THE ROYAL SOCIETY

14 May 2018

Submission to the House of Lords Communications Committee inquiry: The Internet: To Regulate or Not to Regulate?

Summary:

The Royal Society's response:

- Strongly resists a one-size fits all approach to governance of data and its uses. The internet relies on data-enabled technologies to operate. While there are governance challenges that are general in nature, many of them – and their effects – are likely to be specific. For example, the use of data to create personal recommendations for online shopping creates different forms of benefit and risk and involves different types of actors compared to the use of data in online healthcare applications. It would be a mistake to attempt to govern them in the same way.
- Calls for a renewed governance framework for data use to ensure trustworthiness and trust in the management and use of data as a whole. Central to this framework is the overarching principle of human flourishing which reflects the fundamental tenet that society does not serve data but that data should be used to serve human communities. With this overarching principle, this governance framework should be underpinned by a set of high-level principles. All systems of data governance should:
  - Protect individual and collective rights and interests.
  - Ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively.
  - Seek out good practices and learn from success and failure.
  - Enhance existing democratic governance.
- Outlines the need for a stewardship body which would be expected to conduct inclusive dialogue and expert investigation into novel questions and issues, such as those related to the internet, and to enable new ways to anticipate the future consequences of today's decisions.
- Warns against the regulation of machine learning algorithms specifically and advocates a more tailored sector specific approach to regulation.
- Outlines a series of challenges and tensions which must be considered as the capability and prevalence of data driven technologies increases including:
  - Concepts of data governance which are under strain.
  - Balancing the benefits of tailored services and consumer convenience with risks to autonomy.
  - Encouraging innovation while maintaining public confidence and addressing societal needs.

The Royal Society would welcome the opportunity to discuss these issues further with the Committee.

Introduction:

0.1. The Royal Society is the UK's national academy of science. It is a self-governing Fellowship of many of the world's most distinguished scientists working in academia, charities, industry and public service. Its fundamental purpose is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

0.2. The Society's Data Programme is developing policy and promoting debate that helps the UK safely and rapidly realise the growing benefits of data science and digital technologies. This programme brings together leading academics, industry leaders, civil society and data and technology

specialists to better understand the needs of a 21st century data governance system and the challenges associated with changes in data use and society.

0.3. In this response we highlight relevant findings from our work which are pertinent to the Committee's questions, including whether regulation of the internet is desirable or possible, transparency in the use of algorithms, and identify some additional challenges the committee may wish to consider. This draws on previous work of note:

- o In 2017, the Royal Society published their report '*Machine learning: the power and promise of computers that learn by example',* setting out the potential of machine learning over the next five to ten years, and the actions necessary to allow society to benefit fully from the development of this technology.
- o In 2016, the Royal Society and Ipsos Mori completed a public dialogue exercise on machine learning; gaining insights into public knowledge of, and attitudes towards, machine learning.
- o In 2017, the Society collaborated with the British Academy to publish '*Data management and use: Governance in the 21st century'*; highlighting the challenge and existing tensions with data use and the needs for a 21st century governance system.

1. An evolving technological landscape

1.1. An IBM report in 2017 estimated that at the time around 90% of data in the world had been created in the last two years[1]. Data collection activities continue to increase in speed, scale and variety, with the expansion of internet access, applications and capabilities playing a central role in this.

1.2. As the analytic techniques used to process these datasets become more sophisticated, individuals and communities are affected in new and unexpected ways. Fascinating new forms of data analysis such as machine learning have vastly increased the ability to link this data and use the patterns that emerge. Machine learning algorithms are already deployed in a range of systems or situations which shape daily life and use of the internet. Whether it be by detecting instances of credit card fraud, providing online retail recommendations, or supporting search engine functions.

1.3. Uses of data-enabled technologies promise further benefits, from improving healthcare and treatment discovery, to better managing critical infrastructure such as transport and energy. However, history has provided rich illustrations of how the widespread adoption of new technologies without effective public engagement can increase public anxiety, or result in major public controversy, both of which risk hampering potential benefits.

1.4. Further, changes to how data is used and analysed places existing data governance concepts, such as privacy, ownership and consent, under unprecedented strain. Their meanings in policy, law and public discourse have shifted, and will continue to do so in new and unpredictable ways. Uncertainties are accumulating and acting on them is necessary. However, in order to avoid long-term, cumulative and difficult-to-foresee effects, any action must be carefully considered.

2. Context is key – Avoiding a one-size-fits-all approach to data governance

2.1. The internet relies on data-enabled technologies to operate. The Royal Society strongly resists a one-size fits all approach to governance of data and its uses. While there are governance challenges that are general in nature, many of them – and their effects – are likely to be specific. For example, the use of data to create personal recommendations for online shopping creates different forms of benefit and risk and involves different types of actors compared to the use of data in online healthcare applications. It would be a mistake to attempt to govern them in the same way.

2.2. While there may be specific questions about the use of personal data and machine learning algorithms in internet based applications or platforms, these should be handled in a context-specific way, rather than via overarching regulation for all uses.

---

[1] https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

2.3. Practically it would be impossible and also undesirable to try to centralise data governance. Such an approach may inhibit or prevent perfectly reasonable technological developments which would enjoy public support and benefit society. In many cases there are already sector specific regulations that applications should conform to, and which are more appropriate than a one-size-fits-all approach.

2.4. The Royal Society's public dialogue on machine learning highlighted that the nature or extent of public concerns about machine learning and algorithms are linked to the application being considered. Fundamentally, the issues raised in these public dialogues related less to whether machine learning technology should be implemented, but how best to exploit it for the public good. Such judgements were made more easily in terms of specific applications, than in terms of broad, abstract principles, reinforcing the case for a context specific approach.

3. Principles for data governance in the 21st century

3.1. The internet is 'powered' by data, but this is just part of a rapidly changing and evolving data landscape, where big data technologies require us to develop new ways to use and manage data for both online and offline applications. Societies must navigate significant choices and dilemmas: they must consider who reaps the most benefit from capturing, analysing and acting on different types of data, and who bears the most risk.

3.2. While a one size-fits-all approach to data governance is undesirable, governance surrounding the use of data does require a new framework and principled approach to keep pace with the challenges in the 21st century. To ensure the extraordinary opportunities for a data enabled society are realised and that public trust is built, the Royal Society recommends two types of high level responses to data governance specifically.

3.3. First, a renewed governance framework is needed to ensure trustworthiness and trust in the management and use of data as a whole. Central to this framework is the overarching principle of human flourishing. This principle reflects the fundamental tenet that society does not serve data but that data should be used to serve human communities. With this overarching principle, this governance framework should be underpinned by a set of high-level principles. All systems of data governance should:

- Protect individual and collective rights and interests.
- Ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively.
- Seek out good practices and learn from success and failure.
- Enhance existing democratic governance.

3.4. Second, despite the range of actors already carrying out important governance functions in their specific sectors or domains, there is a clear need for a new body to steward the evolution of the data governance landscape as a whole, and to ensure human flourishing. This stewardship body would be expected to conduct inclusive dialogue and expert investigation into novel questions and issues, such as those related to the internet, and to enable new ways to anticipate the future consequences of today's decisions.

3.5. These calls were recognised in the 2017 Budget and Industrial Strategy, where the government outlined plans to create a Centre for Data Ethics and Innovation to enable and ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies. The Society is pleased to see this is in line with our recommendations and welcomes the opportunity to work closely with government in this regard. The newly created Ada Lovelace Institute, established by the Nuffield Foundation to examine ethical and social issues arising from the use of data, algorithms, and artificial intelligence, will also play an important role in ensuring that new technologies can be developed in the way that public want, that exemplifies good practice, and that will allow everyone to benefit.

3.6. As the inquiry notes, there are already a number of public and private actors that regulate activity related to the internet and the use of personal data. The Information Commissioners Office considers data protection and privacy across different sectors, and actors like Ofcom regulate content from streaming services. There are also strong legal structures in place which will be built upon by the Data Protection Bill, currently making its way through Parliament, and the introduction of the General Data Protection Regulation (GDPR), protecting the processing and use of personal data.

3.7. Ensuring that these bodies work collaboratively will be key to the development and preservation of an effective governance framework which enjoys public confidence.

4. Future challenges

4.1. Many of the choices that society will need to make as data-enabled technologies become more widely adopted can be thought of as a series of pervasive tensions, which illustrate the kinds of dilemmas that society will need to navigate. As data enhances our analytical capabilities, notions such as accountability, agency, consent, privacy and ownership are becoming more difficult to maintain. Their meanings in policy, law and public discourse have shifted, and will continue to do so in new and unpredictable ways. As a result, many of the concepts that sit at the core of public confidence in governance are no longer fit for purpose.

4.2. This section outlines a series of tensions and challenges the Committee may wish to consider in the context of regulation, and how it is applied to the internet:

4.3. Existing concepts of data governance under strain. - Consent, ownership, privacy (transparency paradox)
   i. Consent is one of the legal grounds for processing personally identifiable data in the current data protection regime. However, genuine consent is difficult to achieve, and is often not sufficient to ensure adequate protection of individuals' interests. First, the application of consent suffers from what is often referred to as the 'transparency paradox'. Consent requires transparency of what is being consented to. Such transparency has to be meaningful, and the mere disclosure of information is not enough. Anything too long or complex is unlikely to be broadly understood or read yet making a summary widely comprehensible often discards the details that people care about. For example, in the acceptance of terms and conditions when using internet applications or platforms. Second, it is unreasonable to expect an individual to keep track of what data is collected about them and understand how it will be used, and therefore to give meaningful, informed consent.
   ii. Privacy is a deeply complicated, context-specific and multi-layered notion and its different aspects are often conflated. Data is also now often collected without explicit knowledge. It may be gathered from spheres previously thought of as private and combined with other datasets to reveal sensitive or identifiable information. The notion of privacy is also being stress-tested through the increased power of algorithms and their ability to infer and predict behaviour. The ability to draw connections between data is now so advanced that approaches to managing privacy, such as deidentification, may no longer apply. Meanwhile, the balance of risks and benefits to the citizen may play out differently in different contexts, muddying the waters with regards to what constitutes acceptable or unacceptable data use.
   iii. Questions about consent are further complicated by how ownership of different data types is perceived. Data is often co-created and is capable of being silently captured, easily replicated, radically transformed, and cheaply transferred. This bears little resemblance to ownership in the way that one might own a house or a car.

4.4. Tailored public and commercial services vs risks to autonomy
   i. Data-enabled technologies and machine learning applications have made it possible for users to receive a tailored online experience. For example, online music platforms can provide suggested songs or playlists, while online shops will often highlight suggested purchases based

on user's previous activity. Such developments raise a question over where the line is drawn between a tailored service and a risk to an individual's autonomy.

ii. An example of where these concepts have been blurred is in the current controversy surrounding the use of data analytics in political campaigning to target specific groups or areas to the exclusion of others. The potential for personalisation comes with benefits as well as risks to autonomy. It is possible to narrowly target products and services, making it easier for an individual to seek out more suitable services and products. However, in some cases these benefits come with the risks of undesirable statistical stereotyping and profiling. This has an effect on an individual level, which could be restrictive to the way individuals engage in the world around them.

iii. It is also worth noting that statistical profiling is already in use in marketing, insurance, and assessment of threats or policing, so the need to carefully balance manages biases in data is not in itself new.

4.5. Encouraging innovation while maintain public confidence and addressing societal needs

i. Innovative uses of data offer great potential for the UK economy. It is estimated that £66[2] billion of business and innovation opportunities could be generated through effective use of data. To keep step with the pace of change and remain competitive, innovation should be encouraged, and guided so that it addresses societal needs. However, as data-enabled technologies have increasingly large and uncertain social, economic and ethical consequences, getting the balance right will be critical.

ii. Strategic consideration should also be given to the right long-term approach to maximising value from entrepreneurial activity in this space. On the one hand, the recent acquisitions of DeepMind, VocalIQ, Swiftkey, and Magic Pony, by Google, Apple, Microsoft, and Twitter respectively, point to the success of UK start-ups in this sector. On the other, they reinforce the sense that the UK environment and investor expectations encourage the sale of technologies and technology companies before they have reached their full potential. In the case of machine learning, in order to meet the demand across industry sectors, the UK's Industrial Strategy will need to support an active machine learning sector that capitalises on the UK's strengths in this area and its relative international competitive advantages.

iii. With the dominance of a small number of online platforms, creating appropriate mechanisms to apportion value will be a social and technical challenge and one that needs to consider how to balance asymmetries of power between different actors and platforms.

For further information, please contact Becky Purvis, Head of public affairs, becky.purvis@royalsociety.org

---

[2] Parris S et al. 2016 Digital Catapult and productivity: A framework for productivity growth from sharing closed data. Cambridge UK: Rand Corporation. See http://www.rand.org/pubs/research_reports/RR1284.html