Viewpoints and debates in artificial intelligence

Alex Rich

9 April “Lessons for artificial intelligence from the study of natural stupidity”

Rich, A.S., Gureckis, T.M. Nat Mach Intell ***1,*** *174–180 (2019)*

Q: What was your Perspective about?

Machine learning algorithms can behave in harmful and biased ways when applied in high stakes arenas like criminal justice. Often, these algorithms are making decisions that were once left to another set of intelligent but biased agents—humans. Our article 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.

Q: Was there a specific reason or motivation to write the article?

The article came about through a convergence of factors. At the macro level, the alarm bells have been ringing for the last two or three years about the biases and negative impacts of machine learning systems, and there is a need for many perspectives on how to deal with these issues. 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.

Q: How 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 like healthcare where classic solutions may be impossible or unethical.
[1]

Q: Do you have any specific hopes for 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 let AI achieve more flexible and human-like behaviour, it could also

---

[1] The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non- exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the pub- lished form of this manuscript, or allow others to do so, for United States Gov- ernment purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

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.

Cynthia Rudin

13 May "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead"

Rudin, C. Nat Mach Intell **1,** 206–215 (2019)

Q: What was your Perspective about?

The goal of the article 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 reasons that I laid out in the article. 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, or were denied parole or loans.

Q: Was there a specific reason or motivation to write the article?

Yes! Policy makers are struggling with the questions how to regulate the use of machine learning models in practice, and the issue discussed in my article is at the heart of these 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.

Q: How has the topic developed over 2019?

The issue has not resolved. Policy makers need to be aware of these issues, as well as academics who can inform the policy makers. There is still a gap where many academics believe they need to sacrifice accuracy to gain interpretability of these machine learning models, despite the evidence that this is not necessarily true.

Q: Did you receive any surprising responses?

What was surprising to me was the lack of pushback I received. I fear that many people may have misinterpreted the paper, not actually reading its content properly. Some people have cited the work as a reason for not trusting black boxes, but then continued to discuss their work on explaining black boxes. Some people have written that I said we should avoid neural networks in order 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 article 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.

Q: Do you have any specific hopes for AI for 2020?

I have continued hope that policy makers will recognize the danger that society faces if it permits black boxes to make decisions that deeply affect people's 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 Jacoby

11 February "Responsible AI for conservation"

Wearn, O.R., Freeman, R. & Jacoby, D.M.P. Nat Mach Intell ***1,*** *72–73 (2019)*

Q: What was your Comment about?

Our article highlights the need for the AI community and conservation scientists to promote the responsible and ethical use of AI in conservation at a time when global biodiversity is in significant 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.

Q: 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 last few years. This is particularly true in areas such as the processing of data gathered from remote sensors like 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 but also the potentially negative consequence of automation in an attempt to outline where we would ideally like the field to go.

Q: How has the topic developed over 2019?

Anecdotally, we have seen an evolution in the discussions surrounding AI in conservation, with serious recognition now that an ethical question mark hangs over the technology. We predict 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.

Q: Did you get any surprising or useful feedback?

A number of AI ethics researchers reached out to us following publication of the article, 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.

Q: Has your own thinking on the topic evolved?

We realised that the generalisability 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.

Q Do you have any specific hopes for AI for 2020?

As AI becomes more and more ubiquitous in our daily lives, methodological developments are happening at a lightning pace as well as evolving independently across multiple sectors. 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 "Apply rich psychological terms in AI with care"

Shevlin, H., Halina, M. Nat Mach Intell 1, 165–167 (2019) doi:10.1038/s42256-019-0039-y

Q: What was your Comment about?

Our article examined the tendency of many researchers in machine learning and artificial intelligence to describe their systems using the vocabulary of psychology, or what we call rich psychological terms – concepts like 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 the valuable project of finding more novel and informative high-level descriptions of the capacities of artificial systems.

Q: 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.

Q: How has the topic developed over 2019?

There have been a lot of impressive developments in artificial language and reasoning tests over the last year, such as the GPT-2 language model from OpenAI and the striking results obtained by MT-DNN, another language model from Microsoft, on the 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 seem to me of growing importance.

Q: How has your own thinking on the topic evolved?

One point that only came into clarity for me after publication of the article 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.

Q: 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 advancement in this area, with potentially large ramifications for, e.g., the performance of chatbots.

Kanta Dihal

11 february. "Hopes and fears for intelligent machines in fiction and reality."

Cave, S., Dihal, K. Nat Mach Intell **1,** 74–78 (2019) doi:10.1038/s42256-019-0020-9

Q: What was your Perspective about?

Stephen Cave and I analysed around 300 narratives about artificial intelligence (fiction and nonfiction) and categorized the hopes and fears most commonly expressed in them. We showed how these four hopes and four fears are connected, and how losing control means a hope turns into a fear.

Q: 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 Artifical 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?"

Q: How 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.

Q: 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 not able to contribute 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.

Sean ÓhÉigeartaigh

7 January "Bridging near- and long-term concerns about AI"

Cave, S., ÓhÉigeartaigh, S.S. Nat Mach Intell 1, 5–6 (2019) doi:10.1038/s42256-018-0003-2

Q: 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 recognised. The decisions that we make now may have long-term consequences for how AI develops.

Q 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 one side to the other. Given the pace of progress in AI, combining insights from work addressing today's problems with more theoretical or foresight-oriented approaches seems crucial.

Q: 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 at which 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. The coming years will see more powerful systems developed and deployed in an increasingly digitised world; the range of uses and misuses of these advances will require careful forethought.

Q: 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 politicised, 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 politicisation carrying across.

Q: 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.

Q: Do you have any specific hopes for AI for 2020?

I hope to see greater global engagement and cooperation within the research community. 2020 will be the first year in which one of the top-tier machine learning conferences takes place in Africa, with ICLR being hosted in Addis Ababa. I also hope to see continued growth in collaboration between research communities in China, the US 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 "When seeing is no longer believing."

Beridze, I., Butcher, J. *Nat Mach Intell* **1,** 332–334 (2019) doi:10.1038/s42256-019-0085-5

Q: What was your Comment about?

AI is scaling the ability to produce synthetic media (text and audio-visual) and 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.

Q: 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.

Q 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. In the time since the article was published there has been greater coverage of this this topic, a growth in deepfake technology being deployed, and advancements 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 since the article's publication **[link:** https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/#2945d7e02241**].**

Q: 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 UN is also working on addressing the challenges. UNICRI, through its Centre for AI and Robotics, and the Data Science Initiative of the City of The Hague, having hosted a hackathon and workshop on deepfakes and manipulated videos in 2019 and we expect these activities to grow.

Marco Lippi

25 March "Consumer protection requires artificial intelligence"

Lippi, M., Contissa, G., Lagioia, F. *et al. Nat Mach Intell* 1, 168–169 (2019)
doi:10.1038/s42256-019-0042-3

Q: What was your Comment about?

In this paper we explore the ways in artificial intelligence 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 better their rights better and conduct oversight of business practices.

Q: Was there a specific reason or motivation to write the article?

Artificial intelligence 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 developments 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 [link www.claudette.eui.eu/demo]. Many interesting discussions within this inter-disciplinary research group led to this idea of a counter-power for, through the use of AI.

Q: 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.

Q Has your own thinking 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. This would also open the machine learning black-boxes towards more explainable and human-interpretable models.

Q Do you have any specific hopes for AI for 2020?

I hope that the AI community will continue its efforts to develop an 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 "Educational strategies to foster diversity and inclusion in machine intelligence"

Wongvibulsin, S. *Nat Mach Intell* **1,** 70–71 (2019) doi:10.1038/s42256-019-0021-8

Q What was your Comment about?

The article was about the importance of fostering diversity and inclusion in machine intelligence. I wrote about 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.

Q: 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 & Research Tutorials (HEART) program. My experiences with creating and teaching the course 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.

Q: Has your own thinking on the topic evolved?

Beyond the strategies discussed in the Comment (i.e. building a welcoming culture, student as teacher educational models, flipped classrooms, and longitudinal outreach programs), 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 men and women through advancements in machine intelligence that enable the integration of personalized education into daily life activities to excite and educate the next generation of the machine intelligence community.

Q: 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 advancement towards not only the development of increasingly powerful AI algorithms but also for their potential to be integrated into the healthcare system to augment clinical practice and patient care.

Edmon Begoli

7 January "The need for uncertainty quantification in machine-assisted medical decision making"

Begoli, E., Bhattacharya, T. & Kusnezov, D. *Nat Mach Intell* **1,** 20–23 (2019) doi:10.1038/s42256-018-0004-1

Q: What was your Perspective about?

In our article, we advocate for necessity of uncertainty quantification (UQ) research for AI in medical decision making.

Q: Was there a specific reason or motivation for you to write the article?

Yes, 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.

Q How has the topic developed over 2019?

It indeed has. We have seen an increasing focus and interest in "opening the black box of AI", and in a more formal understanding how AI works, and under what conditions it does not.

Q: Has your own thinking changed or evolved?

Yes, it has certainly evolved. We are seeing a very close, inverse connection between the UQ and adversarial AI. We are working now on the set of principles that take the UQ 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.

Q Do you have any specific hopes or expectations for AI for 2020?

For 2020, we hope to see a development of the safety culture in AI research, where an equal attention will be paid to uncertainty quantification, testing, validation, verification, and fairness of the AI models, as it is to their development and demonstrations of the capabilities (under ideal conditions). Also, we hope to see a bit less hype in the media, and a bit more maturation and critical views about the limitations and the capabilities of the AI. We fear the "AI winters" which usually follow the state of unrealistic expectations about the AI, so having a sober, realistic and objective view about the true state of the AI capabilities and its

limitations might prevent the next “AI winter”. We do believe that this time around, AI has its best chance ever to avoid such “change of seasons”, and a drop in public attention, funding, and investments into AI research.

Stephen Cave

7 January “Bridging near- and long-term concerns about AI”

Cave, S., ÓhÉigeartaigh, S.S. Nat Mach Intell 1, 5–6 (2019) doi:10.1038/s42256-018-0003-2

And

11 february. “Hopes and fears for intelligent machines in fiction and reality.”

Cave, S., Dihal, K. Nat Mach Intell **1,** 74–78 (2019) doi:10.1038/s42256-019-0020-9

Q: What were your articles about?

I had two articles in Nature Machine Intelligence last year. The Comment "Bridging near- and long-term concerns about AI" (with Seán S. ÓhÉigeartaigh) argued 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) categorised the four main categories of hopes and fears people have for AI, based on an analysis of a large corpus of fiction and nonfiction works.

Q: 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 which gives a gender theory based reading of Kanta’s and my schema of hopes and fears: ’Helen A'Loy and other tales of female automata: a gendered reading of the narratives of hopes and fears of intelligent machines and artificial intelligence’ [ref 3]

Q: 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’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.

Q: 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 significant bodies of literature. We need more works like Ruha Benjamin’s book ‘Race After Technology’ that forge links between these fields.

Iyad Rahwan

11 February “The evolution of citations graphs in artificial intelligence research”

Frank, M.R., Wang, D., Cebrian, M. et al. Nat Mach Intell **1,** 79–85 (2019)
doi:10.1038/s42256-019-0024-5

Q: What was your Perspective about?

Our article had 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.

Q: Was there a specific reason or motivation for you 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 paper, I was also working on a review article in Nature titled ‘Machine Behaviour’ [ref 1], 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.

Q 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 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.

Q: 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.

Q Do you have any specific hopes for AI for 2020?

The two fields that have traditionally studied human-machine systems were 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 "Robots and the return to collaborative intelligence"

*Goldberg, K. Nat Mach Intell **1,** 2–4 (2019) doi:10.1038/s42256-018-0008-x*

Q: 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.

Q: 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 MIT's HERMES project which uses human instinctive balancing reactions to remotely control a humanoid robot [ref 2].

Q: 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.

Q: 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 with '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 "Evolving embodied intelligence from materials to machines"

Howard, D., Eiben, A.E., Kennedy, D.F. et al. Nat Mach Intell **1,** 12–19 (2019) doi:10.1038/s42256-018-0009-9

Q: What was your Perspective about?

Multi-Level Evolution; a nature-inspired level-based 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 specialise to their tasks all the way from materials to machines.

Q: 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.

Q: How has the topic developed over 2019?

We see that the field of 'material robotics' **[authors: do you mean the field of soft robotics? or something else?]** is gaining a lot of momentum in the research community, and a way of autonomously designing robots using a wide range of materials is therefore more necessary now than before. The underpinning research areas for Multi-Level Evolution continue to mature, and a lot of the conversations we have with collaborators are around how we can work together to realise these architectures.

Q: Do you have any specific hopes for AI for 2020?

I would like to see more cross-field collaboration, and more standardisation. Both are critical to the ability to deploy AI systems (like ours!) at scale.

Jack Stilgoe

2 April "Self-driving cars will take a while to get right"

Stilgoe, J. *Nat Mach Intell* **1,** 202–203 (2019) doi:10.1038/s42256-019-0046-z

Q: 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.

Q: 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

Q: Do you feel the topic has developed over 2019, has the discussion moved on, and if so how?

There have been various self-driving car developers admitting that developing the tech has been harder than they anticipated. I figure 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.

Q: Did you get any surprising or useful 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.

Q: 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.

Q: 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.

Acknowledgments

This work has been in part co-authored by UT- Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The content is solely the responsibility of the authors and does not necessarily represent the official views of the UT-Battelle, or the Department of Energy.

References

1 Rahwan, I., Cebrian, M., Obradovich, N. *et al.* Machine behaviour. *Nature* **568,** 477–486 (2019) doi:10.1038/s41586-019-1138-y

2 Ramos, J., Wang, A/ and Kim, S. IEEE Spectrum, May 2019. https://spectrum.ieee.org/robotics/humanoids/human-reflexes-help-mits-hermes-rescue-robot-keep-its-footing

3 Adams, R. AI & Soc (2019). https://doi.org/10.1007/s00146-019-00918-7).