

Artificial Intelligence and Public Standards Report Follow Up

Name of organisation: Birmingham City Council

December 2023

1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible.	No, not at the moment.
2. How do you demonstrate that you are using AI in ways that are legal and legitimate?	In the event that we use AI, we will utilise existing risk assessment processes such as DPIA's or other governance mechanisms to document that consideration has been given to ensuring that AI is using data in a manner that is legal and legitimate.
3. How do you assess the potential impact of a proposed AI system on public standards (e.g., openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?	We do not consider that AI would replace the human being in the process and therefore would expect that the human being is responsible for the decisions taken and would be accountable for ensuring public standards on openness, accountability and objectivity. The use of AI in the Council should also be transparent to demonstrate openness.
4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.	This is still an emerging issue and challenge within the AI space, and we would expect to continue to engage with the sector to understand the challenges that are faced by organisations when using AI to ensure that bias and discrimination do not affect decision making in a manner that causes harm or detriment to any group or groups of people. Regulatory practice, professional adherence to ethical AI principles and new national and international legislation will assist in the regard.
5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?	We do not see that responsibility for AI systems is any different from responsibilities given to staff within the LA for things such as financial and human resources and as such would expect existing safeguards to support responsibility and accountability obligations in the use of AI.
6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?	An AI system would not be treated any differently from the use or operation of any other system in operation. It would be managed in accordance with best practice and standards available within the sector to ensure that it operates in the manner intended.
7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?	Whilst we do not currently use AI systems, we have in place an Information Assurance Framework and Information Assurance Board, chaired by the SIRO, to provide oversight and assurance on the use and management of data – this board would govern the use of AI and raise any risks / concerns to senior leaders, as necessary.

<p>8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?</p>	<p>Not applicable at this stage. However, any future adoption would seek to ensure that any shift in changes to business processes where AI plays a part would need to incorporate the ability for individual to challenge decision making processes, appeal against decisions and, in accordance, with existing laws, have the right to seek redress.</p>
<p>9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?</p>	<p>Employees would be expected to undergo training in the use of AI as they would with the introduction of any new system / technology that affects the work they do and the decisions that they take.</p>

Cornwall Council (December 2023)

Cornwall Council is not currently using AI to make nor aid decision making. Cornwall Council recognises the potential benefits that AI may bring in the future as well as the inherent risks that it brings. As such Cornwall Council is taking a cautious, case by case and risk based approach in its analysis, investigations and testing the safe future use of AI enabled solutions within its business processes. The Council is establishing core AI principles based around the commonly accepted areas of:

- **Safety, security, and robustness** - AI systems should function in a robust, secure and safe way throughout the AI life cycle, and risks should be continually identified, assessed and managed. AI systems are technically secure, and function reliably as intended throughout their entire life cycle.
- **Appropriate transparency and 'explainability'** - Transparency refers to the communication of appropriate information about an AI system to relevant people (for example, information on how, when, and for which purposes an AI system is being used). 'Explainability' refers to the extent to which it is possible for relevant parties to access, interpret and understand the decision-making processes of an AI system.
- **Fairness** - AI systems should not undermine the legal rights of individuals or organisations, discriminate unfairly against individuals, or create unfair outcomes. Actors involved in all stages of the AI life cycle should consider definitions of fairness that are appropriate to a system's use, outcomes and the application of relevant law.
- **Accountability and governance** - Governance measures should be in place to ensure effective oversight of the supply and use of AI systems, with clear lines of accountability established across the AI life cycle.
- **Contestability and redress** - Where appropriate, users, impacted third parties and actors in the AI life cycle should be able to contest an AI decision or outcome that is harmful or creates material risk of harm.

Cornwall Council is actively developing appropriate policy and governance frameworks which are focussed initially around Generative AI use and ethics including legal, compliance and legitimacy. This includes improving organisation wide understanding, confirming individual responsibilities and accountabilities as well as how to use generative AI capabilities safely maintaining a human in the loop (and accountability) principle for all content generated. The Council is engaging with independent industry experts to assist with and assure our approach.

Public Bodies Progress Update

Name of organisation: Crown Commercial Service (CCS)

November 2023

1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible. If not, please send us a nil return. If yes, please answer questions 2-9.

CCS is at the beginning of its journey with AI. Thus far, CCS has initiated projects to help provide insight to support effective customer engagement and provide a better level of service. CCS has not implemented any automated decision-making.

As the largest public procurement organisation in the UK, CCS has datasets capturing the current and historical procurement activity of tens of thousands of primarily public sector organisations that can use CCS commercial agreements in line with the Public Contracts Regulations (our customers) and supplier organisations. These datasets provide an opportunity to leverage AI to help CCS provide a better and more efficient service. Please see below for details of some of the early CCS AI projects:

- a recommendation system that can be used to generate insight for customers on which CCS commercial agreements might suit their needs. Please note the recommendations are based on the buying behaviour of organisations and not individuals
- experimenting with chatbot and large language model technology to help serve our customers more efficiently while coping with increasing contact volumes
- anomaly detection of supplier reports - helping to decide which suppliers may require auditing
- invoice payment prediction - to help identify organisations that may not pay their next invoice on time, ranked by financial risk

CCS has plans to increase the use of AI to support the organisation's business needs and allow us to deliver a better service to our customers and suppliers.

2. How do you demonstrate that you are using AI in ways that are legal and legitimate?

CCS is committed to the legal and legitimate use of AI. CCS does not use data of natural persons or handle personal data in the AI models, instead using information relating to organisations. CCS has a clear understanding of the implementation of the UK Data Ethics Framework and adheres to the terms and conditions set out within it.

3. How do you assess the potential impact of a proposed AI system on public standards (eg. openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

CCS does not use AI systems to make autonomous decisions, instead using AI as a decision support tool. Workings and outputs are reviewed together with domain experts and

any risks are treated appropriately. All decisions are reviewed to ensure compliance with the Nolan Principles.

4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.

CCS captures data regarding organisations and does not deal with natural persons or handle personal data. CCS reduces any bias in the AI models by ensuring that only features relevant to the problem are taken into account and that the underlying data used to train the model is as objective as possible. Additionally, previous outcomes from AI systems are not used as inputs to prevent any feedback effects that may result in bias based on past decisions.

5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?

CCS currently has a small team responsible for the development of AI systems, each with clearly defined roles and responsibilities.

6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?

When a new AI system is developed or retrained, if it is supervised (trained on labelled data), CCS ensures that a hidden test dataset is separated to measure performance against. If an algorithm is unsupervised (no labelled training data), CCS gauges effectiveness based on the assessment of colleagues with the appropriate domain expertise. If there is any deviation from the expected outcome, this is fully investigated before release.

7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

Currently oversight and scrutiny is dealt with by the CCS team tasked with developing AI systems. CCS aims to have an oversight system in place as part of the delivery of our AI strategy, due by the end of financial year 2023/24.

8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?

This has not yet been required as CCS do not have any AI systems making automated decisions without human oversight.

9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?

CCS's AI team engages in continuous professional development and works closely with data scientists across government to ensure they are aware of potential risks and how to deal with them appropriately.

Department for Transport (September 2023)

1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible. If no, please send us a nil return. If yes, please answer questions 2-9.

Algorithms are used in various analytical processes and in modelling in the Department for Transport. These inform decision making but ultimately decisions are taken by Ministers, not automated. The department publishes a list of business-critical models that provides details of their purposes, the [latest version is available here](#).

The department is developing an internal AI programme which will use algorithmic processes, for example to detect fraudulent grant applications and streamline work processes. To date none of these developmental systems have been used for automated decision making and all processes envisioned will retain a human in the loop.

2. How do you demonstrate that you are using AI in ways that are legal and legitimate?

Any processing of personal data via AI needs to comply with all applicable UK law relating to the processing of personal data and privacy, including but not limited to the UK GDPR and the Data Protection Act 2018.

As AI is a new and innovative technology, a Data Protection Impact Assessment must be completed and signed-off prior to any processing of any personal data via AI. This process enables us to demonstrate compliance with our data protection obligations, identify an appropriate lawful basis and minimise any risks arising, with input from the department's Data Protection Officer and legal advisors.

The Equality Act 2010 also places obligations on organisations in relation to fairness and discrimination which are relevant to the use of AI. An Equality Impact Assessment can help to demonstrate compliance with these obligations.

3. How do you assess the potential impact of a proposed AI system on public standards (eg. openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

As the department explores the creation and use of complex algorithmic tools, it will aim to do so in line with the Government's [Algorithmic Transparency Reporting Standard](#) (ATRS).

The ATRS requires consideration of accountability, rationale for use, impacts on decision-making, risks, risk mitigations and impact assessments. The ATRS is a maturing standard which is being progressively promoted and adopted across the public sector. Reports from its pilot are available [here](#).

Under Data Protection law we also have transparency obligations. At a high level, we need to provide individuals with clear, meaningful information about why we're processing their data, the legal basis for processing it, who we may be sharing it with and how long we're keeping it.

4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.

Bias in AI systems is normally a symptom of bias within the data on which it was trained. Prior to training an AI system, the training data is first assessed to ensure that it provides a balanced, proportional representation of all data sub-groups, without any being under or over represented. Any issues found are remedied through an appropriate method (such as sourcing additional data for any sub-groups that are under-represented).

Once an AI system has been trained, a second assessment for bias is made that is focussed on its predictions. This involves interrogating the predictions to identify any divergence between sub-groups.

5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?

The DPIA process requires a project lead, with responsibility for the AI system, to be identified and documented.

Oversight of internal projects is provided by the department's Data and AI Board. Independent external experts are being recruited to join this board, with the intention that they will provide impartial assessment of how responsibilities for DfT's AI systems are being exercised appropriately.

If a new dataset is created for the use of this project, an information asset owner, with appropriate training, must also be documented. Equally if an existing information asset is used by AI, the IAO with responsibility for that asset would need to approve its use.

6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?

An appropriate performance metric is defined prior to the development of an AI system that provides a measure of how well the system performs against its central objective. This metric is used to select the most performant AI methodology and to provide an indication of how well the system will perform in a real-world setting. The same performance metric is monitored continuously once the system is operational, along with a wider set of measures that will provide early indication of any issues.

7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

The DfT has established an AI board for developing and delivering the Transport AI strategy Strategy, Data and AI Board for organisational AI governance, and additional measures are being considered throughout the commercial life cycle to better facilitate this.

AI algorithms also go through the same governance and assurance processes set up for all traditional coded models and bespoke software development. Technology Governance is through the Architectural Change Board where designs are assured at design and production stages of the development lifecycle. Information Assurance, Data protection, and cyber security is assured at ACB also. Scalability, maintainability and accessibility is also reviewed. Change Assurance Board is used to support changes to services and products in live production.

Where appropriate, high cost and high value products and services, including any future AI models would go through Spend Control and GDS assessments, especially if available to the general public.

Existing technology governance evolves to incorporate any new HMG standards on digital, including AI. AI principles and standards are incorporated into overall technology principles, policies and standards.

8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?

Data protection law applies to all automated decisions and profiling involving personal data and there are additional rules that apply to protect individuals from solely automated decisions that may have significant legal or similar effects. If any decisions of this nature were implemented, we would ensure that our privacy information provided meaningful information about the processing, the rights that applied and how to exercise them.

9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?

We are developing an AI skills and awareness capability plan. This aims to develop training pathways for the various levels of engagement with AI we expect from our officials and ministers, from intelligent customer, all the way to AI data scientist. The aim is to have a robust and rolled out programme for the department by 2024, using existing processes where possible, but not limited to what is currently available, and tailored to the DfT context. This includes specialist training on data ethics which will be provided by the Open Data Institute (ODI). We have brought in two specialist AI PhDs from the University of Southampton and Heriot Watt University to support this work. Specialist areas such as commercial and digital are in the process of identifying additional training needs for their functions and will roll out the relevant training for digital platforms and procurement processes as required.

The DfT has initiated work with the ODI to train a number of staff in DfT as data ethics practitioners. This will allow us to explore some of the data ethics issues facing the department and develop tools and capabilities to constructively discuss data ethics. There is a strong link between data ethics and ethics around AI, as such we will continue to collaborate across the Department on data ethical issues.

The [PPN 06/23 The Commercial Playbooks](#) published in June encourages the use of the DDaT Playbook (which now has a section on AI) when undertaking commercial activities.

DWP PROGRESS UPDATE ON THE ARTIFICIAL INTELLIGENCE AND PUBLIC STANDARDS REPORT

Date received: 2 October 2023

Are you currently using AI to aid decision-making in your organisation?

Yes, the Department for Work and Pensions (DWP) is currently using AI to aid our decision-making. DWP is exploring the use of AI and its potential to support providing more digital services with a human touch in a safe, ethical and considered way. DWP will not automatically determine or deny a payment to a claimant using Artificial Intelligence.

We currently use AI in these areas:

Transaction Risking (TxR). This is a modern standard segmentation engine that looks for potentially fraudulent new claims. It triages inbound UC claims and directs potentially fraudulent claims to a queue for human investigation.

Intelligent Automation Garage (IAG). The IAG uses Robotic Process Automation (RPA) to process Back Office work. When a need for a benefit is not straightforward a multi-disciplinary team evaluates how best to meet the needs of the customer. This is complex and very time consuming - collating and redacting information. We have created automation that now does this for us and what used to take up to three hours per case, is now just a few minutes. Enabled by the IAG, we are reducing mundane manual processing such as processing inbound mail or self-serving simple queries e.g., bundle builder and security advice chatbot.

Job Matching Pilot. Here we are experimenting to understand whether Machine Learning and Natural Language Processing can improve the success rates of customers finding new jobs. We have run pilots assessing how well certain tools can assist claimants and work coaches to find suitable vacancies. Work is underway in UC to gather data on claimants' skills and barriers to help differentiate our services based on labour market proximity segments.

Identifying Vulnerable Customers Quickly. We receive and open tens of thousands of letters a day, which are then read to identify particularly vulnerable people who need our help the most and the fastest. We have built technology that can now read this mail, identify vulnerability, and flag them as the highest priority. Our colleagues can then action these first and make faster decisions to help people more quickly. We can now do this with around 21,000 letters, typed, handwritten and emailed every day.

Alerting to Risks of Harm. This includes scanning inbound contact channels and alerting to potential risks of harm. This happens on both the UC journals and inbound whitelist.

Technology Services maintenance. Our digital infrastructure and networks are scaling the use of AI to optimise availability, predict failures and understand usage patterns.

In addition, DWP is starting to explore the applicability of Generative AI (GenAI) in line with the Prime Minister's Foundation Model Taskforce. We have labelled this an AI Lighthouse Programme and we are examining several options from a range of candidate areas across DWP to enable the safe, considered acceleration of AI and innovation. All candidates share the core principle of having a human in the loop. The concepts we are assessing include:

Heritage System Migration. DWP has proven that it is possible to use GenAI to determine the business logic contained in heritage IT platforms that may not have up-to-date documentation. The next phase of this work is to use this logic and GenAI to write code for this business logic on modern cloud platforms.

Service Centre Agents. How do we break the supply/demand relationship and help existing service centre agents be more productive. Could AI help answer complex policy/guidance questions and/or summarise the call afterwards?

Work Coach Support. We are short of Work Coaches and there is a high volume of inexperience. How do we help put them on the front foot with different cases and make recommendations in the short time they have?

Policy Briefings. Policy teams get a lot of short notice demands for detailed briefings – can we better utilise our corporate knowledge and pull together content for different tasks on-demand?

Colleague Access. Powerful new tools are available in the market. Can we identify appropriate specific uses that we can deploy safely and quickly?

Counter-Fraud. How do we tackle the threats of misinformation and misuse of AI against DWP such as impersonating voices/false images, conduct penetration testing and prepare for more advanced security and fraud challenges?

Onboarding Training. We have a significant recruiting and onboarding challenge ahead of use (c. 20K new frontline staff this FY). Can we leverage Generative AI for Onboarding & Training (broadly across the DWP People & Capability and Learning & Development domains)?

Analysis. Can we improve the speed of Analysis and Forecasting (Finance Group and Analytical Community) and automate financing?

How do you demonstrate that you are using AI in ways that are legal and legitimate?

We have a clear statement of how we use Artificial Intelligence (AI) in DWP in our public-facing [Personal Information Charter \(PIC\)](#). *Specifically in relation to AI, we state: Artificial intelligence (AI) is the use of digital technology to create systems capable of assisting or performing tasks commonly thought to require intelligence. DWP uses AI to help detect and prevent fraud and error. DWP does not use AI to replace human judgement to determine or deny a payment to a claimant.*

In addition, we have good governance and a clear set of principles that we adhere to

1) Our Governance

We have three pillars of governance in place:

Governance – focusing on strategy, guidelines, and standards in coordination with cross-government bodies, our Chief Data Officer holds governance accountability and oversight is provided from the AI Steering Board which has senior representation from the wider department and the AI Assurance and Advisory Group provides independent risk assurance with a range of representation including risk management, legal, data protection, ethics, security, commercial. The AI Steering Board in turn reports into the DWP Data Board. Both bodies have an agreed set of Terms of Reference that has been reviewed and agreed by the AI Steering Board. DWP's Digital Board is also periodically consulted on the progress of our AI projects. The AI Steering Board also manages expenditure on Lighthouse projects (see below) The AI Steering Board has agreed (Sept 2023) to maintain a register of AI projects across DWP.

Enablement – to enable progress by managing risks, selecting vendors and engaging with communities of interest, our Chief Technical Officer is accountable and is supported by a Centre of Excellence; this multidisciplinary team leads on the development of Generative AI, assesses propositions and impacts possible solutions.

Transformation – to bring about lasting benefits, DWP uses a hub and spoke model with **Lighthouse projects** within projects teams are supported by AI Steering Board, AI Assurance Group, Centre of Excellence to align and empower multi-disciplinary teams. Each project is accountable to a Digital Director and a Product/Service Owner Director who in turn include Lighthouse Projects in their programme governance.

2) Generative Lighthouse Approach

DWP has created the Generative AI Lighthouse Programme to safely explore our use of emerging AI technology.

The programme is developing 5 pilot projects using cutting-edge generative AI during the 2023-24 financial year. Inherent in this approach is the oversight by the AI Steering Board (above) which has representation from across the Department.

Generative AI Lighthouse programme enables clear governance, enablement and transformation to progress AI safely, using the right capabilities and technologies. We are looking at real world problems and using concepts to see if they can tackle them.

The Generative AI Lighthouse programme has been recognised by Government's Central Digital and Data Office as an exemplar of how to safely test and accelerate the use of AI in public services.

3) Our Overarching Principles:

We have a set of principles to guide our work and they form part of the DWP Data Strategy and are overseen through an independent AI Advisory & Assurance Group, the DWP AI Steering Board, the DWP Data Board and up to the Executive Team.

- **Human Oversight of Decision Making:** DWP will not automatically determine or deny a payment to a claimant using Artificial Intelligence.
- **Strategic Impact.** We will ensure that any major investment and use of AI technologies will be specifically applied to help progress our strategic priorities rather than simply automating existing services.
- **Fully Explainable.** Any outputs from our use of AI technology should be fully explainable for governance purposes and that accountability for the accuracy and fairness of the system can be maintained throughout its use.
- **Mandatory Safeguards.** Where there is a potential risk of harm or discrimination to claimants, existing safeguards to prevent negative consequences should be updated with new safeguards to mitigate this.
- **Constant Evaluation.** We will regularly test, learn, adapt, monitor, and evaluate our approach and application of these technologies to check they continue to meet both the business requirement and user needs.
- **Data Protection.** We will safeguard the privacy of the individual in our use of AI in line with data protection rules. We may share data with approved third parties acting on our behalf, but this data will not be used outside of the agreed purpose by those partners for their own gain.
- **Ethical, Informed & Transparent.** We will underpin all deployment decisions taken on AI using a clear governance and decision framework to ensure its development is ethical, informed, and transparent. This includes being informed and supported by consultation with independent bodies, academic experts, and research.
- **Standardisation through the Supply Chain.** We will require comparable standards and ethical safeguards of our contracted providers and be transparent in our use of partners.

4) Risk and assurance

We use the 12 challenges in governance as set out by CDDO to guide our governance. We've included the Department of Science, Innovation and Technology's risk register into our AI Assurance & Advisory Group's framework to ensure close alignment with governmental risk and assurance.

How do you assess the potential impact of a proposed AI system on public standards (e.g., openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

DWP have reviewed their Data Protection Impact Assessment (DPIA) against the Information Commissioner's Office (ICO) AI toolkit.

DWP have subsequently updated the DPIA to include additional considerations for AI, such as fairness, meaningful human review, transparency, and 'explainability' as it relates to automated processing. DWP's Advanced Analytics function has a defined, repeatable governance process for the lifecycle of data models.

It is critical that DWP maintains a reputation for responsible use of AI to retain the trust of customers and regulators. When working in Fraud and Error (F&E), the ICO recognise it is not possible for us to be fully transparent as this undermines our ability to identify fraud because full transparency would allow fraudsters to understand how to counter our capabilities. However, our key strategy is ensuring that we continue to adhere to our AI Principles and Personal Information Charter.

DWP has built strong relationships internally with DWP Legal and Data Protection colleagues and externally with regulators including the Information Commissioners' Office (ICO), the NAO and the Government Internal Audit Agency (GIAA) to ensure legal and public standards are adhered to as appropriate.

How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.

As part of implementing instances of AI, teams must complete an "Equality Analysis" assessment which considers aspects of the Cabinet Office's ethics framework and is used to assess whether new models contain bias.

This is then reviewed by the independent AI Advisory & Assurance Board and reported to the Data Board.

All models are subject to Equality Analysis throughout the model's lifecycle. We proactively engage with legal and data protection teams within DWP to effectively mitigate any risks.

DWP has committed to report annually to Parliament on its assessment of the impact of data analytics on protected groups and vulnerable claimants.

How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?

We are currently developing a live list of all AI systems in use and how they are used by the business. This will be made available to senior stakeholders to ensure visibility by October 2023.

We will be developing training for colleagues using these platforms to explain how and where AI is being used.

How do you monitor and evaluate your AI systems to ensure they always operate as intended?

All DWP AI models are subject to rigorous testing and monitoring for bias and accuracy by DWP data scientists and the outcomes are monitored continuously by trained DWP employees and DWP data scientists.

In the past year we have continued to monitor this fairness and impact on customers, especially for Protected Characteristics. Currently this monitoring is done manually but we have projects in place to automate and to provide alerting where appropriate.

The department continues to work closely with the Data Protection Office (DPO) and the National Audit Office (NAO) to ensure appropriate safeguards are in place for the proportionate, ethical, and legal use of data with internal monitoring protocols adhered to.

What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

DWP has built strong relationships internally with DWP Legal and Data Protection colleagues and externally with regulators including the ICO, the NAO and the Government Internal Audit Agency (GIAA) to ensure we have appropriate oversight and transparency.

Our legal stance is to not use AI for automated decision making. A 'human in the loop' is critical and, therefore, we have formally declared "DWP does not use AI to replace human judgement to determine or deny a payment to a claimant." ([Personal information charter - Department for Work and Pensions - GOV.UK \(www.gov.uk\)](#)).

AI models must have a Data Protection Impact Assessment (DPIA) and Equality Analysis (EA). These documents contribute to a wider legal risk assessment that is conducted before AI models can be operationalised.

The Social Security Act 1998 Section 3 allows DWP to pool information from customers and other sources for the purpose of functions relating to social security. Any external data comes with other legal protections from the data controller and other broader legislation like the Digital Economy Act or the Data Protection Act.

The main legal data protection legislation, for AI or any use of data, is the UK General Data Protection Regulation (UK GDPR). GDPR Art 6(1)(e) allows DWP to process data as it is necessary for the performance of a task carried out in the public interest.

[How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?](#)

DWP does not use AI to replace human judgement to determine or deny a payment to a claimant.

A final decision in these circumstances always involves a human agent. Human agent decisions can already be challenged through existing routes.

[Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?](#)

Not yet, but a data literacy programme is planned and will include AI and ethical risks. The One Big Thing training across the Civil Service is the start of this.

Department of Energy Security and Net Zero (September 2023)

It is a nil return from the Department of Energy Security and Net Zero. Though our Analysts and Chief Digital Information Officer would like to note that the Department's Analysts use machine learning techniques where appropriate as part of analytical support to policy development. However, there is currently no use of such algorithms for automated decision making.

Department of Health and Social Care (November 2023)

DHSC does not use AI directly in decision-making. However, AI techniques are now additional tools available to the department and we aim to use them robustly, efficiently and ethically in developing the best insights and processes. DHSC has a mature analysis function with growing capacity and capability in data science, which is allowing us, for example, to explore machine learning models for more accurate public health statistics and to derive trends from unstructured text-based reports to identify areas for further investigation.

Foreign, Commonwealth and Development Office (August 2023)

The FCDO is at an early stage of exploring the potential of AI, and associated technologies, to augment trade-craft, decision-making and service-delivery across our diplomatic, development, consular and corporate operations. The FCDO believes that technology and data have the potential to transform both how we operate and how we deliver impactful outcomes across the globe increasing our efficiency and agility and giving us a competitive advantage in an increasingly complex world. The opportunities these technologies present are varied and span the breadth of the FCDO's work from providing new channels through which to provide Consular services to British citizens and streamlining corporate processes to increasing our ability to seamlessly operate across cultures or enabling staff to model and simulate complex global events and foreign policy scenarios.

The FCDO is working with key partners and others across HMG, including the Central Digital and Data Office (CDDO) in Cabinet Office and the National Cyber Security Centre (NCSC), both to explore the potential of AI and to ensure that any future uses of AI are appropriately assured and governed and that their use is compliant with relevant legislation.

The FCDO is delivering a multi-year Digital, Data and Technology Strategy that is putting in the technological and structural foundations that will enable us to securely exploit these technologies and ensure they are appropriately governed. As part of this, the FCDO has recently established a new dedicated innovation capability (FCDO-x) to enable the FCDO to explore these opportunities and to partner and collaborate with others across the wider HMG innovation ecosystem.

We would welcome the opportunity to learn from the findings of this exercise to ensure we can adopt best practice with regards to AI governance.

**Questions from Lord Evans of Weardale
Chair, Committee on Standards in Public Life
Artificial Intelligence and Public Standards Report Follow Up**

Name of Organisation: HMRC

1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible. If no, please send us a nil return. If yes, please answer questions 2-9.

Yes, HMRC uses both internally created AI outputs as well as those developed by our suppliers to aid decision making, however all these decisions, be they around tax liabilities, debt management, appeals or compliance actions, ultimately end with a human approved decision.

Internally we have used AI tools and techniques for many years, ensuring they are both explainable and transparent. Uses include:

- improving compliance targeting, by predicting taxpayers most likely to be non-compliant that helps identify the riskiest cases faster and improves yield.
- analysing huge volumes of documents, utilising techniques such as anomaly detection, image recognition, and network visualisation.
- improving debt collection, identifying customers who are most likely to get into debt and then proactively supporting them.
- utilising natural language processing to analyse customer contact data and call transcripts, identifying key themes and understanding customer sentiment and satisfaction, along with machine learning techniques to categorise and group feedback responses – using this insight to improve services.

Examples of 3rd party solutions include:

- We're using AI to improve the services we offer to our customers by helping them better navigate our website for speedier resolution to their queries. HMRC's digital assistant automatically helps customers to find the information they are looking for, linking them to an adviser through webchat if it can't locate the answer.
- We've also used AI to help with the digitisation and destruction of physical microfiche and microfilm records, as well as delivering a digital storage and retrieval system.
- Some of the AI solutions are embedded in our existing tools for example we use a 3rd party AI speech-to-text solutions for colleagues who have difficulties with hearing.

Questions from Lord Evans of Weardale
Chair, Committee on Standards in Public Life
Artificial Intelligence and Public Standards Report Follow Up

- We use automated testing tools provided by Eviden in every aspect of quality assurance, test management, automation, predictive analysis and test execution. We currently have over 50 tooling offerings available in the Eviden Application Testing Service (ATS) private cloud. Specific examples include Borders & Trade where we are using the toolset to support performance and validation testing across the Customs Declaration Service. This is to allow import & export declarations to be made when moving goods in and out of the UK and will be the UK's single platform from 2024.
- And like many other organisations we use AI for routing and optimising network space etc.

2. How do you demonstrate that you are using AI in ways that are legal and legitimate?

We've established an AI Ethics Working Group in HMRC to design, agree and ensure implementation of ethical principles and frameworks to be used in process and applied to all stages of development of predictive analytics solutions. The group also agrees the process for deployment of algorithms into operations and assurance process for those in use. We also align with UK GDPR and ensure data protection and security is adhered to via a number of mechanisms, such as Data Protection Impact Assessments, Business Impact Assessments and regular engagement with Security and Information Business Partners

3. How do you assess the potential impact of a proposed AI system on public standards (e.g., openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

When AI products are being considered as potential solutions to business challenges, the scope, content, delivery, and impact of those projects are fully considered ahead of any implementation. One key resource to support this is the Predictive Analytics Handbook, which is a cross government handbook and associated QA checklist, that ensures the developers of any AI solution consider all aspects of assurance and ethics as part of the design and build stage. These aspects include the transparency and explainability of any model, and thus any decisions or suggestions made by the model, in which case additional consideration is needed on that interaction and the ethical viewpoint.: HMRC have a 'human in the loop' for any use of AI that could materially affect taxpayers.

The Professional Standards Committee (PSC) also has a key role in providing oversight of how HMRC administers the tax system and applies policies in accordance with its values. The committee considers how HMRC's actions could affect trust in the tax system and public perception of fairness. It offers critical challenge to how HMRC exercises its powers, supporting fair practice in the use of its powers and safeguards.

Questions from Lord Evans of Weardale
Chair, Committee on Standards in Public Life
Artificial Intelligence and Public Standards Report Follow Up

- 4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.**

We have an established AI Assurance process and AI ethics framework and governance (including the AI Ethics Working Group) to ensure effective, responsible use of our AI models deployed within HMRC. The Group has representation from a wide range of teams (Technology, analysis, risking, strategy) and grades. Additionally, where we use AI in a way that could impact customer outcomes, we always ensure that the result is explainable, there's a human in the loop, and that it complies with our data protection, security, and ethical professional standards.

Also at a foundational layer, when analysts are scoping and developing AI models, a lot of work is done to assure the quality, completeness, and reliability of any base data, this includes identifying, understanding and if appropriate, managing and minimising any potential bias in that data where possible, giving increased confidence to our outputs and recommendations.

- 5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?**

When IT solutions are onboarded into HMRC they go through a rigorous process of IT testing and quality assurance, these processes also ensure the tools are fully supported and documented. This is no different for products incorporating AI. They all require clearly established governance, including the identification of a senior responsible officer and business area lead.

The Predictive Analytics Handbook provides a framework for how to industrialise your model ensuring it is fit for purpose for deployment (where is code hosted, is code easily deployable, how will the model be used, what is archival policy for older predictions). It also aides in the development of a robust monitoring programme to ensure that the model continues to perform as developed and mitigate risks to the business group. It assists with planning for how you will approach model redevelopment, and how to decide between new models versus the current model (the so-called Champion/Challenger approach)

- 6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?**

Questions from Lord Evans of Weardale

Chair, Committee on Standards in Public Life

Artificial Intelligence and Public Standards Report Follow Up

Once an AI system has been developed, tested, and deployed, then the role of the developer is to ensure that that AI system is performant and works as intended. This is achieved in a number of ways, firstly through the monitoring of the performance metrics, and with associated analysis to better understand any changes in those metrics. Secondly where predictive models are deployed, a Champion/Challenger approach is always utilised, this ensures that not only that the performance of the chosen model is monitored and assessed over time, but that alternative models are also assessed, and adopted when they become the best performing model. These approaches ensure that performance metrics are understood, and that model degradation is dealt with effectively as appropriate.

7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

On a day-to-day basis the Data Science & AI Board, chaired by our Chief Data Scientist, is the main governance forum for oversight of data science and AI initiatives, opportunities and risks. The membership of this board is wide and covers all main analytical areas, Transformation and Strategy, compliance and operational excellence. There are several groups that report into this board, including the AI Ethics Working Group, the LLM/GPT Special Interest Group and the Data Science Academy.

The Data Science & AI Board itself feeds into the Strategic Data Capability Board, chaired by our Chief Analyst, and that board reports to Data Committee lead by our 2nd perm sec. Any AI solution would also need to go through our existing IT change framework which includes various assurance and governance controls.

As mentioned above, HMRC also has the Professional Standards Committee (PSC), which is a non-Exec. The PSC provides oversight of how HMRC administers the tax system and applies policies in accordance with its values. The committee considers how HMRC's actions could affect trust in the tax system and public perception of fairness. It offers critical challenge to how HMRC exercises its powers, supporting fair practice in the use of its powers and safeguards. The PSC is chaired by HMRC's Director General for Customer Strategy and Tax Design, Jonathan Athow.

Outside of these AI-specific oversight roles, HMRC has a number of other roles and teams responsible for ensuring secure, fair and appropriate use of our data, right across the department. Roles such as the Data Protection Officer and Chief Security Officer as well as teams including Security and Legal Services, ensure standards and operating processes are established, with appropriate guardrails in place.

At a lower level, the developers, managers, and users of the AI models are also essential to that oversight. As mentioned above the models are continually evaluated from a performance perspective, and amended where necessary, they are also subject to rigorous assurance checks throughout the process, ending with human involvement

Questions from Lord Evans of Weardale
Chair, Committee on Standards in Public Life
Artificial Intelligence and Public Standards Report Follow Up

in assuring and approving any AI-led decision making. All AI solutions require human oversight and approval.

8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?

We currently don't use AI for any automated decision-making where there are interactions with our customers. We do use automated decision-making in areas such as customer services to support how we respond to engagement, and in compliance we are using AI to make decisions on how resources are better targeted and what actions should be taken in relation to specific cases, however for those (compliance) cases we always ensure there is a 'human in the loop' making the final decision. For example, using the outcome of previous interactions to predict which taxpayers are more likely to be non-compliant. This information is reviewed by tax specialists before selecting cases to contact, and existing appeals processes remain in place, if applicable, once decision letters are sent.

9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?

To get the best from AI, we are treating it as a profession, ensuring that we recruit and retain the right people and have the right skills to deploy AI ethically. We have held several conferences and learning events to help develop knowledge and capability in the Data Science and Artificial Intelligence space as illustrated in the list below.

- Digi-Con – Data Science Conference (9 & 10 September 2020)
- Digi-Learn – Learning Event (16 & 17 March 2021)
- 2021 Data Science Conference (7 & 8 December 2021)
- 2022 Data Science Conference (20 - 23 September 2022)
- Data Science Insights – Learning Event (17 May 2023)

External conference attendance, such as Big Data LDN, Operational Research conferences, ONS Data Science conferences and Public Service Data Live are also avenues of learning available to HMRC.

We also hold regular informal learning sessions, where members of our analytical community congregate to share best practice and knowledge, as well as help discuss issues to assist and support data scientists and analysts in the execution of their duties in their respective departments and help with their professional development.

Kent Police (September 2023)

Kent Police are not currently using AI. We are positively exploring options and research but have not yet implemented this within our workstreams. The predictive policing tool to help us decide what crimes are likely to be solvable is called EBIT. However, this is completed using regression analysis. We have successfully recruited a development opportunity within Kent to focus primarily on AI and Robotics within our Change Team so that we can dedicate a resource to explore and implement AI in the very near future, as we know there are so many opportunities out there. We will ensure that these technologies are used in accordance with the Nolan Principles at all stages.



Ministry
of Defence

JAMES CARTLIDGE MP
MINISTER OF STATE FOR DEFENCE PROCUREMENT

4/8/1/4/ap

19 October 2023

Dear Lord Evans,

ARTIFICIAL INTELLIGENCE AND PUBLIC STANDARDS REPORT FOLLOW UP

Thank you for your letter of 4 July 2023 requesting information about the adoption of AI in decision-making functions across Government, and the measures that departments are putting in place to manage associated challenges and risks.

Last year we published the Defence AI Strategy setting out our ambition to become the world's most effective, efficient, trusted, and influential Defence organisation for our size. AI offers massive potential to transform Defence and the UK Armed Forces, with potential applications – including for decision support – ranging from the back office to the front line. We have a duty to use these technologies safely and responsibly, in line with our legal commitments and the values and standards of the society that we serve.

Any Defence use of AI is governed by our AI Ethics Principles, which we have set out in our 'Ambitious, Safe, Responsible' policy. At the centre of our approach is the principle that a human must be accountable, and that that accountability cannot be delegated to a machine, regardless of the way in which AI is used to support decision-making or deliver an effect. Moreover, our default approach to adoption of AI will be through Human-Machine Teaming, combining human cognition, inventiveness and judgement with machine-speed analytical capabilities.

However, while we have high ambitions to adopt AI, it is important to understand that AI is not yet a widely deployed technology for any aspects of our activities. The overwhelming majority of AI-related projects and programmes in MOD are still in Research and Development. We are working at pace to operationalise our policy and implement appropriate technical and procedural control frameworks to continue to ensure that Defence equipment is safe, reliable, and effective for responsible use in pursuit of legitimate military objectives.

Given that the MOD does not yet utilise AI decision-support capabilities operationally, we have interpreted “how **do** you” statements as “how **might** you” for the purposes of providing responses to your questions.

I trust that this information is helpful to the Committee in preparing your update report.

Yours sincerely,

A handwritten signature in black ink, appearing to be 'J. Cartlidge', written in a cursive style.

JAMES CARTLIDGE MP

Annex: MOD’s Response to the Committee’s Questions

- 1. Are you currently using AI to aid decision-making in your organisation?
Please give details if possible. If no, please send us a nil return. If yes, please answer questions 2-9.**

A small number of teams have been authorised to use publicly-available large language models (LLM) such as ChatGPT to support specific tasks. The use-cases typically relate to increasing productivity (e.g. by speeding up the summarising and revising of large document sets) or research, but may nevertheless be considered as very low-level aids to decision-making. The purpose of the authorisation mechanism is to ensure that staff and managers at an appropriate level take responsibility and are accountable for using these tools in line with security rules, appropriate use policies, and guidance relating to the risk of bias, error and other issues associated with LLMs.

AI is not yet being used in the MOD in any formal or structured sense to aid decision-making processes. However, we do have a number of projects at various stages of the R&D maturity pipeline and have high ambition to adopt AI across the Defence enterprise. For example, the Enhanced Command and Control (EC2) Spearhead is demonstrating rapid development, integration and deployment of Artificial Intelligence (AI), Machine Learning (ML) and Automation technologies to support better and quicker decisions, including by automating routine but labour-intensive planning activities to reduce the burdens on military staff and help increase operational tempo.

The purpose of this and other prospective AI tools is to support human decision making. We have provided further information below about the MOD's wider strategic approach to the safe, responsible and effective adoption of AI in response to your below questions.

2. How do you demonstrate that you are using AI in ways that are legal and legitimate?

Defence will always abide by its national and international legal obligations. Any AI-enabled system or capability intended to be used by MOD or the UK military would be governed by the department's robust legal, safety and regulatory compliance regimes.

In many ways, the adoption of novel technologies and capabilities is not a new challenge for the Department; we are well practised at adapting our approaches, trialling and testing new systems in a range of challenging scenarios before they are qualified or trusted to be used operationally. However, we recognise that AI technologies can present different testing and assurance challenges when compared to traditional physical and software capabilities, particularly where systems have a capacity to learn and evolve over time. We set out MOD's approach to the use of AI-enabled systems in the Defence AI Strategy and accompanying Ambitious, Safe and Responsible (ASR) policy, which includes our Ethical Principles for AI in Defence. We are actively examining our processes and compliance regimes to ensure that we address any novel challenges posed by AI and 'learning systems' through appropriate controls throughout the system lifecycle.

Transparency and challenge are central to our approach. We publish information about our strategy and plans to the greatest extent possible (given security considerations) and actively engage with a broad range of external stakeholders to understand concerns, test thinking and scrutinise our approach to the responsible and ethical adoption and use of AI. This includes through an AI Ethics Advisory Panel that brings together voices from Defence, academia, industry and civil society to provide expert advice across the subjects of AI development, AI ethics, military ethics and international law.

3. How do you assess the potential impact of a proposed AI system on public standards (e.g. openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

AI-enabled systems offer great potential to enhance public service delivery and help our people to understand and continue to uphold the highest ethical standards. This could include, for example, through better informed and quicker decision-making, by gleaning novel and objective insights from data to mitigate bias and groupthink, by helping to strengthen the diversity and inclusiveness of our workforce, or by rapidly identifying instances of fraud and system misuse. However,

we recognise that – if poorly or inappropriately applied – AI-enabled systems could introduce or amplify errors (such as bias) that could lead to particularly harmful outcomes in a Defence context.

We are committed to adopting AI safely, responsibly and in ways that maximise the productivity and effectiveness of our people. Our AI Ethics Principles (Human Centricity; Responsibility; Understanding; Bias and Harm Mitigation; Reliability) are central to our approach. All AI systems and AI-enabled capabilities used in Defence – from the battlespace to the back office – must be assessed against these principles in order to identify any potential risks, including any potential impact on public standards, and put in place appropriate mitigating controls.

For example, we are clear that accountability always applies to a human and cannot be delegated to a machine. The ‘Human Centricity’ and ‘Responsibility’ principles require personnel who are involved in the decisions to field and use AI-enabled systems to understand what they are accountable for and have the appropriate context specific training needed to exercise that accountability. The effectiveness of steps taken to educate and train these personnel, and any associated procedural wrapper, would be tested through AI Ethics Risk Reviews. Moreover, Defence has significant expertise in understanding how accountability works on a systemic basis and how to assign accountability to the right level. We apply accountabilities through duty holding and through rules of engagement in order to empower military personnel and capabilities in a manner that achieves military strategic effect. New technological capabilities are adopted within that system of accountabilities.

We are currently working to understand how our existing compliance, risk management and regulatory regimes may need to be adapted to address AI-related challenges and ensure that the AI Ethics Principles are properly assessed and mitigated across the entire lifecycle of an AI system. New direction and guidance will be issued in due course to ensure that our AI-enabled systems must be reliable, fulfil their intended design and deployment criteria and perform as expected.

4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.

Our Ambitious, Safe and Responsible policy paper states that: *“Those responsible for AI-enabled systems must proactively mitigate the risk of unexpected or unintended biases or harms resulting from these systems, whether through their original rollout, or as they learn, change or are redeployed.”*

AI systems have potential to introduce unintended bias into Defence decision-making in various ways, including through poor system design, flawed data sets or improper use. We are taking a multi-faceted approach to mitigating

these risks as part of a 'system of systems' approach that addresses related people, process and technology issues.

We will mitigate the risk that partial or unreliable datasets might introduce unintended bias in three main ways: firstly, by assessing training data sets (particularly open source) to ensure that they are appropriate and any biases or data issues in the data are understood and documented; secondly, by augmenting or otherwise rebalancing datasets that are found to contain excessive bias (e.g. through addition of extra datasets where they exist, or by using 'synthetic' (i.e. generated) data to supplement real data where additional examples will help to mitigate statistical bias or where there just isn't enough real data in the set to enabling learning); and thirdly, through a rigorous verification and validation approach to testing trained models prior to their use operationally. Research is ongoing into the most effective use of verification and validation techniques.

Implementation of the Defence Data Strategy will increase the reliability and accuracy of AI training data (including through work on data governance and by providing the data fabric to manage MOD's data), while the Defence AI Centre is undertaking work on data assurance for reliability and accuracy. This second aspect will involve the curation and onboarding of new data sets (including assessing data to identify any quality or reliability issues, biases or limitations with the data which is then documented), and also defining how training data sets should be labelled in Defence in order to standardise the approach and reduce the potential for inaccurately labelled data.

We are also conscious that bias mitigation cannot always be prevented through assurances of data and that there is also a responsibility on users to apply their own critical judgement when they work alongside AI capabilities. Our AI ethics principle on *Understanding* states that Defence personnel must have an appropriate, context-specific understanding of the AI-enabled systems they operate and work alongside. For example, our policy on Large Language Models (LLMs) is clear that staff who have been temporarily authorised to use an LLM in support of their work must be given guidance on the issues that can arise and must understand that they retain "*overall responsibility for the accuracy and quality of the end product*". The level of understanding will naturally differ depending on the knowledge required to act ethically in a given role and with a given system. It may include an understanding of the general characteristics, benefits, and/or limitations of AI systems. It may require knowledge of a system's purposes and correct environment for use, including scenarios where a system should not be deployed or used. It may also demand an understanding of system performance and potential fail states.

- 5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?**

The MOD's AI Ethics Principle for 'Responsibility' states that "*Human responsibility for AI-enabled systems must be clearly established, ensuring accountability for their outcomes, with clearly defined means by which human control is exercised throughout their lifecycles*". Accountability always applies to a human and can never be delegated to a machine.

We are currently working to operationalise these Principles across Defence. This includes: formally assigning the responsibility for AI Ethics to Senior Accountable Officers in each Command and Enabling Organisation; instituting a robust assurance regime including AI Ethics Risk Reviews to test effectiveness of any control measures; using policies and processes to ensure clear lines of accountability for any development and use of AI systems, in line with our extant systems of accountabilities; and providing training and guidance materials to ensure that our people at all levels of the organisation understand their responsibilities and best practice in the development and use of these new technologies. Our approach continues to be shaped advice from a range of expert external stakeholders, including the Centre for Data Ethics and Innovation (CDEI) and the MOD AI Ethics Advisory Panel, chaired by the 2nd Permanent Secretary.

6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?

The performance and reliability of AI systems is of critical importance in a Defence context given the real-world impacts of our activities. We have repeatedly set out our commitment to the safe and responsible development and use of AI – including through effective systems monitoring and assurance controls – and seek to be as transparent as possible about our approach. For example, we developed our policy approach in collaboration an independent panel of experts (the AI Ethics External Advisory Panel) and regularly engage with external stakeholders to explain and test our thinking. That includes the recent oral evidence provided by the Minister for Defence Procurement to the House of Lords Committee on AI in Weapons Systems, where the minister was accompanied by MOD's Second Permanent Secretary and the Deputy Commander of Strategic Command to provide detailed information across a wide range of related matters.

As we explained to the committee session and have set out in our published policies, the MOD has robust and well-established processes for the testing, evaluation and certification of new systems and military capabilities. However, we recognise that AI has potential to pose particular challenges, not least owing to the unpredictability of some AI systems and the relative difficulty of interpreting how some forms of AI learn and make decisions. Our Defence AI Centre is working closely with Defence Regulators, other government departments, industry partners and external experts to ensure we maintain pace with technology's development, understand these novel challenges and can adapt our policies, processes and assurance regimes so that they remain fit for purpose to manage AI technologies.

This approach is expressed through our AI Ethics Principle on 'Reliability' which states that *"AI-enabled systems must be demonstrably reliable, robust and secure"*.

By way of example, it is clearly important for AI-enabled systems to fulfil their intended design and deployment criteria and perform as expected, within acceptable performance parameters. Those parameters must be reviewed and tested for reliability on an ongoing basis, particularly as AI-enabled systems learn and evolve over time, or are deployed in new contexts. Moreover, appropriate levels of explicability will be mandated throughout the lifecycle of a product's development, and we are investigating and developing new technical approaches that could allow us to assure this, potentially by segregating, containerising and managing AI within deterministic software architectures. Security conditions for AI-enabled systems include following our existing, stringent approaches to cybersecurity, data protection and privacy.

7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

Internal and external challenge are critical to our approach; ensuring that we understand inter-related issues and can assure that any AI-enabled system or military capability is safe, trustworthy and ethical. We work closely with a diverse range of experts to develop and test our approaches, including ethicists, technologists, academics, legal advisers, industry specialists, frontline military personnel, partners across government, international allies and civil society stakeholders. This engagement helps to ensure that our approach to the adoption of AI technologies is transparent, inclusive and founded on a rigorous evidence base.

For example, we established the MOD AI Ethics Advisory Panel (EAP) to bring together a group of experts from Defence, Academia, Industry and Civil Society to advise the MOD's Second Permanent Secretary on the safe and responsible use of AI in Defence. While it does not have formal decision-making powers, the EAP's advice was critical to the development of Defence's AI Ethics Principles and it will be central to ensuring effective implementation over the coming months and years. We continue to seek challenge from the panel to identify gaps or vulnerabilities within Defence where ethical standards might be particularly at risk.

Other examples of our ongoing engagement with external stakeholders and civil society groups include recent workshops with the International Committee for the Red Cross to test current policy positions; regular discussions with international partners and other stakeholder groups through UN-convened forums; and working with Team Defence and a group of industry stakeholders to understand how industry would design an AI Assurance framework for Defence. We also regularly take part in workshops on implementing responsible AI with leading industry experts.

8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?

The impact of AI technologies on people is likely to vary significantly depending on the function and context and will need to be assessed on a case-by-case basis. The AI Ethics Principle on 'Human-Centricity' clearly sets out that *"the impact of AI-enabled systems on humans must be assessed and considered, for a full range of effects both positive and negative across the entire system lifecycle. Whether they are MOD personnel, civilians or targets of military action, humans interacting with or affected by AI-enabled systems for Defence must be treated with respect. This means assessing and carefully considering the effects on humans of AI-enabled systems, taking full account of human diversity, and ensuring those effects are as positive as possible"*.

In addition to longstanding safety and regulatory compliance regimes, the MOD is developing a range of people, process, and technology controls to ensure that the impacts and implications for humans of adopting AI enabled systems in Defence are properly considered and mitigated throughout the system lifecycle. This includes, as examples: detailed direction, guidance and best practice guides; Senior Accountable Officers in each Command and Enabling Organisation that are responsible for assuring that Defence Policy is properly embedded throughout the organisation; AI Ethics Risk Reviews; the use of diverse and multi-disciplinary R&D teams; and a clear focus on human machine interface (HMI) design and effective training to ensure that systems are used ethically and effectively.

As the MOD system sets out clear and transparent accountability systems in respect to particular AI capabilities, the established systems of redress in respect to decisions are expected to be reinforced. As AI systems are operationalised the MOD will continue to apply the principle of 'Human Centricity' to ensure that the impact on those affected by AI systems will be addressed in context, this includes rights of redress.

9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?

The Defence AI Strategy explains how, over time, we will need to recruit talent as well as incubating and uplifting skills across the whole of the Department and UK Armed Forces to drive the effective exploitation and use of AI technologies.

We are currently developing a Defence AI Skills Framework which will identify key skills requirements across Defence. This will be overseen by our Head of AI Profession, who will sit within the Defence AI Centre and will also be responsible for developing our recruitment and retention offer; setting standards for delivery team skills and creating AI career development and progression pathways, with options for skilled generalists as well as deep specialists skills. These products – together with a broad range of supporting guidance and best practice training aids,

including on the operationalisation of our AI Ethics Principles and the responsibilities of key people throughout the development and use lifecycles – will be shaped by a range of expert stakeholders, including the AI Ethics Advisory Panel. As examples:

- Leaders must be equipped with a level of understanding that is sufficient to navigate the hype, seize opportunities and act as intelligent customers – an initial product introducing senior readers to core AI concepts and responsibilities will be released shortly.
- Decision-makers and operators of AI-enabled military capabilities must be trained to understand likely system behaviours in a given situation and context in order to develop appropriate trust and ensure that systems are used ethically and effectively.
- Whilst developers of AI systems need deep specialist knowledge, we are not expecting the whole force to need such a deep technical knowledge – but they do need to understand broadly how the AI systems they operate work, their limitations, and how they should responsibly and effectively use them.

Training and awareness activities must be delivered at the point of need, and repeated on a regular basis to ensure appropriate levels of understanding - including on the application of Defence's AI Ethics Principles - are maintained throughout the development and use of novel AI technologies across the Department.

Question	Response
Name of organisation:	Ministry of Justice September 2023
<p>1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible. If no, please send us a nil return. If yes, please answer questions 2-9.</p>	<p>1. The Ministry of Justice is still at the early stages of assessing where AI might help drive greater efficiency and deliver maximum value for the taxpayer, as part of the Government's digital transformation journey.</p> <p>2. For example, work done to date has been focused on modelling and advanced data analytics within our Performance, Strategy and Analysis group in order to inform policy making, as well as record linking across some line of business applications to de-duplicate our data. Another example, as recently mentioned elsewhere, is the work done around the use of Actuarial Risk Assessment Instruments (ARAI's) in HM Prisons & Probation Service to assess the risks posed by, and needs of, an offender by combining actuarial methods of prediction with structured professional judgement.</p>
<p>2. How do you demonstrate that you are using AI in ways that are legal and legitimate?</p>	<p>3. We are increasing our data ethics capability, including building an ethical framework for AI within the justice system. For instance, MoJ Data Science hub is collaborating with the Alan Turing Institute to build an ethical framework for data science and AI in the criminal justice system; MoJ officials are already representing the UK on the Council of Europe's Committee on Crime Problems; MoJ Security and Information Group has produced Security Guardrails on the use of generative AI and Large Language Models; and, a MoJ-specific risk assessment process has been developed from the Aqua book guidance.</p>

<p>3. How do you assess the potential impact of a proposed AI system on public standards (eg. openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?</p>	<p>4. The UK's legal services regulatory landscape is fragmented, with a significant part of the sector (approximately 37%) not falling under frontline regulation. The Legal Services Innovation team in MOJ is looking at the regulation of AI in legal services together with the Office of AI (OAI) to ensure these complexities are taken into account in OAI's proposed framework for AI regulation. This includes considering how best to approach risks in the areas which are not directly regulated.</p> <p>5. An MoJ-specific risk assessment process (Analytical Quality Assurance) has been developed from the Aqua book guidance, which is a good practice guide for creating analysis for government. Within the AQA checklist there is a risk matrix which factors in the risk level (low to high) and scale of the model to determine the appropriate level of risk assessment procedures that need to be completed. Training is provided to those who use the AQA checklist.</p>
<p>4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views?</p>	<p>6. The Ministry of Justice recognises that the use of AI in the justice system raises important ethical considerations, such as bias in the data used to train algorithms and the potential for automated decision-making to perpetuate existing inequalities. Therefore, any implementation of AI in the UK Ministry of Justice must be done carefully and transparently, with appropriate safeguards in place.</p> <p>7. The MoJ Data Science team holds Algorithm Consultation Panels to consider the technical, ethical, legal and operational considerations for algorithms in development with representatives from across the organisation such as colleagues from operations, data science, digital professions, data ethicists and end users.</p>

<p>5,How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?</p>	<p>8. So far, we have not deployed AI instruments that directly impact the end user of our services without ensuring human assessment of the results. Going forward, and as we continue experimentation, we are exploring other opportunities to strengthen and adapt our governance structure and accountability process to the risk associated with AI and ensure that any further and more complex implementation of AI in the MoJ must be done carefully and transparently, with appropriate safeguards in place.</p> <p>9. Therefore, we are setting up an ‘AI Steering Group’, to agree a set of principles for the use of AI; as well as our governance approach for AI including visibility and transparency, ethical use, issues of bias and discrimination and risk management. The forum will facilitate the agreement of roles and responsibilities across the department with regards to the use of AI.</p> <p>10. We are also in the process of appointing a Chief Technology Officer for Justice Digital, the digital function of the department, who among other things, will be responsible, for coordinating the AI agenda across the department.</p>
<p>6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?</p>	<p>11. There are a number of risk management processes in place which support in identifying issues associated with AI. As mentioned above the AQA checklist provides good practice guidance for producing quality government analysis. When the head of a data science team signs off a model, they check to ensure the AQA checklist has been completed correctly.</p> <p>12. Training is provided to those who use the AQA checklist, within which there are key checks including whether data ethics and protection considerations have been made, calculations have been checked both internally and externally, and that there is sufficient documentation for another analyst to understand what has been done and why.</p>
<p>7. What internal and external oversight mechanisms do you have</p>	<p>13. The Government Internal Audit Agency (GIAA) produced a maturity report covering MoJ’s governance maturity. It highlighted areas of good practice around data ethics and some areas of risk assessment (amongst others) but noted some areas for improvement, including</p>

in place to ensure that your AI systems are properly scrutinised?	further development of an AI strategy and how it would be delivered. There were further recommendations around strengthening our approach in risk management, too.
8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?	14. As mentioned elsewhere in this commission, the MoJ has released public facing publications of some of its analytical studies, which include the algorithms they have produced and how they operate, demonstrating the principles of transparency and accountability over use of analytical models.
9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it?	15. As mentioned elsewhere in this commission, training is provided to those who use the Analytical Quality Assurance (AQA) checklist, within which there are key checks including whether data ethics and protection considerations have been made, calculations have been checked both internally and externally, and that there is sufficient documentation for another analyst to understand what has been done and why.

West Midlands Police (December 2023)

1. Are you currently using AI to aid decision-making in your organisation? Please give details if possible.

Yes in proof of concept stage (2 month duration) – a national police science and technology office funded initiative around providing a better service for 101 non-emergency requests volumes. The pilot has progressed well and we achieved excellent results. We presented the findings to the HO and STAR fund and are looking for ways to fund it moving into production and developing further.

2. How do you demonstrate that you are using AI in ways that are legal and legitimate?

The “Amy 101 Voice Assistant” project was scrutinised at the highest ethical level by the independent OPCC Ethics Committee. For the submission a 30 page ethical paper was written to cover 18 topics of ethics under 3 sub-groups (Data, Behaviour and Confidence in Policing), these 18 topics were then risk rated with mitigations stated. This paper is marked under GSC, but could be suitably redacted and provided upon request.

3. How do you assess the potential impact of a proposed AI system on public standards (e.g. openness, accountability, and objectivity), and ensure that the design of the AI system mitigates any standards risks identified?

See response to Question 2.

4. How do you tackle issues of bias and discrimination in AI systems and decisions? For example, by taking into account a diverse range of behaviours, backgrounds and views.

Lots of work went into these considerations and was reported into the ethics paper accordingly as per question 2 response.

5. How do you ensure that responsibility for AI systems is clearly allocated and documented, and that operators of AI systems are able to exercise their responsibility in a meaningful way?

Currently WMP operates various governance boards to ensure suitable check/challenge accordingly on changes around AI. WMP however does not have an AI framework in-place to formalise such responsibilities. It will be considered as part of the 2 month PoC analysis, currently the team is small – hence changes/responsibilities are tightly controlled and involve business and technology subject matter experts.

6. How do you monitor and evaluate your AI systems to ensure they always operate as intended?

As part of our PoC we are leveraging AWS Lex V2 technology, which includes capabilities around smoke/regression testing on any changes delivered in code. As the PoC is small the

changes are not managed via a CI/CD pipeline, however should WMP look to implement this technology in future we will leverage CI/CD type technologies to ensure mandatory pass/fail type smoke and regression testing is implemented in an automated way. This is all manual checks currently, while robust in nature can take time to implement. The PoC also outputs to dashboards of which performance is monitored daily for trends/patterns accordingly.

7. What internal and external oversight mechanisms do you have in place to ensure that your AI systems are properly scrutinised?

See response to question 2 & 5.

8. How do you enable people to challenge automated and AI-assisted decisions and to seek redress using procedures that are independent and transparent?

This is not covered in the current proof of concept (Amy directs calls to be answered) so there is no decision making outside of a call taking process and we provide choice with the public calling. It is important for this to be considered as part of a wider AI framework that WMP would look to create and adopt where nationally available.

9. Do your employees working with AI undergo continuous training and education about AI and the ethical risks associated with it

Yes for the technical implementation team, albeit informally due to the small scale of the proof of concept we are undertaking.