Harvard Data Science Review • Issue 3.1, Winter 2021

Data Scientists Should Be Value-Driven, Not Neutral

Timothy Francis Clement-Jones¹

¹Select Committee on Artificial Intelligence, House of Lords, United Kingdom Parliament, London, England, United Kingdom

Published on: Jan 29, 2021

DOI: <https://doi.org/10.1162/99608f92.39876cea>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#)

Sabina Leonelli's early review of the use of data science during the COVID pandemic ("[Data Science in Times of Pan\(dem\)ic](#)," this issue) is a very timely piece of research in light of the work that both the European Union and United Kingdom are doing on data strategies, the huge contribution that data science can make to solving societal problems, and the accelerating digital transformation that is now taking place worldwide.

Given the extent and thoroughness of her review, Leonelli is surprisingly modest in making her headline conclusion that "fast data science need not be rushed." Her arguments are valid, particularly as regards the dangers of data collection through surveillance, but her excellent article, as I hope to show, can be mined for many equally, if not more pertinent, inferences—many if not most of them—having wider implications beyond the circumstances of the pandemic, as I outline in the following.

In going beyond COVID implications, I am reassured by her stated aim of stimulating discussion and nothing is more topical than her framing of the role of and relationship between, scientists and politicians, which certainly can and should be extrapolated well beyond COVID.

As a politician myself, I can say she is totally correct in saying that scientists are responsible for producing evidence and evaluating its impact but politicians, while being responsible for interventions based on it, can't simply claim to be 'following the science.' There are conflicts of evidence to be resolved and choices to be made.

Beneath the subtleties of her academic language lies trenchant criticism of U.K. and other government data policies, particularly in terms of the top-down nature and centralization of data collection, and the danger of skewed incentives, giving too much to some of the surveillance-based scenarios or "imaginaries" and too little to some of the more collaborative ones she advocates.

One of the first firm additional conclusions that can be drawn is that, however it is done, those harnessing our data need to build public trust. Many of us are fans of using data for a great many societal purposes, but we need to ensure we retain public trust while doing so, particularly when it comes to the increasing use of surveillance data.

For a start, if we are sharing data across borders and across industries, with the ever-greater rise of machine learning applications, in addition to ensuring high standards of data governance, we need to ensure that the information is authentic. As she describes, there are clear risks inherent in the potential lack of reliability of different methods of data collection and use, whether that use is in prediction or understanding causality.

I believe some of the remedies prescribed by Sir David Omand in his book *How Spies Think: 10 Lessons in Intelligence* (2020) offer valuable insights in terms of how we should do this, particularly in terms of contextualization, using Bayesian probabilistic inference to continually reassess information, interrogation of the facts, and avoiding group think.

Both Sir David and Leonelli emphasize the difference between data and useful, accurate, and actionable information. Cross-checking between different types of research is needed.

There are a number of other ways of tackling a public trust deficit. For public data such as health data, we need a transparent mechanism for its valuation. We also need to fast-track the development of data trusts (or social data foundations as they are increasingly called), designed to provide a sound and trustworthy governance vehicle for storing and sharing public data.

Moreover, as the recent follow-up report from my own House of Lords AI Select Committee has said (2020), as the deployment and use of artificial intelligence (AI) systems, and wider sharing of data accelerates, the public's understanding of that technology, and the ability to give informed consent, could be left behind. Governments must lead the way in actively explaining how data is being used. Being passive in this regard is no longer an option and ties in with Leonelli's stress on the need for community engagement and collaboration in data collection and use.

One of the big questions arising from the conclusions that Leonelli draws from her five imaginaries is whether any of them really are, per se bad, or whether we can consciously and transparently mitigate their risks.

A strong ethical governance framework is crucial in this respect, but in my view, Leonelli gets very close to the definitive answer when she says data scientists should stop feigning neutrality; values are important. Values should strongly underpin each of her imaginaries and should explicitly lie at the core of all data scientists' research and collaboration. That is by far the best form of risk mitigation and would allow us to be much more confident in data scientists using the full range of the imaginaries outlined by Leonelli available to them.

Disclosure Statement

Timothy Francis Clement-Jones has no financial or non-financial disclosures to share for this article.

References

House of Lords. (18 December 2020). AI in the UK: No room for complacency.
<https://publications.parliament.uk/pa/ld5801/ldselect/ldliaison/196/196.pdf>

Omand, D. (2020). *How spies think: 10 lessons in intelligence* New York, NY: Penguin.

©2021 Timothy Francis Clement-Jones. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.