THE NATIONAL ARCHIVES

Artificial Intelligence Policy

IT Security & Assurance / IT Operations

Contents

Policy owner: Security & Assurance (IT Operations)
Version: 0.2 draft
Published: 28/06/2024
Reviewed: 28/06/2024
Next review by: December 2024

1 Introduction

1.1 The broad term 'Artificial Intelligence' (AI) is used in this policy to describe various computer systems that can perform tasks previously requiring human-like intelligence. This includes Natural Language Programming, Machine Learning and Generative AI.

1.2 'Generative AI' is used to describe any type of artificial intelligence that can create new content, including text, images, video, audio, or code. Large Language Models (LLMs) are part of this category of AI and produce text outputs. The abilities of Generative AI have captured the public imagination and create potential applications within The National Archives (TNA).

1.3 AI has the potential to unlock significant productivity benefits. This document aims to help people at The National Archives understand AI, to guide anyone building AI solutions, and, most importantly, to lay out what must be taken into account to use AI safely and responsibly. It is based on a set of ten principles which should be considered in all AI projects.

1.4 Significant risks exist with the use of AI, such as bias, incorrect information, discrimination, data privacy violations. The use of AI tools in an organisation such as ours presents new challenges for our information security and data management and protection responsibilities.

2 Purpose

2.1 The purpose of this policy is to establish guidelines and best practices for the responsible use of artificial intelligence within our organisation. AI technologies have the potential to transform our business processes, enhance decision-making, and improve efficiency. However, it is essential to ensure that AI is used ethically, responsibly, transparently, and in compliance with legal and regulatory requirements.

2.2 This policy builds upon government guidance, Generative AI Framework for HMG, published 18 January 2024, for the use of AI in the public sector and the National AI Strategy by the Office for AI, but adds additional controls specific to the nature of the work carried out by The National Archives. The framework has Generative AI as its focus, but TNA's policy expects staff to apply the same

principles to any use that falls within the broader definition of Artificial Intelligence.

2.3 We recognise that development and deployment of AI at The National Archives is at an early stage and there is activity underway to develop strategies, plans and governance.

3 Scope

3.1 This policy applies to all people using Artificial Intelligence at The National Archives. Its scope covers the use of AI tools for all work tasks (in any location), but also private tasks if undertaken using laptops or other equipment or logins provided by TNA. The scope also includes the potential use of any data in TNA that has been created with the help of AI.

3.2 AI tools may be used at The National Archives in accordance with this policy. Three main types of tool are anticipated:

- Ad-hoc use of web AI services by individuals
- AI tools embedded into IT applications, e.g. Microsoft Co-Pilot
- Tools/services/products developed by TNA with AI components.

3.3 TNA's default assumption is that AI tools are insecure. Unless specifically advised otherwise, **personal data, sensitive data and bulk data must not be used with any AI tool.** Any data that is input into an AI tool must be anonymised and sanitised, for example to avoid references to individual or to TNA/The National Archives.

4 Policy

4.1 General

4.1.1 TNA will follow government policy and guidance on the use of AI.

4.1.2 All users of AI tools should be aware of and follow, where appropriate, the ten common principles in the government Generative AI Framework for HMG for safe, responsible and effective use of generative AI in government organisations. The white paper A pro-innovation approach to AI regulation, sets out five principles to guide and inform AI development in all sectors. This

framework builds on those principles to create ten core principles for generative AI use in government and public sector organisations:

- **Principle 1**: You know what generative AI is and what its limitations are
- **Principle 2**: You use generative AI lawfully, ethically and responsibly
- **Principle 3**: You know how to keep generative AI tools secure
- **Principle 4**: You have meaningful human control at the right stage
- **Principle 5**: You understand how to manage the full generative AI lifecycle
- **Principle 6**: You use the right tool for the job
- **Principle 7**: You are open and collaborative
- **Principle 8**: You work with commercial colleagues from the start
- **Principle 9**: You have the skills and expertise needed to build and use generative AI
- **Principle 10**: You use these principles alongside your organisation's policies and have the right assurance in place

Note: **Principle 4** includes ensuring that for all AI-assisted processes, TNA commit to upholding the primacy of human judgement and ensuring full transparency and openness with our colleagues and trade unions.

4.1.3 Where available, corporately provided tools must be used. If this is not possible and a login is required, staff must authenticate using their TNA email address together with a unique password that follows our password policy.

4.1.4 Outputs from an AI tool must not be used without reviewing it first.

4.1.5 Were any AI tool to have an impact on job roles at The National Archives, there would be engagement with trade unions in line with the existing processes for this.

4.2 Responsibility

4.2.1 All TNA staff are expected to adhere to security best practices when using AI systems. This includes the following:

- **Evaluation of AI tools:** Users will evaluate the security of any AI tool before using it and seek permission from their department Information Asset Owner.
- **Protection of confidential data:** TNA users must not upload or share any data that is confidential, proprietary, or protected by regulation.

- **Access control:** TNA users must use their TNA credentials for any work-related AI use and must not share login credentials or other sensitive information with third parties.
- **Use of reputable AI tools:** Only reputable AI tools will be considered for use in the TNA and staff must be cautious and avoid AI tools developed by individuals or companies without established reputations.
- **Compliance with security policies:** Staff must apply the same security best practices we use for all company and customer data.
- **Data privacy:** Even when approval for a tool has been given staff must exercise discretion when sharing information publicly.
- **Risks:** All users should be aware of the risks, benefits and general characteristics of AI tools before using them. All users of these tools must use them safely and ethically.

4.2.2 Governance for AI follows the principles of the TNA Cloud Services Policy:

4.2.3 Information Asset Owners (IAOs) are responsible for ensuring AI tools are used safely within their department and may approve use of them where appropriate for OFFICIAL data.

4.2.4 IAOs should record their department's use of AI tools, for example in IAO reporting.

4.3 Transparency

4.3.1 All users must be open and honest in their use of AI systems, the purposes they are used for and how they are applied, particularly where AI has contributed to their work. Any outputs produced using AI must be clearly marked.

4.3.2 IAOs are responsible for publishing under the Algorithmic Recording Transparency Standard

5 Risks

5.1 AI systems we develop must be safe and secure. They must be designed to prevent harm to individuals and organisations.

5.2 All users should be aware of the risks and problems in using AI tools and in the responses they give.

5.3 Risks from inputting data:

5.3.1 An AI tool may use any prompt or input in its own learning and repeat it to any other user of that tool. As such, any data given to an AI tool must not be sensitive and be suitable for public access.

5.4 Risks of AI outputs:

5.4.1 AI outputs may be wrong or misleading, fictitious, biased or in an improper style (e.g. US English).

5.4.2 AI outputs may be out of date or imply that past activities are current.

5.4.3 AI outputs may present facts out of context in a misleading way.

5.4.4 Sensitive data could emerge unexpectedly in AI output.

6 Contact

6.1 Users or IAOs with any doubts about an AI tool should contact the IT Security team via the IT Service Desk.

6.2 All incidents or suspected incidents related to AI must be reported as soon as possible following IT security reporting procedures:

7 Review

This policy will be reviewed and updated on a regular basis to ensure that it remains current and effective.