View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository@Nottingham

Dr Ansgar Koene – Written evidence (AIC0208)

Written evidence submitted to House of Lords Select Committee on Artificial Intelligence, "*What are the implications of artificial intelligence?*" inquiry by:
Dr Ansgar Koene, Horizon Digital Economy Research Institute (University of Nottingham).
September 9th 2017.

1. Horizon[1] is a Research Institute at The University of Nottingham and a Research Hub within the RCUK Digital Economy programme[2]. Horizon brings together researchers from a broad range of disciplines to investigate the opportunities and challenges arising from the increased use of digital technology in our everyday lives. Dr. Koene is a Senior Research Fellow at Horizon and is co-investigator on the EPSRC funded UnBias[3] project (EPSRC grant EP/N02785X/1) within Horizon which is studying issues related to non-operationally justified bias in algorithmic systems that control access to information online (e.g. search engines, recommender systems, news feeds). Dr. Koene conducts research as part of the UnBias project. An important part of this work includes the facilitation of multi-stakeholder workshops with industry, civil-society organizations, academics and teachers designed to identify experiences, concerns and recommendations information mediating algorithms. Dr. Koene is chair of the IEEE P7003 working group for the development of a *Standard for Algorithm Bias Considerations*[4], and member of the Internet Society (ISOC UK[5]). Dr. Koene is willing to give verbal evidence if so desired.

Questions

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2. The field of artificial intelligence (AI) started in the 1940s with the seminal work on Cybernetics by Norbert Wiener and Computational Theory by Alan Turing. Since then the progress in AI research has been marked by a succession or rapid expansions followed by 'AI winters' triggered when the limitations of practical applications failed to live up to the hyped expectations raised by the rapid advancement in theories. In contrast to the previous 'AI summers' in the 1960s and 1980s, the current wave of enthusiasm is triggered not so much by fundamental advances in AI theory but primarily by the availability of processing power and huge amounts of data which have made it possible to apply AI to real world services.

3. The growth in processing power has been the result of a combination of continuous improvements in micro-processors (CPUs and GPUs) for performing computations, improvements in computer memory and the growth of cloud computing centres which have allowed internet connected devices like smart phones to 'offload' heavy processing tasks 'into the cloud' as is the case with voice controlled personal assistant services such as Siri and Google Now.

4. Thanks to the wide scale adoption of the internet, the accompanying digitisation of services and the use of digital devices (especially smart phones), the amount of data that can be harvested to train AI systems with has grown exponentially for over a decade, to the point where (in 2015) more than 90% of the world's accumulated

data had been produced in the last 2 years[6]. Figure 1 shows the statistics for global data as estimated by a 2013 report by the UN Economic Commission for Europe.

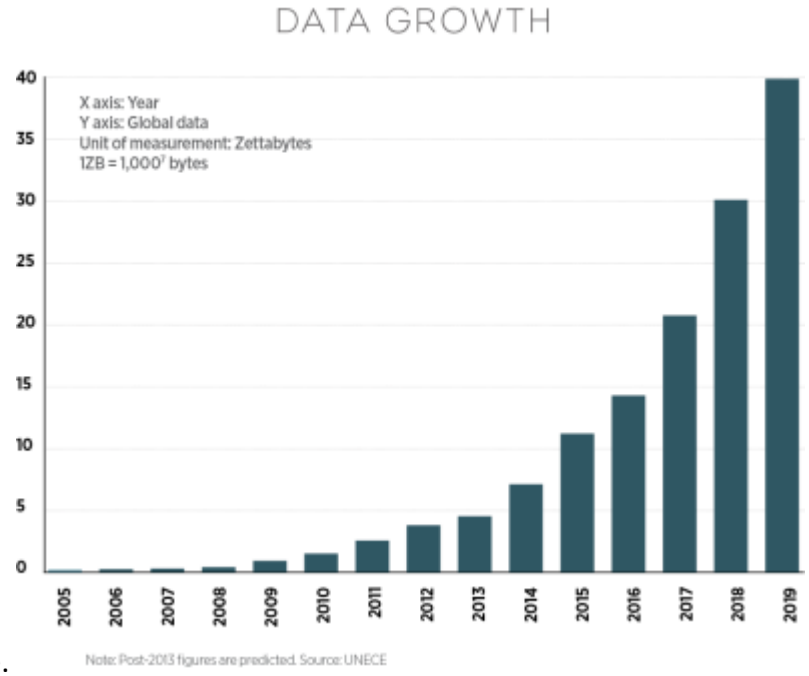

Figure 1: Global Data extimate by UN Economic Commission for Europe

5. For now there is no sign that there will be a significant reduction in the growth of data. While privacy regulation and data protections laws, such as the GDPR may make it slightly more difficult to access some forms of data this will be more than compensated by the growth in connected devices, such as Internet of Things (IoT) devices, and expansion into the developing world.

6. For the next 5 years it is likely that the growth in AI markets will continue and probably expand. Beyond that however, most of the straight forward 'pattern matching', 'data categorization' and 'path finding' kind of applications will have been done. At that point there will be a need for new developments in fundamental AI theory to tackle more open-ended kind of challenges. How AI will develop at that stage will probably depend less on accessing even more data and processing power and more on new scientific breakthroughs in mathematical modelling of complex systems, computational social science, our levels of understanding from the physical sciences and even consciousness research.

2. Is the current level of excitement which surrounds artificial intelligence warranted? Impact on society

7. The current level of excitement surrounding AI is warranted in so far as AI is finally able to automate complex pattern identification and classification problems, enabling organizations (private of public) that hold huge amounts of data to dig through this data to find patterns that can reveal new insights, which in a commercial context can be turned into a competitive advantage. As a result AI is attracting a lot of private sector investment which is boosting the rapid growth in application oriented developments in the field.

8. A leading area of monetization of AI is the personalization of online services, especially advertising. In this context, the new insights that are revealed by the AI analysis of data patterns can include intimate personal information, such as medical conditions, which people may have deliberately been trying not to reveal to a commercial entity.

9. As the application of AI is moving from inconsequential things, such as movie recommendations by Netflix, to more serious matters, such as criminal sentencing recommendations[7], the societal impact will require better oversight to provide the necessary accountability and reliability.

3. How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

10. As part of the UnBias project we have been reviewing case studies of controversies over potential bias in AI practice and scoping the informed opinion of stakeholders in this area (academics, educators, entrepreneurs, staff at platforms, NGOs, and staff at regulatory bodies etc.). It is apparent that the ever-increasing use of AI to support decision-making, whilst providing opportunities for efficiency in practice, carries a great deal of risk relating to unfair or discriminatory outcomes. When considering the role of AI in decision making we need to think not only of cases where an algorithm is the complete and final arbiter of a decision process, but also the many cases where AI play a key role in shaping a decision process, even when the final decision is made by humans; this may be illustrated by the now [in]famous example of the sentencing support algorithm used in some US courts which was shown to be biased[7]. Given the ubiquitous nature of computer based processing of data, almost all services, be they government, public, business or otherwise, are in some way affected by AI decision-making. As the complexity of these algorithmic practices increases, so do the inherent risks of bias as there are a greater number of stages in the process where errors can occur and accumulate. These problems are in turn exacerbated by the absence of oversight and effective regulation.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

11. Commercially, those who are gaining the most from AI are companies such as online platforms that have access to large sources of data because the availability of data is a key driver in the current AI development (see response to question 1).

12. Among consumers those that gain the most are people who match the ‘white male upper-middle class’ interests and demographics of the coders and beta testers who create and validate the AIs. Without consciously intending to do so, developers will naturally be better at optimising the systems to match their own needs and interests. Due to an unfortunate lack in diversity among coders this is likely to lead to systems that disadvantage some groups in society. A start example of this is

provided by Joy Buolamwini, an African-American researcher at the MIT Research Lab, who found that she has to don a white mask because her face is often not detected by generic facial-recognition software used by robotics programs[8].

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

13. The recent research work that we have conducted with young people has highlighted important concerns around algorithm use, including AI, and trust issues. Results from a series of 'Youth Juries'[9] show that many young people experience a lack of trust toward the digital world and are demanding a broader curriculum beyond the current provision of e-safety to help them understand algorithmic practices, and to increase their digital agency and confidence. Current use of AI in decision-making (e.g., job recruitment agencies) appears surprising to many young people, especially for those unaware of such practices. Algorithms are perceived for most young people as a necessary mechanism to filter, rank or select large amounts of data but its opacity and lack of accessibility or transparency is viewed with suspicion and undermines trust in the system. The Youth Juries also facilitated young people to deliberate together about what they require to regain this trust – the request is for a comprehensive digital education as well as for choices online to be meaningful and transparent.

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

14. As indicated in response to question 1 (point 6) the main areas of AI strength currently are in 'pattern matching', 'data categorization' and 'path finding'. Clerical and service sector jobs involving these kinds of information processing, for instance HR admin, are likely to experience rapid automation through AI. Jobs where talk output cannot easily be transformed into something resembling 'sorting into categories' are much less likely to be solved by the current AI methods.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be address. How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

15. No comment

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

16. When discussing bias in AI decision-making it is important to start with a clear distinction between operationally-justified and non-operationally-justified bias. Justified bias prioritizes certain items/people as part of performing the desired task of the algorithm, e.g. identifying frail individuals when assigning medical prioritization. Non-operationally-justified bias by contrast is not integral to being able to do the task, and is often unintended and its presence is unknown unless explicitly looked for.

17. In order to identify good practice related to biases or discrimination, some important processual issues must be taken into account, for example:
    I. In order to understand the scope for AI decision-making in relation to bias adequately and appropriately, it is necessary to engage with, and integrate the views of, multiple stakeholders to understand how AIs are designed, developed and appropriated into the social world, how they have been experienced, and what the concerns surrounding their use are;
    II. Importantly, this undertaking and exploration should be achieved through rigorous research rather than abstract orientations towards good practice in relation to AI: thus, considering examples of the consequences that people have experienced when AIs have been implemented, particular scenarios surrounding their use, and as emphasised in the point above- talking to people about their experiences.
    III. Given the complexities of the landscape in which AI are developed and used- we need to recognise that it is difficult, in some cases impossible, to develop completely unbiased algorithms and that this would be an unrealistic ideal to aim towards. Instead, it is important to base good practice on a balanced understanding and considering of multi-stakeholder needs.

18. The need for 'good practice' guidance regarding bias in algorithmic decision-making has also been recognized by professional associations such as the Institute of Electrical and Electronic Engineers (IEEE) which in April 2016 launched a Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous system[10]. As part of this initiative Dr Koene is chairing the working group for the development of a Standard on Algorithm Bias Considerations[11] which will provide certification oriented methodologies to identify and mitigate non-operationally-justified algorithm biases through:
    I. the use of benchmarking procedures
    II. criteria for selecting bias validation data sets
    III. guidelines for the communication of application domain limitations (using the algorithm for purposes beyond this scope invalidates the certification)

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

19. In principle, AI decisions can be traced, step by step, to reconstruct how the outcome was arrived at. The problem with many of the more complex 'big data' type processes is the high dimensionality of the underlying data. This make it very difficult to comprehend which contributing factors are salient and which are effectively acting as noise (for any given specific decision). Analytic methods for dimension reduction can be used to make this more understandable in many situations, but may need to be applied on a case-by-case basis to appropriately evaluate the important outlying and challenging cases.

20. Similarly, it is important to note that many 'big data' AI algorithms learn from the data they are supplied with and modify their behaviour. We must look not only at the code that constitutes the AI algorithm, but the "training data" from which it learns. Practically this is becoming increasingly difficult as algorithms become

embedded in off-the-shelf software packages and cloud services, where the algorithm itself is reused in various contexts and trained on different data – there is not one point at which the code and data are viewed together.

21. The IEEE Global Initiative (see point 19) are also working to establish a Standard for Transparency of Autonomous Systems[12] which aims to set out measurable, testable levels of transparency. The working group for this standard is chaired by Prof. Alan Winfield[13].

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

22. While there is a need for *meaningful transparency*, this does not require that copyrighted code (or data) is made public. Within the community currently researching this topic, a recurring suggestion is the use of a neutral (or government associated) auditing body that could be tasked with certifying AI systems through a process of expert analysis. This algorithm auditing could be done under a non-disclosure-agreement, protecting the IP, and the individual data. A detailed discussion outlining arguments in favour of such an approach was developed in an open access published paper by Andrew Tutt with the title "An FDA for Algorithms"[14].

23. Even if the copyrighted code is not made public, somehow making aspects of the design of AIs more visible may still be useful. We see how the food industry make elements of their produce accessible for consumers to allow for consumers to make informed decisions about what they purchase. At this point it is difficult to say what is better/worse without full and proper engagement with industry and other stakeholders, as we are currently engaged in through the UnBias project.

24. It is necessary to have a dialogue with industry to understand their genuine concerns surrounding increased transparency, and how a way forward can be forged. There are elements of business procedures which have to be made transparent already (e.g. the requirements for audit, health and safety, etc…) so it is not that they are unaccustomed to such requirements. However, given that there is an element of commercial sensitivity in this context, it is important to see what suggestions they would have to allow for increased transparency.

25. We should be careful that we do not give the impression that commercial interests supersede the rights of people to obtain information about themselves. We should be cautious about assuming industry interests are more important than other ones, and move forward with a balanced approach.

26. Finally, the traditional bargain between society and inventors has been the patent - disclosure to stimulate innovation in return for commercial protection – the question arises as to what role might patents play in transparency. However, the situation concerning software patents is globally complex, but then the issue of algorithmic transparency is rapidly becoming a global issue.

27. What is essential here is to create a *meaningful transparency*: that is a transparency that all stakeholders can engage with, allowing the workings of, and practical

implications of, AI to be accessible across the diverse stakeholder base that experience them.

28. In order to create a meaningful transparency, we need to understand what stakeholders feel such a transparency would have to incorporate for them to be adequately informed, and enable them to engage with the positive and negative implications of algorithms. Though it is unlikely that there would be complete consensus, such stakeholder engagement can provide key insights for the nature and shape of solutions to be developed.
29. Importantly, this meaningful transparency should also relate to a *meaningful accountability*. It is not enough for stakeholders just to understand how AI are developed and how they make decisions. In making things meaningfully transparent, stakeholders should be given some agency to challenge algorithmic decision-making processes and outcomes.

12. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

30. The right to explanation in GDPR is still open to interpretation and the actual practice will become established as cases unfold when enforcement starts in 2018. For example, the right to recourse and to challenge algorithmic made decisions, is restricted to decisions that are made *fully autonomously* by algorithms and that have clearly *significant* impact on the person – it will be some time before we understand how these clauses will be implemented. The recent paper by Wachter et al.[15] puts forward the case that much more is needed to deliver a '*right to explanation*'.

31. The Council of Europe's Committee of Experts on Internet Intermediaries (MSI-NET)[16] is currently also exploring the human rights dimensions of automated data procession techniques (in particular algorithms) and possible regulatory implications. As part of this investigation a preliminary report[17] was published on February 20th 2017 which includes a number of relevant case studies and recommendations that are applicable to the topic of this inquiry.

10 September 2017

---

[1] http://www.horizon.ac.uk
[2] https://www.epsrc.ac.uk/links/councils/research-councils-uk-rcuk/digital-economy-research-rcuk/
[3] http://unbias.wp.horizon.ac.uk
[4] https://standards.ieee.org/develop/project/7003.html
[5] http://isoc-e.org/
[6] http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/
[7] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[8] https://youtu.be/IbnVu3At-0o
[9] http://oer.horizon.ac.uk/5rights-youth-juries/
[10] https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[11] https://standards.ieee.org/develop/project/7003.html
[12] https://standards.ieee.org/develop/project/7001.html
[13] http://people.uwe.ac.uk/Pages/person.aspx?accountname=campus\a-winfield
[14] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994
[15] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469
[16] https://www.coe.int/en/web/freedom-expression/committee-of-experts-on-internet-intermediaries-msi-net-
[17] http://rm.coe.int/doc/09000016806fe644