

A Hippocratic Oath for data science? We'll settle for a little more data literacy

August 22 2019, by Lewis Mitchell And Joshua Ross



Credit: Karolina Grabowska from Pexels

"I swear by Hypatia, by Lovelace, by Turing, by Fisher (and/or Bayes), and by all the statisticians and data scientists, making them my witnesses, that I will carry out, according to my ability and judgment, this oath and

this indenture."

Could this be the first line of a "Hippocratic Oath" for mathematicians and [data scientists](#)? Hannah Fry, Associate Professor in the mathematics of cities at University College London, [argues that mathematicians and data scientists need such an oath](#), just like medical doctors who [swear](#) to act only in their patients' best interests.

"In medicine, you learn about ethics from day one. In mathematics, it's a bolt-on at best. It has to be there from day one and at the forefront of your mind in every step you take," Fry argued.

But is a tech version of the Hippocratic Oath really required? In medicine, these oaths vary between institutions, and have evolved greatly in the nearly 2,500 years of their history. Indeed, there is some debate around whether the oath [remains relevant to practicing doctors](#), particularly as it is the law, rather than a set of ancient Greek principles, by which they must ultimately abide.

How has [data science](#) reached the point at which an ethical pledge is deemed necessary? There are certainly numerous examples of algorithms doing harm – [criminal sentencing algorithms](#), for instance, have been shown to [disproportionately recommend that low-income and minority people are sent to jail](#).

Similar crises have led to proposals for ethical pledges before. In the aftermath of the 2008 [global financial crisis](#), a [manifesto](#) by financial engineers Emanuel Derman and Paul Wilmott beseeched economic modelers to swear not to "give the people who use my model false comfort about its accuracy. Instead, I will make explicit its assumptions and oversights."

Just as prejudices can be learned as a child, the biases of these

algorithms are a result of their training. A common feature of these algorithms is the use of black-box (often proprietary) algorithms, many of which are trained using statistically biased data.

In the case of criminal justice, the [algorithm](#)'s unjust outcome stems from the fact that historically, minorities are overrepresented in prison populations (most likely as a result of long-held human biases). This [bias](#) is therefore replicated and likely exacerbated by the algorithm.

Machine learning algorithms are trained on data, and can only be expected to produce predictions that are limited to those data. Bias in, bias out.

Promises, promises

Would taking an ethical pledge have helped the designers of these algorithms? Perhaps, but greater awareness of statistical biases might have been enough. Issues of unbiased representation in sampling have long been a cornerstone of statistics, and training in these topics may have led the designers to step back and question the validity of their predictions.

Fry herself has [commented on this issue in the past](#), saying it's necessary for people to be "paying attention to how biases you have in data can end up feeding through to the analyses you're doing."

But while issues of unbiased representation are not new in statistics, the growing use of high-powered algorithms in contentious areas make "data literacy" more relevant than ever.

Part of the issue is the ease with which machine learning algorithms can be applied, making data literacy no longer particular to mathematical and computer scientists, but to the public at large. Widespread basic

statistical and data literacy would aid awareness of the issues with statistical biases, and are a first step towards guarding against inappropriate use of algorithms.

Nobody is perfect, and while improved data literacy will help, unintended biases can still be overlooked. Algorithms might also have errors. One easy (to describe) way to guard against such issues is to make them publicly available. Such open source code can allow joint responsibility for bias and error checking.

Efforts of this sort are beginning to emerge, for example the [Web Transparency and Accountability Project](#) at Princeton University. Of course, many proprietary algorithms are commercial in confidence, which makes transparency difficult. Regulatory frameworks are hence likely to become important and necessary in this area. But a precondition is for practitioners, politicians, lawyers, and others to understand the issues around the widespread applicability of models, and their inherent statistical biases.

Ethics is undoubtedly important, and in a perfect world would form part of any education. But university degrees are finite. We argue that data and statistical literacy is an even more pressing concern, and could help guard against the appearance of more "unethical algorithms" in the future.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: A Hippocratic Oath for data science? We'll settle for a little more data literacy (2019, August 22) retrieved 29 June 2024 from <https://phys.org/news/2019-08-hippocratic-oath-science-literacy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.