

**The
Alan Turing
Institute**

**Response to
the National
Data Strategy**

The Alan Turing Institute (“the Turing”) is the UK’s national centre for data science and artificial intelligence (AI). The Turing works alongside major regional, national and international organisations in order to deliver societal benefit from data related technologies, in line with its charitable objectives. The Turing comprises of over 350 academics from the UK’s leading universities: Birmingham, Bristol, Cambridge, Edinburgh, Exeter, Leeds, Manchester, Newcastle, Oxford, Queen Mary, Southampton, University College London, and Warwick; with major international research, development and innovation (RD&I) collaborations, and a convening power that is internationally unique.

The Turing’s position provides insights into the research landscape of data science and AI across academia and industry in the UK. This document provides the Turing’s response to the Government’s National Data Strategy consultation. The Turing’s response combines the perspectives of various Turing staff and researchers affiliated with the Turing, gathered through workshops facilitated around the questions that sit underneath each of the five missions. A list of individuals who contributed to this response can be found in the appendix.

Questions on the overall framework of the National Data Strategy

1. To what extent do you agree with the following statement: Taken as a whole, the missions and pillars of the National Data Strategy focus on the right priorities.

Overall

The stated missions and pillars of the National Data Strategy (NDS) cover a broad range of opportunities and concerns that need addressing. A consensus of across the Turing community was that the relationship between the five “concrete and significant opportunities” and the five “missions” is confusing, and that the five missions should more clearly respond to the five opportunities, and that the use of the four “pillars” could be better explained.

Focus

Given the data access issues that the UK’s response to the COVID-19 pandemic have demonstrated, the Turing felt that the focus of the NDS on use of data, rather than data access, needs to be addressed in the next iteration of the NDS. The UK’s use of data and IT in general during the pandemic has been poor, even by comparison with other GDPR countries. This is mirrored in The Royal Society's DELVE group’s report on "Data Readiness: Lessons from an Emergency", which states "Guidelines for data quality that do exist focus on intentions - or set out the desired end-point - rather than providing mechanisms for improving data accessibility". The aspirations of the NDS would be more realistically achieved if the narrative underpinning the NDS placed a greater emphasis on data access, rather than on data use, which should follow in due course.

Framing

The use of metaphors in the NDS, such as “unlocking the value of data,” “championing the flow of data,” or “data is knowledge”, obscures a clear understanding in the direction of travel for the NDS. For instance, the metaphor of “unlocking the value of data” might be better described as the goal of “unlocking the potential benefits of innovative and responsible data use.” Lack of clarity is likely to create hurdles for gaining public trust, as the public perception of the “big data revolution” is already centred around data extraction and exploitation to personalise advertisements over business models that earn revenue from delivering a product or a service of value, at the expense of individual and community wellbeing. Government should portray the promotion of data use that benefits impacted people and groups, as well as innovators and technologists.

Another framing issue can be found in the tension that is set up between a “pro-growth data regime” and a trusted but burdensome data regime. This description potentially portrays the false narratives that technology regulation, by its very nature, operates as a barrier to accelerated innovation - a narrative often used by tech companies and others seeking to avoid societally beneficial constraints and guardrails. Good governance and responsible innovation produce better data science and artificial intelligence (AI), and more efficient and effective data-

intensive applications, in a way that can help innovators to avoid mistake-ridden, “garbage-in-garbage out” outcomes that make data-driven technologies potentially wasteful and unsustainable.

Definitions

Phrases such as “fit-for-purpose”, “accessible” or “data quality” should be replaced with more precise equivalents, such as accuracy, validity, representativeness, robustness, and being up to date. The next iteration of the NDS would also benefit from a clearer definition of data and distinguish between different types of data, for example, digital, statistical, and other kinds of structured digital information. The NDS should also cater to a wider variety of data that may not be necessarily associated with individuals, for example satellite data, sensor data, process data, asset data, inventory data, etc.

Coherence

The pillars are said to be “highly interconnected”, but it is important to demonstrate what the interconnections are, how they work, and the mutual dependencies and directions of influence.

For example, ‘skills’ is acknowledged as a foundational pillar, however, this consultation does not specifically enquire about this. The Turing feels this is a significant omission, and outlines the most pressing challenges and opportunities in relation to skills and the NDS below:

- **Addressing the lack of agreed definitions and role descriptors in data science and AI**
A central issue in the national skills landscape is that of terminology and nomenclature. While the knowledge, skills and behaviours necessary for success in data science and AI are generally well understood, these are not always applied consistently. An increasing number of programmes in data science and AI across universities now exist; however, their focus and depth are variable, so their contributions towards the national skills deficit are difficult to assess. This ambiguity also perpetuates into the industry, where there is an observable inconsistent use of role descriptors, which complicates skills assessments. There is a need for unified ‘role descriptors’ for roles directly relating to, and adjacent to data science and AI. Advances in this area are critical if government is to effectively monitor the UK’s skills readiness in a robust way (Area Two).
- **Shaping curriculum standards and accreditation**
The Turing has prior success in shaping curricular standards around data science and AI, most notably the introduction of the Level 7 AI Apprenticeship standard, which is in strategic alignment with the forthcoming Industrial AI Masters scheme. The Turing, as a multi-disciplinary national Institute, can play an increasingly important role in unifying the curriculum landscape around data science and AI, because without strong leadership and coordination, discrepancies in definitions of data science and AI core knowledge, skills and behaviours will lead to inconsistent adoption of best practices. 2020 has also brought about revitalised discussions around the professionalisation of data-adjacent roles including ‘data steward’ and the data science occupation among the BCS, The National Physical Laboratory, The OR Society, The Royal Academy of Engineering, The Royal Society and RSS. The Turing could potentially take a coordinating role in

overseeing Chartership and accreditation activities of these and related bodies.

- **Addressing the data deficit around skills supply and demand, which restricts strategic planning and decision making**

The data skills landscape in the UK suffers from an acute data deficit, which restricts strategic planning and decision-making. This leads to challenges in assessing regional and sectoral disparities in training provision, as well as assessing restrictions on historically underrepresented groups. There is a pressing need to support the collection and analysis of high-quality data to help policymakers and providers to identify gaps and opportunities in skills training, in the form of a 'data skills observatory'.

- **Bringing alignment to leadership in the skills landscape**

Challenges in the skills landscape are unlikely to be resolved by industry or academia alone, motivating the need for interventions to stimulate deep and meaningful collaboration and knowledge sharing among these groups. The Data Skills Taskforce represents a forum which uniquely brings together industry, academia, learned societies and professional bodies with the aim of achieving strategic alignment in data skills. This has led to several substantial contributions:

- Collaborative project with DCMS and Opinium on 'Defining the Data Skills Gap' project.
- Work supporting the Office for AI (OAI) and Ipsos MORI in AI skills market analysis.
- The Turing, in collaboration with DCMS, playing a key role in the development of an online skills portal and a self-assessment tool to help organisations – particularly SMEs – to identify their organisational and technical readiness for data, and signpost them accordingly.

The Turing's co-chairing of the Data Skills Taskforce is evidence of the Turing's strong position as a convening power in the national skills landscape.

- **Data and AI education infrastructure as a national resource**

A significant barrier in data science education, and the adoption of data science in industry, is centred around the availability of compute infrastructure to support data science work. Interactive computing environments, including JupyterHub and RStudio Cloud are gaining significant traction in data science education; however, considerable expertise is required in making these tools available reliably and at scale to learners. Turing has significant expertise through its Tools, Practices and Systems Programme and Research Engineering Group, and could pursue such an interactive computing environment as a national resource to conventional learners as well as organisations. Prior efforts in this space include:

- Syzygy (<https://syzygy.ca/>) is a Canada-wide JupyterHub environment for Universities
- Digital Catapult Machine Intelligence Garage (<https://www.digicatapult.org.uk/for-startups/acceleration-programmes/machine-intelligence-garage>) providing Cloud and GPU resources to support adoption of machine intelligence.
- NSF CloudBank (<https://www.cloudbank.org/>) seeks to overcome procurement challenges around compute resources.

2. We are interested in examples of how data was or should have been used to deliver public benefits during the coronavirus (COVID-19) crisis, beyond its use directly in health and social care. Please give any examples that you can, including what, if anything, central government could do to build or develop them further.

The recent pandemic brought to the forefront the need for a data-driven understanding of policy interventions; not only in healthcare, but in society and the economy at large. Examples are presented below of public-purpose data science beyond health and social care:

Project Odysseus – understanding London’s busyness

Project Odysseus is a good example of rapidly and safely pivoting existing data science projects to address UK priority areas. A team of Turing researchers combined multiple large-scale and heterogeneous datasets capturing mobility, transportation and traffic activity in the city of London to better understand “busyness”.

This data included: JamCam cameras, traffic intersection monitors, point of sale counts, public transit activity metrics and aggregate GPS activity from the Turing's London air quality project. This big picture of urban mobility helps London authorities assess the effectiveness of the lockdown measures and compare with post-distancing readings in near real-time. The team is delivering an application programming interface (API) which the Greater London Authority, Transport for London, the London Data Commission, and the Office for National Statistics may query for access to the data sources, the analysis outputs and the statistical models.

<https://www.turing.ac.uk/research/research-projects/project-odysseus-understanding-london-busyness-and-exiting-lockdown>

RAMP – Rapid Assistance in Modelling the Pandemic

The RAMP (Rapid Assistance in Modelling the Pandemic) initiative is being led by the Royal Society to bring modelling expertise from a diverse range of disciplines to support the pandemic modelling community already working on COVID-19 in order to help guide the UK’s response. Through this approach, RAMP aims to enhance modelling capacity in time to create a clearer understanding of different exit strategies from the current lockdown. In providing support for existing research groups, the aspiration is that new models or insights can be created to help inform the work of the Government’s scientific advisors, through data science-based approaches and applying knowledge from related epidemiology domains.

Mark Birkin, Programme Director for the Urban Analytics Programme, was also asked to coordinate a key workstream. Using Devon for the initial calibration, it showed early promise in what was the volunteer phase by reproducing observed infection rates amongst synthetic individuals and assigned to geography, with transmissions modelled in accordance with individual behaviours and movement patterns. Feedback from UKRI stated that RAMP “will become a key national resource” and is now looking at ways to sustainably fund parts of the activity.

Health Foundation initiative on health-related misinformation

In collaboration with the Health Foundation, the Turing's Public Policy Programme commenced a project on health-related misinformation during the COVID-19 pandemic. The was a series of experiments to see which people are most vulnerable to which sorts of misinformation and working out what measures should be taken to minimise the risk. This activity is in the final stages of being able to share its initial findings

3. If applicable, please provide any comments about the potential impact of the proposals outlined in this consultation on individuals with a [protected characteristic](#) under the Equality Act 2010.

The Turing felt responses to this question would be adequately covered by submissions from other organisations or individuals.

4. We welcome any comments about the potential impact of the proposals outlined in this consultation on the UK across all areas, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK.

The NDS portrays the issue of fairness in terms of the potential to “harness data to reach people in need” or to “tackle bias and exclusion”. While this is an important dimension of principles-based data use, it does not consider the more fundamental and sizable problem of how to foster equitable conditions for fair data innovation.

Currently, the UK is hampered by conditions of societal inequality that have created digital divides, as well as entry barriers in resources, education and training for the average person in the AI and machine learning (ML) innovation ecosystems. When the NDS refers to “creating a fairer society for all,” it must therefore better address the way these conditions of structural inequality and asymmetrical resource allocation will be addressed.

Considerations of the enabling conditions for fairness-promoting data innovation environments require a more empirical and holistic approach on how factors of inequality and discrimination operate as determinants of fair outcomes in an increasingly digital society. The NDS should consider this aspect of building equity into technology and thus of fostering downstream social trust more than this iteration.

Mission 1: Unlocking the value of data across the economy

5. Which sectors have the most to gain from better data availability? Please choose all relevant options, below.

This list is based on the Standardised Industry Classification (SIC) framework for categorising work of organisations.

Charity or Non-Profit

Public Sector/Central or Local Government, including Defence

Financial and Insurance Activities

Information and Communication

Professional, Scientific and Technical Activities

Accommodation and Food Service Activities

Administrative and Support Service Activities

Agriculture, Forestry and Fishing

Arts, Entertainment and Recreation

Construction

Education

Electricity and Gas supply

Human Health and Social Work Activities

Manufacturing

Mining and Quarrying

Real Estate Activities

Transportation and Storage

Water Supply and Waste Management

Wholesale and Retail Trade

Other (please specify):

Please give any further details about your selections here:

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

6. What role do you think the central government should have in enabling better availability of data across the wider economy?

Accessible, high-quality, and well-archived data are critical in the progress of data scientific insights and AI/ML technologies, but responsibly opening data also involves privacy optimised, impact aware, and security compliant data sharing. Data sharing is a part of the 'data value chain', which is the process by which value increases as data is transformed into insights that lead to better decision-making and innovative products and services. The opportunities to create value increase when information is shared across organisational and international boundaries.

Therefore, managed accessibility and maximal data integrity is needed for trusted data to be more freely circulated among an ever-widening circle of responsible researchers and other actors so that results can be replicated, and new, societally beneficial insights produced.

- **The NDS should consider all the components of the data value chain**, which also include the data collection and curation process, the choice of data formats and catalogues, the data modelling and the ethical standards and norms that guide data sharing.
- **Government should catalyse and support the development of communities of practice for various data functions across the data value chain**, for example, data need identification, data management, analytics, etc., to promote better collaboration and coordination between local and central government.
- **A reflective and evidence-based, whole society approach to data readiness should be considered as essential to the NDS.** Currently, there is considerable variability in the data maturity of various sectors. There should be a greater emphasis on creating conditions of data readiness across the whole of society to enable adoption. The assessment of data readiness levels and capabilities in various industries should be the basis for a better, standardised and harmonised understanding of data usage rights and responsibilities.
- **The NDS needs a much greater emphasis on the ethical considerations of data sharing and the nuance about the different considerations for different types of data.** Linking data effectively while maintaining the privacy, robustness, security, and integrity of the subjects, as well as assessing the quality of data linkage, especially in siloed government databases are key enablers for data availability. Government should collaborate with organisations, such as the Turing, the Open Data Institute and the Ada Lovelace Institute, to help translate data ethics guidance to practical recommendations.

6a. How should this role vary across sectors and applications?

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

7. To what extent do you agree with the following statement: The government has a role in supporting data foundations in the wider economy.

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Please explain your answer here. If applicable, please indicate what you think this role should be.

The NDS should lay the foundations for more transparency and control, as well as tackle the practical and cultural barriers to sharing publicly and privately held data

These foundations include technologies and architectures that facilitate security or privacy and ease access to data while ensuring accountability; robust quality assurance processes; the development of new data standards; ensuring that various parties have the skills and resources to implement the above.

Government should stimulate the development of trusted research environments

Government should stimulate the development and testing of trusted and safe research environments with trusted organisations. These environments include technologies and architectures that facilitate security, privacy and robust data sharing and processing; the development of common methodologies, tools, practices, and systems across sectors; and demonstrating that data-sharing agreements can be used for creative exploration through testbeds and sandpits with consortia on pre-competitive topics.

UK academia has the expertise needed to encourage this kind of data collaboration across sectors. The Turing, as a national institute, has a trusted and impartial status to convene capability across the UK and direct activity towards UK priority areas. The Institute is uniquely placed to lead the adoption and standardisation of next generation approaches from the research, development and innovation frontier that are ready to become a new part of state of the art, accessible to data practitioners in government, industry and beyond. Examples of the Turing developing the foundations for secure data collaboration include:

- **developing technical standards and tools that facilitate data sharing and access.** Examples include: Open competitive testbed platforms for algorithmic innovation, for example Air Traffic Control – [project page](#) and [impact story](#)); New tools for reproducible analysis practices (e.g. the [Turing Way's](#) work with [Binder Hub](#))
- **connecting data to the compute capability that allows to turn data into insight**, such as information sharing protocols allowing intelligence to be distributed and linked across multiple data owners and the integration of model outputs as data sources for decision making, for example the National Digital Twin
- **developing techniques and tools that allow sensitive data to be worked with productively and securely**, for example, reproducible, reusable software-defined infrastructures for data innovation platforms (e.g. [Data Safe Haven](#)); and privacy-preserving analysis, such as secure federated analysis, privacy-preserving data access ([PDQ](#)) and [privacy-preserving machine learning](#) and the development and evaluation of techniques for encrypted computation (e.g. [SHEEP](#))

These responsible research innovations can become the foundational layer that tackles the barriers to sharing publicly and privately held data.

Government should further fund and support the research and development of privacy-enhancing technologies in the public sector and beyond

Government should further fund applied research on privacy enhancing technologies (PET). Government should also act as an early PET adopter through trials and testbeds. Government departments should identify new opportunities for safe analysis of sensitive datasets across their portfolios, test PETs tools and publicly share their learnings. There is also a need for government and regulators to provide guidelines about the use of PETs in mitigating privacy

and security risks. Furthermore, public sector organisations should be supported and have the level of expertise required to become intelligent users and buyers of PETs. The Information Commissioner's Office (ICO) together with the National Cyber Security Centre (NCSC) should galvanise the development of a community of PET practitioners, as well as disseminate advice and best practice about the maturity and suitability of various PET solutions. Finally, the Turing recommends that government works closely with the research community to convene and support the UK's innovation community developing PETs, to nurture a more dynamic innovation supply chain.

Government and regulators should commission work on data access and sharing such as standardising contractual clauses for data transfers and allocation of risk that are internationally and nationally agreed.

The allocation of risk in contractual agreements refers to who will access the data and how the data will be protected. These processes can be lengthy, especially when the nature of data is unclear and when the data is sensitive, personal and/or confidential. Enabling better access to data may require the standardisation of various data categories which will in turn enable a better assessment of the type of data protection that is required. Standardising terms of data access would significantly reduce friction in data sharing negotiations. This is particularly challenging for charities and Higher Education organisations that prefer not to have indemnities in contracts and frequently must go through an escalation procedure that takes time to navigate.

Contractual limits of liability are another point of contention. When transferring data sets between parties, often unreasonably high caps are included that intend to cover the perceived outcomes of data breaches. The liability levels are so high as to include potential fines from regulators while, in reality, the recoverable amounts under data breaches are much lower. This, again, hampers the conclusion of data sharing contracts. We welcome some standard clauses and/or guidance that could help mitigate this challenge.

Where the Lambert Agreements focused on the management of intellectual property between universities and businesses, nationally and internationally, a new set of model agreements and clauses needs to be created to address data sharing, nationally and internationally.

8. What could the central government do beyond existing schemes to tackle the barriers that small and medium sized enterprises (SMEs) face in using data effectively?

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

9. Beyond existing Smart Data plans, what, if any, further work do you think should be done to ensure that consumers' data is put to work for them?

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

Mission 2: Securing a pro-growth and trusted data regime

10. How can the UK's data protection framework remain fit for purpose in an increasingly digital and data driven age?

The pro-growth agenda underpins the emphasis on reducing the complexity and burden of existing regulatory frameworks, particularly for small and medium-sized enterprises. Simultaneously, there are discussions of the importance of empowering individuals to make informed choices around how they share their data. This places a strong emphasis on individual responsibility. A more suitable alternative would be to emphasise the role of regulation in safeguarding data and protecting individuals' data rights (collectively) rather than necessitating high levels of individual responsibility.

The Turing believes it is important to develop this capacity, both technical and non-technical skills, amongst regulatory bodies to ensure that they are well-equipped to respond to, and anticipate, developments in data science and AI. The Turing and OAI are currently working on a report exploring the opportunities around developing common capacity among the regulators and setting out approaches through which to do this. Knowledge sharing, collaborative working and development of shared tools will be important to ensure that effective regulatory regimes enable good data science and responsible innovation. Furthermore, this is needed to establish the regulatory certainty necessary for the development and adoption of AI.

The next iteration of the NDS should have a stronger emphasis on the importance of ethics and trust. While important ethical principles are noted, the NDS should draw on recent and ongoing research and practice, which is moving beyond ethical principles and developing guidance on ways of putting these into practice. For instance, the Government's official guidance, 'Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector', that was co-produced by the Turing, the OAI, and Government Digital Service (GDS), as well as the guidance on 'Explaining decisions made with AI', which was developed by the Turing in conjunction with the ICO. It is important that the NDS draws upon this existing work and others, to provide a clear set of principles and guidance around the ethical practice.

11. To what extent do you agree with the following statement: the functions for the Centre for Data Ethics and Innovation (CDEI) should be Artificial Intelligence (AI) monitoring, partnership working and piloting and testing potential interventions in the tech landscape?

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Please explain your answer here:

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

11a. How would a change to statutory status support the CDEI to deliver its remit?

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

Mission 3: Transforming government's use of data to drive efficiency and improve public services

12. We have identified five broad areas of work as part of our mission for enabling better use of data across government. We want to hear your views on any actions you think will have the biggest impact for transforming government's use of data.

As stated by the NDS, the COVID-19 pandemic has demonstrated the potential for the societally beneficial marshalling of digital information and data analysis. However, the pandemic has also exposed issues surrounding poorly designed data science and AI innovation practices, digital and material inequality, in the context of techno-solutionist approach. Key challenges remain in effectively unlocking the potential of transactional and administrative data across the public sector:

- **the increasing privatisation and outsourcing of technical skills within government has significantly reduced expertise** from government officials to analyse data and evaluate data science projects.
- **legacy contracts remain significant blockers as government must pay data providers and storers to use or further analyse data.** Proprietary data makes models inaccessible and hidden to the public, allowing for a wide array of ethical issues to emerge due to the lack of reproducibility. These data sets are often curated in a way that do not reflect real-world demographics.
- historically, **government has not used the transactional and administrative data available to it** in the design and delivery of public services.
- government departments often work in siloes, creating data sets, models, and policies for their respective areas. The pandemic has reinforced that this is far from optimal. **Government could have benefitted greatly from a joined-up modelling approach, where the epidemiological models were linked to the socio-economic models and able to run analyses on the system-wide impacts of policy interventions.** Such an approach requires careful thought. In future, work practices will need to be redesigned to

allow for linking up and the sharing of responsible innovation practices across the public sector.

The UK will be able to retain its place as a global pacesetter and thought leader on responsible innovation and robust digital technology governance by ensuring the next iteration of the NDS mandates responsible data collection and use practices, in the context of public distrust and societal polarisation.

Under four of the five areas of work identified in the NDS, the Turing recommends government:

1) **Quality, availability, access**

- **establish secure, smart, scalable, and interoperable long-term data sharing and analytics infrastructure;** Open data is insufficient. Most datasets on data.gov.uk are spreadsheets and PDFs, with metadata provided as unstructured HTML or PDF prose that is not machine-readable. Most repositories are not FAIR (findable, accessible, interoperable, reusable) best practice. Currently, very few datasets can be effectively discovered, reused, or combined with other datasets without significant human intervention. Government should work with dataset providers / creators, infrastructure providers, and researchers to develop and adopt best technical, policy and process practices for a FAIR data ecosystem in the public sector across the UK.
- **help mitigate security and privacy risks.** There is an increased security and privacy risk associated with sharing data across the public sector that need to be addressed. Data reidentification is a serious risk to individuals as data can be linked across a variety of sources, thereby creating a full profile on an individual. Before data sharing is investigated, security measures must first be in place, including, but not limited to encryption, server penetration tests, password protections, and strict access limitations.
- **reach beyond its own administrative datasets.** The private sector, while running their business operations, collects data that is sometimes referred to as “data exhaust”. This data should be within scope for consideration by government to collect, curate, store and make accessible for public sector use.

2) **Capability, leadership, and culture**

- **build capability, leadership, and data literacy** at all levels. Government should create common data science capacity for civil servants and regulators. This includes the provision of training programmes that aim to make the Civil Service AI literate, and a smarter buyer of IT services where data should be properly considered. An understanding of the technical aspects of data science and AI "at all staff levels" is also an understanding of the ethical implications of data sharing and biases that occur in both data sharing and modelling processes. Data ethics training especially related to data usage and sharing, should be made available to all staff levels. Highly technical data scientists and those sharing the data need to be trained on ethical protocols and standards.

3) **Accountability and productivity**

- **generate and cultivate public trust through transparency, accountability, and consent.** The success of any AI/ML innovation project depends on the degree of public

confidence in the safety and responsibility of the innovation has been established that is sufficient to foster its adoption by society at large, as well as the quality and performance of the product.

First, all AI/ML innovation projects should proceed with end-to-end transparency to establish that design, discovery, and implementation processes have been undertaken responsibly and that outcomes are appropriately explainable and can be conveyed in plain language to all affected parties. This entails the adoption of best practices mechanisms for responsible data sharing and the assurance of data integrity taking a holistic, end-to-end, and context-based approach to building AI/ML systems that are explainable-by-design.

Furthermore, all AI/ML innovation projects should proceed with end-to-end accountability to ensure both that humans are answerable for the parts they play across the entire AI/ML design, discovery, and implementation workflow and that the results of this work are traceable from start to finish. This will ensure public confidence that innovation processes prioritise patient and consumer interests from beginning to end.

Finally, members of civil society, domain experts, and other relevant stakeholders should be included in the AI/ML workflow through the institution of independent advisory consortia, which function as sounding boards as well as sense-checks and oversight mechanisms throughout innovation processes.

4) **Ethics and public trust**

- **put ethics at the heart of government data operations.** The Turing would welcome an ethical framework that is greater in scope than "transparency, safeguards and assurance" outlined in the NDS, which would establish new legal and regulatory frameworks to support the safe and ethical use of data across organisational boundaries. Also, upfront ethical standards and norms to guide the data sharing practices in government would avoid problematic situations.

Trust-building through community consultation should be utilised to foster the development of equal and respectful relationships - true partnerships - among researchers, healthcare professionals, and affected individuals and communities. Furthermore, public buy-in should come both from the groups in wider society that are impacted by the products of AI/ML innovation projects and from everyone who is directly affected using these products.

- **adopt ethical principles to create a shared vocabulary for balancing and prioritising conflicting values.** Resolving ethical dilemmas in a diverse society is dependent on building inclusive and well-informed consensus rather than appealing to higher authorities or traditions. This need for consensus-building is especially crucial in the context of AI/ML innovation, where circumstances often arise in which ethical values come into tension with each other. For instance, there may be situations, such as with digital contact tracing, in which the use of data-driven technologies may advance the public interest only at the cost of safeguarding certain dimensions of privacy and autonomy. Trade-offs, in cases like these are inevitable, but the choices being made between differing values should occur through

equitable deliberation, mutual understanding, and inclusive and knowledgeable communication.

To this end, it is especially important to set up procedural mechanisms that enable reciprocally respectful, sincere, and open dialogue about ethical challenges. These mechanisms should help conversation participants speak a common language so that, when an innovation project's potential social and ethical impacts are being assessed and re-assessed, diverging positions can be weighed understood, and suitably considered. The Government's official guidance, '*Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*', that was co-produced by Turing, OAI, and GDS, presents a framework for constructing the aforementioned mechanisms and directly addresses the importance of creating a shared, values-based vocabulary.

- **foster social mobilisation, public engagement, and community-involving participation.** The next iteration of the NDS should consider the aspects of social mobilisation, public engagement, and community-involving participation in data innovation processes. Whilst the NDS states that "the government must be willing to open itself up to scrutiny and increase public engagement", it is essential that this public engagement is not solely one-way communication and that it leads to meaningful impact through informing future practices to demonstrate trustworthiness.

Public engagement should not be pursued as a mechanism for building trust in government use of data but to understand the conditions needed to underpin this trust, and thereby develop and demonstrate trustworthy systems and approaches. Integrating citizens into the various missions of data use, from start to finish, will also not only improve the design and development of societally responsive innovation, but promote and ensure the widespread take-up of new digital technologies.

- **foster equitable data innovation and protect the interests of the vulnerable.** Patterns of social inequity, marginalisation, and injustice are often "baked in" to the data distributions on which AI/ML systems learn. The pandemic should be an inflection point to critically assess and redress elements of the digital divide that still define contemporary society and that help to perpetuate more widespread societal inequities. Policy makers, public officials, data scientists and AI/ML developers should convene with affected stakeholders to ascertain how to include these potentially left-out members of our communities in consequential policies, initiatives, and innovations.

13. The Data Standards Authority is working with a range of public sector and external organisations to create a pipeline of data standards and standard practices that should be adopted. We welcome your views on standards that should be prioritised, building on the standards which have already been recommended.

Government should catalyse the formation of an ecosystem of data owners, providers, and consumers to develop new data standards, such as the globally market-leading Open Banking Standard, especially in transition industries like agriculture and energy. The NDS set the expectation for similar initiatives, for example, Open Energy, Open Supply Chain, Open Manufacturing, Open Bids, Open Digital ID, Open Legal, Open Underwriting, etc. Opening these data monopolies will enable a more diverse ecosystem of organisations in data cleaning, brokering, aggregation, certification, and analytics to flourish.

Mission 4: Ensuring the Security and Resilience of Data infrastructure

14. What responsibilities and requirements should be placed on virtual or physical data infrastructure service providers to provide data security, continuity, and resilience of service supply?

Almost all dedicated data centres, whether on-premise, co-located, or in the cloud, must already comply with security standards, such as ISO27001. Similarly, most software-as-a-service (SaaS) offerings that sell to government or large organisations will also comply with such standards. However, many organisations that use these services will not have their dedicated information governance and data security experts. Therefore, providing clear and easy to follow best practice and guidance for these organisations on how to securely use and build on these data infrastructure and services is critical to maintaining the security of the data processed by these organisations, and NCSC's cloud security principals are one such good example of the type of guidance required.

The standards and guidance for secure and appropriate sharing data across organisations are much less mature than those for securing data within an organisation, and thus an area where regulators setting strong requirements would be useful.

In some instances, it is necessary to share identifiable information across organisations to provide the required service. However, this should be achieved by sharing only the minimum required data for the minimum required time. Some organisations do this well, providing secure federated or delegated access to only the necessary data, for example, the federation of patient records across the various organisations in the NHS and DVLA service for drivers to share their driving license information with third parties. However, many systems do not provide mechanisms to share data with other systems that incorporate strong access control, auditing, and data minimisation capabilities. This lack of technically capable APIs means that many organisations end up either not sharing data at all or inappropriately oversharing data in terms of the number of records or duration of retention.

In many other cases, the value of sharing data across organisations can be achieved without sharing identifiable or sensitive information at all, or at least sharing it only for the minimum time necessary to link records across organisations and generate a shared non-sensitive output. In both the public sector and in industry, there is a concern about organisations receiving access to the sensitive data of others to perform this linking, and this is a service that could be provided by trusted third party organisations, such as the ONS and regulatory bodies.

For personal data, public trust and informed end-user consent is also key to the wider sharing of data for the public good, and legal infrastructure such as data trusts and co-operatives can play a key role in securing this. Looking to the future, privacy-enhancing technologies have been developed for encrypted computation, secure federated analysis and privacy-preserving ML, providing mechanisms for organisations to analyse their combined data without sharing the sensitive data itself. These techniques need to be validated and further developed in real-world

contexts and government could play a key role here by working with academics to evaluate these emerging techniques in areas where data is already shared between government departments or organisations.

14a. How do clients assess the robustness of security protocols when choosing data infrastructure services? How do they ensure that providers are keeping up with those protocols during their contract?

There is an expectation of clients relying on providers achieving and maintaining current evidence of compliance with recommended security standards. However, not all data requires the same level of security. For example, data that has been pseudonymised, so there is a low risk of re-identification does not require the same level of security as the underlying identifiable data it was derived from. Summary statistics from the same data that are impossible to attribute to any individual require even lower levels of security. This range of sensitivity is not currently handled well in many sectors, resulting in limited sharing of even pseudonymised data across organisations.

This is an area where clearer guidance from regulatory bodies would be valuable. However, judging the privacy risk of a pseudonymised data set is very context-dependent, and there is no universally applicable privacy scale, making it difficult for data owners to make informed risk-based decisions. Better advice and guidance are needed, and this is an area where government could again play a key role as an early adopter practitioner working with academia to evaluate privacy and utility measures in real-world contexts and developing practical guidance in using these to make informed assessments of privacy risk.

Secure analysis environments for working with pseudonymised data may also help mitigate residual risk by restricting who can access the data and ensuring that only truly non-identifiable data is taken out of the secure environment. The security and resilience standards for such data does not need to be as high as for the most sensitive identifiable data. Still, these are less developed than those for the most sensitive data. Government should support those working on standards for environments for analysing such data and work to align these efforts across sectors.

15. Demand for external data storage and processing services is growing. In order to maintain high standards of security and resilience for the infrastructure on which data use relies, what should be the respective roles of government and data service providers, their supply chain, and their clients?

External providers of data infrastructure, such as cloud services and SaaS applications, have the scale to have larger teams dedicated to security and resilience than many, if not all, of their clients, and their business models rely on them maintaining high standards for both. In addition, clear and easy to follow good practice guidance should be provided for client organisations on how to use and build on external data infrastructure and services securely is critical to

maintaining the security of the data processed by these organisations, as outlined in Q14. Government can play a crucial role in setting standards here, not just as a regulator of these data infrastructure and service providers, but also as a key buyer of such services.

16. What are the most important risk factors in managing the security and resilience of the infrastructure on which data use relies? For example, the physical security of sites, the geographic location where data is stored, the diversity and actors in the market and supply chains, or other factors.

The Turing felt responses to this question would be adequately covered by submissions from other organisations and individuals.

17. To what extent do you agree with the following statement: The government should play a greater role in ensuring that data use does not negatively contribute to carbon usage?

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Please explain your answer here. If applicable, please indicate how the government can effectively ensure that data does not negatively contribute to carbon usage.

Government should play a key role in reducing the UK's carbon usage, and reductions in carbon usage for data should primarily be driven by the same measures that are used to reduce carbon usage more generally. For example, if using renewable energy is incentivised across sectors, it is expected that data centres would move to preferentially use such sources of energy. There may potentially be a role for government in making the energy cost of data storage and processing more visible to the clients of data infrastructure and their end customers, in a similar manner as has been achieved for energy efficiency for consumer goods market. However, given the variety of computational workloads and a lack of a common standard for the value of a compute unit, further work is required to understand the appropriate measure of energy efficiency or carbon usage that could be applied beyond the underlying energy or carbon cost per unit of compute or storage provided by the low-level infrastructure providers.

Mission 5: Championing the international flow of data

18. How can the UK improve on current international transfer mechanisms, while ensuring that the personal data of UK citizens is appropriately safeguarded?

Additional safeguards in place for personal data protection for citizens

While international data flows are needed for the security of citizens and nations, as well as to foster economic and social prosperity, proportionate and appropriate safeguards are essential. For instance, there is potential for cross-border data sharing that could be used for the exploitation of vulnerable people, such as the tens of thousands of victims of modern slavery. People are as vulnerable in cases when data about one person inadvertently also provides data about someone else without their knowledge or permission.

We encourage the government to involve and engage citizens in the collection and analysis of data on all realms of UK life, including environmental measurements and medical observations – the COVID-19 pandemic is an immediate example. The Turing has had success with engaging the public through citizen juries and other means. Citizen engagement should be a core component of the data strategy as the public is best placed to advise on their personal privacy concerns and to identify preferred safeguards.

Additionally, members of the Turing community have noted that the current way the UK articulates its international data sharing strategy fails to mention group rights or the UK Government's responsibility to non-citizens, and rather focusing on individual privacy and citizen rights.

Data-adequacy arrangements with key strategic partner countries

While currently, it is easier to work with some countries where there are specific arrangements, different countries and agencies have different approaches. Such inconsistency means that researchers are required to mobilise substantial amounts of funding to undertake projects to take advantage of data sharing policies and conditions which are subject to change. For example, there is a limited number of ways of broadly applicable, flexible ways of developing collaborative programs with many countries in the energy sector. Specifically, Turing researchers struggled to successfully collaborate with friendly, familiar organisations such as the US Department of Energy.

The UK needs suitable arrangements for sharing data quickly and establishing appropriate, flexible ways of developing collaborative programmes. Achieving this would involve establishing platforms which facilitate sharing and checking data quality. For instance, there is a need to ensure national security regarding international data transfers involved in larger-scale engineering policy questions.

Clarify the plans to support innovation with its data strategy

Turing researchers noted that the NDS can be abstract in places. For instance, with infrastructures that might underpin different types of international data sharing, be they national data repositories, or other arrangements. The Turing could contribute with suggestions on which parts of UK data infrastructure and training need urgent support to facilitate data reuse under safe conditions, including secure international transfer mechanisms with clear returns for the UK.

Improve on current data transfer arrangements by empowering citizens, researchers, and the commercial sector

The needs of the public sector, industry and academia can be quite different from one another with respect to their use data and the type of international collaborations they engage in. 'Empowerment' might take the form of safeguards and frameworks that do not burden researchers or business with the challenges of working with international partners on collaborations that involve data sharing. The Turing's own experience conducting 'Data Study Groups', which have involved data from multiple countries have consistently demonstrated the considerable time sink resulting from inadequate guidance for international data transfer. Government could help such situations by considering sector-specific incentives to overcome any apprehension caused by regulatory uncertainty.

19. What are your views on future UK data adequacy arrangements (e.g. which countries are priorities) and how can the UK work with stakeholders to ensure the best possible outcome for the UK?

Government should set out a clear mandate for UK values and principles around the international flow of data

The principles need to strike a balance between preserving privacy and public benefit. Such principles would, ideally, signal to prospective international partners the necessity of adherence to a level of standards, which should include and continue to improve upon the existing GDPR framework. Government should also engage with experts who have experience in international data governance, legal frameworks relating to data, IP, and trade rules. Through these efforts, it could develop criteria for data maturity and provide certification when organisations and agencies have met the set standards.

Government should promote international collaborations built on shared principles and trust

The ability to effectively share data internationally is central to being able to work collaboratively with other countries on shared challenges, such as the United Nations' Sustainable Development Goals (SDGs). Successful international collaborations, which involve shared data, require trust and must be guided by a shared, collaborative agenda. There are sectors and geographies whose high levels of risk aversion may hinder progress on significant global challenges, such as ending modern slavery.

The Turing recommends that government prioritises the support of such international collaborations to build infrastructure and competence, in an 'inclusive top-down approach' to safe and ethical international data sharing. Such an approach could also be complemented by UK government support in the localisation of data, of agendas and of development, an area where the Turing could provide technical support.

Additionally, there is a need for capacity building to share and to collaborate on the use of data to achieve a shared agenda. For instance, the shared ambition of different countries to monitor and tackle the defined SDGs faces considerable barriers having to do with the capacity of countries to organise data around these challenges. There is a role for the UK, and the Turing amongst others, in capacity building for developing infrastructure and competence for managing international data appropriately. It is also important to ensure that global data flows respond to local priorities, for example, collecting data related to SDGs could and should also inform local processes.

Government should acknowledge and minimise issues of uncertainty in a post-EU-exit and COVID-19 pandemic world.

Government should determine what potential safeguards are to be put in place post-EU-exit and what enabling data flows might mean in a crisis, such as public health emergencies or terrorist attacks. This will require legislative changes that allow for easier flows of data and entail the relaxing of data privacy for the limited period surrounding the crisis.

Authors

Allaine Cerwonka
Cosmina Dorobantu
Matt Forshaw
Christine Foster
David Leslie
Eirini Malliaraki
Amit Mulji
Martin O'Reilly
Charlie Thomas

Contributors

Mhairi Aitken
Joao Porto de Albuquerque
Morgan Briggs
Christopher Burr
Ian Carter
Charisma Choudhury
Jon Crowcroft
Chris Dent
Vanessa Forster
James Geddes
Nicolas Guernion
Zoe Kourtzi
Vanessa Lawrence
Sabina Leonelli
Helen Margetts
Carsten Maple
Anjali Mazumder
Ben Murton
Katrina Payne
Kate Robson Brown
Michael Rovatsos
James Robinson
Gavin Shaddick
Oliver Strickson
Hitesh Thakrar



[turing.ac.uk](https://www.turing.ac.uk)
[@turinginst](https://twitter.com/turinginst)