

not be a technology available for all. There are some difficult questions about who is likely to benefit and who is making the decisions.

Who's driving?

To think seriously about the ways in which self-driving cars might change the twenty-first century, the history of the motorcar in the twentieth century is a useful guide. The historian Peter Norton describes how, at the dawn of the motor age, a powerful lobby was able to convince American cities to hand over public space and build more roads in order to realize the potential of a new technology — the internal combustion engine car⁴. People without vehicles were relabelled as 'pedestrians' or, if they were in the road, 'jaywalkers'. As the United States and other countries transitioned to embrace cars, there were huge benefits, but also great costs. More than 40,000 people still die every year in the US because of cars, and many US cities are now trying to extricate themselves from car dependency. If we do not engage in a democratic discussion about self-driving cars now, governments will be bumped into making costly changes to suit a particular technology, which they may later regret.

How can we change direction?

The most urgent question is safety. The crash a year ago in which Elaine Herzberg was killed by a self-driving Uber SUV is still under investigation by the US National Transportation Safety Board (NTSB). Their report will help us understand the extent to which algorithmic outputs can be completely explainable. Self-driving car crashes have understandably received substantial public attention. We should not get stuck in a discussion of whether self-driving cars will be more or less safe than conventional cars. Machine learning will be a powerful

way to avoid human error, but if we want to see rapid improvements in safety, as we have with air travel, we also need processes of social learning that will allow social institutions to understand the uncertainties of new technologies and avoid the repetition of mistakes⁵. Advances in computer vision that have enabled the testing of self-driving cars have been remarkable, but some recent studies suggest that image recognition algorithms can be fragile and their failures hard to explain^{6,7}.

When planes crash, investigators use 'black box' data recorders to understand what happened and why, helping ensure that mistakes aren't repeated. The focus is on learning rather than blame. The NTSB has a slogan: anybody's accident is everybody's accident. My collaborators Alan Winfield and Marina Jirotko have argued that artificial neural networks will be hard to validate and verify. They propose that all robot systems should carry 'ethical black boxes' to help accident investigation⁸. Companies will be loath to share their data, but we know machine learning performs most reliably when algorithms can learn from as many experiences as possible, including accidents. The privatization of data could lead to a de facto privatization of transport, from which many will lose out.

Governments need to be clear about what they want from self-driving technology. At the moment, most have prioritized innovation for innovation's sake. They now need to agree principles — perhaps including safety improvements, free flow of traffic, sustainability, liveable streets or universal access to mobility — which will help guide innovators.

Science, according to Harvey Brooks, should be the 'conscience of technology', and machine learning is no different⁹. Researchers will play a vital role, not just in

making systems work, but also in finding new ways for them to be interpretable. This is not just about doing 'ethical' machine learning. It also involves honesty about the limits and downsides of machine learning. Decision makers will rely on machine learning researchers to tell them not just what the technology can do, but also what it can't do. People in transport policy need to know the ways in which machine intelligence differs from human intelligence so that they can make good decisions about the technology. A dialogue between experts in machine learning and those in transport planning will help put the technology in its place, which ultimately is the only way we will see the promised benefits. □

Jack Stilgoe ^{1,2,3}

¹Department of Science and Technology Studies, University College London, London, UK. ²Driverless Futures. ³Turing Institute, London, UK. e-mail: j.stilgoe@ucl.ac.uk

Published online: 02 April 2019

<https://doi.org/10.1038/s42256-019-0046-z>

References

1. Sage, A. *Reuters* <https://www.reuters.com/article/us-autos-autonomous-infrastructure-insig-idUSKCN0WX131> (2016).
2. Mindell, D. A. *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking, 2015).
3. Kahn, J. *Bloomberg* <https://www.bloomberg.com/news/articles/2018-08-16/to-get-ready-for-robot-driving-some-want-to-reprogram-pedestrians> (2018).
4. Norton, P. D. *Fighting Traffic: The Dawn of the Motor Age in the American City* (MIT Press, 2011).
5. Stilgoe, J. *Soc. Stud. Sci.* **48**, 25–56 (2018).
6. Alcorn, M. A. et al. Preprint at <https://arxiv.org/abs/1811.11553> (2018).
7. Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3156–3164 (IEEE, 2015).
8. Winfield, A. F. T. & Jirotko, M. *Philos. Trans. R. Soc. A* **376**, 20180085 (2018).
9. Brooks, H. *Res. Pol.* **23**, 477–486 (1994).

Competing interests

The author declares no competing interests.