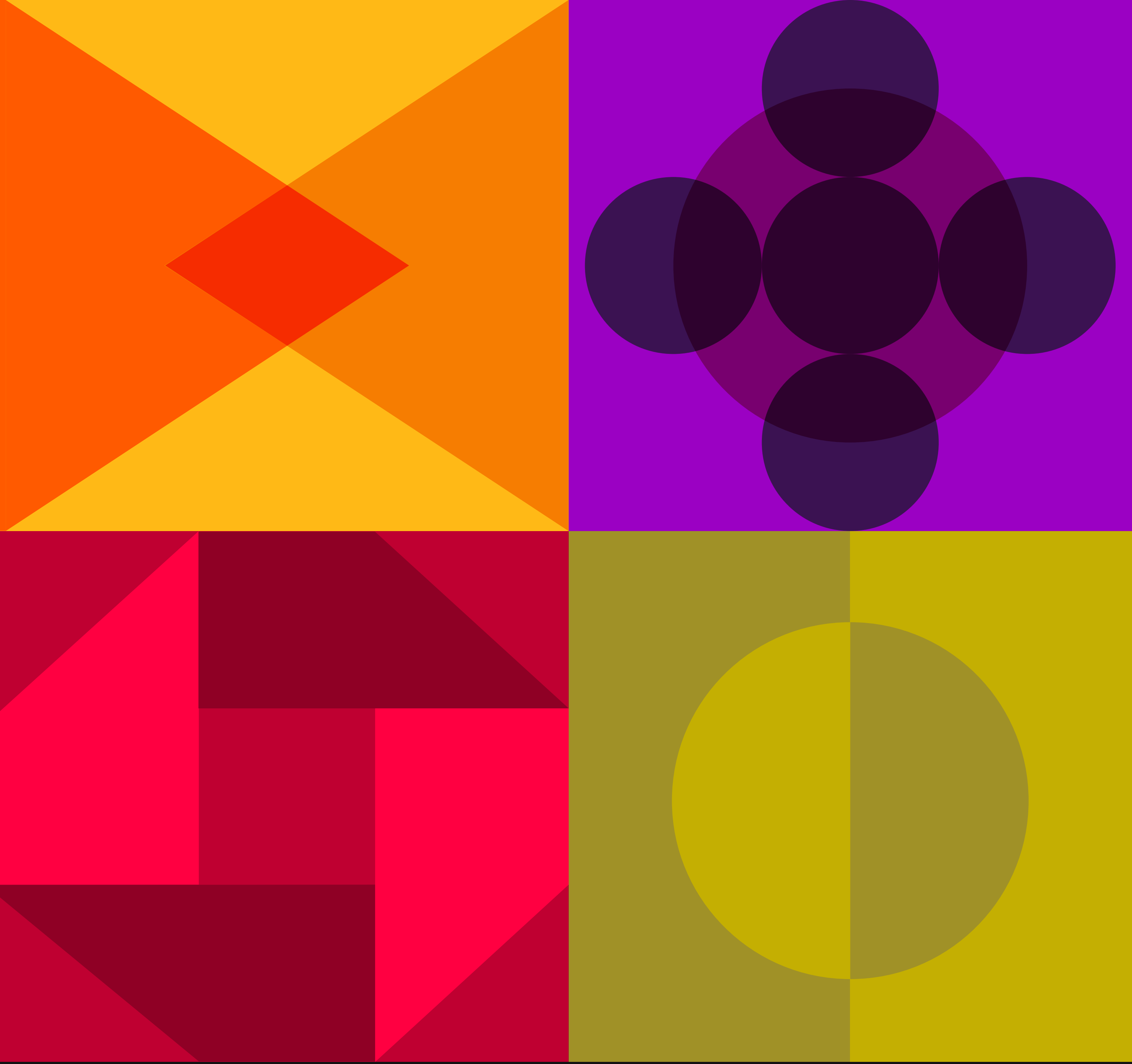


nesta

The future of minds and machines

How artifical intelligence can enhance collective intelligence

AUTHORS
Aleks Berditchevskaia
Peter Baeck

ACKNOWLEDGEMENTS

We are very grateful for the valuable insight and feedback on this research from our colleagues at Nesta, Kathy Peach, Thea Snow, Zosia Poulter, Katja Bejo, Jen Rae, Jack Orlik, Harry Armstrong, Bea Karol Burks, Geoff Mulgan, Kostas Stathoulopoulos, Markus Droemann and Eva Grobbink, as well as the wider Explorations and Research, Analysis and Policy teams.

We would like to thank the LAMA Development and Cooperation Agency for the background research and horizon scan they undertook, which underpin many of the lessons in this report. At LAMA, we would especially like to thank Elena Como, Eleonora Corsini, Stefania Galli, Dario Marmo, Walter Nunziati.

This report is based on the insights and experiences of a whole range of projects and researchers working at the forefront of exploring the AI and CI field through research and practice. We would like to thank the following people and organisations for taking part in the research through interviews and sharing feedback on early stages of the research.

Mollie Zapata, Daniel Perez Rada, Franco Pesce, Karel Verhaeghe, Erik Johnston, Andrew Bogdanovic, Mauro Lombardi, Nathan Matias, Jeff Deutsch, Juliene Corbiese, Walter Lasecki, Hannah Wallach, Anna Noel-Storr, Stefana Broadbent, Dirk Helbing, Carlo Torniai, Mark Klein, Stefan Herzog, Nalia Murray, David Cabo, Vito Trianni, Harry Wilson, Sara Caldas, Louis Rosenberg, Colin Megill, Francis Heylingen, Elian Carsenat, Ashvin Sologar and Zooniverse.

The visual assets for this report were created by Margherita Cardoso of Soapbox. Figures 2 and 3 were created by Lily Scowen for the Collective Intelligence Design Playbook.

The original version of this report was developed with interactive visual elements and is best experienced online at **www.nesta.org.uk/report/future-minds-and-machines**.

A supporting feature, highlighting 20 case studies at the intersection of AI and CI, can be found at **www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies**. A limited number of case studies have been reproduced in the print version of the report.

ABOUT NESTA

Nesta is a global innovation foundation. We back new ideas to tackle the big challenges of our time.

We use our knowledge, networks, funding and skills, working in partnership with others, including governments, businesses and charities. We are a UK charity but work all over the world, supported by a financial endowment.

To find out more visit **www.nesta.org.uk**

If you'd like this publication in an alternative format, such as Braille, large-print or audio, please contact us at: **information@nesta.org.uk**

Published March 2020

The future of minds and machines

How artifical intelligence can enhance collective intelligence

01

Introduction

When it comes to artificial intelligence (AI), the dominant media narratives often end up taking one of two opposing stances: AI is the saviour or the villain. Whether it is presented as the technology responsible for killer robots and mass job displacement or the one curing all disease and halting the climate crisis, it seems clear that AI will be a defining feature of our future society.

However, these visions leave little room for nuance and informed public debate. They also help propel the typical trajectory followed by emerging technologies; with inevitable regularity we observe the ascent of new technologies to the peak of inflated expectations they will not be able to fulfil, before dooming them to a period languishing in the trough of disillusionment.[1]

There is an alternative vision for the future of AI development. By starting with people first, we can introduce new technologies into our lives in a more deliberate and less disruptive way. Clearly defining the problems we want to address and focusing on solutions that result in the most collective benefit can lead us towards a better relationship between machine and human intelligence. By considering AI in the context of large-scale participatory projects across areas such as citizen science, crowdsourcing and participatory digital democracy, we can both amplify what it is possible to achieve through collective effort and shape the future trajectory of machine intelligence. We call this **21st-century collective intelligence (CI)**.

In *The Future of Minds and Machines* we introduce an emerging framework for thinking about how groups of people interface with AI and map out the different ways that AI can add value to collective human intelligence and vice versa. The framework has, in large part, been developed through analysis of inspiring projects and organisations that are testing out opportunities for combining AI and CI in areas ranging from farming to monitoring human rights violations. Bringing together these two fields is not easy. The design tensions identified through our research highlight the challenges of navigating this opportunity and selecting the criteria that public sector decision-makers should consider in order to make the most of solving problems with both minds and machines.[2]

What is in this report

Sections 1 to 3 provide an overview of AI, CI and how they can be brought together to solve problems. Sections 4 to 6 describe the challenges faced by CI and how AI methods can help. Sections 7 and 8 demonstrate how AI is already enhancing CI and helping it to scale, as well as how CI could help build better AI. Building on this, sections 9 and 10 highlight the design questions that need to be considered by anyone wanting to make use of these new innovation methods. The report concludes with a number of recommendations on how to support this new field for policy makers, and those involved in funding, researching and developing AI & CI solutions.[3]

Intended audience

This report is aimed at innovators working in public sector and civil society organisations who have some experience with participatory methods and want to understand the opportunities for combining machine and collective human intelligence to address social challenges. We hope that it can serve as inspiration for funders who care about determining a trajectory for AI that can bring the broadest possible societal benefit.

This report will also be relevant for technology and research communities with an interest in new opportunities for solving real-world problems, in dialogue with decision-makers and members of the public. Ultimately, we aim to stimulate more communication and collaboration between all of these groups.

A short history of humans and machines

Ada Lovelace and Charles Babbage first imagined an Analytical Engine that could translate codes to perform tasks useful to humans in the middle of the 19th century. Ever since, people have sought to understand the relationship between humans and computational machines. As these analytical engines have evolved into more sophisticated tools over the last 100 years, our interest in this relationship has grown exponentially, making the leap from the fringes of science fiction to the top of our daily newsfeeds.

Many academic disciplines are devoted to this question, ranging from a focus on how individuals interact with machines (Human Computer Interaction) to looking at the impact of computers on societal systems (Cybernetics and Cyber-Physical Systems). Different forms of Crowd-Machine Interaction[4] explore how groups of people work together with computers to achieve shared goals or maximise collaborative efforts. The literature ranges from the more practice-based approaches of Crowdsourcing and Citizen Science to the more academic Human Computation and Social Computing. These terms are all used to describe the aggregation of diverse inputs from a crowd towards a specific goal, respectively: to open up innovation and ideation, contribute to scientific discoveries, solve computational problems or stimulate online social behaviours.

The history of contributions across the field of Computer Science has nevertheless lacked a comprehensive overview of the potential of AI to scale and enhance collective human efforts, particularly with reference to real-world case studies. Notable exceptions include the survey of field in 2015 by Daniel Weld,[5] which mostly focused on the uses of machine intelligence on crowdsourcing platforms, and the more recent report by New York University's Governance Lab, *Identifying Citizens' Needs by Combining Artificial Intelligence and Collective Intelligence*.[6]

02

What is artificial intelligence?

The term 'artificial intelligence' has undergone numerous evolutions in meaning since it was first coined by John McCarthy in the 1950s.

More recently, the following definition by Russell and Norvig[7] has gained traction in the research community, for its ability to capture the variety of different AI methods and problem domains:

> *Intelligent agents that receive percepts from the environment and take actions that affect that environment.*

This definition includes everything from the back-end algorithms that power Google's search engines and Netflix's recommender systems to AI-powered hardware systems like robots and autonomous vehicles. The actions or tasks that these AI agents perform are typically considered to require human (or 'natural') intelligence. Common everyday uses of AI include perception of audio and visual cues by personal assistants like Alexa, automated translation between languages by Google Translate and routing apps that optimise navigation in cities, such as Citymapper and Waze.

Over the last 100 years, AI research has been defined by several competing schools of thought, which differ in their assumptions of what ingredients are needed to create an intelligent machine. The two dominant paradigms are Symbolic and Statistical (see Figure 1). Symbolic methods were very popular in the early days of AI and assumed that AI could be programmed by predefining a set of rules for a computational system to follow.

However, since the 90s and 00s, when hardware advances started to amplify machine capabilities by increasing data storage and the speed with which algorithms could carry out computations, statistical methods have come to dominate the field. These methods extract relevant rules and knowledge based on many examples from a specific problem domain. A final class of methods – known as Embodied Intelligence – sits outside both of these paradigms. These methods assume that higher intelligence requires a body or the ability to act in the external world, for example Robotics. While the figure opposite is not exhaustive, it illustrates the main methods currently being used at the intersection of AI and CI.

Figure 1

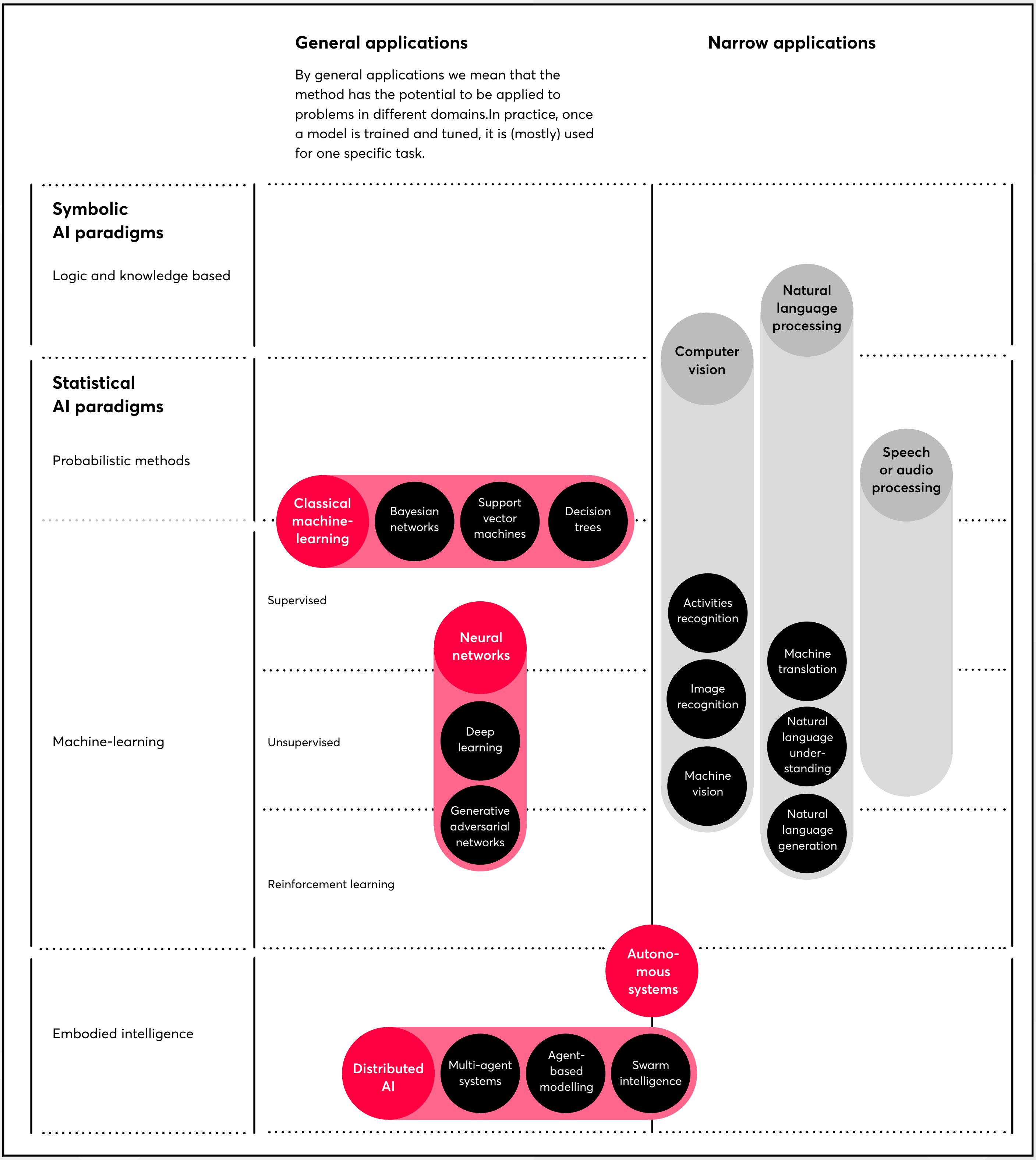


Figure 1 has been adapted from the AI Knowledge Map by Francesco Corea.

Machines that learn from data

Most applications that use AI outside of research labs today are based on machine-learning algorithms, which refers to the broad category of algorithmic methods that improve their performance of a task based on experience and data. Machine-learning algorithms are statistical, which means they typically rely on extracting patterns from very large training datasets to achieve the necessary quality of output before they can be deployed in real-world contexts.

Machine-learning techniques optimise their performance based on different learning paradigms. A lot of CI projects that use AI use supervised learning. This is where an algorithm learns to make predictions by looking at many examples of data that has already been labelled by people. For example, some citizen science projects ask participants to assign labels to images of animals, galaxies or to transcribe scanned documents. This information can be used as a training dataset for machines to learn how to classify similar images that are unlabelled.

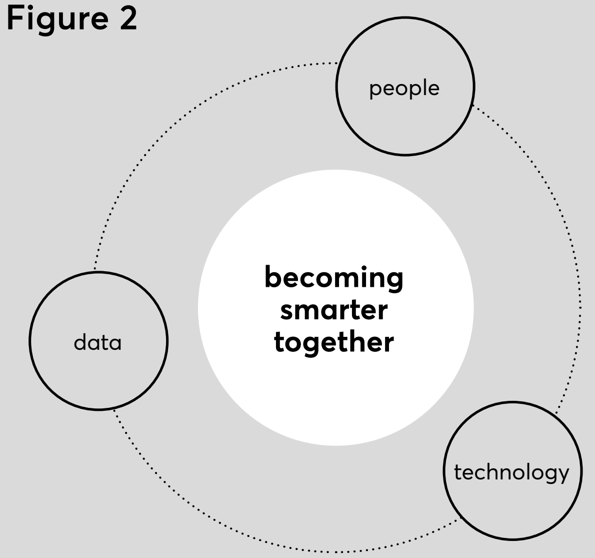
Another common learning paradigm is unsupervised learning. It is used for large complex datasets that do not have labels. Unsupervised learning helps to identify common features of the data in order to make it easier to understand. For example, when citizens are invited to submit ideas for policy interventions or public funds, the resulting dataset can be difficult for officials to process because it contains thousands of ideas that are all written in different styles. In this case, unsupervised techniques may be used to simplify the data by assigning ideas into broader categories, like commonly occurring themes.

Machine-learning is a particularly good fit for CI projects, many of which gather or interpret large datasets. The data is often human-generated content like images and videos, actively crowdsourced through smartphone apps and online platforms or scraped from social media or other public channels, such as the radio, where members of the public are passive contributors.

03

What is collective intelligence?

Figure 2



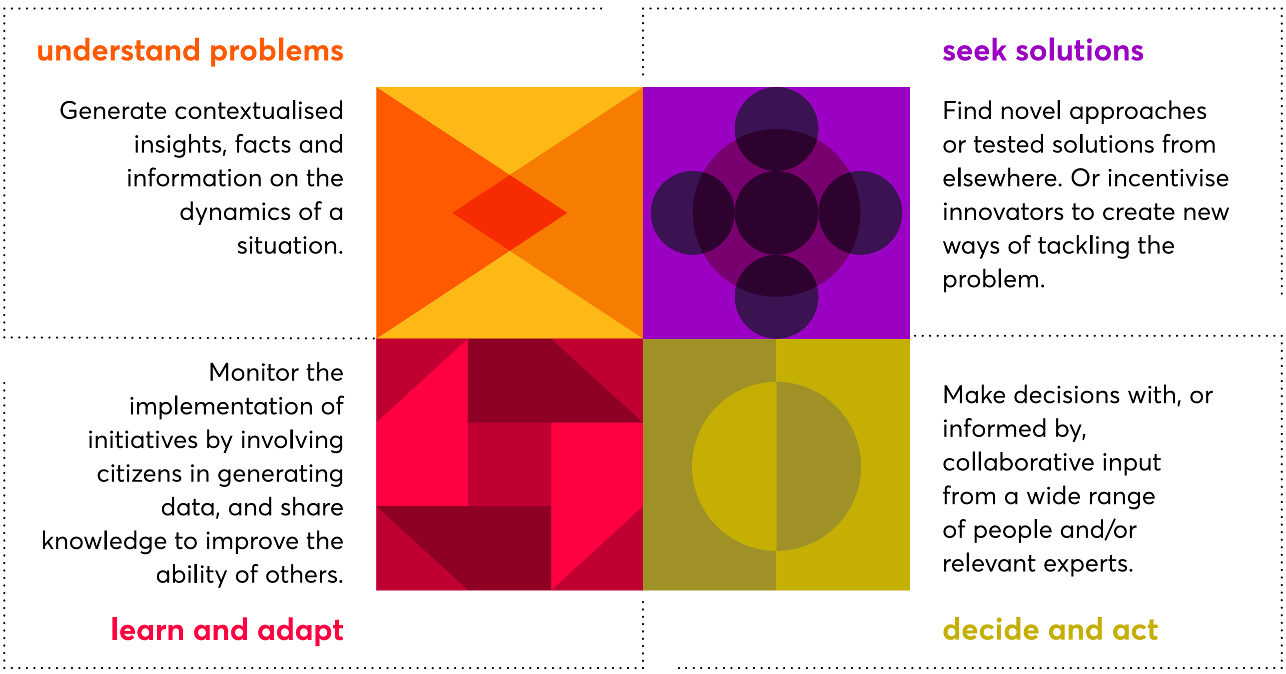
At its simplest, 'collective intelligence' can be understood as the enhanced capacity that is created when people work together, often with the help of technology, to mobilise a wider range of information, ideas and insights. Collective intelligence (CI) emerges when these contributions are combined to become more than the sum of their parts for purposes ranging from learning and innovation to decision-making.

It has been around for a long time, but the rise of new technologies that connect more and more individuals over greater distances to share knowledge and skills has transformed what can be achieved through CI. In the 19th century, it took almost 70 years to crowdsource the 400,000 words that made up the complete first edition of the Oxford English Dictionary.[8] A modern-day equivalent, Wikipedia, receives 1.8 edits per second and sees more than six million new pages created per month.[9]

CI covers a wide range of participatory methods, including crowdsourcing, open innovation, prediction markets, citizen science and deliberative democracy. Some of them rely on competition, while others are built on co-operation; some create a sense of community and teamwork, while others operate on the basis of aggregating individual contributions or microtasks. Academic research on CI is equally varied and draws on many different disciplines, including Social Science, Behavioural Psychology, Management Studies and Computer Science.[10]

Figure 3 highlights four distinct opportunities offered by CI to help decision makers and communities solve complex problems.

Figure 3



All modern examples of CI rely on bringing together people and/or data or insights in some way. Building on this, the core methods in CI can broadly be described as falling into three categories.

Connecting people with people

The oldest form of CI, bringing people together with other people. It can facilitate distributed information gathering, problem-solving, peer learning and prediction-making. Increasingly, this form of CI combines offline engagement with online contributions, which allows inputs from larger groups of contributors. Methods include peer production, participatory prioritisation, deliberation and open ideation.

CitizenLab is a citizen participation platform that allows local governments to reach out to communicate with the communities they serve. Citizens are invited to submit ideas, prioritise public spending or respond to consultations through an online interface, then the results are sent directly to policymakers. The platform fosters connections between local government and citizens to enable more informed and legitimate decision-making.

Connecting people with data

Brings both people and data together and often involves crowds generating, categorising or filtering unstructured data, such as photos or audio recordings. Some methods offer deeper engagement in processes beyond data gathering, to include participants in the scoping, analysis and evaluation phases of projects. Citizen science, crowdsourcing and crowdmapping are typical such methods.

MapWithAI is part of the digital humanitarian effort, hosted by the Humanitarian OpenStreetMap Team where volunteers create maps in regions where this information is missing or changing due to conflict or environmental disaster. Participants from across the world use images collected by satellites to trace features, such as buildings and roads. These maps help humanitarian agencies to plan missions and response activities that save lives.

Connecting data with data

Brings together multiple and diverse datasets to help generate new and useful insights. These methods increasingly make use of non-conventional data sources generated by people, such as posts on social media, mobile phone geolocation and sensor data. Data collaboratives, open-source repositories and open application programming interfaces (APIs) are some of the methods that are typically used in these data-driven CI projects.

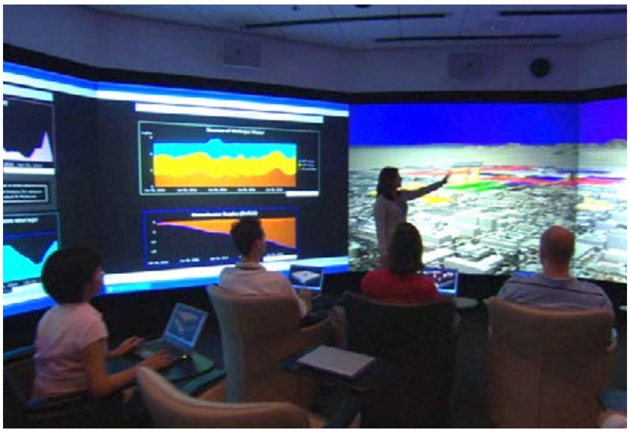
Dataminr is an early warning system that helps organisations to plan and co-ordinate their response to crises. It integrates many non-traditional data sources, including updates posted by citizens on social media, to monitor how situations unfold in real time and create summaries, which are sent as alerts. These updates are tailored to the needs of each client, so they contain the information most relevant to helping them plan and make decisions.

Collective intelligence is more important now than ever before

From the climate crisis, to the displacement and migration of human groups, to rising socio economic inequality, it can sometimes feel like the 21st century has been defined by the rise of increasingly complex problems.[11] Unlike simple problems, which follow more predictable trajectories and have obvious fixes, making progress on these complex problems requires dealing with uncertainty and multiple unknowns, where there isn't just one optimal solution.

This makes them ripe candidates for CI, which draws on a combination of data, technology and diverse human skills to address different aspects of uncertainty.[12]

The Decision Theater[13] developed by Arizona State University is a good example of what this CI looks like in practice. It draws on the complementary resources of data, technology and people to create an interactive simulation experience that supports decision-makers to navigate complex problems. To date, it has been used to devise strategies for dealing with problems from land degradation to co-ordinating aid during emergency response. A combination of big data analysis and high-performance computing is used to create detailed models of public policy scenarios. The models are then translated into interactive data visualisations, which are projected onto multiple screens to help teams of decision-makers explore the impact of intervening at various points. Decisions are ultimately made through group deliberation about competing values, trade-offs and priorities, but it is the technology that enables the decision makers to engage with the complexity of the problem.



The Decision Theater, Arizona State University

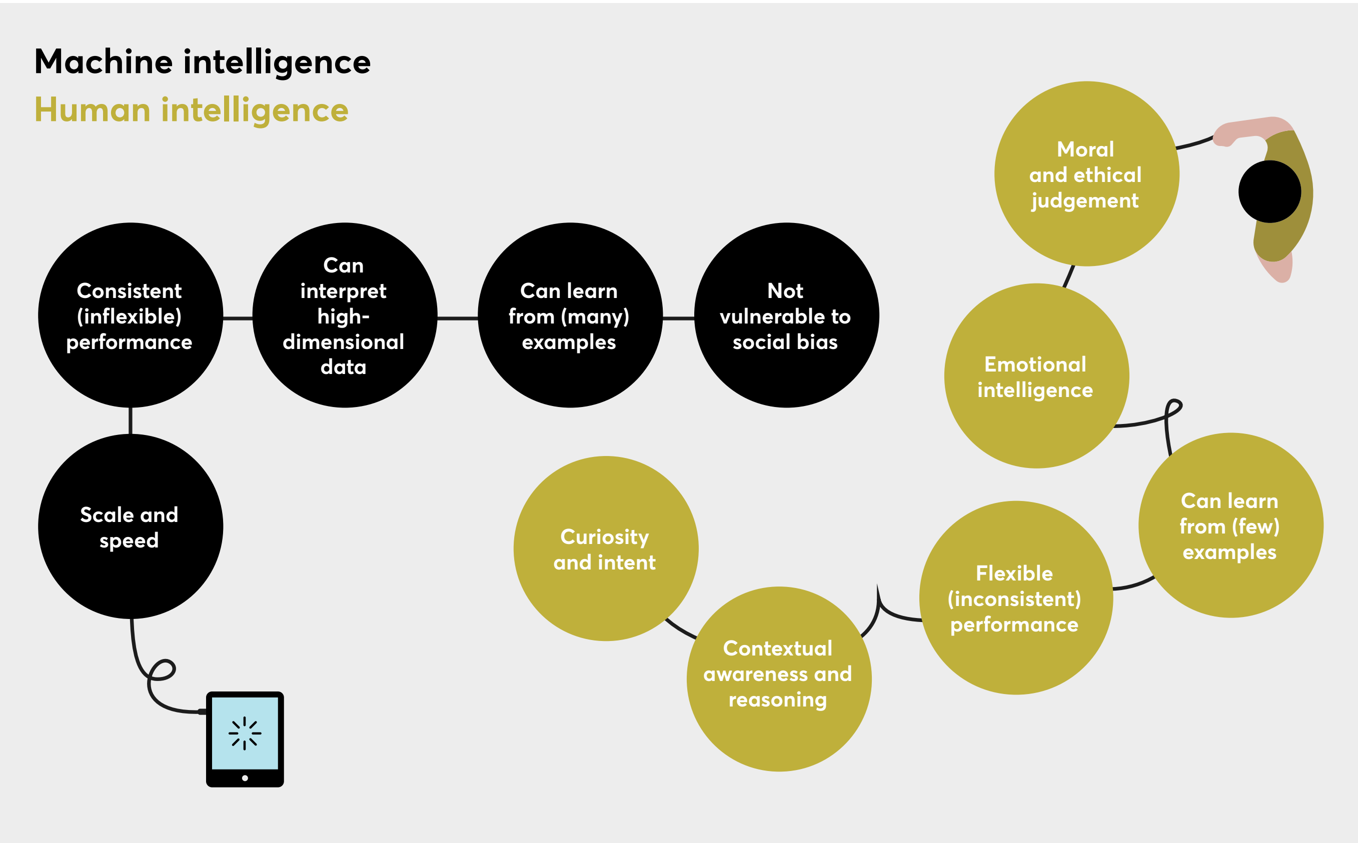
04

How human and machine intelligence are different

Like no other technology that has come before it, AI raises questions about the unique value of human intelligence. After all, if computers can do what we do, some of them with even better results, how special is human intelligence?

Such existential questions have been fuelled in part by the quest towards developing what is known as artificial general intelligence, which aims to recreate true human-like intelligence that is flexible and generalises between different tasks, rather than focusing on a narrow set of specialised tasks like most of the AI that we currently see in the world. Unlike human intelligence, these AI methods show little ability to transfer the skills learnt for solving one type of problem into a different context.[14]

The ambition of recreating human intelligence risks losing sight of the potential gains offered by combining complementary aspects of human and computer capabilities.



Complementary sources of intelligence for real-world problems

AI models are typically most useful where a 'ground truth' is well defined[15] and the data sources that the AI model uses as input do not frequently change. In these situations, the ability of AI to find patterns in huge amounts of data is useful for streamlining decision-making. For example, the medical search engine Epistemonikos[16] uses machine-learning to identify clinical systematic reviews in the academic and policy literature. For many years the definition of a systematic review has been globally agreed by medical practitioners and the formats of studies found in the literature follow a narrow set of templates acknowledged and expected by the sector. These characteristics makes it a perfect candidate for machine-learning because the training datasets are exactly representative of the data that the model will encounter when it is deployed in the real world.[17] Epistemonikos has been used by the Chilean government to increase the efficiency with which policymakers set new health guidelines based on the latest evidence.

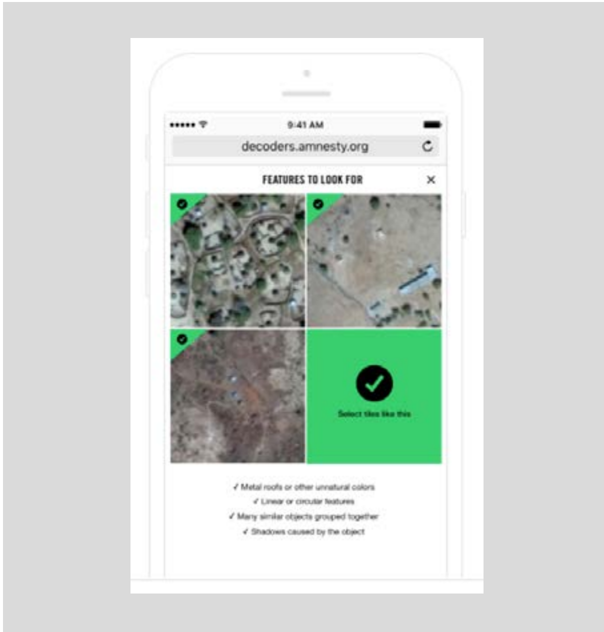
However, in the case of many complex real-world problems, like health epidemics and extreme weather events, the dynamics of the situation might lack historical precedent. This issue can result in so-called 'dataset drift', which means that the data a given model was trained on is no longer equivalent to the real-world situation in which it is used, so there is no guarantee that the model's predictions will be accurate. When it comes to high-stakes decision-making, such as co-ordinating disaster response or managing medical emergencies, this inaccuracy can be particularly dangerous.

In these circumstances, the human ability to adapt to new situations, understand context and update knowledge fills in the data gaps of machines. For example, the Early Warning Project (see page 31) draws on crowd forecasting to plug gaps in between its annual update to the statistical models it uses to estimate the risk of genocide across the world. In this context, the collective provides an alternative source of intelligence that can respond to weak signals[18] and unexpected developments that could influence political decisions.

At the collective level, attributes of human intelligence include the ability to tell collective stories as an act of sense-making and learning. For example, Decode Darfur – an initiative run by Amnesty International for their distributed volunteering community, Amnesty Decoders – asked participants to identify areas of destruction in the settlements of Darfur using satellite images. Apart from demonstrating group accuracy that was similar to classification by experts, the project provided an opportunity for the volunteers to discuss what they were seeing as an act of 'collective articulation of experience'.[19] It is difficult to imagine smart machines taking on this role and yet, when faced with societal challenges, having platforms for discussing and shaping collective values in this way is important.

So far, the well-known Turing test[20] and various specialised benchmarks[21] have been used by industry and academia as measures of the performance of AI systems, but they are primarily measures of relative performance between new AI methods. It is not so easy to develop an absolute intelligence quotient for machines, and the value of distilling intelligence into a single static number is as questionable for AI as it is for human intelligence.[22]

To truly take advantage of 21st-century collective intelligence, where AI is interacting with large groups of people and social systems, we will need to compare and contrast different ways of knowing the world and how best to combine them to solve a given problem. This process requires developing entirely new standards of evaluating performance, including criteria such as the level of human agency, how equally benefits and opportunities are distributed within society and the differential impact of AI on under-represented populations.



05

Making the most of the CI opportunity

Collective intelligence (CI) is powered by mobilising our collective knowledge and resources to solve problems.

Yet, anyone who has tried to organise the effective exchange of ideas in something as simple as a meeting knows how difficult it can be to do it well.

Creating larger-scale and looser online communities that share resources, tasks and goals is even more complicated. After all, there were at least seven failed attempts to create online collaborative encyclopedias before the success story of Wikipedia.[23]

Making deliberate choices that optimise the collective power of groups is known as **collective intelligence design**.[24] Below we outline some of the key principles that underpin CI and the challenges CI projects often face.

Crowds need to be diverse and get value from participation

One of the most important principles to help groups become more than the sum of their parts is diversity. For example, Anita Woolley has demonstrated that gender diversity helps smaller groups to improve their problem-solving ability and that cognitive diversity is vital for creativity[25] and learning[26] in workplace teams. The principle holds even when it comes to larger-scale CI efforts, where it may intuitively seem that diverse contributions could make it harder to distinguish signal from the noise. This is because the 'wisdom of crowds'[27] relies on everyone making mistakes in slightly different ways. For example, in crowd predictions, if some people have a tendency to overestimate, while others are more likely to underestimate, overall their errors cancel each other out, making collective estimates more robust to interference from individual biases.[28]

Another core principle of CI is making sure that the overall goal of the project brings some benefit to the participants, so that they are motivated to contribute. In citizen science, projects tend to offer a range of incentives, such as learning new skills or socialising with peers, which help to attract high-quality contributions from the community members. Allowing people to contribute their views equally and independently, and selecting the most appropriate methods for aggregating diverse inputs for the problem at hand are also important features of successful CI projects. For example, deliberative democracy methods that involve citizens in decision making may choose to adopt the final preferences of the majority or require the group to reach a consensus.

CI design challenges

Even though research has helped us to understand which principles are important for CI, operationalising them can still be challenging. The most common challenges that stand in the way of making the most of CI approaches to solving problems fall into three broad categories:

1. Making sense of the data

The big data at the heart of many CI projects is a double-edged sword. It can help drive better-informed actions and decisions, but without efficient systems for aggregating, organising and combining this data, there is also a risk of getting lost in the noise. Data for CI projects often comes from novel sources, like citizen-generated content on social media, mobile phone call records or satellite images, which require novel approaches to analysis in order to extract insights. Sometimes the data is of variable quality or, in the case of digital democracy projects, may require the integration of multiple competing viewpoints. Organising and prioritising this data to make it searchable and usable towards the goal of a project can be a difficult task for public sector and civil society organisations looking to use CI to address social challenges.

2. Setting the right rules for exchanging information and skills

Realising the benefits of a group's diversity is only possible when there are effective mechanisms for people to access information and skills. The rules governing how group members interact with one another impacts how easily ideas spread and take hold in the group. For example, well-timed breaks to achieve a balance between group work and exploring ideas at the individual level encourage productivity in problem-solving.[29] The structure of a community network can play an important role in determining whether a group is able to exchange information efficiently.[30] CI also relies on having shared open repositories of knowledge, accessible to old and new members alike. Well-known online platforms like GitHub or Wikipedia are often used by communities to document their processes. These function as a source of collective memory, and their absence can lead to the repetition of old mistakes over time.

3. Overcoming human cognitive biases

All people can be subject to biases, which affect their interactions with other people. In certain circumstances these biases are useful; they provide shortcuts[31] to help us manage our cognitive resources more efficiently. However, some social biases can interfere with the principle of diversity. Status and power imbalances between participants can silence those who are lower in the social hierarchy and confirmation bias can lead to certain ideas being devalued because they do not match assumptions held by the group.[32] Apart from biases, some fundamental properties of human cognitive architecture can prevent us from contributing effectively to CI projects. For example, due to limits in working memory and attention, we struggle to keep more than a small number of ideas in mind at any one time and can get easily distracted. These limits are put under even more pressure in CI projects where there is a lot of information to process.

In practice, many of these challenges overlap and can even exacerbate each other. For example, the limits of human attention may be stretched by poor platform infrastructure that overwhelms participants with too much data and information, making it difficult to make meaningful connections and build on the work of others. In recent years, experiments in CI design have expanded to include AI in order to address one or more of the challenges outlined above.

06

How different AI tools can enhance CI

The field of AI comprises a range of methods, at varying levels of maturity, for implementation in real-world contexts. Focusing specifically on the methods that are already being used to enhance CI we have identified broad application areas, which we refer to as '**types of AI**'.

In this section we define each of the types alongside the data they rely on as input and examples of how they are being used to solve problems in practice.

AI is being implemented across a range of CI methods, from citizen science to crowdsourcing and digital democracy platforms. Many well-established CI initiatives, such as the Humanitarian OpenStreetMap Team and Zooniverse, which bring together distributed volunteers for humanitarian action and scientific research respectively, have only started implementing AI into their community platforms in the past couple of years, after existing for a long time without it. Others, like Factmata and Wefarm, which crowdsource expertise to identify misinformation and help farmers to solve problems respectively, have been created as integrated AI and CI platforms from the outset. We also found examples that started with the technology first, such as Swarm AI, where the AI works in real time with human groups to answer questions and make predictions. Swarm AI can be used by any group or team, but it does not have a dedicated community of practice attached to the tool.

It is difficult to carry out a comprehensive mapping of either CI or AI projects separately, let alone together. Most existing examples do not document their use of methods in detail or give insight into any other approaches they considered. Thus, while we have tried to capture the breadth of the field, we cannot claim to be exhaustive. Further progress will require much more transparency and sharing of lessons between all of the actors in the AI and CI ecosystem.

Type of AI	Definition	Types of data accepted as input	Case studies
Computer vision	The ability of a computer to understand, analyse or generate images and/or videos.	User-generated** images and videos Satellite images Drone images	Syrian Archive, MapWithAI, iNaturalist, OneSoil (see page 29), WeRobotics (see page 42), Decode Darfur
Natural language processing	The ability of a computer to understand and translate human-generated text and potentially simulate language.	User-generated** text on social media and blogs Transcribed documents	Wefarm (see page 35), Factmata, Siminchikkunarayku
Speech and audio processing	The ability of a computer to recognise, analyse, manipulate and potentially generate speech or audio signals.	User-generated** audio Sensor data	Mozilla Common Voice project (see page 40), Siminchikkunarayku, Pulse Lab Kampala
Predictions on structured data*	The ability of a computer to analyse structured data using numeric and machine-readable data.	Official data (e.g. databases) Structured taxonomies Participant rankings	Early Warning Project (see page 31), Polis (vTaiwan), Epistemonikos
Distributed AI: autonomous agents	A system where many individual (software) agents react to their local environment and other agents to produce emergent collective behaviour.	Data on past performance	Swarm AI (see page 33), Malmo Collaborative AI Challenge
Autonomous systems: robots	A system situated within and as part of an environment that senses and acts on that environment.	Human teleoperation Video demonstrations	RoboTurk
Recommender system*	A subclass of information filtering system that seeks to predict the 'rating' or 'preference' a user would give to an item.	Participant rankings User-generated** images, videos and text	CAT Lab (see page 44)
Mixed methods			Dataminr, Zooniverse

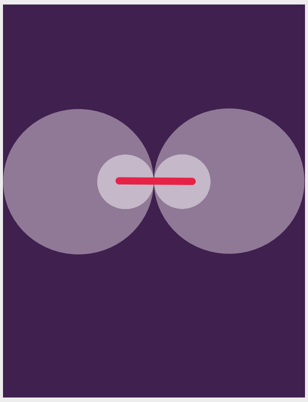
*Predictions on structured data and recommender systems typically rely on algorithms from classical machine-learning. See Figure 1 (page 9) for an overview of the AI methods currently used in CI projects.

** User-generated data inputs can be either actively submitted by participants of a CI initiative or collected by scraping online content and other public information.

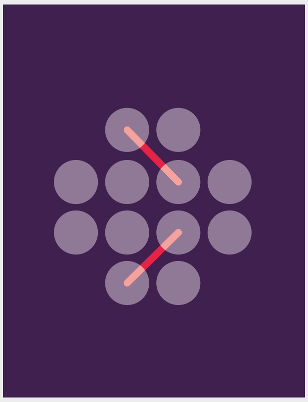
Explore our full AI and CI case studies feature online at
www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies

Added value from AI

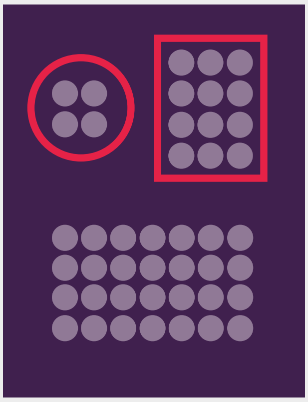
Where AI is already enhancing CI, the value it adds to the collective can be broken down into eight broad categories. These categories address challenges commonly encountered in CI, such as those related to managing data, process and people (see Section 3). These different sources of added value demonstrate that AI functions as a cognitive tool that allows the extension of natural intelligence,[33] from perception and decision-making to creativity and learning. Much like the transition to literacy (both handwritten and print) from predominantly oral cultures first amplified individual and collective intelligence for people, by externalising memory and extending the reach of ideas,[34] the added value that smart algorithms bring to CI comes in many different forms.



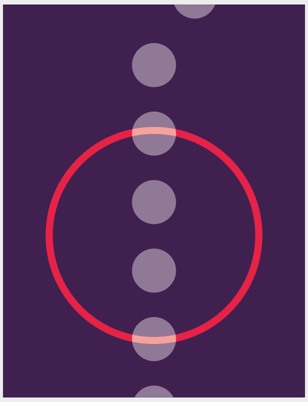
Growth in efficiency and scale of data processing:
Combining and processing vast amounts of structured and unstructured data to enable quicker response.



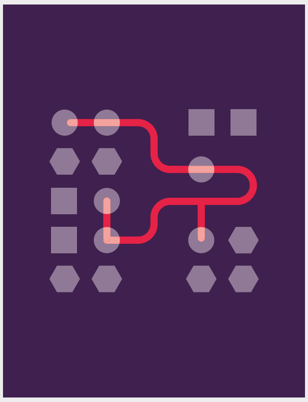
Making new causal connections for complex issues:
Identifying new patterns and relationships between variables in high-dimensional datasets, either to infer causality or make predictions.



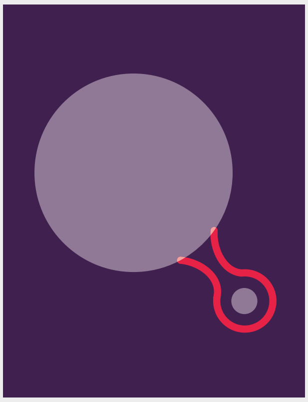
Organising human knowledge and finding structure:
Filtering, clustering and ranking information and helping to group data into categories.



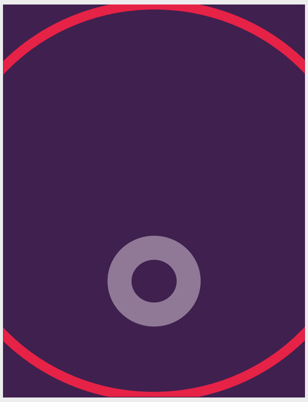
Optimising and prioritising processes:
Streamlining complex systems and processes to optimally satisfy different (competing) resource constraints.



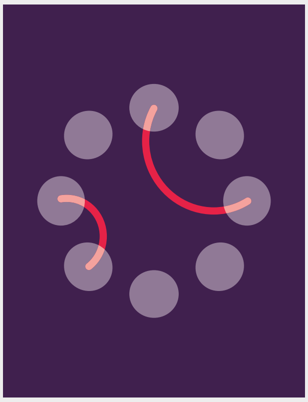
Modelling and visualising complexity:
Creating models of entire systems to expand our capacity for systems thinking, multi-stage processing and an individuals' role within a collective.



Increasing the reach of and access to rare knowledge:
Using rare knowledge and non-traditional data sources to codify specialised expertise and gain new insights.



Transformational creativity:[35]
Expanding the boundaries of the solutions normally considered by humans, generating new ideas or helping to surface old ideas.



Improving interactions between individuals:
Optimising individual contributions and facilitating knowledge exchange between members of a group in real time.

How AI can help address CI project challenges

The table below summarises some of the most common design challenges faced by CI projects, as outlined on page 21. Drawing on the tools currently being used, we provide examples of the potential opportunities offered by AI methods to overcome these challenges and enhance CI.

Common CI challenges	Specific problem	AI opportunity
Making sense of the data	Huge amount of data; many different data sources contain relevant information for a given issue.	Predictive models that identify patterns and estimate the likelihood of outcomes based on many variables.
	Vast amount of big data from novel sources that are difficult to organise and make sense of.	Models that classify data into categories and produce structured content.
	High number of options when seeking the best solutions to problems.	AI systems that match the most appropriate solutions to the problem holder.
Rules for exchanging information and skills	Co-ordination problems in decision-making processes	AI systems that optimise co-ordination between different options.
	Vulnerability to rapid spread of misinformation and negative content across the community network.	AI chatbots to reinforce online community norms or automated detection of negative content.
Overcoming human cognitive biases	Need for novel and creative solutions.	Creative AI that generates unusual and unexplored solutions to well-defined problems.
	Vulnerability to group biases discourages sharing diverse information.	AI bots to mitigate against bias in groups.
	Inability to focus on common goals and understand/estimate collective benefits or risks.	AI agents to suggest actions related to collective risks, benefits and goals.
Spanning multiple challenges	Failure to share and extract lessons from rare knowledge and experience to enable collective level learning.	Machine-learning to identify relevant patterns from past performance to tailor recommendations for learning.
	Need to engage and sustain the involvement of volunteers.	Tailoring participant training and task allocation based on optimal individual ability.
	Unequal contributions between participants, either due to oppressive dominance or social loafing.	Bots to facilitate group interactions and encourage contributions from all group members.
	Difficulty in identifying areas of consensus and disagreement.	Ranking and organising huge amounts of data to understand where overlaps lie.
	Difficulty imagining impacts of local-level actions on systemic issues.	Simulations that create a model of the issue being discussed or an entire system (e.g. traffic flows in a city).

07

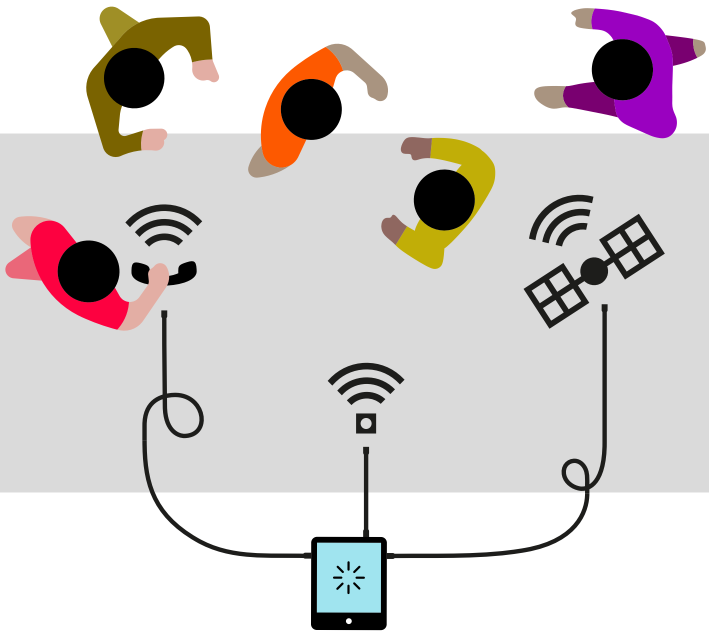
A framework for understanding how AI and CI interact

In spite of the rapidly growing field of projects introducing AI into CI initiatives, there are currently no frameworks for understanding the different forms of interaction between crowds and machines.

This presents a challenge to understanding the field as a whole and how to support its future development.

Based on an analysis of case studies[36] and emerging academic research, we have identified at least four ways we can begin to understand this relationship. Some projects are more advanced and include more than one type of interaction. Although these categories span different levels of maturity and will undoubtedly evolve as the field continues to grow, they provide a starting point for those interested in exploring the current AI-enabled CI landscape and future opportunities.

Machines working on data generated by people and sensors



In the first type of interaction, distributed networks of humans and/or sensors produce data that is used as continuous real-time input for machine-learning algorithms. Data is produced by individuals (or sensors) independently of one another, but the AI works on the aggregated collective data. In these cases, the human contributors interact with the AI passively.

A typical data source is user-generated content online, such as videos, photos and text on social media platforms. Dataminr is one example that uses passive user generated content scraped from the internet to monitor for unexpected events of public interest – such as environmental disasters or public health emergencies – in order to produce early warning alerts for officials who need to plan responses in real time.

Other examples use data collected from remote or on-the-ground sensors, which can vary from satellite data to geolocation data from mobile phones or specialised hardware that measures atmospheric conditions. The latter is used by OneSoil (see opposite page), a non-profit platform that provides real-time insights to help the agricultural community make decisions and plan for the future. OneSoil uses AI to analyse images from European Sentinel satellites and on-the-ground sensors in order to map field boundaries and estimate crop health on farming land.

Case study: Onesoil

Using computer vision, based on supervised machine learning, to support a global community of farmers

The challenge

From crop diversity to weather patterns, farmers need to consider a high number of variables when deciding how to manage their fields. Accurately monitoring the growth of crops and the effects of various interventions can have a significant impact on yield and income. Gathering the relevant information can require costly specialist equipment and can be difficult to maintain over large territories. The ability to integrate multiple vast datasets in real time to generate predictions is also beyond the capabilities of individual farmers and their teams.

The AI and CI solution

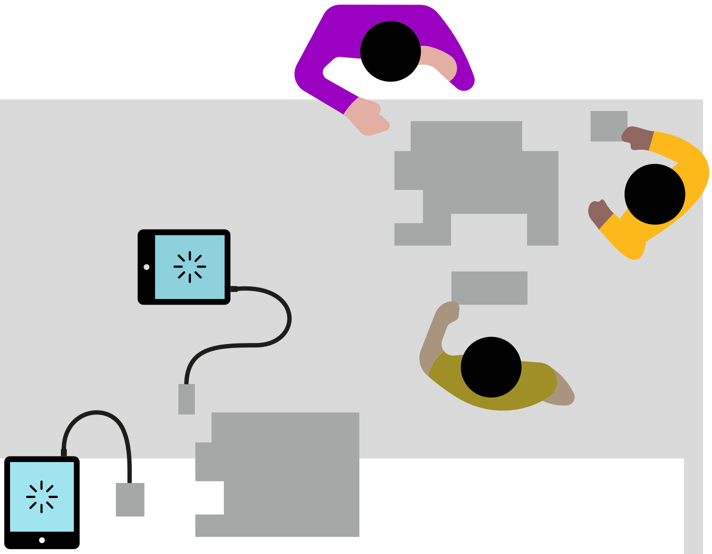
OneSoil is an app that has been developed to help farmers and their teams make better decisions. It uses machine-learning and computer vision to assign field boundaries and crop types to farmland, as well as calculate a vegetation index as a proxy for crop health. The OneSoil algorithms are trained on the high-quality open data obtained through the European Commission's satellite programme, Copernicus. On the platform, this data is combined with localised measurements from on-the-ground sensors tracking various environmental features – such as soil moisture, air humidity and temperature – to give a more comprehensive understanding of the state of every field. This up-to-date information collected by farmers through sensors provides the vital additional detail needed to accurately model the farming environment, predict outcomes and manage interventions.

OneSoil also supports effective team communication and collective learning. All employees, from machine operators to managers, can add information onto the platform, including correcting mislabelled crops and adding manual annotations and image attachments reflecting the current status of the fields.

So what?

The cross analysis of big data from weather sensors and satellite images, alongside active annotations from users, helps farmers predict their lands' productivity, plan and monitor fieldwork and predict the emergence of possible plant diseases and pests. The maps on the platform are updated every two to five days, so farmers can manage their fields using near real-time data. OneSoil has mapped more than 35 million fields in Europe alone and is being actively used for decision-making by the global agricultural community.

Machines and people taking turns to solve problems together



In this form of AI and CI interaction, a distributed network of people actively performs microtasks, while AI is used as an alternative source of intelligence. Participants typically carry out tasks independently from one another, which are aggregated to produce a collective output, such as a 'wisdom of crowds' prediction. In these situations, the AI and distributed crowds work on different parts of the problem. One example is the Early Warning Project, (see opposite page) which uses both crowd forecasting and statistical modelling to generate predictions about the risk of mass atrocities worldwide. In combination, the methods offer complementary insights and counterbalance each other's weaknesses.

Similarly, the Waze app, used by over 115 million drivers all over the world,[37] draws on complementary aspects of human and machine intelligence. It uses AI to learn the day-to-day patterns of its users and map out potential travel routes, which is supplemented with real-time crowdsourced information by the app's users. The hyperlocal and dynamically changing information about road construction projects, traffic conditions and even fuel prices supplied by users is combined with the AI-generated directions to suggest the optimised final route.

Another example is the iNaturalist app, which supports an online social network of nature enthusiasts to log sightings of different species. It uses a combination of computer vision, community feedback and individual expertise to classify observations and generate an open dataset on global biodiversity that is used in scientific research and conservation.

Case study: Early Warning Project

Combining expert assessment, crowd predictions and statistical modelling to anticipate mass atrocities

The challenge

Triggers for humanitarian crises, such as mass genocide, are difficult to predict in advance due to their rare occurrence and the variety of different contributing factors, some of which can change with little advance warning. The ability to accurately estimate the likelihood of genocide and mass atrocities could help better co-ordinate responses and prevent some of the trauma and devastation caused by these crimes.

The AI and CI solution

The Early Warning Project (EWP) tried to address this challenge by improving the early warning system for mass atrocities using a novel combination of crowd forecasting, expert ranking and statistical modelling.

The project was divided into three phases. During the first phase, experts in the field participated in an annual comparison survey, where they ranked pairs of countries by choosing which is more likely to experience a new mass killing.

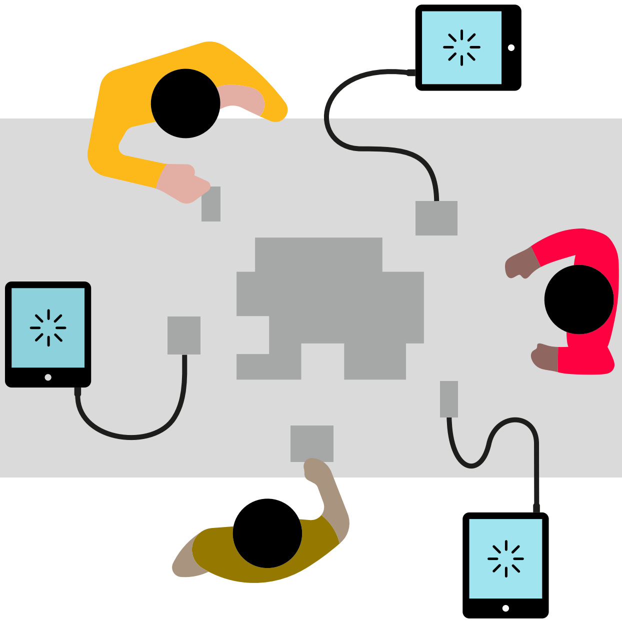
The results from this survey informed the selection of 17 'higher risk' countries, which the EWP tracked in real time using 'crowd forecasting'. Crowd forecasts are calculated by aggregating many individual judgements about the likelihood of events, ranging from the outcomes of political elections to the winning teams for international sporting events. Participants can update their estimate over time depending on how different factors that influence the outcome evolve. In the EWP, crowd forecasting took place over the course of a year. Previous research[38] on crowd forecasting has suggested that a non-specialist crowd can predict geopolitical events more accurately than individual intelligence analysts.

Alongside these human predictions, the EWP calculated a risk assessment score using statistical algorithms, one of which relies on a classical machine-learning method called random forest. The algorithms generated their estimates based on more than 30 different variables from historical datasets, ranging from basic facts about the country, such as population size, to more specific measures of attitudes on human rights and civil liberties.

So what?

The EWP produced a ranked list of more than 160 countries, based on their likelihood of experiencing a mass killing, in order to better target preventative action by governments and charities. Mass atrocities are rare events that have little historical precedent and so the EWP's approach ensures that weak signals from the crowd consensus predictions help address gaps in the statistical risk assessment and expert recommendations. The project is an example of combining different complementary capabilities of humans and AI to inform decision-making.

People and machines solving tasks together at the same time



Instead of taking turns, this form of collaboration happens in real time. AI forms part of the group that is working together on the same task. It places AI and people into a highly interdependent relationship and is reliant on trust and social acceptance of AI. Most existing work on this form has taken place in lab-based experiments or through gaming, such as the Project Malmo virtual environment, which has been built on top of the game Minecraft. The Malmo Collaborative AI Challenge was a competition where artificial agents played a collaborative game with other agents and humans to advance research into co operation between people and AI.

Swarm AI, an online platform developed by Unanimous AI, is a rare real-world example, where groups of people and AI agents work together as part of a closed-loop system to make consensus decisions and predictions in real time, on issues ranging from medical diagnosis to political preferences.

Autodesk's generative design software[39] for collaborative design is another example. In this case, the AI gives designers and other users iterative suggestions for possible permutations of a solution based on the parameters that it is given. Generative design has been credited with extending human creativity by moving beyond the boundaries of the solution spaces that designers typically explore. To date, generative design has mostly been used in industrial manufacturing and product design, but it is easy to imagine the technology transforming public spaces and urban planning based on real-time interaction with a larger community of individuals. A demonstration of the viability of this approach is the Autodesk office in Toronto, which was created using generative design based on parameters specified by 300 employees,[40] as well as other factors.

Case study: Swarm AI

Connecting hybrid human–machine groups to make decisions together in real time

The challenge

Bringing together diverse groups of people to make decisions or predictions is a key goal of many collective intelligence projects and a core principle of the wisdom of crowds. However, the interactions between people who have different experiences and viewpoints can be difficult to co-ordinate. Conflicting viewpoints, interference from social biases and failure to tap into the expertise in the group can all stand in the way of an optimal consensus being reached. We therefore see many attempts to mobilise networked groups of individuals (both online and offline) leading to the opposite outcome, with incendiary behaviours in response between those who disagree and the rising polarisation of views.

The AI and CI solution

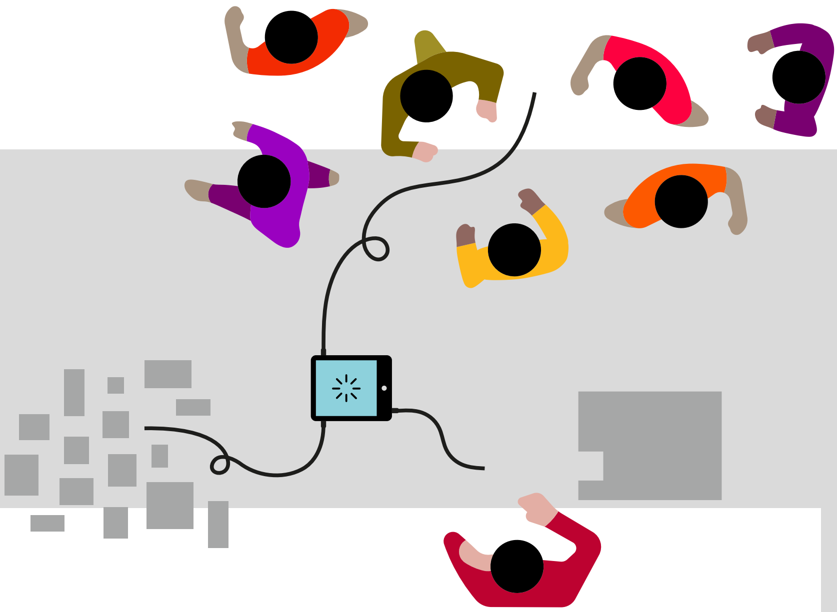
The Swarm AI platform, developed by Unanimous AI, is a rare example of distributed AI and human groups working together on a task in real time.

Swarm AI is inspired by the collective behaviour of natural systems, such as flocks of birds and swarms of bees. Swarm intelligence algorithms moderate the interaction of a group of individuals who are deciding between a set number of options. Each person can log into the online platform at their location. The algorithms are trained on data about behavioural dynamics of groups, rather than on the subjects they are debating. Individuals connect with each other and AI agents to form a closed-loop system where both the machine and individuals can react based on the behaviour displayed by others to change or maintain their preference. In a second step, a neural network model trained with supervised machine-learning uses the interaction dynamics of the participants to generate a conviction index. This index estimates the group's confidence in the final outcome.

So what?

The Swarm AI platform has increased the accuracy of group decisions across a wide variety of tasks, from health diagnostics to forecasting political polls. For example, diagnostic accuracy of a small group of networked radiologists working as a real-time swarm intelligence system reduced errors by 33 per cent in comparison to the individuals on their own, and by 22 per cent compared to an AI-only solution. Unanimous AI has claimed that the Swarm AI system navigates the group towards optimal consensus decisions, which result in higher levels of satisfaction in the group. As of January 2020, Swarm AI has been deployed in low stakes decision-making in commercial settings or in research environments, but the results show promise for applications in the public sector, such as prioritisation of public policy.

Using machines to connect knowledge and tasks in groups



AI can also play a vital role in enabling more efficient and streamlined CI projects by helping people better navigate different kinds of information and tasks. In this type of interaction, AI is used for back-end functionality to enhance the individual capabilities of users to perform tasks or improve their experience. We see this type of AI contribution as 'greasing the wheels' of a CI process.

It can be achieved in many different ways, such as through better matching of individuals in a community who have common interests. For example, the Wefarm app (see opposite page) is a peer-to-peer network of farmers who use it to ask for advice about issues they encounter from others in the community. It uses AI to analyse the requests posted by farmers and matches them to others who are the best qualified to answer. A different version of this interaction is offered by the Syrian Archive, which crowdsources footage of conflict from Syria and has implemented AI to optimise the experience of searching through this rich material. It is used by activists and non-profit organisations to gather evidence of human rights violations. Here, the AI is enabling better matching between people and information.

Other examples are focused on better matching tasks to individual users. We see this in the Gravity Spy project on the Zooniverse platform, which tailors the training process for each new user on the project, resulting in a better performance and an improved experience for the citizen scientist. A slightly different approach has been tried by the Space Warps project, also on the Zooniverse platform, which has started experimenting with intelligent task assignment for participants to optimise co-ordination between different skill levels.

Case study: WeFarm

Using natural language processing (NLP) to enable peer support between smallholder farmers

The challenge

Over one billion smallholder farmers produce 80 per cent of the world's food, and four of the five most traded commodities on earth. Yet the vast majority lack internet connections and access to up-to-date information to help them solve problems or share ideas with their peers. The primary challenges faced by peer networks are co-ordinating between users efficiently and matching needs to existing expertise within the network.

The AI and CI solution

Wefarm is a free peer-to-peer service that enables small-scale farmers to share information via SMS, without the internet and without having to leave their farm.

Farmers in Kenya, Uganda and Tanzania use Wefarm to ask each other questions about anything related to agriculture, then receive crowdsourced bespoke content and ideas from other farmers around the world within minutes. Machine-learning algorithms then match each question to the best suited responder within the network, based on analysis of the content and intent. The project uses natural language processing (NLP) models that can identify three regional African languages – Kiswahili, Luganda and Runyankore – in addition to English. The questions can be asked in any language and messaging is free of charge. If farmers don't have internet access, they can access Wefarm via SMS on their mobile phones.

So what?

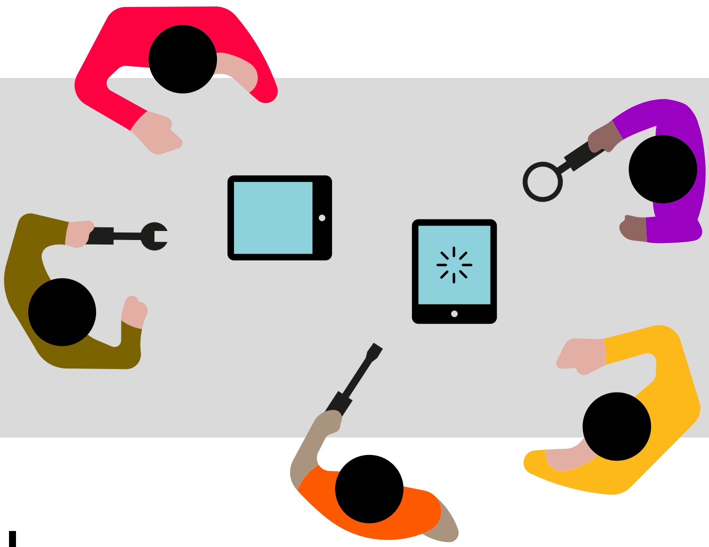
The AI provides an efficient routing between farmers' needs and rare human expertise that is uniquely suited to solving the problem. The NLP models used by Wefarm are some of the first to support regional African languages, which enables access and advice to the broadest possible group of users. As a result, Wefarm has grown to become the world's largest farmer-to-farmer digital network, with almost two million farmers using it in Kenya and Uganda alone, sharing more than 40,000 questions and answers daily.

08

CI for better AI: The fifth interaction

While the use of AI in CI initiatives is relatively new, CI-based methods have played a significant but often unrecognised role in shaping the development of AI. Below we discuss this evolution and how increased recognition of and investment in CI could help develop better AI.

Using CI to audit and support the development of better AI



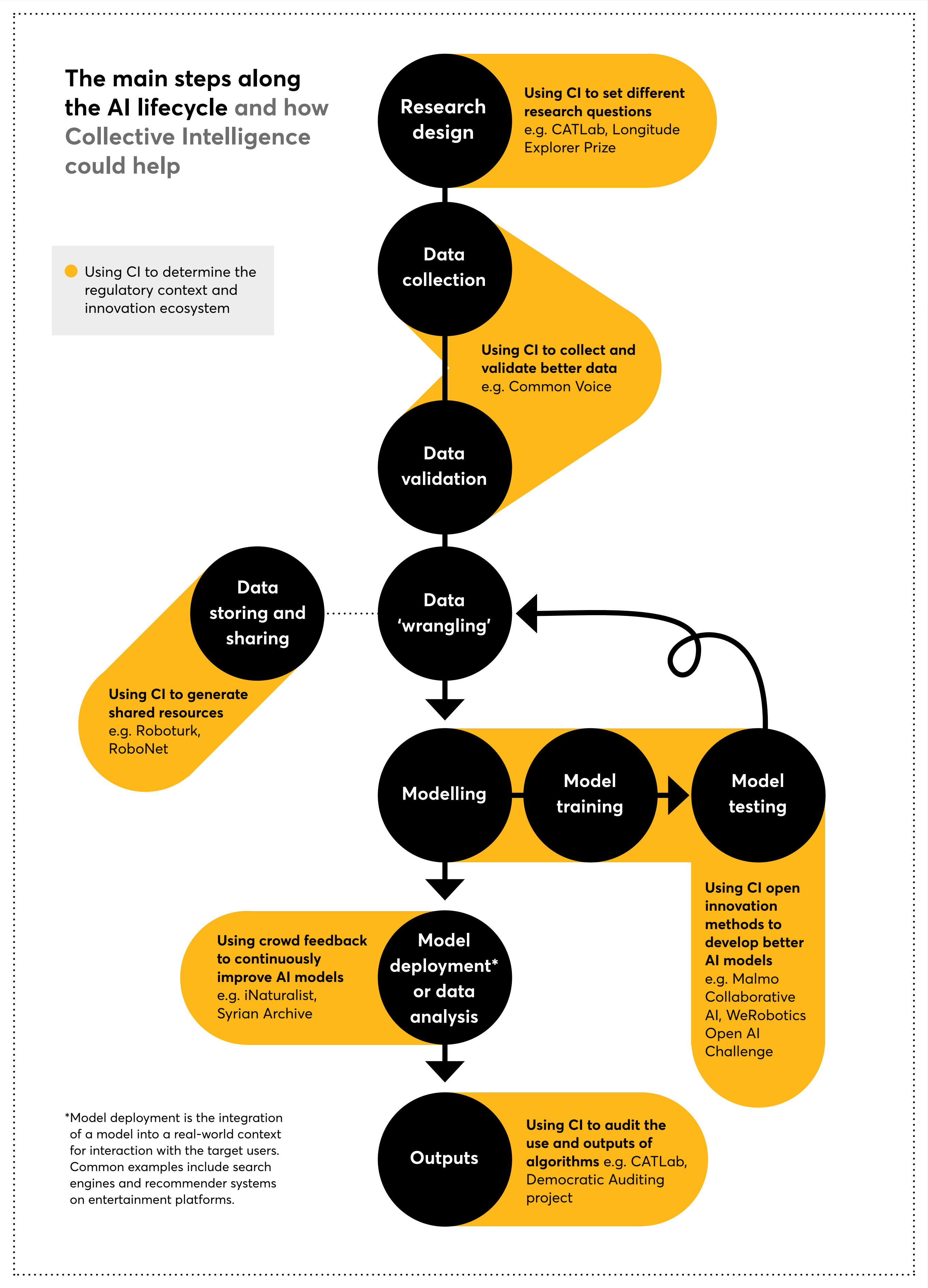
The hidden story of CI in AI

AI today is much more dependent on groups of people than is often appreciated. Unseen collective labour contributes to the latest advances in AI, especially machine-learning approaches that rely on humans to label the large datasets they use as their training material.[41] Whether these activities are compensated through crowd-work platforms such as Amazon Mechanical Turk[42] and Figure Eight[43] or obtained through internet traffic with tools such as reCAPTCHA,[44] they often remain the untold story[45] behind the success of AI. Concerns about the quality of data labels and the motivations and productivity of crowd workers are the focus of hundreds of papers in the machine-learning research community.

More recently, companies have started to enable industry and researchers to outsource the management of crowd workers. Mighty AI, Scale and understand.ai are just a few that claim to have a more specialised labelling workforce.[46] The differences between these crowd-powered endeavours and the mobilisation of crowds by CI projects are nuanced. It often comes down to the framing incentive structures, prioritisation between technology versus participant outcomes and their overall purpose (i.e. the aims they prioritise).

CI projects typically articulate a collective purpose, while paid crowd-labelling efforts emphasise individual gains by paying for each completed microtask. Research has shown that incentivising collective performance[47] can lead to better outcomes by preserving the diversity of individual contributions; in contrast, individual incentives encourage increased conformity and impaired collective performance. On the other hand, some of the challenges faced by researchers working with crowds in either AI or CI show unsurprising convergence. For example, across both fields there is literature that addresses how to optimise participant performance and track the changing motivations of the crowd to maintain engagement.[48]

Figure 4


The main steps along the AI lifecycle and how Collective Intelligence could help
Using CI to determine the regulatory context and innovation ecosystem
Research design
Using CI to set different research questions e.g. CATLab, Longitude Explorer Prize
Data collection
Using CI to collect and validate better data e.g. Common Voice
Data validation
Data storing and sharing
Using CI to generate shared resources e.g. Roboturk, RoboNet
Data 'wrangling'
Modelling
Model training
Model testing
Using CI open innovation methods to develop better AI models e.g. Malmo Collaborative AI, WeRobotics Open AI Challenge
Using crowd feedback to continuously improve AI models e.g. iNaturalist, Syrian Archive
Model deployment* or data analysis
Outputs
Using CI to audit the use and outputs of algorithms e.g. CATLab, Democratic Auditing project
*Model deployment is the integration of a model into a real-world context for interaction with the target users. Common examples include search engines and recommender systems on entertainment platforms.

How collective intelligence could help create better AI

There is no doubt that the costs and potential negative impacts of AI throughout its lifecycle deserve careful attention, but there is also a risk that these issues will overshadow the potential public benefits of the technology.

Collective intelligence offers an alternative path towards an AI-enabled future. By getting more of us involved in questioning AI, scrutinising its impacts and imagining which problems we should be applying it to, we can move towards more values-driven deployment of smart machines. Below, we outline how the principles and methods of CI can be used to counterbalance some of the limitations of AI and ultimately, improve its development.

CI for better training data

Although crowdsourcing is already widely used within the AI community,[49] it is rarely deployed with the explicit purpose of improving training data in order to develop technologies that have a wider public benefit and do not place under-represented groups at a disadvantage. One of the ways to develop fairer AI systems is to train them on datasets that are more representative of the diversity of real-world experience (an underlying principle of CI). This requires a more deliberate effort to involve individuals with rare knowledge, such as members of indigenous cultures or speakers of unusual dialects, in data collection. Mozilla's Common Voice project uses an accessible online platform to crowdsource the world's largest open dataset of diverse voice recordings, spanning different languages, demographic backgrounds and accents. Common Voice aims to open up the AI market and stimulate the development of AI voice assistants that are able to serve the needs of more diverse communities. Other projects focus on a smaller subset of languages; for example, Siminchikkunarayku and Masakhane[50] focus on Peruvian and regional African languages respectively. These projects recognise that developing AI systems based on under-represented languages and voices helps to preserve and increase the reach of this unique cultural heritage. The resulting AI systems are also better able to serve the needs of these groups, who would otherwise be excluded.

Case study: Common Voice

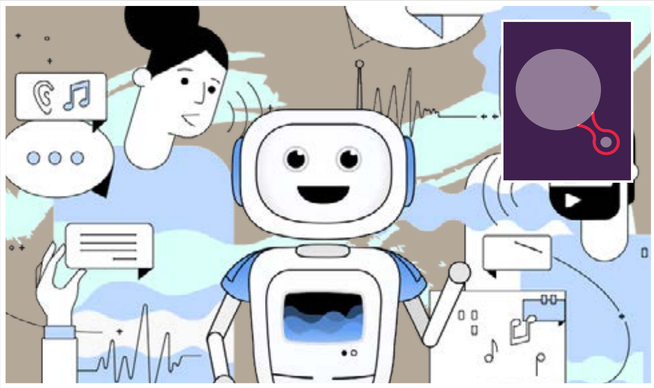
Crowdsourcing voices to train speech recognition software

The challenge

Most of the software and voice data that powers the personal assistants in our smart devices is locked up in privately owned systems. Getting access to good- quality data takes time and money. As a result, the cost of developing speech recognition and other software that relies on voice data is prohibitively high, giving a few companies a monopoly on these services. There is also little transparency about what data has been used to develop smart assistants, meaning that certain populations can remain underserved. These limitations make the technology less effective for some groups, such as non-native speakers with accents, or for languages spoken by small populations.

The AI and CI solution

Common Voice is a Mozilla initiative, which addresses this challenge by developing the world's first open-source voice dataset and a speech recognition engine, called Deep Speech. The concept is simple. Common Voice crowdsources voice contributions through an online platform where users are invited to record themselves reading sentences. All sentences are sourced from texts that are under a Creative Commons license , to ensure they can be freely reused by researchers and entrepreneurs in the future. Users can also listen to and validate the contributions from others to ensure that the data is of high enough quality to train an AI algorithm. The market's leading voice technologies are powered by deep learning algorithms, which can require up to 10,000 hours of validated data to train.



So what?

As of January 2020, users have recorded almost 2,500 hours of their voices in 29 different languages for Common Voice. The aim of the project is to ensure that the data used to train voice recognition tools represents the full diversity of real people's voices. Each data entry contains an audio file with the linked text, as well as any associated metadata about the contributor, if it is available. By making the datasets open, Mozilla is creating opportunities for a wider range of researchers, developers and public sector actors to develop voice technologies that can benefit a wider range of people. This accessibility can help to incentivise innovation and healthy competition for better tools. Mozilla released the first version of Deep Speech in 2017.

Common Voice is an example of how a CI approach to data collection – that emphasises diversity and open access – can be used to improve the development of AI, which in turn has the opportunity to be used for other CI purposes.

Similar initiatives: AI4D-African Language Dataset Challenge,[51] Niger-Volta Language Technologies Institute GitHub for West African languages,[52] Masakhane.[53]

All of these projects also integrate CI principles in their approach to data sharing by developing open, shared repositories of data. When data is rare, it is even more important to institute these practices so that more researchers and organisations can use the data to make progress. Sometimes the data required to train AI models is difficult to collect because it is very specialised. This is the case for autonomous systems in robotics, which rely on physical demonstrations by humans in order to develop machines that reproduce actions in the real (physical) world. Collecting demonstrations of tasks performed by people, whether recorded on video or through interactive displays, can be time consuming and cumbersome. The RoboNet[54] open data repository, which brings together data collected by robotics labs from different institutions and the crowdsourcing platform RoboTurk, are two recent efforts by the robotics community to accelerate progress in the sector using CI.

Using CI to open up innovation

From peer knowledge production to targeted Challenge Prizes, many CI methods are concerned with harnessing the power of groups to generate new ideas or develop new solutions. The computer sciences and tech communities have embraced open innovation practices, such as the peer production of both code and models to accelerate the development of software. Dedicated open innovation platforms such as Kaggle and InnoCentive have made it simple to outsource difficult problems to distributed communities of experts, but it is rare for these communities to take explicit measures to ensure the diversity of their contributions. Some open-source software communities have even been accused of the opposite: promoting harmful power dynamics that discourage newcomers from contributing.[55]

Some exceptions demonstrate how the principles of CI can be integrated into open innovation practices. The global peer learning community coordinated by WeRobotics (see page 42), regularly runs a Challenge Prize known as the Open AI Challenge to stimulate AI development around a specific issue that its members (the Flying Labs network) are facing. During the Tanzania Open AI Challenge in 2018 the organisation worked with local partners and non-profit groups to encourage participation from innovators based in or originating from Africa. Restricting competition entries to groups that are otherwise excluded from the innovation ecosystem can help broaden the range of ideas about how AI is developed and applied to solve societal problems. An example of this approach is Nesta's Longitude Explorer Prize,[56] which has been running since 2014 and targets applicants between 11 and 16 years old. It invites teams from schools in the UK to submit creative ideas for how to use AI to make progress on some of society's biggest challenges, such as the climate crisis and ageing populations.

Case study: WeRobotics Open AI Challenge

Drones, AI and robotics for a local-to-global learning network

The challenge

The development and use of robotics, drone technology and AI have mostly taken place in the Global North. However, it is the Global South that may stand to benefit most from these new tools. WeRobotics is a global initiative of more than 20 innovation labs across Africa, Latin America, Asia and Oceania called the Flying Labs network, set up to explore this opportunity. Their mission is to decrease the digital divide and use the Flying Labs to build skills, expertise and an innovation ecosystem to address local challenges. They primarily use drone solutions to acquire high-resolution local aerial data to support the implementation and impact of local aid, health, development and environmental projects. However, analysing the high-dimensional drone footage is labour- and time-intensive, and most existing AI tools have not been developed to detect the features most needed by the Flying Labs.

The AI and CI solution

In recent years, WeRobotics has run three challenge prize competitions called the Open AI Challenge, which have led to the development of AI tools to help analyse drone footage in Tanzania, the Caribbean and the South Pacific. Understanding local context can be vital to building an AI tool that is able to detect the most relevant features on maps, so the team takes additional steps to encourage participation from local innovators, as well as the international data science and AI community. For the Open AI Tanzania Challenge in 2018, they partnered with Black in AI, DataKind and local universities to attract African data scientists. The winning entry was developed by tuning existing convolutional neural networks (deep learning) to perform segmentation and classification of building types appropriate for the local context.

So what?

In recent years, WeRobotics has run three challenge prize competitions called the Open AI Challenge, which have led to the development of AI tools to help analyse drone footage in Tanzania, the Caribbean and the South Pacific. Understanding local context can be vital to building an AI tool that is able to detect the most relevant features on maps, so the team takes additional steps to encourage participation from local innovators, as well as the international data science and AI community. For the Open AI Tanzania Challenge in 2018, they partnered with Black in AI, DataKind and local universities to attract African data scientists. The winning entry was developed by tuning existing convolutional neural networks (deep learning) to perform segmentation and classification of building types appropriate for the local context.

CI to audit and monitor AI

Some algorithmic methods are already being applied in complex social contexts that rely on value judgements and moral reasoning, such as: who should be granted early release from prison?Who has a right to social welfare? How do we prioritise housing allocation for those in need?[57] Even if AI can help us assess and model these situations, we first need to have a wider conversation about the values we want technology to help promote and the level of responsibility that we think it is appropriate to delegate to autonomous systems as opposed to human judgement. To do this effectively, we need to use methods of collective sense-making that will allow us to interrogate algorithmic systems at all stages of the AI lifecycle – not just at the tail end, where algorithmic performance is judged by the outputs. To date, most public consultations about AI have focused solely on decision-making by AI systems or abstract notions of AI ethics rather than this more holistic practice-based approach. (see Figure 4 on page 38 for potential interventions).

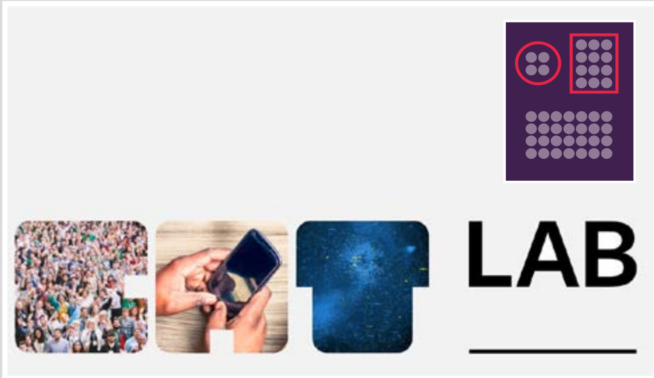
Light-touch auditing of AI performance already occurs in certain AI and CI project pipelines, for example if models are deployed for immediate use by communities in real-world contexts. By setting up a feedback loop from participants, models can continuously update their functions and iteratively improve to better serve the user's needs. For example, iNaturalist, the social network used by nature lovers to learn about, discuss and share images of animals, has integrated an AI model that uses computer vision to help community members accurately classify each animal sighting they enter into the database. The classifications suggested by the model are more accurate for some species than others. When users come across an error in the algorithm's classification, they can report it and assign the correct tag. This feedback loop ensures continuous collective oversight of the algorithm's performance.

The next step is developing more sophisticated approaches to 'collaborative governance' of intelligent machines. CI could play a role throughout the AI pipeline, using active group deliberation to question assumptions and reach agreement about the use and performance of AI. This vision of collectively intelligent governance could take place on many levels, from legally binding multi-stakeholder public–private partnerships where organisations hold each other accountable, to the distributed moderation practices used by online communities. Importantly, these new models of CI-inspired governance to audit and monitor AI need to involve communities that are likely to be affected by the use of AI. CAT Lab (see page 44) uses citizen social science to examine the impacts of emerging technology (including AI) on online communities. In one research project, CAT Lab worked with an active community on Reddit[58] to learn about how communities can mitigate against the negative behaviours fuelled by the platform's algorithms. The research, analysis and interpretation was carried out in collaboration with the communities. Towards Democratic Auditing is a project run by the Data Justice Lab in Cardiff that aims to develop processes for groups to monitor and take action on the use of automated scoring and decision-making systems by public sector organisations.

Auditing practices are also emerging among teams of domain experts in the context of automated professional DSTs. For example, social service workers responsible for child welfare risk assessments in Douglas County in the US have developed a new process (called the Red Team) where they undertake a team-based interpretation of algorithmic risk scores to guide their decision-making.[59]

Case study: CAT Lab

Using CI to mitigate against the negative impacts of AI on online communities



The challenge

Almost all online social networks make use of AI within their platforms, yet the full extent of the algorithms' impact on the communities that use the platforms are still rarely rigorously evaluated in real-world settings. There is evidence to show that negative behaviours of users on popular social networks like Reddit and YouTube are amplified by the use of AI,[1] which can automatically promote the most attention-grabbing or shocking content. Experiments into these trends tend not to be co-designed with members of online communities, even though these individuals hold unique insights about how the platform is used and are important for driving behavioural change that can improve users' experience of online environments.

The AI and CI solution

The mission of the Citizens and Technology Lab (CAT Lab)[1] at Cornell University is to 'work with communities to study the effects of technology on society and test ideas for changing digital spaces to better serve the public interest'. By involving existing online communities in the research, the group can ensure that they pose research questions that are the most relevant to the real-world setting.

Using a collaborative citizen science model, the academics work together with an online crowd to set research questions and design experiment protocols. In 2016, CAT Lab ran a study to test the effect of different community moderation practices on mitigating against harassment and combatting the spread of fake news, in forums hosted by the Reddit platform.

So what?

The research showed that online communities can help to consolidate positive social norms through simple actions, such as regular reposting of community rules in discussions. These interventions increased rule compliance by 8 per cent and significantly increased contributions by new users, both of which can serve as a counterpoint to the negative content promoted by AI. Designing and testing interventions with the communities most affected helps to instil more positive behaviour on social platforms and empowers communities to take part in the governance and monitoring of new technology.

The CAT Lab uses CI to evaluate the impact of digital technologies, including AI, on social interactions online. Since its launch in 2014, the group has completed five studies on some of the most popular online platforms, including Facebook, Reddit and Twitter.

Diversity for better AI

Research on CI has shown time and again that a group's diversity affects how well it is able to solve problems. This poses a significant challenge for AI, where diversity is notoriously lacking. In 2019, an analysis of the participation of women in AI found that of the 1.3 million articles published about AI only 13.8 per cent included a female author. The study[60] also showed that women working in physics, education, computer ethics and other societal issues, and biology, were more likely to publish work on AI in comparison with those working in computer science or mathematics. Papers with at least one female co-author tended to be more focused on real world applications and used terms that highlighted the social dimensions of the research, such as fairness, human mobility, mental health and gender.

This illustrates an important point about widening participation in technology development and how a lack of diversity affects the kinds of questions that we apply AI to. The background and experiences of developers affect the way they frame problems or even how they code algorithms, as these may be based on the assumptions they hold. We stand to gain by taking active measures to embed diversity at all stages of the AI lifecycle, from the collection and validation of the data used to train algorithms all the way through to their interpretation and use. Although adding even more 'messy' real-world contributions may start to slow things down, it will help to make sure that the development of AI does not slide into a trajectory determined by a narrow demographic group. AI can only help to enhance CI if we make deliberate and thoughtful choices to ensure that it reflects the differences we see in society.

09

Common pitfalls of AI and CI integration

Despite many promising opportunities in combining AI and CI, it can be easy to make mistakes when bringing together these two often contrasting methodologies.

Building on lessons from project failures and challenges overcome by successful initiatives, we outline some examples of how the integration of AI and CI can go wrong and the main design tensions anyone in this field should be aware of.

Big tech hubris

Some of the most common failures of AI integration stem from not adequately considering ongoing human interactions and group behaviour when deploying AI tools. One example is Google Flu Trends,[61] which was hailed as a success of search-query-scraping to predict flu prevalence, before it emerged that it was vulnerable to overfitting[62] and changes in search behaviour.[63] Integrating real-time public health data into the predictions or involving health professionals in the tool's development might have circumvented some of these shortcomings.

Even well-intentioned projects like the Detox tool developed by Google Jigsaw and Wikipedia, which used crowdsourcing and machine-learning to identify toxic comments, may only be effective for short periods until 'bad actors' figure out how to counteract them.[64] The vulnerability to such 'gaming' is a common feature of automated methods when they are not updated frequently enough to remain sensitive to dynamic real-world contexts and the potential for negative behaviour from human users. When Microsoft launched a public chatbot called Tay in 2016 as part of their research into conversational AI, they failed to anticipate that some Twitter users would teach the AI agent to make racist comments.[65] Carrying out regular assessments of vulnerabilities and impacts of adversarial actors in any social network is vital to preventing such cases of misuse before they occur or affect too many individuals.

Prioritising marketing over deliberation on platforms

Some social platforms have been criticised for encouraging the spreading of negative comments or amplifying bad behaviour by using algorithms that optimise for features like click-through rates.[66] For example, the popular discussion forum platforms 4chan/8chan have been widely criticised for encouraging, supporting and protecting hate-filled rhetoric, and researchers have claimed that algorithms on Reddit are tuned to incentivise bad behaviour.[67]

Although some of these features are adjustable, changing them still relies on moderator preferences and an active decision by online communities to opt out of default settings. Experiments by CAT Lab (see page 44), have shown that community behaviours such as regularly reposting community rules can help to counteract some of the negative impacts of AI-enabled platform design by reinforcing social norms.

The importance of design features is similarly highlighted by the deliberation platform Polis,[68] where users contribute their ideas and opinions on a discussion topic. Participants rank each other's statements but are not able to directly post replies to any of the ideas. The absence of a comment feature was a deliberate choice by the platform's designers to help promote more open, consensus-driven debate among users. The platform's visualisations cluster similar opinions to help participants understand where their opinion falls in relation to others' and how they contribute to the formation of group consensus. Polis is notably used by the Taiwanese government to help identify areas of agreement between different groups as part of their citizen participation project, vTaiwan.[69]

Reliance on partnerships between organisations

Data collaboratives fall into the category of CI methods that describe an arrangement to contribute and share data between different parties. For many sectors, the delivery of public services is split between multiple organisations, teams or stakeholder groups so, to make the most of the AI opportunity, these parties need to come together to share data or code. However, the deployment of AI in these contexts can encounter long delays (or complete deadlock) due to clashes in institutional processes and values or difficulty in negotiating responsibilities and resources. For example, the New York City Fire Department, which long promised an enhanced AI-enabled version (Firecast 3.0) of its model for predicting fire risks, has faced many difficulties due to organisational culture.

New pressures on the social contract

There is a thin line between mobilising CI for collective benefit and exploiting users' data or manipulating crowds. The use of AI as a method of social surveillance (such as through facial recognition technology) has attracted criticism of the Chinese state and led to regulatory bans in parts of the US and Europe. Public sector organisations should navigate these debates with extra care and attention, given that they often lack the resources to develop their own models in-house or to adequately scrutinise commercial AI systems. As organisations in the public sector seek to make the most of the added value offered by AI in the face of continued low public trust in institutions, it will become increasingly important to re-examine and renegotiate the social contract with the communities they serve. For AI-enabled CI to flourish, it is necessary to build automated systems that use responsible data practices and foreground principles of collective benefit.

10

Design choices at the heart of AI for CI

Too often, debates about the development and use of AI are framed in terms of economic gains without due consideration of its wider impacts on values, such as equality and environmental justice.

There are a number of other significant tensions that are important to consider before venturing into the world of AI and CI design.

Very few of the examples we found had publicly available information about the costs and trade-offs they had considered during the design of their projects, making it difficult for others to learn from their experience. In this section, we explore six design challenges at the heart of AI and CI projects that innovators need to consider when applying human and machine intelligence to help solve social problems.

Optimising the process: project efficiency versus participant experience

Several examples of AI integration in CI projects use data that had been previously classified (labelled) by volunteers to train machine models. These AI models are then deployed within the project to perform the same tasks alongside volunteers. The Cochrane Crowd[70] project, which categorises clinical reviews, and the Snapshot Wisconsin project, which analyses camera-trap images on the Zooniverse platform, have both introduced machine classifiers in this way to identify the simplest cases among their large datasets, leaving the more unusual tasks to their volunteers.

Although this approach helps to advance project goals by completing tasks more efficiently, evaluations of both projects found that it risks disincentivising volunteers by making their tasks too hard[71] and too monotonous.[72] Apart from the potential damage to relationships with volunteer citizen scientists, it may also reduce broader social impacts. These include increased science literacy and behavioural change with regard to environmental issues, which are an important part of CI projects.

One of the great benefits of AI systems is their speed and ability to perform consistently. In contrast, large-scale CI participatory methods can require more time to acknowledge and interpret contrasting values and viewpoints. Projects need to decide on the optimal balance between AI and CI and the different functions they provide depending on the issue at hand. For example, when citizens are brought together to deliberate on contentious or complex issues, as is the case on digital democracy platforms, they may choose to optimise the methods used for transparency and inclusiveness, which may in fact slow down the process. In other contexts, such as anticipating and responding to humanitarian emergencies or disease outbreaks, the high speed of AI methods may be prioritised at the expense of involving a wider array of stakeholders.

Developing tools that are 'good enough'

In AI development, the primary focus is often on developing the best possible technology by, for example, creating the most accurate prediction model. However, developing something that is just 'good enough' can often lead to better uptake and value for money.

Whether choosing between passive or active contributions from the crowd, classical machine-learning or more cutting-edge techniques, you should be guided by the principle of 'fitness for purpose'.[73] This metric will be more or less focused on accuracy, depending on whether AI is being used as a triage mechanism or to issue assessments in high-stakes scenarios, for example making diagnostic assessments. In reality, gains in accuracy are sometimes minimal in comparison to simple models[74] or even traditional statistics.[75]

Some of the most cutting-edge AI methods, such as deep learning,[76] are developed in laboratory settings or industry contexts, where a detailed understanding of how the algorithm works is less of a priority. CI initiatives that involve or affect members of the public carry a higher risk of widespread impact when things go wrong, for example when projects address complex social issues such as diagnostics in healthcare or management of crisis response. This places a higher burden of responsibility on CI projects to ensure that the AI tools they use are well understood.[77]

Some companies that initially develop models using uninterpretable methods can be forced to discard them in favour of simple machine-learning models to meet the stricter accountability norms imposed by working with the public sector or civil society. One example is NamSor, an AI company working on accurate gender classifications who switched to working with simpler, explainable models in order to work with a university partner.[78] Testing new methods in the more realistic and complex virtual environments, such as Project Malmo from Microsoft,[79] and formalising standards of transparency for public sector AI can help to overcome this gap between performance in development and the real world.

Managing the practical costs of running AI and CI projects

Developing CI projects can be costly. Weighing up the ambition to involve the widest group of people and use cutting-edge AI tools will always need to be balanced against the significant costs of doing this. The resource demands will vary depending on the size of the group involved, the depth of engagement and duration.

The integration of AI-data-centric methods adds considerably to these costs through the need to access specialised expertise, additional computational demands and maintain data storage. Some of the costs can be shouldered by private sector tech partners. Examples of this kind of sponsorship include the iNaturalist Computer Vision Challenge, which was supported by Google, and the Open AI Challenges model developed by WeRobotics (see page 42), which are typically financed by a consortium of public and private sector organisations. Another example is the Chinese AI company iCarbonX, which partnered with the peer-to-peer patient network PatientsLikeMe. In 2016, iCarbonX established the Digital Life Alliance,[80] a collaborative data arrangement to accelerate progress on AI systems that can use biological data in combination with the lived experience of patients to support decision making in health for patients and medical practitioners alike.

Recent changes to the use of AI in practice, such as better documentation of training data and new tools for detecting dataset drift[81] or describing model limitations,[82] may help smaller organisations navigate the trade-offs of available technology. However, the extra administration required to implement these practices could create a resourcing burden that larger institutions are able to absorb while smaller organisations, charities and community groups struggle. This may mean that few will be able to take advantage of these methods.

Wider societal costs of popular AI methods

Alongside technical challenges and impact on volunteers, any project working at the intersection of AI and CI needs to take into consideration the potential wider societal cost of its use. A common critique of AI is the strain that its development and deployment can place on the right to decent work,[83] environmental resources and the distribution of wealth and benefits in society. This stands at odds with the problems that we turn to technology to solve, such as social inequality and the climate crisis. In fact, some of the latest AI methods are so computationally demanding that the environmental impact of training them has been estimated as equivalent to the lifetime carbon footprint of five average cars.[84]

Studies into the use of AI systems in areas such as criminal justice and social welfare have shown that smart machines can reinforce historical biases and uphold political value systems.[85] CI, when done well, offers a potential counterpoint to this challenge by opening up public problem solving to citizens and under-represented groups. Unless these issues are targeted from the outset, there is a risk that participatory methods become a vehicle for consolidating existing power hierarchies.

Anatomy of an AI system

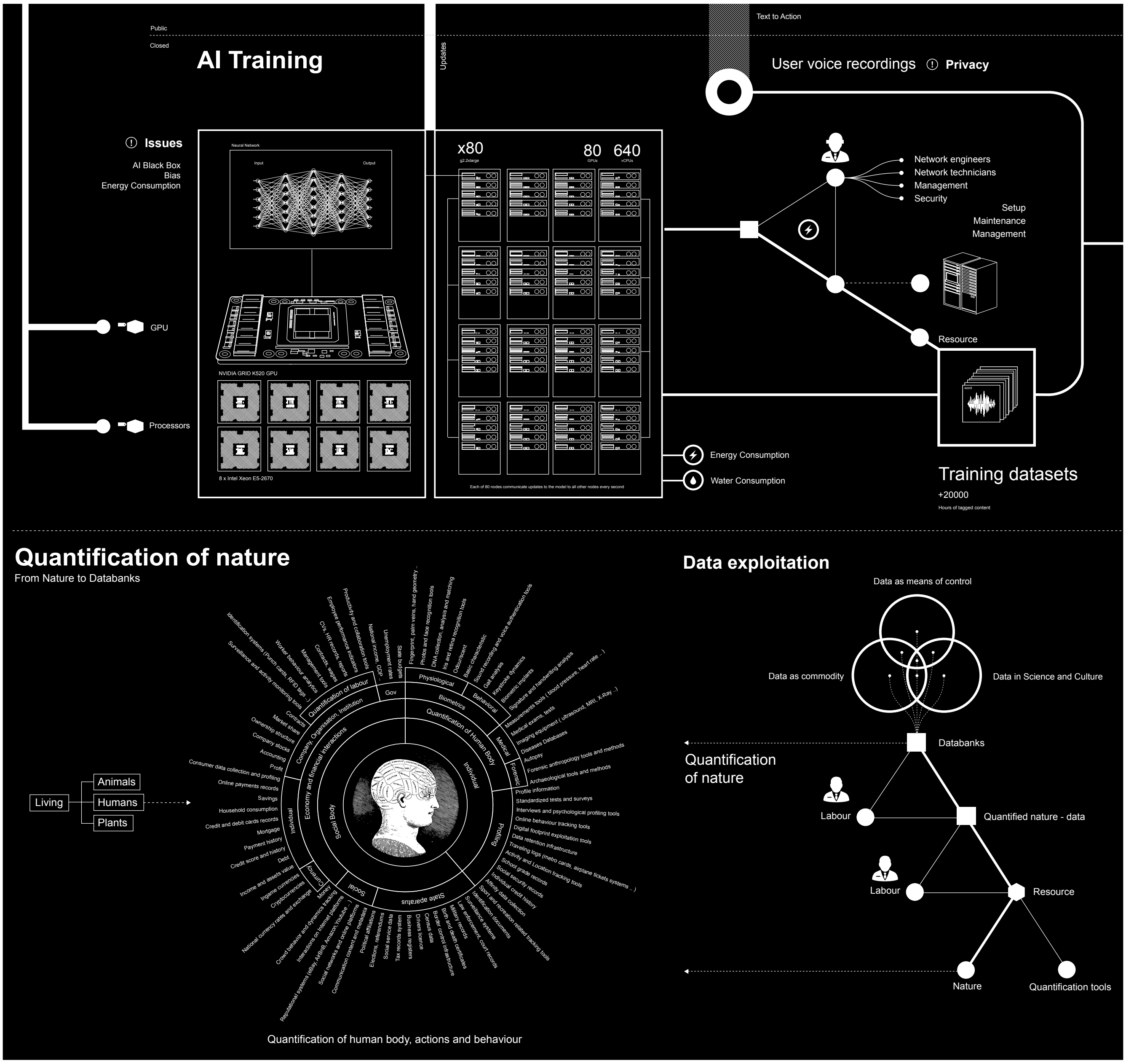


Image credit: Kate Crawford and Vladan Joler (2018)

The 'Anatomy of an AI system' is a detailed mapping of the crowd labour behind the production of an Amazon Echo, developed by Kate Crawford and Vladan Joler. It raises important questions about the right to fulfilling work and acceptable standards for crowd work, as well as highlighting unsustainable resource demands of the supply chain. Transparency about these different trade-offs can help guide more informed decisions and stimulate public debate about the normative values that should govern the design of solutions to public sector problems.

Open data, open-source tools versus vulnerability to misuse

Open data and software are key enablers for CI initiatives. Opening up datasets and involving more individuals in the development of solutions can help to stimulate innovation by generating more creative ideas or spreading the use of existing solutions to new contexts. Publicly available data can be used to develop entirely new CI projects[86] and open up markets to new providers. Precision agriculture company OneSoil (see page 29), for example, applies AI to high-quality free and open-source satellite images in order to map field boundaries, crop types and plant health on farmland in near real time to support farmers' decision making. In some cases, particularly when CI is used for participatory democracy, open data and software provides transparency to ensure the verifiability of public processes. As an example, the entire collective decision making process for the participatory democracy project vTaiwan is conducted using open-source tools.[87]

However, when data is contributed through passive or active crowdsourcing, it may contain sensitive information that is vulnerable to misuse. If placed in the wrong hands, personally identifiable features of data collected in humanitarian contexts or measurements that reveal private health information might create issues for contributing individuals.[88] The use of passive citizen contributions for AI and CI initiatives in the public sector (e.g. mobile geolocations, conversations on social media, voice calls to public radio) may require a recalibration of the social contract to decide what we deem justifiable in the name of public good and collective benefit.

Cultivating trust and new ways of working with machines

At the heart of AI and CI integration sits the need for a high degree of trust between groups of humans and the machine(s) they work with. In comparison with the resources allocated for enhancing AI systems, relatively little attention has been paid to investigating public attitudes to AI performing a variety of tasks (apart from decision-making). As a result there is a limited understanding of trust and acceptability of using AI in a variety of settings, including the public sector. In *How Humans Judge Machines*,[89] Cesar Hidalgo and colleagues describe how they use large-scale online experiments to investigate how people assign responsibility depending on whether people or machines are involved in decision-making. The participants in the experiments differentially judged human decision-making by the intention, whereas the decisions made using automated methods were assessed by the harm/impact of the outcome.

When it comes to decision-making as a group, there is an additional layer of complexity around whether individuals feel they carry responsibility for collective decisions.[90] It's possible to imagine situations where artificial agents help to increase a feeling of shared ownership over consensus decisions but to take full advantage of these possibilities we need to better understand the variety of individuals' attitudes to AI. In high-stakes decision-making, for example when automated decision support tools (DSTs) are introduced to guide the work of social workers or judges, they encounter a wide range of reactions from individual team members, such as ignoring or relying too heavily on the machine's recommendation.[91] Sometimes automation bias (a tendency to place excessive faith in a machine's ability) can cause professionals to dismiss their own expertise or use it as a source of validation.[92] We are only at the beginning of unpicking these complex interactions between people and AI and how well they are explained by our existing theories of trust and responsibility.

11

Recommendations

This paper has illustrated that the AI and CI field, while still nascent, holds significant opportunities for using technology to solve some of our most complex social challenges and how we can better involve people in shaping the future trajectory of AI development.

However, making the most of this opportunity requires a significant shift in how we think about AI policy, R&D and developing partnerships across different sectors and disciplines.

More than £1 billion has been invested by the UK government into the AI Industrial Strategy, and more than £800 million was raised by AI companies in the first half of 2019.[93] This testifies to the fact that AI is one of the most well-funded areas of research and development. In contrast, we only see a fraction of this funding for CI opportunities. Funds to support the overlap between AI and CI are even more rare. It is thus unsurprising that less than 15 per cent of large companies developing AI technology are actively working to manage risks associated with equity and fairness.[94]

The Nesta fund for CI experiments, delivered in partnership with the Omidyar Foundation, the Wellcome Trust and the Cloudera Foundation, is a rare example of a joint AI and CI fund in the UK. Other countries have shown more foresight. The European Commission plans to launch a new €6 million fund[95] for applied research in this area in 2020. In the US, both the Intelligence Advanced Research Projects Activity[96] and the Defense Advanced Research Projects Agency[97] have funded major projects at the intersection of AI and CI. The multibillion investment in AI research and development in China is also well known, whereas experiments in AI and CI in Chinese cities receive less attention.[98]

The UK has recognised AI as a significant policy priority through the Industrial Strategy Grand Challenges and the AI Sector Deal as well as through the creation of the Office For AI. But there is a risk that setting a trajectory determined by industry priorities rather than viewing AI through a CI lens will lead to public backlash and result in a missed opportunity to use AI and CI to both improve the technology and help enhance collective capacity to solve social problems.

Below we present recommendations for policymakers, funders and researchers on how they can make the most of the AI and CI opportunity.

Policymakers

Put collective intelligence at the core of all AI policy in the United Kingdom. The UK Government should adapt AI policy to reflect the widest possible collective benefits from AI applications as well as emphasising diversity and broad participation in the development of AI. Specifically:

- The Office for AI, BEIS and DCMS should promote methods where AI enables citizen innovation and collective action to make progress on the Industrial Strategy missions, particularly ageing and mobility, and to ensure the economic benefits of the AI Sector Deal are balanced with social benefit and application of AI in the public interest.
- The AI Council should use emerging AI and CI models and use cases to inform their work on data, narratives and skills of AI and to reframe the UK's AI opportunity as one that is focused on inclusion and augmenting human intelligence to solve social challenges.
- The upcoming National Data Strategy consultation is an opportunity to use participatory methods to shape future data policy, and to ensure that future AI and CI initiatives are supported by robust data infrastructure and practices.
- The Centre for Data Ethics and Innovation should use its planned State of the Nation report to highlight AI and CI practices and commission a feasibility study for different AI and CI models in the UK public sector context as part of its future work programme.
- The Office for AI and Government Digital Service should update their guidance on using AI in the public sector[99] to include specifications for AI and CI. All procurement of AI tools by central government departments should require vendors to demonstrate CI principles or methods in the development and implementation of tools. The expected guidance on *Social value in government procurement*[100] should reflect this commitment.

Create testbeds for experimentation to accelerate learning. So far, experimentation in AI and CI has been ad hoc. Creating dedicated regional or sectoral testbeds allows for experimentation in real-world settings. This could help stimulate private and public sector collaboration and accelerate learning about best practice in AI and CI for public benefit.

Both local and central government should more clearly exploit the opportunities in using AI and CI to increase the quality and scale of existing methods for involving the public in developing policy and delivering public services. While this applies to most parts of public services, we see a specific opportunity for innovation in three areas:

- **Digital democracy**: Ensure that existing investment in new participatory democracy processes such as the local government citizen assemblies[101] is given an opportunity to create impact by incorporating AI to make better use of citizen's contributions.
- **Environment, energy and climate**: The climate crisis and reducing energy demands are two significant challenges where AI could amplify the impact of collective action. Existing citizen science initiatives on air quality, pollution and biodiversity should be integrated into DEFRA policies. Using AI within these projects could help citizen-generated data to achieve maximum impact. Alongside this, Ofgem should work with citizens and industry to explore AI models of different scenarios for future decarbonisation.
- **Healthcare and wellbeing**: There is a growing need to develop a sophisticated understanding of the socioeconomic determinants of health and to empower citizens to take action to improve their mental and physical wellbeing. NHSx and the NHS AI Laboratory should work with patient and community groups to identify opportunities for developing AI that combines multimodal analysis of health data and non-traditional data sources like open data on GP prescriptions, with lived experience.

Funders

The first major funder to put £10 million into this field will make a lasting impact on the future trajectory for AI and create new opportunities for stimulating economic growth as well as more responsible and democratic AI development.

Launch a new dedicated funding programme for AI and CI research and development. There are currently no large-scale funding opportunities in the UK for AI and CI research and development. This gap could be filled by UKRI and foundations dedicated to solving societal challenges, like the Wellcome Trust, Open Society Foundations and Luminate. Public funders should also focus on integrating AI and CI opportunities into existing AI funding programmes and commit to supporting a significant proportion (at least 20 per cent) of all AI projects that explicitly focus on the involvement of people and/or societal impact. Foundations have a specific opportunity to shift current AI4Good funding by foundations towards a clearer focus on how AI can empower collectives and ensure long-term societal impact.

Invest in new partnerships and governance models for AI and CI experiments. The relative disconnect between the fields of AI and CI along with failures of systems-level co-ordination and governance threaten the success of AI and CI projects.

- Funders need to incentivise interdisciplinary collaboration between the fields of AI and CI by making funding criteria contingent on a partnership approach. The Office for Civil Society, the Knowledge Transfer Network and the Catapults should help to broker public–private partnerships across different sectors.
- Independent organisations like the Ada Lovelace Institute and the Open Data Institute, as well as the Office for AI, should provide guidance on new models of data trusts and oversee public auditing of AI used in the public sector. This will help to ensure the responsible development of AI in the public interest.

Research and practitioners

Looking beyond the research questions raised by the report, there are a number of systemic interventions that are necessary to ensure the continued growth of the field. The academic institutions and technology companies that change the emphasis of their AI research and development programmes towards collective intelligence and applications of AI in the public interest will have 'first mover' advantage and be recognised as global leaders in this emerging field.

Build a new interdisciplinary field and link to real-world practice. The field of AI and CI covers a broad range of subdisciplines in which the UK research community is recognised as one of the global leaders, such as AI and citizen science. However, the UK has no academic institution or discipline dedicated specifically to understanding the field of AI and CI, which limits our understanding of current and future opportunities.

- Institutions working across the relevant fields, such as UCL, the Oxford Internet Institute and the Alan Turing Institute, could advance this agenda through dedicated research programmes. This could build on international lessons from similar initiatives such as the MIT Centre for Collective Intelligence in the US and the UM6P School of Collective Intelligence in Morocco.
- Progress in the field could be further advanced through the creation of a dedicated international academic journal which strengthens the links between different research fields and practice-based CI.

Accelerate progress on AI and CI research by committing to open science and evaluation.

Currently, significant resources are being wasted and efforts duplicated due to the lack of access to existing knowledge and solutions, such as data and software. Existing AI and CI projects do not put enough resources into evaluation and sharing of lessons learnt, which risks the repetition of mistakes within the projects and by others in the field. All researchers and practitioners working in AI and CI should:

- Apply the FAIR principles[102] to data management and follow emerging guidance on sharing data and code to encourage transparency and ensure reproducibility.
- Openly publish on feasibility studies and costs associated with AI and CI solutions.
- Develop new criteria for AI and CI design and new benchmarks to measure performance and impact evaluation (see 'Practice – getting the design of AI and CI projects right').

Practice – getting the design of AI and CI projects right

The field can only evolve through more organisations experimenting with different models of AI and CI and the opportunity to deliver novel solutions to real-world challenges. However, beyond 'just' calling for more experimentation by practitioners with these new methods, we put forward the following criteria that should be considered in any AI and CI project.*

These questions are intended to guide more in-depth consideration of the integration of AI and CI. Practitioners should use them to help plan their projects and as a starting point for project evaluation. The first and last question should always be: Does/did this project really need AI?*

- **The problem**: What issue are you working on? What other methods exist to answer the same question(s)? What are the limitations of current approaches, and can these be addressed by the integration of AI and CI?
- **Performance of AI**: What is the algorithm optimising for? What existing metrics can be used to continuously monitor AI performance, and what additional criteria are needed to measure the impact on CI initiatives?
- **Social acceptance of AI**: Are the participants in the project aware of the use of AI? Will they be consulted about its deployment? Do the participants have a choice to opt out of AI-enabled functions?
- **Transparency of the algorithm**: To what extent is the model interpretable? Is the training data available? Is it possible to verify biases and track model or data set drift?
- **Ability to achieve collective goals**: Does AI enhance the CI initiative's progress towards understanding problems, seeking solutions, making decisions or collective learning? What baseline can this improvement be measured against?
- **Quality of participation**: Is AI enhancing the quality of participation? How does AI change the participatory process (e.g. facilitation, reduction of bias, surfacing new information)?
- **Level of interaction between the crowd and AI**: What different models of AI and group interaction are most relevant to this project? How serious is the context? What are the risks associated with more/less autonomous implementation of AI?
- **Resources and sustainability**: What costs are associated with the project in the short and long terms? How much money will be required for data acquisition, storage and analysis, technical support and community engagement? What is the environmental impact?
- **Partnerships**: Does the project require partnerships between the non-profit, public and private sectors? What governance processes will be followed? What is the value proposition for each side?

*For those new to CI, we recommend Nesta's Collective Intelligence Design Playbook, which features design questions and resources for project development from problem definition to identification of real-world impacts.

12

Bibliography

Aggarwal, I., and Woolley, A. W. (2018). Team Creativity, Cognition, and Cognitive Style Diversity. Management Science. https://doi.org/10.1287/mnsc.2017.3001

Aggarwal, I., Woolley, A. W., Chabris, C. F., and Malone, T. W. (2019). The Impact of Cognitive Style Diversity on Implicit Learning in Teams. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.00112

Bazazi, S., Zimmermann, J. von, Bahrami, B., and Richardson, D. (2019). Self-serving incentives impair collective decisions by increasing conformity. PLOS ONE, 14(11), e0224725. https://doi.org/10.1371/journal.pone.0224725

Benkler, Y., Shaw, A., and Mako Hill, B. Peer Production: A Form of Collective Intelligence. in Malone, T. W., and Bernstein, M. S. (2015). Handbook of Collective Intelligence. MIT Press.

Bernstein, E., Shore, J., and Lazer, D. (2018). How intermittent breaks in interaction improve collective intelligence. Proceedings of the National Academy of Sciences, 115(35), 8734–8739. https://doi.org/10.1073/pnas.1802407115

Boden, M. A. (1998). Creativity and artificial intelligence. Artificial Intelligence, 103(1), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Cherna, M., et al. (2019). Child Welfare Decision-Making Using Predictive Risk Modeling: Two Counties Experiences. https://imgsvr.eventrebels.com/ERImg/02/41/36/4304045/141438-2-48157.pdf. Accessed January 2020.

Denton, E. Retrospectives Session 3: NeurIPS (2019). https://slideslive.com/38921896/retrospectives-a-venue-for-selfreflection-in-ml-research-3. Accessed December 2019.

Diaz L.V.M., Gaytan-Lugo L.S., Rocha M.A.M., Santis A.C. (2013), Crowd-Computer Interaction, A Topic in Need of a Model. In: Collazos C., Liborio A., Rusu C. (eds) Human Computer Interaction. CLIHC 2013. Lecture Notes in Computer Science, vol 8278, Springer, Cham.

El Zein, M., Bahrami, B., and Hertwig, R. (2019). Shared responsibility in collective decisions. Nature Human Behaviour, 3(6), 554–559. https://doi.org/10.1038/s41562-019-0596-4

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press, Inc.

Geoghegan, H., Dyke, A., Pateman, R., West, S. and Everett, G. (2016) Understanding motivations for citizen science. Final report on behalf of UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England.

GLUE benchmark. Available online at https://gluebenchmark.com/leaderboard. Accessed December 2019

Gray, J. (2019). Data witnessing: Attending to injustice with data in Amnesty International's Decoders project. Information, Communication and Society, 22(7), 971–991. https://doi.org/10.1080/1369118X.2019.1573915

Gray, M. L., and Suri, S. (2019). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Houghton Mifflin Harcourt.

Green, B. 'Good' isn't good enough,' in NeurIPS Joint Workshop on AI for Social Good. 2019. https://www.benzevgreen.com/wp-content/uploads/2019/11/19-ai4sg.pdf. Accessed January 2020.

Hernández-Orallo, J., and Vold, K. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 507–513. https://doi.org/10.1145/3306618.3314238

Hidalgo, C., Orghian, D., Albo-Canals, J., de Almeida, F., Martin, N., (2020, forthcoming) How Humans Judge Machines.

Hillen, M. A., Gutheil, C. M., Strout, T. D., Smets, E. M. A., and Han, P. K. J. (2017). Tolerance of uncertainty: Conceptual analysis, integrative model, and implications for healthcare. Social Science and Medicine, 180, 62–75. https://doi.org/10.1016/j.socscimed.2017.03.024

Hong, L., and Page, S. (2012). Some Microfoundations of Collective Wisdom. In H. Landemore and J. Elster (Eds.), Collective Wisdom: Principles and Mechanisms (pp. 56-71). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511846427.004

Lakeman-Fraser, P., Gosling, L., Moffat, A.J. et al. To have your citizen science cake and eat it? Delivering research and outreach through Open Air Laboratories (OPAL). BMC Ecol 16, 16 (2016). https://doi.org/10.1186/s12898-016-0065-0

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. Science, 343(6176), 1203. https://doi.org/10.1126/science.1248506

Makridakis S, Spiliotis E, Assimakopoulos V (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 13(3): e0194889. https://doi.org/ 10.1371/journal.pone.0194889

Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. New Media and Society, 19(3), 329–346. https://doi.org/10.1177/1461444815608807

Mayo, A. T., and Woolley, A. W. (2016). Teamwork in Health Care: Maximizing Collective Intelligence via Inclusive Collaboration and Open Communication. AMA Journal of Ethics, 18(9), 933–940. https://doi.org/10.1001/journalofethics.2016.18.9.stas2-1609

Mellers, B. A., Baker, J. D., Chen, E., Mandel, D. R., and Tetlock, P. E. (2017). How generalizable is good judgment? A multi-task, multi-benchmark study. Judgment and Decision Making, 12(4), 369–381.

Ong, W. J. (2013). Orality and literacy. Routledge.

Rabanser, S., Günnemann, S., and Lipton, Z. C. (2019). Failing Loudly: An Empirical Study of Methods for Detecting Dataset Shift. ArXiv:1810.11953 [Cs, Stat]. http://arxiv.org/abs/1810.11953

Richardson, K. (2002). What IQ Tests Test. Theory and Psychology, 12(3), 283–314. https://doi.org/10.1177/0959354302012003012

Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. ArXiv:1811.10154 [Cs, Stat]. http://arxiv.org/abs/1811.10154

Russell, S., Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Prentice Hall.

Schmidt, F. (2019). Forschungsförderung Working Paper—Detailseite. (n.d.). Hans Böckler Stiftung. Retrieved February 13, 2020, from https://www.boeckler.de/64509.htm

Shirado, H., and Christakis, N. A. (2017). Locally noisy autonomous agents improve global human coordination in network experiments. Nature, 545(7654), 370–374. https://doi.org/10.1038/nature22332

Stathoulopoulos, K., and Mateos-Garcia, J. C. (2019). Gender Diversity in AI Research. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3428240

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. ArXiv:1906.02243 [Cs]. http://arxiv.org/abs/1906.02243

Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., Glasziou, P., Shemilt, I., Synnot, A., Turner, T., Elliott, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., Albrecht, L., Horsley, T., Pearson, L. (2017). Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology, 91, 31–37. https://doi.org/10.1016/j.jclinepi.2017.08.011

Trouille, L., Lintott, C. J., and Fortson, L. F. (2019). Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. Proceedings of the National Academy of Sciences, 116(6), 1902–1909. https://doi.org/10.1073/pnas.1807190116

Weld, D. et al, Artificial intelligence and collective intelligence. in Malone, T. W., and Bernstein, M. S. (2015). Handbook of Collective Intelligence. MIT Press.

Wilson, Linus, Clickbait Works! The Secret to Getting Views with the YouTube Algorithm (2019). Available at SSRN: https://ssrn.com/abstract=3369353 or http://dx.doi.org/10.2139/ssrn.3369353

Vaccaro, M., and Waldo, M. V., J., The Effects of Mixing Machine Learning and Human Judgment. Retrieved February 13, 2020, from https://cacm.acm.org/magazines/2019/11/240386-the-effects-of-mixing-machine-learning-and-human-judgment/abstract

Vaughan, J. W. (2018). Making Better Use of the Crowd: How Crowdsourcing Can Advance Machine Learning Research. Journal of Machine Learning Research, 18(193), 1–46.

Young and Zahurenic (2019). Identifying Citizens' Needs by Combining Artificial Intelligence and Collective Intelligence. New York University Governance Lab. Retrieved from http://thegovlab.org/new-report-identifying-citizens-needs-by-combining-artificial-intelligence-ai-and-collective-intelligence-ci

Endnotes

1. Terms ‘peak of inflated expectation’ and ‘trough of disillusionment’ taken from the Gartner Hype Cycle.
2. The research has been informed by an analysis of 150 existing CI projects that make use of AI, desk research, literature review and interviews with experts working across both fields.
3. The report touches on current debates in fairness and disparate impact but sets aside the detailed examination of AI Ethics, which is beyond scope. Instead, we recommend the publications being produced by the Berkman Klein Center, the Oxford Internet Institute, Data & Society, the AI Now Institute, the FAT ML (Fairness, Accountability, and Transparency in Machine Learning) community and countless others who give these issues the attention and critical discussion that they deserve.
4. Diaz L.V.M et al (2013).
5. Weld, D. et al (2015).
6. Young, A. and Zahurenic, A. (2019).
7. Russell, S. and Norvig, P. (2010).
8. https://public.oed.com/history/ Accessed January 2020.
9. https://stats.wikimedia.org// Accessed January 2020.
10. https://www.nesta.org.uk/blog/mapping-collective-intelligence-research/ Accessed January 2020.
11. According to Dave Snowden’s Cynefin® framework for problem-solving, problems fall into four distinct categories: simple, complex, complicated and chaotic.
12. It has been suggested that there are three sources of uncertainty: probability, ambiguity and complexity. Hillen et al (2017).
13. https://dt.asu.edu/ Accessed January 2020.
14. Apart from transfer learning paradigms.
15. More recently, the existence of the ‘ground truth’ principle has been questioned – alongside claims to neutrality – by the data science/AI communities. (see for example Denton, E. (2019).
16. https://www.epistemonikos.org/en/about_us/methods. Accessed January 2020.
17. The assumed match between the distributions of training and real-world data is known as the IID principle (independent and identically distributed data).
18. A weak signal is any indicator of an emerging issue, that may become significant in the future.
19. Gray, J. (2019).
20. Proposed by Alan Turing in 1950, the test Can Machines Think? evaluates a machine’s capacity to convince a human that they are interacting with another human.
21. For example, GLUE battery of language tasks for natural language processing methods.
22. Richardson, K. (2002).
23. Benkler, Y. Shaw, A. and Mako Hill, B. (2016).
24. nesta.org.uk/cidplaybook
25. Aggarwal, I. and Woolley, A. (2018).
26. Aggarwal, I. et al (2019).
27. Wisdom of the crowds refers to the idea that, on average, a large random group are collectively more accurate at making estimates or more effective at problem solving than any one individual.
28. Hong, L. and Page, S. (2012).
29. Bernstein, E., Shore, J., and Lazer, D. (2018).
30. Shirado, H., and Christakis, N. A. (2017).
31. Known as heuristics.
32. Mayo, A. and Woolley, A (2016).
33. Vold, K. V., and Hernandez-Orallo, J. (2019).
34. Ong, W. J. (2013).

35. This term was first introduced by Margaret Bowden. Bowden, M. (1998).

36. We found an initial longlist of over 150 CI projects using AI. Our final analysis used approximately 50 of these, which captured a range of CI methods across different sectors.

37. https://www.waze.com/en-GB/about Accessed January 2020.

38. Mellers, B. A., Baker, J. D., Chen, E., Mandel, D. R., and Tetlock, P. E. (2017). How generalizable is good judgment? A multi-task, multi-benchmark study. Judgment and Decision Making, 12(4), 369–381.

39. https://www.autodesk.com/solutions/generative-design Accessed January 2020.

40. https://www.dwell.com/article/will-algorithms-be-the-new-architects-095c9d41 Accessed January 2020.

41. This is known as supervised machine-learning.

42. https://www.mturk.com/ Accessed December 2019.

43. https://www.figure-eight.com/ Accessed December 2019.

44. https://www.google.com/recaptcha/intro/v3.html Accessed December 2019.

45. Gray, M. and Suri, S. (2019).

46. Schmidt, F. (2019)

47. Bazazi, S., et al. (2019).

48. Geoghegan, H., (2016)

49. See Wortmann-Vaughan, J. (2018) for a thorough review.

50. https://www.masakhane.io/ Accessed January 2020.

51. https://zindi.africa/competitions/ai4d-african-language-dataset-challenge

52. https://github.com/Niger-Volta-LTI

53. https://www.masakhane.io/#h.p_Mwmja7_sVZxf

54. https://www.robonet.wiki/ Accessed January 2020.

55. Benkler, Y. Shaw, A. and Mako Hill, B. (2016)

56. https://longitudeexplorer.challenges.org/ Accessed January 2020.

57. Eubanks, V. (2018)

58. Reddit is an online discussion platform with over 13 million subscribers. The study described in the text worked with the r/science community.

59. Cherna, M. et al (2019).

60. Stathoulopoulos, K. and Mateos-Garcia, J. (2019).

61. https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/ Accessed February 2020.

62. Overfitting is when models are tuned too precisely to the examples that are used to train them, meaning that they are less able to generalise the new inputs they encounter in the real world.

63. Lazer, D., et al (2014).

64. https://www.vice.com/en_us/article/aeyvxz/wikipedia-jigsaw-google-artificial-intelligence Accessed February 2020.

65. https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/ Accessed February 2020.

66. Wilson, L. (2019).

67. Massanari, A., (2017).

68. https://pol.is/home Accessed February 2020.

69. https://www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/ Accessed January 2020.

70. crowd.cochrane.org Accessed January 2020.

71. Thomas, J., (2017).

72. Trouille, L., Lintott, C. J., and Fortson, L. F. (2019).

73. Lakeman-Fraser, et al. (2016).

74. Rudin, C. (2019).

75. Makridakis S, Spiliotis E, Assimakopoulos V (2018).

76. Deep learning is an advanced subset of machine-learning techniques that learn from unstructured, high-dimensional and interconnected datasets.

77. This concern has been raised for deep learning models, see Rudin, C. (2019).

78. Personal communication from NamSor team.

79. www.microsoft.com/en-us/research/project/project-malmo

80. http://www.pmlive.com/blogs/digital_intelligence/archive/2017/january/patientslikeme_joins_artificial_intelligence_alliance_1183608 Accessed January 2020.

81. Rabanser, S., Günnemann, S., and Lipton, Z. C. (2019).

82. https://modelcards.withgoogle.com/about Accessed January 2020.

83. Gray, M. and Suri, S. (2019).

84. See Strubell, E., Ganesh, A., and McCallum, A. (2019). This estimate is based only on the fuel required to run the car over its lifetime, not manufacturing or other costs.

85. Green, B. (2019).

86. https://cseol.eu/about/ Accessed January 2020.

87. https://medium.com/@afhill/what-can-we-learn-from-the-civic-tech-community-in-taiwan-about-public-discourse-and-engagement-55626d4c1162 Accessed January 2020.

88. https://reallifemag.com/the-next-big-cheap/ Accessed January 2020.

89. Hidalgo, C. et al (2020).

90. El Zein, M., Bahrami, B., and Hertwig, R. (2019)

91. This is known as anchoring.

92. Vaccaro, M., and Waldo, J., (2019).

93. https://www.telegraph.co.uk/technology/2019/09/09/ai-investment-reaches-record-levels-uk/ Accessed February 2020.

94. https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf Accessed February 2020.

95. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-54-2020 Accessed February 2020.

96. https://www.iarpa.gov/index.php/research-programs/hfc?id=661 Accessed February 2020.

97. https://www.darpa.mil/program/2016-11-28 Accessed February 2020.

98. The Alibaba City Brain technology, which has been deployed in Shanghai, Suzhou and Hangzhou, uses AI powered by citizen-generated data to make forecasts and improve public service provision (while also raising concerns about the ethics of deploying, for example, facial recognition technology). See: https://www.cbinsights.com/research/china-artificial-intelligence-investment-startups-tech/ Accessed February 2020.

99. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/861204/A_guide_to_using_AI_in_the_public_sector__Web_.pdf Accessed February 2020.

100. https://www.gov.uk/government/consultations/social-value-in-government-procurement Accessed February 2020.

101. https://www.gov.uk/government/publications/innovation-in-democracy-programme-launch Accessed February 2020.

102. https://www.go-fair.org/fair-principles/ Accessed February 2020.

nesta

58 Victoria Embankment
London EC4Y 0DS

+44 (0)20 7438 2500

information@nesta.org.uk

@nesta_uk

www.facebook.com/nesta.uk

www.nesta.org.uk