Researchers from the Human Centred Computing Research Group
and the Responsible Technology Institute
Department of Computer Science
University of Oxford

19 November 2021

Response to the public consultation on reforms to the UK's data protection regime by the Department for Digital, Culture, Media & Sport

This document provides the response of researchers within the University of Oxford's *Human Centered Computing* Research Group and the *Responsible Technology Institute* to the Department for Digital, Culture, Media, and Sport's consultation on data protection, privacy, and responsible AI development.[1]

Our ability to speak credibly to these issues comes from our various projects working towards building a better future of digital infrastructures and grounded responsible innovation. Specifically, our researchers are spearheading the Oxford Martin School's Ethical Web and Data Architectures (EWADA) project, which explores the vital themes of Data Privacy and Algorithm Accountability; and from the Responsible Technology Institute, the EPSRC RoboTIPS and EPSRC Trusted Autonomous Systems RoAD projects, which address the social impact of responsible robotics in the digital economy and automated vehicle data. Additionally, our work has close ties to the Open Data Institute and the Institute for Ethics in AI grounding our expertise in issues directly tied to the real-world implementation of advanced algorithms at scale.

This response is based on our direct experiences with managing datasets of personal data in research and industry across a diverse variety of user groups; our technical knowledge base about the design and development of algorithms interacting with users throughout their

[1] Oxford Human Centered Computing: https://www.cs.ox.ac.uk/research/HCC/; Responsible Technology Institute: http://www.cs.ox.ac.uk/projects/RTI/

daily lives; and our research into the diverse social and legal impacts of data collection and use.

Our response highlights certain concerns found within the consultation that touch on the core issues of data protection and privacy, responsible research, and balancing support for organisations as well as individuals. The major themes of our response cover the following topics:

- **Data intermediaries and institutions:** Lack of clarity regarding data intermediaries, institutions, and practices put in place to safeguard individuals and support technological growth.
- **AI and responsible innovation:** The opportunities for AI innovation in the UK depend on a robust regulatory regime that encourages highly context-specific risk management. This will be best promoted through maintaining existing measures like Data Protection Impact Assessments, Data Protection Officers, record keeping, and prior consultation, amongst others.
- **Erosion of trust in online tracking:** Excessive box-ticking in the form of consent banners is not a necessary feature of existing data protection and privacy law, but rather a symptom of non-compliance with it.
- **Removal of the balancing test:** The removal of the balancing test for pre-approved legitimate interest purposes will create disproportionate risks for UK citizens, and a false sense of certainty for controllers.

We provide recommendations and suggestions on these themes and statements, intended to help build a sustainable future of AI and data protection within the UK that not only promotes innovation but also advocates for and protects individuals.

Summary of Recommendations

Data intermediaries and institutions:

1. Reinforce and expand data subjects' ability to access data by enabling collective access requests through representative intermediaries (e.g. NGOs and trade unions).

2. Create processes for *certification of data intermediaries* according to fiduciary duty interoperable with data altruism principles in the EU Data Governance Act and make clear what standards they will be subject to following beyond the existing requirements for processors.[2]

3. Empower *certified data intermediaries* with protections and specific standards to act on behalf of groups of data subjects.

4. Uphold the principles of *data minimisation* called for by the ICO in response to the consultation, and explicitly define the ICO's role with respect to the regulation of data intermediaries.[3]

As researchers actively engaged in the development of data intermediary infrastructure via the EWADA project[4], in which we investigate a new decentralised paradigm for data sharing and ownership, we foresee data intermediaries playing an integral role in promoting the agency of individual data subjects and building innovative new technologies and business models. In the era of AI, control over personal and behavioral data is tremendously powerful. We believe institutions like data intermediaries will be critical to guarantee subjects are able to maintain control over their data and how it is used. Data intermediaries should enhance

[2] European Commission. (2020). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European Data Governance. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

[3] Information Commissioner's Office. (6 October, 2021) Response to DCMS consultation: Data a New Direction. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2021/10/response-to-dcms-consultation-foreword/

[4] University of Oxford (2021). The Oxford Martin Programme on Ethical Web and Data Architectures. https://www.oxfordmartin.ox.ac.uk/ethical-web-and-data-architectures/

subjects' ability to exercise their rights, and collectively use their data to counteract exploitation or discrimination. Furthermore, data intermediaries could help ensure sufficient public trust to facilitate data processing that will be essential for innovation and advancement of research for the public good.

AI and responsible innovation:

Our recommendations to alleviate the potential challenges of data protection during AI development are drawn from our research focus on AI accountability. We support the proposal's call for appropriate safeguards regarding usage of personal data (Q1.5.5) as well as the new condition to Schedule 1 to the Data Protection Act 2018 (Q1.5.12). Specifically, we recommend:

1. Treating AI design and development similar to research methodology involving human participants, a field already driven by widely accepted, conscious, and ethical guidelines.

2. Maintain, rather than reform, existing accountability and record-keeping requirements for AI providers, recognising that these underpin the development of a trustworthy AI sector.

3. Recognise that regulation is not a barrier to AI innovation but rather an enabler that provides organisations and individuals with added transparency into safe data practices.

Erosion of trust in online tracking and removal of the balancing test:

1. Recognise that box-ticking is less a result of the existing law, and more a consequence of failure to comply with existing requirements around the selection of lawful basis and the acceptance of electronic means of expressing consent. Any future reform should recognise the potential of *automated consent and compliance*

analysis tools, such as the ones developed by our research group and other researchers to minimise consent interactions while preserving choice[5].

2. Maintain the *balancing test* in the legal text because its removal would create significant risks for the rights and freedoms of UK citizens, and give a false sense of certainty to businesses and other data controllers.

[5] For instance: Zimmeck et al. (2019). MAPS: Scaling Privacy Compliance Analysis to a Million Apps. https://sciendo.com/article/10.2478/popets-2019-0037; Kollnig et al. (2021). A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps. https://arxiv.org/abs/2106.09407; Kollnig et al. (2021). Are iPhones Really Better for Privacy? Comparative Study of iOS and Android Apps. https://arxiv.org/abs/2109.13722. Ekambaranathan et al. (2021). "Money makes the world go around": identifying barriers to better privacy in children's apps from developers' perspectives.

Data intermediaries and institutions

Data intermediaries represent not only an opportunity for innovation and empowerment but also vulnerabilities to privacy and individual freedoms.[6] Data trusts, data cooperatives, personal data stores and other data intermediaries can be effective tools, but without well-defined roles and safeguards, subjects remain vulnerable to exploitation.[7] The consultation asks researchers what roles the government might play in steering, accrediting and regulating the development of data intermediaries (Q1.7.1, Q1.7.2iii-iv), what legal basis data intermediaries might rely on (Q1.7.2i-iii), and how other recommendations of the consultation interact with the interests of intermediaries (Q1.8.1).

As researchers in *Human Centred Computing* and in the *Responsible Technology Institute,* and developers of data intermediary infrastructure under the EWADA project*,* we believe it is essential that data intermediaries be both supported and constrained to act in the best interests and rights of individual data subjects by design. Further, we believe investment, legal distinction, and enforcement of standards in ethics and practices defined in legislation will be essential in ensuring data intermediaries realise their potential (Q1.7.1).

Empowering data subjects and protecting their rights *promotes* innovation. This consultation proposes to lower a variety of barriers and controls for processing and sharing data in the private, public, and third sectors – in particular the proposed removal of requirements for data protection officers (DPOs) and data protection impact assessments (DPIAs) (§163, §167). Eliminating controls for data processing and failing to provide explicit commitments to support the development of more robust and accessible legal and technical mechanisms through which data subjects can exercise their rights, risks endangering individuals' autonomy over their data, and the ability to uncover unlawful data processing and algorithmic harms. The suggestion that removing DPOs is sufficiently countered by an

[6] Brandusescu, A., & van Geus, J. (2020). *Shifting Power Through Data Governance*. Mozilla Foundation Data Futures Lab.

[7] Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*. https://doi.org/10.1093/idpl/ipz014

optional designation to perform a similar function does not mitigate concerns around data subject protection and access.

The proposal further wagers the protection of the individual against the internal commitment of each controller to data protection principles, particularly in high-risk cases: the UK-GDPR requires organisations to undergo a data protection impact assessment where processing is 'likely to result in a high risk to individuals' (§165). The counter-proposal that organisations 'may identify other risk management practices which achieve the intended outcomes' (§166) not only removes what clarity the DPIAs provided but does so with the open acknowledgement that removing this requirement 'may increase the risk that organizations will undertake processing that is high risk without an adequate prior assessment of the impact… [and] may not identify and put in place appropriate safeguards' (§168).

Data intermediaries should not be seen as an alternative to these established forms of risk management like DPOs and DPIAs. Instead, they are designed to act in data subjects' best interests and with governance mechanisms that represent their agency, streamlining the processes of responsible data stewardship while respecting the rights of data subjects. Researchers have already imagined some of the forms such intermediaries could take.[8] Furthermore, data intermediaries also possess the potential to provide a basis of trust necessary to incentivise data sharing from subjects who might otherwise be unwilling to share – providing critical avenues to combat algorithmic harms and private monopolization of data.[9]

[8] See: Balkin, J. (2014). Balkinization: Information Fiduciaries in the Digital Age. *Balkinization*. https://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html;
Colclough, C. J. (2020, November 3). *Towards Workers' Data Collectives*. The Why Not Lab. https://www.thewhynotlab.com/post/towards-worker-data-collectives;
Zhang, A. X., Hugh, G., & Bernstein, M. S. (2020). PolicyKit: Building Governance in Online Communities. Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, 365–378. https://doi.org/10.1145/3379337.3415858

[9] Examples of DIs functioning for data subject advocacy can be seen from labour organization with the Workers Info Exchange, to combating algorithmic harms by OpenSchufa, and even exposing undue influence of misinformation by the Sitra DigiPower investigation.

Data stewardship is a two-way street. It requires not only that controllers responsibly process subjects' data – a responsibility that would be significantly undermined by the removal of DPO and DPIA requirements – but also that subjects might access and make use of their own data in a portable format. Presently, the burden rests on individual subjects to ascertain which controllers might possess their personal data and thereafter exercise their rights (e.g. to request their personal data be made portable or erased).

In the past, Data Subject Access Requests (DSARs) have served as a key instrument for research into the practices of data controllers[10]. Furthermore, while data subjects might more effectively exercise control over their personal data, they rarely have the grounds or the means to benefit from the value of aggregate data sets of which their data may be part. Subjects are inevitably impacted by inferences based on aggregate data sets and require better tools to understand the patterns of the data sets beyond their individual data alone.[11] Helping subjects share in the value their data contains in aggregate, or at least to open opportunities to gain transparency on what data sets contain, should be a central function for data intermediaries. However, data intermediaries will only fulfill this function if they have powerful and accessible means to access data held by controllers.

Accordingly, data intermediaries should be empowered with the ability to broker, manage, negotiate, and facilitate the execution of DSARs or any methods for subject data access in continuance of the UK-GDPR's provision that third parties may be deputised to execute DSARs on subjects' behalf (Q1.7.2). In order to protect bilateral access to data, no fee for DSAR responses should be initiated as the consultation proposes and further, the handling of a large volume of DSARs originating from certified or fiduciary data intermediaries should not be deemed to be 'vexatious' or 'disproportionate' simply because they may be more

[10] Ausloos, J., & Dewitte, P. (2018). Shattering one-way mirrors – data subject access rights in practice. *International Data Privacy Law*, *8*(1), 4–28. https://doi.org/10.1093/idpl/ipy001

[11] Wachter, S., & Mittelstadt, B. (2018). *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* [Preprint]. LawArXiv. https://doi.org/10.31228/osf.io/mu2kf

frequent, thorough, and utilise automated communication when compared to DSARs originating from individual data subjects(§188, §189).

Data intermediaries might play a critical role in empowering individual data subjects to share in the benefits processors enjoy through their use of personal data. In return, data intermediaries, if properly implemented, might help remove some of the need for arduous processes of data stewardship that the proposals aim to alleviate.

In summary, data intermediaries should serve to (Q1.7.1):

- Simplify the process of accessing personal data and knowledge derived from personal data that subjects are entitled to.
- Educate data subjects on how and why they may want to access their data.
- Ensure individual data subjects share in the benefits derived from data they generate.
- Support bottom-up empowerment for data subject collectives and ensure subjects' role in governance over their personal data.
- Reduce risk by limiting the collection of duplicative data sets, adhering to principles of data minimization.
- Act as advocates for better data subject protection, exposing poor data practices of others.
- Work cooperatively with the ICO to mediate complaints on behalf of subjects and aid in identifying areas for enforcement.

AI and responsible innovation

The consultation suggests that (actual or perceived) regulatory barriers to the development of AI in the UK may need to be removed in order to promote innovation. We believe that any such barriers are largely due to misconceptions of the existing regulatory regime, rather than the law itself, and furthermore, that regulation can positively enable better AI systems in the long run. Indeed, the most promising opportunities for the UK AI industry specifically depend, in our view, on strong regulation which encourages private, public, and third sector organisations to be engaged in the careful, contextually-grounded deployment of AI.

To illustrate why, consider that much AI in use today comes in the form of outsourced AI-as-a-Service models, typically based on machine learning (ML). These models are designed for generic use cases, such as image detection, facial recognition, CV scoring and credit risk. ML systems often face substantial shortcomings when deployed in particular contexts; despite performing well in training on a specific test set and given a well-defined task, in practice there are usually nuances to the task that only emerge after deployment. Large scale providers of AI-as-a-Service offer standardised commodity products without any consideration of those nuances. In practice, AI models will always need to be carefully designed and suited to the contexts of deployment, both for performance reasons as well as safety and ethical risks. Trying to shoehorn a generic model into a specific context might end up indirectly leading to more work downstream dealing with the unexpected errors and side-effects.

Because AI performance and risk management is highly dependent on context, current AI-as-a-Service models often don't actually safely 'scale' in practice. We anticipate that the most significant opportunities for innovation and value-creation in the AI sector, in the UK in particular, will not arise in the context of existing dominant technology firms who have already cornered the markets for the AI-as-a-Service model provider for generic tasks. Instead, it may come from a larger number of smaller, more nimble providers who specialise in the human and ethical challenges of embedding AI in particular contexts in a safe, effective, and legally compliant way.

Existing accountability requirements – including Data Protection Officers, Data Protection Impact Assessments, prior consultation with the ICO, and record-keeping – are all key to supporting the development of such organisations, because they encourage contextual, case-by-case risk assessment and mitigation. A diverse and dynamic range of AI providers specialising in the deployment of tailored AI systems in particular contexts is likely to have a greater awareness of relevant best practices and codes for those contexts (e.g. codes of research practices focused on human subjects as seen within research organisations that reference the UKRIO's Code of Practice for Research[12]). They also represent a more realistic growth opportunity for the UK AI sector, than attempts to compete with the already globally dominant tech firms who provide generic AI services. Their competitive advantage would lie in their understanding of the organisational, human and regulatory considerations arising in particular contexts that are necessary to create effective, safe, ethical and legally compliant systems in practice. It is therefore our view that removing the existing accountability requirements of the UK-GDPR would not help but rather hinder such developments.

Erosion of trust in online tracking

Our research has long studied first and third party data collection and tracking across the web, apps and IoT – especially as part of the EPSRC-funded SOCIAM and X-Ray projects since 2015. We understand the frustration of many individuals with the current amount of consent banners on the web. Privacy policies are impossibly long, and hide pertinent information by overwhelming the user. They are clearly not an ideal approach. However, removing data subjects' options not to be tracked as they use the web and apps is not the solution. In our view, the amount of consent banners on the web is not an indicator of overly strict consent requirements under the UK-GDPR, but of lack of meaningful compliance and enforcement to the detriment of data subjects.

Under the UK-GDPR, consent is supposed to be free and informed, rather than coerced and blindly given. Consent given because there's no other option (e.g. 'nudging' and forcing

[12] UKRIO: Code of Practice for Research. https://ukrio.org/publications/code-of-practice-for-research/3-0-standards-for-organisations-and-researchers/3-7-research-involving-human-participants-human-material-or-personal-data/

users into consent), or by making rejection more difficult than consent, does not uphold the legal requirements. Only last week, the Belgian Data Protection Authority confirmed this view and ruled that some of the most common 'consent' pop-ups on the web are against the GDPR[13].

Meanwhile, promising technical measures could be utilised to improve the situation with regard to consent banners. First, as noted in Recital 66 of the 2009 amendment to the ePrivacy Directive, 'the user's consent to processing may be expressed by using the appropriate settings of a browser or other application'. This would drastically cut down the number of consent interactions users would have to make online. Second, there already exists a wide range of privacy-preserving analytics technologies (e.g. ACRA, Matomo) that provide businesses with valuable insights into the performance of their products and whose use would have much less impact on the fundamental rights of affected UK citizens and are hence much less likely to require such consent if configured correctly.

Removing the balancing test would provide false confidence to controllers

A proposal which cuts across all three of the above areas is to remove the balancing test as a requirement of the legitimate interest lawful basis, in a select set of conditions. While the suggested list of processing activities includes many which would often be justifiable under the existing balancing test requirement, it is not hard to conceive of certain circumstances in which even these reasonable objectives might not be justified in light of the rights and interests of the data subjects. For instance, while de-identifying personal data through pseudonymisation or anonymisation might often pass the balancing test, there are circumstances where it would be strongly against the data subject's interests, such as in cases where data subjects later need to be able to access a copy of their data to verify their status to a third party, or support a legal claim. In such cases, the de-identification process

[13] Irish Council for Civil Liberties (2021). Tracking-industry body IAB Europe told that it has infringed the GDPR, and its "consent" pop-ups used by Google and other tech firms are unlawful. https://www.iccl.ie/news/online-consent-pop-ups-used-by-google-and-other-tech-firms-declared-illegal/

may mean the controller is no longer able to satisfy this right because they can't link the subject to their data.

More generally, removing the balancing test would likely do very little to alleviate the overall legal ambiguity inherent to any processing affecting the rights and interests of individuals, regardless of data protection law. It is a feature and asset of the legal system – especially in the UK's common law jurisdictions – that any law (including data protection law) requires case-by-case analysis. Removing the balancing test (Q1.4.1) might thus give a false sense of certainty for data controllers, once implementations of the new data protection regime get challenged in court. Controllers would still need to consider whether the processing has a legitimate purpose, is necessary and proportionate, and also does not infringe other laws, including equality, consumer protection, employment, competition and others; in other words, they will still have to engage in something very much akin to a balancing test to assess whether the intended processing would be lawful. The removal of the balancing test would create immense risks for UK citizens, by wrongly creating the impression for some data controllers that it would then be lawful to ignore other rights wherever the proposed specific legitimate interest conditions are relied upon.

Conclusion

In summary, as researchers, we are concerned about: the removal and reduction of personal data protection measures; the lack of appreciation that regulation has a positive role to play in encouraging effective, safe, and compliant AI development in the UK; the lack of commitment to novel structures to support data subjects; the erosion of trust in online tracking; and the false sense of confidence that may result from the removal of the balancing test for certain legitimate interest purposes. In particular, we are concerned that the diminishment of protections will make individuals and groups more vulnerable, as well as undercut the stakeholder engagement necessary for positive growth and innovation.