



Where and when AI and CI meet: exploring the intersection of artificial and collective intelligence towards the goal of innovating how we govern

Stefaan G. Verhulst¹

© Springer-Verlag London Ltd., part of Springer Nature 2018

1 Where and when AI and CI meet

This paper seeks to explore the intersection of Artificial Intelligence (AI) and Collective Intelligence (CI), within the context of innovating how we govern. It starts from the premise that advances in technology provide policy makers with two important new assets: data and connected people. The application of AI and CI allows them to leverage these assets toward solving public problems. Yet both AI and CI have serious challenges that may limit their value within a governance context, including biases embedded in datasets and algorithms, undermining trust in AI; and high transaction costs to manage people's engagement limiting CI to scale.

The main argument of this paper is that some of the challenges of AI and CI can in fact be addressed through greater interaction of CI and AI. In particular, the paper argues for:

- Augmented Collective Intelligence where AI may enable CI to scale;
- Human-Driven Artificial Intelligence where CI may humanize AI.

Several real-world examples are provided throughout the paper to illustrate emerging trends toward both types of intelligence; and their applications to solve public problems or make policy decision differently.

The author would like to acknowledge the research support from Andrew Young and Mastoureh Sadeghnia, both at the GovLab.

✉ Stefaan G. Verhulst
sv39@nyu.edu

¹ The GovLab, NYU Tandon School of Engineering, New York University, 6 MetroTech Center, Brooklyn, NY 11201, USA

2 Data and connected people

As the technology, research and policy communities continue to seek new ways to improve governance and solve public problems, two new important assets are occupying increasing importance: data and connected people. Leveraging data and people's expertise in new ways offers a path forward for smarter decisions, more innovative policy-making, and more accountability in governance (Verhulst 2017). Yet, unlocking the value of these two assets not only requires increased availability and accessibility of those assets (through, for instance, open data or open innovation), it also requires innovation in methodology and technology.

The first of these innovations involves **Artificial Intelligence** (AI). AI offers unprecedented abilities to quickly process vast quantities of data that can provide data-driven insights to address public needs (Ng 2015). This is the role it has for example played in New York City, where Fire-Cast, leverages data from across the city government to help the Fire Department identify buildings with the highest fire risks (Rielan 2015). AI is also considered to improve education, through the creation of virtual tutors and improved learner self-direction and assessment (Kurshan 2016); urban transportation, through predictive analytics on stresses to transport infrastructure like train equipment (Basu 2016); humanitarian aid, through improved understanding of refugees' demographics and the resultant targeting of resources (Smith 2017); and combat corruption¹, by modeling optimal, legitimate government service delivery strategies, and eventually automating the delivery of some government services. Artificial intelligence is also being used to deliver personalized health treatments, provide psychological support to Syrian refugees through the use of chatbots, improve the accessibility of internet content for people with visual impairments, more accurately predicting crop yields through

¹ <http://theorylamp.com/2016/08/29/can-artificial-intelligence-stop-corruption-in-pakistan/>.

the automated analysis of satellite imagery, and more (Castro and New 2016).

The second area is Collective Intelligence (CI). Although it receives less attention than AI, CI offers similar potential breakthroughs in changing how we govern, primarily by creating a means for tapping into the “wisdom of the crowd” and allowing groups to create better solutions than even the smartest experts working in isolation could ever hope to achieve. For example, in several countries patients’ groups (Nicholas and Broadbent 2015) are coming together to create new knowledge (Addario 2017) and health treatments (Weiner 2014) based on their experiences and accumulated expertise. Similarly, scientists are engaging citizens in new ways to tap into their expertise or skills, generating citizen science (Wynn 2017)—ranging from mapping our solar system² by inviting whoever interested to help NASA make maps of scientifically interesting features in our Solar System, to manipulating enzyme models³ in a game-like fashion, where people can get involved in an online puzzle video game, and eventually the highest scoring solutions are analyzed by researchers.

Neither AI nor CI offer panaceas for all our ills; they each pose certain challenges, and even risks. The effectiveness and accuracy of AI relies substantially on the quality of the underlying data⁴ as well as the human-designed algorithms used to analyze that data (Verhulst 2017). Given AI’s reliance on “training data” to inform automated decision-making, the collection, processing, sharing, analysis, or use of low quality data can derail the effectiveness of AI implementations, and any inaccuracies or challenges arising from low quality data are likely to compound over the course of this data life cycle. Among other challenges, it is becoming increasingly clear how biases against minorities and other vulnerable populations can be built into these algorithms. For instance, some AI-driven platforms for predicting criminal recidivism significantly over-estimate the likelihood that black defendants will commit additional crimes in comparison to white counterparts (Mattu et al. 2016). The need for incused algorithmic scrutiny is increasingly recognized, with a growing literature⁵ (Srinivasan et al. 2017) on the topic examining issues related to algorithms used by information intermediaries, governance challenges, tools for the transparency and accountability of algorithms, and studies of particularly opaque and harmful uses of algorithms in the consumer finance and criminal justice realms. While the literature is indeed growing, the field has not yet found clear and widely implementable solutions to the challenges posed

by our increasing reliance on algorithmic decision-making, and those challenges are only likely to grow as we move toward more and broader use of AI.

In theory, CI avoids some of the risks of bias and exclusion present in a number of the implementations of AI to date, because it is specifically designed to bring more (diverse) voices into a conversation. But ensuring that the multiplicity of voices adds value, not just noise, can be an operational and ethical challenge (Standing and Standing 2017). Questions still remain both regarding how to effectively surface the most useful and relevant and expertise during crowdsourcing or collective intelligence efforts, and how to ensure, for example, that the often free labor provided by participating individuals is not handled in a exploitative manner. As it stands, effectively and ethically identifying the signal in the noise in CI initiatives can be time-consuming and resource intensive, especially for smaller organizations or groups lacking resources or technical skills.

Despite these challenges, however, there exists a significant degree of optimism – evidenced by, for example, Nesta CEO Geoff Mulgan’s recent book *Big Mind: How Collective Intelligence Can Change Our World* (Mulgan 2017), and MIT professor Max Tegmark’s *Life 3.0: Being Human in the Age of Artificial Intelligence* (Tegmar 2017)—surrounding both these new approaches to problem-solving. Some of this enthusiasm is likely hype, with AI and CI representing “shiny objects” that are believed capable of solving many if not all of the world’s problems. Some of this enthusiasm, however, is merited—CI and AI do offer very real potential for rapidly bringing more evidence and perspectives to bear for decision-making and problem-solving processes, and the task facing both policymakers, practitioners and researchers is to find ways of harnessing that potential in a way that maximizes benefits while limiting possible harms.

In what follows, I argue that one potential avenue for addressing the challenges facing the impactful and ethical use of AI and CI described above may involve a greater interaction between AI and CI.

These two areas of innovation have largely evolved and been researched separately until now.⁶ However, I believe that there is substantial scope for integration, and mutual reinforcement. It is when harnessed together, as complementary methods and approaches, that AI and CI can bring the full weight of technological progress and modern data analytics to bear on our most complex, pressing problems.

To deconstruct the above statement, I propose three premises toward establishing a necessary research agenda on the

² <https://cosmoquest.org/x/>.

³ <https://fold.it/>.

⁴ <http://thegovlab.org/why-we-should-care-about-bad-data/>.

⁵ <http://thegovlab.org/algorithmic-scrutiny/>.

⁶ There is not much research evidence on the interaction between AI and CI. However some research has been done on the connection between AI and CI, such as here: <https://homes.cs.washington.edu/~weld/papers/ci-chapter2014.pdf>.

intersection of AI and CI that can build more inclusive and effective approaches to governance innovation. As opposed to more traditional “government”, governance innovation refers to the idea that increased availability and use of data, new ways to leverage the capacity, intelligence, and expertise of people in the problem-solving process, combined with new advances in technology and science can transform governance.

3 Premise I: toward augmented Collective Intelligence: AI will enable CI to scale

While CI is built around the idea that groups of citizens or experts can be smarter and more effective than individuals, scaling CI initiatives can be difficult. This is largely because of the transaction costs involved in CI. Unlike the more automated processes of AI, which needs minimum or no human intervention, CI typically involves a substantial degree of human effort in curating, [e.g., human effort (volunteers) involved in creating and editing all the entries in Wikipedia], in inviting and enabling participation by particular groups of individuals or institutions. In addition, significant effort can be required to triage signal from noise (resulting from input quantity and quality). All of this means that CI can be fairly labor intensive, and hard to automate in the same way that AI is automated.

Could the capabilities represented by AI—including making sense of complex systems using intelligent analysis—help to optimize, eliminating unnecessary human effort, CI initiatives and overcome some of these scaling challenges? If implemented effectively, automation through AI could indeed save time and effort, leading to what we call “Augmented Collective Intelligence”. Consider, for instance, platform Notice and Comment (N&C)⁷ that not only enables comments to policy proposals but, more importantly, seeks to generate insights from the big quantity of comments received to regulatory data leveraging an AI tool called Regendus.⁸ Similarly, the paradigmatic, yet ailing (Halfaker et al. 2012) example of collective intelligence, Wikipedia, has started to use AI bots (Merrill 2015) to help edit articles, identify and clean up vandalism, categorize and tag.

Besides automation, AI can also help by identifying communities, via, for example, analyzing data from community assets mapping,⁹ expert mapping, and GIS, with something relevant and valuable to offer CI initiatives. In addition, techniques like sentiment analysis (also known as emotion AI

or opinion mining) (Medhat et al. 2014), which is a computational study of people’s emotions and opinions towards an entity, can also lower the burden of those charged with acting upon the CI-generated insights; such techniques can at least partially automate and generally improve the processes of gauging, analyzing and even acting upon the inputs received from participants in CI initiatives.

4 Premise II: toward Human-Driven Artificial Intelligence: CI will humanize AI

Much of the concern surrounding the expansion and evolution of AI revolves around its perceived “inhumanity” (resulting from automation or lack of human judgement when making sense of datasets). Dystopian scenarios examine the possibility of a “robot takeover,” and, less fantastically, an increasingly wide range of areas with significant impact [for example, military decision-making (Allen and Taniel 2017), banking and even driving (Markoff 2017)] are undergoing a somewhat discomfiting human-to-machine transition mainly as a result of the belief that machines lack the crucial human intuition, which leads to unpredictable consequences. AI is a black box. Its very advantage (the ability to make decisions beyond the reach of humans) is also cause for concern of inhumane decisions. Despite calls for greater algorithmic transparency, the fact remains that even the creators of AI algorithms do not understand the actions or results thrown up by their creations. For instance, the recent “misbehaving” conversation bots,¹⁰ made facebook researchers shut down the relevant AI system. They had to pull the plug on the AI systems “because things got out of hand”, according to the researchers. The AI agents that were programed to use English to converse with each other, later stopped using English and created a new language that only AI systems could understand, which defied their purpose. This led Facebook researchers to interfere and force the AI system to speak English.

CI has a potentially valuable role to play here, too. For example, introducing a human element into AI through coordinated CI efforts could help surface biases embedded in datasets and demystify the analytics performed on those datasets—one of the rationales behind a commercial service called CrowdFlower.¹¹ CI could also increase the legitimacy of AI initiatives through a collaborative design process (which can itself be vetted by CI) to ensure that AI interventions that raise ethical or other concerns are developed

⁷ <https://www.noticeandcomment.com/about-us.aspx>.

⁸ <http://www.regendus.com/>.

⁹ http://healthpolicy.ucla.edu/programs/health-data/trainings/Documents/tw_cba20.pdf.

¹⁰ <http://www.firstpost.com/tech/news-analysis/facebook-researcher-shut-down-ai-bots-that-started-speaking-in-a-language-unintelligible-to-humans-3876197.html>.

¹¹ <https://www.crowdfunder.com/>.

carefully, or not at all. MoralMachine,¹² for instance, gathers “human perspective on moral decisions made by machine intelligence”. More generally, introducing a human element to AI could increase the legitimacy of AI in the public eye, and mitigate some of the emerging concerns and opposition surrounding the field.

These are just some of the ways in which CI can help mitigate the risks posed by AI. Interestingly, emerging research suggests that CI may have a role to play not only in increasing the ethical legitimacy of AI, but even AI’s effectiveness. Researchers at MIT (Hardesty 2017), for instance, have experimented with using crowdsourced expertise to identify the main “features” of big datasets; in this way, CI is the first point of entry into data, which is then considered more deeply using traditional, automated AI. Similarly, other researchers,¹³ including Professor Sandy Pentland, have been applying social techniques of learning used by humans to create more intelligent neurons in AI algorithms; the effort is to make individual neurons learn from each other in much the same way that humans use social and cultural contexts to make decisions. Both of these examples suggest that, in addition to increasing legitimacy and trust, CI may in fact have a role to play in enhancing the capabilities of AI—much in the same way AI can help scale up CI by saving time and effort, for example, (see previous section), leading to Augmented Collective Intelligence.

5 Premise III: open governance will drive a blurring between AI and CI

For different reasons, open governance—understood as processes for more equitable and participatory pathways in decision-making and problem-solving—can appear to be at odds with both AI and CI. AI initiatives can be biased, closed and opaque. Similarly, CI efforts can be driven by narrowly defined communities, with the result that traditionally marginalized or disenfranchised groups (which may in fact possess relevant expertise) can be excluded. Both these scenarios run counter to the *openness* principle, i.e., transparency, at the core of open governance.

Rather than acting in opposition, the methods and values of open governance can in fact be embedded into both AI and CI, in the process helping both these innovations move closer together. For example, efforts to introduce greater openness into AI, by minimizing the effects of existing biases embedded in AI, are likely to lead to more integration with CI which, as described above, can help increase

the inclusiveness and transparency of AI initiatives, by introducing a human element to help surface biases embedded in datasets and increasing the legitimacy of AI initiatives in the public eye. Likewise, as CI seeks to move beyond limited communities of participation, AI can play an essential role in identifying and curating actors and stakeholders that may widen the collective conversation while at the same time ensuring the relevance of their inputs and a high signal-to-noise ratio. In these ways, open governance and its underlying principles has a valuable role to play in the efforts to both strengthen AI and CI, as well as in bringing these two strands of innovation closer together.

6 Establishing a research agenda

This paper sought to highlight the potential and challenges of artificial and collective intelligence for governance innovation and argues that a greater interaction of AI and CI may help addressing some of the existing challenges. Although there are some early initiatives that seek to experiment with a closer interaction of AI and CI, there is hardly any mapping of existing practice and little evidence exists of their impact and the conditions that contributed to this.

Thus we end with a call for more interdisciplinary research on the interaction between AI and CI which could determine how we solve problems in the future. In particular, the following questions remain unanswered:

Questions as it relates to Augmented Collective Intelligence:

- How can AI scale CI? What attributes of AI can make a difference?
- How can we ensure that the introduction of AI into CI initiatives does not introduce new biases or increase the risk of making decisions based on bad data?
- How can we ensure that wide, diverse audiences are able to participate in AI-led Collective Intelligence initiatives?
- What use cases could act as testbeds for further experimentation into Augmented Collective Intelligence?

Questions related to Human-Driven Artificial Intelligence:

- Can CI legitimize AI processes by bringing in a more human element, and, if so, to what extent does that legitimizing function improve upon other pathways, e.g., expert review panels for AI algorithms?
- What strategies could allow for a collective governance process to minimize the power asymmetries created by AI?

¹² <http://moralmachine.mit.edu/>.

¹³ https://www.edge.org/conversation/alex_sandy_pentland-the-human-strategy.

- Can CI offer greater ability to understand and explain the decision-making and outcomes of AI processes?
- Can CI play a role in integrating values and ethics into the design and functioning of AI processes?

Questions on how the intersection of AI and CI can contribute to Open Governance:

- How can the values embodied in open governance be integrated into the design of AI-meets-CI initiatives?
- How can AI and CI be used in concert to create more active citizenship?
- Do we need new institutions to help push forward new governance approaches that may emerge from joint AI-CI initiatives?
- Is there a need for an AI meets CI Magna Carta articulating principles around risk management, redress systems, and accountability; duties of institutions across the AI/CI value chain; and prohibitions related to the expanded use of AI and/or CI?
- How compatible (if at all) are the metrics of success for, respectively, AI, CI and open government?

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

References

- Addario B (2017) Patient Power: Crowdsourcing in Cancer. The Huffington Post, New York
- Allen G, Taniel C (2017) Artificial Intelligence and National Security. Belfer Center for Science and International Affairs. Harvard Kennedy School, Cambridge
- Basu M (2016) Exclusive: Singapore's Plans for Predictive Transport. GovInsider, Singapore
- Castro D, New J (2016) The Promise of Artificial Intelligence: 70 Real-World Examples. Information Technology and Innovation Foundation
- Halfaker A, Geiger S, Morgan J, Riedl J (2012) The rise and decline of an open collaboration system. *Am Behav Sci* 57(5):664–688
- Hardesty L (2017) Crowdsourcing big-data analysis. MIT News Office, Cambridge
- Kurshan B (2016) The future of artificial intelligence in education. *Forbes Magazine*, New York
- Lin C, Weld D, Bragg J, Mausam (2015) Artificial intelligence and collective intelligence. In: Malone T, Bernstein M (eds) *The Collective Intelligence Handbook*. MIT Press, New York
- Markoff J (2017) A Guide to Challenges Facing Self-Driving Car Technologists. *The New York Times*, New York
- Mattu S, Larson J, Angwin J, Kirchner L (2016) How We Analyzed the COMPAS Recidivism Algorithm. ProPublica
- Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 5:1093–1113
- Merrill B (2015) The Bots Who Edit Wikipedia (And The Humans Who Made Them). *MakeUseOf*
- Mulgan G (2017) *Big Mind: how collective intelligence can change our world*. Princeton University Press, Princeton
- Ng A (2015) What AI Can and Can't Do. *Harvard Business Review*, Boston
- Nicholas L, Broadbent S (2015) *Collective intelligence in patient organisations*. Nesta
- Rieland R (2015) How data and a good algorithm can help predict where fires will start. *Smithsonian.com*
- Smith C (2017) Humanitarian efforts could be aided by AI. *Imperial College*, London
- Srinivasan P, Young A, Verhulst S (2017) The GovLab selected readings on algorithmic scrutiny. Available at <http://thegovlab.org/algorithmic-scrutiny/>
- Standing S, Standing C (2017) The ethical use of crowdsourcing. *Business Ethics: A European Review*
- Tegmar M (2017) *Life 3.0: being human in the age of artificial intelligence*. Random House, New York
- Verhulst S (2017) From #Resistance to #Reimagining governance: 6 shifts that can improve the way we solve public problems. *OpenDemocracy.net*
- Weiner M (2014) The potential of crowdsourcing to improve patient-centered care. *Springer Int Publ* 7(2):123–127
- Wynn J (2017) *Citizen science in the digital age: rhetoric, science, and public engagement*. The University of Alabama Press, Tuscaloosa