



## Open Research Online

### Citation

Suran, Shweta; Pattanaik, Vishwajeet; Kurvers, Ralf H. J. M.; Hallin, Carina Antonia; De Liddo, Anna; Krimmer, Robert and Draheim, Dirk (2022). Building Global Societies on Collective Intelligence: Challenges and Opportunities. Digital Government: Research and Practice, 3(4), article no. 31.

### URL

<https://oro.open.ac.uk/85655/>

### License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

### Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

### Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding



## Building Global Societies on Collective Intelligence: Challenges and Opportunities

SHWETA SURAN and VISHWAJEET PATTANAİK, Tallinn University of Technology

RALF KURVERS, Max Planck Institute for Human Development

CARINA ANTONIA HALLIN, IT University of Copenhagen and Massachusetts Institute of Technology

ANNA DE LIDDO, The Open University

ROBERT KRIMMER, Johan Skytte Institute for Political Studies

DIRK DRAHEIM, Tallinn University of Technology

Digital disruptions caused by the use of technologies like social media arguably present a formidable challenge to democratic values and in turn to collective intelligence. Challenges such as misinformation, partisan bias, polarization, and rising mistrust in institutions (including mainstream media) present a new constant threat to collectives both online and offline—amplifying the risk of turning ‘wise’ crowds ‘mad’ and rendering their actions counterproductive. Considering the increasingly important role crowds play in solving today’s socio-political, technological, and economical issues, and in shaping our future, it is vital to protect crowd-oriented systems against such disruptions. In this commentary, we identify time-critical challenges and potential solutions from emerging work on diversity, transparency, collective dynamics, and machine behavior that require urgent attention if future collective intelligence systems are to sustain their indispensable role as global deliberation instruments.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; • **Information systems** → **Collaborative and social computing systems and tools**; • **Social and professional topics** → **Government technology policy**;

Additional Key Words and Phrases: Collective intelligence, citizen science, democracy, digital disruptions, emerging research, information systems

### ACM Reference format:

Shweta Suran, Vishwajeet Pattanaik, Ralf Kurvers, Carina Antonia Hallin, Anna De Liddo, Robert Krimmer, and Dirk Draheim. 2022. Building Global Societies on Collective Intelligence: Challenges and Opportunities. *Digit. Gov. Res. Pract.* 3, 4, Article 31 (December 2022), 6 pages.

<https://doi.org/10.1145/3568169>

S. Suran and V. Pattanaik contributed equally to this work.

The work of R. Krimmer was financed in part through the European Union’s Horizon 2020 research and innovation programme under grant agreements 857622 and 959072.

Authors’ addresses: S. Suran, V. Pattanaik, and D. Draheim, Information Systems Group, Tallinn University of Technology, Akadeemia tee 15a, Tallinn, Estonia, 12618; emails: {shweta, vishwajeet.pattanaik, dirk.draheim}@taltech.ee; R. Kurvers, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany, 14195; email: kurvers@mpib-berlin.mpg.de; C. A. Hallin, Business IT Department, Collective Intelligence Research Group, IT University of Copenhagen, Copenhagen, Denmark, 2300, and MIT Center for Collective Intelligence, Massachusetts Institute of Technology, Cambridge, MA 02142; email: caah@itu.dk; A. De Liddo, Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom, MK7 6AA; email: anna.deliddo@open.ac.uk; R. Krimmer, Center for IT Impact Studies, Johan Skytte Institute for Political Studies, Tartu, Estonia, 51003; email: robert.krimmer@taltech.ee.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.

2639-0175/2022/12-ART31 \$15.00

<https://doi.org/10.1145/3568169>

## 1 INTRODUCTION

The concept of **Collective Intelligence (CI)**—that is, “groups of individuals acting collectively in ways that seem intelligent” [17, p. 3]—harks back millennia and has continued to evolve in concert with the novel communication means of the different eras. The progression of CI changed dramatically with the advent of the Internet (and the social web), the latter offering unprecedented means for mobilizing collectives, fueling the emergence of a plethora of novel CI applications [24]. These CI applications can be seen across a wide and diverse range of services and activities such as knowledge accumulation and exchange (e.g., Wikipedia [17]), crowd-sourcing platforms for tackling global issues like climate change (e.g., Climate CoLab [12]) and the COVID-19 pandemic, and global medical diagnosis initiatives (e.g., Human Dx [6]). A newer class of CI applications that have emerged recently are citizen-owned **Civic Techs (CTs)** [11] (e.g., vTaiwan, CitizenLab, Participedia, and ProDemos) and government-owned **Citizen Engagement Platforms (CEPs)** [22] (e.g., Decide Madrid, Le Grand Débat National, MyGov.In, and Rahvakogu). Designed to strengthen the relationship between citizens and governments, the emergence of CTs and CEPs has only further accentuated the role of CI systems in citizens’ lives worldwide [24].

Social media has become a crucial component in most of these novel CI applications. ‘Activities’ such as liking/disliking, commenting, and sharing have become a prevalent means for users to express themselves, and platforms like Facebook and Twitter have become a foremost means for mobilizing collectives—necessitating both private and public organizations to rely on said social media ‘activities’ and platforms for harnessing citizens’ CI. Consequently, challenges encountered on social media platforms (like echo chambers, misinformation, and botnets) [21] are now bleeding into CI systems, threatening their functionality and usability [19]. This can have far-reaching consequences, as these systems, which typically offer great potential to do good and promote democratic values, could be coaxed to vicissitude—and wreak havoc.

To protect collectives against malevolent individuals and collectives, and to ensure that future CI systems continue to play their multifaceted role in shaping and being shaped by the needs of global societies, we believe it is imperative to draw out design policies that would allow future CI systems to be more robust against such disruptions. The critical research gaps identified in this commentary build upon the previously proposed ‘generic’ framework for CI systems, which describes crowd-oriented ICT systems through their 24 unique attributes or components [24].

## 2 GENERIC CI FRAMEWORK

The ‘generic’ CI framework (as conceptualized by Suran et al. [24]) consists of 24 essential components that enable CI in ICT systems. Designed by aggregating 12 state-of-the-art CI frameworks and models, the ‘generic’ framework describes CI systems as having four primary components (staffing, process, motivation, and goals) each containing its own sets of types, interactions, and properties. The authors state that combinations of these components promote CI in ICT systems, and thus through these components, developers and stakeholders cannot only understand existing CI systems but also develop new systems [24]. Unfortunately, like most CI frameworks, the ‘generic’ CI framework also only focuses on attributes and components that are imperative for inculcating CI in ICT systems, and not so much on the social media activities being carried out on said CI systems.

We would like to argue that given the increasingly important role social media platforms (and their activities) play in the mobilization and harnessing of CI, social media activities ought to be included in CI frameworks at some point. However, disruptions being caused by the malevolent use of social media necessitates that solutions to tackle such disruptions must also be included in future CI frameworks. To this end, we find four essential aspects, namely good deliberation, power to the crowd, role of machines, and collective dynamics—emerging solutions that should be included in CI frameworks if systems designed using said frameworks are to sustainably play their current role in our societies.

### 3 GOOD DELIBERATION

Among the many roles CI initiatives play in today’s socio-political landscape, democratic deliberation is arguably the most critical—with citizens worldwide seeking ways to actively engage in their own governance [10]. To assure that such deliberations (e.g., on CTs and CEPs) are meaningful and beneficial to all, citizens not only need the opportunity to share their ideas, opinions, and beliefs (already enabled to some extent through social media ‘activities’) but also environments that are free from the aforementioned threats to CI [19]. To achieve this, it is instrumental to compare the core aspects of CI and deliberation. As defined by Mansbridge, deliberation in public sphere can be ‘minimally and broadly’ understood as “mutual communication that involves weighing and reflecting on preferences, values, and interests regarding matters of common concern” [18, p. 27] [8, p. 5]. The aspects of this definition we find most critical are ‘mutual communication’ and ‘weighing and reflecting.’ We would like to argue that from a CI perspective, these two aspects correspond to two key aspects of CI, namely diversity and transparency—which require more attention from CI researchers.

#### 3.1 Diversity

In the context of CI, diversity refers to “diversity of opinions” [24]—that is, the collective should be composed of individuals holding diverse, even conflicting, personal beliefs and backgrounds [17]. This is critical, as individuals with diverse opinions, when provided with a safe and respecting environment [24], can perform better in some tasks [5], achieve more effective deliberations [19], and reach *modus vivendi* [17]—thus shielding the collective against polarization [10]. Unfortunately, few design policies of CI systems actively promote dialogue between diverse opinions [24]. In addition, many systems employ algorithms (e.g., recommendation systems) that (in)advertently exacerbate individuals’ behavior of surrounding themselves with others with similar tendencies—promoting homophily [20], thereby preventing collectives from experiencing diverse opinions. A better understanding of such policies, interfaces, and systems that facilitate, rather than disrupt, the exchange of diverse opinions is needed [20].

#### 3.2 Transparency

Current discussions on transparency in CI or crowd/citizen-oriented systems primarily come from the social sciences and e-governance communities, which typically focus on ‘transparency of data’ in citizen-government interactions (e.g., by focusing on open government data [2]). We find that ‘transparency in processes and identities’ are equally important in the context of crowd-oriented systems [24]. We argue that for the design of self-organizing and adaptive CI systems, transparency of systems’ processes (including goals and norms) at all levels (micro, meso, and macro) is equally important [24]. It should be noted that in the current context, while using the term *transparency*, we are adhering to three metaphors of transparency as described by Ball [4]: ‘transparency as a public value or norm of behavior to counter corruption’ (or, malevolent behavior), ‘transparency as open organizations,’ and ‘transparency as a means to evaluate effectiveness and efficiency’ of CI systems.

Keeping these metaphors in mind, when discussing ‘transparency in processes and identities,’ system stakeholders ought to provide their users comprehensive and understandable explanations of the systems’ processes [15]. This would allow users to make better-informed decisions when contributing to the system; thus, promoting trust between systems and their users—enabling the latter to conduct high-quality deliberations while making the systems less prone to abuse [15]. This is especially vital as trust plays a key role in maintaining the systems’ users’ motivation to contribute to the systems, giving the systems time to self-organize, adapt, and evolve [24].

We believe that transparency in identities of the members of the collective is equally important, as it plays a critical role in tackling cascades of misleading, biased, and inaccurate information. User reputations can, for example, help build trust and respect among system users, making transparency in identities *sine qua non* for self-organizing systems [24]. To achieve this, CI systems typically adopt one—or several—of three approaches. The

first way is to link CI users' profiles to their real-world identities (as with Twitter's blue verified badge or Google's knowledge panel). The second approach is to look at a user's reputation based on an individual's influence (through their followers, as on Twitter), or their contributions based on up/down-votes (as with Reddit's karma or Kaggle's progression system). The third approach is to adopt contribution labeling, where CI users' profiles are labeled based on textual/sentiment analysis of their contributions (e.g., Stack Overflow's top tags/communities, Quora's knows about, or R/ChangeMyView's delta).

Each of these approaches does, however, come with its own drawbacks: disclosing users' real-world identities can expose individuals to punishments or actions from influential stakeholders with conflicting interests; platforms with follower-based user reputations can (and tend to) self-organize into partisan structures, which can lead to a greater divide between ideologically opposed communities [21], while reputations based on up/down-votes can be manipulated with fake accounts and bots [19]; and finally, labeling individuals based on their opinions, although the least intrusive, can possibly also negatively impact users' wellbeing—the effects of these different approaches on users and systems, however, remains underexplored in CI research. For example, what effects these approaches have on crowd composition, dynamics, and behavior (and vice versa) and which approaches are best suited for what CI environments and objectives are largely unknown.

#### 4 POWER TO THE CROWD

Many of today's CI systems are business models focusing more on user engagement (similar to social media platforms) than user welfare (or problem solving). Like social media platforms, CI systems too often end up capturing and storing large volumes of user data, presenting users with tailored visualizations of outcomes and results, typically decided by the systems' stakeholders. This may limit the power of the crowd, because under the right conditions, collectives can come up with innovative uses of accumulated data themselves. Web sleuths on Reddit and Facebook (through their tenacious persistence and unique heuristics approaches), for example, helped solve the 20-year-old mystery of a missing teenager 'Grateful Doe' (Jason Callahan), or the more recent Reddit rally over GameStop. Facilitating data access to users should, however, be done with care, as individuals with malicious intentions can use these to create 'counter-visualizations' to mislead other users, as illustrated by coronavirus skeptics in the United States who attempted to defy public health guidelines on COVID-19 through social media [14], or as in the case of r/findbostonbombers where Redditors turned into mobs and made false allegations against unsuspecting and innocent high school students [26].

We find that although open-data access (considering its benefits as understood in science and governance) should be the norm for CI systems, it must be accompanied by policies that promote media literacy [14], transparency in identities and processes, and integration of easy-to-use data-analysis (and visualization) tools as well as moderated discussion forums. Doing so would help in establishing trust between CI systems' stakeholders and users, and enable users to further interact with each other and share their experiences when looking at the platforms' accumulated data—thus promoting emergent behavior, which is a prerequisite for CI systems to sustain themselves over time [24]. Furthermore, other system features that can help promote emergent behavior, especially in the current digital landscape, need to be identified and studied in CI research.

#### 5 ROLE OF MACHINES

A friend-foe of CI recently bothering researchers is **Machine Intelligence (MI)** ("realized through artificial intelligence and machine learning" [17, p. 76]) [19]—especially pertaining to the heated discussions on biases in predictive policing and concerns around black-box recommendation systems. Furthermore, concerns about machines becoming bosses of their human counterparts have led researchers to ponder about how to establish a balance between humans and machines. The key question here is: who leads whom, to what extent, and why? As research on the 'Global Brain' [7] (viz. Superminds [16]) suggests, it is possible for humans and machines to work in tandem, to create a higher intelligence and achieve human goals through hyperconnectivity [16, 17];

however, given that the idea of combining human intelligence with MI is still being developed, there is still a lot more to be researched—particularly frameworks surrounding task and responsibility sharing between humans and machines [16]. In addition, given the tremendous organizational and societal implications these superminds (specifically their MI counterparts) can have, it is important to look beyond the ‘explainability’ of such systems and examine them through the emerging field of “machine behavior” (described only as “the scientific study of behavior exhibited by intelligent machines”) [23]. This research is time critical, as we already see concrete organizational solutions (e.g., Blackrock’s Aladdin, Genpact’s Cora, and IBM’s Collaborative AI) that have understood the full potential of superminds. Such exclusivity to major players may create a new digital divide, and therefore it is critical to understand and unlock this potential for everybody [1].

## 6 COLLECTIVE DYNAMICS

And then, there are the fundamental questions, “How do collectives self-organize, adapt, and evolve?” [24], “Why do some online communities fail, while others succeed?” [17] and “Why does the crowd sometimes turn to self-destructive behaviors?” [19], which are still largely unanswered. Although researchers have attempted to answer some of these questions, most of the work has only focused at either the micro-, meso-, or macro level (e.g., [9]); research linking these levels, studying their interactions, is scarce. In this context, we believe that to understand how individuals make decisions in CI systems (and in doing so shape these systems) requires an explicit understanding (and modeling) of the users’ cognitive processes [25], and future studies should aim to better understand an individual’s cognition, including how and when individuals search, take up, and share information on CI systems—as well as how these processes shape an individual’s CI informational landscape, and how this, in turn, feeds back on an individual’s cognition.

## 7 MOVING FORWARD

Finally, considering the profound ways social media ‘activities’ and recommendation algorithms affect the collective’s cognition, we find that understanding collective dynamics also requires both exploratory and explanatory analysis of the intertwined effects the social media ‘activities,’ individuals’ cognition, and ‘machines’ have on each other. And that more ‘crisis-focused’ studies on CI behavior (similar to [9, 13, 25]) need to be conducted, to bring together insights from the emerging fields of “collective behavior” [3] and “machine behavior” [23]. Ultimately, it is the amalgamation of adoption of policies and guidelines from research, and investigation of emerging crisis disciplines involving humans and machines, that could protect both collectives and their deliberation instruments against ongoing digital disruptions.

## ACKNOWLEDGMENTS

We are grateful to A. Berditchevskaia and R. Srivastava for valuable feedback.

## REFERENCES

- [1] Andrés Abeliuk, Daniel M. Benjamin, Fred Morstatter, and Aram Galstyan. 2020. Quantifying machine influence over human forecasters. *Scientific Reports* 10, 1 (Dec. 2020), 15940. <https://doi.org/10.1038/s41598-020-72690-4>
- [2] Judie Attard, Fabrizio Orlandi, Simon Scerri, and Sören Auer. 2015. A systematic review of open government data initiatives. *Government Information Quarterly* 32, 4 (2015), 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>
- [3] Joseph B. Bak-Coleman, Mark Alfano, Wolfram Barfuss, Carl T. Bergstrom, Miguel A. Centeno, Iain D. Couzin, Jonathan F. Donges, et al. 2021. Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences* 118, 27 (July 2021), e2025764118. <https://doi.org/10.1073/pnas.2025764118>
- [4] Carolyn Ball. 2009. What is transparency? *Public Integrity* 11, 4 (2009), 293–308. <https://doi.org/10.2753/PIN1099-9922110400>
- [5] Dan Bang and Chris D. Frith. 2017. Making better decisions in groups. *Royal Society Open Science* 4, 8 (Aug. 2017), 170193. <https://doi.org/10.1098/rsos.170193>
- [6] Michael L. Barnett, Dhruv Boddupalli, Shantanu Nundy, and David W. Bates. 2019. Comparative accuracy of diagnosis by collective intelligence of multiple physicians vs individual physicians. *JAMA Network Open* 2, 3 (March 2019), e190096. <https://doi.org/10.1001/jamanetworkopen.2019.0096>

- [7] Abraham Bernstein, Mark Klein, and Thomas W. Malone. 2012. Programming the global brain. *Communications of the ACM* 55, 5 (May 2012), 41–43. <https://doi.org/10.1145/2160718.2160731>
- [8] Andre Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark Warren. 2018. *The Oxford Handbook of Deliberative Democracy*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198747369.001.0001>
- [9] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proceedings of the ACM on Human-Computer Interactions* 2, CSCW (Nov. 2018), Article 32, 25 pages. <https://doi.org/10.1145/3274301>
- [10] John S. Dryzek, André Bächtiger, Simone Chambers, Joshua Cohen, James N. Druckman, Andrea Felicetti, James S. Fishkin, et al. 2019. The crisis of democracy and the science of deliberation. *Science* 363, 6432 (March 2019), 1144–1146. <https://doi.org/10.1126/science.aaw2694>
- [11] Yu T. Hsiao, Shu-Yang Lin, Audrey Tang, Darshana Narayanan, and Claudina Sarahe. 2018. vTaiwan: An Empirical Study of Open Consultation Process in Taiwan. Available at <https://doi.org/10.31235/osf.io/xyhft>.
- [12] Joshua Introne, Robert Laubacher, Gary Olson, and Thomas Malone. 2011. The climate CoLab: Large scale model-based collaborative planning. In *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems (CTS’11)*. 40–47. <https://doi.org/10.1109/CTS.2011.5928663>
- [13] Bertrand Jayles, Clément Sire, and Ralf H. J. M. Kurvers. 2021. Impact of sharing full versus averaged social information on social influence and estimation accuracy. *Journal of the Royal Society Interface* 18, 180 (July 2021), 20210231. <https://doi.org/10.1098/rsif.2021.0231>
- [14] Crystal Lee, Tanya Yang, Gabrielle D. Inchoco, Graham M. Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI’21)*. ACM, New York, NY, 1–18. <https://doi.org/10.1145/3411764.3445211>
- [15] Philipp Lorenz-Spreen, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. 2020. How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour* 4, 11 (Nov. 2020), 1102–1109. <https://doi.org/10.1038/s41562-020-0889-7>
- [16] Thomas W. Malone. 2018. *Superminds*. Oneworld Publications, London, UK.
- [17] Thomas W. Malone and Michael S. Bernstein (Eds.). 2015. *Handbook of Collective Intelligence*. MIT Press, Cambridge, MA.
- [18] Jane Mansbridge. 2015. A minimalist definition of deliberation. In *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies*. World Bank, 27–50. [https://doi.org/10.1596/978-1-4648-0501-1\\_ch2](https://doi.org/10.1596/978-1-4648-0501-1_ch2)
- [19] Geoff Mulgan. 2018. *Big Mind*. Princeton University Press, Princeton, NJ.
- [20] Niccolò Pescetelli, Alex Rutherford, and Iyad Rahwan. 2021. Modularity and composite diversity affect the collective gathering of information online. *Nature Communications* 12, 1 (Dec. 2021), 3195. <https://doi.org/10.1038/s41467-021-23424-1>
- [21] Pew Research Center. 2020. *Many Experts Say Digital Disruption Will Hurt Democracy*. Technical Report. Pew Research Center: Internet, Science & Tech. <https://www.elon.edu/u/imagining/surveys/future-of-democracy-2020/>.
- [22] Sarbeswar Praharaj, Jung Hoon Han, and Scott Hawken. 2017. Innovative civic engagement and digital urban infrastructure: Lessons from 100 smart cities mission in India. *Procedia Engineering* 180 (2017), 1423–1432. <https://doi.org/10.1016/j.proeng.2017.04.305>
- [23] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W. Crandall, et al. 2019. Machine behaviour. *Nature* 568, 7753 (April 2019), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- [24] Shweta Suran, Vishwajeet Pattanaik, and Dirk Draheim. 2020. Frameworks for collective intelligence. *ACM Computing Surveys* 53, 1 (May 2020), 1–36. <https://doi.org/10.1145/3368986>
- [25] Alan N. Tump, Timothy J. Pleskac, and Ralf H. J. M. Kurvers. 2020. Wise or mad crowds? The cognitive mechanisms underlying information cascades. *Science Advances* 6, 29 (July 2020), eabb0266. <https://doi.org/10.1126/sciadv.abb0266>
- [26] José van Dijck and Thomas Poell. 2013. Understanding social media logic. *Media and Communication* 1, 1 (2013), 2–14. <https://doi.org/10.17645/mac.v1i1.70>

Received 17 February 2022; revised 9 August 2022; accepted 2 October 2022