DOI:
10.18742/pub01-069

Data: a new direction

A call for evidence from Department for Digital, Culture, Media & Sport

Executive summary

Chapter One

Research Purpose (answers to Q1.2.1 - Q1.2.5)

- Facilitating innovation and research is a commendable objective of the new direction; this objective needs to be balanced with objectives such as being at the service of the public, openness, transparency, and traceability by the public and the communities from whom the data is gathered. Unfortunately, the present definition of research does not cater for this balance; in particular, the purpose of research is completely overlooked, and the openness and transparency are not included in the definition of the research for which data access is to be facilitated.
- In addition to formal definitions, mechanisms of *oversight* and *public investigation* of the research method should be put in place to ensure that the above-mentioned objectives are observed throughout the research process. Otherwise, relaxing the rules for accessing, storing, and processing personal data can lead to a severe breach of public trust and threats to public security and privacy.

Data fairness (Q1.5.1)

- Data fairness is affected by the decisions made by AI practitioners and developers during the project ideation, development, and deployment due to the lack of clearly defined policies, regulations, and oversight across the system development lifecycle. Many existing tools and guidelines for creating Ethical and Fair AI applications are designed for internal self-assessments (e.g., developers, quality assurance, managers). There is an absence of methods to externally assess fairness and bias, particularly following system deployment leading to public decay of trust in AI systems.
- We advocate for an initial systematic assessment of these existing tools and frameworks. The output of this assessment will lead to designing a unified, comprehensive policy framework for Practicing Fair and Ethical AI involving the scientific community, the industry and diverse end-user groups. The policies must be designed considering the diversity of stakeholders of AI systems (e.g., AI practitioners, developers, consumers). Further, we encourage this comprehensive policy framework capture nuanced definitions for fairness and bias as well as methods for ensuring the longevity of compliance.

Building trustworthy AI systems (answers to Q1.5.5 - Q1.5.7)

- We somewhat disagree with the proposal that the government should permit organisations to use personal data more freely, subject to appropriate safeguards, for the purpose of training and testing AI responsibly. We propose that such permission should only be given under clear allowances for scrutiny and transparency.
- We think that the government should provide support to research in the area of explainable AI. The current stage of research does not guarantee developers and researchers to generate meaningful and understandable explanations.

Automated decision-making and data rights (answers to Q1.5.14 - Q1.5.17)

- We strongly disagree with the proposal that the government should remove the Art. 22 'Right to Explanation' in current data protection legislation (Q1.5.17), since this would be unlikely to deliver anticipated benefits, undermine innovation and even imply significant risks.

- However, we tend to agree that the government could helpfully clarify the limits and scope of definitions in relevant provisions ('solely… automated processing' and 'legal effects… or similarly significant') (Q1.5.14). These provisions are only 'sufficiently future-proofed' (Q1.5.16) to the extent that the government continues to explore the balance between innovation and safeguards. We have suggestions as to how and why this can be done.

Public trust in the use of data-driven systems (answers to Q1.5.19 - Q1.5.20)

- Our response to this section connects to relevant parts of Chapter 4 below. We observe that it is unlikely that the proposed reforms will build public trust in the use of data-driven systems unless they are used as an opportunity to signal a clear policy 'reset'. Public trust depends on better demonstrating how technology is helping to improve public services.
- The government should consider further legislative changes 'to enhance public scrutiny of automated decision-making and to encourage the types of transparency that demonstrate accountability' only in combination with these proposed data protection reforms and at the same time, not separately or afterwards (Q.1.5.19). It is important that the government be seen to reflect seriously on prominent recent controversies such as those involving immigration systems, A-level grading and the NHS datastore (Q1.5.20).

Innovative Data Sharing (Q1.7.1 - 1.8.1)

- We think that the government should lend its support to responsible data intermediaries, to overcome common barriers to data sharing, such as lack of trust in organisations that share data, and legal uncertainties about where and how data can be shared. As a collaborator or funder of data sharing schemes, government should continue to invest in data assets such as open and shared data, as well as programmes that facilitate data sharing, the development of data sharing mechanisms, as well as evidence for its efficacy.
- We believe that the proposals in 'Reducing barriers to responsible innovation' would impact on people who identify with the protected characteristics under the Equality Act 2010 (EA). People who fall in these categories should have the option to exclude characteristics protected by the EA.

Chapter Two

Privacy Management Programmes (Q2.2.1)

- We disagree with the notion that the accountability framework of the UK GDPR is too prescriptive to be flexible and risk-based. The current accountability framework regards tools like record-keeping and data protection impact assessments as core elements of a flexible risk-based approach. The resources, cost and effort required to demonstrate accountability can be lowered by taking advantage of modern automated solutions. Automated detective controls can assist organisations in implementing a systematic compliance strategy.

Chapter Four

Building Trust and Transparency Programmes (Q4.4.1-4.3.3)

- As observed above, our response to this chapter links to the previous section on 'Public trust in the use of data-driven systems'. We think that these reforms can only really help develop public trust in relevant technology if they are used to signal a clear policy 'reset' in terms of the way it is being used to deliver improved public services.
- We therefore strongly agree that compulsory transparency reporting on the use of algorithms in decision-making for public authorities, government departments and government contractors using public data will improve public trust in government use of data, drawing on fresh, policy-oriented legal research to advance some suggestions as to how to develop mandatory transparency reporting of this kind.

Citation: Trustworthy Autonomous Systems Hub, THuMP and PLEAD projects (2021) *A Response to: Data: a new direction: call for evidence.* DOI: https://doi.org/10.18742/pub01-069.

Contact details: Justyna.Lisinska@kcl.ac.uk or Mark.Kleinman@kcl.ac.uk

Table of Contents

Response authors

Professor Adriane Chapman: Professor of Computer Science and Head of the Digital Health group, at the University of Southampton. Adriane is internationally recognised for her work on provenance. She is a Fellow of the Alan Turing Institute. Adriane is a co-Investigator in the Trustworthy Autonomous Systems Hub and Principal Investigator of the Infer<Proven>ance and PROVAnon projects.

Archie Drake: Research Associate at King's College London specialising in technology law and policy. Archie has worked in various advisory positions across central and local government. His past research work includes an investigation into new forms of data for dispatch decisions with the London Ambulance Service and he is currently working on legal aspects of the EPSRC Trust in Human-Machine Partnerships (THuMP) project.

Dr Dong Huynh: Research Fellow at King's College London with experience in tracking the flows and use of data with provenance technology in a wide range of applications. He led the development of the demonstrator showcasing provenance-driven explanations for automated decisions at explain.openprovenance.org.

Professor Richard Hyde: Professor of Law, Regulation and Governance and Deputy Head of the School of Law at the University of Nottingham. Richard is interested in the regulation of new technologies. He is a co-Investigator in the Trustworthy Autonomous Systems Hub.

Dr Perry Keller: Reader of Media & Information Law at King's College London, a specialist in information law, including legal issues relating to data privacy and security; access rights to information; and freedom of expression. His current research concerns privacy and data protection law responses to governmental and commercial applications of profiling and predictive analytics to citizen and consumer personal data.

Professor Mark Kleinman: Professor of Public Policy at King's College London and lead of the policy strand of the TAS Hub. Mark has been a government adviser at the No. 10 Strategy Unit, a senior civil servant, and a policy director for three London Mayors. He has held academic posts at the University of Cambridge and the London School of Economics, as well as a chair in international social policy at the University of Bristol.

Dr Justyna Lisinska: Research Fellow at King's College London, working on developing a policy programme for the Trustworthy Autonomous System hub (TAS-hub). She has experience of working within the Cabinet Office and working within a knowledge brokerage at the nexus of evidence and policy with a deep understanding of opportunities and challenges.

Professor Luc Moreau: Professor of Computer Science and Head of the Department of Informatics, at King's College London. Luc is renowned for his work on provenance (both provenance of data and provenance of things). Luc was co-chair of the W3C Provenance Working Group, which resulted in a standard model for provenance. Luc is deputy director of the TAS Hub, and Principal Investigator of the PLEAD and THUMP projects.

Professor Mohammad Reza Mousavi: Professor of Software Engineering at the Department of Informatics (King's College London). Mohammad has extensive experience in and engagement with the connected and autonomous vehicles (CAVs) domain and is the principal investigator of the Verifiability Node in the UKRI TAS Programme.

Dr Mohammad Naiseh: Research Fellow in Trustworthy Autonomous Systems at the University of Southampton. Mohammad has extensive experience in current explainable AI research and applications, is supporting and leading work packages in two agile projects: 'Inclusive Autonomous Vehicles', and 'Trustworthy Human-Swarm Partnerships'.

Professor Sarvapali D. Ramchurn: Professor of Artificial Intelligence and Director of the UKRI Trustworthy Autonomous Systems (TAS) Hub. Sarvapali has over 19 years of experience in developing AI solutions for systems involving large numbers of autonomous vehicles and humans.

Dr Diego Sempreboni: was a Research Fellow in Computer Science at King's College London. Diego brought his experience in security analysis and research methodologies to focus on the trust and security aspects in the TAS Hub. He is now working as Product Security Engineer for a Cybersecurity company.

Professor Elena Simperl: Professor of computer science at King's College London working on socio-technical aspects of data science and AI. Elena led two public sector incubator programmes supporting small and medium businesses to innovate with open and shared data. She is currently the PI of the ACTION project, exploring citizen-led data collection, and MediaFutures, applying a mix of arts and AI to tackle misinformation online.

Professor Sophie Stalla-Bourdillon: Professor in Information Technology Law and Data Governance within Southampton Law School at the University of Southampton and Senior Privacy Counsel and Legal Engineer at Immuta. Her research focuses upon platform responsibility, data governance and data sharing models, and privacy and data protection, with a special emphasis on the legal and ethical implications of data operations within analytics environments.

Dr Gefion Thuermer: Researcher at King's College London, working on human-data interaction, citizen engagement, and data sharing. Author of the data sharing toolkit, as part of the Data Pitch project.

Dr Niko Tsakalakis: Research Fellow at the University of Southampton, working on the implications of disruptive technologies for data protection and on ways to engineer data protection in the design of complex IT systems. For the PLEAD project, Niko ran the analysis for explainability requirements and how this could be operationalised by a provenance driven explanation system.

Professor Ivan Tyukin: Professor of Artificial Intelligence at the University of Leicester. He is a Turing AI Acceleration Fellow and a Co-Investigator of the Verifiability Node in the UKRI TAS Programme. He is a member of the All-Party Parliamentary Group on AI's data governance task force which examines the economic, social, and ethical implications of AI.

Dr Sachini Weerawardhana: Post-doctoral Research Associate at King's College London. Sachini's research interests are in human-aware automated planning, intervention planning, human-agent interactions and explainable planning systems.

About the TAS Hub, TAS Node on Verifiability, THuMP project, PLEAD project, Data Pitch project, ACTION project, and MediaFutures project:

The UKRI TAS Hub (EP/V00784X/1), assembles a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. The role of the TAS Hub is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice.

The UKRI TAS Node on Verifiability (EP/V026801/2), brings together a multi-disciplinary and diverse team of researchers with expertise in AI, robotics, human-computer interaction, systems and software engineering, and testing. Our goal is to develop a unifying framework that will integrate rigorous verification techniques for autonomous systems. Our framework will support the heterogeneous and adaptive nature of verification techniques, their scale, and their levels of abstraction: from requirements and planning to coding and control algorithms to actual hardware and robotic implementation.

The Trust in Human-Machine Partnerships (THuMP, EPSRC EP/R033722/1) project at King's College London aims to advance state of the Art in explainable, intelligent decision support. It focuses on allocating resources in critical domains, bringing together AI and social science researchers to develop and test methods for explaining the reasoning behind plans and actions recommended by data-backed AI-driven systems. THuMP considers how collaborative, interactive decision-making can foster trust in AI systems, as users gain confidence in decisions reached through mutual understanding.

The PLEAD project (Provenance-driven and Legally-grounded Explanations for Automated Decisions, EPSRC EP/S027238/1) brings together an interdisciplinary team of technologists, legal experts, commercial companies and public organisations to investigate how provenance can help *explain the logic that underlies automated decision-making* to the benefit of data subjects as well as help data controllers to demonstrate compliance with the law, under research conducted jointly by the University of Southampton and King's College London. Explanations that are provenance-driven and legally grounded will allow data subjects to place their trust in automated decisions and enable data controllers to ensure compliance with legal requirements set on their organisations.

Data Pitch (H2020 732506) was a European open innovation programme bringing together corporate and public-sector organisations that have data with startups and SMEs that work with data. It was centred around a competition with several tracks which described shared data challenges, and a virtual accelerator programme to help startups and SMEs develop solutions to meet these challenges. Data Pitch delivered 26 challenges with shared data in domains such as health, tourism, environment or skills, and supported 46 companies in 13 countries, which created 112 jobs and more than €22 million in impact.

The ACTION project (Participatory science toolkit against pollution H2020 824603) delivers an accelerator for 16 citizen science pilots from the UK and six other European countries, which collected and analysed data about light, soil, water, noise, and air pollution. ACTION worked with citizen-led teams to design mechanisms for crowdsourced data collection, curation, preservation and governance, which are described in the ACTION toolkit to be published at the end of 2021. https://actionproject.eu

The MediaFutures project (H2020 951962) explores how approaches from science, technology, and the arts can come together to transform how people engage with data and AI in science education, high-quality journalism, and democratic processes. The project delivers an accelerator and a residency programme, which in its first year supported 19 startups and artists tackling challenges such as online polarisation, climate change denial, and cyberbullying. Data is used as a material and as an enabler for novel products, services, artworks and experiences.

List of acronyms

ADM	Automated Decision-Making
AFR	Automated Facial Recognition
AI	Artificial intelligence
DPIAs	Data Protection Impact Assessments
EA	Equality Act 2010
PLEAD	Provenance-driven and Legally-grounded Explanations for Automated Decisions
SDG	Sustainable Development Goals
SMEs	Small and mid-size enterprises
TAS-hub	Trustworthy Autnonomous Systems Hub
THuMP	Trust in Human-Machine Partnerships
TIGRR	Taskforce of Innovation, Growth and Regulatory Reform
XAI	Explainable artificial intelligence

Chapter 1 Reducing barriers to responsible innovation

Research Purposes

Q1.2.1. To what extent do you agree that consolidating and bringing together research-specific provisions will allow researchers to navigate the relevant law more easily?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible.

This reduces the complexity of navigating through legal documents and unifies the procedures across the field. With this unified framework, both the processes are streamlined, and thorough scrutiny of the processes is made possible. The legislation will evolve, but a consolidated policy facilitates consistency and transparency. However, there are other complicating factors in the use of data, such as IP rights and data sharing agreements, as well as domain-specific aspects pertaining to health-, justice-, geo-spatial data. Although consolidating legal provisions is helpful, we are unsure if all of these aspects are considered within the remit of such legal provisions.

Q1.2.2. To what extent do you agree that creating a statutory definition of 'scientific research' would result in greater certainty for researchers?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible.

We need to have a clear-cut definition of scientific research before consolidating the legislation around research provisions. It is important to consider that in different fields, there are different standards and processes for what constitutes scientific research. Many pieces of research on data science techniques are not subject to the processes that concern medical research (Pryor, 2012).

Q1.2.3. Is the definition of scientific research currently provided by Recital 159 of the UK GDPR ('technological development and demonstration, fundamental research,applied research and privately funded research') a suitable basis for a statutory definition?

○ Yes
○ No
○ Do not know

Please explain your answer, providing supplementary or alternative definitions of 'scientific research' if applicable.

A more nuanced definition with different levels of being at the service of the general public and generating knowledge for the public goods is needed. The term "privately funded" refers to the source of funding rather than the purpose of research. There is no consistent classification of the types of research in this definition. However, scientific research is not limited to research for public benefit. Commercial organisations can label their activities as scientific research by using users' data for their commercial gain or fitting into their agenda. There have already been examples of misuse of data and data-oriented research against the public good. The most famous example is the Cambridge Analytica scandal, where users' data was used for targeting campaigns to influence voters during elections without their consent. On the other hand, Google Deepmind (part of the Alphabet group) has been sued for using data to undertake research to develop tools for clinicians (for cancer diagnosis) but at the expense of patient privacy, for which they were challenged in the courts and caused public outcry, thus highlighting the lack of concern such companies have for public trust. Another example is the common practice of over-engineering systems to pass the legal emission test by using the publicly available data from norms and standards. In such processes, vehicles are over-engineered to perform exceptionally well during test scenarios, while in actual driving scenarios beyond the test regime, they emir excessive number of harmful particles. This manifested itself in the Volkswagen emission scandal but is by no means restricted to this single case. In the Volkswagen scandal, the company intentionally programmed their engines to activate emissions controls only during laboratory testing. Research with commercial incentives may be against the public interest (e.g. environmental concerns) and relaxing the use of public data for such type of research is questionable. We have seen several examples of such activities, which we call cyber-physical doping (Biewer et al. 2021), and we advocate a notion of technical transparency in data use to make such activities amenable to scrutiny. Disregarding such provisions for transparency and public scrutiny may not only breach public trust but also lead to serious security and privacy issues (Kuner, 2017).

We, therefore, not only suggest a different definitional approach than stated in Recital 159 of the UK GDPR, but also we recommend developing an oversight mechanism for ensuring transparency and traceability in the use of data for research.

Q1.2.4. To what extent do you agree that identifying a lawful ground for personal data processing for research processes creates barriers for researchers?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible, including by describing the nature and extent of the challenges.

In the short term, this may appear to facilitate use of data by researchers for research purposes. However, it effectively creates the perception that all control of personal data is removed from users, which over time will erode the trust and willingness of individuals to provide quality and correct data. This will erode both the data available for research and in the health domain, creating an incentive for individuals to provide poor information for their own health treatment (Hussain-Gambles, 2004). Mascalzoni et al., (2009) found that consent would be less likely to be given when personal data regarding income, education, and occupation is linked to health data. Instead, a better model of consent is needed. The Great North Care Record collects patient health data from a range of sources and makes them accessible in one place (Lau, 2017). Consent is managed on an opt-out basis and only approved researchers can access anonymised patient health data (Konstantinidis et al. 2020; Konstantinidis Holt & Chapman, 2021).

Q1.2.5. To what extent do you agree that clarifying that university research projects can rely on tasks in the public interest (Article 6(1)(e) of the UK GDPR) as a lawful ground would support researchers to select the best lawful ground for processing personal data?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible.

We need to have a process with proper oversight (of diverse stakeholders) to make sure that public interest is pursued in a university research project. Using data should be transparent and traceable, we should make sure that diversity in the population used for gathering data and in the analysis, techniques used for building predictive and classifying models are safeguarded.

Data fairness

Q1.5.1. To what extent do you agree that the current legal obligations with regards to fairness are clear when developing or deploying an AI system?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible.

There should be guidance to support AI practitioners and developers to use current fairness and bias mitigation tools in the AI development process, and on how to choose between a large pool of these tools. Fairness and bias are nuanced concepts, especially for domain-specific AI systems aimed at niche user groups. There should be clear policy provisions in place that practitioners/developers can follow to define fairness and bias mitigation for these types of domains and to communicate the system's limitations to end users prior to deployment. We also recommend that AI practitioners and developers be given incentives to learn about incorporating fairness into the design/development process of AI systems at design time, as well as preserving fairness after deployment. These incentives could come in the form of training/certificate programs for practitioners and developers, funding, improved visibility and public acceptance for systems designed by adhering to such provisions. Furthermore, setting up a framework for periodical evaluations of the individuals and entities involved in AI development will encourage longevity of compliance.

A recent review of 39 Frameworks/Guidelines for creation Ethical and Fair AI Systems (Ayling and Chapman, 2021) that have a focus on practical applicability found:

- Stakeholder types directly using the tools are clustered around the product development phase of AI (developers, delivery, quality assurance), with the output from the tools (reporting) being used by management Decision Makers.
- There is little participation in the assessment or audit process by certain stakeholder groups (Voiceless, Vested Interests and Users) who are not included in the process of applying the tools or interacting with the outputs as tools for transparency or decision-making. Perhaps most surprising is how little inclusion there is of Users/Customers in these tools.
- Nearly all of the tools are for Internal Self-assessment, with only the IEEE standards requiring any kind of external verification, and the two examples of public registers providing explicit transparency.
- Techniques and practices deployed by other forms of Impact Assessment (like EIAs) are not present or rarely suggested in ethical AI impact assessments (Participation process, Baseline study, Life-cycle assessment, Change measurement or Expert committees.)
- The output from the tools can provide documentation for Oversight from external actors, but as the majority are Internal activities, there is generally no process or requirement for the wider publication of the results of these tools.
- A third of the Impact Assessment tools focus on Procurement processes for AI systems from 3rd party vendors, indicating the need for not only producers of AI products to engage with ethical assessment, but also the customers for these products, who will be the ones deploying the products.

However, Ayling and Chapman (2021) also noted that there was not a clear reason for organisations to engage with these guidelines and that regulation in this area is needed.

Building trustworthy AI systems

Q1.5.5. To what extent do you agree that the government should permit organisations to use personal data more freely, subject to appropriate safeguards, for the purpose of training and testing AI responsibly?

○ Strongly agree
○ Somewhat disagree
○ Neither agree nor disagree
○ Strongly disagree

Many oganisations are already using personal data to train and test their machine learning systems. For example, Sensyne and Deepmind are UK-based companies that have negotiated deals with NHS Trusts to obtain personal data in the past to train their algorithms, potentially without the permission of the data owners. This has led to Deepmind being taken to court for its use of personal data without consent. There are three issues to address with the question (i) whether is the system that uses personal data is defined as being artificially intelligent instead of being simply an automated decision making system (ii) whether the organisations that use the data are open and transparent about the objectives of the AI system being trained, and (iii) whether the providers of the data understand the technology to be applied to the data.

There should be no distinction between an AI-based system and any other system that applies an algorithm to data. AI is based on the use of algorithms that may employ a range of methods to make decisions, for example, neural networks, rules, linear programming, etc.. The focus should be placed on the likely impact on the owners of the data and the public at large rather than on the technology.

Many organisations do not currently have the internal expertise to understand the value of the data they own. An example of this is where Deepmind used personal data from the Royal Free NHS Trust to develop tools for clinicians. While Deepmind employs hundreds of data scientists, the Royal Free NHS Trust does not have the same inhouse capability to match it. This asymmetry in expertise can lead to exploitation and misrepresentation of the intent to use the data, which eventually led to a challenge in court in the case of Deepmind.

We recommend such permission should only be given under clear allowances for scrutiny and transparency. The nature of use, its purpose, and the general principle governing the use have to be made clear to ensure the right balance between innovation facilitation and personal data protection. If the government wants to allow using personal data more freely for training AI, the process should be transparent and traceable. We need to make sure that diversity is safeguarded, and there should be oversight by a diverse community of stakeholders. Various stakeholders, including various groups in the general public, should have the means to scrutinise how their data have been used in training AI systems and whether sufficient measures have been taken to avoid bias.

Q1.5.6. When developing and deploying AI, do you experience issues with identifying an initial lawful ground?

Please explain your answer, and provide supporting evidence where possible.

We foresee the need of transparency principles that mandates a higher level of transparency by the data controllers regarding their processing activities, including data collection, algorithmic decision-making process and sharing of data. The information provided to the data subject should be sufficient and meaningful. However, the opacity or black-box nature of machine learning and deep learning algorithms make it difficult to provide sufficient and meaningful information. The research in the area of explainable AI is not yet able to guarantee meaningful and sufficient information to the data subject. Current advances in explainable AI research mainly focused on increasing the explainability of AI models to help data scientists and ML engineers to debug AI (Naiseh et al., 2021), whereas the focus on generating understandable and meaningful explanations to the public is still limited (Miller, 2018). Further, it remains open how we would construct the 'ideal algorithmic explanation' and how these explanations can be embedded in the AI development process. We suggest supporting innovation and research in the area of human-centred explainable AI.

Q1.5.7 When developing and deploying AI, do you experience issues with navigating re-use limitations in the current framework?

Please explain your answer, and provide supporting evidence where possible.

While Article 89(1) of the UK GDPR indicates that processing data requires appropriate safeguards for the right and freedom of data subject. The main issue is related to a lack of guidance procedures in the oversight of the use and re-use of data. This is particularly essential when the data subject has not provided specific consent to the use of the data or the data are not directly collected from the subject.

Automated decision-making and data rights

Q1.5.14. To what extent do you agree with what the government is considering in relation to clarifying the limits and scope of what constitutes 'a decision based solely on automated processing' and 'produc[ing] legal effects concerning [a person] or similarly significant effects?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible, including on:
○ The benefits and risks of clarifying the limits and scope of 'solely automated processing'
○ The benefits and risks of clarifying the limits and scope of 'similarly significant effects'

We welcome this initiative as a timely reflection on how legislation might better contribute to the vigour and direction of AI innovation in the UK. With regard to clarifying the limits and scope of what constitutes 'a decision based solely on automated processing' and 'produc[ing] legal effects concerning [a person] or similarly significant effects', **we somewhat agree.**

There would be substantial benefit in clarifying the idea of 'solely… automated processing', especially with a view to ensuring that good explanations are provided for processing which involves interactions between machines and people. This echoes the ICO's opinion that the scope of Article 22 should be extended to cover partly automated decisions (ICO, 2021).

For example, it might help to focus on a broader category definition of processes of algorithmic decision-making without meaningful human involvement. This would help shift to a more constructive focus on how to support meaningful human involvement, in the form of interactions with AI systems.

Our research has started to demonstrate the potential to scale and expand beyond 'solely automated processing' using technological solutions. The PLEAD project has developed automated computable explanations that have been applied equally to partly automated decisions, with various degrees of human involvement, and can be extended to cover manual decision-making processes. The Thump project has been developing innovative approaches in explainable artificial intelligence (XAI), a field of computational methods and tools to interpret and explain AI systems.

The research work undertaken in PLEAD demonstrates that computable explanations can help to achieve procedural fairness for both fully and partly automated decision making. In the example of COVID given in the consultation report (p. 38), the PLEAD system would have been able to attach automatically computed explanations to each decision, explaining to the GPs reviewing the decisions how the clinical image of each patient contributed to their inclusion or exclusion from the list. It could also be used on demand by the reviewing GPs to construct explanations based on any number of criteria that they would select, allowing them to query the results according to their needs. And, of course, it could produce auditable explanations about the operation of the system to allow audits and troubleshooting, assisting in a robust strategy for accountability as it will be discussed in Q2.2.1.

Thump project research work has focused on cutting-edge approaches capable of developing peoples' trust as they interact with automated systems:

- **Information provenance:** Ensuring that all the 'information, data dependencies and processes underpinning a decision' are captured in a standardised way, enabling XAI to link up across organisational boundaries (ICO & Alan Turing Institute, 2020).
- **Planning:** Developing generally-applicable, non-domain specific XAI 'as a service' in this industrially important branch of AI (Cashmore et al., 2019; Fox et al., 2017).
- **Argumentation:** Developing XAI as conversational interactions between humans and machines, allowing humans to question machine decisions using familiar language (Canal et al., 2021; Miller, 2018).
- **Visualisation:** Approaches to user-centred design of XAI system interfaces (Borgo et al., 2018).

So the government can draw and build upon insights and expertise from current UK research projects when moving to clarify the limits and scope of what constitutes 'a decision based solely on automated processing'.

There would also be benefit in clarifying 'legal… or similarly significant effects', although it is less readily apparent how this could usefully be accomplished if effects continue to be addressed only in terms of individuals. There are opportunities to strengthen the protection of people's rights and reduce the complexity of assessing when human intervention was 'meaningful' (see above).

Thump project work especially highlights the need to move beyond the individual as the unit of assessment for human impacts in XAI:

- We are conducting user studies on how people interact with automated explanations in practice, based on a method proven to generate useful insights (Heath & Luff, 2018). These insights can then be used to (re)design and develop explanations, helping organisations implement them successfully and responsibly.
- Ultimately, automated decision-making has societal effects which deserve to be taken seriously (Keller, 2019) and which have become prominent in UK national policy dialogue in recent years (British Institute of International and Comparative Law & Dickson Poon School of Law, King's College London, 2021).

We would certainly recommend that the government consider alternative approaches to this problem. Given the global, interconnected nature of digital technologies, national legislative divergence is difficult to justify in the absence of incontrovertible benefits. Instead of focusing on legislative adjustments, we would recommend that the government concentrate on establishing participatory processes for more careful and thorough development of socio-technical standards for automated decision-making including further XAI advances. The government could: either refer these questions of legislative scope and clarity to the Law Commission, as the House of Lords recommended in 2018 (House of Lords Select Committee on Artificial Intelligence, 2018); or convene a public inquiry using the Inquiries Act 2005 to prevent recurrence of the kinds of harms that have started to occur in recent years as automated decision-making has become more prevalent.

In addition to the above, a risk-based approach needs to be considered. All models are abstractions, and as such are incorrect for a segment of uses. On top of this, the systems are dependent upon data, which have known biases including systemic and institutional biases that are not easy to remove. The use of AI for a decision based solely on automated processing needs to be carefully weighed with the impact of that decision. If an individual will lose their rights, liberties or life, then such a system should not be used. Even in lower-stakes decisions, a recourse to human review and intervention must be available.

Q1.5.16. To what extent do you agree with the following statement: 'In the expectation of more widespread adoption of automated decision-making, Article 22 is (i) sufficiently future-proofed, so as to be practical and proportionate, whilst (ii) retaining meaningful safeguards'?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible, on both elements of this question, providing suggestions for change where relevant.

We somewhat disagree with the statement that Article 22 is (i) sufficiently future-proofed, so as to be practical and proportionate, whilst (ii) retaining meaningful safeguards, in the expectation of more widespread adoption of automated decision-making. It needs to be properly implemented. As the ICO has observed in its response (ICO, 2021), developing Article 22 as a useful and effective legislative measure requires continuing investment in 'detailed analysis and evidence' on the best ways to support humans in engaging with AI-enabled automated decision-making. Discussions about the need for legislative reform in this area should include clearer acknowledgement of the investments the UK government and business are making in XAI[1].

[1] This also links to the question of record-keeping in Section 2.2 of the consultation (Q2.2.11). As the Competition and Markets Authority have noted, record-keeping is essential for explanations and XAI development (CMA, 2021) and provenance standards provide an important tool in this respect (FN 177).

Q1.5.17. To what extent do you agree with the Taskforce on Innovation, Growth and Regulatory Reform's recommendation that Article 22 of UK GDPR should be removed and solely automated decision making permitted where it meets a lawful ground in Article 6(1) (and Article 9-10 (as supplemented by Schedule 1 to the Data Protection Act 2018) where relevant) and subject to compliance with the rest of the data protection legislation?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

The Government believes that AI and automated decision-making is likely to increase in the coming years and recognises that organisations will need "the right tools to build trustworthy and fair AI systems." The Government is, therefore, attempting to identify compliance challenges for organisations who wish to develop and deploy AI systems.

In this respect, the Government considers that the scope and application of Article 22 of the UK GDPR are subject to uncertainty and that, where Article 22 applies, its safeguards may not be practicable or proportionate, especially the obligation for a meaningful human review.

Although we welcome the Government's intention to clarify the application of Article 22 (see above), we disagree that increased take-up of automated decision-making will render its safeguards unserviceable. In an effort to reduce complexity for organisations that deploy automated decision-making, the Government is considering a proposal by the Taskforce of Innovation, Growth and Regulatory Reform (TIGRR) to remove the right to human review from Article 22. We share the concerns of the ICO about the removal of this right to human review.

It appears from the TIGRR report that two benefits are anticipated from removing Article 22. Neither of these benefits are realistic.

TIGRR's first anticipated benefit is making it easier and cheaper for organisations to 'automate routine processes because [Article 22 means] they must also have a manual process'. This ignores the growing field of explainable artificial intelligence (XAI), which has provided a wide range of computational methods and tools to interpret and explain AI systems (Vilone & Longo, 2020). Article 22 does not require a manual process.

TIGRR's second anticipated benefit is the exclusion of 'meaningful human review', in favour of fully automated decision-making where this might 'perform better'. This ignores well-known issues with defining AI system performance. It reflects a harmful, machine-centric view which focuses on theoretical performance and tends to consider humans as an inconvenience, rather than considering human-machine interactions as a question of partnership prioritising human goals (Jennings et al., 2014).

TIGRR's assessment underestimates the contribution of computational methods and tools in assisting efforts to comply with data protection obligations and the further benefits they offer. For example, the PLEAD project has developed tools for computable explanations that can assist in the transparency, understandability and trustworthiness of automated decisions (Huynh et al. 2021). Two different types of explanations can be supported: explanations that target the recipient of an automated decision to explain the reasons behind a decision and the decision-making process; and explanations that target the decision-maker, their auditor(s) or supervisory authority to assist in reviewing the decision by providing accounts of the decision-making process. These explanations use provenance data models (Moreau & Missier, 2013) to trace back decisions to data sets, processes, people and organisations. Provenance data can be linked up across organisations boundaries, offering a common vocabulary to express information throughout the lifecycle of the AI. Their value in explaining responsibility, safety and performance in automated processing was included in the ICO's authoritative guidance (ICO, 2020) and was also described as "an additional tool that could provide the basis for standards in record-keeping for explainability." (CMA, 2021). Similar efforts focus on different technologies to automate other aspects of compliance with data protection.[2] A benefit of computational methods such as those proposed by PLEAD is that they are entirely automated, once deployed for the specific context. They introduce a small computational overhead when systems run while providing benefits for consumers and businesses, in near real-time.

TIGRR's proposal also ignores significant risks from the removal of Article 22 as follows:

- Disincentivising XAI innovation in the UK. Article 22 has 'provided added impetus to solve the problem of explainable AI systems' (House of Lords Select Committee on Artificial Intelligence, 2018). Removing it risks withdrawing a significant spur to XAI research and development in the UK.
- Undercutting UK competitiveness on XAI. Removing Article 22 in the UK would place firms at a disadvantage relative to those in the European Union on XAI, as well as relative to US actors known to be making major investments in XAI including 'big tech firms' like Google, Microsoft and IBM as well as the US government (DARPA, 2018).
- Promoting irresponsible technological innovation. Removing Article 22 would tend to promote the exclusion of humans from decision-making processes which have significant outcomes for humans.
- Undermining social support for AI technology over the longer term. XAI is recognised as a computational technique useful for the development of trust in AI systems (Canal et al., 2020). Removing Article 22 would tend to promote applications with greater capacity for undermining trust, increasing the likelihood of a backlash and undermining the potential for wider groups of people to contribute constructively as part of AI innovation processes.

Public trust in the use of data-driven systems

Our contribution in response to this section and to relevant parts of Chapter 4 below (Delivering better public services) draws on some fresh legal research aiming to develop reflections for policymakers, based on the observation that these issues have become increasingly controversial in the UK recently in ways that are further undermining public trust in relevant technology (Drake et al., 2021).

[2] See, for example, the outputs of the project 'Business Process Re-engineering and functional toolkit for GDPR compliance' (BPR4GDPR) at https://www.bpr4gdpr.eu about a compliance ontology and Laurens et al. (2020) modelling automated solutions for Data Protection Impact Assessments.

To quote directly from the article:

- The *proposed reforms do apparently offer some potential for a 'reset'* from the vicious cycle of mistrust (for example, if compulsory public sector transparency reporting Q4.4.1 is developed into a strong and effective measure including in relation to procurement, or if data protection reforms are accompanied by proper measures beyond data protection Q1.5.19-20).
- However, *it is far from clear that a reset is intended*. The proposed reforms are 'puzzling' as an accountability framework substitution (Boardman & Nevola, 2021). In the context of low trust and growing contestation, ideas for new arrangements will seem vague and unreliable compared to specific plans to remove perceived legal obstacles to the present direction of AI-related decision-making (eg much more detailed proposals Q4.4.4-7 on 'public interest' processing, 4.4.8 'streamlining' matters for the police on biometrics, 4.5.1 on 'public safety' justifications). Whatever else, the government is dismantling accountability mechanisms. For example, it is proposed not just to abolish Data Protection Impact Assessments (DPIAs) (Q2.2.7-8) but to remove Art.30 GDPR record-keeping requirements (Q2.2.11). Such measures appear likely to 'hinder effective enforcement', despite the claim that the risks will be 'minimal'.[3]
- Rather than developing Automated Decision-Making (ADM) implementation standards constructively, these proposals therefore appear – especially as the first step in AI-related reforms – to *perpetuate underlying assumptions and trends in UK AI policy in a corrosive manner*. Prominent amongst relevant assumptions is the idea that it is legal constraints that are problematic, especially in data protection, rather than policy decisions. The Art. 22' Right to Explanation' is emblematic of this issue, as noted in the Background. The current proposal to remove this provision (Q1.5.17) is an unhelpful priority. Whatever its shortcomings, recent efforts to clarify its effect in guidance have represented a step forward for AI awareness and technical standards. Removal will not make AI compliance substantially easier. Actors determined to contest AI-related decision-making have a range of other grounds at their disposal with which to demand different kinds of explanations.
- From a legal contestation perspective, therefore, the reforms as currently proposed seem *highly unlikely to achieve the stated aim of increasing public trust in ADM*.

Q1.5.19. Please share your views on what, if any, further legislative changes the government can consider to enhance public scrutiny of automated decision-making and to encourage the types of transparency that demonstrate accountability (e.g. revealing the purposes and training data behind algorithms, as well as looking at their impacts).

The government should consider further legislative changes 'to enhance public scrutiny of automated decision-making and to encourage the types of transparency that demonstrate

[3] Equally some might interpret the proposal to merge biometrics and surveillance camera oversight into the ICO (Q5.8.1-2) as a further, more formal discouragement to proactive regulation (see Discussion above).

accountability' only in combination with these proposed data protection reforms and at the same time, not separately or afterwards. As the House of Lords recommended in 2018, the appropriate steps given levels of obscurity and confusion in this field would be to:

- Request that the Law Commission' consider the adequacy of existing legislation to address the legal liability issues of AI and, where appropriate, recommend to Government appropriate remedies to ensure that the law is clear in this area'; and
- Seek the National Audit Office's advice on regulators' responsibilities and resourcing. (House of Lords Select Committee on Artificial Intelligence, 2018).

Q1.5.20. Please share your views on whether data protection is the right legislative framework to evaluate collective data-driven harms for a specific AI use case, including detail on which tools and/or provisions could be bolstered in the data protection framework, or which other legislative frameworks are more appropriate.

Legal actions on a range of use cases illustrate that there is limited value to picking apart the legislative framework for data protection or trying to distinguish it from other relevant frameworks. In practice, data protection provisions overlap and interact with a wide range of law in fields including public administration, discrimination, human rights, contract etc. The starting point should be evidence of where automated decision-making has resulted or might result in serious harm to people in ways that are not currently properly addressed. The government should be demonstrating that it is drawing lessons from instances of harm and improving rules on that basis. For example, drawing from the many cases considered in the latest legal research:

- The inability to determine whether the law was properly applied in UK visa systems over 2011-14, because of English testing services and the Home Office's reliance on a contractor's automated systems for identifying cheating, suggests that public procurement standards need to develop to accommodate technological change.
- The Bridges appeal of 2020 on the use of Automated Facial Recognition (AFR) systems by the police suggests value to a broad, participatory process for reflection on their implementation, rather than the continuing unquestioned support apparent in government policy.[4]
- The A-levels grading controversy of 2020 suggests a need to reflect on the implications of digital change for educational assessment decisions over the longer term.
- The 2020-21 Uber & Ola judgments in the Netherlands on explanation of automated decisions suggests a similar need to reflect on the implications in employment.
- The 2021 controversy over contracting in relation to the proposed NHS 'datastore' suggests that the government needs to communicate the anticipated benefits of relevant changes to decision subjects more effectively.

[4] R (On the Application Of Bridges) v South Wales Police [2020] EWCA Civ 1058; [2020] 1 WLR 5037

Innovative Data Sharing Solutions

Q1.7.1. Do you think the government should have a role enabling the activity of responsible data intermediaries?

○ Yes
○ No
○ Don't know

Please explain your answer, with reference to the barriers and risks associated with the activities of different types of data intermediaries, and where there might be a case to provide cross-cutting support). Consider referring to the styles of government intervention identified by Policy Lab - e.g. the government's role as collaborator, steward, customer, provider, funder, regulator and legislator - to frame your answer.

It would be useful for the government to lend its support to responsible data intermediaries for several reasons:

- Trust is a big barrier to data sharing. Something like a regulator-approved 'responsible data intermediary' scheme would go a long way to overcoming this hurdle. This would require a clearer definition of what constituted responsible (and by extension, irresponsible) data sharing.
- Clear definitions, as would be provided by including the recitals in the legal text, would also help to address the second common issue, which is (legal) uncertainty. Given the potential retributions for getting it wrong, organisations may be hesitant to engage in data sharing; clear guidance and government support and regulation would help clear up these uncertainties and thus allow more organisations to share data responsibly and confidently.
- Government could act as a collaborator and funder, but also a regulator and legislator in this context.
    - As a collaborator, government should continue to invest in providing access to key data assets. This includes open data, whose broad impact has been widely documented (Open Data Institute, 2015), but also shared data, which cannot be released in the public domain. The UK has led the open government data revolution worldwide; it should do the same by sharing data with other key data stakeholders and enabling mechanisms for business to government data sharing (Walker et al., 2019).
    - As a funder, government should invest in programmes that pilot and explore different data sharing mechanisms, including trusts, cooperatives, and corporate and contractual mechanisms (Ada Lovelace Institute, 2021) across a range of sectors to understand sociotechnical challenges and collect evidence of impact of data sharing. One such programme was Data Pitch (2017 to 2019, funded by H2020). Data Pitch implemented an open innovation programme to explore the use of corporate and contractual mechanisms to share data between solution providers (startups, SMEs) and data providers (in the public and private sectors). Data Pitch was a data sharing intermediary; it worked with data providers to define 28 business challenges that could be addressed with shared data. It set up a programme of funding, mentoring and support for 46 small and medium businesses in 13 countries, which developed solutions to these challenges, adding 112 new jobs and unlocking more than 22 million € in efficiency savings, sales, and additional investment (Godel et al., 2019; Thuermer et al., 2019). A second such programme is ACTION (2018 to 202), which focuses on citizen engagement in data collection, analysis and use to

tackle various forms of pollution. Citizen-led data work requires participatory data stewardship to ensure data is maintained and used ethically and equitably (Thuermer, 2021), however, the processes, tools, and technologies to realise such participatory governance formats are not yet well understood (Reeves & Simperl., 2019); this is a missed opportunity, given the social and environmental impact of citizen data observatories and the role they can play in improving coverage and granularity of data used to inform Net Zero policy and SDG reporting. Lastly, in the MediaFutures project (2020 to 2023), artists and start-ups use open or shared data to address issues of misinformation, showcasing the value such data can have for current challenges.

Q1.8.1. In your view, which, if any, of the proposals in 'Reducing barriers to responsible innovation' would impact on people who identify with the protected characteristics under the Equality Act 2010 (i.e. age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation)?

The proposal to allow wider use of data for research may interfere with these characteristics, especially where they change over time (e.g. disabilities develop, marriages dissolve, pregnancies do not last forever), but remain recorded against identifiable individuals. Where data collection and sharing is intended under this new regulation, data subjects who fall in these categories should have the option to exclude characteristics protected by the EA.

Chapter 2 Reducing burdens on businesses and delivering better outcomes for people

Privacy management programmes

Q2.2.1. To what extent do you agree with the following statement: 'The accountability framework as set out in current legislation should i) feature fewer prescriptive requirements, ii) be more flexible, and iii) be more risk-based'?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Deployment of the computational tools mentioned in our replies to Q1.5.14 - Q1.5.17 beyond cases of automated decision-making can alleviate the burden of demonstrating compliance with the current accountability framework.

The Government has expressed concern that current accountability obligations may pose 'a significant and disproportionate administrative burden' for organisations, especially for ones undertaking low-risk processing. The Government instead proposes the take-up of flexible and risk-based privacy management programmes, in the hope that this 'should result in a more coherent, comprehensive and systemic approach to accountability'. Privacy management programmes comprise different building blocks, dealing with internal governance structures, data assets, risk assessment, breach reporting, staff training, and ongoing monitoring, among others (para. 156). We note that each of these blocks already has a corresponding obligation under the UK GDPR, making it unclear how a privacy management programme will significantly reduce the administrative burden on organisations. In any case, we believe that importance should be given to available tools that can assist organisations in operationalising accountability, whether under the existing framework or a privacy management programme. In PLEAD, we have illustrated how computable explanations can be used as detective controls to facilitate compliance with most data protection goals (e.g., transparency, data minimisation, storage limitation, accuracy, procedural fairness…) and support a systematic accountability strategy (Dong Huynh, Stalla-Bourdillon, Moreau, 2019)

Finally, we regard with scepticism the Government's proposal to remove the requirements for data protection impact assessments, prior consultation, record keeping and breach notification, which are viewed by data protection authorities as integral parts of effecting data protection by design.[5] Instead, focusing on the tools used to embed preventive, directive, detective, and corrective controls within systems and processes, and, creating a knowledge base by mapping these tools to privacy threats will streamline many aspects of compliance, significantly reducing the administrative burden for, e.g., risk assessment, transparency reporting and subject access requests.

[5] See, e.g. CNIL (2018):2; Danezis (2014):11; ICO (2019):193.

Chapter 4 Delivering better public services

Our contribution in response to this chapter and to Q1.5.19-20 above (Public trust in the use of data-driven systems) draws on some fresh legal research aiming to develop reflections for policymakers, based on the observation that these issues have become increasingly controversial in the UK recently in ways that are further undermining public trust in relevant technology (Drake et al., 2021).

Building Trust and Transparency

Q4.4.1. To what extent do you agree that compulsory transparency reporting on the use of algorithms in decision-making for public authorities, government departments and government contractors using public data will improve public trust in government use of data?

○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Please explain your answer, and provide supporting evidence where possible.

There are very good reasons to believe that compulsory transparency reporting of this kind will improve public trust in government use of data. Activists are already using other transparency mechanisms, including Freedom of Information and the common law duty of candour as well as DPIAs, to obtain relevant disclosures in a reactive and piecemeal way. Proactive reporting obligations would address the growing sense that public actors are automating decision-making in secret and badly.

Q4.4.2. Please share your views on the key contents of mandatory transparency reporting.

Mandatory transparency reporting should first and foremost involve proactive disclosure of the fact that a relevant decision process will be or has been automated in part or in full. Reporting should then also include a description of the purpose and lawful basis of the automation, a basic description of how decision automation has been implemented and details of how decisions can be challenged if mistakes have been made. There is no need for transparency reporting rules to specify 'meaningful information', detailed information on system logic etc. because the point is to be upfront with the public about which decision processes are being automated, rather than rushing ahead to open the AI' black box'. Or, in terms recently used by the Alan Turing Institute considering AI in Financial Services, the target should be 'process transparency' rather than 'system transparency' (Ostmann & Dorobantu, 2021).

Q4.4.3. In what, if any, circumstances should exemptions apply to the compulsory transparency reporting requirement on the use of algorithms in decision-making for public authorities, government departments and government contractors using public data?

There would clearly be a need for certain exemptions to apply to a compulsory reporting requirement. But it is important that exemptions should be limited in two major respects, at the risk of undermining the purpose of signalling a policy 'reset':

- Criminal justice and immigration must not be subject to an exemption (except possibly on proven national security grounds), given their importance to trust dynamics.
- Government contractors must not be subject to an exemption based on commercial confidentiality, since detailed system information is not required.

References

Ada Lovelace Institute (2021) *Participatory data stewardship*. http://www.tandfonline.com/doi/abs/10.1080/01944366908977225

Ayling, J. and Chapman, A. (2021) Putting AI ethics to work: are the tools fit for purpose?. *AI and Ethics*, pp.1-25.

Biewer, S., et al. (2021) Conformance Relations and Hyperproperties for Doping Detection in Time and Space. *Logical Methods in Computer Science*. DOI: https://arxiv.org/pdf/2012.03910.pdf

Boardman & Nevola (2021) *Changing direction? UK consults reforms to its data protection law*. Bird & Bird http://www.twobirds.com/en/news/articles/2021/uk/changing-direction-uk-consults-reforms-to-its-data-protection-law

Borgo, R., Cashmore, M., & Magazzeni, D. (2018) T*owards Providing Explanations for AI Planner Decisions*. IJCAI 2018. IJCAI/ECAI 2018 Workshop on Explainable Artificial Intelligence (XAI). https://www.semanticscholar.org/paper/Towards-Providing-Explanations-for-AI-Planner-Borgo-Cashmore/9c6ac98ee83ddad5bf3758a7337e0222c9c9db9c

British Institute of International and Comparative Law & Dickson Poon School of Law, King's College London (2021) *Contesting automated decision making in legal practice: Views from practitioners and researchers*. https://www.biicl.org/blog/25/contesting-automated-decision-making-in-legal-practice-views-from-practitioners-and-researchers

Canal, G., et al. (2020) Building Trust in Human-Machine Partnerships. *Computer Law & Security Review*, 39, 105489. https://doi.org/10.1016/j.clsr.2020.105489

Canal, G., Krivic, S., Luff, P., & Coles, A. (2021, June 8) *Task Plan verbalisations with causal justifications*. ICAPS 2021 Workshop on Explainable AI Planning. https://openreview.net/forum?id=4T9oO_MrhEm

Cashmore, M., et al. (2019, July 12) *Towards explainable AI planning as a service*: 2nd ICAPS Workshop on Explainable Planning.

CMA. (2021) *Algorithms: How they can reduce competition and harm consumers*. GOV.UK. https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers

CMA (2021) *Research and analysis. Algorithms: How they can reduce competition and harm consumers*. Footnote 177, https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers.

CNIL (2018) *Privacy Impact Assessment* (PIA): Methodology, https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf

Danezis, G. et al., (2014) *Privacy and Data Protection by Design – from policy to engineering* (European Union Agency for Network and Information Security (ENISA))

DARPA (2018) *Explainable Artificial Intelligence*. https://www.darpa.mil/program/explainable-artificial-intelligence

Drake, A., et al. (2021) Legal contestation of artificial intelligence-related decision-making in the United Kingdom: Reflections for policy. I*nternational Review of Law, Computers & Technology*. https://doi.org/10.1080/13600869.2021.1999075 *(accepted 25 October 2021)*

Fox, M., Long, D., & Magazzeni, D. (2017) *Explainable Planning.* ArXiv:1709.10256 [Cs]. http://arxiv.org/abs/1709.10256

Heath, C., & Luff, P. (2018) The Naturalistic Experiment: Video and Organizational Interaction. *Organizational Research Methods*, 21(2), 466–488. https://doi.org/10.1177/1094428117747688

House of Lords Select Committee on Artificial Intelligence (2018) *AI in the UK: ready, willing and able*. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Hussain-Gambles. M, Atkin. K, & Leese. B. (2004) Why Ethnic Minority Groups Are Under-Represented in Clinical Trials: A Review of the Literature. *Health Soc Care Community*, 12(5), 382-388.

Huynh TD, Stalla-Bourdillon S and Moreau L. (2019) *Provenance-based Explanations for Automated Decisions* (Final IAA Project Report).

Huynh, TD. and et al., (2020) Addressing Regulatory Requirements on Explanations for Automated Decisions with Provenance—A Case Study. *Digit Gov*: Res Pract Article 16e

ICO (2020) Explaining decisions made with AI - Part 1: The basics of explaining AI (Project EpxlAIn, v1,0,44, 2 December), p. 62.

ICO (2019) Guide to the General Data Protection Regulation (GDPR). v. 1.0.654, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf

ICO (2021) *Response to DCMS consultation "Data: A new direction"*. https://ico.org.uk/media/about-the-ico/consultation-responses/4018588/dcms-consultation-response-20211006.pdf

ICO & Alan Turing Institute (2020) *Explaining decisions made with AI*. ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/

Jennings, N. R., et al. (2014) Human-agent collectives. *Association for Computing Machinery*. https://doi.org/10.1145/2629559.

Kuner, C. et al. (2017) The rise of cybersecurity and its impact on data protection, *International Data Privacy Law*, Volume 7, Issue 2, May 2017, Pages 73–75, https://doi.org/10.1093/idpl/ipx009

Keller, P. (2019) Participatory Accountability at the Dawn of Artificial Intelligence (SSRN Scholarly Paper ID 3448315). *Social Science Research Network.* https://papers.ssrn.com/abstract=3448315

Konstantinidis, G. Holt, J., & Chapman, A. (2021) *Fine-grained consent in databases*, *VLDB 2022*.

Konstantinidis, G., et al. (2020) The need for machine-processable agreements in health data management. *Algorithms*, 13 (4), 1-21.

Lau, A. (2017) *The Privacy Box*. Technical report, Public Interest Advocacy Centre.

Godel, M., Landzaat, W., and Perkins, R. (2019) *Data Pitch evaluatio*n. London Economics. https://datapitch.eu/london-economics-data-pitch-evaluation-final-pdf/

Mascalzoni, D., et al. (2010) Comparison of participant information and informed consent forms of five European studies in genetic isolated populations. *European Journal of Human Genetics*, *18*(3), pp. 296-302.

Miller, T. (2018) Explanation in Artificial Intelligence: Insights from the Social Sciences. ArXiv:1706.07269 [Cs]. http://arxiv.org/abs/1706.07269

Moreau, L. & Missier, P. (2013) PROV-DM: The PROV Data Model. *World Wide Web Consortium, W3C Recommendation. https://www.w3.org/TR/2013/REC-prov-dm-20130430/*

Ostmann, F., & Dorobantu, C. (2021) *AI in Financial Services*. Alan Turing Institute. https://doi.org/10.5281/zenodo.4916041

Open Data Institute (2015) *Open data means business: UK innovation across sectors and regions*. http://theodi.org/open-data-means-business-uk-innovation-sectors-regions

Pryor, G. (2012) *Managing research data.* London, UK, Facet Publishing.

Reeves, N. T., & Simperl, E. (2019). *Efficient, but Effective?: Volunteer Engagement in Short-term Virtual Citizen Science Projects*. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–35. https://doi.org/10.1145/3359279

Sion, L., et al., (2020) DPMF: A Modeling Framework for Data Protection by Design.*15 Enterprise Modelling and Information Systems Architectures*.DOI:
DOI: https://doi.org/10.18417/emisa.15.10

Thuermer, G., Walker, J., & Simperl, E. (2019). *Data Sharing Toolkit*. Data Pitch.

Thuermer, G. *Participatory Science Toolkit against Pollution* (ACTION) (2021), https://actionproject.eu/toolkit/.

TIGRR (2021) *Taskforce on Innovation, Growth and Regulatory Reform independent report* (pp. 52–53). https://www.gov.uk/government/publications/taskforce-on-innovation-growth-and-regulatory-reform-independent-report

Vilone, G., & Longo, L. (2020) Explainable Artificial Intelligence: A Systematic Review. ArXiv:2006.00093 [Cs]. http://arxiv.org/abs/2006.00093

Walker, J. et al. (2019) *Decision making processes for data sharing: a framework for data trusts*. ACM WomENcourage 2019: Celebration of Women in Computing, MAXII Rome, Rome, Italy. 16 - 18 Sep 2019

Naiseh, M., et al. (2021). Explainable recommendations and calibrated trust: two systematic user errors. *Computer*, 54(10), pp.28-37.