A new method for ethical data science

March 21 2019, by Danil Mikhailov

Artificial Intelligence is transforming our world, sometimes in ways that its creators did not intend. In Wellcome Data Labs we are developing a new method of applying approaches from the social sciences to the way AI algorithms are produced to solve data science problems. The goal is to avoid potential negative consequences of the algorithms by identifying them earlier in the development process.

There have been attempts to set out such a way of working already. An example is Catalina Butnaru's excellent post proposing a new Agile ethics process. There is much to recommend this approach, not least that it is systematic and aligned closely in its steps to well-known steps of agile software development methodologies.

However, Butnaru does not address the mechanics of how her suggested Agile ethics process could be managed. Is it the team of data scientists and engineers themselves who are responsible for following the steps? Or their product manager? Or the UX team? Or a separate team to the engineers that audits their work?

We have been thinking about such questions a lot, since we are keen to test out how ethical approaches can be applied to the work of data scientists in practice and not just in theory.

The key challenge we set ourselves is: how to apply a process such as Butnaru's, or one of the other rival methodologies, in a way that measurably reduces ethical issues, like inadvertent bias, but does not reduce the energy and effectiveness of our Agile product teams?

We think this can be done by encouraging social scientists to work as part of interdisciplinary teams with software developers and data scientists, adopting their agile and iterative methodologies.

I have outlined some of the challenges of doing this. For example, the difficulty of getting social science researchers to work at the same speed and to the same rhythm as the software developers and data scientists. However, there is a potential template to follow by learning from the successful integration of the User Experience discipline into the software development workflows.

There is an additional challenge, though. Relying on a user researcher embedded in a product team to steer that team through an Agile ethics methodology on their own introduces the risk of them losing objectivity. This is a well-known issue in ethnographic research, where there is an active tension between a researcher's role as an impartial observer and the alternative of being an active participant.

A less technical way of looking at it is that people, fundamentally, are team players: they want to fit in and may find it difficult to criticise the work of their close colleagues. They might also become subject to 'group think' without realising it.

In Wellcome Data Labs we have worked out a paired approach to Agile ethics which is intended to resolve this issue. Our proposed methodology has three steps:

1. Embedding within Data Labs a user researcher with a background both in working as part of Agile product teams and in carrying out social sciences research. This embedded researcher will have the explicitly defined objective of testing the algorithmic models the software developers and data scientists are working on from the point of view of their possible social

impact.

2. They will adjust and develop their analysis iteratively to match the speed of the technology work and feed their emergent conclusions back to the data scientists to steer the course of their work.

3. The embedded researcher will be paired up with another social scientist outside the team to provide an objective critique and the necessary checks and balances on their analysis.

All three parts of the proposed methodology are equally important.

- Not embedding the researcher in the team would make it hard for them to have a close enough knowledge of what the data scientists are doing.
- Not iteratively retesting and rewriting their analysis of possible social impact will fail to match the rhythm of the technological development – the key proposed advantage of this methodology.
- Finally, the pairing is designed to prevent the embedded researcher risking a loss of their professional detachment and objectivity, which is a risk precisely because they are so closely embedded within the technology teams.
- This whole approach is an experiment in itself and we are not at all certain that it will work. However, that is exactly what makes it exciting to us. We hope it will help us become better aware of the biases being introduced by the algorithms that we develop and minimise any potential negative unintentional consequences of the tools the team produces.

This is important because Wellcome, as a significant funder of scientific research, has a notable impact on the academic and health industries. And Wellcome Data Labs' analysis feeds into Wellcome's decision making process. Any unintended biases in the algorithms my team produces that can impact Wellcome's decisions, could have a ripple

effect on the decisions of more funders, which in turn could cascade down to secondary impacts on other industries and the wider society. We have a responsibility to get it right.

Provided by Wellcome Trust

Citation: A new method for ethical data science (2019, March 21) retrieved 18 April 2024 from https://phys.org/news/2019-03-method-ethical-science.html