CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

RESPONSE TO THE CONSULTATION ON THE NATIONAL DATA STRATEGY

November 2020

1. This response to the National Data Strategy (NDS) – which is an ambitious and extensive framework – is limited to its intersection with the proposed Online Harms framework, the area in which we have most expertise but which is only referenced in passing in the NDS.

2. We note the Secretary of State's enthusiastic, pro-tech and pro-growth foreword which repeats the narrative first set out in the UK's 2017 Digital Strategy with its ambition for the UK to be both the best place to start a digital business and the safest place in the world to go online:

   "The strategy is a central part of the government's wider ambition for a thriving, fast-growing digital sector in the UK, underpinned by public trust. We want the UK to be a nation of digital entrepreneurs, innovators and investors, the best place in the world to start and grow a digital business, as well as the safest place in the world to go online. We will set out more on how we propose to support a digital drive for growth in our Digital Strategy, which we will be publishing in the Autumn."
   (Rt Hon Oliver Dowden, NDS Foreword).

3. The Government's repeated commitment to make the UK the safest place in the world to go online has not been progressed since it was first made in the previous Digital Strategy. There is no mention in the NDS on how issues like fairness, transparency, governance and public trust in the use of data will deliver this commitment to online safety, especially given the problematic use of data by some actors in the online ecosystem. At the time of writing, we still await the final Online Harms proposals from DCMS. With the exception of the announcement of the £2.6m pilot project to use AI to detect harms, and passing references to the Competition and Markets Authority's major piece of work on Digital Markets, there is no significant consideration of how the NDS can help underpin, or intersect with, the Government's "world-leading" Online Harms policy or legislation, which has been in development in the same Department for three years. This is a concern.

Background to our work

4. Our work on a statutory duty of care – developed over the past two years and published in the form of blogs, articles and papers[1] – has informed the Government's proposals in its Online Harms White Paper. We would commend our full paper, published in April 2019, to the NDS team for reference.[2] With particular relevance to stimulating economic growth and innovation while protecting consumers, citizens and vulnerable users, we also would refer the NDS team to our correspondence with the then

[1] https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

[2] See our full detailed paper (April 2019): https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

DCMS Secretary of State in December 2019 on how the UK might develop a coherent approach to digital regulation[3] and our recent blog on “regulatory interlock”, which will assist the joining up of regulatory oversight across a wide range of online and data-driven harms and enable Ofcom (if confirmed as the Online Harms regulator) to address these through the Online Harms regulatory framework[4].

5. Our proposed regulatory regime – a statutory duty of care enforced by an independent regulator - is intended to reduce harm to people from social media and other online services. Our approach is systemic, rather than content-based, preventative rather than palliative. While our focus is primarily on the reduction of harm caused by the design and operation of social media and other “public” online spaces (eg where users interact with one another, or where content can be targeted at individuals), its principles apply as much to the design of all data-driven services that use data and algorithms to serve up content to individuals or to make decisions on what content or services they are offered online.

6. At the heart of our regime is a duty of care set out in Parliament in statute, which would require most companies that provide social media or online messaging services in the UK to protect people in the UK from reasonably foreseeable harms that might arise from use of these services. The approach is risk-based and outcomes-focused. A regulator would have sufficient powers to ensure that companies delivered on their statutory duty of care.

Health and Safety at Work Act 1974 applies to software

7. While we were developing our approach to a statutory duty of care for online harms we began to wonder whether the Health and Safety at Work Act 1974 applied to software supplied for use at work. We worked with Lord Stevenson who asked the government a question in Parliament. Baroness Buscombe replied unequivocally:

> ‘Section 6 of the Health and Safety at Work etc. Act 1974 places duties on any person who designs, manufacturers, imports or supplies any article for use at work to ensure that it will be safe and without risks to health, which applies to artificial intelligence and machine learning software. Section 6(1)(b) requires such testing and examination as may be necessary to ensure that any article for use at work is safe and without risks but does not specify specific testing regimes. It is for the designer, manufacturer, importer or supplier to develop tests that are sufficient to demonstrate that their product is safe.’[5]

8. Section 6 of the Act requires testing of items supplied for use at work by the person supplying them to ensure the items are safe. The broad intention seems to be that a business owner might buy something that performs a task in their production process (in 1974 this would likely have been a complex machine) but the purchaser isn’t competent to assess in full how the item operates. So the burden falls upon the maker of the item or the importer to certify that it is safe through testing. The act is drafted in general terms to future proof its effect. The HSE notes that ‘Section 6(1) of the HSW Act places a general health and safety obligation on anyone in the supply chain’.[6]

[3] https://www.carnegieuktrust.org.uk/news/draft-online-harm-bill-press/

[4] https://www.carnegieuktrust.org.uk/blog/online-harms-interlocking-regulation/

[5] See Lords Hansard - Industrial Health and Safety: Artificial Intelligence: Written question - HL8200 https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Lords/2018-05-23/HL8200/

[6] https://www.hse.gov.uk/work-equipment-machinery/uk-law-design-supply-products.htm]

9. The analogies with so-called AI or ML software with many claims seem strong – indeed the 'black box' nature of such software is seen by some as a feature not a flaw. Suppliers of software to government should be asked to assess whether it is safe to use and to certify that as part of the procurement process. In this context safety would mean that there is low risk of a reasonably foreseeable harm – such as the use of the software not compromising a breach of someone's rights under the Equalities Act. Government's role as a buyer will help drive safety certification for AI software through the supply chain. Rigorous certification would lead to better selection of training data and study of the outcomes of the software's operation.

10. One method of managing risk would be for the government to work with industry to provide external standards and procedures with which to manage risk of harm from using third party software of uncertain pedigree to manipulate citizen data on a large scale.

Applying a duty of care to online services

11. Everything that happens on a social media or messaging service is a result of corporate decisions: about the terms of service, the software deployed and the resources put into enforcing the terms of service and maintaining the software. Decisions on data use, online targeting and algorithmic decision-making lie at the heart of the design of these services. But these design choices are not neutral: they may encourage or discourage certain behaviours by the users of the service.

12. Substantive indicative evidence of online harm is emerging from advocacy groups and from regulatory bodies such as the CMA, Ofcom and ICO[7]. We have worked closely with the Centre for Data Ethics and Innovation, who recommended in their review that data-driven online targeting be integrated in the Online Harms "duty of care".[8] Data-driven targeting can be a source of consumer detriment[9] while the use of recommender algorithms is a means to rapidly spread illegal or harmful material (such as extremist content, hate speech, child sexual abuse imagery, disinformation).[10]

13. We noted in our submission to the CDEI's consultation on Online Targeting that, despite these emerging concerns, there are recurrent arguments for not regulating social media and online companies because they are in some way "unique" or "special: a complex fast-moving area where traditional regulatory approaches will be blunt instruments that stifle innovation and required platform operators to take on the role of police and/or censors. Another is that the technology is so new, sufficient evidence has not yet been gathered to provide a reliable foundation for legislation. CDEI was set up to build up an evidence base to inform policymaking and regulatory decision-making, while protecting and encouraging innovation in the digital sphere. We are disappointed that the NDS does not draw on its work more explicitly in its consideration of how data-driven online targeting might be applied, particularly in a public sector context

[7] https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/online-nation; https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study

[8] https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations

[9] https://www.which.co.uk/policy/digitisation/2659/control-alt-or-delete-the-future-of-consumer-data-main-report

[10] Zeynep Tufecki, 'You Tube, the Great Radicalizer', New York Times, 10 March 2018, available: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html?smid=plshare; anthropological research suggests that those coding recommender algorithms see their function as 'hooking' users; that these algorithms operate as a trap: N. Seaver, 'Captivating algorithms: Recommender systems as traps' (2018) Journal of Material Culture, available: https://journals. sagepub.com/doi/10.1177/1359183518820366]; DCMS Select Committee report on Disinformation and Fake News: https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/

14. As we argue strongly throughout our work, the traditional approach of not regulating innovative technologies needs to be balanced with acting where there is good evidence of harm. We have found the precautionary principle[11] a particularly useful framework to consider when Governments and their advisers need to enable potentially hazardous commercial activity to proceed relatively safely but also require a bulwark against short term political attempts to ban things in the face of moral panic.

15. In data-driven services that use targeting or algorithmic decision-making, just as in social media networks and platforms, the harms may be multiple, and may be context- or service- specific, while the speed of their proliferation makes it difficult for policymakers to amass the usual standard of long-term objective evidence to support the case for regulatory interventions. The software that drives online services is updated frequently and the vast majority of changes are invisible to most users. Tweaks to the software that companies use to decide which content to present to users may not be discernible. Features visible to users change regularly. External researchers cannot access sufficient information about the user experience on a service to perform long term research on service use and harm.

16. Evidencing harm in this unstable and opaque environment is challenging, traditional long-term randomised control trials to observe the effect of aspects of the service on users or others are nearly impossible without deep cooperation from a service provider. We believe that, as with harms relating to social media, there is "good reason to believe that harmful effects may occur to human[s]", as the ILGRA put it, despite the uncertainties surrounding causation and risk.

17. We therefore see our duty of care proposals – risk-based, outcome-orientated and operating at a system level – as a suitable regulatory approach for the minimisation of harm that may arise to online users from targeting or algorithmic decision-making. Under this approach, a straightforward risk assessment should tell the owners and operators of any given service whether the design and application of the algorithms they deploy or the method of targeting users with content or services are likely to cause reasonably foreseeable harm or detriment to those users, particularly vulnerable groups. The NDS provides no detail, in its section on "Creating a Fairer Society For All", on how such basic risk assessment principles should be baked into this ambition.

18. For example, while it is true that "data-driven online profiling technologies can help identify potentially vulnerable web users (such as people suffering from gambling addiction), and target support or prevent them from seeing potentially harmful content", a truly systemic, fair and equitable approach to harm reduction would require service providers to undertake risk assessment on how the design and systems of its service might reduce the risk of foreseeable harm to all vulnerable users. We are concerned that the focus on this example of how data-driven online profiling might reduce harm – without reference to the wider debates on Online Harms and the ethical and privacy considerations that arise from this approach – suggests an unbalanced and ultimately detrimental understanding of the nature and prevalence of existing online targeting, which should be addressed first through risk-based, proportionate regulation.

19. Within a duty of care regulatory regime, the regulator would seek evidence of, and take into account the impact of, the steps taken by service providers to reduce harm – in this example, both through detrimental online targeting as well as "for good" targeting to reduce individual risk. This is not a one-off action but an ongoing, flexible and future-proofed responsibility that can be applied effectively to fast-moving technologies and rapidly emerging new services in all data-driven sectors.

[11] https://webarchive.nationalarchives.gov.uk/20190701152341/ https://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm

20. On one specific point relating to online targeting, we would refer the NDS team to the February 2019 declaration by the Council of Ministers of the Council of Europe on the manipulative capabilities of algorithmic processes:

> ‘Data-driven technologies and systems are designed to continuously achieve optimum solutions within the given parameters specified by their developers. When operating at scale, such optimisation processes inevitably prioritise certain values over others, thereby shaping the contexts and environments in which individuals, users and non-users alike, process information and make their decisions. This reconfiguration of environments may be beneficial for some individuals and groups while detrimental to others, which raises serious questions about the resulting distributional outcomes…. The Council of Ministers encourages member States to assume their responsibility to address this threat by [inter alia] considering the need for additional protective frameworks related to data that go beyond current notions of personal data protection and privacy and address the significant impacts of the targeted use of data on societies and on the exercise of human rights more broadly…. taking appropriate and proportionate measures to ensure that effective legal guarantees are in place against such forms of illegitimate interference.’[12]

While not legally binding, such declarations are sometimes used as a guide to interpretation by the European Court of Human Rights.

21. Finally, we note that there is a specific proposal relating to investment in a pilot project to address Online Harms:

‘to improve systems for detecting and addressing online harms, the government is launching a £2.6m programme that will help companies to develop AI-based solutions to tackle these issues ever more effectively .. we will review and upgrade the data infrastructure that underpins the monitoring and reporting of online harms such as child sexual abuse, hate speech and self-harm and suicide ideation.'

22. Without further detail on this proposal, the problem it seeks to solve or the involvement to date of civil society and social media companies in its design – and without it being in any way apparently connected to the wider Online Harms policy development or the draft Codes of Practice relating to child sexual abuse, which we understand the Government will publish soon – we are unable to comment on whether this is an appropriate response. We would ask that expert groups – from academics through to civil society organisations with a track record in expertise in relation to Online Harms – are fully involved in the development of this pilot to ensure that it does, in itself, not inadvertently cause undermine the protection of vulnerable users at risk of the harms identified.

23. In addition to our extensive work on online harms, the Trust has also delivered a small number of other data-related projects, with insights pertinent to this consultation. Overall, we support the ambition for greater accountability and transparency of data use. However our work with Involve and Understanding Patient Data has shown clear challenges in how data is shared both between and within existing public services, with no clear framework for assessing the level of public benefit and risk such sharing would create[13]. We proposed an 18-question framework to help services make better decisions about when data should and shouldn’t be shared, with three tests (purposeful, proportionate and responsible) necessary for public service providers to gain the social licence to share and use data more widely. This framework is intended help professionals weigh up the purpose of

[12] Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 Decl (13/02/2019) 1 https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

[13] https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2018/04/11143400/Data-Public-Benefit-FINAL-REPORT.pdf

sharing data against the potential for harm and help public service providers have conversations with the public about data sharing.

24. Similarly, there needs to be an equally critical look at the role and implementation of predictive analytics and automation in public service decision making, specifically prior to systems going live and being used with the public. While we are seeing a rise in national and international interest around the use of predictive analytics, machine learning and automated systems, the available evidence suggests that there has so far been a mixed responses from government in the UK in uses of algorithmic and automated systems in public services. Some departments and agencies have implemented these programmes, some are piloting them, and others have cancelled the use of these systems altogether. Our forthcoming work with the Data Justice Lab[14], analyses over 50 cancelled systems in the UK and internationally. It demonstrates a breadth of factors influencing decisions to pause or cancel systems including: legal challenges; civil society critique; concern about privacy, fairness, bias, discrimination; political intervention; and the systems sampling not working well or meeting expectations.

25. In terms of public services more broadly, there is a range of opportunities that can support the public to become more data literate[15]. For example, public libraries, as safe spaces in communities that enable access to information, knowledge and culture, have a clear role to play when it comes to data privacy, if they are given the right support and resources to do so.

Carnegie UK Trust
November 2020
Contact: maeve.walsh@carnegieuk.org

[14] https://www.carnegieuktrust.org.uk/project/automating-public-services/

[15] https://www.carnegieuktrust.org.uk/publications/leading-the-way-a-guide-to-privacy-for-public-library-staff/