AI reflections in 2019

There is no shortage of opinions on the impact of artificial intelligence and deep learning. We invited authors of Comment and Perspective articles that we published in roughly the first half of 2019 to look back at the year and give their thoughts on how the issue they wrote about developed.

Alexander S. Rich

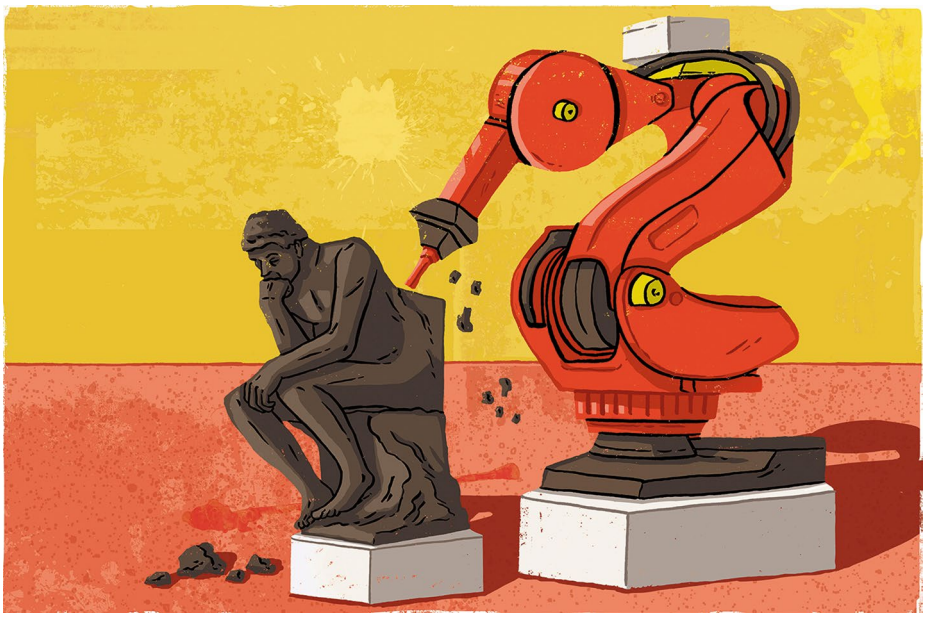
9 April; Rich, A. S. & Gureckis, T. M. Lessons for artificial intelligence from the study of natural stupidity. *Nat. Mach. Intell.* **1**, 174–180 (2019)

What was your Perspective about?
Machine learning algorithms can behave in harmful and biased ways when applied in high-stakes arenas such as criminal justice. Often, these algorithms are making decisions that were once left to another set of intelligent but biased agents — humans. Our Perspective lays out the literature on human learning and decision-making biases, and argues that understanding why these biases develop in humans can help us prevent them from emerging in machines.

Was there a specific reason or motivation to write the article?
The Perspective came about through a convergence of factors. At the macro level, the alarm bells have been ringing for the past few years about the biases and negative impacts of machine learning systems. On a personal level, I was shifting from being an academic psychologist to an industry machine learning practitioner. This gave me a unique vantage point to look back on how five decades of research into human biases can add to this conversation.

Has your own thinking on the topic evolved?
Spending time working on machine learning use cases in industry has made clear to me the pressing need for practical tools to identify and prevent algorithmic bias. This is particularly true for issues that occur due to choice-contingent feedback, which we discuss in the second section of the Perspective. There are many potential methods to address choice-contingent feedback from the reinforcement learning and causal inference literature, but there's a lack of accessible software or writing to guide use cases in domains such as healthcare where classic solutions may be impossible.

Do you have any specific hopes for artificial intelligence (AI) for 2020?
One trend I'm excited by is the growing interest in causal inference and causality within the machine learning community (see, for example, Judea Pearl's *The Book of Why*). Not only might causal reasoning lead to more flexible and human-like behaviour in AI but also it could be a key to preventing some of the biases discussed in our paper by letting algorithms account for the real-world data-generating processes behind the data.



Credit: Ikon Images/Eva Bee/Alamy Stock Photo

Cynthia Rudin

13 May; Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019)

What was your Perspective about?
The goal of the Perspective was to help people realize that there is a big difference, perhaps even a chasm, between inherently interpretable machine learning models and explaining black box models. Black box models (with or without explanations) are problematic for the reasons that I laid out in the Perspective. Many people have already suffered from decisions affecting their lives that were based on black box models, for example, as they were given extra prison time, were denied parole or loans.

Was there a specific reason or motivation to write the article?
Yes! Policymakers are struggling with how to regulate the use of machine learning models in practice, and the issue discussed in my Perspective is at the heart of such policy questions. Since 2016, there has been a huge effort towards explaining black box models, but not nearly as much effort in building interpretable models. Part of the problem is that many people (many smart people!) don't actually understand that an interpretable model and an 'explained' black box model are different and not equally valuable. A black box still requires you to trust the dataset that it was constructed from. Also, an explanation cannot fully explain the black box — otherwise no black box would be needed, only the explanation.

How has the topic developed over 2019?
The issue has not resolved. Policymakers need to be aware of these issues, as do academics who can inform the policymakers. There is still a gap where many academics believe that they need to sacrifice accuracy to gain interpretability

of machine learning models, despite the evidence that this is not necessarily true.

Did you receive any surprising responses?
The lack of pushback I received surprised me and I fear that many people may have misinterpreted the paper, or not actually read its content properly. Some have cited the work as a reason for not trusting black boxes, but then continued to discuss their work on explaining black boxes. Others have written that I said we should avoid neural networks to avoid black box models, but this is not what I stated; I even gave an example of an interpretable neural network. It is unfortunate that some people have not bothered to read the Perspective properly even when going to the trouble of mentioning it in their own writing. The misunderstandings and confusion about terminology (interpretable versus explainable) are precisely the reasons why I wrote this paper.

Do you have any specific hopes for AI for 2020?
I have continued hope that policymakers will recognize the danger that society faces if it permits black boxes to make decisions that deeply affect human lives. If we explain black box models instead of replacing them with interpretable models, we are just giving more authority to those who want to use black boxes, despite the inherent risks. I simply hope it stops. And I hope it stops before something bad happens on a very wide scale.

David M. P. Jacoby, Robin Freeman and Oliver R. Wearn

11 February; Wearn, O. R., Freeman, R. & Jacoby, D. M. P. Responsible AI for conservation. *Nat. Mach. Intell.* **1**, 72–73 (2019)

What was your Comment about?
Our Comment highlights that the AI community and conservation scientists need to promote the responsible and ethical use of AI in conservation, at a time when global biodiversity is in substantial decline. We argue for better, more diverse metrics of algorithm success and greater transparency of training data, in addition to ethics statements in research articles detailing both the generalities and limitations of use.

Was there a specific reason or motivation to write the article?
The use of machine learning in the automated processing of biological data has exploded in the past few years. This is particularly true in areas such as the processing of data gathered from remote sensors such as camera traps that can generate thousands of images at a single study site. With more and more research articles vying to gain the highest measures of predictive accuracy from image data, we felt it was a good time to reflect on where the field might go. We wanted to highlight both the exciting opportunities on offer and the potentially negative consequence of automation, particularly given the current fragility of even the most sophisticated deep neural network pattern-recognition tools[1], in an attempt to outline where we would ideally like the field to go.

How has the topic developed over 2019?
Anecdotally, we have seen an evolution in the discussions surrounding AI in conservation, with serious recognition that an ethical question mark hangs over the technology. We predict that this trend will continue into 2020, with much more restrained and nuanced reporting of algorithm performance, as well as greater discussion of mechanisms for ethical oversight of algorithms.

Did you get any surprising or useful feedback?
A number of AI ethics researchers reached out to us following publication of the Comment, which better connected us with parallel discussions on AI ethics outside of conservation. However, we think that many conservationists and ecologists remain unaware of the ethical dilemmas of this new technology, and greater discussion of these issues needs to be fostered in conservation-specific journals, not just in the machine intelligence literature.

Has your own thinking on the topic evolved?
We realized that the generalizability of methods across habitats, species and scenarios is something that hasn't been well addressed in our field yet. It's increasingly clear that the creation of methods to identify or predict are only a small part of the application of these methods in day-to-day practice in conservation.

Do you have any specific hopes for AI for 2020?
From our perspective, we hope that 2020 brings a greater synergy between the biological and conservation realms and the wider AI community. To avoid post hoc ethical considerations in response to unintentional use or misuse, factoring in ethical considerations or even just thinking about the real-world consequences during AI development is key.

Henry Shevlin

8 April; Shevlin, H. & Halina, M. Apply rich psychological terms in AI with care. *Nat. Mach. Intell.* **1**, 165–167 (2019)

What was your Comment about?
Our Comment examines the tendency of many researchers in machine learning and AI to describe their systems using the vocabulary of psychology, or what we call rich psychological terms — concepts such as agency, creativity and understanding. We caution against employing these terms too liberally, on the grounds that it makes communication harder among different branches of cognitive science, risks kneejerk reactions from policymakers and stakeholders, and potentially leads us away from finding more novel and informative high-level descriptions of the capacities of artificial systems.

Was there a specific reason or motivation to write the article?
As cognitive scientists working at the intersection of AI and animal cognition, we observed a dramatic difference in the methods and standards of evidence employed in the usage of psychological vocabulary between these two fields, and noticed that liberal use of rich psychological terms was commonplace in AI research.

How has the topic developed over 2019?
There have been a lot of impressive developments in artificial language and reasoning tests over the past year, such as the GPT-2 language model from OpenAI and the striking results obtained by the multi-task deep neural network (MT-DNN), a language model from Microsoft, on the General Language Understanding Evaluation (GLUE) benchmark. As artificial systems come closer to aping human performance on tasks like these, the question of whether the underlying mechanisms are appropriately described in the same terms as those in humans seems of growing importance.

Has your own thinking on the topic evolved?
One point that only came into clarity for me after publication of the Comment concerns a 'local research minima' problem. In short, there is a danger that companies and researchers keen to emulate human-level performance may over-invest in models that come close to replicating human abilities yet which differ in their fundamental architecture, meaning that moving from near-human-level to full-human-level performance may be impossible without fundamental changes being implemented in the system. By adopting stricter standards

for our application of psychological terms to such systems, we might make such misallocation of resources less likely.

Do you have any specific hopes for AI for 2020?
A key aspect of human language that has long fascinated me is conversational pragmatics — mundane utterances such as 'I'm tired' are important for human communication, but pose a major challenge for achieving even near-human-level performance in natural language understanding. Speaking to AI researchers, I have the impression that many teams regard this as a key area for future work, and I am hopeful that 2020 may bring some advances in this area, with potentially large ramifications for, for example, the performance of chatbots.

Kanta Dihal

11 February; Cave, S. & Dihal, K. Hopes and fears for intelligent machines in fiction and reality. *Nat. Mach. Intell.* **1**, 74–78 (2019)

What was your Perspective about?
Stephen Cave and I analysed around 300 narratives about artificial intelligence (fiction and non-fiction) and categorized the hopes and fears most commonly expressed in them. We show how these four hopes and four fears are connected, and how losing control means a hope turns into a fear.

Did you get any surprising or useful feedback?
Among other responses, the categorization presented in our paper informed a survey conducted by the BBC about perceptions of AI among the British public. Teaming up with Kate Coughlan from the BBC, we presented the results of this survey at the AAAI/ACM conference on Artificial Intelligence, Ethics and Society in 2019, in a paper titled 'Scary robots'. This title is the reply one participant gave us when asked in the survey, "How would you describe AI to a friend?"

Has your own thinking on the topic evolved?
We are now expanding our work on AI narratives in a global context, deploying it around the world in translation, as part of our Global AI Narratives research project. We are bringing together international experts on such narratives to enable comparative work and foster intercultural understanding.

Do you have any specific hopes for AI for 2020?
I hope that the global AI debate will continue to become more inclusive. The dominant voices in this space have much to learn from regions that are currently hindered in contributing as widely, due to linguistic, financial, political or cultural barriers. I hope that in 2020 the network we are developing through our Global AI Narratives work will be even stronger and more visible.

Seán S. ÓhÉigeartaigh

7 January; Cave, S. & ÓhÉigeartaigh, S. S. Bridging near- and long-term concerns about AI. *Nat. Mach. Intell.* **1**, 5–6 (2019)

What was your Comment about?
A divide has emerged between communities of researchers working on present-day challenges related to AI (such as algorithmic bias and interpretability) and issues that may emerge further in the future (such as large-scale impacts on labour market and artificial general intelligence). We argue that these topics have more links, and benefit more from collaboration between research communities, than is often recognized. The decisions that we make now may have long-term consequences for how AI develops.

Was there a specific reason or motivation to write the article?
We were concerned to see that separate communities and forums developed around near- and long-term concerns, topics we think are intrinsically linked. We also noticed the sometimes dismissive attitudes from scholars on both sides. However, combining insights from work addressing today's problems with those from foresight-oriented approaches seems crucial.

How has the topic developed over 2019?
Forums such as the Partnership on AI have provided valuable opportunities for researchers working on different timescales, and on a broad range of topics, to share insights and methodologies. For example, discussions have started around publication norms — should a new AI system with dual-use potential be openly released? Good arguments were made for and against OpenAI's decision to delay release of their impressive GPT-2 language model. At a time when advances are being made in many areas with the potential to enable misinformation and manipulation — from deepfakes to political targeting — it was good to see a sophisticated debate of these issues.

Did you get any surprising or useful feedback?
One interesting piece of feedback was an observation that many present-day or near-term issues are by their nature deeply politicized, as they relate to specific actors, power structures and existing inequalities. In exploring the links to longer-term issues, it will be important where possible to avoid this politicization carrying across.

Were you excited by any development in AI in 2019?
I've been particularly excited to see progress in applying AI to global scientific challenges such as renewable energy and biomedical sciences. Some standout examples include DeepMind improving the predictability of Google's wind energy production and a number of groups applying AI to protein folding.

Do you have any specific hopes for AI for 2020?
I hope to see greater global engagement and cooperation within the research community. It will be the first year in which one of the top-tier machine learning conferences takes place in Africa, with ICLR (International Conference on Learning Representations) being hosted in Addis Ababa. I also hope to see continued growth in collaboration between research communities in China, the United States and Europe. With global politics becoming increasingly fractious, it is more important than ever that researchers work together across borders and cultures to develop beneficial AI.

James Butcher

29 July; Beridze, I. & Butcher, J. When seeing is no longer believing. *Nat. Mach. Intell.* **1**, 332–334 (2019)

What was your Comment about?
AI is scaling the ability to produce synthetic media (text and audio-visual) and it is making it easier for individuals to create doctored content. This has a number of concerning implications and poses an increasing security threat. The solution requires technical and governance approaches.

Was there a specific reason or motivation to write the article?
We wanted to promote awareness of the issues, synthesising the developments and challenges in a succinct and detailed manner. We also sought to contribute to the conversation by highlighting potential solutions.

How has the topic developed over 2019?
The topic exploded in 2019 as the public is becoming increasingly aware of the capabilities and potential threats of AI. Since the Comment was published, there has been much more coverage of this topic, a growth in deepfake technology being deployed and advances made for potential

solutions. There has even been an example of fraudsters using AI to impersonate a CEO's voice and demand a fraudulent transaction of €220,000.

Do you have any specific hopes for AI for 2020?
Since publication, some major technology companies have started dedicating resources explicitly for tackling the problem of deepfakes. We hope to see greater cooperation between stakeholders to design appropriate solutions to address the challenges that AI-enabled synthetic media poses. The United Nations is also working on addressing the challenges. UNICRI (United Nations Interregional Crime and Justice Research Institute), through its Centre for AI and Robotics, and the Data Science Initiative of the City of The Hague, hosted a hackathon and workshop on deepfakes and manipulated videos in 2019, and we expect such activities to increase.

Marco Lippi, Przemyslaw Palka and Paolo Torroni

25 March; Lippi, M. et al. Consumer protection requires artificial intelligence. *Nat. Mach. Intell.* **1**, 168–169 (2019)

What was your Comment about?
In this Comment, we explore the ways in which AI and data analytics can be used to empower consumers in the digital marketplace. We look at how AI-powered tools for the analysis of contracts, ads and algorithms can help the civil society exercise their rights better and conduct oversight of business practices.

Was there a specific reason or motivation to write the article?
AI is currently being presented as a source of threats to consumers, whose data is being constantly collected, analysed and used by companies to increase their sales and influence consumer behaviour. These threats are real, but do not account for the whole picture. We believe researchers should work together to use AI to empower consumers. We started a fruitful collaboration two years ago between a team of computer scientists and a team of lawyers (experts in consumer law) for the automatic detection of potentially unlawful or non-compliant clauses in terms of service and privacy policies. Many interesting discussions within this interdisciplinary research group led to this idea of a counter-power for consumers, through the use of AI.

How has the topic developed over 2019?
The topic is continuously evolving. There is a growing interest in the analysis of ethical problems in AI, and consumer protection from unfair uses of AI is becoming a major societal challenge. We believe that more attention should be paid to the ways in which AI can in turn be used by consumers to tackle the ethical and regulatory challenges in the digital marketplace.

Has your own thinking on the topic evolved?
We are convinced that research in the field of AI and law should attempt to integrate neural and symbolic approaches in AI: this is a crucial step to combine data-driven tasks, such as detection or categorization, with high-level reasoning tasks, which require formalization and an exploitation of some background knowledge.

Do you have any specific hopes for AI for 2020?
I hope that the AI community will continue its efforts to develop the ethics of the discipline, so that research will focus on relevant societal challenges and on the improvement of the quality of life of citizens.

Shannon Wongvibulsin

28 January; Wongvibulsin, S. Educational strategies to foster diversity and inclusion in machine intelligence. *Nat. Mach. Intell.* **1**, 70–71 (2019)

What was your Comment about?
The Comment is about the importance of fostering diversity and inclusion in machine intelligence. I describe strategies to build an accessible educational and mentorship structure to promote a sustainable infrastructure for active participation and long-term success in the machine intelligence community for individuals of all backgrounds.

Was there a specific reason or motivation to write the article?
During the fall of 2018, I had the privilege of designing and teaching a course to introduce undergraduates to cutting-edge engineering research and its societal impact through the Hopkins Engineering Applications and Research Tutorials (HEART) programme. My experiences as well as the feedback I received from the students and faculty motivated me to share my ideas for attracting individuals from a broader range of backgrounds to join the machine intelligence community through educational strategies that foster diversity and inclusion.

Has your own thinking on the topic evolved?
Beyond the strategies discussed in the Comment (that is, building a welcoming culture, student as teacher educational models, flipped classrooms and longitudinal outreach programmes), I believe more efforts will be essential in inspiring the next generation of individuals to join the machine intelligence community. In particular, there is enormous potential to capture the attention of young people through advances in machine intelligence that enable the integration of personalized education into daily life activities.

Do you have any specific hopes for AI for 2020?
Although much exciting progress has been made in AI, concerns still remain about its usability, transparency and safety, especially in medicine. I hope that in 2020, progress will be made towards addressing barriers to the clinical translation of AI developments. Through increasingly multidisciplinary teams, I am hopeful for advances towards not only the development of increasingly powerful AI algorithms but also their potential to be integrated into the healthcare system to augment clinical practice and patient care.

Edmon Begoli

7 January; Begoli, E., Bhattacharya, T. & Kusnezov, D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat. Mach. Intell.* **1**, 20–23 (2019)

What was your Perspective about?
In our Perspective, we advocate for the need of uncertainty quantification research for AI in medical decision-making.

Was there a specific reason or motivation to write the article?
It is based on our own work and experience in developing systems, capabilities and methods in both uncertainty quantification and in medical AI and recognizing that these two disciplines need to converge.

How has the topic developed over 2019?
We have seen an increasing focus and interest in 'opening the black box of AI', and in a more formal understanding of how AI works and under what conditions it does not.

Has your own thinking on the topic evolved?
We are seeing a very close, inverse connection between uncertainty quantification and adversarial AI. We are now working on the set of principles that take uncertainty quantification for AI a bit further in terms of practice, and also in connecting it with the ways to understand exploitability of AI. We are worried about the abuse of AI, and about the exploitation of AI-based decision-making, because we do not fully

understand how to quantify the uncertainty and the limitations of the AI-based models that support that decision-making.

Do you have any specific hopes for AI for 2020?
For 2020, we hope to see development of the safety culture in AI research, where equal attention will be paid to uncertainty quantification, testing, validation, verification and fairness of the AI models, as to their development and demonstration of the capabilities (under ideal conditions). Also, we hope to see less hype in the media, and a bit more maturation and critical views about the limitations and the capabilities of the AI. Having a sober, realistic and objective view about the true state of the AI capabilities and its limitations might prevent the next 'AI winter', which would be accompanied by a drop in public attention, funding and investments into AI research. We believe that this time around, AI has its best chance ever to avoid such a 'change of seasons'.

Gisbert Schneider

27 February; Schneider, G. Mind and machine in drug design. *Nat. Mach. Intell.* **1**, 128–130 (2019)

What was your Comment about?
Machine intelligence offers fresh opportunities for pharmaceutical research. This Comment highlights one of the most relevant questions in drug discovery, namely, which molecule to make next, and to what degree contemporary AI can assist the medicinal chemist in this regard.

Was there a specific reason or motivation to write the article?
Machine intelligence saw its first heyday in pharmaceutical discovery in the 1990s. At the time, the expectations were unduly high but rather abruptly curbed. Now, AI seems to be back for good. I wrote the Comment to revisit some of the big challenges that modern AI must address to be successful. As a scientific community, we should remember the mistakes from the past and try not to make them again.

Do you feel the topic has developed over 2019?
Judging from the rapidly increasing number of critical articles on the topic, and the lively discussion, it is my impression that the field is maturing and the dust is slowly settling. Industry has cautiously begun to integrate the available tools into their discovery pipelines. There still are voices claiming the 'magic of AI'. However, only the prospective experimental application of AI-supported drug design will help us identify those algorithms and their applicability domains that bear the greatest potential.

Has your own thinking on the topic evolved?
Speaking to many researchers with and without a background in AI has supported my initial assumption that the appropriate mindset seems to determine success or failure of applied AI in drug discovery. Methods evolve over time, but they should be embedded within a collaborative discovery framework.

Were you worried or excited by any development in AI in 2019?
I've seen a few recent papers ballyhooing AI as 'the solution' to the challenges we are facing in drug discovery and personalized healthcare. We should be more cautious, avoid over-selling of modest achievements and separate commercial interest from fact-based scientific reporting. There is virtue in humility, and we should let the results of a scientific experiment speak for themselves.

Do you have any specific hopes for AI for 2020?
I certainly hope to see many prospective applications of AI in pharmaceutical research. It is still a long way before we will know the potential and limitations of many of these methods for drug discovery and development. Fostering collaborative interdisciplinary thinking and designing good experiments will be indispensable.

Stephen Cave

7 January; Cave, S. & ÓhÉigeartaigh, S. S. Bridging near- and long-term concerns about AI. *Nat. Mach. Intell.* **1**, 5–6 (2019)
11 February; Cave, S. & Dihal, K. Hopes and fears for intelligent machines in fiction and reality. *Nat. Mach. Intell.* **1**, 74–78 (2019)

What were your articles about?
I had two articles in *Nature Machine Intelligence*. The Comment 'Bridging near- and long-term concerns about AI' (with Seán ÓhÉigeartaigh) argues for mending the divide between communities concerned with the near-term risks of AI and those concerned with the longer-term risks. The Perspective 'Hopes and fears for intelligent machines in fiction and reality' (with Kanta Dihal) categorizes the four main categories of hopes and fears people have for AI, based on an analysis of a large corpus of fiction and non-fiction works.

Did you get any surprising or useful feedback?
I have been very pleased at how both articles have provoked debate and further work. I was particularly delighted at the publication of a recent paper by Rachel Adams that gives a gender theory-based reading of Kanta's and my schema of hopes and fears[2].

Has your own thinking on the topic evolved?
My own thinking on the frenzied discourse around AI has increasingly been informed by critical perspectives like Rachel Adam's, including not only gender theory but also postcolonial and critical race theory. I think many of us still take key concepts around AI and its impacts too much at face value. An example is the concept of intelligence itself, which is highly value-laden and has a dark history entwined with eugenics and colonialism.

Do you have any specific hopes for AI for 2020?
I hope that 2020 will see an increasing range of scholars and communities engage with debates about our future with AI. Although machine intelligence poses new challenges, it also exacerbates a range of existing ones, such as the oppression of certain communities, on which there are already substantial bodies of literature. We need more works like Ruha Benjamin's book *Race After Technology* that forge links between these fields.

Mona Sloane and Emanuel Moss

9 August; Sloane, M. & Moss, E. AI's social sciences deficit. *Nat. Mach. Intell.* **1**, 330–331 (2019)

What was your Comment about?
Our Comment argues that AI designers should enlist ideas and expertise from a broad range of social science disciplines, including those embracing qualitative methods, to reduce the potential harm of their creations and to better serve society as a whole.

Was there a specific reason or motivation to write the article?
As social science researchers studying approaches to machine learning and AI, we were regularly attending academic computing conferences where we would hear aspects of society described through very quantitative frames. This might manifest as a numerical index being used to represent a complex social phenomenon such as the 'sentiment' or 'toxicity' of online speech acts, or as a ranking system for 'successful' conversations in a chatbot application, or any number of risk scores used to predict recidivism, propensity towards fraud and so on. In thinking about how much reduction

of social complexity is needed to compress all the complexity of social life into these quantitative measures, we realized that there is a massive potential for cross-disciplinary engagement with practitioners of qualitative methods, who could properly contextualize the use of quantitative measures and help draw reasonable bounds around the use of quantitative measures, so that they aren't used beyond the scope in which they are appropriate and are well calibrated.

How has the topic developed over 2019?
The reliance on quantitative measures for representing complex social phenomena has, if anything, increased apace in machine learning over the past year, and so we see our call for more qualitative social science in machine learning and AI as being as relevant as before. This is particularly true in conversations around fairness and bias in machine learning, where the focus has been on developing more quantitative fixes to the problem of dataset imbalances and constrained optimization problems, rather than the pursuit of just and equitable applications of machine learning technologies.

Were you worried or excited by any development in AI in 2019?
We were happy to see critical discussions being put on the table and being discussed not only by researchers but also across industries, policymakers and communities. Importantly, key works in the area of critical AI studies are led by those who tend to be marginalized in the tech communities. What is worrying is that a lot of money is put into technological research on AI; we need to see more investment into understanding the social, economic and ecological implications of AI innovation.

Do you have any specific hopes for AI for 2020?
We hope that we will see more nuanced 'impact assessments' and that we will be able to include the voices of those most affected by AI technologies, rather than technologists themselves. We also hope that we will be able to leave behind the narrative of the 'global AI race' and shift our focus onto more nuanced discussions and actions about AI technology, the climate crisis, colonialism, global inequality and so on.

Iyad Rahwan

11 February; Frank, M. R. et al. The evolution of citations graphs in artificial intelligence research. *Nat. Mach. Intell.* **1**, 79–85 (2019)

What was your Perspective about?
Our Perspective has two main messages. First, AI research seems to be getting more insular over time, being less connected to research in the social sciences. Second, it seems that private companies (Internet giants) are becoming a dominant player in AI, raising questions about the extent to which academic institutions can keep up in the future.

Was there a specific reason or motivation to write the article?
We believe that understanding the evolution of AI research is important to quantify and predict its impact on society. It is also important to understand to what extent other fields of research, especially the social and behavioural sciences, are connected to recent AI developments. In parallel with writing this Perspective, I was also working on a Review Article in *Nature* titled 'Machine behaviour'[3], which was an invitation to scientists from all disciplines to help us understand the behaviour of intelligent machines and the collective behaviour of human–machine systems. So I wanted to understand to what extent these fields were connected in the past, and how this trend was evolving.

How has the topic developed over 2019?
There is a growing recognition that AI scientists need to learn more from other fields. In fact, response to the 'Machine behaviour' Review Article, which resonated strongly with quantitative social and behavioural scientists in particular, was mostly positive. I am excited that more scientists from outside computer science are now taking AI seriously as an object of study in their own fields.

Has your own thinking on the topic evolved?
I think there is an important role for both quantitative and qualitative social science. Quantitative social scientists can help computer scientists measure and model the behaviour of human–machine systems with precision. But it is very difficult for them to build comprehensive models of complex phenomena, for example, of how algorithms might amplify biases that are more systemic. So a collaboration between quantitative 'machine behaviourists' and qualitative scholars from the field of science, technology and society would be very fruitful, and would help us cover blind spots, while also putting qualitative claims to the test whenever possible.

Do you have any specific hopes for AI for 2020?
The two fields that have traditionally studied human–machine systems are human–computer interaction (within computer science) and science technology and society, which has its roots mostly in the fields of policy, history and philosophy of science and technology. Now, other fields, such as economics, political science, biology and psychology, are beginning to enrich our understanding of human–machine systems, and I hope this trend will accelerate.

Ken Goldberg

7 January; Goldberg, K. Robots and the return to collaborative intelligence. *Nat. Mach. Intell.* **1**, 2–4 (2019)

What was your Comment about?
Robots are increasingly collaborative with humans and with each other. The Comment reviews how four growing and increasingly overlapping subfields of robotics research are influencing this trend: co-robotics, human–robot interaction, deep learning and cloud robotics.

How has the topic developed over 2019?
This trend has grown as companies realize how AI and robot systems can benefit from having humans in the loop. An example is the Massachusetts Institute of Technology's HERMES (Highly Efficient Robotic Mechanisms and Electromechanical System) project, which uses human instinctive balancing reactions to remotely control a humanoid robot[4].

Has your own thinking on the topic evolved?
I started using the term 'complementarity' to describe systems where AI and robots complement human skills, allowing humans to focus on what we do best: dexterity, creativity, intuition, empathy and communication.

Do you have any specific hopes for AI for 2020?
I am an optimist and believe that advances in AI can inspire what I call 'wide learning' for humans (in contrast to 'deep learning' for machines). Wide learning has the potential to expand human learning opportunities along three dimensions: people skills, cognitive diversity and lifelong learning.

David Howard

7 January; Howard, D. et al. Evolving embodied intelligence from materials to machines. *Nat. Mach. Intell.* **1**, 12–19 (2019)

What was your Perspective about?
The Perspective is about multi-level evolution — a nature-inspired approach to designing robots that combines materials discovery, evolutionary robotics and diversity-based machine learning. It will open up opportunities to create bespoke, highly performant robots that specialize to their tasks all the way from materials to machines.

Was there a specific reason or motivation to write the article?
The question of how to incorporate materials discovery and selection into robot design kept cropping up, and we saw an opportunity to develop our thoughts on how to bridge the gap between machine learning-based materials discovery and machine learning-based robotic design. We wanted to provoke discussion by proposing a simple, viable architectural framework. In particular, we saw high-throughput materials science, materials modelling, and advanced additive and subtractive manufacture as enabling technologies for robotic manufacture. This opened up a world of possibilities for designing robots in a large range of morphological configurations, and based on a plethora of candidate materials. We also saw what 'multi-level evolutionary' architectures could do for us in terms of getting robots to act naturally and robustly in challenging natural environments, which is still a challenging unanswered research question.

How has the topic developed over 2019?
We see that the field of 'material robotics' is gaining a lot of momentum in the research community, and a way of autonomously designing robots using a wide range of materials is highly sought after. The underpinning research areas continue to mature, and a lot of the conversations we have with collaborators are around how we can work together to realize these architectures.

Do you have any specific hopes for AI for 2020?
I would like to see more cross-field collaboration, and more standardization. Both are critical to the ability to deploy AI systems (like ours!) at scale.

Luciano Floridi

7 May; Floridi, L. Establishing the rules for building trustworthy AI. *Nat. Mach. Intell.* **1**, 261–262 (2019)

What was your Comment about?
I argue that the report *Ethics Guidelines for Trustworthy AI* — published by the European Commission's High-Level Expert Group (HLEG) on 8 April 2019 — is a good step in the right direction to develop and evaluate the responsible development of AI system. As a member of the HLEG, I felt it was crucial to discuss its contents and value in a timely and publicly very visible manner.

How has the topic developed over 2019?
The topic of ethical frameworks for AI has moved on rather quickly, as now there are Organisation for Economic Co-operation and Development Principles and Beijing Principles, just to mention some other important initiatives. The good news is that available frameworks cohere with each other quite substantially, thus offering the opportunity to align any organization's values with international expectations, instead of having to elaborate its own.

Has your own thinking on the topic evolved?
I am now looking into the impact of external auditing on companies developing or using AI as part of their core business. In particular, I am a member of EY's AI advisory board to address the ethical challenges posed by AI. And EY serves as auditor to some of the most important tech firms, so its position about the ethics of AI may be influential. I think this is one of the next, main challenges.

Were you worried or excited by any development in AI in 2019?
I was excited by the increasing realism with which autonomous vehicles are being discussed, with fashionable and distracting topics such as the 'trolley problem' fading away and real issues attracting much more attention: the discussion is now focused on where and how transports may be subject to different level of automation, and with what impact, socially and environmentally.

Do you have any specific hopes for AI for 2020?
I hope that increasing clarity about regulations and ethical frameworks will enable private and public actors to design, develop and deploy AI in contexts such as healthcare or climate change, where the opportunity costs of doing nothing, due to lack of certainties, are mounting. I expect AI to become increasingly explainable, more reliant on synthetic data whenever possible and more 'unexcitingly' widespread as a normal technology, with useful solutions further crowding out sci-fi speculations.

Jack Stilgoe

2 April; Stilgoe, J. Self-driving cars will take a while to get right. *Nat. Mach. Intell.* **1**, 202–203 (2019)

What was Comment about?
For self-driving cars to work, their developers need to do more than just make improvements to machine learning. Others need to be involved too, which will take time. The 'race' to develop self-driving cars could lead to bad decision-making.

Was there a specific reason or motivation to write the article?
Hype about self-driving cars was building. The understandable enthusiasm for the technology, and for its status as a real-world application of AI, was leading to some important issues being overlooked. I felt that the responsible development of AI needed to include some broader considerations.

How has the topic has developed over 2019?
There have been various self-driving car developers admitting that developing the tech has been harder than they anticipated. I figure that they were just postponing consideration of some of the hard questions. In some places, and with some systems, drivers have actually been taken out of cars (see Waymo in Arizona). However, driverless cars may still, for most people in most places, be a distant possibility.

Did you get any surprising feedback?
I was surprised that most comments were supportive. It suggests that the community welcomes critical engagement from other disciplines, which is a good sign.

Were you surprised or worried by any development in AI in 2019?
I remain extremely concerned about the possibility of AI widening inequality, and I don't see enough people within the AI community talking about this, which could mean that new injustices happen by default.

Do you have any specific hopes for AI for 2020?
I would like to see the nascent discussion about AI ethics develop into a mature discussion about AI and power. I want people to take seriously the question 'who benefits from AI?' and, if they don't like the answer, think about how to improve the governance of the technology. ❐

Alexander S. Rich[1], Cynthia Rudin[2], David M. P. Jacoby[3], Robin Freeman[3], Oliver R. Wearn[3], Henry Shevlin[4], Kanta Dihal[4], Seán S. ÓhÉigeartaigh[4], James Butcher[5], Marco Lippi[6], Przemyslaw Palka[7], Paolo Torroni[8], Shannon Wongvibulsin[9], Edmon Begoli[10], Gisbert Schneider[11], Stephen Cave[4], Mona Sloane[12], Emmanuel Moss[13], Iyad Rahwan[14],

Ken Goldberg[15], David Howard[16], Luciano Floridi[17] and Jack Stilgoe[18]

[1]Flatiron Health, New York, NY, USA. [2]Duke University, Durham, NC, USA. [3]Institute of Zoology, Zoological Society of London, London, UK. [4]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK. [5]United Nations Interregional Crime and Justice Research Institute, Centre for Artificial Intelligence and Robotics, The Hague, the Netherlands. [6]University of Modena and Reggio Emilia Reggio, Emilia, Italy. [7]Yale Law School, Center for Private Law, Information Society Project, New Haven, CT, USA. [8]University of Bologna, Bologna, Italy. [9]Johns Hopkins School of Medicine, Baltimore, MD, USA. [10]Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. [11]ETH Zurich, Zurich, Switzerland. [12]Institute for Public Knowledge, New York University, New York, NY, USA. [13]Data and Society Research Institute, New York, NY, USA. [14]Center for Humans and Machines, Max-Planck Institute for Human Development, Berlin, Germany. [15]UC Berkeley, Berkeley, CA, USA. [16]CSIRO, Brisbane, Australia. [17]Oxford Internet Institute, University of Oxford, Oxford, UK. [18]University College London, London, UK.

Published online: 17 January 2020
https://doi.org/10.1038/s42256-019-0141-1

References

1. Heaven, D. *Nature* **574**, 163–166 (2019).
2. Adams, R. *AI & Soc*. https://doi.org/10.1007/s00146-019-00918-7 (2019).
3. Rahwan, I. et al. *Nature* **568**, 477–486 (2019).
4. Ramos, J., Wang, A. & Kim, S. Human reflexes help MIT's HERMES rescue robot keep its footing. *IEEE Spectrum* https://go.nature.com/3580D0M (2019).