

# Data Ethics, AI and Responsible Innovation

Nick Hood

05 November 2020



# Contents

Introduction	5
Week 1: Ethics and Law	7
References	13



# Introduction

These are my notes from participation in the Edinburgh Data Ethics MOOC, running in November and December 2020. Any errors here are my own responsibility.

Nick Hood  
@cullaloe

This document was last updated on 05 November 2020 at 20:24.

## Course overview

**Week 1 Ethics and Law** What are ethical values? Can we rely on legal regulation? What are the most pressing issues facing data driven industries?

**Week 2 Crime and Justice** Should we use predictive policing and sentencing algorithms? How can biases sneak into such algorithms? How can we remove them?

**Week 3 Home and City** What are the promises of smart homes and cities? How can they impact our privacy and freedom? How can we design them to protect those values?

**Week 4 Money and Markets** Can future tech lead to a world without money? Can algorithms help us fairly distribute resources? How could we design a fair AI?

**Week 5 Life and Health** Should we keep genetic databanks? Would you trust an AI doctor? What are the principles of responsible research and innovation?



# Week 1: Ethics and Law

This starts with some housekeeping, introductions to the tutors and setting up a profile within the course. Dr. Ewa Luger gets us started by considering a “broad overview” of **ethics**, using a radio interview about the Cambridge Analytica (CA) scandal as a resource to get us thinking before we embark on a more detailed introduction to ethics.

## Ethical ‘Issues’

In the clip, problematised issues with the CA matter included:

- the way data was gathered from people
- the fact that the data was then passed on to another party
- that the data was now available for commercial use
- that the data was used for political purpose
- the FaceBook (FB) login on an app gave permissions for all of users’ FB data to be used and shared
- people aren’t outraged, or even bothered by this enough to take action

**Discussion forum task** Think back over your professional life and identify one example of an ethical issue that you personally experienced. In the discussion forum, write a brief summary (around 50-100 words) describing this issue - try to explain the context, what happened, and why you felt it was an ethical issue. After you’ve done this, comment on two other posts.

I had to take a quick look at the forum first to gauge the level of ethical matters that were being shared. There seems to be a discernible difference between **moral**, **legal** and **ethical** matters in the revelations on the forum. My own example came from quite a selection of professional experience that ranged from legal but immoral, moral but illegal, unethical but moral and legal. I picked an example from a while back.

I was project manager in a business that made flight crew training simulators. These devices were multi-million-dollar complex machines that used specially designed and manufactured circuit boards

mounted in standard racks to drive the various components of the device. I had an internal review one morning that failed badly when the equipment that had been working well the night before suddenly failed to function. It was extremely embarrassing for my team. We discovered later that day that in the night, another project manager had switched a key circuit board for a faulty one from his own machine.

The other examples I commented on or read seemed to me to be not so much ethical issues as difficult choices or situations. This is causing me to wonder if I have a secure understanding of the term, “ethical”. Hopefully the next section will clear that up for me.

## The challenge of data-driven innovation

Issues arising from the 2018 Cambridge Analytica scandal are identified as:

- the value derived from data is not evenly distributed
- power rests in the hands of very few
- the potential for harm and inequality is high
- the potential impact of bad actors

These are, in my view, issues of power arising from wealth inequity and not particularly related to anything technological. The difficulty is that technology amplifies the effects (as it amplifies inequities in a classroom – this much we have learned in moving to “Digital First” pedagogies resulting from the COVID measures).

The basic principles of ethics are described as, “how we should address these issues, whilst minimising harm, ensuring morally good outcomes and fairness, and protecting human autonomy.”

## Data at the heart of it

For the purposes of the course, this definition of ethics is set in the context of *data-driven innovation* which makes use of large amounts of data to train algorithms that make predictions or provide insights for decision making. Machines are increasingly now involved in the decision making too, and in taking actions. These machines are described as *intelligent*.

## Visibility of rationale

These machines include, of course, neural networks and we know from the way that these pattern-matchers work, that other than the input and output layers,



it is not possible to determine the path of the rationale – the specific choices of selections being made in the intermediate steps – that underlies the outputs.

## Ambient intelligent systems

Interactional moments are perceived to provide friction in the operation of systems, therefore designers are removing these for a smoother experience – think of entering your password every time you wanted to look at your phone, now replaced with fingerprint or face recognition. The result is that such interactions become invisible to the user: it is the environment we interact with in a natural way. Unnaturally, *the environment is reading us to discern intent*.

## Human → AI interaction

This is problematised in a video clip, suggesting that this interaction is not yet defined. I rather think that Turing (1950) had a clear enough view of what it might look like, but recent innovations have increased *system opacity* such that understanding how the system makes choices based upon how it perceives us is difficult. In some sense, we are already familiar with this idea with our less tech-savvy relatives who need support working with machines. It gets harder when you don't know there's a machine.

## Utility

Comparison with basic utilities like water, power, etc., is made from a very privileged first-world perspective: *we only notice them when they fail*. These systems were unregulated in their early days, too, and some of the horrors of human behaviour (Edison, for example) are long forgotten. As data-driven innovation moves to infect our lives by becoming a utility, the question is asked, “should it be regulated?” First steps have been made in this, of course, and we see the return of interactional friction with things like the cookie quiz on GDPR-compliant websites. The monetisation of data has led to the term, “surveillance capitalism”, which describes how our identity and behaviours have become the product in a high-value industry.

## Policy lag

Clearly, policy makers are well behind the curve when it comes to data-driven innovation. GDPR and similar legislation have added friction yet advances are made at speed. The additional reading (Wagner, 2019) is critical of policy makers and suggests that ethics is what corporations do to avoid government interference.

## Values

What we value in society is a cultural attribute. Values are therefore dependent on context and are not absolute: I have often said that every principle has a price attached to it, and that if the price is right, all principles are for sale. This has caused strong reaction when vocalised that way, especially from those who feel that they themselves hold high principles in comparison with others. Clearly, passions are aroused when deeply-held values are challenged, as the Guardian article on Peter Seeger illustrates.

Examples are discussed in the course material to illustrate the point that values depend very much on context. This statement is a call to action, perhaps, talking of moral dilemmas and conflicting values:

“Our job, as ethical actors, is to identify what they are, and then negotiate how best to instantiate and balance them.”

Research is as robust as you’d expect for the social sciences, with this, from a study of a small sample of people ( $N < 700$ ) in three countries with similar (colonial) cultural pedigree:

“... people endorse the same values to a similar extent across countries and also instantiate them similarly.” (Bardi et al., 2018)

Not much of a surprise there, given the evolutionary path of the handful of people studied.

## Case studies

Project Maven and the Google employee revolt against it is presented as the first case study of how corporations can make choices that do not align with the values of the stakeholders, in this case, the staff. That project sat in stark contrast to Google’s “Don’t be evil” mantra, still enshrined in its code of conduct (cf. the characterisation of ethics as an escape from regulation by Wagner (2019)). The military application of data-driven innovation is clearly compelling, and others are emerging, including Clearview AI’s face recognition app, which the company has worked hard to present as ethical use of technology.

## What are ethics?

A branch of moral philosophy: a codification of habits that are valued within the context of a culture: ergo, no right or wrong answers, just human custom. *Normative* ethics are those that define what we *ought* to do, without concerning themselves with what we *actually* do. Within that class of ethics are 3 perspectives on how we should act:

**Deontological ethics (rules)** rules of duty and obligation – perhaps universally agreed upon, like not killing each other, or hurting animals. Some things are always right, others, always wrong, no matter the consequences.

**Teleological ethics (consequences)** focus on the outcome of actions – actions for the greater good, perhaps, or the end justifying the means. *Consequentialism* is the view that the moral quality of a choice is decided solely by its outcome.

**Virtue ethics** are about judgement of people’s character or moral fibre.

The obvious issue here is that life is complicated, isn’t it? We can’t make a simple set of rules because the complexity of life requires us to make choices sometimes that go against those simple rules for a better end. Is there any point, then, in having rules in the first place? Well, clearly, because we want everyone else to make choices that don’t harm or disadvantage us.

## The trolley problem

**Scenario 1** A runaway train is travelling on a railway track towards 5 people. You can’t warn them, but there is a lever that you can operate to switch the tracks. The problem is that there is a person on the other track. **Do you pull the lever or not?**

**Scenario 2** As above, but now there are no points – instead, a large person standing a bridge over the tracks. **Do you push Fatty into the path of the train to save the five?**

I chose yes in both scenarios.

“The needs of the many outweigh the needs of the few.” – Spock

Interestingly, although I was with the majority in the first scenario, I was in the minority in the second.

## Making moral decisions

So, are ethics about choices, or outcomes? Rules or consequences? How does the trolley problem change when the 5 are convicted rapists? Children? When Fatty is a scientist working on a cure for Leukemia? When the 5 are Mountain People (insert your own lower-class of person)? Tories?

MIT are crowdsourcing moral acceptability with their moral machine project in which you are presented with multiple scenarios (like driverless car choices) in which you decide the choice the vehicle should make, usually resulting in somebody or something dying. It is easy to become quickly abstracted from the awfulness of the choices you make in this game.

Guidelines in ethical frameworks for machine choices are broadly divided into four categories:

- Do good
- Minimise harm
- Respect human autonomy
- Be just or fair

Transparency of the algorithm and accountability are increasingly being emphasised. These frameworks have become regarded as inadequate as they offer a way for corporations to hide behind them in what is called, “ethics washing”.

# References

- Bardi, A., Holloway, R., Lönnqvist, J.-E., Bevington, P., Hanel, P. H. P., Maio, G. R., Soares, A. K. S., Vione, K. C., De Holanda Coelho, G. L., Gouveia, V. V., Patil, A. C., Kamble, S. V., & Manstead, A. S. R. (2018). *Cross-Cultural Differences and Similarities in Human Value Instantiation*. <https://doi.org/10.3389/fpsyg.2018.00849>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Wagner, B. (2019). Ethics As An Escape From Regulation. From “Ethics-Washing” To Ethics-Shopping? In M. Hildebrandt (Ed.), *Being profiled* (pp. 84–89). <https://doi.org/10.1515/9789048550180-016>