ETHICS IN MATHEMATICS DISCUSSION PAPER 3/2018

University of Cambridge

How the UK and France are dealing with the AI revolution

A comparison of the approaches the two governments are taking

This discussion paper compares two parliamentary documents on Artificial Intelligence written in 2018: 'Algorithms in Decision-Making,' written by the Science and Technology Committee for the UK government, and 'For a Meaningful Artificial Intelligence,' written by a team headed by the Fields Medalist Cédric Villani for the French government.

Maurice Chiodo | Lara Gordon
mcc56@cam.ac.uk | lara.isobel.gordon@gmail.com

How the UK and France are dealing with the AI revolution
by Maurice Chiodo and Lara Gordon

EiM Discussion Paper 3/2018
University of Cambridge
Ethics in Mathematics Project

Maurice Chiodo is a research mathematician at the University of Cambridge, where he runs a lecture course on ethics in mathematics and is PI of the University of Cambridge *Ethics in Mathematics Project*. You can contact Dr Chiodo at mcc56@cam.ac.uk.

Lara Gordon is a third-year mathematics undergraduate student at the University of Cambridge and was part of a team led by Dr Maurice Chiodo who were working on the *Ethics in Mathematics Project* over the summer of 2018. You can contact Lara at lara.isobel.gordon@gmail.com.

Abstract

In this paper we look at two parliamentary reports, by the UK and France, on artificial intelligence. The first, titled "Algorithms in Decision-Making," is a report by the Science and Technology Committee ordered by the House of Commons in the UK (Lamb, et al., 2018). The second, "For a Meaningful Artificial Intelligence," (Villani, et al., 2018) is a report by a team assembled by Fields Medallist and Member of the French Parliament Cédric Villani, under the instruction of the French Prime Minister Édouard Philippe.

Both reports were published in 2018, and offer insight into how the United Kingdom and France are thinking about and dealing with the growth and impact of Artificial Intelligence in society. We compare and contrast these two reports, investigating how each addresses common points, and identifying points that are addressed in one report but not the other. We then give our own analysis and commentary, from the point of view of our work on ethics in mathematics, on issues addressed in and arising from these two reports.

Contents

Introduction 2
1. Summary of the British report 4
1.1. Who wrote the report? 4
1.2. What does the British report say? 5
2. Summary of the French report 6
2.1. Who worked on the French report? 6
2.2. What does the French report say? 7
3. A comparison of the two approaches 8
3.1. Being part of the algorithms revolution 8
3.2. Using AI 11
3.3. Research 14
4. A comparison of ethical issues 15
4.1. Data protection 15
4.2. Algorithmic bias 16
4.3. Transparency and accountability 18
4.4. Impacts on jobs and employment 20
4.5. Ecological considerations 21
4.6. Governance of ethics in AI 22
5. Conclusion 23
6. Main references 24
7. Further Comments 25
7.1. Researchers, data scientists and mathematicians. 25
7.2. AI we can't switch off 25
7.3. The UK select committee interview of Cambridge Analytica. 25
7.4. Predictive policing algorithms 26
7.5. A search engine is a distortion machine 26
7.6. Portability of data 27
7.7. The privatisation of censorship 28
7.8. Taking copyright law into their own hands 29
7.9. Building auditable algorithms which are more constrained, but also more transparent. . 29
7.10. Shifting the culture of "It's not my problem" 30
7.11. Proposing legislation and accreditation 30
7.12. The borderless nature of algorithms 31

How the UK and France are dealing with the AI revolution.

Introduction

The notion of an *algorithm, a* precise step-by-step sequence of mathematical instructions to be followed to carry out a particular task, has been part of human knowledge for many centuries. With the improvements in speed and cost of computation over the last 70 years, algorithmic processes are becoming increasingly prevalent in society. They can bring about substantial increases in efficiency, but sometimes their implementation has other costs, such as introducing bias into processes unexpectedly. Historically, the steps of an algorithm were clearly laid out by its designer; usually someone with a technical background such as a mathematician or computer scientist. However, we are now entering an era wherein such design methods are being replaced by *machine learning*, whereby the individual steps and parameters of an algorithm are not laid out explicitly by the developer, but instead *learned* through analysing a large data set to "figure out" what these steps and parameters should be[1]. Let us be more precise: the developer sets up some *incentives* for the algorithm to adjust itself against (such as specifying how "close" to correct it has been when given some testing data to learn from), and the algorithm modifies itself to meet these incentives as closely as possible. At the end of this learning process, the algorithm is then able to process *arbitrary* input data[2] and (hopefully) give meaningful output. An algorithm formed in this way is often described as being a demonstration of *Artificial Intelligence;* an algorithm that has inferred things for itself from data sets. An example of this would be an algorithm to infer gender from a photograph of a face; the algorithm is "fed" a large data set of pictures of faces to learn from, each one with an associated gender tag. The algorithm reads one picture from the training data set, and then tries to "figure out" what characteristics of a face point to whichever gender. It then "computes" the gender of the picture, and then compares its computation to the gender tag. If it is wrong, it makes some small internal changes to its steps and parameters, and then reads the next sample picture. After doing this on thousands, or millions, of pictures, these algorithms can become quite (but not completely) efficient at determining gender from a picture. That is, when given a picture with *no* gender tag attached, the algorithm can compute the gender of the person in the picture[3]. We use the term *machine learning* to denote the

1 The data that the algorithm uses to "learn" from is often referred to as *training data.*

2 This is *not* the original set of training data, but instead the *actual* inputs that we wish to compute things about.

3 Another example could be the following: taking as a training set the medical scans of thousands of patients with a particular type of cancer. Then, when the algorithm has "learned" all the necessary parameter and steps, it can be given a scan of a patient who is yet to be diagnosed, to compute whether

process of producing an algorithm in this "self-learning" manner. Throughout this document we will refer to *Artificial Intelligence* (AI) to mean *an algorithm developed via machine learning*.

In recent years, it has become increasingly clear to governments and citizens that Artificial Intelligence is, and will, play a larger and larger role in society as time progresses. In light of this, both the British and French governments (independently) ordered reports relating to the growth and impact of AI on society. The British produced a document titled "Algorithms in Decision Making" (Lamb, et al., 2018), prepared by the Science and Technology Committee, commissioned by the House of Commons (hereafter referred to simply as the *British report*). The French produced a document titled "For a Meaningful Artificial Intelligence" (Villani, et al., 2018), prepared by a specialist team assembled and led by the Fields Medallist Cédric Villani, commissioned by the French Prime Minister Édouard Philippe (hereafter referred to simply as the *French report*). These two reports address and comment on various aspects of AI, how it might impact society, what consequences and outcomes their respective national governments should be aware of and act upon, and give some medium to long-term projections on where this technology might be headed.

The two reports have slightly different remits, and cover slightly different topics, but they are similar enough in both aim and scope that we thought it would be worth writing a comparison of the two. We note that the French report is substantially longer and more detailed than the British report.

Given how close the UK and France are in terms of geographical location, population, and GDP, it is interesting to compare how these very similar nations are addressing the big questions arising from AI. We have approached this from the point of view of our own work on *Ethics in Mathematics*[4], and are concerned ultimately with the ethical issues arising from the development of AI and what potential *human* benefits, and drawbacks, may arise. Thus, a large part of our paper focuses on ethical issues.

Our paper is structured as follows:

We begin by giving a brief summary of each report in sections 1 and 2. These include outlining who is working on each report, and then giving a short summary of the main points and topics covered in each report. We then go into a deeper comparison of the two reports in section 3, starting off by listing various technical topics addressed in both reports and comparing what is, and isn't, presented in each report regarding them. These topics are:

1. Being part of the algorithms revolution
2. Using AI
3. Research

or not that *new* patient does or doesn't have that particular type of cancer. This "diagnosis" done via such a machine-learned algorithm may be incorrect, but is hopefully correct with high likelihood.

4 The *Cambridge University Ethics in Mathematics Project:* https://www.ethics.maths.cam.ac.uk/

As our interest lies predominantly around the *ethical* issues that arise when doing mathematical/technical work, we present and address, in section 4, the ethical issues addressed by the two reports. These ethical issues are:

1. Data protection
2. Algorithmic bias
3. Transparency and accountability
4. Impacts on jobs and employment
5. Ecological considerations
6. Governance of ethics in AI

After making a short conclusion in section 5 on the comparison of the two reports, and providing references in section 6, we use section 7 to provide some further comments, presenting our own analysis and commentary on the issues arising from the two reports. This section contains our own original ideas, motivated by, and relevant to, the content of the British and French reports.

Throughout this paper we refer to algorithms many times. As this is a paper motivated by our study of Ethics in Mathematics, we are concerned with the *use* of algorithms, and not just the technical nature of what they do. We note here that algorithms can be used in many ways, all of which raise their own ethical challenges. We look at d*ecision-making algorithms;* those that "decide" what to do in certain instances, such as whom to admit to university, or what prison sentence to give to a convicted criminal. We look at i*dentification algorithms;* those that identify a characteristic of a particular thing, such as whether someone has cancer, or whether a credit card transaction is fraudulent. And, we look at p*redictive algorithms;* those that predict what an entity or system will do in the future, such as where crime is likely to occur (as part of police deployment tools), or what value certain shares will take on the stock market in the future. These are all places where algorithms are used, and although these uses are quite different, it is still an algorithm that is being used to "do" something.

The comparison part of this paper was (predominately) written by Lara Gordon as part of an undergraduate summer project, with the comments at the end added by the project supervisor: Maurice Chiodo. Both authors were fully involved in the overall production of this paper.

We wish to thank Piers Bursill-Hall for suggesting that we write this paper and for his helpful comments throughout, Toby Clifton for his thorough proof reading and insightful suggestions, and Dennis Müller for his editorial recommendations.

1. Summary of the British report

1.1. Who wrote the report?

Understanding the background of those who worked on the British report provides good context for the content within it. The British report is one of many written by the Science and Technology

Committee.[5] This committee is expected to write several reports a year, on a variety of topics under the broad umbrella of "Science and Technology"[6], so specific expertise on algorithms and AI can not necessarily be expected[7]. Unfortunately, it is unclear to us what exact expertise those on the Science and Technology Committee possess, and their membership changes somewhat frequently. What we do know is that they are all MPs in the UK parliament, so any specific expertise may well be by chance. As the title of the British report suggests, it focuses on the use of algorithms in decision making, including the use of artificial intelligence.

1.2. What does the British report say?

The British report came about from the work of Dr Stephanie Mathisen, a policy officer at "Sense about Science"[8], as a result of questioned she raised to the Science and Technology Committee. She felt that the extent to which decision-making algorithms can exacerbate or reduce biases needed to be investigated, and that there is a need for decisions made by algorithms to be challenged, understood and regulated. Under recommendation from this same committee, the Centre for Data Ethics & Innovation[9] has recently been established. The British report provides several "themes and challenges" for this newly established body to consider as they begin their work. The report is written with an understanding that the growth of big data and machine learning has had a large effect on algorithmic decision making in recent years and is likely to continue changing the way decisions are made. Therefore, machine learning algorithms are the focus of the analysis and ideas presented, with additional reference to the data issues that go hand in hand with this.

To summarise the issues discussed, the committee recognises that data play an important role in the development of algorithms and that generally such algorithms are more effective and valuable if more data are available, and therefore a recommendation is made for the creation of "data trusts". The purpose of these would be to show that the government is supporting the "algorithms revolution". The committee also recognises that with the development of machine learning algorithms there are data protection issues to consider. They quote the newly instated EU General Data Protection Regulation (GDPR) as an example of helpful protection for those affected by algorithms and for those whose data are used in the development of algorithms (Lamb, et al., 2018, p.4).

The question of bias within algorithms is also raised (Lamb, et al., 2018, p.18). There is a large focus on the training data required by machine learning algorithms being the cause of bias. There is also a suggestion that teams developing algorithms should contain people with backgrounds

5 To read these reports, visit: https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/

6 https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/publications/

7 In 2018 the reports they produced covered areas such as flu vaccines, genomics, and e-cigarettes.

8 http://senseaboutscience.org/team/dr-stephanie-mathisen/

9 https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation

from a sufficiently wide cross-section of society, to avoid inherent bias appearing in the algorithms being produced. The Committee recommends that the Centre for Data Ethics & Innovation investigate accountability tools in order to integrate algorithmic decision making into society (Lamb, et al., 2018, p.32). This includes the potential use of regulatory bodies. Transparency of an algorithm is claimed to be a key underpinning for algorithm accountability, and there is discussion about how to achieve as much transparency as possible.

2. Summary of the French report

2.1. Who worked on the French report?

The mission assigned to Villani was to set up a French and European Strategy for Artificial Intelligence. Villani was given six months, from 8th September 2017 to 8th March 2018. Villani gives a foreword in the French report summarising his thoughts on being asked to head the mission. He makes it clear that AI is becoming a vastly important part of modern society, and that it is relevant to mathematicians like himself; even as a pure mathematician, his research on optimal transport has been cited in articles about AI. It is for this reason, as well as his underlying interest in the subject, that he was unsurprised that he was asked to head up this mission. Villani began by assembling the following diverse team, including some members with highly technical training, to work on the project:

- Villani himself is a mathematician and 2010 Fields Medallist and has held several positions at several foreign universities. He is also a member of the French Parliament and is vice-president of the Parliamentary Office for Scientific and Technological Options Assessment.
- Marc Schoenauer is Principal Senior Researcher with the French Institute for Research in Computer Science and Automation and was previously a full-time researcher with the French National Research Center working at the Applied Maths Laboratory.
- Yann Bonnet, General Secretary to the French Digital Council, is an engineer by training.
- Charly Berthet is a lawyer and Head of Legal and Institutional Matters at the French Digital Council.
- Anne-Charlotte Cornut is Rapporteur of the French Digital Council.
- François Levin is a philosophy graduate and currently Head of Economic and Social Affairs at the French Digital Council.
- Bertrand Rondepierre is an engineer in the Cops de l'armement working for the French defence procurement agency.
- Stella Biabiany-Rosier is the executive assistant of the French Digital Council.

Villani describes this as "a 'dream team' of seven highly competent individuals of diverse backgrounds, dedicated full time to the task force," stating that this was crucial for the project's success.

2.2. What does the French report say?

In contrast to the British report, one of the main purposes of the French report is to find a way for France to become a major global player in Artificial Intelligence advancement (Villani, et al., p.6). Whilst France is at the forefront of worldwide exploration on mathematics and AI, the report recognises that the country's scientific progress does not always translate into concrete industrial and economic applications (Villani, et al., p.10). Part of their plan is to improve the resources available to researchers by setting up a supercomputer exclusively for AI research and development, to make public research careers more attractive, and to change the university courses available for people to take, for example joint Law and AI degrees could be created (Villani, et al., p.13).

The report also discusses the social issues surrounding the integration of AI into society (Villani, et al., p.81). The mission has identified issues, such as staff feeling alienated by the technology they may need to work with. They give potential solutions to these problems, which may arise should an "AI take-off" occur. One potential solution offered is a formal education and lifelong learning overhaul to equip people with skills in creative thinking that are becoming more vital for AI development. They say it is important to recognise issues such as these to keep society running smoothly in a technological transition.

For our purposes, Chapter 5 (Villani, et al., p.112) is particularly relevant, entitled "What are the Ethics of AI?" The French committee has started to build an ethical framework for the development of AI. They lay out five principles:

- **Greater transparency:** They recognise that explaining the decisions made by machine learning algorithms[10] is difficult for technical reasons, because of the way such algorithms are built, and the way that they operate. It also raises questions as to whether algorithms whose decisions cannot be explained should be used in situations as crucial to the life of an individual as access to credit, employment, accommodation, justice or health. They also make the point that accountability of technology is one of the conditions for its social acceptability and that if the technology is not socially accepted, many of the benefits it could bring to society could be lost.

- **Incorporating ethics into the training of engineers and researchers working on AI:** They recognise that technical people are not getting enough training in ethical issues which are likely to become only more prevalent in technical areas in the future. The article says that these people have a certain amount of responsibility with regards to the effects of the artificial intelligence they create but recognises that they are not necessarily equipped to consider these issues fully. Therefore, they fully recommend that this type of training is included in the education of engineers and researchers.

- **Protection of rights and freedoms:** They recognise that current legislation on data protection generally only deals with personal data and that many issues raised by the use of algorithms now constitute a "blind spot" of the law. In the context of deep learning, data are used on a massive scale to produce correlations which could affect whole groups of

10 In particular, deep neural networks.

individuals. The legislation does not account for how data is used in these deep learning algorithms.

- **Staying in control of AI**: They make the important point that artificial intelligence systems are capable of making mistakes, and depending on where the technology is being used, this could have disastrous consequences. They discuss situations where it would and wouldn't be appropriate to use AI algorithms based on this fallibility.
- **Specific governance of ethics in AI:** They discuss the requirement to set up advisory committees and initiate international debate on ethics for digital technology and artificial intelligence, recognising that the importance of considering ethics in these areas has become increasingly significant in recent years.

3. A comparison of the two approaches

3.1. Being part of the algorithms revolution

We will begin with a discussion of how the UK and France plan to evolve technologically to be part of the AI revolution. At this stage, we will not consider the ethical issues. This section aims to provide context for the issues discussed later. A strong desire to develop and promote artificial intelligence technology should go hand in hand with a strong desire to do this in an ethically considered way.

For both committees, there is a recognition that due to the development of machine learning, data lie at the heart of the AI revolution. Data are essential for machine learning algorithms. Based on a vast amount of test data, the algorithm "self-determines" what values certain parameters should take, and subsequently uses these parameters to make decisions based on later input data it is given. Generally speaking, the more test data there are, the more accurate the algorithm will be. It is, of course, not the only factor affecting the accuracy of a machine learning algorithm, but it is something that both reports have decided to focus on as an area where their respective governments can help with the permeating of AI-powered decision-making into society. Governments and large private companies often have large quantities of data at their disposal, meaning that smaller organisations are at a fundamental disadvantage with regards to AI development and usage. The British committee recommends setting up "data trusts" to facilitate the sharing of data between multiple organisations (Lamb, et al., 2018, p.15). This would be done in a way that "ensures that the proper privacy protections and other relevant protections are in place, that there is a governance of the data, which ensures that the voices of interested parties are represented in that governance, and that there is a fair sharing of the value that can be derived from those data (Lamb, et al., 2018, p.16)." Linked to this is a suggestion that the public sector has a "lack of confidence", thinking that "the magic lies within the private sector (Lamb, et al., 2018, p.17)." As an example, in 2015, the Royal Free NHS Foundation Trust signed an agreement with Google DeepMind Health to give them access to 1.6 million personal identifiable records with no monetary gain in return (Lamb, et al., 2018, p.17). They did not understand that the algorithms rely on the data to be effective. The technological community knew that they had more knowledge about the value of data, and therefore took advantage of the NHS, exploiting the knowledge asymmetry. Therefore it is suggested that the data trusts could provide a means of striking

agreements between the public and private sectors so that the data sharing is mutually beneficial. The 2017 Autumn budget announced a £75m investment to take forward recommendations from the 2017 "AI Review", which includes investing in the exploratory work involved in setting up these data trusts. This is also part of the government's plan to more effectively utilise the data that they have at their disposal. What is meant by "exploratory work" is not discussed and a time frame illustrating when these data trusts may be put into use is also not given. Even though the data trusts are mentioned frequently in the document, the lack of specific details regarding their implementation makes it difficult to determine the importance that the British committee is placing on them. If data sharing was considered a high priority for UK's role in the algorithmic revolution, more specific details might be expected.

Aside from data sharing, few other issues are covered with regards to the UK being a major player in the algorithmic revolution. It is noted in the document that digitisation of the NHS has not yet occurred; the aim is for the NHS to have paperless records by 2023. It is difficult to take the governments plans for the use of machine learning within the NHS seriously when that organisation still holds vast amounts of its records on paper. The British report mentions that the lack of NHS digital data is slowing the development of AI algorithms (Lamb, et al., 2018, p.12). We think that the British government should prioritise the digitisation of NHS records over national development of AI in the UK, especially when considering AI-applications to healthcare.

The French document also recognises that data sharing is key to the development of AI (Villani, et al., p.19). They believe it is crucial to accelerate the development and implementation of a policy for making data publicly available. There is already a policy in France for open data: it comes under the umbrella of the Law for a Digital Republic (Villani, et al., p.19). This also provides personal data protection above and beyond that provided by the GDPR. In addition to this current policy, the French committee suggests that "the authorities need to initiate new methods of data production, collaboration and governance through the provision of 'data commons' (Villani, et al., p.19)." This idea seems very similar to the British notion of data trusts. The aim is to provide incentives for data to be shared, pooled and in some cases made public. There is also a recognition that data become more valuable when they are contextualised and cross-referenced, and for society to make best use of this their circulation needs to be promoted; an observation the British report did not make. However, they do fail to add that although contextualised data can be more valuable, they can also be more dangerous. Additionally, sharing data can have positive impacts on safety and security, but it can also have negative impacts which are not mentioned.

In the context of autonomous cars, by making data relating to possible driving scenarios available, manufacturers will be able to achieve a higher level reliability in their self-driving algorithms. The French report mentions that in the UK, the Open Data Institute has encouraged full access to private data to stimulate economic growth (Villani, et al., p.25) although interestingly this is not mentioned the British report. In France, the idea of making data publicly available is not a hypothetical one. In the field of transport and mobility, efforts are already being made to aggregate public and private data (Villani, et al., p.26). The French company Transdev has the aspiration of becoming the international "Wikipedia" of open data. It aims to collect and compile data, clean it up and put it in an open format to allow for innovation of AI within the transport

and mobility industry (Villani, et al., p.26). The report makes it clear that they believe that it is the responsibility of the public authorities to initiate and support the data sharing revolution.

Interestingly, neither report discusses the issue of data protection. Both committees appear to have plans for large-scale data collection/sharing but don't provide suggestions as to how this data should remain protected. Given the large number of massive data breaches from the large industry that have occurred in the last ten years[11], data security is a huge problem faced by the technological industry. Despite this, the issue is not considered at all by either report. Big data is being treated as a universal solution to many of society's problems, and yet its flaws are not being discussed in as much detail as perhaps they should be.

The French committee has plans for France to become a major player in the AI industry. These include promoting France's capabilities and successes to raise their visibility on the global AI stage. They also aim to support future purchasers of AI solutions. The idea is that in advising on the specific requirements of the client, they can steer them away from the international leaders in the field towards smaller-scale stakeholders who may be able to offer a more effective and tailored solution (Villani, et al., p.32). Additionally, the French team propose the establishment of "innovation awards" for small businesses, to help diminish the perceived risk clients have when doing business with a less well-known companies (Villani, et al., p.33). Such a proposal aims to break the stronghold that large tech companies currently have on AI development and deployment in industry.

Here we have given an outline of just a few of the ideas presented by the French committee, who go into much more detail in their report. These plans are not directly relevant to our discussion on the ethics of artificial intelligence so we will not go into any further depth on this topic. However, their detail clearly demonstrates that France has a high level of determination to succeed in the field. They have thought seriously about their strategy. They demonstrate a holistic understanding of the current state of AI globally and, as we will see later on in this paper, their depth of thought extends into the ethical issues surrounding AI where they have carefully considered some deep and important issues.

Our own remark about big data:

In the 19th and 20th centuries, asbestos was widely adopted in buildings across the western world. It was promoted as cheap, abundant, possessing excellent properties with few drawbacks. However, it turns out that asbestos is extremely harmful and has been identified as carcinogenic to humans. Moreover, it is very expensive and time-consuming to remove from the infrastructure of a building. Big data appears to be headed in a similar direction. It is being hailed as a solution to many of society's problems. However, it may prove to cause huge problems in the future, either through direct invasion of personal privacy, or more subtly through collectivised harm as occurred with the work of Cambridge Analytica. It will be nearly impossible to remove it later, given how easy it is to copy, move and store data. There is a possibility that big data could be the asbestos of the 21st century and yet its drawbacks are not being discussed, even by well-intentioned governments.

11 Most notably Sony, Ashley Madison, TalkTalk, Facebook, Google Plus, Equifax, Uber, Yahoo, but also countless others: https://en.wikipedia.org/wiki/List_of_data_breaches

3.2. Using AI

Interestingly, the British report does not give any indication as to areas where it may be inappropriate to use AI. The French report suggests that AI is capable of making mistakes, and so presents a discussion on situations where using AI is appropriate and where it is not. The British committee examined the way that machine learning algorithms are affecting three sectors in particular – in healthcare, criminal justice and social media.

Within healthcare, they recognise that there are many potential benefits, including earlier and faster diagnosis, supporting preventative medicine and guiding complex treatments (Lamb, et al., 2018, p.11). However, there is also a recognition that algorithms need to be used with care. A recent NHS breast cancer screening programme allowed a large number of women to be automatically invited for screening at the appropriate time. However, there was a technical error in the algorithm, resulting in all women aged 68-71 missing out on receiving an invitation for screening. This error is described as a "straightforward coding error" (Lamb, et al., 2018, p.27) and generally dismissed; they believe that making the coding public would have resulted in the error being spotted sooner (Lamb, et al., 2018, p.27).

Historically, it has not always been the case that errors in open-source software are found quickly. For example, Heartbleed was a serious security bug in the OpenSSL cryptography library. It was introduced into the software in 2012 but not realised publicly until 2014[12], and was a software bug of catastrophic proportions. Thus, making the code public does not guarantee that any errors will be detected quickly. Moreover, if no incentives are provided, such as bug bounties, then making the code open-source can potentially make the system *less* safe. It means that the code is available for malicious hackers to analyse, therefore making it easier for them to identify weaknesses. Meanwhile, members of the software development community might not even look at it because they have few incentives to find and fix bugs. Furthermore, the way the British report dismisses the error as "straightforward" doesn't help us to see how the NHS intends to ensure that potential errors in larger, more complicated algorithms do not go on to detrimentally affect the healthcare of vast numbers of people.

Instead of discussing this, the British report decides to focus on the issue of digitising the NHS to aid adoption of AI in healthcare. Whilst digitisation of the NHS should definitely be a priority, doing this in order to rapidly deploy AI in the NHS is somewhat disconcerting. Errors in automatic decision making can have profound effects on a very large number of people, as the breast screening case demonstrates. These can be more severe than error from a human decision, because humans tend to make decisions on a very local scale, whereas automated processes are deployed more widely. Therefore the use of AI in areas as crucial and life-changing as the NHS needs to be carefully considered. Moreover, such considerations should be made by people who are sufficiently well-versed in both the technology and the social impacts that it could have.

From the French perspective, healthcare is one of the key areas that they wish to address with their AI industrial policy. Amongst other areas, they aim to look at early detection of diseases, "personalised, preventative, predictive and participatory healthcare", and the elimination of

12 http://heartbleed.com/

medical deserts (a populated region more than 60 minutes away from the nearest acute-care hospital) (Villani, et al., p.44). Details of how they plan to use AI in these areas are not specified.

Many of the issues relating to using AI in healthcare are discussed in a more general sense in the French report, and therefore we will also save our treatment of them for our more general discussion on ethics in AI.

In the UK criminal justice system, AI algorithms are already being used for facial recognition by some police forces. Issues about reliability and potential racial bias in this technology have already been raised, and therefore "people are not arrested solely on the basis of matches made by facial recognition software (Lamb, et al., 2018, p.13)." Kent Constabulary have been using a commercial algorithm - a predictive policing tool called "PredPol" (an abbreviation of "Predictive Policing") - to help identify areas where offences are likely to take place (Lamb, et al., 2018, p.13). The report neglects to mention the possibility of a feedback loop arising here. Even if the algorithm correctly identifies areas where crime is likely to be committed, more resources consequently supplied there will likely mean that more incidences of crimes are detected. This could lead to some areas being over-policed, which could infringe on peoples right to privacy in these areas. Additionally, Durham Constabulary is already using an algorithm called HART to "assist decision making relating to whether a suspect could be eligible for a deferred prosecution (Lamb, et al., 2018, p.13)." Supposedly, using these algorithms has achieved a "consistency in decision making" (Lamb, et al., 2018, p.13) for targeted interventions for offenders. What is consistent, or whether the consistency achieved is of the positive type, is not mentioned.

The British report recognises that there is a potential for racial bias within the facial recognition algorithms and other algorithms used in the criminal justice system. Moreover, it raises the issue that algorithmic data analysis or decision making often cannot be cross-examined or questioned. To deal with these issues, they suggest "a single [independent] oversight body and regulator for the use of police databases and algorithmic analysis in criminal justice (Lamb, et al., 2018, p.14)." Whilst they believe that it is "imperative" that algorithms used in the criminal justice system are not unfairly discriminatory, they do mention that Durham Constabulary warned against demanding "hypothetical perfection (Lamb, et al., 2018, p.18)." Few details are given in terms of what can be regarded as an "acceptable" level of bias. To reduce the risk of discrimination against poorer areas, Durham Constabulary removed a "postcode" field in their HART algorithm (Lamb, et al., 2018, p.21). However, it should be noted that postcode is not the only proxy for the wealth of an individual. According to Urwin [13], the HART program takes an input labelled "CustodyMosaicCodeTop28" when doing its computation, which is defined as "The 28 most common socio-geo demographic characteristics for County Durham (Urwin, p.98)." It also takes in "FirstPropertyOffenceAge", which is defined as "The suspect's age at first property offence regardless of juvenile or adult (Urwin, p.98)." These two parameters might be heavily correlated to wealth. This should be reasonably obvious for "CustodyMosaicCodeTop28". For first property offence age, it is much more likely that someone from a low-wealth area will have stolen

13 Urwin, S. *"Algorithmic forecasting of offender dangerousness for police custody officers : an assessment of accuracy for the Durham constabulary model",* PhD thesis, Cambridge (2016). Retrieved from: https://www.crim.cam.ac.uk/global/docs/theses/sheena-urwin-thesis-12-12-2016.pdf/view

something as a youth than someone from a high wealth area. Returning to predictive policing generally, the British report also expresses a desire for the use of algorithms in the criminal justice system to be restricted to advisory roles. How to do this in such a way that the judge or police officer using this algorithm places the right amount of faith in the algorithm, for example, not too much so that the algorithm becomes effectively instructive as opposed to advisory, is not discussed. Interestingly, the idea of *not* using algorithmic decision making in areas as difficult as the criminal justice system is not mentioned.

France is currently not using predictive policing algorithms, but according the French report, French and European police services are looking into the possibility of using such algorithms to aid crime prevention (Villani, et al., p.123). They reference the impact such algorithms have had in the U.S. and use this to suggest that caution should be taken when developing and implementing this kind of technology. They suggest that there is a possibility that these algorithms could infringe on fundamental liberties, including the right to a fair trial. Importantly, the French committee recognises that these systems are not perfect and are capable of making errors, the consequences of which could be devastating for individuals who are wrongly assessed. This gives rise to another related issue, that the judges and police officers who use the algorithms need to be able to identify when an error is potentially being made. It is easier for a judge to follow an algorithm's output than to carefully go through a case in detail. However, judges are also not perfect decision-making entities; otherwise the use of algorithms would not need to be considered. The French report contains a reminder that the French Data Protection Act of 1978 declared that "no court or other decision involving legal consequences for an individual can be taken solely on the basis of the automated processing of personal data intended to define the profile of the person concerned or to assess certain aspects of his personality," adding that "an individual has the right to know and to challenge this information and the logic underlying the automated processing when these results are denied him (Villani, et al., p.124)." This shows that issues involved in algorithmic decision making have long been in the minds of the French governing bodies. The case of Loomis v Wisconsin[14] in the US demonstrates that other countries have differing rules and ideology on this. Loomis was a convicted criminal who was given a harsh sentence partially because of the COMPAS risk score he received[15]. It was ruled that Loomis did not have the right to be given an explanation for the COMPAS score he received. The French committee also states that it is essential that responsibility can be attributed to a human being. We will go into more depth about the more general issues of accountability later, but it is worth mentioning the clear line that the French committee team take with regards to accountability in the context of predictive policing algorithms.

With regards to algorithms being used to more effectively target news and advertising, the issue of consent for use of personal data is raised by the British committee. They also mention that distorted filtering in search engines have been built into algorithms. In fact, the British report

14 https://harvardlawreview.org/2017/03/state-v-loomis/

15 COMPAS is an algorithm designed to assign criminals a score based on how likely they are to re-offend, and was implemented in various U.S. States. http://www.equivant.com/solutions/inmate-classification

mentions that Google was fined by the European Commission (Lamb, et al., 2018, p.14) for manipulating its algorithms to demote rival shopping services in its search results and therefore allowing its own shopping service to feature more prominently. Yet, only three pages later in the same report the British mention that it was a Google-owned company, DeepMind Health, that the Royal Free NHS Foundation Trust gave data to, which the British report cites as an example of the private sector making good use of big data to develop AI (Lamb, et al., 2018, p.17). This is a lesson demonstrating the need for regulation and legislation of AI (and, more generally, algorithmic processes in society); the people developing algorithms cannot necessarily be trusted, as they have their own (corporate, institutional) interests in mind, which may not align with the wider interests of society. The British report also recognises that the major social media platforms have been able to secure strong market positions by using algorithms in their services with the aid of the large amount of data they have available to them, with the suggestion that this leaves smaller businesses at a disadvantage. Aside from recommending that data be shared to support smaller businesses there is very little discussion about whether private entities should have the power to influence the news and advertising that the public are able to see. The French report only discusses targeted advertising briefly in the context of biased algorithms.

Our own remark on the idea of antitrust laws for data companies:
There is the issue that for many companies, data lie at the heart of the business model. Therefore, it is likely that some large companies would be reluctant to offer data to support smaller businesses who potentially make up some of their competition. The concept of antitrust laws for data companies is not mentioned by either document; that is, using law to restrict the amount of data one organisation can hold. This is one option that achieves a similar result but enforces it by law instead of relying on the willingness of companies to cooperate.

3.3. Research

Both documents suggest investing in AI research, but the research areas within AI that they intend to focus on are a little different. The UK Government has recently announced that it intends to spend £11 million on research projects "to better understand the ethical and security implications of data sharing and privacy breaches (Lamb, et al., 2018, p.32)." Furthermore, the British committee state that they encourage funding for research into understanding how to get the best out of AI whilst mitigating potential risks. Generally speaking, it appears as though the British wish to focus their research efforts on understanding how to utilise the AI tools that are currently available, by understanding how one might deploy them in a transparent, accountable and fair manner.

In contrast to this, the French committee suggests much bigger plans for French research into AI. They recognise that although France plays a large role in AI research, they see the following issue: increasingly, researchers are becoming more inclined to work for private companies rather than in public research (Villani, et al., p.6). Therefore, they wish to make academic research institutes more appealing places of work, which could involve salary increases and reduction in administrative formalities. Additionally, they recommend that more resources are supplied to allow public research to achieve its goals. Furthermore, they realise that AI requires a lot of computing power and therefore recommend that a customised, dedicated supercomputer is set up for AI research. They also wish to set up a network of coordinating Interdisciplinary Institutes for Artificial Intelligence within public higher education systems nationwide because they

recognise that applications of AI are not just confined to the technology sector. They recognise that public research is not the only way forward and suggest partnering with some private partners and support them in making their own technological breakthroughs (Villani, et al., p.63). It is clear from reading the document that France intends not only to research the ethical consequences of AI and how to implement it in society, but they also wish to become one of the countries leading the technological revolution.

4. A comparison of ethical issues

It is interesting to see how each of the two reports envisage how their respective countries should approach the ensuing technological change from a practical perspective. However, it is crucially important to examine how they are approaching the ethical issues that go alongside the implementation of AI and algorithmic decision-making into society. We will see here that whilst there are some similar themes, it is clear that the British committee, and the French committee, have different perspectives with regards to these ethical issues.

4.1. Data protection

Both committees do realise that data protection is becoming an increasingly important issue. The British report continuously refers to the new GDPR as a solution to problems which have previously occurred in this area, such as in the case of Cambridge Analytica. They say that the GDPR will "provide helpful protections for those affected by algorithms and those whose data are subsumed in algorithm development (Lamb, et al., 2018, p.4)." There is recognition that there are some grey areas which could result in individuals being left unprotected, but details of this are not mentioned. Notably, there is no mention of the wider-reaching issues that could arise from the inferred patterns of human behaviour extracted from aggregated data, and the subsequent damage that this could cause to society. The only comment they make regarding these "grey areas" is to recommend the Centre for Data Ethics & Innovation keep the operation of the GDPR under review in relation to its impact on algorithmic decision-making, with the view that this could lead to amendments in the UK's data protection legislation if required (Lamb, et al., 2018, p.38). It is important to note here that development of the GDPR began in 1995, yet it was only finalised and implemented in 2018.[16] The GDPR was not designed to solve today's problems, hence it only focusses on personal data. How the GDPR is discussed in the Science and Technology Committee report suggests a naïvety about what has changed in the intervening 23 years, and about other problems that need to be addressed with regards to data security.

The French committee is one step ahead of the British; they acknowledge that their own French Data Protection Act and the GDPR deal solely with personal data. With the developments in AI and the way it uses data potentially leads to a "blind spot" in the law because it does not only use personal data. This is a key observation. Aggregated data can be used to infer meaning and patterns, and this can be used to harm all of society, including those whose personal data was not part of the training data set. The choice of data training sets for AI development can have a large impact on groups of people as well as individuals. They recognise that the only way that

16 https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en

individuals can decide to not have their data collected by a specific company is to simply not use the services provided by them, although this is not always a viable option. This power currently remains in the hands of regulators and legislators. The French committee also suggests to make class action in data-protection cases more effective; it is currently extremely limited because, although injured users can be recognised as victims, they cannot receive any compensation.

Our own remark on the secondary effects of data breeches:

Neither report goes into detail about the issue that data privacy is not just about protecting the personal data of the person uploading/providing it. It is completely possible, and indeed has already transpired, that individuals who have nothing to do with a particular service provider/platform can still suffer consequences as a result of others who are willing participants. This has been happening for years with supermarket loyalty cards: A small proportion of shoppers sign up for a card and from that limited data set the shop can determine which products to stock, which items to put on offer and so on, in a way that affects all customers. Most seriously, this happened with Cambridge Analytica. They first created a Facebook app called "This Is Your Digital Life" for which 270,000 users subscribed. Due to a peculiarity of the options available to app developers, the app was able to harvest not only the personal information of the 270,000 users who subscribed[17], but also the (approximately) 87 million "friends" of all these users[18], and thus we already see an extended violation of privacy. Further to this, Cambridge Analytica then applied machine learning algorithms to this data set of 87 million people and created a predictive tool that could profile people and deduce things about their voting preferences, life ambitions and preference, etc., all through their Facebook profile. Cambridge Analytica then created many different targeted political adverts, and directly advertised to all Facebook users in the UK/US during the Brexit/Trump campaigns. This advertising was low-cost, accurate, and manipulative. Thus, from the lax data concerns of 270,000 people, the entire Facebook user base in the US and UK[19] were affected; a scale factor of nearly 1000. Such things also happen in other instances, like with predictive policing/sentencing, and happens in just about every occurrence of machine learning algorithms. Users who do manage to keep themselves away simply end up being affected because of the actions of others around them, and there is little they can do about it as individuals. As humans often behave in quite predictable and regular ways, information about how a small subset of society behaves can be used to predict, and manipulate, how most of society behaves.

4.2. Algorithmic bias

The British report highlights ways in which an algorithm might become biased; through the use of inappropriate training data, a lack of data, correlation disguised as causation or via an algorithm development team which might not be fully representative of society. They recognise that bias will exist in some form because of the way the algorithms work, giving more weight to some information and less to other information. However, the document states that "some forms of bias can nevertheless extend beyond what is acceptable (Lamb, et al., 2018, p.18)." There is

17 *https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/*

18 *https://newsroom.fb.com/news/2018/04/restricting-data-access/*

19 *Nearly 250 million users: https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/*

also a recognition that current legislation is not equipped to protect individuals from discrimination caused by automated decision-making, and yet they make no recommendation with regards to what needs to change to help such individuals. This theme continues, many problems are identified but few solutions are offered. The only general suggestion alluded to is that there should exist a system of certification of algorithms and that ethics boards should oversee algorithmic decisions. Specific solutions are not offered to deal with individual issues concerning algorithms.

In terms of inappropriate or insufficient training data, they suggest that this is probably the most likely place for bias to come from in an algorithm, and they provide many examples of where this has proved to have negative impacts. However, a solution or a plan to reduce this source of bias is not offered. They suggest that there needs to be a push to achieve greater diversity in algorithmic development teams. However, they provide no mechanism to achieve this, nor do they discuss the specific ways in which bias could be introduced into an algorithm via the team developing it. More detail here would have been appreciated because the lack of specification about how the bias is introduced suggests a naïve approach to the issue.

In producing an algorithm via machine learning, the development team give the algorithm incentives or goals to allow it to "learn" from the training data provided to it. Internal bias or misunderstanding can manifest itself when setting these incentives. Additionally, diversity in the team does not solve the whole problem. If a diverse team is developing an algorithm for a situation or group of people that they have little understanding about, similar issues can arise. Despite identifying that the algorithmic development teams can be a source of bias, there isn't any mention in the British report of providing them with training to help them identify how their actions may introduce bias into the algorithm. Unfortunately, such knowledge does not normally form part of their formal education. It is suggested that it is the job of the newly formed Centre for Data Ethics & Innovation to consider the issues proposed by the report.

The French report gives many examples in which we have already witnessed the negative effects of bias in existing AI algorithms, like Google's targeted advertising offering fewer well-paying jobs to women, and predictive policing algorithms suggesting higher levels of surveillance in poorer, predominantly African-American, areas. They recognise that examples such as these raise fears about AI amongst the general public (Villani, et al., p.116). This could lead to a slowdown in its development meaning that society may not reap all of the potential benefits AI has to offer. They suggest that issues with bias should be fixed as a matter of priority. They recognise that through the use of biased data, there is the potential for existing social inequalities to become embedded in deep learning algorithms. Under current European legislation, a Private Impact Assessment (PIA) is required to be carried out in the case when algorithmic data processing reveals a risk of discrimination. A PIA requires those who process personal data to find out the potential impact of their activities on those concerned. The French committee recommends that with regards to AI, the PIA is paired with a similar measure more relevant to AI to force those creating AI to consider the social consequences of the technology they produce.

Although not necessarily in the context of bias, the French committee makes it clear that the researchers and engineers working on the algorithms need to have ethical training incorporated into their education. This is becoming crucial because the developers are the people with the best

understanding of how the algorithms are put together, and therefore are the intermediaries between humanity and the technology. They hope this will better prepare the technicians to deal with any ethical issues which might arise in the development of the technology and teach them to deal with potential moral issues responsibly. They hope that by training researchers and engineers to be more responsible, it could lead to the development of more responsible technology. One suggestion for educating specialists is to introduce a major/minor system into higher education: for example, an individual could major in computer science and minor in law. Here the French committee has made the important observation that it is not sufficient to teach lawyers a little bit of AI or to teach computer scientists a little bit of law. It is becoming the case that these disciplines cannot be kept separate because there needs to be communication across the fields to ensure that policies and law are in line with the current capabilities of technology.

4.3. Transparency and accountability

In the British report, there is an understanding that transparency of an algorithm would make the issue of accountability a simpler issue to deal with. However, transparency is not necessarily easy to achieve especially in the case of machine learning algorithms. The document questions whether full transparency in these algorithms is even possible. It is well-understood by computer scientists that algorithmic transparency is a necessary, but far from sufficient, mechanism for understanding how an algorithm works. Moreover, in the case of machine learning algorithms, it is almost impossible for someone to infer anything meaningful from the code. The British committee reiterate the issue of data protection here; if an algorithm which relies on vast amounts of personal data is made fully transparent, then the document suggests that some of this private data could be discoverable. This is potentially one of the few places where the British committee is actually being over-cautious. It would be quite difficult to reverse-extract personal data from an algorithm, especially a machine learning algorithm where the data manifests itself in the form of a large matrix of machine-learnt parameters. Globally, we are still in the early stages of development of methods to achieve transparency in certain machine learning algorithms, but it is important to consider potential issues like this during the development phase even if they should be lower down on the list of concerns.

The British report also raises questions about the requirement to publish the code for algorithms. However, the committee recognises that this raises concerns regarding trade secrets, copyright and that knowing the code for an algorithm makes it easier for people to undermine its effectiveness. The issue of transparency in an algorithm is closely tied with the concept of "right to explanation" and the ability to contest a decision made by an algorithm. The British report mentions that in France, the digital-economy minister recently said that the French government should not use any algorithm which cannot be explained. The British report does not make a recommendation of this kind, and also notes that the UK Government has not gone beyond the GDPR in this respect, meaning that individuals cannot currently challenge a result of an algorithmic decision. As the French committee points out, if the public cannot scrutinise and test the processes by which they are being judged, or at least see that it is possible to scrutinise them, then they will lose faith in them and cease to trust them. If these are processes which are integral to the functioning of society, then this starts to break down the social structures we have set out to protect and improve. However, they do state that "transparency must be a key underpinning of algorithm accountability (Lamb, et al., 2018, p.30)." A decision is not reached on what form this

transparency should take, whether it would involve a full understanding of how the algorithm works or whether an explanation of the decision reached would suffice, or the tractability of achieving transparency by these means.

Regarding accountability in a context aside from transparency, the British committee recognises the difficulty in assigning accountability. The people designing and making the algorithms have a certain amount of responsibility with regards to how the algorithms are designed but cannot be held fully accountable for how they are used. If the chain of accountability is not properly established then it will not be clear whom to blame when things go wrong. Not only is this unacceptable to society, but it also means that developers can do whatever they like without restriction or repercussion, as there are no disincentives or consequences for their actions. It is suggested in the British report that standards are required in order to begin designing a system for accountability. The principles suggested mainly focus on the proper implementation of the GDPR and ensuring that data rights are maintained to protect the privacy and security of individuals. There is an excessive focus on the individual, without a proper understanding of the mass harm that can come from such systems beyond the individuals whose data is fed into them. There is also the suggestion that there should be a system for auditing algorithms with the idea of reducing the number of inherently biased algorithms being put into use. Sociologist Professor Daniel Neyland suggested the use of certificates and third party "seals" for audited algorithms could help address "the contemporary limitations of accountability and transparency in algorithmic systems (Lamb, et al., 2018, p.26)." Although this is a nice idea, the practicalities of such "seals" are not discussed. The "black box" nature of many machine learning algorithms makes auditing and certification a difficult, if not impossible, task. The committee recommends that "the Government should immediately task the Centre of Data Ethics & Innovation to evaluate [auditing and certification] tools and advise on which to prioritise and on how they should be embedded in the private sector as well as in government bodies (Lamb, et al., 2018, p.27)." They provide this recommendation without any discussion as to whether the task they have assigned is even technically possible.

Opening the "black box" is the first ethical issue considered by the French committee. They recognise that this is an extremely difficult task for many branches of AI and that it "constitutes a real scientific challenge (Villani, et al., p.115)." This comment highlights the difference in understanding of the French and British committees. France is taking the hard line with regards to a lack of transparency: "we cannot allow certain important decisions to be taken without explanation," (Villani, et al., p.115) listing employment, credit, accommodation, justice and health as areas where using unexplainable algorithms could not be justified. They give serious treatment to the issue of transparency in light of their claim that an algorithm should not be used if it cannot be explained, and transparency is one of the key steps in being able to explain what an algorithm does. Similar to the British, the French report also considers the concept of auditing AI, which they believe should not just be restricted to government agencies. The team readily state that the auditing protocols are still in their infancy and businesses who develop algorithmic tools do not want to give away their intellectual property to third parties, therefore making the issue of auditing a difficult one. They suggest auditors may be satisfied to attempt to determine the fairness of an algorithm by feeding it with false input data and examining the output, stating it is not always useful or necessary to analyse the source code for an algorithm. Though feeding in test

cases is a good start, it is definitely far from sufficient, because knowing how an algorithm behaves on test data does not always effectively demonstrate how the algorithm will behave given "real world" data. Moreover, there are already cases of industry-wide cheating of testing scenarios, whereby knowing the details of the test enabled people to cheat the test (most notably: the Volkswagen emissions scandal[20] from 2008-2015). The French report suggests that one of the main problems with regards to public auditing is getting access to the training data used to make the AI algorithms, which leads back to the data sharing mechanisms described earlier. However, despite these recommendations they note that research urgently needs to be carried out into understanding AI. They identify three specific areas that should be looked into: how to produce models that are more easily understood, how to produce user interfaces which are more intelligent[21], and understanding the "black box" mechanisms with the aim of producing satisfactory explanations of decisions made.

4.4. Impacts on jobs and employment

The British report does not raise the issue of the potential impact on jobs and employment that new AI technology could have. One of the reasons for this could be that the British report only focusses on algorithmic decision-making which has a narrower remit than AI as a whole, but it is still an issue relevant to this area. It seems odd to neglect the impact on jobs and employment, when implementing algorithms for decision making potentially means job losses for people who were previously employed to make such decisions. We will discuss the potential impact on jobs and employment here because the French report has dedicated approximately 20 pages (Villani, et al., p.80-p.99) of their report to this, and it is certainly an important issue to consider, given the technology-driven world we now live in. They recognise that historically when transitions of this type have occurred, it has often disproportionately affected vulnerable people within society. Therefore they wish to make a conscientious effort to avoid similar problems occurring with the rise of AI. The French report references several studies which warn of a wide decimation of jobs, including one (Frey et al.[22]) which predicted that 47% of US jobs were at risk of disappearing over the next 20 years. Given current technological capabilities and the types of jobs on the market, it is likely to be the lower-skilled occupations that will be automated first. Additionally, the types of jobs that the introduction of AI could potentially create are likely to be mainly highly skilled jobs (those involved in the development of AI), along with some more basic jobs in domestic services. This is likely to lead to a greater polarisation in the job market than currently exists. The jobs that AI could do are still in question as the technology is still young, but the same issues apply as when we consider the impact that automation might have on society.

The French report breaks down the problem by defining some more specific objectives. These include: avoiding the potential excessive polarisation of the job market and inequality associated

20 *"The Volkswagen emissions scandal explained",* https://www.theguardian.com/business/ng-interactive/2015/sep/23/volkswagen-emissions-scandal-explained-diesel-cars

21 They do not specify what *intelligent* means.

22 Frey, C., Osborn, M., *The future of employment: How susceptible are jobs to computerisation?,* Technological Forecasting and Social Change **114**, 254-280 (2017). https://www.sciencedirect.com/science/article/pii/S0040162516302244

with that, and finding a way to work with machines, otherwise known as developing complementarity between humans and machines. What this latter point entails is in itself a difficult issue as it might include following orders from AI, losing control of processes, and delegating decisions to AI, which are not necessarily conducive to a pleasant working environment. Therefore, ways to work with AI that are desirable to the human workforce need to be promoted. This will involve gaining a better understanding of the major macroeconomic effects of AI technologies in order to effectively judge the best course of action.

As a starting point, education is identified as a key area that needs to be adapted to better train individuals for the new types of work they may be expected to do. As part of this, training AI specialists is a top priority. This includes developing interfaces between courses in AI and in other subjects. Public employment and vocational training policy schemes should also be adapted as they are currently "not sufficiently factoring in the need to urgently and specifically target certain jobs and individual profiles (Villani, et al., p.86)."

Within the 20 pages on this topic, the French committee comes up with a five-point plan to help society deal with the impact that AI could have on jobs and employment. We will very briefly give an overview of these five points; needless to say, the French report goes into much greater detail than what we give here. Firstly, the need to anticipate the impacts on employment and test out initial transition schemes. This involves setting up public think tanks and performing trials in specific local areas on new models to deal with automation. Secondly, the idea of complementarity between humans and machines needs to be developed within organisations and working conditions need to be regulated. This may involve legislative reform of working conditions. Thirdly, training overhaul needs to occur. This involves adapted methods of training as well as training content. They suggest encouraging creativity and innovative teaching practices to develop new experimental teaching methods. Fourthly, new methods for funding vocational training need to be explored and tested. Finally, training AI talent at all levels must occur to meet the needs of the workplace, which involves creating new courses and training programmes. As already mentioned, this may involve the introduction of more hybrid courses like AI with Law.

France clearly has big plans for AI. They are taking its potential impact seriously, even recommending a complete overhaul of their entire education system. This level of thought and dedication to helping society into the upcoming era of AI is vital if we are not going to be thrown by the wave of technological change coming our way over the next few years.

4.5. Ecological considerations

Once again, there is an area relevant to ethics in AI which the British report does not allude to. Again, this could be because the remit of the British report is narrower and therefore they do not consider this an issue that they should consider. However, the ecological impacts of AI require further investigation, and the French report goes into great detail about this. This is also an issue which is not often talked about in the media regarding AI and therefore we will discuss it briefly here.

The French report cites the American Association of Semi-Conductor Manufacturers as saying that the global demand for data storage capacity (which is only growing with the development of AI) will exceed the world's available production of silicon by 2040. Furthermore, the energy

required for all such computation will exceed the world's available energy production. We can clearly see that energy issues need to be discussed to allow the development of AI to continue, and to prevent the world being plagued by energy and material shortages. The French committee suggests that although AI development may lead to many problems ecologically, it can also offer a lot of solutions. It may allow better management of resources, and lead to the development of new ways to protect the environment. Gains in potential energy saving need to offset the potentially negative impacts AI could have in this area. They state "a truly ambitious vision for AI should, therefore, go beyond mere rhetoric concerning efficient use of resources; it needs to incorporate a paradigm shift toward a more energy-efficient collective growth which requires an understanding of the dynamics of the ecosystems for which this will be a key tool (Villani, et al., p.102)." The document goes on to give a plan for France to champion sustainable and ecological artificial intelligence. In brief terms this involves making the issue part of the international agenda, designing more energy efficient AI and releasing ecological data to allow AI to be part of the solution. France sets a good example here with regards to discussing the wider impacts that the development of AI can have. There are global issues to consider which should not be neglected during this technological boom. It is imperative that all nations consider the total impact of AI on society.

We note that further societal issues that could be (but are not) considered are the cultural effects (AI might start to have creative output, in the form of novels or music. Indeed, it has already created an artwork that was sold for USD$432,500[23]), genetic effects (AI assisting with online dating) and religious effects (people potentially following "commandments" from AI and "believing in" AI, without understanding any deeper rationale behind them), to name just a few.

4.6. Governance of ethics in AI

Regarding governance of AI, the UK has recently established the Centre for Data Ethics & Innovation which is an oversight and ethics body which should monitor and address issues within Big Data and Artificial Intelligence *(Lamb, et al., 2018, p.3)*. The Government wishes to place the organisation on a statutory footing; this would mean that it would be required to report annually to Parliament on the results of its work. The British report lays out many recommendations for themes and challenges that the new Centre for Data Ethics & Innovation should address as it begins its work. Importantly, they have been tasked with "overseeing the future development of algorithms and the 'decisions' they make *(Lamb, et al., 2018, p.9)*." Recommendations include examining the bias built into algorithms, evaluating accountability tools, charging ethics boards with oversight of algorithmic decisions and potentially continuing research into algorithms generally. More specific details about these recommendations are not provided. Given the importance of these issues and the rather generous time provisions the committee had to prepare their findings, it is surprising that the committee gives no further detail beyond simply pointing out where some problems lie.

23 *"Is artificial intelligence set to become art's next medium?",* https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

As it stands, France do not currently have a specific national advisory committee for the ethics of digital technology and artificial intelligence, but the French committee deems this a necessary step because of the significant role that ethics plays in the AI debate *(Villani, et al., p.128)*. They say that this committee would be responsible for coordinating public debate in an accessible and constructed way within a legal framework. They believe that the committee could provide clarity on technological choices made by researchers and those in industry through independently developed opinions. The French committee stresses the importance of debate within society, not just in France but on an international scale with the potential for a network of national ethical committees.

Within both France and the UK, the concept of ethical committees for AI is in its early days. It will be interesting to see the work of these committees in the coming years and how effective they are at promoting ethical discussion and potentially dealing with some of the ethical issues that arise in AI.

5. Conclusion

Even though the two reports aim to serve a similar purpose for their respective governments, they contain some very different ideas and illustrate very different goals with regards to the advancement of AI. The French committee appears to have much more ambition for development (potentially from the broader scope of their remit) than the British committee. Moreover, they demonstrate a much deeper thought process with regards to the potential issues that may arise, especially with regards to the wider national and global effects like impacts on jobs and employment and ecological considerations, which the British report does not give any thought to. Comparing the two reports shows a potential naïvety in the British approach. Their ideas and considerations lack depth and imagination. The issues they cover are not new; they discuss issues with data privacy and bias and little else. The GDPR was introduced in Europe to address issues with data privacy because it is a long-known issue. The potential presence of bias is one of the most prolific criticisms circulated by the media with regards to AI. Moreover, the British committee gives very few recommendations for additional changes that could be implemented to help further address these issues. Generally speaking, the only recommendations they make are in the form of suggesting problems for the newly formed Centre for Data Ethics and Innovation to think about, and the British committee seems to think of this new organisation as the solution to most of their AI-related problems. Additionally, the British don't entertain the idea of simply *not* using AI in certain circumstances as perhaps the only way to deal with the potential negative impact it could have on society. They seem to suggest AI is inevitable, which is not the case if the big economic powers choose not to use it. As we have discussed in this report, the French report addresses a much wider variety of issues and gives many recommendations for the French Government to not only help the country through the transition into the new technological age, but also recommendations to help France be one of the countries leading the revolution. The British committee could learn a lot from the French committee.

6. Main references

Lamb, N., *et al., Algorithms in Decision-Making,* Science and Technology Committee, (2018). Ordered by the UK House of Commons. Retrieved from https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

Villani, C., et al., (2018). *For a Meaningful Artificial Intelligence,* (2018). Retrieved from https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Note: all other references are included as footnotes throughout the text.

7. Further Comments

It is important that we take this opportunity to go beyond just a comparison of these two reports. As trained mathematicians, we are able to provide further observations, analysis, and input into the discussion at hand. Therefore we have provided some additional comments relating to, and often moving beyond, the content of the two reports.

7.1. Researchers, data scientists and mathematicians.

It seems that these reports use the terms *researchers* and/or *data scientists* without fully defining them. Indeed, "researcher" invokes an image of someone in a lab coat, and "data" leads people to assume some sort of computer science. The lay reader may fail to make the connection that these terms often refer to trained mathematicians. The term "data scientist" hardly existed 10 years ago, and yet now it is mathematicians who these data science firms are targeting in their recruitment and hiring processes, as mathematicians are the ones who have the necessary skills and understanding to comprehend the complexity of data and how to extract value from them. Mathematicians are trained to see patterns, and processing big data is simply pattern-matching. It is therefore important that mathematicians receive the necessary ethical training required for such important work, as a large proportion of them end up working directly on AI projects with big data.

7.2. AI we cannot switch off

One widely-held misconception about AI is that we can always turn it off. However, that is no longer the case. We already have the ability to implement machine-learned algorithms in a way that cannot be stopped: it is already that case that "unstoppable transactions" are being deployed on the Etherium blockchain in the form of DAOs[24]. We cannot turn these off, as the blockchain is a completely decentralised system and no one entity possesses the computational power to take it over, for the physical means to shut it down. Thus we need to foresee negative outcomes before they hit us. Having Etherium (a store of wealth) do things in an unstoppable automated way, guided by AI, means that one of our stores of wealth is now operating outside of our absolute control. This means it could take wealth from entities we do not want it to take from (such as a government), or give wealth to entities we do not want it to give to (such as criminal organisations); in both instances, we have no way of undoing these. There is no "kill switch" here; turning it off requires us to destroy a large percentage of the (decentralised) Etherium miners, which no single institution, organisation, or government currently has the capacity to do.

7.3. The UK select committee interview of Cambridge Analytica.

It is fascinating to note that when the British government interviewed various key players in the Cambridge Analytica (CA) scandal, they did so through the Digital, Culture, Media and Sport Select Committee[25]. Though this Select Committee did an excellent job of investigating the Cambridge Analytica scandal, it demonstrated that, somehow, the UK government misunderstood where many of the underlying problems of CA lay. Cambridge Analytica was made up mostly of data

24 *"What is a DAO?",* https://www.coindesk.com/information/what-is-a-dao-ethereum/

25 *"Disinformation and 'fake news': Interim Report Contents",* https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/36314.htm

scientists (about 100 of them), comprised of mathematicians, computer scientists, physicists, etc. These people have nothing to do with culture, media, or sport. They were trained in technical sciences, and made up most of the organisation. Their work was entirely mathematical/scientific. Thus, it would have been much more appropriate for the interviews to be done through the Science and Technology Select Committee; the authors of the British report. Though they make reference to the CA scandal several times (Lamb, et al., 2018, p.5,6,10,12,16,35,37,38,39,41,46), they never acknowledge that it was completely within their remit to investigate the scandal in detail and advise on it. Such is the stealth of the work of mathematicians, that even those charged with investigating, understanding and scrutinising them cannot see when and where they are operating.

7.4. Predictive policing algorithms

The issue of predictive policing algorithms is a complicated one, and referred to by both reports. Both reports comment on the issue of fairness in these algorithms, mentioning that certain areas may end up being over-policed and thus making up a disproportionate part of arrests. However, even if such algorithms are correctly identifying the places where crime is most likely to be committed, it is highly likely that a feedback loop may form. We mentioned this issue in the comparison of the reports, because neither of them addressed it. We will now take this opportunity to discuss it in a little more detail.

Identifying places with a low crime rate does not mean there is no crime occurring there, and vice-versa: a high crime rate does not mean crime will definitely occur there[26]. However, any resource-limited policing unit (many of which are) will, acting rationally, dispatch the greatest number of police officers to the places where the algorithm says crimes have the highest chance of being committed. Such an algorithm working "perfectly" and being used (rationally) like this will mean that more crime is registered in those high crime areas. Eventually, the crime machine learning algorithm (which presumably will be frequently updated with new data) will "learn" that it is an area of even higher crime, and divert even more resources there. And so a feedback loop is created: more arrests leads to more policing leads to more arrests[27]. The same principles apply to credit reporting done with AI, university admissions done with AI, and so on; feedback loops can quickly be created. It is surprising that neither report deals with feedback loops in much detail, as these form a serious part of the social threat of AI.

7.5. A search engine is a distortion machine

It seems like the British report takes the position that there is a "correct" algorithm to implement in any give case, scenario, or for any given problem. In cases of mathematical calculations, for example in calculating interest payments, this is true. However, in many cases, it is not. The best example is that of an internet search engine, like Google. The British report mentions that Google was fined for distorting its algorithms to preference its own shopping service over those of its competitors. However, neither report seems to acknowledge that a search engine is simply a

26 We have simplified things slightly here, as such predictive policing tools consider not just location, but also time of day, day of the week, etc. But our discussion remains valid.

27 Ensign, D., et al., *Runaway Feedback Loops in Predictive Policing*, Proceedings of Machine Learning Research **81**, 1–12, (2018). http://proceedings.mlr.press/v81/ensign18a/ensign18a.pdf

distortion machine, by design. It looks through trillions of web pages, picks out all the ones that it "thinks" are relevant, and then orders them in some "optimal" way. This means that out of potentially billions of search results, the user is presented with ten at a time on a page. Assuming we accept that a "search result" is a webpage containing all the search words, is there any "correct" way to order these billion pages? Of course not! Google *chooses* a way to order them, which just so happens to be convenient for many people[28]. Every search engine does this; it takes these results, and imposes some artificial, human-built (possibly through AI) ordering on them, and this ordering directly effects what information you do or don't receive. Therefore, a search engine distorts what you see; it is a distortion machine, and there is no way around that fact. If the EU can distinguish between good distortion and bad distortion, to the extent that it issued fines to Google for manipulating its search results to benefit its own shopping services of those of its competitors, then we need to come up with some clear guidelines on what levels of distortion are acceptable to society. The problem is that such algorithms can be extremely opaque, or worse still, operating as proprietary software, and the way that humans interact with these algorithms is not well understood. What is the effect on society if the algorithm, for some purely technical reasons, always ranks pages of a certain political party higher than others? The question of how to deal with this issue, which is present by design, is not addressed in these reports. However, it is definitely something that should be part of such discussions.

7.6. Portability of data

Both reports mention the importance of data sharing, since small AI companies that don't have access to large data sets are simply not able to compete with large companies that do. This means that these large companies keep getting bigger, with few challengers. In addition to this substantial competitive advantage, consumers are often wedded to existing platforms like Facebook because it can be extremely difficult to migrate to a new platform without starting afresh, unless the consumer is willing to do a "manual" transfer of their data (an extremely laborious task for those who have an extensive profile built up over several years) as well as forgo the the network of contacts they have already established. Therefore, laws which make the porting of data easier, and enforce interoperability between platforms, would reduce friction and improve competition. There are two relevant examples from the telecommunications industry of this.

The first of these is set up of phone networks. As has been the case for a long time, people can call each other even though they have different phone network providers. But this was not always the case, as in (Mueller[29], p.389) we see that around 1904-14, more than 55% of the US population lived in cities or towns where there were two unconnected telephone exchanges.

The second of these is the porting of phone numbers. Nowadays, consumers in the UK and EU can quite simply and quickly port their phone number from one provider to another, but this is only

28 Google's original PageRank algorithm gave non-unique solutions; how did Google choose which one to use?

29 Mueller, M., *Universal service in telephone history: A reconstruction*, Telecommunications Policy **17**, No. 5, 352-369 (1993). https://www.sciencedirect.com/science/article/pii/030859619390050D

thanks to various legislation that was passed to enforce this[30]. Previously, if a consumer wanted to change provider, they needed to change their number, which was a huge hassle and a disincentive.

It seems almost ridiculous now to think of a world where people on different networks cannot call each other, or a world where changing phone provider means changing number. But that *was* the case, historically, and it is *exactly* what is happening with Facebook and other social media platforms. A general method to ensure the portability of data would need to take the form of both legislation and standardisation. This has clearly worked in the past, even in the context of internet communication, with email serving as yet another good example[31]. Thus this could be explored for other communication methods and platforms. At the moment, portability of data is an inhibiting factor to the data sharing solution proposed by both reports.

7.7. The privatisation of censorship

The privatisation of censorship is something that neither committee addresses. Though not strictly within their stated remits, it is a highly relevant issue as AI is being used by various tech companies to decide what material should be censored on their platforms. When public discourse takes place on private ground such as platforms and search engines, the public and the state lose all control over what can or cannot be said. Recently several platforms unilaterally decided to remove radio host Alex Jones from their systems for purported hate speech[32]. Moreover, some of these platforms have automated systems to detect fake news and hate speech[33]. Though such processes have probably been designed with the best of intentions, has it now become the duty, or indeed the right, of large platforms to be deciding what should or shouldn't be censored, especially if it is carried out in an opaque way? Undoubtedly, some clever AI has been deployed to tackle these colossal tasks. Society cannot know that these processes are not engaging in an overreach of censorship; silencing a valid voice can be just as bad as, if not worse than, letting through an unacceptable one. These firms have comparable population saturation to traditional media outlets, but they act on a higher level than those media, holding near-complete monopolies over what they do with virtually no competition. When Facebook, Twitter, or Google decide to implement such changes,[34] they are engaging in censorship in an unaccountable way. AI gives

30 https://www.ofcom.org.uk/phones-telecoms-and-internet/information-for-industry/numbering/number-portability-info

31 In this instance, mutually-agreed standardisation was all that was necessary for email, but this was because it was part of the process from the onset.

32 *"Alex Jones and Infowars Content Is Removed From Apple, Facebook and YouTube",* https://www.nytimes.com/2018/08/06/technology/infowars-alex-jones-apple-facebook-spotify.html

33 *"F8 2018: Using Technology to Remove the Bad Stuff Before It's Even Reported",* https://newsroom.fb.com/news/2018/05/removing-content-using-ai/

34 Both Google and Facebook have recently made changes to "de-rank" certain searches, and certain types of news:
"*Google to 'de-rank' Russia Today and Sputnik",* https://www.bbc.co.uk/news/technology-42065644
"*Facebook news feed changes could affect your business's post engagement",*

them the ability to carry out a task that would be impossible to do by hand, yet there is no associated responsibility assigned to them to get it right, for whatever "right" may mean here.

7.8. Taking copyright law into their own hands

On a similar note, various platforms, such as YouTube, have now implemented algorithmic take-down mechanisms to prevent occurrences of things like copyright infringement[35], which is sometimes demonstrably unjust, as in the case of five copyright infringement claims against a publication of white noise[36]. This is a clear example of YouTube determining, and then policing, copyright law. Given how big YouTube actually is, their algorithm now defines copyright infringement to some degree[37]. They are the judge, jury and executioner, all wrapped in one, and the public has no recourse here. YouTube decides, via AI, what to allow or remove, and the issue is settled. An actual case of copyright infringement would need to be brought before the courts, have evidence presented and scrutinised, be judged by peers, and be open to public scrutiny; the YouTube process contains none of that. Moreover, there is no way for members of the public to prove that YouTube is wrong and force them to reinstate the removed content. We cannot see the algorithm they have implemented, and even if we could see it, we would probably be unable to scrutinise it with any certainty, especially if it was developed through machine learning. Also, YouTube is a private platform, and we have no inherent right to put up content there. The problem with this last point is that, even though we have a choice and are free to use another platform, none have quite as large a user base as YouTube, and thus not as large a reach. These technology firms, through the use of AI, are acting as a de-facto policing and legal system, with no form of public oversight. Moreover, the scalability of AI means that this can be effectively implemented on massive platforms (say, ones with a user base of over 10% the internet connected population). AI, combined with a large user base, allows a technology company to start carrying out the functions of the state, at a fraction of the cost. This is an important point, and one that is not picked up on by either report.

7.9. Building auditable algorithms which are more constrained, but also more transparent.

The French report makes it clear that more research is needed to understand AI algorithms, as they are currently extremely difficult to scrutinise, and making them open-source will only be a token gesture to achieve such a goal. It may be possible, and might end up being entirely necessary, to build highly restricted auditable algorithms, even through machine learning. It would mean that the algorithms would need to be developed in a highly structured way, and we

https://www.techrepublic.com/article/facebook-news-feed-changes-could-affect-your-businesss-post-engagement/

35 "*How Content ID works"*, https://support.google.com/youtube/answer/2797370?hl=en

36 *"White noise video on YouTube hit by five copyright claims",* https://www.bbc.co.uk/news/technology-42580523

37 *"When your YouTube video becomes a corporate profit center",* http://www.latimes.com/opinion/op-ed/la-oe-0628-witt-youtube-copyright-20150628-story.html

may even need to give up Turing-completeness[38] in the process, which would undoubtedly irk many developers. However, what we end up with may be algorithms which are less powerful, but socially acceptable. Future generations may look back on early 21st century algorithmic development in the same way we look back on medieval medical techniques; poorly-understood tools in society, where those operating them act without restraint, that often cause more harm than they were trying to deal with. Modern cars must comply with safety and environmental standards, which make the car more expensive, but which society finds more acceptable as a whole. The aviation industry implements volumes of rules and specifications, and every digital process in an aeroplane must be *provably* correct (which of course makes the programming much harder). The same approaches may be needed for AI.

7.10. Shifting the culture of "It is not my problem"

One thing that the French committee fails to address is that some members of the technical community, such as mathematicians and computer scientists, are somewhat dug-in to the idea that "We just do the mathematics, or write the code, or produce designs satisfying the stated requirements. How they are used is not our problem". This requires an educational shift, where students in such technical disciplines are taught how their work impacts society. However, and possibly more pressingly, it also requires a cultural shift. This would involve the technical communities become more aware of their role in society, and accepting moral accountability for their own actions.

Ethical awareness and a broad perspective on societal impact are not "rules" that people are forced to follow by law. They are things that people do because they believe in them. To get people to do caring things, they first need to be taught to care. Moreover, all the educational systems we have in the technical disciplines are designed – likely inadvertently – to train people to not care about the wider perspective or impact of their work. Technical courses usually teach people to focus on narrow problems, which comes with its own set of drawbacks. Simply bolting on a side degree in law, or literature, or philosophy, to a mathematics or computer science or engineering degree, as suggested by the French committee (Villani, et al., p.13), is not enough; these need to be woven together as one. This is a somewhat daunting task; rather than have each field (say, mathematics and law) teach their discipline separately, they would need to come together to prepare an entire syllabus and teaching methodology in a homogenised way. Failing to do so makes it easy for students to perceive the two fields as distinct skill sets with little overlap. Thus, when solving problems, they may choose to look a them through a mathematical lens, *or* a legal lens, but not for a good solution overall.

7.11. Proposing legislation and accreditation

Neither the French committee, nor the British committee, propose any legislation governing AI development, though perhaps it was beyond their remit to specifically address the creation of

38 Most standard programming languages are *Turing-complete*. This means that, for *any* general algorithmic process that can be envisaged, it is possible to write code to carry out that process within that programming language. The issue here is that auditing a general algorithmic process is extremely difficult. By using a programming language that is *not* Turing-complete, we might lose some programming capability, but at the same time be able to understand, explain and audit our algorithms more effectively.

new laws. However, rather than even mentioning the need for new laws, they both seem to rely on various institutes or organisations (e.g.: the Turing Institute, or the Centre for Data Ethics & Innovation) to come up with an ethical framework. If it is genuinely the case that AI ethics is becoming increasingly important, and moreover that there are unlikely to be laws governing its development (which, in a sense, is reasonable: these are extremely complex issues), then it is surprising that neither report proposes an accreditation or licensing mechanism for algorithmic and AI developers. A doctor, or lawyer, cannot practise without a license, and there are penalties for doing so. Moreover, these professional licensing bodies lay out the ethical framework by which their members must abide. At the moment, anyone with a Raspberry Pi can start doing machine learning. Any startup of circa. twelve people can get some data sets, do some machine learning on them, and sell their products to industry. If AI is indeed so important, then we should not be letting everyone do it. Professional accreditation mechanisms for other high-level jobs exist for a reason: people in those professions do things that the average citizen cannot understand or be expected to understand, and if they do not act responsibly and carry out their job with care then there is potentially a large amount of harm to society. These exact reasons carry over to machine learning and algorithmic development.

7.12. The borderless nature of algorithms

One issue that both teams miss is the transnational nature of the impact of algorithms. Even if the UK and France set up an excellent framework for the development, auditing, and implementation of algorithms and machine learning, there is the issue of the fluidity of data and algorithms. An entity based abroad, operating outside the regulatory framework of the UK or France, might provide services to a British or French company. We will start seeing offshore algorithmic processing, much like offshore banking for tax avoidance. Moving data and algorithms costs practically nothing, is practically instant, and can be done in a practically untraceable manner[39]. Without mechanisms to curtail such activities, we will see offshore tax havens fill up with computer scientists and mathematicians doing AI. This very thing happened with Cambridge Analytica, where they used multiple companies and multiple work-forces based in multiple jurisdictions to get around all sorts of electoral advertising rules[40]. This is happening, now, and will continue to happen. The genuine efforts and actions of the British or French committees and governments could all be for nothing, unless they give serious consideration to this issue.

39 By sending encrypted data over obfuscated networks like Tor.

40 *"Cambridge Analytica is dead – but its obscure network is alive and well"*, https://www.theguardian.com/uk-news/2018/may/05/cambridge-analytica-scl-group-new-companies-names

Recent Titles in the Publications of EiM

EiM Discussion Papers

DP 3/2018
M. Chiodo, L. Gordon
How the UK and France are dealing with the AI revolution

DP 2/2018
D. Müller
Is there Ethics in Pure Mathematics?

DP 1/2018
M. Chiodo, P. Bursill-Hall
Four Levels of Ethical Engagement

About EiM

The Cambridge University Ethics in Mathematics Project does research into the ethical questions that arise when doing mathematics. We hope to teach mathematicians about the impact of mathematics in society, and to give them some of the tools and insight that they will need to prevent any harm from taking place.

New Titles
Please consult our website for up-to-date information about EiM publications and ongoing research.
www.ethics.maths.cam.ac.uk

About EiM Discussion Papers

EiM Discussion Papers is part of the *Cambridge University Ethics in Mathematics Project*. The scope of these discussion papers is to stimulate informed discussion around ethical issues in mathematics, the teaching of ethics to mathematicians, and a corpus of case studies.

All discussion papers are subject to internal peer review.

To submit a discussion paper or to contact the coordinators, please send an email to ethics-dp@maths.cam.ac.uk.