MVP: AI Common AI Risks and Mitigations

The AI Digital Government Exchange Working Group

The Digital Government Exchange (DGX) Working Group on AI was established in December 2022 to share experiences, build collective approaches and exchange knowledge of Artificial Intelligence (AI) within the DGX network.

The working group is currently chaired by the UK's Government Digital Service (GDS). It comprises members from Australia, China, Germany, Israel, Japan, New York, Singapore, South Korea, and Sweden and includes World Bank and World Economic Forum representation.

The objectives of the DGX Working Group on Artificial Intelligence (AI) were to:

- develop draft AI Ethicist roles
- develop possible approaches to mitigate common risks in AI projects
- leverage the expertise of its members to create real-world use cases for the DGX community

Methodology

The group approached this project using agile practices, ran through several co-working sessions to understand the unique experience of DGX member country contexts and followed agile best practices through continuous improvement and iteration. This year's work includes two MVP AI Ethicist roles, mitigations to common risks paper, and a set of use cases based on feedback.

Minimal Viable Product (MVP)

In the first part of this paper, we have created a Minimum Viable Product (MVP) - a piece of good practice to help teams consider what to do when they build or buy AI projects. The areas in this MVP are based on the common thematic areas derived from DGX working group members and secondary research, looking outward at good practices from the international GovTech and AI communities.

This MVP is for:

- Senior Civil Servants who want to understand more about how their teams can adopt AI responsibly and appropriately and to help them make decisions for which they are responsible
- programme managers who are looking at building AI into their product or service
- procurement officials who may be involved in buying AI-based technology to help them understand what's expected when working with private sector organisations
- officials developing programmes that will include an AI workstream

Using the MVP

The MVP can be used in 4 ways:

1. to help inform governments on the common challenges and considerations they are working with as they implement their AI initiatives.
2. to understand where there are unique or country-specific challenges based on context, structure or other factors
3. to understand and share risk mitigation strategies when working with AI
4. to provide a framework for teams to assess AI projects

It is essential to reflect that this work is an MVP; we expect to iterate it based on feedback from across the DGX network.

Defining AI

Using the definition outlined previously in the MVP on good practice for new AI products and services for governments and central government departments, Artificial Intelligence (AI) refers to systems that imitate or mimic human intelligence to perform actions. Some AI systems can constantly improve themselves through interactions with users and the information they collect.

AI systems work by initially absorbing labelled training data, looking for data patterns, and using these patterns to "generate outputs, such as content, predictions, recommendations, or decisions influencing the environments they interact with."[1]

Some examples might include;

- using chat bots to provide real-time guidance for common government services, reducing wait times
- tracking the spread of a disease like COVID, and ensuring that appropriate resources are available to treat patients in areas that might need it most, or treating patients with heart conditions[2]
- reducing the work-load in agencies by using machine-based risk assessment in areas like business registration and income tax
- receiving automated benefits for specific life events like childbirth, retirement and bereavement
- AI supported language processing to support writing documents

Set out below are some common risks and mitigations in running AI projects.

Managing Bias

AI systems can inherit biases from their data, leading to unfair outcomes for specific groups. AI models are only as unbiased as the data they use. If the training data is biased, the AI

[1] REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

[2] 3D heart scans on the NHS to speed up disease diagnosis

model may learn and reinforce cultural, racial, gender, regional, and linguistic diversity, leading to discriminatory outcomes.

- **develop robust ethical frameworks and governance mechanisms to guide AI systems' design, development, and deployment**. Engage domain experts, stakeholders, and impacted communities to inform the creation of these frameworks, ensuring that they reflect diverse perspectives and uphold relevant principles like fairness or transparency.
- **carefully curating the training data used to train AI models.** Bias mitigation techniques like constant user testing and UR to mitigate bias and assumptions can reduce the impact of biased data on the AI model's learning process. Regularly auditing and testing AI models and input data for biases during and after development. An audit should be a part of the assurance activity. Bias-detection tools and techniques, such as adversarial testing and fairness-aware machine learning, can assess the fairness and robustness of the AI model. Regular audits and user testing should involve input from diverse stakeholders, including individuals from different demographic groups, to comprehensively evaluate biases from multiple perspectives.
- **consider including Civil Society Organisations (CSOs) in the audit process of AI projects**. CSOs can
    - provide diverse perspectives, ensuring marginalised communities' interests are considered.
    - act as independent watchdogs to ensure that AI projects are audited thoroughly and objectively.
    - create awareness to engage and educate the public on AI projects. A CSO may include educating the public about the potential impact of AI on society, raising awareness about the importance of auditing AI projects, facilitating public discussions and debates on the ethical implications of AI.
    - gather valuable feedback, perspectives, and insights to inform the audit process and ensure that AI projects align with societal values and interests.

- **embrace an exhaustive data collection process encompassing a wide spectrum of demographics and attributes**. Strive to encompass the intricate tapestry of humanity by including diverse samples representative of different races, genders, ages, socio-economic backgrounds, and other variables like language. Such inclusivity engenders a more accurate reflection of a multifaceted society, guarding against the risks of underrepresentation or overrepresentation and consider aspects related to digital divide issues (those that have access and those that do not, and those who do not wish to engage on services that include AI).
- **engage in diligent scrutiny to identify and address biases that may permeate the training data**. Employ rigorous methods to detect and quantify biases, with specific attention to protected attributes such as race, gender, and age. Once identified, institute effective mitigation strategies to rectify these biases, including data preprocessing techniques and careful curation to ensure fair representation.
- **establish a perpetual cycle of monitoring and evaluation to gauge the performance of AI models**. Regularly assess their outputs across various subgroups, attentively scrutinising for disparate or unjust outcomes. This ongoing

vigilance enables prompt intervention and necessary adjustments, safeguarding against perpetuating bias-induced disparities.

- **leverage people's expertise to supervise the functioning of AI models**. Deploy human reviewers and subject matter experts who understand the potential biases in the data.
- **encourage open communication, and establish mechanisms for responsibility** by cultivating an environment where transparency is valued, decisions are justified, and accountability is upheld at all levels is essential.

> ***The Australian Tax Office's (ATO)*** *has an AI-powered virtual assistant, Alex, to answer questions and provide tailored information to ATO clients. Within the first 18 months, Alex engaged in more than two million conversations with an 88% first contact resolution rate, exceeding the industry benchmark of 60-65%. Alex has further contributed to an 8-10 per cent reduction in contact centre call volumes and provided the ATO with real-time feedback on client engagement trends and website information gaps.* [cs-australian-tax-office-en-au.pdf]

Security

AI systems are vulnerable to hacking and other cyberattacks, and malicious actors can use AI for malicious purposes. AI systems are vulnerable to cyberattacks, and their complexity makes detecting and preventing attacks difficult.

- **ensure robust cybersecurity measures** for protecting the data that the AI trains on and the infrastructure used by AI systems. This includes implementing robust encryption protocols, regularly patching software vulnerabilities, and using multi-factor authentication to prevent unauthorised access. Regular security audits and penetration testing can help identify and address potential vulnerabilities in the AI system. It may be helpful to consider a lifecycle approach that deploys various security approaches throughout the development and deployment of the model and the underpinning data.
- **consider incorporating AI-based security mechanisms into the AI system to enhance its resilience against cyberattacks**. AI algorithms can detect and prevent anomalies or suspicious activities in the system, such as detecting patterns of unauthorised access or abnormal data inputs. Machine learning techniques can continuously update and adapt the AI system's defence mechanisms against emerging threats.
- **foster a culture of cybersecurity awareness and training among all stakeholders involved in developing and operating AI systems**. This includes educating developers, operators, and users about the risks of cyberattacks on AI systems and providing training on best practices for securing AI technologies to build a culture of security and support better detection and prevention of cyberattacks on AI systems.
- **develop strict access controls and authentication mechanisms to prevent unauthorised access to AI systems**. Limiting access to only authorised personnel

and monitoring and auditing access logs can help detect potential security breaches and plus limit AI exposure to additional systems, and limit the outputs of AI back to itself where needed (to prevent runaway processes)

- **AI requires large amounts of data, often including personal data.** This can raise concerns about data privacy and the potential misuse of personal information. Privacy risks in this space are multifaceted and demand careful consideration. One concerning aspect is the inadvertent exposure of private data, as AI systems, driven by powerful algorithms, may unintentionally unravel personal information. Additionally, the manipulation of AI through prompt injection poses a grave threat, potentially enabling the extraction of confidential data.
- **self-regulation, AI lacks the ability to fact-check its outputs,** compounding the reputational risk. Ensuring that data storage, management and processes meet any existing data privacy policies such as GDPR and existing information classification needs is essential to managing personal data.

Creating Regulation

AI raises new regulatory challenges; no universal, international legal framework or voluntary frameworks currently addresses these concerns.[3]

Mitigating the regulatory challenges posed by the rapid development and deployment of AI technologies requires proactive measures by governments.

- **a robust and adaptive regulatory framework can help ensure that AI is developed, deployed, and used responsibly,** balancing innovation with societal concerns and safeguarding the interests of individuals and communities.
- **establishing multi-disciplinary task forces or regulatory bodies that bring together experts from various fields, including technology, law, ethics, and policy, to develop comprehensive and adaptive regulatory frameworks for AI**. These task forces can conduct thorough assessments of ethical, social, economic, and legal implications and provide recommendations for regulations that promote the responsible and beneficial use of AI while addressing potential risks and concerns.
- **foster international collaborations and partnerships to develop harmonised regulatory frameworks for AI.** Given AI technologies' global nature and potential impact, coordination among governments and international organisations can help establish consistent and coherent regulations across jurisdictions. Collaborative efforts can include sharing best practices, exchanging information, and harmonising regulatory approaches through agreements or treaties, ensuring that the regulatory challenges of AI are addressed in a coordinated and cooperative manner.
- **actively engage with stakeholders from academia, industry, civil society, and the public to solicit input and feedback on AI regulations.** Public consultations, stakeholder forums, and participatory decision-making processes can ensure that diverse perspectives are considered in the regulatory process. Stakeholder and public engagement will create a more inclusive and effective set of regulations that reflect the needs and concerns of various stakeholders.

[3] A first step in AI regulation is the use of AI in the EU will be regulated by the AI Act, the world's first comprehensive AI law: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

Australia's SmartGates (abf.gov.au): *AI is used in many ways to improve customer's and citizen's interactions with government and support government decision-making. The Australian Border Force (ABF) uses SmartGates which allows for the biometric identification of travellers and rapid facilitation of low-risk travellers through the border. By using facial recognition software and ePassports, powered by AI, to process individuals, this enables the ABF to focus their efforts and resources on high-risk travellers.*

Singapore adopts a balanced approach to AI governance (https://www.imda.gov.sg/about-imda/research-and-statistics/sgdigital/tech-pillars/artificial-intelligence). *The government developed practical and implementable frameworks, guidelines, and tools for industry's voluntary implementation. These include the Model AI Governance Framework (published in 2019 and updated in 2020), Implementation and Self-Assessment Guide for Organisations, and 2 volumes of Compendium of Use Cases to showcase the successful implementation of responsible AI to inspire other companies to do the same.*

Accountability

The use of AI raises questions about accountability and responsibility. It can be difficult to determine who is responsible if something goes wrong with an AI system or if it produces negative consequences.

Mitigating the impact of questions about accountability and responsibility in AI requires clear guidelines and frameworks that define various stakeholders' roles, obligations, and liabilities in developing, deploying, and using AI systems.

- **establish legal and ethical standards that outline the responsibilities and accountabilities of the AI ecosystem's developers, its users, third party suppliers if necessary, and other stakeholders.** These standards can also include data collection and use guidelines, model development, transparency, fairness, explainability, and bias mitigation. Such standards can provide a basis for determining accountability and responsibility in case of negative consequences arising from AI systems.

- **implement robust governance mechanisms for teams and suppliers (if relevant feedback) to ensure AI systems' transparency, traceability, and explainability.** This includes documenting the AI system's design, development, and deployment processes. Implementing algorithmic impact assessments, audits, and third-party certifications can hold stakeholders accountable for outcomes. Establishing channels for reporting and addressing concerns, grievances, and complaints related to AI can provide a mechanism for addressing issues of accountability and responsibility.

The Australian Energy Market Operator (AEMO) *has partnered with Microsoft to leverage AI and machine learning to build new digital tools to enhance forecasting and grid reliability[1]. For example, the AEMO is currently developing an energy simulator for the entire Australian east coast network to*

ensure that the grid's operation is simultaneously optimised for security and efficiency. AEMO sparks data driven, AI infused energy transformation – Microsoft Australia News Centre

- **establishing transparent decision-making processes is critical to ensure public sector accountability in AI projects.** This includes clearly defining decision-making roles and responsibilities, documenting the rationale behind decisions, and making relevant information and documentation publicly accessible wherever possible. Additionally, implementing mechanisms for external audits, evaluations, or reviews of AI projects can provide an additional layer of accountability, ensuring that the decision-making processes are fair, unbiased, and in line with relevant regulations and policies.
- **invest in training:** Responsible AI development practices, such as data collection, model training, validation, and deployment and monitoring, can also be covered in staff training to ensure they are equipped with the knowledge and skills necessary to develop AI systems responsibly.
- **invest in a diverse and multidisciplinary team** with expertise in relevant fields, such as data science, ethics, law, and policy, to ensure accountability in AI projects. A team with diverse perspectives and expertise can better identify potential biases, ethical concerns, and unintended consequences associated with AI systems. It can also facilitate comprehensive risk assessment, mitigation, and informed decision-making throughout the project lifecycle.
- **engage with stakeholders and the public:** Involving stakeholders and the public in the decision-making processes of AI projects can foster transparency, accountability, and trust. This includes constant user testing and also against humans in the case of automations, soliciting feedback, conducting public consultations, and engaging with relevant experts, civil society organisations, and communities affected by the AI projects.

Singapore government's training efforts for public service officers. *Prompt Engineering is the process of crafting input instructions for generative AI, and it is a critical skill for officers who want to effectively harness LLMs to solve everyday problems.*

To facilitate and accelerate the learning curve, the government rolled out a Prompt Engineering Playbook for public service officers. This playbook will provide officers with a step-by-step guide to master prompt engineering, including best practices, practical examples and advanced tips and tricks.

In addition to the Prompt Engineering Playbook, the Singapore Civil Service College had also rolled out a Prompt Engineering Course to provide officers with hands-on training in crafting effective prompts for LLMs.

Assurance

Implement assurance processes (relative to the project's level of risk) to ensure that AI projects are being developed under relevant laws, regulations, and ethical standards. This can involve regular audits, reviews, and quality assurance checks to ensure the project progresses as planned.

As artificial intelligence (AI) becomes more ubiquitous, building assurance into AI projects to manage risk is increasingly important. AI systems are complex and challenging to understand, making them particularly susceptible to unintended consequences. By implementing a robust assurance process, organisations can help ensure that their AI systems are safe, reliable, and trustworthy.

In June 2023, Singapore's Minister for Communications and Information announced the launch of the AI Verify Foundation to harness the collective power and contributions of the global open-source community to develop AI testing tools for the responsible use of AI.

AI Verify, currently a Minimum Viable Product, helps organisations validate their AI systems' performance against internationally recognised governance principles through standardised tests and process checks. The AI Verify toolkit is a single integrated toolkit that conducts technical tests, records process checks, and generates test reports. Vertical plug-ins, such as sector-specific testing tools and/or new testing algorithms, can be built upon AI Verify. (https://aiverifyfoundation.sg/what-is-ai-verify/).

Several steps can build assurance into AI projects. The first step is establishing objectives and requirements for the AI system. This involves defining the problem the AI system intends to solve, the outcomes and performance metrics, and involving stakeholders.

Once the objectives and requirements have been established, developing the AI system is next. This involves selecting the appropriate algorithms and models, as well as collecting and preparing the data that will be used to train the system. It is essential to ensure that to the extent possible, the data is representative of the problem space and does not contain biases that could influence the performance of the AI system.

As the AI system is developed, it is essential to conduct regular reviews to ensure it meets the objectives and requirements. Qualified individuals with technical expertise and domain knowledge should conduct reviews. Reviews should cover various areas, including the AI system's performance and accuracy, the data quality, and the robustness of the algorithms and models.

Reviews should be conducted at several critical stages of the AI project. The first review should be completed before the AI system is deployed to ensure it functions as intended and is aligned with the objectives and requirements. Subsequent reviews should be conducted periodically to ensure the AI system meets the objectives and needs and identify potential issues or risks.

If a third party does the review, it is important to ensure that the reviewer has the necessary qualifications and expertise. The reviewer should have a deep understanding of the technical aspects of the AI system and domain knowledge in the area in which the AI system will be deployed. In addition, the reviewer should have experience in conducting audits and reviews of complex systems.

The UK's Department for Science, Innovation and Technology's approach to AI Assurance techniques : *Referencing a recent White Paper on AI regulation, the UK's Department for Science, Innovation and Technology (DSIT) sets out how it manages and applies a Portfolio of AI Assurance Techniques. The portfolio features a range of case studies illustrating various AI assurance techniques being used in the real-world to support the development of trustworthy AI.*

An audit should evaluate:

1. **performance of the AI system against the stated outcomes.** This includes assessing the accuracy of the system and its ability to handle a range of inputs and scenarios.
2. **quality of the data that is used to train the AI system**. This includes assessing whether the data is representative of the problem space, whether it contains biases, and whether it is properly labelled and annotated.
3. **robustness of the data processing algorithms, hyperparameters and models used in the AI system**. This includes assessing whether the algorithms are susceptible to adversarial attacks, handle missing or incomplete data, and are scalable and efficient.
4. **evaluate whether the AI system is aligned with ethical considerations**, such as fairness, accountability, and transparency (this is dual meaning, so could clarify - transparency of development, or transparency of decision making - the latter being harder to calculate for complex AI models.)

United Kingdom's (UK's) National Health Service (NHS): *All AI projects go through the Digital Technology Assessment Criteria (DTAC) for health and social care. "It is the national baseline criteria for digital health technologies entering and already used in the NHS and social care. The DTAC brings together legislation and good practice in these areas. The DTAC is designed to be used by healthcare organisations to assess suppliers at the point of procurement or as part of a due diligence process, to make sure digital technologies meet our minimum baseline standards. For developers, it sets out what is expected for entry into the NHS and social care."*

https://transform.england.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/

Managing and Establishing Risk Thresholds

Organisations can adopt proactive risk management practices, including comprehensive testing, validation, and monitoring of AI systems to identify and address potential biases,

risks, and unintended consequence**s.** Implementing robust mechanisms for ongoing monitoring and evaluation of AI systems' performance can help detect and mitigate any adverse impacts and hold relevant stakeholders responsible for their roles in developing and using AI technologies.

Risk assessment is critical in ensuring that AI projects are only undertaken if the potential benefits outweigh the risks. The criteria for assessing risk should be appropriately comprehensive and consider a range of factors, including the potential impact of the AI system, the likelihood of adverse events, and the ability to mitigate or manage risks.

- **establish a threshold for risk to ensure that AI projects are only undertaken if the potential benefits outweigh the risks**. This threshold should be based on an objective assessment of the project's potential risks and benefits, defining acceptable levels of risk and ensuring that any potential risks are identified and addressed early in the project lifecycle. Relevant laws and regulations and ethical considerations should inform the assessment.
- **consider the project's potential benefits, as well as the potential harms and risks**. An acceptable level of risk should be determined based on carefully balancing these factors to maximise the potential benefits while minimising the potential risks. The acceptable level of risk should be informed by relevant laws and regulations, as well as ethical and moral considerations.
- **conduct a risk assessment at the outset of the project, and it should be updated periodically throughout the project lifecycle**, including a thorough evaluation of the potential risks associated with the AI system and an analysis of the likelihood and potential impact of each risk. All assessments should be based on carefully evaluating relevant data and information and conducted by a team with the necessary expertise and qualifications.
- **implement a range of risk mitigation and management strategies throughout the project lifecycle.** These strategies include implementing robust testing and quality assurance measures, establishing clear policies and procedures for using the AI system, and providing training and support to users
- **establish clear communication channels** so that users and the project team can report any issues or concerns related to the AI system.
- **developing a comprehensive approach that applies best practices in project management,** risk management, and accountability.

General Criteria for measuring risk

- **assess the potential impact of the AI system**. This includes the potential for the AI system to cause harm to individuals or groups and the potential for the AI system to have unintended consequences. The impact of the AI system should be evaluated in terms of both the severity of the impact and the number of people who could be affected.
- **establish the likelihood of adverse events**. This includes the likelihood of the AI system failing or malfunctioning and the likelihood of the AI system being misused or exploited. The likelihood of adverse events should be evaluated based on the

complexity of the AI system, the quality of the data used to train the system, and the potential for human error or malicious intent

- **measure risk is the ability to mitigate or manage risks.** This includes detecting and responding to adverse events and implementing measures to prevent or reduce the likelihood of adverse events. The ability to mitigate or manage risks should be evaluated based on the availability and effectiveness of tools and technologies and the expertise and resources available to the team.

The United States of America's National Institute of Standards and Technology (NIST) *recently published the AI Risk Management Framework (RMF). The RMF offers practical guidance on the necessary capabilities that organisations should have to ensure confident innovation and effective management of AI risks. The NIST's RMF not only identifies the key principles but also provides practical guidance specifically focused on risk management activities associated with each principle. Link: https://www.nist.gov/itl/ai-risk-management-framework*

Implementing agile practices to mitigate risk

One effective way to ensure that AI projects are accountable is to use Agile methodologies that place emphasis on regular check-ins, feedback loops, and iterative development and conducting regular user research. This helps identify issues early in the project lifecycle, which can be addressed before they escalate into major problems. This includes defining clear project objectives, roles, and responsibilities, establishing a timeline, and setting up regular reporting and review mechanisms. Additionally, engaging with stakeholders, including users and relevant experts, throughout the project lifecycle can help ensure that the project remains aligned with its intended goals and is subject to appropriate scrutiny.

A checklist should be developed to ensure that all relevant stakeholders are involved in the project and that appropriate accountability measures are in place.[4] This checklist should include the following:

- clear project and programme objectives, expected outputs and outcomes
- identification of all relevant stakeholders and their roles and responsibilities
- a plan for regular reporting and monitoring of the project's progress against the expected outputs and outcomes
- mechanisms for ensuring transparency and accountability in decision-making processes, including a process for identifying and mitigating potential risks

[4] ***DGX: MVP - good practice for new AI products and services for governments and central government departments, 2022*** *https://www.developer.tech.gov.sg/our-digital-journey/digital-government-exchange/files/mvp-ai-good-practice-for-governments-and-central-govt-departments.pdf*

Managing Vendors and Suppliers

Establish ground rules for working with vendors or suppliers to ensure they are held accountable for their work. This can involve defining expectations for their involvement in the project. These ground rules should include:

- **the vendors' or suppliers' roles and responsibilities in the AI project**. This should consist of their specific tasks, deliverables, and timelines.
- **communicate the ethical and legal obligations that vendors or suppliers** must adhere to in developing and deploying AI systems. This can include requirements related to bias mitigation, data privacy, transparency, and accountability.
- r**egular reporting and monitoring mechanisms to track the progress of the work carried out by the vendors or suppliers**. This can include regular check-ins, progress updates, and milestones.
- **contractual obligations related to ethics, data privacy, bias mitigation, and other relevant aspects of AI development into vendor or supplier contracts**. Ensure these obligations are legally binding and specify the consequences for any breaches or violations.
- **require vendors or suppliers to provide documentation and explanations of their choice of algorithms**, data processing workflows and AI models used in the project. This can include details about the data used, the model's decision-making process, and any potential biases or limitations.
- **regular audits of the work done by vendors or suppliers to ensure compliance with established ground rules and contractual obligations.** This can involve reviewing their processes, methodologies, and outputs to identify potential issues or deviations.
- **a culture of open communication and encouraging vendors or suppliers to raise any concerns or challenges they may encounter during the project.** This can help address any issues early on and ensure that the work is carried out responsibly and accountable.
- **provisions in the contract for termination or dispute resolution in case of disagreements or breaches of contractual obligation**s. This can provide a mechanism for addressing accountability issues formally and legally.
- **regularly review and update the ground rules for working with vendors or suppliers to ensure they remain relevant and practical throughout the AI projec**t. This can include revisiting ethical and legal obligations, reporting mechanisms, and contractual obligations as needed.
- **seek legal expertise when drafting contracts and establishing ground rules** for working with vendors or suppliers to ensure they are legally sound and enforceable.

Misinformation

AI-powered tools, services and models can propagate false or misleading information like "deep fakes' of voice and image. The use of these models could be unintentional or deliberate.

Misinformation is a growing problem in our society, with false information being spread through various platforms and technologies. Education is one of the primary ways to develop critical approaches to identify and avoid misinformation.

Overall mitigations:

- **develop clear, concise, and accurate messaging that is easy for the public to understand.** This messaging should be developed in collaboration with experts in the field, including scientists, health professionals, and communication specialists.
- **prioritise educating the public about the dangers of misinformation and how to recognise it.** This can be done through public service announcements, social media campaigns, and educational programs.
- **develop fact-checking resources that are easily accessible to the public**. These resources should be regularly updated and widely promoted.
- **monitor social media platforms for misinformation and take action to remove false content**. This can be done by working with social media companies to develop policies and algorithms to identify and remove false content.
- **encourage media literacy by promoting critical thinking and providing resources to help individuals evaluate the accuracy of the information they encounter.** One such measure is identifying and showing the sources of the results. It is essential to inform individuals when they are using an AI system and to make them aware of the potential biases and limitations of the system.
- **legislative levers can also be used to enforce notifications on AI systems.** This can include labelling or promoting ideas, even with AI systems, identifying and connecting rules from existing systems to existing legislation.
- **data governance is another crucial aspect of mitigating the risks of misinformation; clean and accurate data will train AI systems** and reduce the risks of misinformation. It is critical to have robust data governance practices in place to ensure that the information is accurate.

On a more practical level, incremental steps can also be taken to mitigate the risks of misinformation.

- **provide proactive notification on AI systems** so that users are aware of when AI is being used.
- **include creating a public inventory** that the government can use to track AI systems and developing general education on AI and potential use cases.
- **create and publish best practice guides, build local peer communities** on AI systems, and create test questions for data quality to ensure that information is clear and accurate.
- **build trust with the public by being transparent about the decision-making processes and communicating openly and honestly with the public**. This can help to mitigate the spread of misinformation and improve public trust in government institutions. Involving CSOs to help check and provide feedback on information to ensure a robust data program in place could also be helpful.

Sustainability

- **focus on optimising the energy efficiency of AI infrastructure**. This can be achieved through various means, such as using energy-efficient hardware, optimising algorithms and models for reduced computational requirements and employing intelligent power management techniques.
- **implement sustainable practices within data centres that house AI system**s. Use energy-efficient cooling and power management technologies, consider renewable energy sources for powering data centres and explore advanced cooling techniques like liquid cooling to reduce energy requirements. Additionally, promote recycling and responsible disposal of electronic waste generated by AI infrastructure.
- **conduct life cycle analysis to assess the carbon footprint of AI systems.** This involves evaluating the environmental impact across the entire life cycle, from manufacturing and deployment to decommissioning. Additionally, organisations can invest in carbon offset programs to compensate for the CO2 emissions produced by AI operations, supporting initiatives such as renewable energy projects or reforestation efforts.

***UK's approach to Cloud Computing and its** green IT strategy.*

In 2013, the UK government introduced the Cloud First policy, encouraging public sector organisations to consider cloud computing solutions before making any new IT investments. The policy aims to exploit the benefits of cloud computing, including cost savings, scalability, and improved efficiency.

The Government Digital Service (GDS) and the Central Digital and Data Office (CDDO) plays a crucial role in driving the adoption of cloud computing within the UK government. It provides guidance, standards, and best practices to government departments and agencies for adopting cloud-based solutions securely and effectively.

The UK's green IT strategy aims to minimise the environmental impact of IT operations across the public sector. It focuses on reducing energy consumption, carbon emissions, and IT infrastructure and services waste. Cloud computing plays a significant role in this strategy by enabling the consolidation of data centres, optimising resource utilisation, and promoting shared services.

The G-Cloud framework was established to simplify public sector organisations' procurement of cloud services. G-Cloud acts as an online marketplace, where government entities can browse and purchase pre-approved cloud services from a pool of accredited suppliers. This framework helps ensure that cloud solutions meet security and regulatory requirements.
The UK has also been actively consolidating its data centre infrastructure to reduce energy consumption and improve overall efficiency. By migrating applications and services to cloud providers, government organisations can reduce the number of physical data centres, leading to significant energy savings and reduced carbon emissions.

Cloud service providers typically operate large-scale data centres that leverage advanced infrastructure and energy-efficient technologies. By migrating workloads to cloud environments, the government can take advantage of these energy-efficient data centres, reducing its overall carbon footprint and energy consumption compared to traditional on-premises infrastructure. The government actively encourages cloud service providers to demonstrate their commitment to sustainability and environmental responsibility. When procuring cloud services, departments

consider the environmental credentials of the providers, including their use of renewable energy, energy-efficient infrastructure, and sustainable practices.