DWP Artificial Intelligence Security Policy

Chief Security Officer

Department for Work & Pensions

ThinkSecure

The DWP Artificial Intelligence Security Policy is part of a suite of policies, designed to promote consistency across the Department for Work and Pensions (DWP), and supplier base with regards to the implementation and management of security controls. For the purposes of this policy, the term DWP and Department are used interchangeably.

Security policies considered appropriate for public viewing are published here:

https://www.gov.uk/government/publications/dwp-procurement-security-policies-and-standards

Security policies cross-refer to each other where needed, so can be confidently used together. They contain both mandatory and advisory elements, described in consistent language (see table below).

Table 1 – Terms

Term	Intention
must	denotes a requirement: a mandatory element.
should	should denotes a recommendation: an advisory element.
may	denotes approval.
might	denotes a possibility.
can	denotes both capability and possibility.
is/are	is/are denotes a description.

Table of Contents

DWP Artificial Intelligence Security Policy

Overview

DWP promotes and advocates the use of Artificial Intelligence (AI) in a measured and controlled manner. AI tools can enhance services by automating routine tasks, enhancing data analysis, speeding up information retrieval and aid creativity.

This policy sets out user responsibilities when using AI Tools and aims to ensure:

- Data protection and information management processes are followed,
- User accountability for AI usage,
- Accuracy of information generated by AI Tools,
- Transparency when using AI Tools.

This policy has been created using a risk-based approach to ensure that the policy statements align with the organisational risk appetite.

Purpose

This security policy establishes end user responsibilities when using AI Tools for official DWP business. The policy aims to ensure that AI Tools are used securely by outlining acceptable use of AI, the need to check accuracy of results, the need for transparency and user accountability for AI generated content. The policy also addresses the data protection responsibilities of Users when utilising AI Tools.

Scope

This policy applies to:

a) DWP employees and contractors using DWP equipment or software licenced by DWP. Referred to as “Users” throughout the document.
b) All AI Tools, including but not limited to generative AI, machine learning and large language models.

The Third Party Artificial Intelligence Security Policy (Annex A) details requirements for suppliers and their third parties when utilising AI Tools for DWP business and/or in the provision of services to DWP.

This policy does not replace legal or regulatory requirements.

Definitions

Artificial Intelligence (AI) – A computer programme which “learns” from data and may perform tasks usually carried out by humans.

AI Tool – Any computer software that uses artificial intelligence in its processing.

Approved AI Tools – AI Tools that reside on the DWP infrastructure and that have approved by The Digital Design Authority (DDA), which is the lead DWP governance board for the approval of new digital tools (further information can be found in Annex B).

Online AI Tools – AI Tools which are accessed on the internet through a web browser.

AI Output – Any content that is generated by an AI tool.

Generative AI – A form of AI which generates content based on the inputs of users.

Large Language Model – A form of generative AI that examines very large datasets to produce context related output.

Prompt – An instructive command or question which directs AI to perform tasks.

Machine Learning - A form of AI which self learns without following explicit instructions.

Policy Statements

1. Approved AI Tools and Online AI tools that can be accessed on DWP devices may be used where there is a business requirement to do so.
2. Where there is a business need to use any of the following information with an AI tool, users must ensure that approval has been granted by the relevant governance board (see annex B):
    - Data that is classified OFFICIAL with the –SENSITIVE handling caveat or above (Please refer to the DWP Security Classification Policy),
    - Personal data,
    - Non-public DWP code (for example, code relating to DWP systems which has not been cleared for the public domain).

    Users must not upload the information above unless explicit approval has been granted.
3. The Data Protection Impact Assessment process must be followed where users are:
    - Using an AI tool to process personal data,
    - Introducing an AI tool into an existing business process which involves processing personal data,
    - Using the output of an AI tool in a way that will affect people.
4. An Equality Analysis is required where personal data relating to protected characteristics is being processed by an AI tool, in line with the Equality Act (2010).
5. Online AI Tools may initially be blocked. Where users have a business need to access a blocked Online AI tool a request can be submitted to unblock the website (Unblocking Webpages).
6. Users must not attempt to access DeepSeek AI on DWP devices.
7. DWP Standards of Behaviour apply when interacting with AI Tools.
8. Users must ensure to the best of their ability that the data they provide to AI Tools is accurate.

9. Users must check the accuracy, reliability, and credibility of AI output, to verify to the best of their ability that it does not contain any misinformation, or bias.

10. AI Tools must be logged on the Algorithmic Transparency Recording Standard (ATRS) where they meet the criteria outlined below:
    a) Have a significant influence on a decision-making process which effects the public, or,
    b) Directly interact with the general public.

    Where these criteria are met, but the user has a strong justification not to disclose their use of AI on the ATRS, an ATRS exemption can be sought from the Department of Science Innovation and Technology (DSIT). An exemption must be approved prior to use of the AI Tool.
    Further information can be found here: Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies - GOV.UK (www.gov.uk).

11. Users must declare the use of AI Tools for official business if requested (for example, via a Freedom of Information (FOI) request, Subject Access Request (SAR) or for assurance purposes).

12. When using AI Tools, the Information Management Policy applies.

13. AI Tools must be used in a way that ensures data protection principles and individuals rights are upheld. Further information (including guidance on automated decision making) is provided by the Data Protection Officer's team and can be found here: Artificial Intelligence and Data Protection.

Accountabilities and Responsibilities

a) The DWP Chief Security Officer is the accountable owner of the DWP AI Security Policy and is responsible for its maintenance and review, through the DWP Deputy Director for Security Policy and Central Services.

b) Line managers must ensure that employees are aware of their responsibilities when using AI Tools.

c) It is the responsibility of all users to ensure that misuse of AI tools is reported to their line manager and, if required, to the Security Incident Response Team (see compliance section b).

d) It is the line manager's responsibility to take appropriate action where non-compliance to policy is identified as detailed in the DWP Discipline Policy.

Compliance

a) All DWP employees, whether permanent or temporary and including suppliers and third parties who use DWP devices and/or licenced software must be aware of and comply with DWP's security policies and standards.

b) Failure to report a security incident, potential or otherwise, could result in disciplinary action and, in the most severe circumstances, result in dismissal. A security incident is the attempted or actual unauthorised access, use, disclosure, modification, loss or destruction of a DWP asset (or a supplier asset that provides a service to the Authority) in violation of security policy. The circumstances may include actions that were actual, suspected, accidental, deliberate, or attempted. Security incidents must be reported as soon as possible. DWP users must report security incidents via the DWP Security Incident Referral Webform; third parties and suppliers must follow the DWP Security Incident Management Standard (SS-014).

c) The DWP Security and Data Protection directorate will regularly assess for compliance with this policy and may need to inspect systems, information and documentation to facilitate this. All DWP employees, whether permanent or temporary and including suppliers and third parties who use DWP devices and/or licenced software will be required to support this activity.

d) DWP may monitor business and personal use of DWP information and communication systems to ensure compliance with DWP policies and standards, further information can be found in the DWP Employee Privacy Notice.

e) An exception to policy may be requested in instances where a business case can be made to undertake an activity that is non-compliant with DWP security policies and standards. If an individual is aware of an activity that breaches DWP security policy or standards, they should notify the Security Policy and Standards team immediately and, if required, the Security Incident Response Team (see compliance section b).

Annex A – Third Party Artificial Intelligence Security Policy Scope

This section details the requirements for suppliers and their third parties when using AI Tools for DWP business. This applies to suppliers and their third parties who meet any of the of the following criteria:

1) Whose systems or services store, handle, or process DWP corporate and/or customer information,
2) Who are involved in the provision and lifecycle management of hardware for DWP,
3) Whose systems or services reside within DWP networks,
4) Who use AI Tools when carrying out work on behalf of DWP.

Policy Statements

1. When utilising AI Tools for DWP business, the following information must not be used without explicit approval from DWP:
    - Data that is classified OFFICIAL with the –SENSITIVE handling caveat or above (Please refer to the DWP Security Classification Policy),
    - Personal data,
    - Non-public DWP code (for example, code relating to DWP systems which has not been cleared for the public domain).
2. Suppliers and their third parties must follow the Data Protection Impact Assessment process via their contract manager where they intend to:
    - Use an AI tool to process personal data
    - Introduce an AI tool into an existing business process which involves processing personal data
    - Use the output of an AI tool in a way that will affect people.
3. Suppliers and their third parties must ensure to the best of their ability that the data they provide to AI Tools is accurate.
4. When using AI Tools for DWP business, on DWP networks or when providing services for DWP, suppliers and their third parties must check the accuracy,

reliability and credibility of information generated by AI, to verify to the best of their ability that they do not contain misinformation or bias.

5. Suppliers and their third parties must be transparent when using AI Tools for DWP business or on DWP networks. This includes:
   - Declaring where AI Tools have been used,
   - Declaring which AI Tools will be/have been used,
   - Providing evidence of how AI output was achieved,
   - Informing DWP where AI will be incorporated into existing services/software.
6. When using AI Tools, the DWP Information Management Policy applies and must be utilised in accordance with data protection legal requirements.
7. AI Tools must be used in a way that ensures data protection principles and individuals rights are upheld. Further information (including guidance on automated decision making) is provided by the Information Commissioners Office and can be found here: Explaining decisions made with AI | ICO Artificial Intelligence and Data Protection.

Accountabilities and Responsibilities

1. The DWP Chief Security Officer is the accountable owner of the DWP AI Security Policy and is responsible for its maintenance and review, through the Security and Data Protection directorate.
2. Contract managers must ensure that suppliers and their third parties are aware of their responsibilities when using AI Tools.
3. Suppliers and their third parties must report any suspected or actual abuse of AI Tools that affect DWP information or assets to contract managers.
4. Contract managers must report any suspected or actual abuse of AI Tools by suppliers and their third parties that affect DWP information or assets via DWP Place portal.

Third Party Compliance

a) Suppliers and their third parties, whether permanent or temporary, have security responsibilities and must be aware of and comply with DWP's security policies and standards.

b) A security incident is the attempted or actual unauthorised access, use, disclosure, modification, loss or destruction of a DWP asset (or a supplier asset that provides a service to the Authority) in violation of security policy. The circumstances may include actions that were actual, suspected, accidental, deliberate, or attempted. Security incidents must be reported as soon as possible. Third parties and suppliers must follow the DWP Security Incident Management Standard (SS-014).

c) All actual or suspected security incidents and data breaches must be raised with the DWP Contract Manager.

d) The DWP Security and Data Protection directorate will regularly assess for compliance with this policy and may need to inspect systems, information and documentation to facilitate this.

e) If for any reason suppliers and their third parties are unable to comply with this policy or require use of technology which is outside its scope, they must discuss this with the contract manager who must then contact the DWP Security Policy and Standards Team to discuss a possible exception to policy.

Annex B – Governance Boards and Associated teams

This section outlines the governance boards and business areas that can be engaged in relation to the implementation and use of AI Tools. Where approval is granted for use in a pilot or proof-of-concept, it may be necessary to engage further before implementation on a larger scale. Please note that this list is not exhaustive and there may be other governance boards specific to your area.

Digital Design Authority (DDA)

This is the lead DWP governance board for the approval of new digital tools.

Change Portfolio Management Office (CPMO)

The CPMO acts as the information, advice, management, and governance hub for all of Service Delivery and non-programme change.

AI Delivery Board

The AI Delivery board acts as the formal sign-off board for AI implementation across DWP.

Digital Security Risk Management (DSRM)

A team within Digital Group that assess and manage the risks of all technological services across DWP and are a primary contact point if you require a risk assessment.

Enterprise Security Risk Management (ESRM)

A team within Security and Data Protection (S&DP) that assess and manage the risk of enterprise level systems and are a primary contact point if you require a risk assessment.

DWP Data Protection Officer's (DPO) Team

A team within Security and Data Protection (S&DP) that monitors information management within DWP and are the primary contact point if you require a DPIA.

Data Transfer Team

A team that manages the transfer of data outside DWP systems. They ensure that data transfers comply with DWP's security policies and standards.