August 27, 2017

SCHOOL *of* INFORMATICS

The University of Edinburgh
The Informatics Forum
10 Crichton Street
Edinburgh EH8 9AB

Telephone +44-131 651 3441
Secretary +44-131 651 3443
Fax +44-131 651 1426
E-mail rbf@inf.ed.ac.uk
URL www.inf.ed.ac.uk

House of Lords

Dear Lords and Ladies

CALL FOR EVIDENCE - ARTIFICIAL INTELLIGENCE

PRELIMINARY STATEMENTS

Expertise from The University of Edinburgh

Since the mid-1960s, The University of Edinburgh has been actively researching Artificial Intelligence (hereafter AI), with the contributions of an estimated 2000 person years of academic and research staff, and an estimated 1000 PhD and 2000 MSc students, all specialising in AI topics. The School of Informatics (with colleagues from other Schools) is the largest academic AI group in Europe. The School's wide-ranging and extensive effort has been applied to advancing general AI methods and a range of AI specialisms, including natural language processing, machine learning, robotics and computer vision, planning, knowledge representation and reasoning. The results of this research have also been widely applied in many sectors. While developing their own research agenda, there is still considerable interaction between the specialisms, in part driven by shared underpinning technologies (*e.g.* probabilistic modelling, data science methods, deep 'neural' networks, machine learning, knowledge representation and reasoning).

Although there has already been a huge investment by the UK in AI research, training and technology transfer, we collectively believe that the development of AI still remains an exciting long-term endeavour, and AI will be one of the defining technologies of the future.

Definition of Artificial Intelligence

AI, as currently realised, is not what is seen on television or in the cinema. It is a pervasive and powerful technology, but it is not yet a general purpose technology. It is currently deployed as a performance enhancing component in a range of highly specialist applications. These can be reasonably straightforward tasks (*e.g.* simple precision agriculture, car driver emergency braking, camera face detection, smart-phone predictive text, speech transcription and generation, smart search, enhanced household appliances). Or they can be more complex decision-making processes such as natural language understanding, machine translation, self-driving cars, IBM's Watson and personal assistants like Apple's Siri. Some are clever applications, but their abilities do not go beyond their narrow domain.[1] The methods underpinning these applications are not new and magic technologies. Instead, they are cleverly engineered collections of advanced computer algorithms, including optimisation, search, knowledge representation, data mining, machine learning, sense data analysis, as well as deep networks and robotics. The result is that AI is better defined by its

[1] A. Darwiche, Human-Level Intelligence or Animal-Like Abilities, CACM, in review.

applications than by its underpinning technologies.

The "intelligence" of an application needs to be distinguished from the "autonomy" of an application. The former gives the application enhanced capabilities; the latter gives the application independent decision making and the ability to act. AI applications have varying degrees of intelligence. Few (to date) have autonomy, and that autonomy is usually closely constrained and formulaic, *e.g.* in an autonomous vehicle, a business-to-business purchasing agent, a stock-trading agent. Conversely, there are many autonomous and semi-autonomous systems such as self-guided missiles, nuclear power plant emergency shutdown systems, and aircraft autopilots that do not exhibit the kinds of intelligence AI is concerned with.

Many of the House of Lords consultation questions are not simply AI questions. Issues of privacy, liability, economic displacement, monopoly, transparency, governance, licensing are relevant to the broader modern economic ecology; AI is only one component. For example, a largely automated factory invokes many of the same issues, but need not be based on substantial AI elements.

Where the real dangers of AI lie

The real and current dangers of AI do not lie in superhuman AI, irrespective of what one sees in the cinema or hears in the media.[2] There have been some major AI successes where performance is close to or exceeds human skill, *e.g.* autonomous vehicles, hand written character recognition, speech transcription, speech generation, partial machine translation, partial text understanding, and selected areas of medical diagnosis. Each of these very narrow competences is the product of 30-50 years of research by hundreds, if not thousands, of scientists and engineers. The human mind is claimed to be the most complex mechanism known to humans - replicating its sophisticated and general capabilities is far beyond current capabilities.

Nonetheless, there are genuine dangers arising from widespread use of AI.

1. The ability to compute at high speed and large-scale means that significant disasters can arise from automated reasoning errors or inadequate understanding of the fragility of complex interconnected systems before humans can intervene (*e.g.* the 2010 stock market "Flash Crash" exacerbated by automated High Frequency Trading algorithms). Similar vulnerabilities arise elsewhere because it is hard to predict all consequences of complex interacting systems. This is especially the case when the algorithms within each system are commercial or military secrets, as was the case with the stock trading systems involved in the flash crash.
2. Economies can decline by being out-competed given the additional leverage of AI in algorithmic decision making and automation (*e.g.* business-to-business sourcing of cheapest materials, large-data analysis of trends and other business information, agent-based modelling of economic scenarios, flexible factory automation). This is in addition to the risks of losing UK income due to the improved commercial prospects of competing products whose performance is enhanced by AI methods (*e.g.* predictive text in mobile telephones).
3. Social unrest can increase dramatically due to reduced opportunities for meaningful employ-

[2] A. Bundy, Smart Machines Are Not a Threat to Humanity, CACM, 2017.

ment, as a consequence of automated manufacturing capability (and the concentration of wealth to those who can afford to invest in it), and of the displacements of middle-level skilled labour replaced by automated service systems (*e.g.* travel agents, sales executives).

4. Smart, but indiscriminant weapons. They will have limited targeting mechanisms, and will be prone to incorrect decisions. Automated object recognition algorithms have advanced greatly but performance typically varies from 10-90%, depending on the types of objects and number of categories. Even a 1% false positive rate could have a devastating impact on civilian populations. For example, consider the consequences arising already from the wide-spread use of land-mines, which are passive weapons mainly affecting non-combatants. With a little AI, they can actively respond to or even seek targets, based on heat-signatures and movement. They are vulnerable to being hacked or left behind, possibly damaged, after a conflict, causing unintended damage long after the original conflict.
5. Social unrest could increase dramatically due to the speed of change and innovation. AI methods could be adopted widely and at large scale due to their economic advantages. Consider the impact of "smart-phones" on different social generations. Human society has not experienced this rate nor type of change previously.
6. Social problems could arise due to widespread ignorance about the capabilities of AI enhanced systems. People are familiar with the inadequacies of speech understanding systems (*e.g.* the humourous Burnistown "11" elevator video[3]). But most people are unaware of the AI enhancements in the products and systems that they use. Instead, the understanding of the non-specialist is largely shaped by the mainstream media (newspaper humour and scare stories, high-profile "end of the world" statements, exciting but unrealistic movies). The consequences of the lack of understanding are unrealistic fears and unrealistic expectations.

Specific Recommendations

We note that these recommendations arise in the context of the discussion on AI, but are, in fact, also relevant to non-AI technologies, such as data collection, storage and analysis, data science, advanced manufacturing, video surveillance, and social media.

I **Economics and Employment**: Because digital objects can be easily replicated and distributed, popular products can easily lead to concentrated wealth-generation by a few dominant market actors, as seen by the rise of *e.g.* Microsoft, Google, Amazon, eBay, and Facebook. Thus, innovative models of wealth and benefit distribution are needed. Bill Gates suggested to tax robots[4], but but this could be extended to more general AI systems. In response to the displaced human labour, we advocate an increase in training and employment opportunities in human-based services (*e.g.* healthcare, ageing, teaching, social care, activities, tourism). It's particularly important for people who would previously have been employed in low-skill jobs that will cease to exist. Any 'living-wage' would need to be set at a level that enables people to participate in these services and the general economy. Substan-

[3]https://www.youtube.com/watch?v=sAz_UvnUeuU

[4]https://en.wikipedia.org/wiki/Robot_tax

tial economics research will be needed to develop models for how an economy with decreasing amounts of human labour might work and how the benefits of the society will be distributed. Global equality issues will become more urgent (and consequences, such as migration).

II **Safety**: Most AI is embedded in products and systems, which are already largely regulated and subject to liability legislation. It is therefore not obvious that widespread new legislation is needed. Systems with embedded AI should be covered under standard recall and fault recourse mechanisms. Manufacturers should demonstrate due diligence, as with any other product. There are existing models of risk and methods for standards and their verification. These need to be enhanced, but not necessarily replaced. Additional legislation may be needed: 1) to provide a framework for requiring satisfaction of specified standards as part of the licensing for deployment of critical AI systems, 2) for situations where multiple independent AI components are integrated into a larger system (either on a single device or across a network), and 3) to address the issue of computer speeds, wherein actions happen at a time scale far faster than humans can respond or intervene. Enhanced developments in cybersecurity are needed to make AI apps safer and less hackable. A particular worry is the use of embedded apps by companies with limited experience with software development and computer security (*e.g.* car and household appliance manufacturers).

III **Privacy and Use of Personal Data**: Because of the potential for widespread collection and automated collation of personal data, and their subsequent use in automated decision-making systems (*e.g.* insurance pricing, social benefit determination), it is likely that additional legislation will be needed to govern activity around discovery, access, deletion and correction of personal data. For example, one should have access and control over to the data collected by an "always-on" personal digital assistant. Another issue concerns what information can be uploaded for corporate analysis from these personal digital assistants. The provenance of training data should be explicit, and should be "fair", *e.g.* representative of the variety of the human population recorded in the dataset.[5] Legal and regulatory systems need to be enhanced (esp. considering the widespread "illegal" use of software "cookies" in the EU).[6]

IV **Education and Public Awareness**: People need to understand broadly the capabilities and limitations of AI enhanced systems and products, and how these capabilities are expected to advance with time. They should also be familiar with their rights and risks. Introduction could occur at school, but given the changes that will occur post-schooling, some web-based public information mechanism would be essential. Royal Institution lectures[7] are useful, but appeal to a narrow audience. Broader dissemination and engagement mechanisms are needed, particularly since the effects are likely to affect less skilled labour harder and earlier.

[5] Royal Society Machine Learning Report 2017.

[6] *e.g.* https://www.theguardian.com/technology/2015/mar/31/facebook-tracks-all-visitors-breaching-eu-law-report

[7] Prof. Chris Bishop, October 2016 Royal Institution "Discourse" on Artificial Intelligence

RESPONSE TO SPECIFIC QUESTIONS

A **Current and future state of AI**

Current state: narrowly specialised AI applications are becoming pervasive, *e.g.* auto-correcting and predictive phone app text. However, there is no clear boundary between an AI-based application and other well engineered computing applications. Most AI enhancements are emerging as a convergence of 30-40 years of academic research (autonomous vehicles, natural language understanding, automated translation), large datasets providing examples of many variations of the recorded phenomenon (customer preferences, automotive faults), improved machine learning and data mining techniques, and cheap desktop supercomputing.

Future developments: lots of increasingly sophisticated, embedded, special purpose applications will provide improvements to personal efficiency and informedness. There will be gradually improving general question answering systems. and widespread medical discovery and diagnosis systems. Increasingly capable and general object recognition systems and reliable and commercially feasible 2-legged mobile robots will follow. These developments are likely to accelerate the competition and gains of major companies and national entities with the resources to invest in research, development and deployment of AI systems. There is likely to be a thriving ecosystem of small AI players. These developments are also likely to accelerate the concentration of wealth in these companies and countries, leading to increased social problems, including migration pressures.

Human level general intelligence[8] *AI is much further in the future.*

B **Is excitement warranted?**

Yes and no. There are increasing numbers of special purpose, incrementally useful applications. These will keep increasing. Although AI has seen many "hype cycles" and reassessments (and there are likely to be more), there have also been real gains, to the point that elements of AI technology are present in almost anything involving a computer.

C **Preparing the public and their understanding**

As noted above, there should be relevant and regularly updated exposure at school level, and public awareness media for all ages. There could be training courses for "application advisers", much as there are financial advisers at present. They could advise on which apps to use, how to connect and use them, and how to stay safe. This could be a commercial skill, but with training supported and encouraged by government.

D **Who will gain most/least?**

Under current economic models, it seems likely that the big winners will be the organisations that have the resources to invest in AI technologies (national, military, or commercial). As a software technology, AI applications are essentially infinitely replicable. There could eventually be a small set of apps competing in most product areas (*e.g.* current software for most things other than smart-phone apps). The producers of these apps will become very wealthy because of the ease of production and distribution (*e.g.* sale of Microsoft software, small transaction or usage license fees). In terms of the public good, everyone is likely to benefit from products and services with improved and more personalised services. Improved

[8] Which itself is not well defined nor understood.

transport, energy distribution, manufacture and agriculture could reduce production costs. Improved medical diagnosis would benefit all. The reduced need for many types of semi-skilled human capital and the training cost and pre-requisites for high-skilled labour are likely to lead to an increasing pool of underemployed and low wage people.

E **Data Monopolies**
Large personal datasets can be collected by any AI-based service, *e.g.* most data science-based services. But even non-AI based web services will collect large amounts of personal data, so issues concerning data monopolies are not just AI issues. Central concerns include: whether data can be sold, data security, provenance, and different levels of access and detail.

F **Ethical Issues**
The core issues are safety, liability, and fairness. We have a concern for the potential for near-instantaneous disasters (*e.g.* like the stock-trading disaster), their scale, and responsibility for when they do occur. We need standards and legal liabilities for AI enhanced products, but based on the product itself, rather than on any particular aspect of the AI. The fair use of data is not strictly an AI issue: privacy, consent, diversity and the impact on democracy are issues that arise in the context of general big data, cybersecurity, and social media.

G **Transparency of AI**
It would be ideal if an AI system could provide an intelligible justification for its reasoning. However, except for the simplest of rule-based systems, this is rarely possible. Logic and proof-based systems can reason based on hundreds or thousands of steps. Probability based systems generally use hypothesised causal relations with probabilities learned from collating possibly only a few, or possibly millions of instances. The recently developed deep learning methods tend to out-perform other methods, but their decision processes are numerical, and are generally completely unintelligible. *What seems more feasible is to only licence critical AI systems that satisfy a set of standardised tests, irrespective of the mechanism used by the AI component. Equally, one could question whether most human decision making is transparent and highly accurate.*

H **Role of government**
Any legislation that affects the deployment of AI systems will need to be agreed internationally, otherwise the UK acting alone risks leaving itself at an economic disadvantage. The UK is well placed to further invent, develop, and exploit AI methods; any legislation should ensure that this supportive environment continues. Existing technical, economic, and social legislative mechanisms are adequate for the moment to cover most AI as well as other computer-based areas, such general software liability, databases and privacy, and cybersecurity.

We appreciate being consulted on this issue and hope that our statement provides a helpful contribution. We welcome the opportunity to continue the discussion if desired.

Yours sincerely,
Prof. Robert Fisher FIAPR FBMVA and colleagues: Prof. Alan Bundy FRS FRSE FREng FACM, Prof. Simon King FIEEE, Prof. David Robertson, Dr. Michael Rovatsos, Prof. Austin Tate FREng FRSE FAAAI, Prof. Chris Williams FRSS