AI reflections in 2021

For a third year in a row, we followed up with authors of several recent Comments and Perspectives in *Nature Machine Intelligence* about what happened after their article was published: how did the topic they wrote about develop, did they gain new insights, and what are their hopes and expectations for AI in 2022?

Cameron Buckner

23 November 2021; Buckner, C. Understanding adversarial examples requires a theory of artefacts for deep learning. *Nat. Mach. Intell.* **2**, 731–736 (2020)

What was your Perspective about?

Recent experimental results have shown that the 'non-robust' features which render deep neural networks vulnerable to adversarial attack may predict category labels in natural data, despite these features being inscrutable to humans. I argued that although this is an intriguing result, these features cannot be relied upon for trustworthy inferences until we determine whether they are processing artefacts. In general, artefacts (such as Doppler effects or lens flares) do carry information about signal sources, but can produce errors if overinterpreted.

Was there a specific motivation to write the article?

I found that the debate over non-robust features often deadlocks on a simple dichotomy: whether they reflect signal or noise. The concept of an artefact is, interestingly, in between these two extremes. Artefacts capture real patterns in the signal source which may be useful, but they can also lead to erroneous inferences in the way they distort or exaggerate those patterns. These distortions can be harmless or even beneficial for some purposes, but disastrous for others. I was particularly intrigued about how this issue intersected with one of the most profound questions in the last two centuries of philosophy of science: what makes a feature 'real' or 'projectible' onto future cases in inductive inference? The recent successes of inscrutable deep learning suggest that this question may no longer be arbitrated by the limitations of human cognition. But this raises the disturbing possibility of scientific progress that by definition does not extend human understanding of the natural world.

Did you get any surprising feedback?

I have found that the idea of artefacts that are distinctive to deep neural network architectures resonated with many

Credit: ElenaBs / Alamy Stock Vector

audiences. There are now more ideas about how, for example, the hyperparameters for convolution operations might interact with high-frequency features to create processing artefacts. Many audiences have suggested interesting potential sources of such artefacts, such as subtle repetitive head movements in functional MRI data.

What are your hopes or expectations for AI for 2022?

I hope that public-facing debates over deep learning will stop oscillating between two extreme views: on the hand that the Singularity is imminent (it's not) and on the other that deep learning is mere 'statistics' or 'curve-fitting', of no relevance to intelligence. The brain can also be described as mere 'neural firing'. What matters is how the statistical processing or neural firing is organized and what behaviour can be accomplished with it. We have ample evidence from psychology and neuroscience that deep neural networks are interesting models of some aspects of human intelligence—specifically perceptual abstraction—but also that something more must be added if they are to scale up to higher cognitive processes. I expect that architectural innovations—in particular, combining multiple modules that play roles attributed to different cognitive faculties such as memory, imagination and attention—will unlock further progress towards more human-like intelligence.

Risto Miikkulainen and Stephanie Forrest

18 January 2021; Miikkulainen, R. & Forrest, S. A biological perspective on evolutionary computation. *Nat. Mach. Intell.* **3**, 9–15 (2021)

What was your Perspective about?

Evolutionary computation is a form of computation inspired by Darwinian evolution in natural systems. In this article, we evaluated how closely evolutionary computation today captures what is known about biological evolution. We identified opportunities to improve evolutionary computation and also places where biological understanding falls short.

Was there a specific motivation to write the article?

In a recent survey article by several members of the community, it was found that evolutionary computation often discovers surprising solutions. However, to rival biological evolution in creativity, it seemed that surprise alone is not sufficient: the solutions need to be useful, economical and robust at least. That led us to look more carefully into the similarities and differences between computational and biological creativity.

How has the topic developed over 2021?

In July 2021 we organized a workshop sponsored by the Santa Fe Institute on the Frontiers of Evolutionary Computation as

part of a program led by Melanie Mitchell, Melanie Moses and Tyler Millhouse on the Foundations of Intelligence. The workshop brought together evolutionary biologists, evolutionary computation experts and various philosophers, computer scientists and biologists. Several speakers at the workshop questioned the value of computation as a model of natural systems, for example, by arguing that evolutionary processes cannot be cleanly separated from the physical substrates in which they are embedded. And a number of speakers highlighted the many ways in which intelligence cannot be separated from evolution.

How has your thinking on the topic evolved?
It has become clear that innovation in biology arises under conditions that are typically different from those currently used in evolutionary computation: instead of elite solutions, there are large populations; instead of strong selection, there is weak selection and neutral changes; instead of measurable benchmarks, there are multiple objectives and extensive time. It is a fascinating challenge to bring such thinking into evolutionary computation.

Has the COVID-19 pandemic affected your research?
Fortunately, in computational fields, many activities can be conducted online, and much of our work was already virtualized before the pandemic. Online activities make broader participation possible: we can more easily attend a variety of talks and conferences, and research teams can include people from different locations, which is inspiring and may promote creativity. On the other hand, everything is planned and scheduled, so serendipitous connections are less likely, and deep dives at the whiteboard are harder. We hope to regain some of the advantages of working face to face in the future.

What are your hopes or expectations for AI for 2022?
We hope that this is the year that the important role of evolutionary computation in AI is recognized more widely, just as scientists have recognized how closely natural intelligence and evolution are intertwined. We also hope to see more large-scale experiments with evolutionary computation, beyond the impressive results that have been achieved in evolving deep learning architectures.

Silvia Milano
17 June 2021; Milano, S., Mittelstadt, B., Wachter, S. & Russell, C. et al. Epistemic fragmentation poses a threat to the governance of online targeting. *Nat. Mach. Intell.* **3**, 466–472 (2021)

What was your Perspective about?
Online targeted advertising is consumed individually, and as such isolates individual consumers. This produces a phenomenon that we call epistemic fragmentation, which has the effect of making it more difficult to identify harms caused by online targeted advertising, especially cases of unfair exclusion from positive opportunities, or targeting based on behavioural vulnerabilities of individual consumers. We argue that regulators should address epistemic fragmentation if they want to achieve more effective governance of online targeting.

Was there a specific motivation to write the article?
Regulation of online targeted advertising is being discussed in the UK and worldwide. Independently of how online harms are defined, monitoring will play an important role, because regulators need reliable means of finding out when and how advertisements break the rules. However, current strategies are inadequate to this task. On the one hand, educating individual consumers to make more informed online choices does not protect them from potential discrimination for example, and raises the bar for seeking redress. On the other hand, giving tech platforms the responsibility to vet and monitor ad campaigns risks ceding too much power to private actors that are not transparent or publicly accountable. This highlights the need to create a shared public space where online targeting can be subjected to democratically agreed rules.

How has the topic developed over 2021?
During the height of the pandemic and the US elections, there was a lot of focus on political messaging and the potential for online targeting to polarize users and to facilitate the spread of fake news. Epistemic fragmentation is a general concern for recommendation and personalization systems that filter the information available to individuals. In an epistemically fragmented network, individuals may be more vulnerable to exploitation, because it is harder to recognize when they are being harmed.

Has your own thinking about the topic evolved?
I now think epistemic fragmentation is a widespread problem in online ecosystems. It is also a source of injustice when it blocks avenues for diverse communities to share experiences and contribute to the regulation of AI systems that impact them. I am thinking about how we could make these systems more robust and aligned to public service values, for example in the case of news recommender systems.

Were you surprised or worried by any development in AI in 2021?
The announcement from Facebook about their plans to create the 'Metaverse' gave me pause: virtual reality in our daily lives may be a long way away, but a move towards it will make epistemic fragmentation even more acute in our everyday experience, which is something we need to think about. I was also worried by the Google AI Ethics team debacle, signalling how we can't leave it to tech giants to set the agenda around the ethical issues arising from AI.

What are your hopes or expectations for AI for 2022?
I expect more interest and work on trustworthy and truthful AI. I hope for better recognition of the political nature of AI at every level and more research on accountable AI.

James Zou
17 June 2021; Abid, A., Farooqi, M. & Zou, J. Large language models associate Muslims with violence. *Nat. Mach. Intell.* **3**, 461–463 (2021)

What was your Comment about?
We studied stereotypes embedded in large language models like GPT-3, and found that GPT-3 persistently associates Muslims with violence. We explored methods to reduce such harmful stereotypes in the language model, which is critical as such models become widely used.

Was there a specific motivation to write the article?
Large language models are some of the most exciting recent developments in AI. They can potentially transform many AI applications including chatbots, search engines and healthcare. At the same time, these models are extremely large (GPT-3 has over 170 billion parameters), they are trained on massive text corpora, and we don't have a great understanding of their behaviours. We believed that it was especially important to systematically audit language models to identify and mitigate potentially harmful stereotypes that it picked up from training data.

How has the topic developed over 2021?
In 2021, there is a growing focus and attention on responsible AI in the research, industry and public policy communities.

There is increasing recognition that powerful AI algorithms, especially language models, can contain problematic biases due to their training data. The discussion now is on how to rigorously evaluate these flexible models, which is much more challenging than evaluating simple classifiers, and how to mitigate their bias. Many leading organizations developing language models, like Hugging Face and OpenAI, now have dedicated teams doing important work on responsible AI.

Has your own thinking on the topic evolved?
I think we need to shift from one-off model evaluation to continuous monitoring of AI models after deployment. The increased flexibility of the models means that they can behave unexpectedly when the data stream changes in practice. Automated or human-in-the-loop frameworks that continuously test AI models over time will be very helpful in ensuring reliability.

Were you excited by any other development in AI in 2021?
I'm particularly excited by the emergence of data-centric AI in 2021. Data-centric AI focuses on developing scalable methods to help us curate, clean and mitigate bias in datasets used to develop and evaluate AI models. Having reliable data pipelines is critical for developing trustworthy AI, it has been relatively understudied in the AI community.

Has the COVID-19 pandemic affected your research?
The pandemic has made it easier to interact widely—I had fun days where I spoke 'in' the UK in the morning, in New York during lunch and then taught my class at Stanford in the afternoon. But it's also become more challenging to have deep and creative conversations, which benefit from more informal interactions.

What are your hopes or expectations for AI for 2022?
I hope and expect that there will be much more activity in data-centric AI in 2022. Especially as model-building becomes more automatic (for example, with AutoML), methods to systematically improve data pipelines and mitigate spurious correlations in the data will be essential for trustworthy AI. We recently organized a workshop to build a community around data-centric AI and released Data-centric AI Benchmark, which is a suite of hundreds of self-contained data puzzles. Each puzzle contains a dataset, a particular data-pipeline task (such as data cleaning or data

selection) and the ground-truth solution. We encourage researchers to compete and submit their methods to tackle these data tasks, so we can start to develop best practices for data-centric AI.

Carina Prunkl
17 February 2021; Prunkl, C. E. A., Ashurst, C., Anderljung, M., Webb, H., Leike, J. & Dafoe, A. Institutionalizing ethics in AI through broader impact requirements. *Nat. Mach. Intell.* **3**, 104–110 (2021)

What was your Perspective about?
Our article addressed the challenges of community governance in AI research. We investigated the pros and cons of introducing obligatory 'broader impact statements' for researchers submitting to machine learning conferences. More precisely, we compared such broader impact requirements with similar, already existing governance measures (such as institutional review boards and funding applications) and identified associated risks and challenges. Finally, we offered a list of tentative suggestions on how to maximize the probability of success of such impact requirements.

Was there a specific motivation to write the article?
Yes! NeurIPS 2020 had just announced that they were going to ask researchers to include a broader impact statement as part of their article submissions. It soon became clear that this was a controversial step and that there was a lot of demand for discussion—both on whether broader impact requirements were the right means to address ethical and social challenges from AI and on what a successful implementation of such requirements would look like. It is crucial that we understand previous lessons learnt in order to implement new governance measures as best possible.

How has the topic developed over 2021?
The community is certainly still in the process of figuring out what the right way forward is. NeurIPS 2021 replaced the broader impact requirement with a 'Paper Checklist' and guidance that is used as part of the review process. One section explicitly asks whether authors have considered any potential negative societal impacts in their submission and provides some guidance as to what responses to the question could look like. However, the organizers stress that answering 'no' to questions is not in itself a ground for rejection. In this sense, the requirement has become weaker than it was in 2020, which may seem at first sight like a step back. However, we need to better understand the challenges associated with

asking researchers to reflect on societal impacts, and will likely need many more iterations of impact requirements before we find one that is both effective and agreed upon by the community. This is why, in our Perspective, we emphasized the importance of establishing dedicated forums for deliberation on researcher norms. The community will need to come together and jointly decide what governance measures are appropriate to address challenges emerging from AI research.

Did you get any surprising feedback?
We were positively surprised by how much resonance we got from the AI research community and conference organizers. We were also approached by some members of organizing committees to chat about the article and its insights. Personally, I was surprised to find that many researchers seem to be cautiously in favour of having a broader impact statement. When we have presented the paper at workshops, most audience members so far have indicated that they approve, some are unsure and only very few are completely against it. Although there certainly exists a selection effect (the workshop themes are typically such that they attract a particular subset of researchers), this nevertheless indicates that, overall, researchers are keen to engage with the impacts of their work.

How has your own thinking evolved?
I feel even more strongly than before about having deliberation forums available to AI researchers. There should be much more dedicated space for reflection on past, current and future governance measures. Such forums can provide a more representative model for how opinions within the community are distributed (as opposed to social media). Ultimately, they can also give legitimacy to any future governance attempts by making the entire process both more transparent and more democratic.

Christopher Irrgang
17 August 2021; Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021)

What was your Perspective about?
We surveyed the recent rise of AI in Earth and climate sciences and contrasted the current limitations of data-driven AI with those of physics-based Earth system models (ESMs). Based on this assessment, we proposed a framework for evolving new AI and classical ESM approaches into

a combined research field—neural Earth system modelling—that aims towards building self-learning and self-validating model–network hybrids. We argued that in climate-change-related decision-making informed by AI research, explainability and interpretability of AI models are essential.

How has the topic developed over 2021?
Research this year shows that we are starting to understand how and when AI can help solve problems in Earth and climate sciences. AI is no longer just a novel method to try out because previous approaches have failed for certain tasks. There is now a clear vision how AI can fit into current state-of-the-art studies and how it can complement current process-based models, at least for several problems in Earth system science. At the same time, I think the discussion has highlighted the limitations of AI for climate prediction problems. Resolving these limitations and further evolving the capability of AI in climate sciences will be the next big leap.

Did you get any surprising feedback?
I was excited by the very positive feedback we received for the Perspective and was humbled by the many chances to meet people around the world (virtually, of course) who work on this topic. The discussions were extremely stimulating.

Were you excited by any development in AI in 2021?
For me, DeepMind's precipitation nowcasting system was an incredibly exciting development from this year.

Has the COVID-19 pandemic affected your research?
During the first months of the pandemic, many tasks, to-do lists and research plans fell overboard with the adaptation to a new and mostly virtual working environment. This time period allowed me to take a step back from the previously established routines and to think about long-term research ideas that go beyond immediate research plans. I consider this opportunity very precious.

What are your hopes or expectations for AI for 2022?
In terms of climate science, I hope that we will see more applications of AI with high impact on usability. We have seen various conceptual studies that show how AI can in principle support classical data analysis or modelling techniques. But I am looking forward to upcoming use cases with a true 'wow' effect that demonstrate how this promising technique can help us to better predict and cope with the changing climate.

I. Glenn Cohen
20 April 2021; Babic, B., Gerke, S., Evgeniou, T. & Cohen, I. G. Direct-to-consumer medical machine learning and artificial intelligence applications. *Nat. Mach. Intell.* **3**, 283–287 (2021)

What was your Perspective about?
Direct-to-consumer applications raise unique concerns, as consumer users can be risk averse about their health outcomes and limited in their statistical and medical literacy. To determine the benefits and costs of these applications to patients and the healthcare system, we need to consider such behavioural factors and how they interrelate to the specificity and sensitivity of the applications themselves.

Was there a specific motivation to write the article?
I had the opportunity to collaborate with an amazing set of co-authors on a series of related papers looking at different facets of problems of AI in medicine. The Perspective allowed us to build on an approach for regulators we call adopting the "system view" of AI and machine learning and to look at a particularly interesting corner of the market—direct-to-consumer applications. We have seen a proliferation of these in the market in the last several years, and we felt that the behavioural aspects of their use had not yet received a sufficiently detailed exploration.

Has the COVID-19 pandemic affected your research?
Yes, in the sense that many of the issues I have been interested in, regarding meeting healthcare needs outside the traditional healthcare setting of a hospital or physician office, have taken on an increased role during the COVID-19 pandemic, as many sites were shut for non-emergency care. As governments have relaxed regulatory restrictions on telemedicine during the pandemic and investors have put more money in the space, all the issues raised in the paper have become more salient. In particular, even if the risk borne by any given individual from failures of such applications may be low, the aggregate cost to public healthcare systems and private insurers can be quite large.

What are your hopes or expectations for AI for 2022?
In part as a result of changes in work patterns since the pandemic, an increase in telemedicine and investment in home health, we are likely to see increased development and, more importantly, integration of various types of patient health assessment

outside traditional healthcare settings into the diagnostic ecosystem. This will take a myriad of forms, including more emphasis on wearable sensors, at-home biospecimen collection and perhaps 'ambient intelligence' of some form, depending on the setting. Patient expectations about the quality of this information and what physicians will do with it, though, may not match its actual use or its optimal use in healthcare settings. I see 2022 as a year of further trying to bridge that gap.

Hao Su, Robin R. Murphy, Russell H. Taylor and Axel Krieger
18 March 2021; Su, H., Di Lallo, A., Murphy, R. R., Taylor, R. H., Garibaldi, B. T. & Krieger, A. Physical human–robot interaction for clinical care in infectious environments. *Nat. Mach. Intell.* **3**, 184–186 (2021)

What was your Comment about?
We identified three major areas where robots can improve patient care and safety for healthcare providers in the combat against infectious diseases: diagnostic procedures, interventional procedures and bedside care. To tackle clinical challenges in these areas, highly flexible and versatile medical robots are needed. Exploring research topics in physical human–robot interaction, including sensing, manipulation and autonomy, can enable such advances.

Was there a specific motivation to write the article?
Since the outbreak of COVID-19, a diverse range of robotic systems have been deployed in the field to manage public health and infectious diseases. One of us (R.R.M.) organized the Robotics for Infectious Diseases Consortium to help bring together researchers and document uses throughout the world. Our team includes frontline clinicians as well as roboticists who have developed both medical devices and robotic systems to mitigate the pandemic. Furthermore, two of us (R.H.T and R.R.M) were co-organizers of a National Academy of Engineering and Computing Consortium workshop on the role of robotics in infectious disease crises. We wanted to synthesize our experiences. We aimed to write a position paper summative of the field for roboticists and broader audiences, especially policy-makers, to understand the major technological barriers in robotics for clinical care during a pandemic.

Did you get any surprising or useful feedback?
Perhaps the most frequent and noteworthy feedback from the community was that our

paper helped to introduce the opportunities and challenges of robotics for clinical care to researchers who had not previously considered working in this area. It has been heartening to see how engaged the robotics community is, aiming to make real contributions to protect healthcare workers, handle the surge in patients and enable hospitals and medical care to keep functioning.

How has your own thinking on the topic evolved?
The biggest use of robots in clinical care was to protect healthcare workers by allowing them to work remotely and handle the surge in patients by offloading mundane tasks such as disinfecting, transporting bio-waste and delivering meals and medicine. We found that some of the most critically needed robots are for more capable infectious materials handling, lab automation and endotracheal intubation. Timely development and widespread effective deployment of such advanced tools requires that multiple issues are addressed. These include basic hardware and physical capabilities, autonomy and intelligent control systems, and human–machine communication, all of which pose research and implementation challenges. There are also various systems issues, among others low-cost manufacturing requirements, and the need for training and IT resources, that need to be tackled to integrate robots in existing work flows.

Were you surprised by any developments in AI or robotics in 2021?
The biggest surprise is how many of the robots being used already existed. Of the 338 documented cases of robots in use for the pandemic in 48 countries, 73% were commercially available. The remainder were robots that were modified either to fit a particular application or to fulfil a new need, like autonomous remote nose swabbing. The concern is that in the rush to meet emerging needs, innovative robotics technology or copycat robots may not be sufficiently reliable to be put into operation. As with developing new vaccines, robots need rapid and thorough testing.

What are your hopes or expectations for AI for 2022?
We hope that a coherent national or international strategy will be developed to increase preparedness to use robotic systems in future emergencies. This strategy should include a role for research to address knowledge and capability barriers, as well as the broader issues that affect adoption, such as reliability, human–robot interaction and

trust. We would like to see more incentives for accelerating the transition from research prototypes to replicated and deployed systems for emerging crisis applications. In particular, medical insurance and regulators should permit reimbursement of hospitals for using robots. We also hope that robotics researchers will have more opportunities to collaborate with clinicians to understand and prioritize the most critical research questions.

Mirko Kovač
10 November 2020; Miriyev, A. & Kovač, M. Skills for physical artificial intelligence. *Nat. Mach. Intell.* **2**, 658–660 (2020)

What was your Comment about?
Our article focusses on the concept of 'physical artificial intelligence', which we define as a synthesis method for the development of lifelike robots. The core idea is that by defining research questions at the interface of scientific disciplines and by co-evolving contributions at these interfaces, we can create robots that have unprecedented capabilities akin to those of natural organisms.

Was there a specific motivation to write the article?
The nascent field of soft robotics needs an integrated development methodology that combines contributions from material science, robot design, learning-based control and bio-hybrid actuation. An integrated framework that defines how to co-evolve such contributions that are not solely related to one discipline is missing, and this leaves a knowledge and skill gap in the field. The article attempts to offer a view on how the next generation of roboticists could be educated to enable them to advance the field and develop technological innovations in robotics and AI.

Did you get any surprising or useful feedback?
Yes, the article received a lot of attention, and I was contacted by researchers who were inspired by the integrated vision we presented. I felt very happy and encouraged that there is so much resonance in the community on the question of how we can better work together across disciplinary boundaries. It reinforced my view that institutional and community-level support structures are required to support researchers doing interdisciplinary work.

What are your hopes or expectations for AI for 2022?
I hope that work presented in interdisciplinary journals such as *Nature Machine*

Intelligence will inspire researchers from disciplines that are not traditionally related to robotics and AI development, in particular within the fields of material science, chemistry and synthetic biology. I am convinced that together we can make a step change in the field and create novel, lifelike and benevolent robots for the benefit of society.

Jathan Sadowksi
19 October 2020; Sadowski, J. & Andrejevic, M. More than a few bad apps. *Nat. Mach. Intell.* **2**, 655–657 (2020)

What was your Comment about?
We argue that approaches for tackling the ethical issues arising from applications of AI in society must move beyond a reactive approach. Instead, we must proactively confront the role of political structures and power relations in establishing which imperatives, whose interests and what goals influence the development of AI and machine learning systems in the first place.

Was there a specific motivation to write the article?
The motivation for this article was the notable rise in public discussions about the ethics of AI. Not just by academics, but also by companies and governments seeking to be the stewards of what 'ethics' means in these debates and applications. However, we saw very little discussion of the role of politics—of power dynamics, social structures, conflicting values, relations of authority. This article was meant to inject these critical concerns into the discussions about AI and society.

Did you get any surprising or useful feedback?
Although the article itself was well received by colleagues, the most surprising and useful feedback came from *Nature* itself. The opinion editor for the flagship *Nature* journal read this article and approached me about writing a commentary for a special issue of *Nature*. This gave me an opportunity to further develop my thinking and refine my argument for a much broader, more generalized audience. Which then received even more positive feedback and opened opportunities.

What are your hopes or expectations for AI for 2022?
My expectation is that AI will continue to become an increasingly widespread and consequential technology, with various applications integrated deeper into our everyday lives—not just through the use of

devices powered by AI, but also through public and private institutions using AI to make automated decisions that influence our lives in many different and important ways. My hope is that AI will become a subject of critical inquiry, that its applications won't be treated as magical or inevitable but as the result of human choices and contingency. Behind every application is a bunch of people designing and building the technology for specific means and ends. And in front of every application is a decision that needs to be made about how to use that technology—if at all.

Vidushi Marda

11 March 2021; Marda, V. & Narayan, S. On the importance of ethnographic methods in AI research. *Nat. Mach. Intell.* **3**, 187–189 (2021)

What was your Comment about?

In the article, which I co-wrote with Dr. Shivangi Narayan, we argue that to truly understand the societal impact of AI, we have to focus on qualitative methods such as ethnography which provide crucial insights into the actors and institutions that wield power through the use of these technologies.

Was there a specific motivation to write the article?

Technologists have traditionally prioritized quantitative methods that focus on algorithmic outputs, data on outcomes, and datasets. We wanted to demonstrate how and why it is important to prioritize qualitative methods like ethnography in addition to current modes of inquiry. We saw the value in doing so for a few reasons: first, access to quantitative data is not a luxury that researchers in the majority of the world have—standard operating procedures do not always exist, data on outcomes is not disclosed, models and

datasets are not shared even through right to information requests. What do researchers do in these cases? How can we move towards algorithmic accountability and demonstrate the societal impacts of these technologies? Second, reflecting on learnings from field work in New Delhi, we argue that quantitative methods can tell us what happened in the case of a particular algorithmic system, but qualitative methods reveal how and why some outcomes occur, and who makes crucial decisions.

Thirdly, there is a tendency to relegate work from non-Western contexts to the realm of 'case studies', but through our research, learnings and reflection, we wanted to demonstrate that what we learnt in New Delhi has important lessons for researchers across the world, in the Global North *and* Global South, to understand and build on in their own local contexts.

What are your hopes or expectations for AI for 2022?

I have two hopes. First, I hope that researchers, practitioners, developers and policy-makers working on machine learning recognize that the most difficult questions we grapple with, from labour rights to climate change, and from the future of work to algorithmic oppression, can be answered not only by computer scientists, lawyers and policy-makers but also by those in adjacent fields of expertise and, most importantly, through the lived experiences of people. We must defer to and learn from other disciplines in order to better understand how technologies cause real people harm, beyond the realm of AI ethics. Second, I hope that expertise, experiences, knowledge and narratives from the majority of the world are seriously considered and that researchers engage with them. We have to challenge dominant narratives that have hitherto emanated from a few jurisdictions,

and continue highlighting questions of power and accountability in our work on AI—we cannot address bias in machine learning systems without addressing bias in the narrative surrounding these technologies as well. ❐

Cameron Buckner[1], Risto Miikkulainen[2,3], StephanieForrest[4,5],SilviaMilano[6],JamesZou[7], Carina Prunk[8], Christopher Irrgang[9], I. Glenn Cohen[10], Hao Su[11], Robin R. Murphy[12], Russell H. Taylor[13], Axel Krieger[14], Mirko Kovač[15,16], Jathan Sadowski[17] and Vidushi Marda[18]

[1]Department of Philosophy, The University of Houston, Houston, TX, USA. [2]The University of Texas at Austin, Austin, TX, USA. [3]Cognizant Technology Solutions, Austin, TX, USA. [4]Arizona State University, Tempe, AZ, USA. [5]The Santa Fe Institute, Tempe, AZ, USA. [6]Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter, UK. [7]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [8]Institute for Ethics in AI, University of Oxford, Oxford, UK. [9]Helmholtz Centre Potsdam, German Research Centre for Geosciences GFZ, Potsdam, Germany. [10]Harvard Law School, Cambridge, MA, USA. [11]Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA. [12]Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. [13]Department of Computer Science, The Johns Hopkins University, Baltimore, MD, USA. [14]Department of Mechanical Engineering, The John Hopkins University, Baltimore, MD, USA. [15]Materials and Technology Center of Robotics, Empa — Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland. [16]Aerial Robotics Lab, Department of Aeronautics, Imperial College London, London, UK. [17]Emerging Technologies Research Lab, Faculty of Information Technology, Monash University, Caulfield, Victoria, Australia. [18]Article19, Bangalore, India.