Experiments in collective intelligence design for social impact
Second cohort: 2020-2021

Collective Intelligence Grants Programme
ACKNOWLEDGEMENTS

This report is based on the findings from 15 collective intelligence experiments that were co-funded (but not conducted) by Nesta, the Wellcome Trust, Omidyar Network and the Patrick J. McGovern Foundation.

The experiments were carried out by: Grzegorz Rutkiewicz, Patrycja Tempska, Maria Dowgiallo, Michal Wroczynski, Gniwosz Leliwa and Maciej Brochocki (Samurai Labs), Christian Ochieng, Ruslan Masinjila, Mohamed Alsayed and Cleopa Otieno (Kenya Flying Labs), Ulrike Hahn and Jason Burton (Centre for Cognition, Computation and Modelling, Birkbeck University of London), Maria Camps, Lydia Garcia, Lin Lin and David Bermejo (Spotlab), Johannes Mutter, Jeffrey Pickett, Helen van Baal, Alexander Davies and Mario Lyon (new by Augmented Thinking), Thomas Chesney (University of Nottingham), Swee-Hoon Chua (University of Tasmania), Robert Hoffman (RMIT University), Bo Percival, Felix Delattre and Ramya Ragupathy (Humanitarian OpenStreetMap Team), Jacapo Margutti (Netherlands Red Cross), Bastian Greshake Tzovaras and Mad Price Ball (Open Humans), Marc Santolini, Leo Blondel and Thomas Landrain (Just One Giant Lab), Alexa Hagerty, Alexandra Albert, Igor Rubinov and Joe McCraw (Dovetail Labs), Vito Trianni, Ralf Kurvers, Valentina Presutti, Andrea Giovanni Nuzzolese and Alessandro Russo (Institute for Cognitive Sciences and Technologies of the Italian National Research Council), Irving Lin (Human Dx), Robert Trigwell, Pritivi Hirani, Nuno Nunes, Eduardo Zambrano, Beakov Gabena, Elvina Zirima, Pablo Gallego, Tasha Shupert, Debora Gonzalez and Lorenzo Guadagno (International Organization for Migration), Alessia Paglialonga, Marco Zanet, Marco Muselli (Institute of Electronics, Computer and Telecommunication Engineering of the Italian National Research Council), Aziz Guergachi (Ryerson University), David Barber, Rebecca Theal, Rachael Morkem (Queen’s University, Canada), Karim Keshavjee (University of Toronto), Robert Kyba (Strategic Global Counsel, Toronto), Ling Tan, Usman Haque and Usamah Khan (Umbrellium), Iram Quasashi (Loop Labs), Franz Miralles (Swisscontact), Rafael Lindemann (RIMISP - Latin American Center for Rural Development), Carmen Tapia and Jesus Chumacero (Banco de Desarrollo Productivo), Martin Murillo (IEEE Humanitarian Activities Committee), Merihan Alhafnawi, Paul O’Dowd, Edmund Hunt, Severin Lemaignan and Sabine Hauert (University of Bristol) and other team members that might not be listed here.

We are very grateful to the following individuals at the organisations who co-funded this grant programme. Haidee Bell and Zaichen Wallace-Lu at the Wellcome Trust, Eshanthi Ranasinghe (formerly) at Omidyar Network, and Claudia Juech and Linda Tsai at the Patrick J. McGovern Foundation.

We would like to thank our colleagues at Nesta’s Centre for Collective Intelligence Design for their involvement in supporting and managing the grants programme over this period of time, specifically Joana Murta Rosa, Eva Grobbink, Camilla Bertocin, Aleksandra Berditchevskaia and Peter Baeck.

We would like to thank the members of our advisory group who help assess applications for the Collective Intelligence Grants. Julien Cornebise, Marc Driscoll, Louise Francis, Erik Johnston, Gitte Kragh, Chris Lintott, Massimo Menichinelli and Lucy Robinson. And the experts from other teams at Nesta who contributed to the assessment process, in particular Rob Fuller, Cath Sleeman, Juan Mateos-Garcia and Samuel Hanes.

We would also like to acknowledge the many hours of work contributed by participants, volunteers and citizen scientists without whom these experiments would not have been possible.

The articles in this report were written by Ian Steadman and the layout for this report was designed by Mike Green.

ABOUT NESTA

We are Nesta, the UK’s innovation agency for social good. We design, test and scale solutions to society’s biggest problems. Our three missions are to give every child a fair start, help people live healthy lives, and create a sustainable future where the economy works for both people and the planet.

For over 20 years, we have worked to support, encourage and inspire innovation. We work in three roles: as an innovation partner working with frontline organisations to design and test new solutions, as a venture builder supporting new and early stage businesses, and as a system shaper creating the conditions for innovation.

Harnessing the rigour of science and the creativity of design, we work relentlessly to change millions of lives for the better.

Find out more at nesta.org.uk

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

Published October 2021

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091. Registered as a charity in Scotland number SCO42833. Registered office: 58 Victoria Embankment, London, EC4Y 0DS.
Haidee Bell, Strategic Design and Innovation Lead, and Zaichen Mallace-Lu, Strategic Design and Innovation Manager

At Wellcome, we recognise the need for collective knowledge to understand and tackle the urgent global health challenges we target with our work: mental health, infectious disease and the health impacts of climate change. Our public participation work considers how community action and knowledge can augment the work of science and data, and we’ve been keen to learn from the field of collective intelligence to explore how different forms of expertise and intelligence can complement one another and in ways that are directed at action and impact.

We are investing in the programme to support and learn about collective intelligence in health and through research. We’re hopeful that this can help point to the value of the collective in enhancing science and allow us to learn about human-level effects, such as how collective intelligence prompts people to participate and to act.

Claudia Juech, Vice President of Data and Society

Patrick J. McGovern Foundation seeks to advance data use and AI to create a thriving, equitable, and sustainable future for all. The experiments supported by this programme demonstrate the value of combining communities’ deep knowledge of their needs and how to address them with sophisticated data analysis techniques. The portfolio of projects shows that community involvement adds benefits to every step of the data lifecycle (collection, analysis and use), across geographies, and among fields as varied as online behaviour and farming. Supporting Nesta’s Collective Intelligence Grants programme is an opportunity for philanthropies to join forces to seed more discoveries across a broad spectrum of use cases for more changemakers to build upon. Our hope for the future of the field is for collective intelligence to help guide AI to be more inclusive and representative of the populations that data-for-social-impact projects aim to serve, ensuring that data solutions not only will be transformative but more importantly, will be just.
Sonny Bardhan, Head of Strategy

Omidyar Network is a social change venture that reimagines critical systems, and the ideas that govern them, to build more inclusive and equitable societies. A key aspiration is to build a global technological ecosystem that reaches and works for everyone: one that balances innovation with responsibility.

Part of our approach also involves scanning for emergent signals and trends and conducting experiments on new avenues of change. So we were pleased to be able to invest in Nesta’s Collective Intelligence Grants programme to support the incubation of an emergent area of technology that seeks to combine human and machine intelligence to address social challenges.

Kathy Peach, Co-Director of the Centre for Collective Intelligence Design

People and machines bring different strengths and weaknesses to the table. Harnessing the complementary strengths of artificial intelligence and human collective intelligence is a powerful combination for solving social problems. Yet research and experimentation is needed for understanding how best to design and deliver projects that use human-machine cooperation.

Our partnership with the Wellcome Trust, Patrick J. McGovern Foundation, and Omidyar Network has enabled us to do just this, by helping to grow the emerging field of collective intelligence design through 15 ambitious experiments. We hope that these experiments may prompt other funders to do more too. There is currently a vast imbalance between the seemingly limitless funding for AI and the resources dedicated to collective intelligence. Yet the stakes could not be higher. Progressing collective intelligence is in many ways humanity’s grandest challenge since there’s little prospect of solving climate change, epidemics or conflict without progress in how we think and act together.

As the experiments funded by our grants programme show, collective intelligence has huge potential for society, and we hope that this will encourage more people to test this approach and more funders to direct their resources to collective intelligence.
When millions of people were forced into working and collaborating remotely from early 2020, it not only caused an accompanying explosion in the use and development of new technologies for productive and effective collaboration – it also made many of us more aware than ever of how technologies influence our thinking and creativity.

The work of the Centre for Collective Intelligence Design (CCID) at Nesta rests upon the premise that collective human intelligence combined with machine intelligence is more powerful than either in isolation. The second round of Collective Intelligence Grants, awarded to 15 diverse teams from institutions across the globe, were designed to explore and test this idea in new ways.

The grantees fall into broad groups, each centred around a collective intelligence dynamic or challenge. The first group of grantees looked at ways in which AI can complement human crowd intelligence. Madrid’s Spotlab successfully corralled large numbers of untrained people, via a simple game, to produce medical diagnostic data which is on par with that generated by medical professionals. Rome’s ISTC-CNR leveraged crowdsourced data from a medical training platform and used it to train an AI that can reduce clinical misdiagnosis – the third-most common cause of death in the US.1 Samurai Labs, meanwhile, grappled with how humans and AIs might work together to moderate online spaces and reduce cyberviolence – and discovered that sometimes machine creativity can outperform the human variety.

The second group of grantees explored ways to encourage more inclusive collective decision making. In Bristol, swarms of robots were deployed in a shopping centre to study how technologies might decrease polarisation around tough topics like tackling climate change. Meanwhile, the team at Birkbeck deduced ways in which algorithmically adjusting how information flows between people in digital spaces can improve the wisdom of the crowds, and be applied towards collective forecasting of future events.

For many of the defining problems of the coming century, negative group dynamics play a key role – from responding to environmental crises, to the challenge of collectively living within earth’s planetary boundaries. The third group of grantees looked into ways in which positive behaviours and collective action can spread, and be encouraged, within and between groups. Umbrellium’s air pollution workshops with East London residents, for
example, encouraged the idea of ‘mutually assured construction’, with individual residents connected together to collectively make environmentally friendly lifestyle changes – empowering them to feel part of the solution to a systemic issue, rather than at its mercy. Nottingham University’s team, meanwhile, explored ways in which improving inter-group communication might mitigate the tragedy of the commons, with potential future applications towards problems like overfishing and water table depletion.

And finally, the fourth group focused on ways in which collective intelligence and AI can be applied to data – gathering, classifying, and analysing it, to either address deficiencies in existing systems in terms of bias and accuracy, or to improve public awareness of its application. In Bolivia, Swisscontact turned the usual development model upside-down by bringing Andean farmers into the world of meteorology, instead of imposing new technologies upon them – their crowdsourced reports of microclimate changes and pest outbreaks helped power better forecasts across the region. And for Just One Giant Lab and the Open Humans Foundation, when they saw thousands of people coming together to collectively tackle pandemic problems – from developing better diagnostic tools to manufacturing masks – they were able to develop an automatic skills matchmaking system for them, so that in future their self-organisation might be more efficient and effective.

As with all good science, not every hypothesis was proven, nor every experiment a success. And the pandemic had a direct impact on many of these projects – as many grantees, who had to hurriedly redesign in-person workshops for Zoom (if not cancel them altogether), can attest. But just as collective intelligence – people working together, with the help of technology, to mobilise a wider range of information, ideas and insights – has played a key role in how governments and civic bodies have responded to the pandemic, it has also influenced the findings our grantees have delivered.

Taken together, these experiments reveal a number of ways in which human crowds can work together with, and through, digital technologies towards more positive, productive ends. Some findings already have real-world implications, while others point in emerging directions which should be explored – some of which our third round of Collective Intelligence grantees will pursue over 2021.
Introduction: Experiments in collective intelligence design

Collective intelligence design is the art and science of bringing together diverse groups of people, data (including information or ideas) and technology. Since Nesta’s CCID was established in 2018, we have been exploring how human and machine intelligence can be combined to develop innovative solutions to social challenges. A key part of this has been researching what works when it comes to collective intelligence design, including supporting others to generate new insights through our grants programme.

Our grants programme funds practical experiments to advance knowledge on collective intelligence design and application in fields with public benefit. In the first programme round launched in 2018, we funded 12 experiments with up to £20,000 each, and their results were published in our report ‘Combining crowds and machines’. In the second programme round, the findings of which are detailed in this report, we increased the funding and support available by partnering with three co-funders – the Patrick J. McGovern Foundation, the Wellcome Trust and Omidyar Network.

The total amount of funding available for the second round of collective intelligence experiments was £500,000. With this we supported 15 diverse experiments with grants of up to £31,250 each.

What is in this report

This report comprises a collection of articles based on the experiences of the second cohort of grantees developed by culture and technology journalist Ian Steadman. It summarises the experiments funded, highlights the main findings and outlines their relevance for the field of collective intelligence.

The 15 experiments are divided into four categories. Chapter 1 presents those experiments that explored how to facilitate better AI-crowd interaction. Chapter 2 covers the experiments that looked into how we can make more inclusive, collective decisions. Chapter 3 details the experiments that investigated how positive behaviour spreads and how collective behaviour change can be encouraged. Finally, Chapter 4 features experiments on how to use collective intelligence for better data. The report concludes with a call for more funding in the area of collective intelligence design.

Intended audience

This report is primarily aimed at practitioners and innovators who want to apply collective intelligence to address social challenges. We hope, however, that the insights will also inspire more funders, decision makers and academics to take this research further.
We believe that to tackle complex problems we need to mobilise all the resources of intelligence available to us. That’s why we funded five experiments that sought to understand how to best combine the complementary strengths of machine intelligence and collective human intelligence.

Exploring AI-crowd interaction
Making more accurate open maps with the help of an AI assistant

Experiment 1: Can machine learning increase the quality and speed of human mappers in areas underrepresented on open source maps?

Who is behind this experiment? Humanitarian OpenStreetMap Team in collaboration with Netherlands Red Cross.

Key finding: The AI-editor increased the speed of the beginner mappers, but also led to low quality mapping data being amplified. For advanced mappers, AI reduced the speed and quality of their work. New tools, particularly those using AI, need to be integrated into established mapping practices gradually.

Who is this relevant for?
• Any collective intelligence initiative that seeks to introduce AI into an established human-led practice.
• Anyone relying on up-to-date information, such as public health authorities or humanitarian organisations.

What was the experiment?
This experiment tested whether using AI-generated map features can increase the quality and speed of mapping by humans on OpenStreetMap. A total of 72 beginner and veteran mappers participated in mapathon experiments where they were randomly assigned either a traditional map editing tool (called ID Editor) or an AI-assisted map editing tool (called RapID editor). Participants were also randomly assigned alternating geographic locations (San Francisco or Uganda).

What did they learn?
AI-assisted mapping increased the speed of beginner mappers, but had little impact on the quality of their work. This meant more low quality data was produced by beginner mappers overall. For advanced mappers, the AI-assisted editor reduced the mapping speed and quality of mapping data. This highlights the risk of introducing new tools into well established mapping practices.

Maps have power. During the European Age of ‘Discovery’, the colonial powers used maps to both guide and legitimise their actions. Today, maps are increasingly digital products – and the choices that go into what is and isn't mapped reflect the biases and incentives of those who do the mapping. San Francisco and London are more thoroughly mapped than rural areas in countries like Indonesia or Kenya, and that can have serious implications for disaster relief. How can you help people if you don’t even know where they live?

OpenStreetMap (OSM) was founded in 2004 in response to the trend for geospatial data being locked within proprietary systems. "Imagine Google Maps and Wikipedia had a baby," explains Bo Percival. He's director of technology and innovation for the Humanitarian...
OpenStreetMap Team (HOT), an NGO that leverages OSM’s community of contributors to build better maps in areas where that can make a difference for things like vaccine distribution, environmental activism, and disaster relief – a key use case for OSM in the years since its launch. But, also like Wikipedia, improving its accuracy is often a slow and laborious manual process – which could cost lives in urgent situations.

“OpenStreetMap Team (HOT), an NGO that leverages OSM’s community of contributors to build better maps in areas where that can make a difference for things like vaccine distribution, environmental activism, and disaster relief – a key use case for OSM in the years since its launch. But, also like Wikipedia, improving its accuracy is often a slow and laborious manual process – which could cost lives in urgent situations."

This isn’t the first attempt to improve OSM contributions with AI assistance, but for a mixture of reasons it hasn’t worked in the past. “There’s resistance to this in the OSM community because people are afraid that it may have a negative impact. And we have seen some examples of that, where companies come in and their model is not great – so we want to build an evidence base around the use of AI, to break through the fear.”

HOT’s experiment tested their assistive AI’s impact on speed and accuracy with 24 beginner and 48 veteran mappers, who were tasked to map buildings from images of San Francisco and rural Uganda in one-hour-long mapathons. They found a clear benefit in terms of speed with beginners in both contexts – although more so with the US images. Beginner mappers using the AI-assisted tool created 82 per cent more data compared to the traditional editor. However, there was little impact on mapping quality, which meant more low quality data being produced by beginners. With veteran mappers there was actually a decrease in speed and quality when using the
AI-assisted editor. However, Percival believes that this reflects that it was more disruptive to unlearn existing methods first than it was to learn from scratch — something which would correct itself given more time. There was also more acceptance from the community than HOT expected. “Many of the communities in lower- and middle-income countries were really open to it. That was a nice surprise, because the loudest voices of resistance weren’t actually coming from the places that suffer most from lack of data.”

While the results clearly demonstrated that AI assistants were worth investigating further as a mapping tool, they also provided valuable insights into underlying issues with OSM’s workflows. The speed increase led to bottlenecks in the editorial verification process, for example. It also illuminated a deeper issue with OSM’s crowdsourcing model.

“The theory is that collective intelligence raises the bar, but that assumes that new data is better than what it replaces. Maybe someone comes in and maps a building they’ve visited, but OSM has no way of privileging that kind of local knowledge, and then an AI comes in and overwrites it. The mappers that have been around for a long time are right to be concerned that uncontained models might overwrite years of work and careful attention to data quality, and understandably want to protect those enormous human efforts.”

The model also had to be trained, but the free open data that was used (due to budget limitations) had its systemic errors replicated by the model: “In San Francisco obviously there’s good quality data, but you can imagine that a terraced house on a hill in San Francisco looks nothing like a hut in rural Uganda. The adaptability of the model fails very quickly.”

The team was also prevented from investigating whether an open AI model developed by an NGO partner was comparable to the corporate solution used for the trials – because the NGO’s staff were deployed to the Philippines to provide aid after a typhoon.

“What we were really looking at is the collective intelligence of people in different spaces,” argues Percival. “Someone sitting in Europe will be mapping the home of someone in rural Tanzania, for example – and, if I’m honest, people in Western countries can have a massive bias, thinking that people in developing countries don’t have intelligence if they disagree with ‘us’. It’s an example of how collective intelligence doesn’t necessarily always improve things, because there’s history and context that might be totally ignored.”

HOT hopes that the future of AI-assisted mapping won’t amplify this bias, but instead correct it, by making sure that it includes and prioritises the underrepresented voices that can speak to the truth on the ground.

“The loudest voices of resistance weren’t actually coming from the places that suffer most from lack of data.”
Harnessing the wisdom of crowds for more accurate medical diagnostics

Experiment 2: How can we increase the accuracy of collective diagnosis from medical professionals on the HumanDX platform?

Who is behind this experiment? Istituto di Scienze e Tecnologie della Cognizione (ISTC) at the Italian National Research Council (CNR) in collaboration with Max Planck Institute for Human Development (MPIB) and the Human Diagnosis Project (Human Dx).

Key finding: A collective diagnosis, aggregated from the individual diagnoses from medical professionals, is more accurate than individual diagnoses.

Who is this relevant for?
- Practitioners and researchers designing collective intelligence platforms (particularly those that deal with open-ended questions, such as in the field of digital democracy).

What was the experiment?
The experiment explored whether AI tools known as knowledge graphs could help physicians on a large medical crowdsourcing platform, called the Human Diagnosis Project (HumanDx), produce more accurate medical diagnoses. The experiment aggregated diagnoses from groups of up to ten medical professionals by using knowledge graphs, and compared this to individual diagnoses to understand which method (collective or individual) was most accurate.

What did they learn?
The aggregated result or ‘collective diagnosis’ was more accurate (92 per cent accurate) compared to the individual diagnoses (57.5 per cent accurate). Previous attempts to aggregate medical diagnoses on the HumanDx platform required expert raters for the aggregation of individual diagnoses, however this experiment also demonstrated that the procedure can be fully automated.

What is a knowledge graph?
Knowledge graphs are a way to organise the world’s structured knowledge, and a way to integrate information extracted from multiple data sources. They consist of nodes, edges, and labels. Anything can act as a node, for example, people or a company. An edge connects a pair of nodes and captures the relationship of interest between them, for example, a friendship between two people or a customer relationship between a company and person. The labels capture the meaning of the relationship, for example, the friendship relationship between two people.

---

a. HumanDx is an online system combining the collective intelligence of medical professionals and trainees with machine learning to allow collaboration on any case, question, or other medical topic.
The third most common way for someone to die in the United States is a preventable medical error—and one of the most common errors is misdiagnosis. Finding ways to help doctors and nurses avoid such mistakes would save thousands of lives every year. But how? One solution might lie in offering patients not just a chance of a second diagnosis, but the expertise of thousands of doctors around the world simultaneously.

"With medical diagnostics you’d expect that experts would come out with similar things, but the variability is huge...that’s a richness that we should exploit."

"I think the driving motivation for me is that while misdiagnosis is known to be a big problem, it also has all kinds of downstream effects," explains MPIB’s Ralf Kurvers. "The most obvious is death—an estimated quarter of a million people in the US die each year from misdiagnosis—but people receiving the wrong treatment can incur other serious health complications, which in turn can lead to a loss of faith in the medical system. On the other hand, there are all kinds of human biases that play a part in preventing doctors from harnessing their optimal decision making powers."

"So we wanted to explore the classic theory of the wisdom of the crowd for this," adds ISTC-CNR’s Vito Trianni. "Put together information from many different experts—do you get a better answer?"

While there’s already been research that’s found polling larger groups of clinicians worked well in some limited scenarios, like assessing mammograms, the team was interested to see if the principle could be expanded to more general, open-ended situations, where there are no binary or predefined answers. The collective intelligence of many doctors, applied at scale, could provide the foundation for an AI ‘second opinion’ system.

The team joined forces with Human Dx, which has developed an online platform where medical practitioners can post cases and

HumanDx’s mobile application
crowdsourcing input from thousands of other doctors around the world – to the benefit of medical professionals and trainees across public and private medical institutions in more than 100 different countries. “Imagine there’s a patient coming in,” says Kurvers. “A 20-year-old male with a cough. The physician goes through a number of diagnostic steps to figure out what’s going on – like a CT scan – and makes a diagnosis. When they post key information about the case on Human Dx, other physicians can type in what they think it is or select it from a menu.”

The information available on Human Dx’s dataset is wide-ranging, and often has discrepancies in terminology, spelling, and other ambiguities arising from differences in language, training, and experience – as well as human error. This makes it perfect as a representation of how varied the process of medical diagnosis can be, albeit while also offering a considerable data aggregation and processing challenge. “I mean, how do you calculate the accuracy of a list of diagnoses, right? That’s not a trivial thing,” says Kurvers.

“We had 1,500 medical cases – a small subset of the Human Dx database – but that still provided around 40,000 different medical terms,” says Trianni. The system they built to analyse those terms, CROME AI, was able to construct a linguistic model that could automatically match them against the SNOMED CT database, a collection of standardised clinical terms used by major health systems like the UK’s NHS. It means that all the different ways in which a diagnosis might be listed – an acronym, a geographically specific term, and so on – could be linked together into a network of concepts, also known as a knowledge graph.

Owing to the matching between Human Dx data and SNOMED CT, the team could investigate their collective intelligence hunch. “In short, what we found is that we could observe the wisdom of the crowd,” says Trianni. “The accuracy increased substantially with larger groups.”

Individual guesses from doctors were correct only 57.5 per cent of the time, on average, but accuracy was as high as 92 per cent when aggregating the diagnoses of ten users. Users who self-reported as experts in certain specialities were also observably better than non-experts at diagnosing the same cases.

Yet there are plenty of questions left unanswered – for example they saw the potential of CROME AI being able to give more weight to the answers of Human Dx users with a good track record, but the team was not able to identify such good users in the available dataset. The team believes that can be solved with larger datasets in future research. Overall, though, the results are encouraging. CROME AI suggests that it is possible to automate the structuring and aggregation of knowledge in open-ended domains – something which is traditionally extremely difficult – opening to real-world applications beyond just healthcare, in fields such as policymaking and law. It could also be used in situations where the collective intelligence of non-experts may be just as valuable.

“My take is that open-ended domains are complex to master in general,” says Trianni. “With medical diagnostics you’d expect that experts would come out with similar things, but the variability is huge, and it’s difficult to deploy automated systems. However, that’s a richness that we should exploit. This is the challenge that we need to address in order to make systems that can harness collective intelligence.”
Encouraging creativity with a search engine that delivers unexpected results

Experiment 3: Can a serendipity-inducing recommendation algorithm improve a group’s creative problem solving ability?

**Who is behind this experiment?** neu (Augmented Thinking) in collaboration with City, University of London.

**Key finding:** While creativity wasn’t clearly enhanced by the serendipity-inducing algorithm, nor was it inhibited. People need time to make sense of unfamiliar connections in order to find them helpful.

**What was the experiment?**
The experiment tested the capacity of a serendipity-inducing recommendation algorithm to increase innovation in tackling complex social problems. In the experiment, innovators were tasked with developing solutions to two social issues (plastic pollution and an ageing population). Some had access to a regular search engine, while others were provided with a serendipity-inducing recommendation algorithm. It tested what kind of search result, accurate vs unexpected, was most beneficial for creative solution generation.

**What did they learn?**
The experiment highlighted that participants need time to make sense of unfamiliar connections in order to find them valuable; introducing distantly related suggestions early on confused ‘novice’ users. The evaluation of submitted solutions by participants who used the serendipity-inducing recommendation algorithm were negligibly rated better by the judging panel in comparison to the control group using the regular search engine. Whilst there isn’t a clear outcome from the experiment, the research demonstrated that recommending distantly related concepts offers at least as good an ‘innovation experience’ as those that were more expected.

Alexander Fleming discovered penicillin in a petri dish left out by mistake; the microwave oven was invented after Charles Spencer, a radar operator, realised a bar of chocolate in his pocket had melted; Charles Goodyear invented vulcanisation after dropping some rubber on a hot stove. Serendipity has been key to many innovations – but could it ever be harnessed deliberately? That’s what neu, a collective of designers, artists, and developers, wanted to explore.

“This came from a personal frustration that I experienced while a student at the Royal College of Art,” explains Johannes Mutter. “We were designing a climate-controlled hydroponics unit for urban agriculture, but after a few months realised that two other students had been working on the same thing.” They coined the term ‘design waste’ to describe what had happened – not just a waste of effort and resources, but also of a chance to collaborate.
They set out to understand its causes, in order to prevent it happening – and not just in physical spaces where creative work happens, like universities or laboratories. “In digital spaces [like search engines and ecommerce sites], it’s much worse,” says Mutter. “They’re often based on the principle of making accurate recommendations – if you’re watching The Magnificent Seven on Netflix, it’s going to show you other Westerns. But creative people need to make wild associations.” With such a frequently heavy focus on delivering familiar, comfortable results, it leaves little room for serendipitous discovery.

Neu set out to design a recommendation engine that would inspire serendipity rather than seek out similarity, and which could be used as a creative tool by inventors. At its heart is a knowledge graph “structured like the human mind, by both association and hierarchy,” containing thousands of technological and design concepts – from the theoretical to the engineered – in hundreds of thousands of different permutations. (“It’s like a Lego set. Even when you’ve assembled the building you can still see the bricks, and build a million other things with them that weren’t in the instructions.”)

The original corpus was sourced from existing technical literature, as well as contributions from 109 members of the public across 34 countries. And, as time goes on, there have been further contributions by each new wave of users – their expertise, insights, and ideas coming together into a form of crowdsourced intelligence.

For their Nesta experiment, the team built a platform – Fridays for innovation – to encourage innovators to spend their Fridays launching side projects to tackle social impact or sustainability-related issues, using the knowledge graph’s tools and concepts for inspiration. There were two design challenges in particular: reducing plastic pollution in South East Asia and managing the UK’s ageing population, to be dealt with over a four-hour-long hackathon format. Some participants were given access to a search engine, plugged into the knowledge graph, that was intended to make serendipity more likely; the others, in the meantime, were only given a normal search engine. The recommendation engine was also tweaked in different ways (like how it presented results, or how autocomplete suggestions appeared) in order to find the optimal mix of ‘expected’ and ‘unexpected’ results that would encourage maximum creativity.

Unfortunately, the outcome was inconclusive – while there were minimally better results with the groups guided by serendipity, they were not significantly so. The pandemic was possibly to blame, as the sessions were held remotely instead of in-person and participants relied on search engines like Google for extra information despite being asked not to. Most of the participants were students, too, which skewed how much prior knowledge the groups possessed. However, one important lesson for the future is that it takes time to adjust to a

The Fridays for Innovation platform and search engine
serendipity-based recommendation system. Participants were liable to find unexpected ideas confusing or unhelpful until they’d had a few minutes to adjust.

“There was still one incredible part of the early phase of the experiment, though,” Mutter says. “During the start of the pandemic in Germany there were #WirVersusVirus hackathons for developing low-cost ventilators and masks, with around 45,000 people participating, and we joined as a supporter and gave them access to our database. Within 24 hours we’d gathered more than 2,000 new solutions. It was a perfect example of collective intelligence, in that there was a huge amount of people with expertise in many, many areas – but it also introduced entirely new challenges for us around vetting the quality of all this new information.”

They also collided with an unavoidable reality of the creative world: IP. As much as open access to patents is key for addressing many global issues, there are plenty of experts who, understandably, are reluctant to open their notebooks and share their favourite ideas. “That’s their Holy Grail, where they have their most important ideas – and we had something of a false vision about whether people would share them with us.” While Mutter floats the possibility of compensating creatives for those ideas, part of the platform’s strength is that it offers a structure where non-experts – less concerned with owning their ideas – can also contribute to collective knowledge at scale.

Nevertheless, Mutter believes that an association-based database and recommendation algorithm could have profound applications in other areas with dense, interlinked knowledge bases – like law.

“Collective intelligence isn’t something you achieve – it’s something you strive for. Gary Kasparov has shown that the combination of a human and a machine is better at competitive chess than either working alone, but it’s all about how you combine them together. And maybe, for the combination of artificial intelligence and creative human intelligence, the best interface we have is still the conversation. It’s the best way to stop you from going in the wrong direction – and for pointing you in a direction that’s more promising.”
Volunteers and AI working together to counter cyber violence

Experiment 4: Can an AI-based detection system and humans work together to reduce the level of online harassment on Reddit?

**Who is behind this experiment?** Samurai Labs.

**Key finding:** Overall, human volunteers were less creative and engaged than anticipated, meaning it is more efficient to use an AI-powered bot that intervenes and responds automatically to cyber violence. Normative interventions were more effective at reducing cyber violence than empathetic responses.

**Who is this relevant for?**
- Social media companies.
- Practitioners and researchers designing collective intelligence platforms (where there may be a risk of cyberbullying).

**What was the experiment?**
The experiment tested whether cyber violence on Reddit can be significantly decreased by interventions delivered by volunteer community-users when working in partnership with an AI-based detection and notification system. It also tested the effectiveness of two types of verbal interventions used by volunteers – normative and empathy-based.

**What did they learn?**
Users in the normative treatment group had a significantly reduced count of attacks compared to the empathetic group. However, there were numerous challenges including volunteers being less engaged and less creative in their responses than anticipated, and the negative impact on volunteers' wellbeing as many received abuse. Overall, the team felt it was more efficient to create an AI-powered bot, which detects personal attacks and responds automatically rather than work with human volunteers.

The last few years have seen social media platforms become more popular than ever, with the conversations they host having powerful real-world consequences. However, juggling the right to freedom of expression against the need to moderate such spaces – and address serious issues like hate speech, cyberbullying, and death threats – is a difficult challenge. Gdynia’s Samurai Labs has been exploring whether moderating online crowd interactions with the help of an AI assistant may offer a solution.

“We build AI to detect and prevent online violence,” says Patrycja Tempiska. “Samurai grew out of a project more than ten years ago to teach computers how to understand human language, and since then we’ve helped detect cyberbullying, suicidal ideation, grooming behaviour, and potential school shootings. We’re always searching for effective strategies that don’t resort to straightforward censorship. These are issues close to our hearts – we know how prevalent they are online, and we want to help solve them.”
One of their most recent of those experiments was ‘James’ – a chatbot deployed on a ‘men’s rights’ forum, pretending to be a normal user. When James detected users posting attacks at each other, ‘he’ was programmed with more than 100,000 different responses with which to try and convince them to stop. By testing two different approaches – emphasising either empathy (“Hey, think about how they must feel when you say that”) or behavioural norms (“Hey, this isn’t how people like us are supposed to behave”) – there was a 20 per cent reduction in interpersonal attacks. But with their Nesta experiment they were curious to explore the dynamics they’d observed more fully, especially when it came to James’ lack of creativity.

“Humans should do much better than bots,” says Tempska. “They’re more creative, they’re better at understanding context, and they can stay on topic.” An AI assisting humans should, they felt, be better for moderating interpersonal attacks in online spaces than either humans or AIs alone, and deliver better results than James’. “And from my perspective as a psychologist, it’s a dream to be able to take clinical practice and apply it on a big scale,” adds Maria Dowgialło. “It could be a way to scale kindness.”

They focused on Reddit, whose subject-specific forums – or ‘subreddits’ – are moderated by human volunteers with warnings and bans. Instead of that “punitive” approach, their AI scanned through millions of conversations across the whole site over six months and alerted a team of 20 volunteers whenever personal attacks were detected. A volunteer would then dive in and intervene along either empathetic or normative lines, as James had, although with the freedom to come up with their own responses rather than rely on a script. Freed of the burden of having to seek out and assess problematic posts, it would enable more innately human traits like creativity.

Their hypothesis had been that (much like with James’ interventions) normative interventions would be effective, but empathetic ones wouldn’t – and that’s exactly what they found. “Our assumptions were validated and invalidated in a number of ways,” says Tempska.

They have several explanations as to why. Most crucially, their assumption that humans would be more creative than an AI was incorrect. “With a few exceptions, the volunteers were really repetitive – some were accused of being bots,” says Tempska. “But with James we joked that he passed the longest-lasting Turing test ever, because near the end of his eight-month run a moderator actually invited him to join the moderation team.”
Empathetic responses required greater creativity compared to normative and punitive responses, but there were general issues with volunteer enthusiasm throughout the experiment. Fewer people signed up than hoped, and the drop-out rate was higher than expected – to the extent that the team started paying people with Amazon gift cards to try and keep them engaged. Some volunteers also reported a significantly negative psychological toll from being exposed to toxic content for multiple hours per day – an issue which many social networks have struggled to mitigate with their own human moderators.

It also didn’t help that the participants lacked a connection with the people they were replying to. “With James, he was in only one forum for eight months,” says Dowgiallo. “Even the people who disliked him adjusted to the fact that this guy was running around telling everyone to be kind. But with our Nesta experiment we were targeting users across Reddit, and our volunteers were outsiders to those communities.”

Overall, the participants were less effective than James had been on his own. However, the results were still significant, and provide evidence that AIs should be investigated further for their potential to automate (either partially or fully) the detection of cyber violence. Samurai’s AI is able to detect verbal aggression with more than 93 per cent precision, and can be implemented in other contexts where language detection and comprehension is required at scale.

“We’ll be conducting more experiments,” says Tempska. “Like seeing whether women respond better to empathy, for example. But I think we wouldn’t repeat this approach, recruiting and using volunteers this way. Instead, we recently developed a moderator bot for [chatroom app] Discord which makes proposals that human moderators can then choose to act on. They’re trained people who know their community well, who are engaged, and who have an incentive to moderate.”

‘Although our goal in the long-term is autonomous moderation – to take the biggest burden off of moderators’ shoulders – it remains the case that we always need to fine-tune any collective intelligence strategy to its context, rather than abandon the idea of human and machine collaboration altogether.”

“With a few exceptions, the volunteers were really repetitive – some were accused of being bots.”
Using serious games to train AI models for medical diagnosis

Experiment 5: Can citizens playing online games be as effective as physicians in training AI models for image diagnosis?

Who is behind this experiment? Spotlab.

Key finding: AI models that have been trained on medical images annotated by non-expert adults and school children are as accurate at diagnosing diseases as AI models trained on expert-based annotations.

Who is this relevant for?
- National policymakers, practitioners and researchers in health.
- Citizen science projects and crowdsourcing platforms.
- International and humanitarian organisations.

What was the experiment?
This experiment tested whether AI models trained by the general public could outperform medical specialists in analysing medical imagery (digitised blood samples) for the diagnosis of global diseases such as malaria. The AI models were trained by volunteers (adults and school children) performing medical image analysis while playing serious online games (through a website or app) for global diseases diagnosis. The accuracy of models were then tested against the performance of medical specialists.

What did they learn?
The experiment found that AI models trained on images annotated by both adults and school children can obtain similar results to ones trained on expert-based annotations, of around 93 per cent accuracy. The team found that the minimum number of different responses (or annotations) from adults and school children needed to achieve a level of accuracy comparable to experts was 20. In fact, responses from 20 school children alone could achieve a similar level of accuracy to experts.

Millions of people are affected every year by diseases which go undiagnosed, simply because some healthcare systems struggle with limited capacity. Spotlab has been using a game called SpotWarriors to see if the wisdom of a crowd of non-experts is comparable to the expertise of trained doctors – and, in the process, empower people to become part of a global healthcare solution.

“Microscope diagnosis – looking at blood, skin, etc, under a microscope – is the gold standard for several diseases,” explains Spotlab’s Lydia Garcia. “But it requires specialist time and resources. It’s a very slow process.”

So, what to use instead? “An AI model,” says Lin Lin. “We thought it would be faster. But the problem there is that training AI models requires a lot of data, and getting someone to label all that sample
data is also difficult because doctors are so busy. So we asked, ‘What if we get a group of normal people like us to label the images, and train the model with that data?’ – and we prove it’s possible.”

So how do you get a group of people without expertise to label medical images? The answer is a game called SpotWarriors. After showing players example microscope images of certain diseases, it then asks them to spot the same signs of infection in other images. “In each image, if the player attempts an identification – identifying a blood cell with a malaria infection, for example – and we know what the correct answer is, they either score or lose points,” says Lin. “But if they select a part of the image where we’re unsure, it’s marked with a question mark for further verification.”

The concept was first developed in 2013 by Spotlab’s Miguel Luengo, with its first iteration, Malariaspot. “He found that 22 non-expert players were equivalent to one specialist,” says Garcia. “Every player makes mistakes, but correct guesses tend to be clustered, so you can detect accuracy in the noise.”

For the SpotWarriors project, the team wanted to test their hypothesis on both kids and adults. They ran 50 workshops in schools, with more than 1,000 children taking part (and this was made much more difficult by the severity of the COVID-19 pandemic in Spain). “The teens were very competitive,” says Garcia. “They were like, 'Oh, now I'm a doctor! That's really fun. This is a beautiful experiment in some ways, because it empowers people. One of our partners, Elena, is an expert on soil-transmitted parasites – but her mother is so obsessed with playing that she's more of an expert than her daughter now. It's awesome.'
Since people’s guesses tend to cluster around the correct answer in each image in the game, the team could use those averages – from across 700 different ‘levels’ – to train a neural network in how to spot the same diseases from the same symptoms. They took a different dataset of 7,600 diagnostic images, and set the neural network loose on it to see how accurate it would be. The result? 93 per cent of the time it correctly diagnosed the relevant illness – and not only was that accuracy near-identical regardless of when either guesses from only adults or only kids were used to train the neural network, it was on par with the accuracy displayed by medical professionals in Spain and Kenya, who took part via a tele-microscopy platform. Lay people, even kids, can actually compare to experts when it comes to diagnosing certain diseases.

The fact that only 700 images are used to train the model is a limitation – expanding the dataset would likely improve the model even further. The team has floated the idea of having hospitals upload new diagnostic images to SpotWarriors in the future whenever an in-house expert is unsure of the correct diagnosis – say, they’re only 90 per cent certain – and letting players collectively tackle it instead. Those guesses, in turn, would help improve the model further still by becoming part of its training dataset.

“We also learned a lot about the importance of game design here,” says Garcia. “With parasites, we designed the game to be like Tinder – you swipe if you see one. But with malaria, we made it more like scrolling around on Google Maps and placing pins. We got better results with the former.” The game also only works with diseases where the key markers fit in a box on a screen. But it has potential. Wherever people kill time by looking at their phones – on the subway, in bed, on the toilet – they could be analysing pictures on their phone, helping to train models like this.

“Everyone is an agent of change,” argues Garcia. “This game gives us a chance to promote principles like solidarity and empathy. It doesn’t always have to be that solutions come from politics – they can be out there in society.”
As we become more polarised in our views, and challenged by the need to make rapid decisions on emerging issues, it can be hard to ensure that we listen to diverse perspectives. But research shows that embracing diversity of experience and opinions is key to better decisions and solving problems more effectively. Two of our grantees tested new methods for more inclusive collective decisions.
Hacking the wisdom of crowds to improve collective forecasting

Experiment 6: Will algorithmically moderating a social network to maintain diversity of opinion improve collective forecasting?

Who is behind this experiment? Centre for Cognition, Computation and Modelling (CCCM) at Birkbeck (University of London).

Key finding: Moderating deliberation in social networks with different rewiring algorithms influences the accuracy of forecasts, though there was no statistically significant main effect between the network treatments.

Who is this relevant for?
• Policymakers.
• Anyone working in fields where decisions need to be taken that have an objectively correct answer (e.g. medical diagnostics or intelligence analysis).

What was the experiment?
This experiment explored how people behave in collective forecasting scenarios, and the role that technology can play in supporting collective deliberation and overcoming polarisation online. Birkbeck developed and tested different rewiring algorithms – programmable rules for manipulating who communicates with whom – as a way of steering deliberating groups towards more accurate forecasts. During the experiment, participants were tasked with forecasting the probability of ten real-world events occurring. They then viewed the forecasts of others and were invited to revise their own forecasts. Different algorithms were used to influence communication between group members, for example one promoted the most extreme opinions while another promoted the most average opinions.

What did they learn?
While Birkbeck did not observe a statistically significant effect between different network treatments, all three algorithms measurably improved the quality of the average forecasts made by those groups. Birkbeck’s experiment on moderating deliberation in social groups serves as proof of concept for using rewiring algorithms as tools for influencing the accuracy of collective predictions in different ways.

The ‘wisdom of crowds’, where the average opinion of a group of people is often more accurate than any individual member alone, has been noted for centuries. Social media is perhaps the largest crowd of all, with billions of users engaged in trillions of conversations. The team at the CCCM at Birkbeck (University of London) were interested in whether it might be possible to take advantage of how those conversations flow back and forth in digital spaces, and leverage that collective wisdom towards predicting the future.

“There’s been a lot of research in the last few years to try and understand how social networks have influenced people, so that maybe we could undo some of the negative things that seem to keep happening,” explains CCCM’s Ulrike Hahn. “And while there’s this historic literature on the wisdom
of the crowd and aggregating judgments, there’s also more recent research that asks, “If I try to put together the right people, and let them talk to one another, can I make their judgments more accurate? And if I amplify the people who are better at forecasting future events, then can I improve the quality overall?”

The team wanted to see if they could combine these two areas of inquiry. A group of participants who are asked to discuss a simple issue – like whether England would win the Euros, for example – might well deliver a strongly predictive guess, once their opinions are averaged out. Even if you can’t know in advance which beliefs will turn out to be more accurate than others, they hypothesised that it should still be possible to understand how exposure to certain kinds of opinions from other people influences how a person then makes their own judgements. Improving the information that each individual person gets exposed to could nudge them towards better forecasts about future events – and the effect should be even more profound when averaged across whole groups of people.

For their experiment, the team wanted those discussions to take place within an online forum they’d built where they could control the ways in which people were able to speak with each other. Algorithms could remove or add different connections between people in different ways over a series of rounds, effectively ‘tuning’ how much information any one group member would be able to see about everyone else as they were making their own iterative guesses about the future.

“It’s not unlike what happens on Twitter, where there’s a recommendation algorithm for promoted content,” explains CCCM’s Jason Burton. “But this could also be a civic technology in the context of digital democracy, where people use online platforms to discuss and vote on different issues, and see how other people are voting.” If you can deliberately induce a better version of the wisdom of the crowds, it opens up all kinds of possibilities for improving collective forecasting in many different contexts.

They first ran some simulations to test the idea, and used the data that was generated to derive the tuning algorithms for the main experiment. Human participants (recruited via Mechanical Turk) logged on to a bespoke group discussion platform and made predictions on ten real future events – like “Both teams in this year’s Super Bowl will score more than 20 points”, “Joe Biden’s approval rating will be higher than 55 per cent after three weeks as US president”, and “Bitcoin will be valued at less than $30,000 on 8 February 2021”. They also explained why they were making their decisions, and those rationales were shared with the rest of the group in different ways.
“So now we’ve got this network, we can intervene on it,” says Hahn. “Who should talk to whom? Who should we give some more connections to? Who should we give fewer? Then we can look, at the end, whether our changes made forecasts more accurate or not.” Alongside a control group (whose interconnections were randomly generated at the start of each trial and left ‘static’ throughout), there were three different algorithms applied to the other groups: one that repeatedly tried to promote the most extreme, outlying opinions; one that did the opposite, promoting the most average opinions; and a final one which randomly changed the connections over time, until everyone had been connected together at least once.

The team found that all three algorithms measurably improved the quality of the average forecasts made by those groups – they’d managed to improve on the wisdom of the crowds. At the same time, the guesses by the participants in the control group actually got worse the more that they communicated with each other. “In general, the participants were bad at predicting,” says Burton. “These weren’t experts, but it seems like these algorithms would work even better if people had more pre-existing knowledge they could bring into the discussions.” They could even see groupthink emerge in the control group, as participants with stronger – yet less correct – opinions hijacked discussions, steering everyone else.

“We’re still a long way from this kind of algorithmic ‘steering’ being rolled out over the ‘high stakes’ scale of a social network’s news feeds, “but I think it’s a completely new direction of research,” Hahn argues. “We’ve only started to scratch the surface on the different contexts in which these algorithms work.” At the same time, there may already be applications in smaller-scale contexts – such as in focus groups.

“This is collective intelligence in the sense of a sort of hybrid intelligence, with machines supporting human thinking and processes,” says Burton. “We’re using these new digital tools to provide the scaffolding for groups to work together.”

“This kind of work helps you understand where you can build practical tools that will help with prediction problems that you care about,” says Hahn. “But there’s also this spinoff with respect to the hybrid human-computer intelligence systems that we’ve already created, like social media. Thinking about it this way allows you to understand not just how those things could be better, but also why we might’ve made them worse.”
## How swarms of robots can improve group debates

### Experiment 7: Can a swarm of robots interacting with humans facilitate social interaction and help people reach informed consensus?

<table>
<thead>
<tr>
<th>Who is behind this experiment?</th>
<th>University of Bristol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key finding:</td>
<td>Initial results show robot swarms can be used to engage people on challenging topics, serve as a prompt to launch conversations, and empower introverts to share their opinions.</td>
</tr>
</tbody>
</table>
| Who is this relevant for?       | • Deliberative democracy experts and facilitators.  
                                 | • Conference organisers, pollsters and anyone looking to design better ways of enabling groups of people to make decisions on polarising topics. |

### What was the experiment?

The team built a swarm of 100 small robots and tested whether the robots could help a crowd to reach inclusive and informed consensus by communicating opinion diversity. The robots were deployed in two experiments: 1) a controlled experiment with university students, and 2) a real-world deployment at a shopping centre in Bristol. Participants could input their responses to a question into the robot. The robot could then display the result for other participants to see before they respond to the same question.

### What did they learn?

The experiments show that robots are an effective tool to diffuse and influence opinions within group settings. The first experiment asked about the percentage of online vs in-person teaching the university should offer, and the percentage of coursework vs final exams modules should be given. The results show that participants changed their opinions during the experiment while viewing the opinions of others. In the real-world deployment, the public were asked “What is the most important action in your opinion that people should take to help control climate change?”. Participants would often review a number of opinions before choosing which to update, suggesting their desire to override opinions they disagreed with. Conversations were often triggered during the experiment, highlighting that the robots were an effective tool to engage the public on a difficult topic.

When groups of people get together to make complex decisions, many factors can influence what happens. In contexts such as citizens’ assemblies, it’s vital that as diverse a range of perspectives and opinions are shared as possible – but at the same time, the larger the group, the more difficult it is for any one member to track every key point. Communication technologies, meanwhile, are just as liable to facilitate polarisation as they are to encourage more democratic outcomes. A team from the Bristol Robotics Laboratory (BRL) were curious whether they could find a way to mitigate this issue – with a swarm of robots.
We wanted to see if we could help people have more meaningful conversations on challenging topics,” explains BRL’s Sabine Hauert. “Let’s say you’re in a large group that’s trying to decide something, and you want the input from the crowd. Typically they’ll chat among themselves, but that can lead to opinions being stuck in silos; or, if the crowd listens as people give speeches, some people might avoid sharing their opinions because they’re nervous or shy.” These dynamics can apply to all kinds of real-world situations where important decisions must be made – like workshops, focus groups, juries, and even legislative sessions.

“We wanted to see if we could help opinions travel and mix more freely in those situations, and allow crowds to be a bit more clever about how they reach a consensus. So we built a swarm of robots.”

The team – primarily led by PhD student Merihan Alhafnawi – built prototype robots running on low-cost credit card-sized computers (Raspberry Pis), each about the size of a smartphone and communicating with each other via Bluetooth, to deploy in an initial test experiment with around 40 students broken into smaller groups. Each robot has a small touchscreen on top, which displays different contextual information as they drive around in random patterns – maybe a question that a participant can answer, or the response that someone else already gave to that question. The idea is that the swarm gradually becomes a reflection of a group’s collective opinions as they’re formed, which in turn can influence further discussions in new ways.

“To understand how the robots’ design changed the dynamics of their discussions, we asked them, for example, what percentage of courses should be online versus in-person,” says Hauert. “It added to the conversations they had with their mates – whether you go with the majority or vote against it, it triggers you to ask them how they voted, and helps those who might prefer to be anonymous. And throughout, we studied the experiment’s design – how the robots moved, if introverts or extroverts were more engaged, whether it made their collective views less or more polarised over time. We’re still analysing that data, but it’s promising – there was a lot of excitement about having the robots involved.”
They then conducted a second experiment in Cabot Circus, a large shopping centre in central Bristol, by letting loose swarms of dozens of robots in a small arena. They asked people walking past to stop and answer a question about what they thought was the “most important action” that people could take “to help control climate change” – and then choose from a range of options, like recycling more, insulating their homes, or becoming vegan. Their choices were displayed on one of the screens to form part of the swarm of overall opinion.

“We started out thinking this was about consensus formation,” explains Hauert, “but when we brought in Séverin Lemaignan, an expert on human–robot interaction, he helped us realise that it wasn’t just about having crowds agree. It was about how robots help human interaction in any kind of brainstorming setting, including where it might be more useful to inspire tension rather than consensus, and climate change allows that tensioning of opinion.”

“And what was interesting is that there were no right or wrong answers. That led to a lot of conversations. People who normally would never chat with us saw the robots and came over – even the kids had opinions. Or we had people arguing that what we were doing was pointless, because climate change required much larger systemic actions than we were asking about. Some people even tried to trick the system by manipulating the votes. But everyone was still looking at the screens to see what others entered, and that made them think.”

The experiments can provide a model for how swarm robotics might be applicable in other, more serious group settings – like citizens’ assemblies. It also shows how ordinary people can be brought into the policy debates around huge issues like climate change in a new way, and at the same time provide policymakers with richer, more nuanced information about public opinion on divisive topics.

COVID-19 limited the number of interactions they could have with members of the public, but the data is still providing valuable insights into how robot swarms can be used to influence group consensus-making. Future experiments will explore other ways to gather and display information – from using the touchscreens as smart post-it notes to forming ‘expressive swarms’ that combine together into larger displays. “We’re thinking quite artistically about what types of performances we could do in future,” says Hauert. “But fundamentally, these tools are meant to be aids. They shouldn’t change the underlying mechanisms of how we form a consensus.”

“[Robot swarms are] not usually used in social settings, but I think this shows that can change. We might normally only think of collective robot intelligence, and collective human intelligence, but the fun thing here is the crossover. It lets us ask how they can benefit each other.”
The importance of understanding collective behaviour in relation to disasters such as floods and pandemics is obvious. It is also significant in tackling many of the complex challenges that we are grappling with in the 21st century, from rising obesity levels to living within planetary boundaries.

Three of our grantees investigated how positive behaviour spreads and how collective behaviour change can be encouraged.
Inspiring better health through peer-to-peer learning

Experiment 8: Can positive deviance and data-driven segmentation help patients with poor diabetes control learn from those ‘like them’ who are successfully managing their disease?

**Who is behind this experiment?** Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni (IEIIT) at the Italian National Research Council (CNR).

**Key finding:** Indicative findings suggest that segmentation and peer-to-peer workshops can be effective ways of connecting people to share experiences of how they manage their type-2 diabetes. Peer-to-peer approaches may be more effective among those who manage their condition through lifestyle modifications.

**Who is this relevant for?**
- Policymakers in health.
- Medical professionals.
- Technologists in digital health.

**What was the experiment?**
This experiment tested whether successful outliers (positive deviants) and patient segmentation could drive behaviour change among type-2 diabetes patients. It used (unsupervised) machine learning algorithms to segment diabetes patients based on their Electronic Medical Records and whether they prefer lifestyle modifications or medication to manage their condition. Virtual peer-to-peer group workshops involving patients with ‘good’ and ‘bad’ disease control were used to test if patients could learn from the success of others in similar segment clusters.

**What did they learn?**
The segmentation found that patients who prefer lifestyle modifications (who also tended to be older and have a higher number of comorbidities) had better control of their diabetes. While the team struggled to recruit enough participants for the peer-to-peer workshops, those who took part shared positive feedback on their experience of participating. The workshops also indicated that a peer-to-peer approach may be more effective for patients using lifestyle modifications, as patients in this segment reported better learning and higher motivation to set a goal and act in the future compared to the medication segments.

In the 1970s, aid organisations noticed that some parents in impoverished neighbourhoods always seemed to find a way to provide balanced, healthy diets for their children – often by ignoring received wisdom and inventing new meals out of unlikely ingredients. They were early examples of what sociologists came to call ‘positive deviants’, or people who go against the grain of their local community in ways that end up being beneficial to both themselves and others.
“It’s been studied in areas like social dysfunction, education, and addiction,” explains IEIIT-CNR’s Alessia Paglialonga. “And we wanted to look at type-2 diabetes. We know that a small proportion of patients with diabetes can get good control of their biomarkers – blood glucose and cholesterol under control, blood pressure within clinical limits – but the majority find it really difficult.”

For their experiment, the team planned to take an established idea – peer-to-peer patient support groups, where people with similar conditions meet regularly and share their experiences – and place positive deviants (whom they call ‘champions’) inside them. Unlike direct advice from a doctor or friend which might be ignored or waved away, the positive example of someone who has managed to independently figure out how to improve their health can influence others in more subtle and indirect ways. The hope was that the champions’ ‘good control’ techniques – like better sugar monitoring, losing weight, and exercising more regularly – would rub off on the others, and potentially offer a model for the treatment of other chronic diseases beyond diabetes.

The plan was to find around 40 diabetics, sourced from a pool of more than 1,000 patients provided by a network of family doctors in Canada linked up with Queen’s University’s Department of Family Medicine. The patients would be split up into six different peer-to-peer groups, each containing a champion and meeting for around an hour-and-a-half per session. The experiment was designed in collaboration with an expert in peer-to-peer group discussions and market research; four of the six groups had patients considered similar to each other in terms of behaviour and attitude, while the other two control groups contained a more random mix.

“They’d have no idea if there’s a champion or not among them,” says Paglialonga. “It’d be just four to six people sitting in a circle talking.” (Champions also had their role explained to them during the recruitment process, but were otherwise unaware of any other information about who they would be ostensibly helping in each group). The four key groups were also designed to further test whether applying ‘segmentation’ – a concept usually only used in marketing, referring to breaking a target audience down into sub-groups based on key demographic similarities – would also be more effective than a one-size-fits-all approach.

Unfortunately, while preparations began as far back as spring 2020, the search for participants took significantly longer than expected because of the need to secure ethics approval from two different universities in Canada, and to negotiate data confidentiality agreements with relevant authorities. Drumming up enthusiasm was difficult, requiring revisions of their invitation language in letters and phone calls (which in turn required another round of ethics approvals). The COVID-19 pandemic forced the workshops to be redesigned for...
Zoom, too. As of July 2021, only three of the intended six groups have managed to meet – and the team is considering compensating participants for their time, after a number of dropouts and no-shows. The study will likely not finish until late autumn 2021.

“What we have so far is the individual feedback, collected during the workshops,” says Paglia longa. “The good news is that people were excited. Some even asked if they could keep doing sessions every week – they’d wanted to talk with their peers like this already. They suggested improvements, too, like pairing up via a buddy system. Women were also especially enthusiastic, particularly those who had experienced similar support during pregnancy. But, overall, we’re finding that even when people are previously aware of the health information being discussed, it matters hugely in terms of engagement that it’s coming from this different angle of personal experience. Most of our participants reported that they felt they were learning from this programme.”

The team is currently planning future iterations of the study to continue to investigate how interactions with positive deviants can rub off on others – such as adding ‘objective’ health tracker data into the mix, to bolster ‘subjective’ advice. Eventually, through a combination of qualitative and quantitative analysis, the hope is to be able to algorithmically identify meaningful behavioural patterns among groups of patients automatically using AI.

From that, champions can be identified, and further peer-to-peer support groups can be coordinated more easily. They also plan to develop a digital platform for peer-to-peer support and goal setting.

Overall, Paglia longa is generally enthusiastic about what’s been achieved, despite the challenges: “This has required a truly multidisciplinary approach – combining clinical knowledge, biomedical engineering, data analysis, and human behaviour expertise – compared with how siloed research normally is. And I think it shows how, when it comes to health, we still have so much to learn. Individual health comes from a collection of biological components, but it also comes from the influence of collective external factors, like your relationship with your doctors, friends, and family, and yourself. And even though we didn’t have as many people as we’d wanted, it’s important that we found a common way for them to be open about these parts of life so freely, and in turn to be open to what others have experienced.”
How to help groups share scarce resources more fairly

Experiment 9: How do groups make decisions to share limited resources among themselves, and how does behaviour spread across social groups?

Who is behind this experiment? University of Nottingham, RMIT University and University of Tasmania.

Key finding: Different levels of social connectivity influenced the ability of participants to manage the shared resource. Communication (strongly) and connectivity (modestly) enhanced sustainable resource use.

Who is this relevant for?
- Local governments and national policymakers.
- Emergency response teams and others tasked with coordinating the distribution of limited resources.

What was the experiment?
The scarcity of common resources, such as water, is set to become a more pressing and global problem in the future. This experiment explored group behaviour in relation to a common shared resource. The task was a common resource dilemma where participants (members of the public) in groups could individually claim from a common resource but faced the prospect of receiving nothing if the total of their claims exceeded a certain threshold. The experiment took place on an online experiment platform with a total of 470 participants. It tested whether different levels of social connectivity and communication within and between overlapping social groups could improve coordination and prevent resource depletion.

What did they learn?
The experiment results found that different conditions influenced the ability of participants to manage their common resource without exceeding the threshold. Communication (strongly) and connectivity (modestly) enhanced sustainable resource use, and that these two factors mutually reinforce one another. However, external shocks reduced the ability of participants to stay within the threshold.

From overfishing to droughts, it doesn’t matter how aware we are of environmental problems – when everyone takes as much as they want from nature, the collective impact is very often disaster. A team of researchers were curious to see if it might be possible to reduce overexploitation of common resources by improving social connectivity within large groups, so that people would be more likely to act in the best interests of their wider community, rather than themselves as individuals.
“When you’re using any kind of resource, it’s often easier for you to do what you want and cross your fingers that everyone else worries about it instead,” says Nottingham University’s Thomas Chesney. “That is exactly the tragedy of the commons, and we were interested in how social networks might impact it – whether networked people, getting information from different groups they’re a part of, might moderate their behaviour.”

“Essentially, if you give under-connected people better information so they know how ‘the commons’ is being used, will they use less of it themselves?”

We’re all members of many different networks in real life – families, school friends, colleagues, church acquaintances, gym buddies, and so on – and, while information might flow within those networks fairly freely, it can be inconsistent in how it flows between them. For their experiment, the team used a piece of software called ‘The GRID’ to model this kind of overlapping social patchwork as interconnected ‘neighbourhoods’. Developed on spec for the purposes of this study, participants can join a GRID session from anywhere in the world, while the software automatically generates and randomly sorts them with four other people into neighbourhoods behind the scenes. The connections between neighbourhoods can be tuned up or down – from isolated islands of people to very dense, overlapping groups – to test the impact of different degrees of social interaction.

Inside, The GRID looks like a kind of multi-dimensional chessboard, with players linked together like characters in a game. “It’s like, ‘I know you, you know him, he knows somebody else, and therefore I have a connection to them as well,’” explains Chesney. People – students in the first round of experiments, users of Mechanical Turk in the second – were asked about some abstracted, non-specific resource, where even the total quantity was only given as a rough estimate. (Using a specific resource, like fish or oil, would encourage participants to lean on prior biases.) They were then asked the experiment’s key question – how many resources do you want to claim for yourself?

“It’s not just a simple Q&A, because your answers impact everybody else. My outcomes depend on your decisions, and your outcomes depend on my decisions, and if you take too much the whole thing evaporates. Everybody gets zero.” (This was not a hollow threat – participants were paid for their time, and received more for achieving better outcomes.)
The experiment tested the impact of three different factors: social connectivity (with some participants only in one group, while others were in multiple groups); communication (some participants could talk to each other via SMS, while others were siloed); and 'external shocks' (sudden resource shortages, without warning). Throughout, all three factors had clear impacts. Better connectivity and communication both led to more sustainable use of the resources by participants – the latter more strongly than the former – while resource shocks made it harder to avoid exhausting the resources altogether.

The team believe that their results have implications in a number of different fields beyond environmental management – including disaster response, where resource triage is crucial. After publication, the team also plans to publish the code for The GRID so that other researchers can use it. (“It’s a big deal that we helped build a platform where people from everywhere in the world can take part in research like this at the same time – it’s one of the major contributions of this study,” argues Chesney.) For example, The GRID could open up new ways to investigate the psychological concept of ‘weak’ versus ‘strong’ ties: “If you're looking for a job, there's no point asking your best friend – you all know the same people, and you've got the same information,” explains Chesney. “A weak tie is someone that you know, but you don’t know all of their friends. We could potentially manipulate these weak ties with the software and identify them more easily.”

RMIT’s Robert Hoffman points to a passage from Michael Polanyi’s *The Republic of Science* to illustrate how these collective intelligence insights can be further applied, both in the sciences and outside them:

> “How can an aggregate of specialists possibly form a joint opinion? Of course, each scientist who is a member of a group of overlapping competences will also be a member of other groups of the same kind, so that the whole of science will be covered by chains and networks of overlapping neighbourhoods. This network is the seat of scientific opinion, not held by any single human mind, but one which, split into thousands of fragments, is held by a multitude of individuals, each relying on the consensual chains which link them to all the others through a sequence of overlapping neighbourhoods.”
Sustaining behaviour change through collective action on air pollution

Experiment 10: Can a combination of collective environmental assessment and collective action sustain behaviour change among local citizens to decrease air pollution in Tower Hamlets?

Who is behind this experiment?
Umbrellium in collaboration with Loop Labs and Tower Hamlets Council.

Key finding:
Communication and collaboration among citizens led to the sustained adoption of actions to decrease air pollution.

Who is this relevant for?
• Local authorities.
• Citizen science and other collective intelligence initiatives that want to encourage people to participate in a variety of tasks or projects.

What was the experiment?
The experiment tested whether collective environmental assessment and collective action would enable people to sustain behaviour change for actions that are known to reduce air pollution, even though the direct individual effects of these actions on air pollution might not be immediately noticeable. Forty-six people living in East London participated (split into six groups) over a four month period, and were provided with a ‘menu’ of possible actions they could take that are known to improve air quality (e.g. switching the lights off, eating a vegan diet and taking public transport). Participants had to self-report their actions on a weekly basis. The Experimental groups could communicate and collaborate within their groups, while the Control group could not.

What did they learn?
The Experimental groups performed better than the Control group in terms of variety of actions undertaken, adherence to the action and overall calculated impact. While the Experimental groups required more time from the research team and incurred higher SMS charges (for communication), the carbon impact achieved was more than double per unit cost of the Control group (2.22kg offset per £ spent to 1.08kg offset per £ spent).

There has never been greater awareness of the environmental challenges that face the world – but, at the same time, it also feels increasingly impossible for us as individuals to play our part in fixing them. Umbrellium, an urban design studio based in London, has been exploring ways in which to bridge that gap by connecting people together within their local communities, empowering them to understand how their actions as individuals can still translate into effective collective action against systemic issues like pollution.

“This experiment is the latest in a series of similar projects on air quality we’ve done over the last five or six years, called Pollution Explorers Collective Action (PECA),” explains Usman Haque. “It’s about getting people involved in proactively doing stuff to improve air quality. Not just being informed,
but feeling a sense of agency – because when something feels huge and systemic, we can’t get out of that mental loop. And I think that’s one of the biggest problems with air quality – it’s invisible, and anything we do tends to be invisible. The whole point of PECA is to find ways to counteract that discouraging aspect.”

PECA is structured around workshops in target areas. “In cities, normally we actually do know what the problem is, where it is, and even how to solve it,” says Haque. “A local authority will invest money to do something – the typical thing is to put out a bunch of sensors, do a map, and show people where air quality is worse or better.” But these maps are often inaccurate and poorly promoted, and can leave people feeling at the mercy of a problem which seems inescapable. “It removes that sense that you, too, are part of the problem.”

For their Nesta experiment, they engaged with 46 diverse participants in Tower Hamlets and Hackney, East London, for four months from January 2021. (The original start date, in December 2020, was delayed by the third COVID-19 lockdown in England; the participant pool was also reduced from 80, and in-person workshops were moved to Zoom.) Participants were presented with a ‘menu’ of 15 ‘challenges’ they could take on, along with information about their impact on air pollution – from the simplest (switching off lights) to the most impactful (going vegan).

Umbrellium also built an SMS-based social platform, which the participants used to talk with each other outside of their regular catch-ups. It made it possible for them to work together as a team – for example, if someone wasn’t able to spend a day being vegan as planned, they could ask someone else to take it up for them. It converted behavioural changes into a communal experience, which in turn would encourage the participants to feel more engaged and empowered about their ability to influence air pollution. The Control group, by comparison, just had the menu of challenges and regular feedback – “the kind of experience a ‘standard’ top-down local authority-style programme might usually deliver”, says Haque.
"Weekly and monthly, we'd feed back to the experiment groups, 'Oh, your group did this much – if you scale that up to the level of your borough, it would have been the equivalent of removing 20,000 car journeys this month.' We were constantly trying to show that, although this is a systemic thing, small changes at scale have this kind of effect."

Results were positive. The Experimental groups tried more of the challenges than the Control group, and the changes were 'stickier' too: “They were inspired by each other to try new things. Going vegan sounds pretty extreme for someone who isn't, but if somebody tried it for a week and reported success, that percolated back through the group and others would try it. Even if it was just for one meal a week, it was having a greater incremental impact over time. They tended to stick to their behavioural change pledges for longer.”

The Control group was both less enthusiastic and less ambitious – mostly sticking to the easier challenges, like taking shorter showers or reducing food waste. The Experimental groups, by contrast, were increasingly communicative as the experiment went on, creating positive feedback loops of encouragement. The more participants in a group talked to each other, the more ambitious, enthusiastic, and long-lasting were the behavioural changes in their group.

By the end of the project each experiment group had saved an average of 1,600kg of CO2 emissions; for the control group, it was only 400kg. Haque argues that this means the experiment justifies being more expensive than 'typical' local authority measures: “Per pound spent, the experiment groups achieved twice as much as the Control group. It's more bang for your buck.”

The next stage is to scale the model up to include hundreds of participants, using a grassroots-style structure led by local organisers. They also want to expand the PECA SMS platform to other communication platforms, like WhatsApp. “I sometimes talk about the idea of 'mutually assured construction', rather than mutually assured destruction,” says Haque. “How can you take the frictions and dilemmas of cooperation and turn them, so that people can tackle issues together?"

“When we were first looking at the callout for the collective intelligence projects we understood it to be more about algorithms that would work in conjunction with people, but what emerged from this project was different: Using a technological framework to connect people so that their collective intelligence emerges from those connections. And it actually results in a sense of collective agency, which is very different from one where you’re abdicating your responsibility to a piece of technology, like an app or an algorithm.”
Collective intelligence
for better data

Around the world there has been a data revolution driven by advances in information technology. But for many developing countries and many complex issues there are still data gaps. Collective intelligence approaches that involve people in generating or classifying data can help create more localised and real-time information, address bias in existing data sets, and audit or monitor AI systems. Five of our grantees explored collective intelligence for better data.
Bringing the public into AI ethics debates

Experiment 11: Can games be used to engage the public in complex debates about emotion recognition systems and the underlying social research?

Who is behind this experiment? Dovetail Labs.

Key finding: A 'serious game' was an effective way to engage the general public in considering the risks of emotion recognition systems.

Who is this relevant for?
• National policymakers and regulators.
• Technologists developing AI tools.

What was the experiment?
The team sought to research public understanding of and views towards emotion recognition technologies (ERTs), and to spark a more thoughtful public conversation about the potential societal impacts. Dovetail Labs developed a website called Emojify to engage the public in experiencing the limits and implications of ERTs through animation, games, and question prompts.

What did they learn?
During the four-week experiment period, the site received a total of 15,207 unique visits, meeting its aim of increasing public awareness in ERTs and serving as the basis for more targeted deliberative projects on the topic in the future. Participant feedback indicated that the majority remained sceptical about ERT systems as a result of taking part, with greatest concern centred on the (in)accuracy of the technology and its potential to oversimplify the complexity of human emotions.

While facial recognition technologies have become increasingly commonplace over the last decade, an increasingly popular version of it has experts and activists alarmed by its potential to perpetuate social injustice. ERTs claim to be able to use facial expressions to discern a person’s internal emotional state. They’re already having a material impact on human lives, from law enforcement identifying ‘suspicious’ behaviour to students being disciplined for ‘inattentiveness’ in class. Yet public awareness and understanding of ERTs – let alone the controversial science which underpins them – is low.

ERTs, like all forms of AI-based facial recognition, are prone to bias, discrimination, and misidentification when dealing with people who differ from an assumed white European masculine ‘norm’ of appearance.9, 10 However, ERTs are also premised on a theory – first articulated by psychologist Daniel Ekman in the
1960s – that human emotions are biological, rather than differing between cultures. In particular, there are six ‘fundamental’ emotions (joy, sadness, fear, anger, surprise, and disgust) which are not just universally experienced, but also expressed in universally consistent facial expressions. This simplistic theory has been increasingly challenged (if not outright debunked) in recent years – but it hasn’t stopped the wave of ERTs from companies such as Microsoft, IBM, and Amazon which claim to accurately identify those ‘universal’ expressions.11

“It’s a reductionist understanding of human emotions,” argues anthropologist Igor Rubinov. “Its real-world application should be managed incredibly carefully, if not paused indefinitely, while we have a conversation about this.”

“The umbrella issue here is how to bring the communities that are impacted by these technologies into the discussion, so they can have a say,” says Alexa Hagerty, an AI ethics researcher at the University of Cambridge. “Right now, those conversations are almost entirely framed by the people who develop this technology. Machine intelligence may be amazing at pattern recognition, but human intelligence is uniquely reflective and wise. That’s the collective intelligence we wanted to bring into this.”

Recent studies and surveys have found that while there may be widespread awareness of facial recognition technologies, there’s relatively little understanding of how they work – let alone their pitfalls. Through Nesta, the Dovetail team was connected with Alexandra Albert, a citizen science researcher at University College London, to collaborate. Together they decided to build a site with ‘serious games’ – that is, games that are both compelling and informative. Serious games offered a way to reach a broad, global audience, while still encouraging players to reflect on what they were being asked to do, and how their own lives may be impacted outside of the game by the same ideas. For the team, effective citizen science isn’t just about raising awareness. It also has to be something which spurs consideration, and then social action as well.

The site they developed, emojify.info, gives players hands-on experience of interacting with an ERT, framed around the question, “Is emotion recognition technology ‘emojifying’ you?” The first game, ‘Wink/Blink’, shows images of winks or blinks and asks players to guess which is which – thereby demonstrating how an ERT model can’t do the same, because it can’t
understand the contextual information which differentiates the two. ‘Fake Smile’, meanwhile, gets players to trick the model by rapidly moving through different facial expressions, thereby showing how external expressions aren’t necessarily an expression of internal emotional state. After playing, users are asked questions to gauge their new understanding of ERT, and then are provided with a large collection of links for further reading.

The site received more than 40,000 unique hits (15,207 during the experiment period) after being promoted on social media and reported on in countries around the world, including Brazil, Israel, Japan, and India. “During the pandemic, too, we were aware of the issues surrounding research burden and the general sense of exhaustion,” adds Albert. “We had to convey a lot of information, while keeping it engaging and fun. But, overall, there was a lot of shock and concern in the responses – as well as humour.” Overall, participation was significantly higher than they had anticipated.

The results demonstrated that their choice of a wide-reaching, quickfire format was an effective way of soliciting rich qualitative data. The nuanced personal reflections that were collected justified their hypothesis that individuals aren’t easily reducible to simplistic data points or categories when it comes to debating technology ethics.

“We also learned the importance of design,” says Rubinov. “You can’t just throw information at people and say, ‘Here, do something with this’. You have to think through the interactions, especially as remote digital projects like this become more common in the future.”

Players engaged with the moral dimensions of AI in ways that they normally never would, and the team feels that their approach offers a good model for future efforts to engage the public in meaningful democratic discussions of AI ethics – something which will be crucial for policymakers and technologists in the coming years as all kinds of AI, including ERTs, are deployed in the real world. They’re also planning to use the data gathered so far as the basis for designing further (and deeper) deliberative projects in the future.

“We can no longer say that AI ethics is only the domain of specialists. It must involve everyone. There’s a real hunger there – people absolutely want to be involved in this.”

“We can no longer say that AI ethics is only the domain of specialists. It must involve everyone. There’s a real hunger there – people absolutely want to be involved in this.”
### Experiment 12: Can community data and machine learning help reduce cholera outbreaks?

**Who is behind this experiment?** Kenya Flying Labs in collaboration with Kenya Red Cross Society.

**Key finding:** It was not possible to test the accuracy of Kenya Flying Lab’s model. However, it did highlight the potential role that community knowledge and combining data sources can play in disease surveillance and prediction.

**Who is this relevant for?**
- Public health authorities.
- Humanitarian organisations.

**What was the experiment?**
The experiment tested whether a public health surveillance system combining crowdsourced local knowledge, satellite imagery and machine learning could improve understanding of cholera outbreaks in Kenya. It took place in two wards in Kajiado County in Kenya, and the team engaged 20 young people in workshops to gather local knowledge and participate in labelling satellite imagery for potential risk factors for cholera including community water points, public and homestead toilets and waste sites.

**What did they learn?**
In 2020, Kajiado County did not record any confirmed cases of cholera during the course of the experiment, and no historical data exists against which to test the accuracy of the model. Kenya Flying Lab’s model will need further development and testing before it might be able to predict a real-world cholera outbreak.

---

In Kenya, 71 per cent of people have no access to basic sanitation. Diseases like cholera – which are otherwise easily controlled with access to clean water, proper toilet facilities, and covered sewers – spread as a result. Could collective intelligence offer a new way to prevent disease, and save lives? Robotics hub Kenya Flying Labs (KFL) certainly thought so.

“Cholera predominantly affects peri-urban settlements,” explains KFL’s Mohamed Alsayed. “These are places with tinned-roof houses, waste heaps, and open sewerage, and where people live in homesteads with five or six houses sharing outdoor water points, bathrooms, and toilets. We wanted to put together all the information that we already knew about these areas, and see if we could predict cholera outbreaks in advance.”
Such an approach is nothing new – John Snow’s 1854 map of cholera cases in Soho, London, was maybe the earliest example of how data tracking could uncover the source of a disease outbreak. However, working backwards from existing cases is one thing – being able to anticipate those cases before they happen would be far more valuable. “If you’re able to predict an outbreak at point A, and point B is downstream of point A, then you could send oral cholera vaccines to point B ahead of time.”

Since cholera is a waterborne disease, it spreads most readily in areas lacking in indoor toilets, and where the areas where people defecate and get their drinking water might overlap – and for that reason, governments and aid agencies tend to roll out relief efforts in areas prone to outbreaks immediately after heavy rainfall and flooding. But that simplistic approach is both too reactive, and too inconsistent – and the cost when it fails is too high. In Kenya, cholera infects tens of thousands of people each year, with around two per cent of cases leading to death. The KFL team’s hope was that a kind of ‘cholera hyper-surveillance’ – a predictive model kept updated with as much information as possible, as frequently as possible, down to a hyper-local level – would offer a new route to outbreak prevention. Their pilot trial focused on two small neighbourhoods in Kajiado County, in southern Kenya, where cholera outbreaks happen frequently.

Key to making their model work was the decision to bring communities into the data gathering process. The KFL ground teams engaged directly with local people in their target areas via workshops – and it both significantly improved the quality of the data, and generated enthusiasm for the project on the ground. “A lot of the time, it was the difference between wasting a day or two collecting irrelevant data and not,” says Alsayed. Bringing collective knowledge into the data-gathering process is a key difference of the project relative to more traditional top-down approaches.
Team members also went out into the field to directly verify key features that had been identified from satellite imagery – like public toilets and open sewers. Merging all of this information together allowed the team to map what they call the ‘sanitation state’ for each location. “And once we had the locations and conditions for all these points, we could combine them into a qualitative index which was used to train a machine learning algorithm, that in turn made [cholera outbreak] predictions.”

Unfortunately, though, there were a number of issues that held the team back from fully realising their ambitions. Some were due to force majeure: there was no heavy rainfall in Kajiado County throughout the whole of 2020, and thus no cholera outbreaks for the model to predict. The COVID-19 pandemic greatly restricted the number of local residents they could engage with, too. And, perhaps most frustratingly for a robotics lab, one of their key project components never got off the ground – the drones.

“The drones would have helped with acquiring that high-resolution imagery, for mapping infrastructure,” says Alsayed. “But unfortunately we didn’t get our permits in time from the government – so we had to rely on satellite imagery instead, where lots of features are harder to identify. I’m convinced we would’ve seen better results otherwise.” Nevertheless, he thinks that they managed to capture around 80 per cent of important features thanks to efforts to gather information at ground level.

KFL’s model will need further development and testing before it might be able to predict a real-world cholera outbreak. But it still managed to produce some valuable new insights into how factors such as temperature, humidity, topography, and even economic development can influence the chances of an outbreak, which has its own predictive utility. But their concept of ‘cholera hyper-surveillance’ is about more than just the processing of data – it’s about melding the collective intelligence of a community affected by a disease with the digital intelligence of machine learning.

“Planning and conducting this experiment helped us to put a name to something that we had already been subconsciously doing,” explains Alsayed. “We now have a more structured approach to collective intelligence, and we’ll continue to practise it within our humanitarian work.”
Experiment 13: Can social matchmaking on a citizen science platform lower the barriers to patient-led research?

Who is behind this experiment? Just One Giant Lab (JOGL) and Open Humans Foundation (OHF).

Key finding: The collective design and implementation of a health tool resulted in high levels of engagement among users and low dropout rates compared to other mHealth tools.

Who is this relevant for?
• Patient communities and health researchers.
• Anyone using a collective intelligence approach that involves crowdsourcing or matching individuals to specific tasks or projects.

What was the experiment?
The original experiment design aimed to test whether matchmaking between skills and needs among patients on a collective intelligence platform would increase engagement in patient-led research projects. Patients were to receive targeted notifications (an email) related to specific projects – including information about the required skills, location, and topic. The aim was to understand if engagement is driven by these different factors. To run the experiment, the team combined JOGL (a platform for DIY biology-makers), and Open Humans (a patient-led research platform), enabling the two groups to collaborate to develop solutions for managing health conditions.

What did they learn?
Prior to the email notification intervention, the community organically established an OpenCovid19 programme and began working on COVID-19 related projects. One project culminated in the development of Quantified Flu, a health tool for monitoring physiological symptoms (e.g. increased heart rate and temperature) as an early warning for COVID-19. The collective design and implementation of Quantified Flu resulted in highly engaged users and low drop out rates compared to other mHealth tools. The matchmaking system found that directing participants with specific skills to projects that require that skill (e.g. coding) significantly increased the likelihood of the participant engaging with the project. Project popularity (i.e. other participants indicating their support, such as through ‘likes’) and project activity (number of posts) also increased engagement.

For many people with chronic illness, they know more about their health than any doctor. Patients with long-term health problems have always banded together to support one another, and at times even produce their own research. This dynamic has accelerated in the social media era, yet the larger a community gets, the faster it can lose focus – and momentum. JOGL and the OHF wanted to fix that.

“There are some very successful patient-led research projects out there,” explains OHF’s Bastian Greshake Tzovaras. “But they tend to be happy little accidents. No-one’s looking at how these groups succeed, but they exist because patients are interested in improving their quality of day-to-day life, more so than a lot of institutions.”
OHHF's Open Humans platform was launched in 2015 as a community hub where people can aggregate and share their medical data – but it lacks a way for users to share their skills and connect on that basis. JOGL, on the other hand, acts as a social network where people can meet and launch open research projects, as well as secure micro-grants to fund them. Both organisations are premised on the idea that even if a community contains a huge amount of expertise, it'll go to waste if its members don't self-organise to make best use of it. “So we thought it would be interesting to combine the two approaches, by improving collaboration between people who are conducting medical research projects,” says JOGL’s Marc Santolini.

Their experiment was intended to connect users from both platforms together, and create a kind of ‘matchmaking’ service for patient-led groups in need of members with vital expertise on how to organise and run research projects. “Can we go from a few happy accidents,” asks Greshake Tzovaras, “to scaling this for other patient communities? Open source software typically starts from an individual building something to address their own need, and we wanted to similarly support a broad range of people to answer their own research questions. It could have been hormone treatments during gender transition, or hayfever remedies, or people with type-1 diabetes investigating nutrition – but then COVID-19 happened, and our focus shifted immediately.”

It steered them towards building what became the OpenCovid19 Initiative – an umbrella for a number of new community initiatives which formed during the early months of the pandemic, in 2020. “It helped that we already had around 9,000 people enrolled on Open Humans,” says Greshake Tzovaras. “But the thing with COVID-19 is that it affected everyone, and I think that’s why we had the luck of seeing a lot of people who traditionally wouldn’t fall within the ‘patient’ category becoming involved.” Its participants met and collaborated online and off, both directly on JOGL and Open Humans, and via platforms like Slack and GitHub.

“We’re trying to redesign a framework that’s defined the past two centuries...one where you need the journal, the peer review, the grant system – instead, we’re empowering the community.”

The team studied two OpenCovid19 sub-projects in particular. The first, Quantified Flu, sought to crowdsource medical data from users’ wearable devices and feed it into a new tracker tool which could identify common early signs of COVID-19, flu, and common cold infections. “They brainstormed what they would like to do – and we, as facilitators through JOGL, found a number of software and web developers who came in and supported the development of the tools the patients wanted,” says Greshake Tzovaras. “It was the best-case scenario, really.”
The second project – CoughCheck – was less successful. The plan was to similarly crowdsource audio of people coughing, and train a machine learning algorithm that could then work as a diagnostic tool – people coughing into their phones to see if they had COVID-19. The concept failed to work, while the participants were also mostly the ‘wrong’ people, with too many marketers and project managers, and too few developers – itself a demonstration of the importance of matchmaking skills correctly within patient-led communities. The wave of early enthusiasm was followed by burnout across OpenCovid19, too – more than 3,000 people were taking part simultaneously in March and April 2020, but within a few months many had dropped out.

The matchmaking was also intended to have been automatic, but the pandemic struck before any system could be built and the job was largely performed by humans instead. It took until the end of February 2021 for the automatic system to be developed and tested, but after it was deployed the team was able to study the factors determining why some groups were more likely to attract new members than others. In particular, more popular projects (with a visibly excited and engaged membership) would in turn encourage even more people to become involved, in a similar way to how successful crowdfunding campaigns operate. Volunteers were also drawn to communities where they could see others like them, with similar skills, already taking part. They’re insights which can be applied to helping other self-organised patient-led groups in the future.

“ln a way our project benefited from this chaos,” adds Greshake Tzovaras. “With these newly formed communities, you don’t have to design for serendipity – it happens by default – but as things calm down that also makes collective intelligence harder. As communities become more mature, they typically close down to newcomers and get stuck in set workflows. Keeping them in an open-minded state is another challenge we need to explore.”

“In a way, we’re trying to redesign a framework that’s defined the past two centuries,” says Santolini. “One where you need the journal, the peer review, the grant system, etc – instead, we’re empowering the community, and in the long-term we’ll see more facilitation of micro-volunteering between researchers and patient communities.”

“I think that’s the right direction – decentralising science, decentralising matchmaking, avoiding the concentration that happens now in funding and journals, and at every stage getting yourself as much as possible out of your bubble.”
Crowdsourcing weather and pest alerts in the Andes

Experiment 14: Can the crowdsourcing of climate data by farmers create a more effective meteorological alert system to help Bolivian farmers better adapt to extreme weather events?

Who is behind this experiment? Swisscontact in collaboration with Banco de Desarrollo Productivo and Latin American Centre for Rural Development (RIMISP).

Key finding: Crowdsourcing increased farmers’ engagement in early warning systems, meaning they were more likely to take action to prevent pest outbreaks.

Who is this relevant for?
- Development agencies.
- Humanitarian organisations.
- Anyone relying on up-to-date or hyperlocal information (e.g. in disaster prevention, climate change adaptation, or emergency response).

What was the experiment?
The team tested whether the local climate data and crowdsourced data on crop diseases and pests can create a more effective alert system to help rural Bolivian farmers better adapt to extreme weather events. Three communities in remote Bolivia were equipped with low-cost weather stations that received hyperlocal weather data and access to an additional crowdsourcing system that enriches and validates pest and diseases outbreak alerts.

What did they learn?
The results demonstrated that temperature and humidity forecasts using local information through the low-cost weather stations are more accurate than official available forecasts. In addition, farmers that crowdsourced reports on pests and disease outbreaks were more likely to act upon the information they receive than farmers who are not involved, including being more engaged in crop monitoring, irrigating their crops more and applying more pest repellent. The team did not find any statistically significant results in terms of crop loss.

The Andes’ microclimates can cause wildly different weather conditions in areas only a few miles apart. However, the world’s climates are changing – and many smallholder farmers, who for centuries have relied on the predictability of their specific local weather, are struggling to adapt. Convincing farmers to put their trust in modern technology (and meteorology) would help, but that’s far easier said than done. Instead of imposing new solutions from above, this experiment sought to make the farmers themselves a part of the solution to their own problems – collective intelligence applied in the pursuit of the democratisation of weather forecasting.

“We tested if crowdsourced weather forecasting was feasible in rural farming communities in Bolivia,” explains RIMISP’s Rafael Lindemann.
While 80 per cent of rural farmers in Bolivia have smartphones, their research has shown that they’re almost exclusively used for calls, or messaging with services like WhatsApp – weather forecasts aren’t accessed at all, because they’re just not accurate enough down to the level of smallholder farms.

“And with climate change, patterns of rain are changing and farmers don’t know what to do,” says Swisscontact’s Franz Miralles. “These communities are suffering from a lack of reliable information.”

Their plan was to take weather forecasts from an existing service (Meteoblue) and improve it with data collected from small weather station units installed across multiple farms. Further data – both microlocal weather reports, and information about pests and disease outbreaks – would also be contributed by farmers via a crowdsourcing platform called KoboToolBox. Better forecasts, tailored to small microclimates, would then be possible, and help them (and other farmers) avoid simple mistakes like irrigating crops before rainfall, which might lead to floods, crop failure, and pests swarming on standing water. They would also be able to get accurate alerts about more severe weather, like flooding. But getting farmers to engage took some convincing.

“Beyond the pandemic, we encountered various difficulties,” says Lindemann. “Political unrest made farmers wary of our presence. It was also hard to engage with women, as their use of mobile phones is more limited – but we had to find ways to encourage women’s participation, as their engagement is key in promoting collective action across whole communities, rather than just with local male leaders.”

The key to breaking through was gaining farmers’ trust, and showcasing the reliability and utility of the forecasts and alerts. “They could see the cost-benefit ratio flip,” says Lindemann. “Being able to plan ahead was huge, and seeing that they were changing their habits was an incredible achievement.” Even though there was no statistically significant difference between participating farmers and their neighbours when it came to crop loss, they still found that being part of the crowdsourcing system made them much more likely to listen to alerts about things like pest outbreaks, and take appropriate preventative measures. The programme was such a success at generating enthusiasm that ten neighbouring communities have since requested to also take part in this collective intelligence network, so they too can receive better weather forecasts and alerts.
Still, there were also significant complications. Beyond the inevitable – and unforeseeable – problems that emerge when deploying prototype tech in the field for the first time, Bolivia’s COVID-19 lockdown kicked in at just the wrong time. “We lost the winter,” says Miralles. “The alerts we were planning were mostly for winter weather, and the summer is much more complicated in terms of heatwaves and plagues.”

The cost of the weather stations was also more prohibitive than had been expected – $2,000 per unit, instead of $500 in the US or Europe – because there were no manufacturers within or near Bolivia, and import taxes were high. It made breakdowns more costly and time-consuming to fix, too. “We eventually found a Bolivian entrepreneur, Jorge Poma, who learned how to manufacture them from YouTube videos,” says Lindemann. “We’re partnering with him to improve our supply chains for the future. Seventy per cent of our effort was spent on these devices, and using their data to enhance Meteoblue’s forecasts. For the scaling-up process we’re connecting our weather stations to a crowdsourcing-based alternative called Weather Underground instead. It’ll be a great step forward, with cost-free hyperlocal forecasts for each station, and a more efficient process overall.”

Regardless, the enthusiasm of the farmers has convinced the team that projects like this have significant potential to disrupt the development sector in South America. “We’re often working with partners within the system – NGOs, development banks,” says Miralles. “But, for me, collective intelligence like this is evolving development work in new ways. Having local people involved is important, and they’re not just receiving information, they’re also making it more reliable. So this isn’t just an experiment – I think it’s helping our partners to see the light. We’re changing the system.”

Lindemann agrees: “Collective intelligence has the potential to transform the way information flows. In conventional development efforts experts transfer knowledge and information to farmers, but farmers often don’t adopt the recommendations, partly because they don’t apply to their specific context. Instead, the information generated in collective intelligence systems like this is distributed, making it locally relevant – and it’s also participatory, making farmers more engaged.”

“A participant farmer applying pesticides to lettuce crops

“Having local people involved is important, they’re not just receiving information, they’re making it more reliable. So this isn’t just an experiment...we’re changing the system.”

“We have to democratise this data, and this much more horizontal, two-way relationship causes a power shift. And it also causes these farmers to realise that they can use their phones for more than just WhatsApp. They become the cells in this new system of crowdsourcing, and that unlocks so many more other new opportunities.”
Harnessing community feedback and AI to validate tools in humanitarian operations

Experiment 15: Can collective intelligence improve the tools that support humanitarian response by involving affected communities and AI to validate tool outputs?

Who is behind this experiment? International Organization for Migration (IOM).

Key finding: Affected communities and Key Informants in Iraq had diverging views towards some aspects of return conditions covered by the Return Index (RI)\(^b\). This highlights the value of widening participation in the return planning process.

Who is this relevant for?
- Development agencies and humanitarian organisations.
- Local authorities.

What was the experiment?
IOM piloted the use of collective intelligence to assess how the Return Index for Iraq could be improved by involving affected communities and AI. The Return Index is a tool designed to measure the severity of conditions in locations of return and is based on information provided by Key Informants. Affected communities were engaged through a digital platform and asked whether they agree or disagree (or to provide a comment) with the Return Index scores across a range of indicators. A pre-trained machine learning algorithm was used to analyse polarity in responses.

What did they learn?
While the platform didn’t garner sufficient responses needed for significance testing, it did indicate differences in opinion between affected populations and Key Informants. There was mutual agreement on the status of security, housing and employment, but disagreement on questions around reconciliation and business.

Since 2014, more than three million Iraqis have been displaced within the country due to violent conflict.\(^{14}\) Now that the presence of Islamic State of Iraq and the Levant (ISIL) has been driven out, efforts are underway to undo the disruption caused to millions of lives. However, after so many years of being uprooted, can the families who were displaced from their homes reclaim the lives they left behind?

Responding to this kind of humanitarian challenge is why the International Organization for Migration (IOM) was established in 1951, and its staff have decades of experience in aiding internally displaced people (IDPs).

b. IOM’s Return Index collects data on areas displaced Iraqis are returning to, to predict locations where they are most likely to resettle. It is based on a network of 2,000 Key Informants (including community leaders, local authorities and security forces) across locations of return.
Gathering and analysing on-the-ground data about returning IDPs is an essential first step in managing the after-shocks of conflict. It means funding and labour for reconstruction efforts can be targeted where it’s needed most. Otherwise, essential services – what’s left of them – can collapse under spikes in demand.

“People are only going to return home if their kids can go to school, if they can get access to healthcare,” explains IOM’s Robert Trigwell. “When conditions are better than where they’ve been staying – a relative’s house, a hotel, a camp, etc – then they’ll make a call if the journey is worth it. It’s the post-humanitarian challenge, in a way – the rebuilding and recovery phase.”

IOM’s Iraq Return Index is designed to reflect these realities down to the granular level of local administrative wards. Data is gathered across 16 categories – relating to livelihoods, basic services, safety perceptions, and social cohesion – by aid staff working with a network of over 2,000 Key Informants, like local community leaders, civil servants, and members of the security forces. For policymakers, NGOs, and the Iraqi authorities, it’s a guidebook – it takes a complex range of interacting factors and condenses them down to a simple green-amber-red traffic light system, to show whether an area is safe to return to. However, it’s an imperfect approach. “We’re communicating this information to charities, local and federal governments, donors, the UN, and others involved in the return effort – it’s a wide array of different audiences. But it’s all one-way communication. We take that information from the Key Informants, process it, but we often don’t utilise the knowledge of local communities to validate our assumptions.”

IOM’s experiment was designed to see if they could make use of digital platforms to get input from a diverse range of IDPs. By making the process of building the Index more participatory and democratic, it could also provide their data scientists with opportunities to validate their existing Index models. Bringing affected communities into the data collection and analysis process is something that rarely happens in the NGO world, where there’s a “top-down power imbalance. It’s a kind of community calibration of our model.”

Aware that smartphone ownership levels are high in Iraq, for their pilot project the IOM team developed a mobile survey site. Focusing on five of the Index’s key areas – schools, health, education, livelihoods, and safety and security – responses and location data were anonymised, and the questions were geared toward letting people disagree with the ‘official’ assessment of their area. “In other data initiatives, we often only hear from people who can speak on behalf of their community, like mukhtars,
local leaders,” explains Trigwell. “So we may not capture youth voices, or certain gender dynamics. An elder male might say their community is safe, but would an 18 year old girl agree? They might say they’ve cleared ordnance from the fields, but do the farmers agree? Maybe the school is listed as ‘rebuilt’, but actually only the school for teenagers. You go through the questions and agree or disagree with the views of the Key Informants, and explain why.”

The survey ran online for six weeks, from late March until May 2021. Respondents saw the unique results for their local ward already contributed by Key Informants, and their additional opinions then formed a kind of ‘peer review’ of the Index, based on their actual lived experiences. IOM data scientists then used NLP on that data – both quantitative (agrees and disagrees) and qualitative (their reasons why) – to validate the Index. While the Index matched what was happening on the ground with regards to security, housing, and employment, it also revealed disagreements about reconciliation and business.

There were a number of challenges, however, as the COVID-19 pandemic limited their ability to engage with people face-to-face, and so the survey was only offered to the 100,000 followers of IOM Iraq’s Facebook page. Approximately 700 respondents participated, and, while it provided a good snapshot of the perspectives of many IDPs, the low response rate limits how useful it can be for further statistical analysis.

Nevertheless, the IOM team still learned valuable lessons when it comes to trying to use digital technologies to encourage input and participation from a wider range of people. For some IDPs, anonymity empowers them to make their opinions known. For others, they’re still more comfortable talking face-to-face, and a mixture of digital and in-person techniques will be more appropriate in future. More importantly, however, the pilot marks the first step towards a broader redesign of these kinds of information-gathering systems within the NGO world – both in terms of bringing affected populations more directly into the systems that influence their lives, and in terms of how community feedback can be used to design fairer, more inclusive data models for tools like the Index.

Trigwell sees the project as a successful pilot which could be expanded in the future – both within Iraq and beyond. “I’ve worked in this sector for ten years,” says Trigwell. “We talk about engaging affected populations, but it’s kind of like this buzzword, and is more complex than just asking people for consent to collect their data. How we use that data has to be participatory, and we have to be accountable. This is the first time I’ve actually seen some sort of peer review on our data analysis processes – and the peers are the affected populations.”
Support is needed to establish the right partnerships to do collective intelligence well

Collective intelligence projects typically require a diverse range of technical and people skills. Ensuring the right range of capabilities will (in many cases) require partnerships – for example, between academic organisations and NGOs/public bodies, to leverage the full range of technical skills and expertise to make projects a success. This was clearly demonstrated by the range of partnerships needed to make the experiments in our programme possible, and the additional brokerage of new relationships that Nesta provided to many grantees. They are also important to testing and scaling cutting edge collective intelligence methods developed in lab-based environments in a real-world context. There is an important future role for Nesta and other funders to play in facilitating and enabling successful collaborations – through co-creation opportunities, match-making and funding.

More research is needed on how to effectively recruit participants and sustain engagement

Many of our experiments struggled with recruiting a sufficient number of participants, or maintaining engagement in the project once recruited. Some grantees experimented with different marketing approaches or incentives (financial and non-financial), but there is a significant gap in understanding how to effectively tap into public motivations for participating in collective intelligence projects and getting incentives right. Given the centrality of public participation to collective intelligence initiatives, this is an area in need of further research.

In the experiments led by Dovetail Labs and Spotlab, serious games emerged as a particularly effective tool to engage the crowd; both experiments received very high levels of public engagement from around the world. Setting participants a ‘challenge’ to complete also led to sustained engagement over the longer time frames, as was the case with Umbrellium’s experiment. Public engagement in these experiments may also have been driven by the salience of the experiment topic due to project timings (i.e. Spotlab’s health experiment during a pandemic) and media attention (i.e. media coverage of facial recognition technology driving participation in Dovetail Labs’ experiment). Understanding how non-financial engagement mechanisms (e.g. gamified approaches) and triggers (e.g. public sentiment) can be leveraged will be essential to scaling collective intelligence projects.

Targeted investment is needed to fund innovation in tools for collective decision making

During the application process for the second and third rounds of Collective Intelligence Grants, we have seen growing numbers of proposals that seek to use a crowd for gathering or classifying data to train machine learning models. Although these experiments are important, we believe there is currently a
critical innovation gap in tools for collective decision making. The recent proliferation of citizen assemblies demonstrates the appetite for new methods of decision making, but this has not yet translated into the development of new digital tools.

The findings from a number of the experiments testing tools to improve collective decisions indicate potential directions and applications for future research. The University of Nottingham’s experiment tested how altering network connectivity and communication influences group behaviour around a common resource. Birkbeck’s experiment similarly demonstrated the importance of connectivity for group decision making, by manipulating connections between individuals resulting in more accurate forecasts. While the University of Bristol explored the role that robot swarms can play in large group decision making scenarios. More research is needed to optimise these techniques and tools, and test them in different contexts and with different audiences.

Collective intelligence can be a powerful approach to increase decision accuracy, legitimacy, and ensure that opinion diversity is reflected in decisions. These experiments have tended to explore this in lab-based environments, however we need to test how this translates in real-world contexts. Dedicated funding should also support R&D in AI-enabled tools that can help increase the scale of deliberation, overcome biases, increase the complexity of issues that can be dealt with, and reduce the cost and time required to do this.

**Practical experience is needed to understand how best to integrate collective intelligence tools into established workflows**

A number of our experiments highlighted that integrating new tools and AI into established workflows can be disruptive, rendering them less helpful than anticipated. For example the Humanitarian OpenStreetMap Team found that the AI-assisted editor reduced the mapping speed and quality of mapping data among advanced mappers. And neu (Augmented Thinking) found that introducing distantly related suggestions early on confused ‘novice’ users of their serendipity-inducing recommendation algorithm – allowing users time to familiarise themselves with the tool was important for realising the tool’s benefits and for it being perceived as helpful. The full potential of these tools may not have been achieved in these experiments due to users needing more time to adjust to using them, or requiring more information on how to use them. When delivering collective intelligence initiatives, user research could be leveraged to understand how best to integrate tools into the process, or additional training may be required to empower users (and participants) with the knowledge to effectively use new tools and techniques.

In the same vein, these experiments demonstrate how the results varied between participants depending on their experience of a particular practice (i.e. advanced versus beginner mappers in the case of the Humanitarian OpenStreetMap Team experiment) or familiarity with a new tool (i.e. those who were new to the serendipity-enhancing recommendation system in neu’s experiment). Users are not a homogenous group who can be expected to use and interact with a new tool (or an AI) in the same way; greater consideration for the multiplicity of experience may be needed when integrating new tools into workflows in the future.
More research is needed to design and test systems that enable positive collective behaviours

A number of the experiments in this programme explored the role of digital tools and technology in helping to drive positive collective behaviours. Umbrellium and IEIIT-CNR both explored whether accountability between peers could lead to sustained individual behaviour change in the contexts of air pollution and diabetes management respectively. In addition, the University of Nottingham tested how different levels of social connectivity influenced the behaviour of groups and their ability to manage a shared resource. Future research in this field should continue to explore systems and tools that can drive positive collective behaviours. Overcoming the societal challenges of our time, from climate change to future pandemics, will undoubtedly require tools that can support changes in collective behaviour and collective social action.

Further exploration is needed to develop cooperative human-machine systems

A number of recent studies have shown that AI-only teams often outperform human-AI teams. And this appeared to be the case for some of the experiments conducted in this programme. For example, in Samurai Labs’ cyberviolence experiment the human volunteers tended to respond less creatively than expected, resulting in an AI-bot being deemed more fitting for the task of delivering verbal interventions.

However, there are widespread (and well known) risks to AI, and growing concern around leaving the design and deployment of ‘AI for Good’ systems to technologists alone. As Stanford University’s AI100 Study Panel Report notes, “Perhaps the most inspiring challenge is to build machines that can cooperate and collaborate seamlessly with humans and can make decisions that are aligned with fluid and complex human values and preferences.” This will require innovation in cooperative AI systems that matches or overtakes the interest in adversarial AI, and in particular, gives more attention to AI and human cooperation in the context of group problem solving. Some of the most important challenges we face in society rely on collective actions and decisions in the face of competing priorities – it is time we invested in the technology that can help us collaborate better to improve our chances of success.
Conclusion

The breadth of the experiments supported through Nesta’s Collective Intelligence Grants programme illustrates the diverse applications of collective intelligence and its potential to solve complex societal challenges.

Nesta’s Collective Intelligence Grants programme has created a space for organisations to test new ideas in collective intelligence, and specifically on crowd-machine cooperation. It enabled non-profit organisations, universities and companies to develop new and unique approaches and collaborate in new ways. Despite its potential, collective intelligence design is still a nascent area for research funding and is dwarfed by investments in AI.

Funding for experimentation is crucial for accelerating learning in the field, to push the boundaries of existing practice and knowledge. This year, Nesta announced a third round of funding through the Collective Intelligence Grants programme. For this round, we are delighted to co-fund three experiments with a total funding pot of £90,000. However, it remains the case that there are currently no dedicated large-scale funding opportunities in the UK for collective intelligence research and development. Our hope is that these experiments may prompt much bigger funders to direct their resources to collective intelligence.

Funders who want to solve the complex global challenges of our time from climate change to misinformation will recognise the urgency of making progress in how we, as humans, understand, think and act together. Collective intelligence is about combining the best of human and machine intelligence to do that.
# More information on the experiments

For more information about the experiments, please reach out to the grantees directly.

<table>
<thead>
<tr>
<th>Grantee</th>
<th>Contact</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Humanitarian OpenStreetMap Team</td>
<td>Bo Percival</td>
<td><a href="mailto:bo.percival@hotosm.org">bo.percival@hotosm.org</a></td>
</tr>
<tr>
<td>2. Istituto di Scienze e Tecnologie della Cognizione (ISTC-CNR)</td>
<td>Vito Trianni</td>
<td><a href="mailto:vito.trianni@istc.cnr.it">vito.trianni@istc.cnr.it</a></td>
</tr>
<tr>
<td></td>
<td>Ralf Kurvers</td>
<td><a href="mailto:kurvers@mpib-berlin.mpg.de">kurvers@mpib-berlin.mpg.de</a></td>
</tr>
<tr>
<td>3. neu (Augmented Thinking)</td>
<td>Johannes Mutter</td>
<td><a href="mailto:johannes.mutter@network.rca.ac.uk">johannes.mutter@network.rca.ac.uk</a></td>
</tr>
<tr>
<td>4. Samurai Labs</td>
<td>Patrycja Tempska</td>
<td><a href="mailto:patrycja@samurailabs.com">patrycja@samurailabs.com</a></td>
</tr>
<tr>
<td>5. Spotlab</td>
<td>Lydia Garcia</td>
<td><a href="mailto:lydia@spotlab.org">lydia@spotlab.org</a></td>
</tr>
<tr>
<td>6. Birkbeck, University of London</td>
<td>Ulrike Hahn</td>
<td><a href="mailto:u.hahn@bbk.ac.uk">u.hahn@bbk.ac.uk</a></td>
</tr>
<tr>
<td></td>
<td>Jason Burton</td>
<td><a href="mailto:jburto03@mail.bbk.ac.uk">jburto03@mail.bbk.ac.uk</a></td>
</tr>
<tr>
<td>7. Bristol University</td>
<td>Sabine Hauert</td>
<td><a href="mailto:sabine.hauert@bristol.ac.uk">sabine.hauert@bristol.ac.uk</a></td>
</tr>
<tr>
<td>8. Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni (IEIIT-CNR)</td>
<td>Alessia Paglialonga</td>
<td><a href="mailto:alessia.paglialonga@ieiit.cnr.it">alessia.paglialonga@ieiit.cnr.it</a></td>
</tr>
<tr>
<td>9. University of Nottingham</td>
<td>Thomas Chesney</td>
<td><a href="mailto:thomas.chesney@nottingham.ac.uk">thomas.chesney@nottingham.ac.uk</a></td>
</tr>
<tr>
<td>10. Umbrellium</td>
<td>Usman Haque</td>
<td><a href="mailto:usman@umbrellium.co.uk">usman@umbrellium.co.uk</a></td>
</tr>
<tr>
<td>11. Dovetail Labs</td>
<td>Alexa Hagerty</td>
<td><a href="mailto:alexa@dovetaillabs.com">alexa@dovetaillabs.com</a></td>
</tr>
<tr>
<td>12. Kenya Flying Labs</td>
<td>Mohamed Akasha</td>
<td><a href="mailto:mohamed@flyinglabskenya.or.ke">mohamed@flyinglabskenya.or.ke</a></td>
</tr>
<tr>
<td>13. Just One Giant Lab and Open Humans Foundation</td>
<td>Bastian Greshake Tzovaras</td>
<td><a href="mailto:bastian@openhumans.org">bastian@openhumans.org</a></td>
</tr>
<tr>
<td></td>
<td>Marc Santolini</td>
<td><a href="mailto:marc@jogl.io">marc@jogl.io</a></td>
</tr>
<tr>
<td>14. Swisscontact</td>
<td>Franz Miralles</td>
<td><a href="mailto:franz.miralles@swisscontact.org">franz.miralles@swisscontact.org</a></td>
</tr>
<tr>
<td>15. International Organization for Migration</td>
<td>Robert Trigwell</td>
<td><a href="mailto:rtrigwell@iom.int">rtrigwell@iom.int</a></td>
</tr>
</tbody>
</table>
Endnotes

5. https://www.humandx.org
11. https://emojify.info