applications can, for example, disincentivize individuals from contributing to image-based knowledge commons (for example, Flickr and Shutterstock) for the same reasons that LLMs can disincentivize contributions to text-based

BOX 3

Open research questions

How can LLMs boost CI rather than replace it?
Understanding the dynamics of trust and user interaction in LLM-facilitated scenarios is crucial to ensuring that LLMs become an integral part of human CI rather than replace it altogether. This requires exploring the acceptance and effectiveness of LLMs in collaborative contexts[191,192].

How can the detrimental homogenization of collective knowledge via LLMs be avoided and diverse views, including those of minority groups, be represented?
For the effective integration of LLMs and CI, research is needed to develop methods for evaluating whether diverse perspectives are represented and respected during training, fine-tuning and application to prevent unintended biases.

How can people's functional diversity be maintained or even enhanced when interacting with LLMs?
The use of LLMs for different cognitive tasks could narrow people's attentional resources to a single source, thereby reducing opportunities for a group to benefit from the functional diversity of its members. Research is needed to identify how LLMs can expand rather than narrow the scope of approaches to tasks.

How should credit and accountability be apportioned when collective outcomes are co-created with LLMs?
As LLMs integrate into collective processes, it will be crucial to establish frameworks for explainable credit and accountability allocation to guide responsible usage, incentivization and oversight.

How should policy measures geared towards CI stewardship be designed and enforced?
Measures such as transparency reports, third-party oversight, usage restrictions and standardized protocols can promote responsible LLM use in CI applications, prevent monopolization by dominant providers and ensure a diverse CI landscape. However, such measures can have an effect only if they are enforceable.

commons (for example, Wikipedia and Stack Overflow), as described under 'Disincentivizing individuals from contributing to collective knowledge commons'[120]. Similar to the way LLMs might be used by malicious actors in the production of false or misleading information (Section 3.4), multimedia-generating AI can also aid such actors, but potentially with even more dangerous effects given the evidential primacy of what we see and hear over what we read[170].

Beyond those intuitive parallels between LLMs and image-generating AI, it seems hasty to add further speculation on the capability of generative AI at large to reshape CI. We do not currently foresee reasons to restrict our call for truly open models, computational resources for independent researchers and third-party oversight of model usage to LLMs rather than generative AI more broadly. However, analysing the implications of any new AI technology requires a nuanced evaluation of strengths and weaknesses relative to particular use cases, which necessarily goes beyond purely technical benchmarking to include assessment of societal impacts and acceptability.

Conclusion

With the rapid uptake of LLM applications such as ChatGPT, the online information environment is undergoing a transformation (Fig. 1). How this transformation will affect individuals' information search, reasoning and decision-making is not yet clear, but proactively anticipating potential unexpected effects that may emerge at a collective level is vital (see Box 3 for open research questions). As tools for and of CI, LLMs can simultaneously power new, heightened CI and threaten society's ability to collectively solve problems. By bringing together the perspectives of researchers from diverse disciplines across academia and industry, this paper synthesizes current thinking and motivates further investigations into the intersection of LLMs and CI.

References

1. Hu, K. ChatGPT sets record for fastest-growing user base—analyst note. *Reuters* (2 February 2023).
2. Mitchell, M. & Krakauer, D. C. The debate over understanding in AI's large language models. *Proc. Natl Acad. Sci. USA* **120**, e2215907120 (2023).
3. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: can language models be too big? In *Proc. 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (Association for Computing Machinery, 2021).
4. Weidinger, L. et al. Taxonomy of risks posed by language models. In *Proc. 2022 ACM Conference on Fairness, Accountability, and Transparency* 214–229 (Association for Computing Machinery, 2022).
5. Bubeck, S. et al. Sparks of artificial general intelligence: early experiments with GPT-4. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2303.12712 (2023).
6. Hilbert, M. What is the content of the world's technologically mediated information and communication capacity: how much text, image, audio, and video? *Inf. Soc.* **30**, 127–143 (2014).
7. Malone, T. W. & Bernstein, M. S. *Handbook of Collective Intelligence* (MIT Press, 2015).
8. Flack, J., Ipeirotis, P., Malone, T. W., Mulgan, G. & Page, S. E. Editorial to the inaugural issue of *Collective Intelligence*. *Collect. Intell.* **1**, 263391372211141 (2022).
9. Malone, T. W., Laubacher, R. & Dellarocas, C. The collective intelligence genome. *IEEE Eng. Manage. Rev.* **38**, 38–52 (2010).
10. Suran, S., Pattanaik, V., Yahia, S. B. & Draheim, D. Exploratory analysis of collective intelligence projects developed within the EU-Horizon 2020 framework. In *Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019: Proceedings* Pt 2 (eds Nguyen, N. T. et al.) 285–296 (Springer, 2019).
11. Suran, S., Pattanaik, V. & Draheim, D. Frameworks for collective intelligence: a systematic literature review. *ACM Comput. Surv.* **53**, 1–36 (2021).
12. Woolley, A. W. & Gupta, P. Understanding collective intelligence: investigating the role of collective memory, attention, and reasoning processes. *Perspect. Psychol. Sci.* **19**, 344–354 (2024).
13. Salminen, J. Collective intelligence in humans: a literature review. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1204.3401 (2012).
14. Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations* (W. Strahan and T. Cadell, 1776).
15. Hayek, F. A. The use of knowledge in society. *Am. Econ. Rev.* **35**, 519–530 (1945).
16. Hastie, R. & Kameda, T. The robust beauty of majority rules in group decisions. *Psychol. Rev.* **112**, 494–508 (2005).
17. Surowiecki, J. *The Wisdom of Crowds* (Anchor Books, 2005).
18. Dasgupta, P. & Maskin, E. On the robustness of majority rule. *J. Eur. Econ. Assoc.* **6**, 949–973 (2008).
19. Arrow, K. J. et al. The promise of prediction markets. *Science* **320**, 877–878 (2008).
20. Simon, H. A. *Rational Decision-Making in Business Organizations* (American Economic Association, 1978).

21. Vélez, N., Christian, B., Hardy, M., Thompson, B. D. & Griffiths, T. L. How do humans overcome individual computational limitations by working together? *Cogn. Sci.* **47**, e13232 (2023).
22. Gupta, P., Nguyen, T. N., Gonzalez, C. & Woolley, A. W. Fostering collective intelligence in human–AI collaboration: laying the groundwork for COHUMAIN. *Top. Cogn. Sci*. https://doi.org/10.1111/tops.12679 (2023).
23. Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W. & Woolley, A. W. Quantifying collective intelligence in human groups. *Proc. Natl Acad. Sci. USA* **118**, e2005737118 (2021).
24. Hong, L. & Page, S. E. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl Acad. Sci. USA* **101**, 16385–16389 (2004).
25. Bang, D. & Frith, C. D. Making better decisions in groups. *R. Soc. Open Sci.* **4**, 170193 (2017).
26. Harrison, D. A., Price, K. H., Gavin, J. H. & Florey, A. T. Time, teams, and task performance: changing effects of surface- and deep-level diversity on group functioning. *Acad. Manage. J.* **45**, 1029–1045 (2002).
27. Mohammed, S. & Angell, L. C. Surface- and deep-level diversity in workgroups: examining the moderating effects of team orientation and team process on relationship conflict. *J. Organ. Behav.* **25**, 1015–1039 (2004).
28. Phillips, K. W. & Loyd, D. L. When surface and deep-level diversity collide: the effects on dissenting group members. *Organ. Behav. Hum. Decis. Process.* **99**, 143–160 (2006).
29. Condorcet, N. *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix* (Chelsea, 1785).
30. Grofman, B., Owen, G. & Feld, S. L. Thirteen theorems in search of the truth. *Theory Decis.* **15**, 261–278 (1983).
31. Page, S. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* New Edn (Princeton Univ. Press, 2008).
32. Hong, L. & Page, S. E. in *Collective Wisdom* (eds Landemore, H. & Elster, J.) 56–71 (Cambridge Univ. Press, 2012).
33. Ladha, K. K. The Condorcet jury theorem, free speech, and correlated votes. *Am. J. Polit. Sci.* **36**, 617–634 (1992).
34. Kameda, T., Toyokawa, W. & Tindale, R. S. Information aggregation and collective intelligence beyond the wisdom of crowds. *Nat. Rev. Psychol.* **1**, 345–357 (2022).
35. Laan, A., Madirolas, G. & De Polavieja, G. G. Rescuing collective wisdom when the average group opinion is wrong. *Front. Robot. AI* **4**, 56 (2017).
36. Lyon, A. & Pacuit, E. in *Handbook of Human Computation* (ed. Michelucci, P.) 599–614 (Springer New York, 2013).
37. Landemore, H. & Page, S. E. Deliberation and disagreement: problem solving, prediction, and positive dissensus. *Polit. Phil. Econ.* **14**, 229–254 (2015).
38. List, C. The theory of judgment aggregation: an introductory review. *Synthese* **187**, 179–207 (2012).
39. Centola, D. The network science of collective intelligence. *Trends Cogn. Sci.* **26**, 923–941 (2022).
40. Baron, R. S. So right it's wrong: groupthink and the ubiquitous nature of polarized group decision making. *Adv. Exp. Soc. Psychol.* **37**, 219–253 (2005).
41. Hahn, U., Von Sydow, M. & Merdes, C. How communication can make voters choose less well. *Top. Cogn. Sci.* **11**, 194–206 (2019).
42. Hahn, U., Hansen, J. U. & Olsson, E. J. Truth tracking performance of social networks: how connectivity and clustering can make groups less competent. *Synthese* **197**, 1511–1541 (2020).
43. Jönsson, M. L., Hahn, U. & Olsson, E. J. The kind of group you want to belong to: effects of group structure on group accuracy. *Cognition* **142**, 191–204 (2015).
44. Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl Acad. Sci. USA* **114**, E5070–E5076 (2017).
45. Zollman, K. J. S. The communication structure of epistemic communities. *Phil. Sci.* **74**, 574–587 (2007).
46. Zollman, K. J. S. The epistemic benefit of transient diversity. *Erkenntnis* **72**, 17–35 (2010).
47. Zollman, K. J. S. Network epistemology: communication in epistemic communities. *Phil. Compass* **8**, 15–27 (2013).
48. Smaldino, P. E., Moser, C., Pérez Velilla, A. & Werling, M. Maintaining transient diversity is a general principle for improving collective problem solving. *Perspect. Psychol. Sci.* **19**, 454–464 (2023).
49. Bak-Coleman, J. B. et al. Stewardship of global collective behavior. *Proc. Natl Acad. Sci. USA* **118**, e2025764118 (2021).
50. Suran, S. et al. Building global societies on collective intelligence: challenges and opportunities. *Digit. Gov. Res. Pract.* **3**, 1–6 (2022).
51. Tsvetkova, M., Yasseri, T., Pescetelli, N. & Werner, T. Human-machine social systems. *Nat. Hum. Behav*. https://doi.org/10.48550/arXiv.2402.14410 (in press).
52. Cui, H. & Yasseri, T. AI-enhanced collective intelligence. *Patterns* https://doi.org/10.48550/arXiv.2403.10433 (in press).
53. Ovadya, A. 'Generative CI' through collective response systems. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2302.00672 (2023).
54. Zaremba, W. et al. Democratic inputs to AI. *OpenAI* https://openai.com/blog/democratic-inputs-to-ai (2023).
55. Lorenz-Spreen, P., Oswald, L., Lewandowsky, S. & Hertwig, R. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nat. Hum. Behav.* **7**, 74–101 (2022).
56. Du, Y., Li, S., Torralba, A., Tenenbaum, J. B. & Mordatch, I. Improving factuality and reasoning in language models through multiagent debate. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2305.14325 (2023).
57. Wu, Q. et al. AutoGen: enabling next-gen LLM applications via multi-agent conversation. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2308.08155 (2023).
58. Yoffe, L., Amayuelas, A. & Wang, W. Y. DebUnc: mitigating hallucinations in large language model agent communication with uncertainty estimations. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2407.06426 (2024).
59. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at *arXiv* http://arxiv.org/abs/2108.07258 (2022).
60. Törnberg, P. ChatGPT-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot learning. Preprint at *arXiv* http://arxiv.org/abs/2304.06588 (2023).
61. Rathje, S. et al. GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl Acad. Sci. USA* **131**, e2308950121 (2024).
62. Goldstein, D. G., McAfee, R. P. & Suri, S. The wisdom of smaller, smarter crowds. In *Proc. 15th ACM Conference on Economics and Computation* 471–488 (Association for Computing Machinery, 2014).
63. Mannes, A. E., Soll, J. B. & Larrick, R. P. The wisdom of select crowds. *J. Pers. Soc. Psychol.* **107**, 276–299 (2014).
64. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
65. Hahn, U. Collectives and epistemic rationality. *Top. Cogn. Sci.* **14**, 602–620 (2022).
66. Lafont, C. Deliberation, participation, and democratic legitimacy: should deliberative mini-publics shape public policy? *J. Polit. Phil.* **23**, 40–63 (2015).
67. Landemore, H. *Can AI Bring Deliberation To The Masses?* (Stanford Institute for Human-Centered Artificial Intelligence, 2022).

68. Cohen, R. L. Procedural justice and participation. *Hum. Relat.* **38**, 643–663 (1985).
69. Greenberg, J. & Folger, R. in *Basic Group Processes* (ed. Paulus, P. B.) 235–256 (Springer New York, 1983).
70. El Zein, M., Bahrami, B. & Hertwig, R. Shared responsibility in collective decisions. *Nat. Hum. Behav.* **3**, 554–559 (2019).
71. Fishkin, J. S. *When the People Speak: Deliberative Democracy and Public Consultation* (Oxford Univ. Press, 2009).
72. Steiner, I. D. Models for inferring relationships between group size and potential group productivity. *Behav. Sci.* **11**, 273–283 (1966).
73. Steiner, I. D. *Group Process and Productivity* (Academic Press, 1972).
74. Hill, G. W. Group versus individual performance: are *N*+1 heads better than one? *Psychol. Bull.* **91**, 517–539 (1982).
75. Almaatouq, A., Alsobay, M., Yin, M. & Watts, D. J. Task complexity moderates group synergy. *Proc. Natl Acad. Sci. USA* **118**, e2101062118 (2021).
76. Straub, V. J., Tsvetkova, M. & Yasseri, T. The cost of coordination can exceed the benefit of collaboration in performing complex tasks. *Collect. Intell.* **2**, 263391372311569 (2023).
77. Zhu, W. et al. Multilingual machine translation with large language models: empirical results and analysis. In *Findings of the Association for Computational Linguistics: NAACL 2024* (eds. Duh, K., Gomez, H. & Bethard, S.) 2765–2781 (Association for Computational Linguistics, 2024).
78. Bawden, R. & Yvon, F. Investigating the translation performance of a large multilingual language model: the case of BLOOM. In *Proc. 24th Annual Conference of the European Association for Machine Translation* (eds. Nurminen, M. et al.) 157–170 (European Association for Machine Translation, 2023).
79. Berdejo-Espinola, V. & Amano, T. AI tools can improve equity in science. *Science* **379**, 991 (2023).
80. Katsnelson, A. Poor English skills? New AIs help researchers to write better. *Nature* **609**, 208–209 (2022).
81. Romero-Olivares, A. L. Reviewers, don't be rude to nonnative English speakers. *Science* https://doi.org/10.1126/science.caredit.aaz7179 (2019).
82. Ramírez-Castañeda, V. Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: the case of Colombian researchers in biological sciences. *PLoS ONE* **15**, e0238372 (2020).
83. Brabham, D. C. Crowdsourcing as a model for problem solving: an introduction and cases. *Converg. Int. J. Res. N. Media Technol.* **14**, 75–90 (2008).
84. von Hippel, E. in *Handbook of the Economics of Innovation* (eds Hall, B. H. & Rosenberg, N.) Vol. 1, 411–427 (Elsevier, 2010).
85. Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V. & Lakhani, K. R. The crowdless future? Generative AI and creative problem solving. *Organ. Sci.* **0**, 1–19 (2024).
86. Girotra, K., Meincke, L., Terwiesch, C. & Ulrich, K. T. Ideas are dimes a dozen: large language models for idea generation in innovation. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.4526071 (2023).
87. Argyle, L. P. et al. Out of one, many: using language models to simulate human samples. *Polit. Anal.* **31**, 337–351 (2023).
88. Jiang, H., Zhang, X., Cao, X., Breazeal, C., Roy, D. & Kabbara, J. PersonaLLM: investigating the ability of large language models to express personality traits. In *Findings of the Association for Computational Linguistics: NAACL 2024* (eds. Duh, K. et al.) 3605–3627 (Association for Computational Linguistics, 2024).
89. Salewski, L., Alaniz, S., Rio-Torto, I., Schulz, E. & Akata, Z. In-context impersonation reveals large language models' strengths and biases. In *Adv. Neur. Inf. Process. Syst. 36 (NeurIPS 2023)* (eds Oh, A. et al.) 72044–720579 (2023).
90. Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C. & Gilbert, E. How AI ideas affect the creativity, diversity, and evolution of human ideas: evidence from a large, dynamic experiment. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2401.13481 (2024).
91. Fink, A. et al. Stimulating creativity via the exposure to other people's ideas. *Hum. Brain Mapp.* **33**, 2603–2610 (2012).
92. Doshi, A. R. & Hauser, O. Generative artificial intelligence enhances creativity. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.4535536 (2023).
93. Herzog, S. M. & Hertwig, R. The wisdom of many in one mind: improving individual judgments with dialectical bootstrapping. *Psychol. Sci.* **20**, 231–237 (2009).
94. Herzog, S. M. & Hertwig, R. Harnessing the wisdom of the inner crowd. *Trends Cogn. Sci.* **18**, 504–506 (2014).
95. Schilling, M. A. & Green, E. Recombinant search and breakthrough idea generation: an analysis of high impact papers in the social sciences. *Res. Policy* **40**, 1321–1331 (2011).
96. Porciello, J., Ivanina, M., Islam, M., Einarson, S. & Hirsh, H. Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nat. Mach. Intell.* **2**, 559–565 (2020).
97. Weitzman, M. L. Recombinant growth. *Q. J. Econ.* **113**, 331–360 (1998).
98. Hibbing, J. R. & Theiss-Morse, E. *Stealth Democracy: Americans' Beliefs about How Government Should Work* (Cambridge Univ. Press, 2002).
99. Rosenberg, S. W. in *Deliberative Democracy* (eds Elstub, S. & McLaverty, P.) 98–117 (Edinburgh Univ. Press, 2014).
100. Achen, C. H. & Bartels, L. M. *Democracy for Realists: Why Elections Do Not Produce Responsive Government* (Princeton Univ. Press, 2017).
101. Sunstein, C. R. On a danger of deliberative democracy. *Daedalus* **131**, 120–124 (2002).
102. Neblo, M. A., Esterling, K. M., Kennedy, R. P., Lazer, D. M. J. & Sokhey, A. E. Who wants to deliberate—and why? *Am. Polit. Sci. Rev.* **104**, 566–583 (2010).
103. Maćkowiak, B., Matějka, F. & Wiederholt, M. Rational inattention: a review. *J. Econ. Lit.* **61**, 226–273 (2023).
104. Shim, J. P. et al. Past, present, and future of decision support technology. *Decis. Support Syst.* **33**, 111–126 (2002).
105. Donohoe, H., Stellefson, M. & Tennant, B. Advantages and limitations of the e-Delphi technique. *Am. J. Health Educ.* **43**, 38–46 (2012).
106. Dalkey, N. & Helmer, O. An experimental application of the Delphi method to the use of experts. *Manage. Sci.* **9**, 458–467 (1963).
107. Tetlock, P. E., Mellers, B. A., Rohrbaugh, N. & Chen, E. Forecasting tournaments: tools for increasing transparency and improving the quality of debate. *Curr. Dir. Psychol. Sci.* **23**, 290–295 (2014).
108. McAndrew, T. et al. Early human judgment forecasts of human monkeypox, May 2022. *Lancet Digit. Health* **4**, e569–e571 (2022).
109. Lin, X. Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. *Adult Learn.* **35**, 56–166 (2023).
110. Argyle, L. P. et al. AI chat assistants can improve conversations about divisive topics. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2302.07268 (2023).
111. Hadfi, R. et al. Conversational agents enhance women's contribution in online debates. *Sci. Rep.* **13**, 14534 (2023).
112. Small, C. T. et al. Opportunities and risks of LLMs for scalable deliberation with polis. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2306.11932 (2023).
113. Fishkin, J. et al. Deliberative democracy with the online deliberation platform. In *7th AAAI Conference on Human Computation and Crowdsourcing*, https://www.human-computation.com/2019/assets/papers/144.pdf (Association for the Advancement of Artificial Intelligence, 2019).

114. Miller, K. A moderator ChatBot for civic discourse. *Stanford HAI* https://hai.stanford.edu/news/moderator-chatbot-civic-discourse (2020).
115. Jackson, M. O. & Xing, Y. Culture-dependent strategies in coordination games. *Proc. Natl Acad. Sci. USA* **111**, 10889–10896 (2014).
116. Coavoux, M., Elsahar, H. & Gallé, M. Unsupervised aspect-based multi-document abstractive summarization. In *Proc. 2nd Workshop on New Frontiers in Summarization* (eds Wang, L. et al.) 42–47 (Association for Computational Linguistics, 2019).
117. Angelidis, S., Amplayo, R. K., Suhara, Y., Wang, X. & Lapata, M. Extractive opinion summarization in quantized transformer spaces. *Trans. Assoc. Comput. Linguist.* **9**, 277–293 (2021).
118. Suhara, Y., Wang, X., Angelidis, S. & Tan, W.-C. OpinionDigest: a simple framework for opinion summarization. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* (eds Jurafsky, D. et al.) 5789–5798 (Association for Computational Linguistics, 2020).
119. Bakker, M. et al. Fine-tuning language models to find agreement among humans with diverse preferences. In *Adv. Neur. Inf. Process. Syst. 35 (NeurIPS 2022)* (eds Koyejo, S. et al.) 38176–38189 (2022).
120. Huang, S. & Siddarth, D. Generative AI and the digital commons. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2303.11074 (2023).
121. Veselovsky, V., Ribeiro, M. H. & West, R. Artificial artificial artificial intelligence: crowd workers widely use large language models for text production tasks. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2306.07899 (2023).
122. del Rio-Chanona, M., Laurentsyeva, N. & Wachs, J. Are large language models a threat to digital public goods? Evidence from activity on Stack Overflow. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2307.07367 (2023).
123. Farič, N. & Potts, H. W. Motivations for contributing to health-related articles on Wikipedia: an interview study. *J. Med. Internet Res.* **16**, e260 (2014).
124. Javanmardi, S., Ganjisaffar, Y., Lopes, C. & Baldi, P. User contribution and trust in Wikipedia. In *Proc. 5th International ICST Conference on Collaborative Computing: Networking, Applications, Worksharing* (eds Joshi, J. & Zhang, T.) https://doi.org/10.4108/ICST.COLLABORATECOM2009.8376 (Institute of Electrical and Electronics Engineers, 2009).
125. Adaji, I. & Vassileva, J. in *Social Informatics* (eds Spiro, E. & Ahn, Y.-Y.) 3–13 (Springer International, 2016).
126. Blincoe, K., Sheoran, J., Goggins, S., Petakovic, E. & Damian, D. Understanding the popular users: following, affiliation influence and leadership on GitHub. *Inf. Softw. Technol.* **70**, 30–39 (2016).
127. Franceschelli, G. & Musolesi, M. Copyright in generative deep learning. *Data Policy* **4**, e17 (2022).
128. Desai, S. C., Xie, B. & Hayes, B. K. Getting to the source of the illusion of consensus. *Cognition* **223**, 105023 (2022).
129. Noelle-Neumann, E. The spiral of silence: a theory of public opinion. *J. Commun.* **24**, 43–51 (1974).
130. Wikipedia: five pillars. *Wikipedia* https://en.wikipedia.org/w/index.php?title=Wikipedia:Five_pillars (2023).
131. Yasseri, T. & Kertész, J. Value production in a collaborative environment: sociophysical studies of Wikipedia. *J. Stat. Phys.* **151**, 414–439 (2013).
132. Hecht, B. & Gergle, D. The tower of Babel meets Web 2.0: user-generated content and its applications in a multilingual context. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* 291–300 (Association for Computing Machinery, 2010).
133. Graham, M. The problem with Wikidata. *Atlantic* (6 April 2012).
134. Clemen, R. T. & Winkler, R. L. Limits for the precision and value of information from dependent sources. *Oper. Res.* **33**, 427–442 (1985).
135. Davis-Stober, C. P., Budescu, D. V., Dana, J. & Broomell, S. B. When is a crowd wise?. *Decision* **1**, 79–101 (2014).
136. Herzog, S. M., Litvinova, A., Yahosseini, K. S., Tump, A. N. & Kurvers, R. H. J. M. in *Taming Uncertainty* (eds Hertwig, R. et al.) 245–262 (MIT Press, 2019).
137. Kurvers, R. H. J. M. et al. How to detect high-performing individuals and groups: decision similarity predicts accuracy. *Sci. Adv.* **5**, eaaw9011 (2019).
138. Palley, A. B. & Soll, J. B. Extracting the wisdom of crowds when information is shared. *Manage. Sci.* **65**, 2291–2309 (2019).
139. Walzner, D. D., Fuegener, A. & Gupta, A. Managing AI advice in crowd decision-making. In *ICIS 2022 Proceedings* https://aisel.aisnet.org/icis2022/hci_robot/hci_robot/7 (Association for Information Systems, 2022).
140. Padmakumar, V. & He, H. Does writing with language models reduce content diversity? Preprint at *arXiv* https://doi.org/10.48550/arXiv.2309.05196 (2023).
141. Kleinberg, J. & Raghavan, M. Algorithmic monoculture and social welfare. *Proc. Natl Acad. Sci. USA* **118**, e2018340118 (2021).
142. Campbell, C. M., Izquierdo, E. J. & Goldstone, R. L. Partial copying and the role of diversity in social learning performance. *Collect. Intell.* **1**, 263391372210818 (2022).
143. Toyokawa, W., Whalen, A. & Laland, K. N. Social learning strategies regulate the wisdom and madness of interactive crowds. *Nat. Hum. Behav.* **3**, 183–193 (2019).
144. Almaatouq, A., Rahimian, M. A., Burton, J. W. & Alhajri, A. The distribution of initial estimates moderates the effect of social influence on the wisdom of the crowd. *Sci. Rep.* **12**, 16546 (2022).
145. Brinkmann, L. et al. Machine culture. *Nat. Hum. Behav.* **7**, 1855–1868 (2023).
146. OpenAI. GPT-4 technical report. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2303.08774 (2023).
147. Semnani, S., Yao, V., Zhang, H. & Lam, M. WikiChat: stopping the hallucination of large language model chatbots by few-shot grounding on wikipedia. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (eds. Bouamor, H., Pino, J. & Bali, K.) 2387–2413 (Association for Computational Linguistics, 2023).
148. Goldstein, J. A. et al. Generative language models and automated influence operations: emerging threats and potential mitigations. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2301.04246 (2023).
149. Simon, F. M., Altay, S. & Mercier, H. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harv. Kennedy Sch. Misinform. Rev*. https://doi.org/10.37016/mr-2020-127 (2023).
150. Musser, M. A cost analysis of generative language models and influence operations. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2308.03740 (2023).
151. Kapoor, S. & Narayanan, A. *How to Prepare for the Deluge of Generative AI on Social Media* (Knight First Amendment Institute, 2023).
152. Solaiman, I. The gradient of generative AI release: methods and considerations. In *2023 ACM Conference on Fairness, Accountability, and Transparency* 111–122 (Association for Computing Machinery, 2023).
153. Warso, Z. & Keller, P. Open source AI and the paradox of open. *Open Future* https://openfuture.eu/blog/open-source-ai-and-the-paradox-of-open (2023).
154. Kapoor, S. et al. On the societal impact of open foundation models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2403.07918 (2024).
155. Widder, D. G., West, S. & Whittaker, M. Open (for business): big tech, concentrated power, and the political economy of open AI. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.4543807 (2023).

156. *National Artificial Intelligence Initiative* https://www.ai.gov/nairrtf/ (National Artificial Intelligence Research Resource Task Force, 2024).
157. *Artificial Intelligence Act* (European Parliament, 2023).
158. Kapoor, S. & Narayanan, A. *Generative AI Companies Must Publish Transparency Reports* (Knight First Amendment Institute, 2023).
159. Mökander, J., Schuett, J., Kirk, H. R. & Floridi, L. Auditing large language models: a three-layered approach. *AI Ethics* https://doi.org/10.1007/s43681-023-00289-2 (2023).
160. Chang, Y. et al. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.* **15**, 39 (2024).
161. Knott, A. et al. Generative AI models should include detection mechanisms as a condition for public release. *Ethics Inf. Technol.* **25**, 55 (2023).
162. Berditchevskaia, A., Malliaraki, E. & Peach, K. *Participatory AI for Humanitarian Innovation: A Briefing Paper* (Nesta, 2021).
163. *Meta's Third-Party Fact-Checking Program* https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking (Meta Journalism Project, accessed 29 March 2024).
164. Porter, E. & Wood, T. J. The global effectiveness of fact-checking: evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proc. Natl Acad. Sci. USA* **118**, e2104235118 (2021).
165. Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. Fact-checking: a meta-analysis of what works and for whom. *Polit. Commun.* **37**, 350–375 (2020).
166. Carnahan, D. & Bergan, D. E. Correcting the misinformed: the effectiveness of fact-checking messages in changing false beliefs. *Polit. Commun.* **39**, 166–183 (2022).
167. Ecker, U. K. H. et al. The psychological drivers of misinformation belief and its resistance to correction. *Nat. Rev. Psychol.* **1**, 13–29 (2022).
168. Cai, A. et al. DesignAID: using generative AI and semantic diversity for design inspiration. In *Proc. ACM Collective Intelligence Conference* (eds Bernstein, M. et al.) 1–11 (Association for Computing Machinery, 2023).
169. Griebel, M., Flath, C. & Friesike, S. Augmented creativity: leveraging artificial intelligence for idea generation in the creative sphere. *ECIS 2020 Research-in-Progress Papers* https://aisel.aisnet.org/ecis2020_rip/77 (Association for Information Systems, 2020).
170. Wittenberg, C., Tappin, B. M., Berinsky, A. J. & Rand, D. G. The (minimal) persuasive advantage of political video over text. *Proc. Natl Acad. Sci. USA* **118**, e2114388118 (2021).
171. Radford, A. et al. Language models are unsupervised multitask learners. *OpenAI Blog* (2019).
172. Hoffmann, J. et al. Training compute-optimal large language models. In *36th Conf Neural Information Processing*, https://proceedings.neurips.cc/paper_files/paper/2022/file/c1e2faff6f588870935f114ebe04a3e5-Paper-Conference.pdf (NeurIPS, 2022).
173. Ouyang, L. et al. Training language models to follow instructions with human feedback. In *Adv. Neur. Inf. Process. Syst. 35 (NeurIPS 2022)* (eds Koyejo, S. et al.) 27730–27744 (2022).
174. Lee, A., Miranda, B. & Koyejo, S. Beyond scale: the diversity coefficient as a data quality metric demonstrates LLMs are pre-trained on formally diverse data. Preprint at *arXiv* http://arxiv.org/abs/2306.13840 (2023).
175. Atari, M., Xue, M. J., Park, P. S., Blasi, D. E. & Henrich, J. Which humans? Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/5b26t (2023).
176. Cao, Y. et al. Assessing cross-cultural alignment between ChatGPT and human societies: an empirical study. In *Proc. First Workshop on Cross-Cultural Considerations in NLP (C3NLP)* 53–67 (Association for Computational Linguistics, 2023).
177. Haller, P., Aynetdinov, A. & Akbik, A. OpinionGPT: modelling explicit biases in instruction-tuned LLMs. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2309.03876 (2023).
178. Levy, S. et al. Comparing biases and the impact of multilingual training across multiple languages. In *Proc. 2023 Conference on Empirical Methods in Natural Language Processing* (eds Bouamor, H. et al.) 10260–10280 (Association for Computational Linguistics, 2023).
179. Arora, A., Kaffee, L.-A. & Augenstein, I. Probing pre-trained language models for cross-cultural differences in values. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2203.13722 (2023).
180. Dietterich, T. G. in *Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science*, vol 1857, https://doi.org/10.1007/3-540-45014-9_1 (Springer Berlin Heidelberg, 2000).
181. Grossmann, I. et al. AI and the transformation of social science research. *Science* **380**, 1108–1109 (2023).
182. Bail, C. A. Can generative AI improve social science? *Proc. Natl Acad. Sci. USA* **121**, e2314021121 (2024).
183. Aher, G. V., Arriaga, R. I. & Kalai, A. T. Using large language models to simulate multiple humans and replicate human subject studies. In *Proc. 40th International Conference on Machine Learning* 337–371 (Proceedings of Machine Learning Research, 2023).
184. Dillion, D., Tandon, N., Gu, Y. & Gray, K. Can AI language models replace human participants? *Trends Cogn. Sci.* **27**, 597–600 (2023).
185. Epstein, J. M. & Axtell, R. *Growing Artificial Societies: Social Science from the Bottom Up* (Brookings Institution Press, 1996).
186. Bonabeau, E. Agent-based modeling: methods and techniques for simulating human systems. *Proc. Natl Acad. Sci. USA* **99**, 7280–7287 (2002).
187. Park, J. S. et al. Social simulacra: creating populated prototypes for social computing systems. In *Proc. 35th Annual ACM Symposium on User Interface Software and Technology* (eds Agrawala, M. et al.) 1–18 (Association for Computing Machinery, 2022).
188. Gao, C. et al. S^3: social-network simulation system with large language model-empowered agents. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2307.14984 (2023).
189. Horton, J. J. *Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?* Working Paper 31122 (National Bureau of Economic Research, 2023).
190. Chen, L., Zaharia, M. & Zou, J. How is ChatGPT's behavior changing over time? *Harvard Data Sci. Rev.* **6**, https://doi.org/10.1162/99608f92.5317da47 (2024).
191. Burton, J. W., Stein, M. & Jensen, T. B. A systematic review of algorithm aversion in augmented decision making. *J. Behav. Decis. Mak.* **33**, 220–239 (2020).
192. Glikson, E. & Woolley, A. W. Human trust in artificial intelligence: review of empirical research. *Acad. Manage. Ann.* **14**, 627–660 (2020).

Acknowledgements
We thank D. Ain for her meticulous editing. We also thank all participants of the summer retreat at the Center for Adaptive Rationality, Max Planck Institute for Human Development, who provided helpful feedback on the original conceptualization of this work. J.W.B. was supported by an Alexander von Humboldt Foundation Research Fellowship. T.Y. was funded by the Irish Research Council (grant no. IRCLA/2022/3217). S.M.H. and A.B. are funded by the European Union's Horizon Europe Programme (grant agreement ID 101070588) and UKRI (project no. 10037991). E.L.-L., S.M.H. and U.H. were funded by the Deutsche Forschungsgemeinschaft (project no. 458366841). S.L. was supported by Science Foundation Ireland (grant no. 12/RC/2289_P2). R.H.J.M.K. is funded by the Deutsche Forschungsgemeinschaft (project no. 45836684).

Competing interests

M.A.B. is employed by Google DeepMind. S. Huang and D.S. are employed by the Collective Intelligence Project. L.F. is affiliated with the Lamarr Institute of ML and AI, an associated partner in the OpenGPT-X project via Fraunhofer IAIS. The other authors declare no competing interests.

Additional information

Correspondence should be addressed to Jason W. Burton.

Peer review information *Nature Human Behaviour* thanks Giovanni Luca Ciampaglia and John Meluso for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Department of Digitalization, Copenhagen Business School, Frederiksberg, Denmark. [2]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. [3]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [4]Center for Cognitive and Decision Sciences, University of Basel, Basel, Switzerland. [5]Google DeepMind, London, UK. [6]UCL School of Management, University College London, London, UK. [7]Centre for Collective Intelligence Design, Nesta, London, UK. [8]Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany. [9]Bonn-Aachen International Center for Information Technology, University of Bonn, Bonn, Germany. [10]Lamarr Institute for Machine Learning and Artificial Intelligence, Bonn, Germany. [11]Collective Intelligence Project, San Francisco, CA, USA. [12]Center for Information Technology Policy, Princeton University, Princeton, NJ, USA. [13]Department of Computer Science, Princeton University, Princeton, NJ, USA. [14]School of Sociology, University College Dublin, Dublin, Ireland. [15]Geary Institute for Public Policy, University College Dublin, Dublin, Ireland. [16]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [17]Department of Psychological Sciences, Birkbeck, University of London, London, UK. [18]Science of Intelligence Excellence Cluster, Technical University Berlin, Berlin, Germany. [19]School of Information and Communication, Insight SFI Research Centre for Data Analytics, University College Dublin, Dublin, Ireland. [20]Oxford Internet Institute, Oxford University, Oxford, UK. [21]Deliberative Democracy Lab, Stanford University, Stanford, CA, USA. [22]Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA. ✉e-mail: jb.digi@cbs.dk