Digital Cooperation – call for contributions

Response by the Institute for Applied Data Science, Queen Mary University of London

Richard Ashcroft, Professor of Bioethics & Fellow, School of Law

Dr Adrian Bevan, Head of Particle Physics Research Centre & Turing Fellow, School of Physics and Astronomy

Lord Clement-Jones, former member of the Lords Artificial Intelligence Committee and Chair of Council, Queen Mary University of London

> *'The High-level Panel on Digital Cooperation was established by the UN Secretary-General in July 2018 to advance proposals to strengthen cooperation in the digital realm, and contribute to the broader global dialogue on how interdisciplinary and cooperative approaches can help ensure a safe and inclusive digital future.'*

The purpose of this document is to provide input to the High-level Panel on Digital Cooperation regarding plausible ways to strengthen cooperation in the digital realm, and contribute to the broader global dialogue on how interdisciplinary and cooperative approaches can help ensure a safe and inclusive digital future. The underlying issue is that technology has rapidly advanced in recent years and in doing so, existing structures have failed to keep pace. There are three key aspects that the panel has requested input on and we consider these in turn in the following. The aspects are:

1. What values and principles should underpin cooperation in the digital realm?
2. How can stakeholders cooperate more effectively in the digital realm, and how can marginalised stakeholders be included?
3. How can cooperation among stakeholders be improved in areas such as inclusive development; inclusive participation in the digital economy; data; protection of human rights online; human voice and human agency in the digital age; digital trust and security; and building the capacity of individuals, institutions and governments for the digital transformation?

Introduction

To start understanding the multitude of issues pertaining to digital co-operation, one needs to understand the landscape of stakeholders and their current as well as ideal/practicable relationships. From this one can consider the key ingredients for a framework built around values to guide stakeholder engagement. Stakeholders come from all communities on the planet and have a range of scales from the individual to multinational corporations. Here, wealth is not a significant discriminating factor. Issues affecting or marginalising those without access to technology in the developed world can also affect those from less affluent parts of the world. Those with access to technology face a different set of issues, and individuals are generally marginalised to some degree in favour of a select few large affluent stakeholders. We consider cooperation between stakeholders ranging from private individuals of any age, companies operating in regions or nationally, local and national governments, non-profit organisations, and large global multinational corporations.

Digital technology in this age has broad meaning and is not restricted to any particular niche. While AI is the zeitgeist, technology encompasses a broad base of issues beyond this. For small samples of data we are able to deploy smart data techniques while many organisations and research laboratories focus on large data samples, referred to as big data. Data science algorithms may be computationally transparent and cheap, or conversely may be difficult to understand and expensive to compute. Understanding bias and accuracy of predictions of algorithms is as important as ethical considerations surrounding their use. Record keeping in databases or using blockchain naturally leads to the need for an understanding of cooperation between the entities creating, updating or accessing those records. Algorithms may be embedded in hardware, for example computer control in transport systems for rail networks, cars as well as commercial and military aeroplanes and more generally the internet of things. Fundamentally, it is not sufficient to develop a technology without regard to the consequences of its use or misuse. Issues surrounding digital cooperation encompass this broad ensemble of technologies.

Some regions have established advisory and legal frameworks to protect society and organisations, for example the General Data Protection Regulation Compliance (GDPR) of the EU [1]. However, this legislation does not go far enough to fully protect the rights of stakeholders. Where advisory frameworks exist, such as the UK Data Science Ethical Framework (UK-DSEF) [2], or the IEEE ethically aligned design [3], they are guidelines that are limited in scope, discuss points of consideration and work toward best practice. While some aspects of those frameworks are aligned generally with good data science irrespective of the entity doing that work, others are orthogonal to the business models driving many large companies. For example, only collecting and using the minimum of data required, as recommended in the UK-DSEF, is at odds with generating marketing revenue from speculatively analysing all available data on an individual to predict spending patterns.

One thing which should be explicit in cooperation between companies, government and users is the nature of the business model that is in play. It is often not clear to the businesses themselves at the outset and emerges as the business develops and may change significantly as it grows. This can mean that what the user thinks the business is providing can be very different from what the business itself and its investors think it is doing. Examples include Google which to users is mainly a search engine and email service provider but generates its revenues from data analysis and advertising. Amazon's operations in service provision for the cloud contrasts with its image as a giant convenient retailer.

Every stakeholder has an agenda, and those will vary wildly. For example, private individuals are often exposed to lengthy user agreements written in complex and unfamiliar language that effectively occludes the intention of how data could be used. These agreements often require that the individual relinquishes intellectual property rights on their data. Early adopters of technology have accepted this as acceptable in return for access to systems whereas larger organisations can afford to negotiate or develop bespoke arrangements to retain those rights. This imbalance underpins the business case of many tech companies. While the outcomes may be considered acceptable by early adopters of technology without a concern for consequence, ramifications of that action take time to become fully apparent. Potential issues have and will continue to arise with more widespread and long-term use of technology. Having identified unethical or inappropriate use of information it becomes necessary to modify what is considered acceptable in order to re-normalise practice. This will be an iterative process, requiring attention for each significant step forward in the development and application of data science. Generally, it may be neither practical nor financially beneficial for a company to implement an ethical and appropriate use of data that considers every possible stakeholder.

The definition of a set of common values underpinning data cooperation will require establishing a suitable and appropriate balance between many factors, including those indicated above. Establishing the basic principles of data cooperation in this context is a complex problem that will ultimately focus on generalisations. To start this process, we can and should take inspiration from existing frameworks, ideals and declarations using ethical reasoning for guidance.

Values and principles to underpin cooperation

Ethical behaviour should guide the principles of data cooperation in the context of existing frameworks. Where those do not go far enough to exemplify or protect individual concerns, there should be an effort to reaffirm or extend those constructs in the context of the digital age. Here we consider several possible mechanisms that could be relevant.

The pre-digital age rights of individuals were encoded in the Universal Declaration of Human Rights (UHDR) [4] written in 1948. Since the UDHR was written, technology has advanced beyond all recognition. It would be prudent to reassess the validity of the articles in the declaration and their appropriateness in the context of the zeitgeist. This may in turn lead to better protection of the individual. A number of the articles naturally lend themselves to extension in the digital realm, for example, if one considers article 27.2, which states:

> "Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author."

This right is entwined in a number of aspects of the use and misuse of social media. Furthermore the style and verbosity of user agreements helps to obscure potential use cases of data posted on social media. The digital society has chosen to relinquish a basic human right time and time again with little or no consideration for future consequences, trading this for access to social media resources. As a result the right to be forgotten has been enshrined as article 17 in the GDPR. This issue can particularly affect younger generations who embrace technology and in doing so may generate issues surrounding future employability. The UK independent think tank *Doteveryone* has described a Digital Understanding Model that recommends the development of new codes of practice surrounding such issues [5] that would help raise awareness of potential issues of using technology with society at large. While the articles of the UDHR may be sufficient for the digital age, we may benefit from contextualising their meaning, or clarifying ways in which we may be potentially waiving rights, in the digital age as a part of reaffirming commitment to maintaining the underlying rights. We believe that the Universal Declaration of Human Rights remains fit for purpose, however it needs to be contextualised for modern day purposes.

Recommendation: UDHR should be reaffirmed in the context of the digital age, so that online and offline rights of individuals are respected equally.

There already exists guidance on public/private partnership policies in medicine and public health to combat disease (and drug regulation). This could be translated into a model for the digital area. However, while we probably know what good public health looks like based on decades of global work in this area, there isn't the same basic understanding of AI.

There are industry sectors who will not be seen as responsible partners – see the World Bank or World Health Organisation (WHO) for guidelines on these partnerships. [6]

Guidance on good digital cooperation between industry and Government could usefully learn from work on partnerships between health and industry. Where trusted records are required,

technologies such as blockchain could be deployed; and understanding the interactions required for blockchain use would naturally require dialogue with the broad spectrum of relevant stakeholders.

Recommendation: Explore establishing responsible partnerships. There are industry sectors who are seen as responsible partners (see World Bank/WHO guidelines [6] on such partnerships).

The overarching process of assessing data cooperation should be rooted in the consideration of the ethical and legal nature related to use of data. There are multiple levels for consideration as the context in which data is used can lead to very different consequences for individuals. For example, high level information analogous to an individual being photographed in the background of a major landmark, and processed within a photo-sharing website is at first glance unlikely to constitute a significant departure from pre-digital data archival contexts. The ease of access to the data in the modern age, raises a potential concern when considering issues such as witness protection programmes where individuals are relocated. The processing of large data sets using modern facial recognition software could ultimately result in the ability to compromise the safety of such individuals. Such technology could be developed with the aspiration of helping identify missing persons, people with amnesia, criminals or terrorists. An unintended consequence of this could be compromising the safety of people in witness protection. The EU coordinated plan on AI [7] (2018) calls for "ethics by design", which could help understand such issues surrounding emerging technology, and guide the development for appropriate safeguards. Organisations engaged with data analytics should consider ethical reviews of their work or aspirations. For example, this has been done recently by the West Midlands Police force in the UK. A report has been prepared by the Alan Turing Institute Data Ethics Group alongside the Digital Ethics Panel for Policing [8].

Access to a tranche of personalised data can provide a unique signature of an individual, which if analysed could be considered invasive. An example of this is illustrated in the case of an individual performing a web search for a word that was part of a previous address to find a result identifying the long-since forgotten street and house number that they lived in as a child. In this case the search algorithm output leads to an extremely personalised result. The fact that algorithms can learn sufficient information about some individuals, as in this case, so as to provide uncannily personal responses is an unintended consequence of the predictive power of the model used. Similar issues may arise in the use of anonymised health records, which may unwittingly allow the identification of individuals. It is likely that there will be instances where the provision of model predictions from personalised data samples will have the potential to cross ethical boundaries, and this is something that should be explored in order to identify means of mitigation.

Recommendation: Ethical considerations should always be applied to technology, in terms of design and use. Organisations should drive the ethical review of their use of technology.

Stakeholder cooperation and mitigating marginalisation

The pace at which technology is evolving is unprecedented. This necessitates frequent reviews of policy to ensure that it remains relevant. This is particularly pertinent for issues surrounding understanding what algorithms are doing and the potential for marginalisation of parts of society as technology outpaces their ability to learn or adapt.

The think tank *Doteveryone* has studied digital technology in relation to society [5]. This group has focussed on reviewing the understanding of technology and how it tailors (or biases) output provided to the user. They report that 62% of users surveyed did not understand that social media sites tailor responses to individuals. The gap in understanding what it is that algorithms used by

technology companies are doing from a high-level transparency perspective, is found to be significant. This group call for:

1. ***New codes of practice for design and consent*** *in the technology industry, so that products and services do the hard work are to be understandable*
2. ***A central, trusted and independent source of information*** *with clear, up to date plain English explanations of the key aspects of digital understanding*
3. ***Public engagement to support digital understanding at all levels of society*** *- not just for children and with a specific focus on digital leadership for public institutions*

The group also proposes a model of Digital Understanding for individuals, consumers, workers and society spanning from cursory to deep understanding through awareness, discovery and questioning. The issue of online data privacy has also been studied by the Carnegie UK Trust [9]. This report found that there is no consensus on definitions, interpretations and boundaries regarding data privacy and that there is significant public concern about data privacy.

There is a second aspect of understanding that is worth raising here – the complexity of algorithms can easily result in them not actually addressing the issue intended. Biased models can compound transparency and understanding of services. Technology companies have a duty of care to ensure that sufficient attention is paid during the development and trialling of algorithms in order to understand that they behave as intended. Due to the complex nature of many modern algorithms, understanding models will be a technological challenge to be faced. Due diligence in this area will help ensure that claims made are justified, and that biased models can be avoided, or deployed with appropriate warnings in order to avoid false advertising or misuse. This is required for users to fully consent to submit their data in and ethical sense and to do otherwise would be unethical. This issue has been extensively discussed elsewhere and one set of examples can be found in [10].

Access to technology in the developed world depends on the personal wealth of individuals. Some parts of society will have no individual access to technology resource. Governments should work to ensure that this resource gap is mitigated in order to provide appropriate access for the population. The nature of the digital revolution means that it is not just school level education where this is imperative, but it also affects the adult population as the nature of work will change with the advance of technology. This issue of a resource gap is accentuated in poorer parts of the world. The technological divide between a village where the priorities may be establishing water and sanitation vs that of a developed world city will continue to grow as technological advancements increase that divide. Developed countries and leading companies should consider investing in bridging this gap. Helping the developing world from a commercial perspective could be seen as a long-term investment in developing future markets, while from a national perspective this is working to ensure global stability. This is not just the right thing to do, but major stakeholders in technology have a vested interest in ensuring that this divide is bridged. A successful example of this ethos can be seen with the use of mobile phones in the developing world. One can trek for several days into remote parts of the world to see locals able to communicate with neighbouring villages that would otherwise require several hours of walking. Good use of technology can significantly enhance quality of life. Similarly, we can expect that data science will be able to enhance quality of life in the developing world, as long as we work to bridge the digital divide. Examples of work to bridge the divide in health and care in Wales are discussed in [11]. At the time this report was written 85% of people in Wales used the internet; hence 15% did not. Case study exemplars reported range from the use of tablets and virtual reality technology for people in care homes to engage with the wider world, to provision of free public Wi-Fi to remove a barrier to access. That report makes 18 recommendations around the issue of digital inclusion.

The importance of bridging the divide is highlighted by well publicised examples of superhuman AI performance. For example, Deep Mind's use of reinforcement learning with game AI and Alpha Go has highlighted the potential of that technology to solve complicated problems. Research into developing models that can learn from solving a sequence of problems is ongoing. Such algorithms are able to benefit from past experience, in analogy with the way that humans and animals learn new skills and then improve upon them. While the goal of a general AI remains a long way off, we already see decision-making algorithms that are able to play different genres of computer game. While in favour, AI is not the only data science method that is used for predicting outcomes or to govern decision making. Care needs to be taken to understand the implications of using a given technology for an application. As a technology matures and moves into broader applications it will bring new opportunities and new challenges to be addressed. Opportunities can be enhanced by incorporating the necessary expert causal knowledge with data into AI systems. Here great care should be taken to consider the ethical ramifications of using such systems. The steps required to mitigate marginalising parts of society will constantly evolve in step with the evolution of such technology.

Recommendation: Establish methods to ensure poverty-stricken regions of the planet can access technology to avoid impossible to overcome barriers to wealth.

As in previous industrial revolutions, the data science revolution that is taking place will transform the energy footprint of humanity. This raises sustainability concerns especially linked to the use of some technologies that are resource intensive. For example, deep learning is a significant step forward in technology, however it is computationally expensive to deploy. From an ecological perspective one should deploy such technology only when there is a clear benefit of doing so, relative to less computationally expensive options; i.e. avoid adopting an approach of always using a sledgehammer to crack the proverbial nut. This is a general observation that applies broadly to all applications of data science. Climate change is currently the biggest challenge facing humanity. For example, this will lead to the marginalisation of many island states as they start to disappear as seas rise. The onset of climate change as a result of widespread use of computationally expensive methods not powered by renewable sources of energy would exasperate the impending humanitarian disasters that are anticipated. Minimising the carbon footprint of data processing is a key issue to focus on.

Recommendation: Embed sustainability considerations in the design of data science applications to ensure that methods use the minimum amount of resource to address a given issue.

One particular issue of concern that could be considered under the remit of this call is in the form of understanding unintended or malign consequences of data science including AI. As with the Manhattan project that led to the development of nuclear weapons, and ultimately provided windows to benign applications of that technology, progress in data science has been and will continue to be disruptive. Technology developed for a particular application or domain context can be expected to have dual potential uses. With data science it is possible that benign developments for one domain can be malignant when applied to other domains. An example is illustrated with the recent Cambridge Analytica data episode [12]. While we may not yet need such a mechanism today, there may soon come a time when the global community finds it desirable to establish international protocols for oversight of this technology. Global disruption through data breaches and malware attacks indicate that the concern is not just constrained to government level activity, but also includes a criminal element.

Recommendation: Stakeholders should be mindful of potential misuses or destructive applications of their technologies and develop strategies for constraining dual use technologies.

Improving stakeholder cooperation

The need to legislate the right to be forgotten in the EU highlights the fact that while mechanisms such as the UDHR are in place, they are not always acknowledged. When there is a lack of mutual understanding, ethical practice, or willingness to comply with an existing mechanism, then legislation forcing entities to conform will follow. Stakeholders with a vested interest in any given aspect of data cooperation need to have a forum to engage in dialogue around technology. This will help all stakeholders understand developing issues and contextualise those with respect to existing mechanisms. We propose that a global Data Cooperation Forum is established as a means of enabling dialogue between stakeholders both working in and using data science. This forum should be sufficiently frequent in order to minimise the lag between the data science innovation and dialogue on policy surrounding the deployment of that technology. Given the pace of developments this forum could meet on an annual or bi-annual basis.

Recommendation*: Establish a world forum on Data Cooperation as a place for all concerned stakeholders to engage in dialogue and explore current and emerging issues.

In the UK the Hall-Pesenti report [13] has recommended the creation of Data Trusts, a proposal that has been welcomed by the UK Government [14]. The purpose of the Data Trust is to provide a safe mechanism for the exchange of data sets.

The UK Government has recommended that where appropriate, publicly funded projects, especially where the data are not sensitive, engage with SMEs to bridge the gap of providing large data sets for processing so that algorithms and workflows may be developed. Data Trusts could help SMEs and other small stakeholders bridge the digital divide [14]. This concept could be extended to provide a global mechanism to provide access to large non-sensitive data sets suitable for SMEs to develop their algorithms with a minimal entry barrier for access. This would help to level the playing field between large dominant organisations and small entities attempting to establish emerging technologies or bring them to market.

Recommendation: Extend the Data Trust concept to be a global mechanism.

Over the past year the Carnegie UK Trust has considered the issue of internet harm reduction [15]. In their report they remark that the case for regulation has increased during 2018. They make a series of recommendations surrounding this complicated issue. Among other things they propose "the regulators should have the power to draw up codes of practice with industry and civil society or to approve already existing codes" and that a super complaints mechanism be introduced, to empower individual right of action. Broader implications for conduct are also discussed in [15].

Data pervades many aspects of modern life around the world. Our existence is entwined with algorithms. Understanding of data and the implications of its use should, from an ethical perspective, be part of basic education received. However, most people entering the work force in this area have no formal training in ethics. We recommend that where appropriate educational institutions should not only educate their students in data science but also provide appropriate ethical training to help establish a deeper understanding of data use and its implications. This will help mitigate stakeholder marginalisation. The pace of technological development means that education should not be restricted to any particular age group as the global workforce also has need

to better understand the use of data science and its implications. When developing educational resources care should be taken to choose exemplars that do not exclude or otherwise marginalise, parts of the population.

Recommendation: Where possible, data science should be part of basic education, and this should include the consideration of ethical issues.

Transparency surrounding data collection, the creation and use and algorithms has the potential to alienate organisations using data from the public at large. It is important that data science and data analytics practitioners work to ensure transparency in all aspects of their use of data and algorithms. This will help those organisations engage with society, and where applicable, their customer base, to communicate their work openly and without prejudice. This sentiment parallels the "ethics by design" aspiration noted by the European Commission in [7].

Recommendation: The use of data science should be accompanied by a commitment to transparency.

The proposed Data Cooperation Forum noted in Recommendation* could be used to understand how best to implement the remainder of these recommendations.

SUMMARY

The following is a summary of our recommendations:

1. **UDHR should be reaffirmed in the context of the digital age, so that online and offline rights of individuals are respected equally.**
2. **Explore establishing responsible partnerships. There are industry sectors who are seen as responsible partners (see World Bank/WHO guidelines [6] on such partnerships).**
3. **Ethical considerations should always be applied to technology, in terms of design and use. Organisations should drive the ethical review of their use of technology.**
4. **Establish methods to ensure poverty-stricken regions of the planet can access technology to avoid impossible to overcome barriers to wealth.**
5. **Embed sustainability considerations in the design of data science applications to ensure that methods use the minimum amount of resource to address a given issue.**
6. **Stakeholders should be mindful of potential misuses or destructive applications of their technologies and develop strategies for constraining dual use technologies.**
7. **Establish a world forum on Data Cooperation as a place for all concerned stakeholders to engage in dialogue and explore current and emerging issues.**
8. **Extend the Data Trust concept to be a global mechanism.**
9. **Where possible, data science should be part of basic education, and this should include the consideration of ethical issues.**
10. **The use of data science should be accompanied by a commitment to transparency.**

References

[1] General Data Protection Regulation outline

[2] UK Government Guidance: Data Ethics Framework

[3] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

[4] http://www.un.org/en/universal-declaration-human-rights/

[5] People, Power and Technology: The 2018 Digital Understanding Report, https://doteveryone.org.uk

[6] World Health Organisation guidelines on Public-private partnerships (PPPs)

[7] EU Coordinated Plan on Artificial Intelligence COM (2018) 795 final

[8] Ethics Advisory Report for West Midlands Police (2017)

[9] On line Data Privacy from Attitudes to Action: an evidence review (2018)

[10] Weapons of Math Destruction, Cathy O'Neil, Penguin (2017)

[11] Digital Inclusion in Health and Care in Wales (2018)

[12] Facebook–Cambridge Analytica data scandal

[13] Growing the Artificial Intelligence Industry in the UK, Professor Dame Wendy Hall and Jérôme Pesenti

[14] Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able? CM 9645 (2018)

[15] Carnegie UK Trust, Internet harm reduction (2019)