

Communicating AI: the role of researchers in enabling public conversations about AI

Note of discussions at a Royal Society workshop

Why AI communication matters

AI research is advancing at pace. Many people now interact with some form of AI-enabled technology on a daily basis – on social media, with virtual personal assistants, or through other autonomous systems – with further applications in development across a range of domains. AI research therefore increasingly affects people’s daily experiences of the world – the services they use, the news they see, the products they buy. While those deploying AI systems promise improved access, increased efficiency, and better-tailored experiences for their users, the scale and influence of these systems bring questions about the societal implications of AI technologies. In recent years, hype surrounding the potential of AI to boost economic growth and promote societal wellbeing has been tempered by growing understanding of the risks associated with AI technologies, often as a result of their disparate societal impact. Prominent examples of this impact in recent years have included news stories about:

- Women being less likely than men to be shown adverts for high-paid jobs through search engines¹.
- Racial disparities in how algorithmic risk assessment tools in the justice system predict the likelihood of recidivism².
- Mis-use of personal data for micro-targeting of political advertising³.

Public dialogues show that awareness of the term ‘machine learning’ is not high – a 2016 survey by the Royal Society showed only 9% of people surveyed had heard the term, despite this technology being the key driver behind many recent advances in AI. The term ‘artificial intelligence’ is more widely recognised, but comes with cultural perceptions that can influence public debate in different ways.

In this context, the way researchers, policymakers, and publics talk about AI technologies matters. It can direct attention to (or away from) areas of opportunity or concern, and it can influence how societies respond to technological advances. It can enable technological development, or hold it back.

A well-founded public dialogue about AI is key to creating an environment of careful stewardship for the development of AI technologies. AI researchers play an important role in this dialogue. As the field experiences a rapid expansion – illustrated by the scale of conferences such as NeurIPS – there are growing opportunities to bring researchers into public dialogues about AI.

1. Datta, A., Tschantz, M., and Datta, A. (2015) Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015 (1), 92 – 112. Available at: <https://content.sciendo.com/view/journals/popets/2015/1/article-p92.xml> (accessed 7 May 2020).

2. See, for example: MIT Tech Review (2019) AI is sending people to jail – and getting it wrong. Available at: <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/> (accessed 7 May 2020).

3. See, for example: The Guardian (2018) Cambridge Analytica Scandal ‘highlights need for AI regulation’. Available at: <https://www.theguardian.com/technology/2018/apr/16/cambridge-analytica-scandal-highlights-need-for-ai-regulation> (accessed 7 May 2020).

2019's NeurIPS communications practicum

The Royal Society is the UK's national academy of sciences. The Society's fundamental purpose, reflected in its founding Charters of the 1660s, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity. Reflecting this mission, its policy activities on data and digital technologies seek to advance these areas of science and technology for the benefit of society.

Building a well-founded public dialogue about AI technologies will be key to creating trustworthy AI technologies. Since the launch of its machine learning project⁴, the Royal Society has been creating spaces for public discussion about AI technologies, and their implications for society⁵.

To help advance discussions about the action needed to build a well-founded public dialogue about AI technologies, on 13 December 2019 the Royal Society and NeurIPS Conference co-convened a 'practicum' on communication skills. This note summarises discussions at the event. It is not intended as a verbatim record, and does not reflect an agreed position by workshop participants or the Royal Society.

Understanding what the audience brings to the table

In the absence of widespread public awareness of technologies such as machine learning, most people's views about AI will be shaped by the narratives that are part of our shared cultures. The ideas about AI technologies that are pervasive in public consciousness – typically that AI is an embodied, super-human intelligence – are shaped by hundreds of years of stories that people have told about humans and machines, and their places in the world. This cultural hinterland shapes how AI is portrayed in media, culture, and everyday discussion; it influences what societies find concerning – or exciting – about technological developments; and it affects how different publics relate to AI technologies.

BOX 1

Lessons from studying the narratives that influenced the development of other emerging technologies

A series of workshops in 2017 – 18⁶ explored the narratives prevalent in public discourse surrounding climate change, nuclear power, and genetic modification. These found the following lessons for those seeking to build a well-founded public debate about AI:

- Popular excitement and concerns about an emerging technology both influence public and policy debates.
- In navigating hype and uncertainty, it can be helpful to have a range of credible scenarios for how a technology might develop, which can be used in a public conversation. Clear models can illustrate uncertainties, alternative futures, and the implications of different interventions.
- Narratives of extreme fear can have potentially beneficial outcomes, for instance in ensuring safety concerns are considered at an early stage.
- Technology can be a lightning rod for broader social debates. It is important to understand which broader concerns or interest may be at play, and how these are bundled with questions about a specific technology. In these debates, perceptions of who benefits and who is at risk from technology development are important.
- Levels of public trust in scientists and technologists influences the perception and reception of new technologies.
- The language used to communicate scientific research is influential. Terminology has different meanings and effects in different communities.

4. Royal Society (2017) Machine learning: the power and promise of computers that learn, available at www.royalsociety.org/machine-learning (accessed 7 May 2020).

5. Such as the You and AI lecture series, available at: <https://royalsociety.org/topics-policy/projects/machine-learning/you-and-ai/> (accessed 7 May 2020).

6. Taken from Royal Society (2018) Portrayals and perceptions of AI and why they matter, available at: www.royalsociety.org/topics-policy/projects/ai-narratives/ (accessed 7 May 2020).

The utopian or dystopian visions of the future that define many of today's AI narratives tend to conjure specific images of AI – for example, a single, embodied intelligence – that do not well-represent the forms of intelligence that machine learning supports. They also often fail to convey the risks that are associated with today's AI, such as the potential to further marginalise vulnerable communities.

Alongside this cultural legacy, a community's social history can also play a role in shaping their responses to AI, and to the researchers developing it. AI could support a range of new tools to support vulnerable or marginalised communities, from systems that help preserve indigenous languages⁷ to projects that enable the provision of vital services⁸. To ensure that such systems work well for their users, it is vital that they are co-created with the communities they are intended to serve.

Effective communication with these communities requires researchers to listen to individual and community needs, and to understand the different cultural perspectives on AI and data they bring, before seeking to construct datasets or AI systems. While there may be enthusiasm for the use of AI within these communities, they often have a history of interactions with researchers that influence trust in such interactions today. Co-creating research challenges can also help identify new research directions that benefit both researchers and affected communities.

Being accountable to the audience

A range of stakeholders play a role in shaping public conversations about AI: companies highlighting their latest research or applications, universities promoting the impact of their work, governments raising the profile of national achievements, individual researchers discussing their work, and journalists investigating the implications. Each of these communication activities will have different objectives, and be subject to different pressures, incentives, and timescales for impact. In this crowded landscape, the incentives at play often favour eye-catching headlines that over-state or over-hype the significance of any individual research advance.

Hype about the potential of AI can contribute to expectations that technological advances and new applications will come quickly. While technological capabilities are advancing, there remain significant challenges in developing real-world AI systems that can be applied in complex domains, such as healthcare. Unrealistic expectations of progress can have significant impacts on individuals: for example, individuals could invest significant hope in the possibility of AI systems delivering cures for diseases, based on reported advances in AI that are, in reality, years from finding true clinical application. They can also have implications for how societies support the development of AI technologies, with repeated disillusionment about the rate of progress potentially laying the foundations for the withdrawal of funding or policy support – a shift that the AI research community has experienced previously in its history.

A key challenge for researchers seeking to contribute to public discussions is therefore the need to convey nuance, creating a clear message that allows space for excitement about what could follow a particular advance or application, while acknowledging the limitations of any individual paper. In this context, the presence of a range of voices that are able to provide accurate background information about – or de-bunking myths around – advances in AI can help improve the evidence base for public debate.

7. See, for example: <https://fnel.arts.ubc.ca/augmented-and-virtual-reality-for-cultural-and-linguistic-continuity-reigniting-the-many-voices-of-a-communal-bison-hunt/> (accessed 7 May 2020).

8. Many such projects were explored in the NeurIPS 2019 AI for social good workshop, the details of which are available at this link: <https://aiforsocialgood.github.io/neurips2019/> (accessed 7 May 2020).

Communicating about AI – lessons from 2019’s NeurIPS communications practicum

2019’s NeurIPS communications practicum highlighted the following lessons for researchers seeking to communicate about their work more effectively:

- Walk the audience through: providing background information about the fundamental concepts in your field can help people understand your findings.
- Think carefully about language: avoid jargon, which can make your work inaccessible, or over-loaded terms that can mis-convey the purpose of your work, or the capabilities of AI technologies.
- Recognise the needs of the newsroom: journalists are often working to different time pressures or incentives. Presenting accessible background information can help them create accurate stories, and preparing stories or anecdotes to humanise your work can help make an interview more compelling.
- Find alternative images: finding interesting images to accompany news articles about AI is a continuing challenge for many communicators. If there are images of work in your lab or in the field, these can be useful in helping frame articles about your research – making the research and its context more tangible to the reader.
- Create different channels for communication: a range of channels exist to talk about your work, each with different potential audiences and advantages. For example, research featured in an institutional press release can be explored in more depth via blogs.
- Respect your audience: while often not technical experts, your audience brings a range of other expertise and experiences to the table. These experiences shape how individuals react to your work – both positively and negatively. This cultural context plays an important role in effective communication.
- Co-create systems with affected communities: when seeking to create AI systems for the benefit of specific communities, it is vital to engage from an early stage with those communities, co-defining research challenges and negotiating ways of working.

Appendix: contributors

The Royal Society would like to thank speakers who participated in this event

David Duvenaud	Assistant Professor, University of Toronto
Zoubin Ghahramani FRS	Professor of Information Engineering, University of Cambridge, and Chief Scientist, Uber
Marzyeh Ghassemi	Assistant Professor, University of Toronto
Karen Hao	AI reporter, MIT Technology Review
Jeremy Kahn	Senior Writer, Fortune Magazine
Neil Lawrence	DeepMind Professor of Machine Learning, University of Cambridge
Zach Lipton	Assistant Professor, Carnegie Mellon University
Michael Littman	Professor in Computer Science, Brown University
Catherine Régis	Canada Research Chair in Collaborative Culture in Health Law, University of Montreal
Caroline Running Wolf	Executive Director and Co-Founder, Buffalo Tongue, Inc
Michael Running Wolf	Software Development Engineer, Amazon
Craig Smith	Founder Eye on AI, and former correspondent, New York Times
Hanna Wallach	Senior Principal Researcher, Microsoft Research New York City

© The Royal Society. Issued: May 2020 DES6826

The text of this work is licensed under the terms of the Creative Commons Attribution License which permits unrestricted use, provided the original author and source are credited. The license is available at: creativecommons.org/licenses/by/4.0