

# Room for improvement

Reflecting on 2020 brings into focus clear challenges for the year ahead, including for AI research that contemplates its broader societal impact.

Starting a new year feels less frivolous than usual. The next few months will be testing, as the world is in a race to subdue SARS-CoV-2 and well-planned vaccination programmes will be needed before the virus can wreak further havoc. The past year has given much to reflect on for AI research, including the lingering question of what its [role](#) could be in fighting the pandemic. Following last year's [‘AI reflections in 2019’](#), we asked several authors of Comments and Perspectives published in the latter half of 2019 and the first half of 2020 to discuss how the topic of their article has developed over the past year. We also asked authors about their hopes and expectations for AI in 2021. The resulting [‘AI reflections in 2020’](#) in this issue brings together viewpoints from researchers in AI who are writing about a variety of topics, including data science, AI ethics, robotics, cybersecurity and privacy-preserving AI. A common thread is that the pandemic has impacted everyone's work in some way. We hope you agree that the article provides an interesting snapshot of an unusual year through the lens of AI research.

In some cases, projects did not get the attention they deserved, as they were overshadowed by the pandemic emergency. For instance, one of the authors of the [‘AI reflections in 2020’](#) article, Aimon Jamjoom, writes about the survey he and colleagues put together on public opinions about liability in surgical robotics: who should be held responsible when a patient is harmed in a robot-assisted operation — the surgeon, hospital, the robot manufacturer or someone else? The [questions](#) raised by the survey are thorny and will need tackling in the near future.

Another challenge for 2021, mentioned among others by author Miguel Luengo-Oroz, is the carbon footprint of AI. The fight against climate change is reaching an [important stage](#) and AI research can, on the one hand, contribute positively with [‘AI for good’](#) applications such as optimizing

energy infrastructure and modelling climate change predictions. On the other hand, researchers have to recognize that AI is itself a [significant emitter of carbon](#).

One topic that has only been amplified by events of 2020 is AI ethics, and many authors of our [‘AI reflections’](#) article touch on AI ethics in some way. An underlying theme of their contributions is that what is needed most is solidarity: minorities and vulnerable communities should be able to benefit from advances in AI and be protected against harmful automated decision-making processes.

Last month, attention on AI ethics reached a high point at the annual Neural Information Processing Systems (NeurIPS) meeting, which took place online. Earlier in 2020, the conference organizers introduced a requirement for all authors to include a [broader impact statement](#) to reflect on ethical aspects and potential future societal consequences of research, including positive as well as negative effects. It has been left largely up to authors how to interpret this requirement, and browsing through a selection of accepted NeurIPS papers shows a variety of broader impact statements: some papers include only a brief statement, to indicate that the authors believe there are no foreseeable societal consequences, while many other papers focus mainly on positive impacts and mention how the advance may reduce problems with inherent bias in datasets. A few papers provide interesting considerations from authors who discuss potential negative outcomes that may not be immediately apparent from the work. For instance, one paper about a generative neural network approach to produce realistic images intended for uses such as photograph restoration acknowledges in the broader impact statement there are also possible harmful outcomes such as when [‘deepfakes’](#) are used in online harassment, defamation and identity theft.

The experiment with broader impact statements has caused debate. One worry is

that AI researchers may in general not be well placed to appreciate and reason about the potential impacts of their work, an issue that was explored during a day-long [workshop](#) at NeurIPS, which discussed the necessary involvement of various stakeholders in the process. While it seems clear that better guidelines for authors are required, the requirement for a broader impact statement is a welcome wake-up call for the AI community. AI research has developed fast over the past two decades, affecting much of society, such as experimenting and using human data with astonishingly little ethical oversight. There is a growing gap between AI research and biomedical research, which has a large amount of community consensus, guidelines and ethical oversight, for instance in [stem cell](#) research and [clinical trials](#). Similarly, the AI community needs to come together with further focussed discussions and initiatives to provide clear rules and guidelines in research standards, particularly in areas directly affecting humans.

Since last year, *Nature Machine Intelligence* has been asking authors for a broad ethical impact discussion for papers that are sent out for review where the main application involves the identification, classification or detection of humans or groups of humans. An example can be found in a [paper](#) in this issue on automated facial affect analysis and recently in a [paper](#) on detecting socio-economic patterns from satellite images of urban areas.

The topic is likely to develop further in 2021, and at high pace. The pandemic has engendered reflection and introspection, and a growing awareness of the role of science and technology in the public sphere. Scientists must respond with greater transparency and accountability on the potential applications of their methods and findings. Hold tight for another intense year. □

Published online: 19 January 2021  
<https://doi.org/10.1038/s42256-021-00294-2>