

Artificial Intelligence Secure Implementation standard

ABSTRACT

This document provides a structured approach to assessing and managing security risks associated with AI adoption in law enforcement. It ensures that AI technologies are evaluated at every stage of their lifecycle, meeting security and operational requirements.

This standard adheres to the National Policing Community Security Policy Framework and is a suitable reference for community members, notably those who build and implement IT systems on behalf of national policing.

ISSUED:	May 2026
PLANNED REVIEW DATE:	May 2027
DOCUMENT OWNER:	National Chief Information Security Officer
DISTRIBUTION:	Members of the Policing Community of Trust
COPYRIGHT:	All content – copyright Police Digital Service
DOCUMENT HANDLING:	OFFICIAL – FOR PUBLIC RELEASE
POLICY VALIDITY STATEMENT	
This standard is due for review on the date shown above. After this date, policy and process documents may become invalid.	
Readers should ensure that they are consulting the currently valid version of the documentation.	

DATE:	27/04/26
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Contents

Community Security Policy Commitment 3

Introduction 3

Definitions 3

Purpose..... 5

Audience..... 5

Scope..... 6

Requirements 6

Communication approach 24

Document Compliance Requirements 24

Equality Impact Assessment 24

Appendices 24

Review 25

Related documents..... 25

 Vendor compliance and transparency – Microsoft Copilot 25

Document Information..... 26

 Document Location:..... 26

 Version History 26

 Approvals..... 26

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Community Security Policy Commitment

National Policing and its community members recognise that threats to policing information assets present significant risk to policing operations. National Policing and its community members are committed to managing information security and risk and maintaining an appropriate response to current and emerging threats, as an enabling mechanism for policing to achieve its operational objectives whilst preserving life, property, and civil liberties.

This risk assessment in conjunction with the National Policing Community Security Policy Framework and associated documents sets out National Policing requirements for the use of Artificial Intelligence in Policing.

Introduction

As Artificial Intelligence (AI) becomes increasingly integrated into law enforcement, it is essential that policing organisations have a structured process to assess and manage the risks associated with AI technologies. While AI has the potential to significantly enhance the work of the policing community, it must be developed, deployed, and operated securely and responsibly.

Without proper oversight, AI introduces unique security risks and operational challenges that require careful consideration. To ensure AI technologies function ethically and securely, the policing community must proactively identify and mitigate these risks at every stage of AI adoption. A secure-by-design approach is essential, embedding security from the outset and enforcing it throughout the AI system's lifecycle. This approach ensures cyber resilience, making security a core requirement throughout the entire lifecycle of AI technologies.

As policing organisations begin integrating AI technologies across their operations, it is important that they apply the recommendations from this risk assessment to ensure AI is implemented securely and responsibly.

Definitions

There are many definitions available for AI. For the purposes of this document, we have adopted and expanded on the definition provided by the NPCC endorsed – 'Principles for Using Artificial Intelligence in Policing', written by Science & Technology in Policing. While we reference AI throughout this document, it is intended that reference to AI, covers all of the below.

What is Artificial Intelligence?¹

There is no definitive definition of Artificial Intelligence (Alan Turing Institute, 2021), and AI is often used to refer to related applications such as automation, neural networks, and machine learning. To bring clarity for policing, we adopt the following definitions:

¹ NPCC - Principles for Using Artificial Intelligence (AI) in Policing

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

- **Artificial intelligence (AI)** refers to a machine that learns, generalises, or infers meaning from input, thereby reproducing or surpassing human performance. An example is using image analysis to determine whether a video contains sexual activity with a child. The term AI can also be used loosely to describe a machine's ability to perform repetitive tasks without guidance.
- **Machine learning (ML)** refers to algorithms that leverage new data to improve their ability to make predictions or decisions, without having been explicitly programmed to do so. ML is a widely used form of AI that has contributed to innovations such as speech recognition and fraud detection.
- **Advanced Data Analytics (ADA)** uses subject matter expertise and techniques that are typically beyond those of traditional business intelligence to extract insights and make recommendations from complex data. The techniques vary widely, from data visualisation to complex linear models to language analytics. An example is the use of Risk Terrain Modelling to quantify environmental factors that shape risk mapping and resource deployments.

There is other related AI terminology, which start to overlap with the above, but are included here for completeness. The above and below are considered the most common, but there are others. The additional definitions are:

- **Generative Artificial Intelligence (GenAI)** is artificial intelligence capable of generating text, images, or other media, using generative models. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.²
- **Large Language Models (LLM)** are a subset of GenAI, where an algorithm has been trained on a large amount of text-based data, typically scraped from the open internet, and so covers web pages and - depending on the LLM - other sources such as scientific research, books, or social media posts.³ Examples include OpenAI ChatGPT and Google Gemini.
- **Natural Language Processing** is a computer's attempt to "understand" spoken or written language. It must parse vocabulary, grammar, and intent, and allow for variation in language use. The process often involves machine learning.⁴

² [Generative artificial intelligence - Wikipedia](#)

³ [ChatGPT and LLMs: what's the risk - NCSC.GOV.UK](#)

⁴ [AI Glossary: Artificial intelligence, in so many words | Science](#)

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Purpose

This risk assessment should empower policing to leverage artificial intelligence responsibly and ethically.

It is intended to help policing organisations meet Community Security Policy requirements in a relatively new and fast evolving technology area.

It is important that policing can be innovative with this technology, and so this risk assessment seeks to provide the guardrails, so that innovation can be carried out safely and securely and does not put policing at unnecessary risk of losing public trust and confidence through consequential data loss or loss of policing services.

Audience

Members of the Policing Community of Trust.

More specifically the standard is targeted at, architects, developers, data scientists and security experts tasked with designing and building solutions, applications and plugins leveraging AI related technologies.

The following should also be aware of the content of this standard, in order that they can provide appropriate oversight and governance of the use of AI related technologies within policing:

- Senior Information Risk Owners (SIROs)
- Information Asset Owners (IAOs)
- Information & Cyber risk practitioners and managers
- Auditors providing assurance services to PDS or policing.

Finally, Policing’s reliance on third parties means that suppliers acting as service providers or developing products or services for PDS or policing, should also be made aware of and comply with the content of this standard, in relation to their work on Policing systems and data.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Scope

In scope for this risk assessment are cyber security considerations and requirements for the:

- Acquisition and implementation of solutions that incorporate AI.
- Development of solutions with integrated AI.
- Use of AI technologies, e.g. ChatGPT, Gemini and LLaMA.

Not all the requirements listed below will apply in the same way across each use case. Forces should apply this assessment proportionately, identifying whether a requirement is met directly, needs to be evidenced by the supplier, or acknowledged as a residual risk to be documented and managed appropriately.

This risk assessment is intended to be used in conjunction with the NPCC AI Principles outlined in the Covenant for Using Artificial Intelligence (AI) in Policing to ensure that both cyber security and ethical and legal considerations are comprehensively addressed when adopting AI technologies in policing.

Requirements

This section details the minimum requirements for the acquisition, development and use of AI to protect policing from the loss of confidentiality, integrity or availability of the data or loss of availability of the systems and services it relies upon to meet policing outcomes.

The newness of AI technologies and their potential for driving significant change suggests that these systems demand an entirely fresh set of control requirements to monitor, police and curate them. This is not necessarily the case. These systems in many ways are similar to existing technologies in use across policing organisations today, making them manageable in the same way as those familiar systems.

Most control requirements relating to the acquisition and use of AI and AI-based technologies can be found in the existing standards which have been written to support the National Policing Community Security Policy and Principles. These guidelines apply to AI technologies in the same way they apply to other digital solutions. However, AI technologies and AI integrated technologies can present a unique set of vulnerabilities, and so the requirements below are to address these.

These requirements are not intended to replace those in other standards and while there will be duplication, they are documented here, as they are particularly pertinent to the acquisition, development, and use of AI-based technologies.

This risk assessment and requirements do not standalone; it is important that existing cyber security standards are considered during the acquisition and development of AI-based solutions.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
1	Adoption		
1.1	Before adopting an AI technology, policing organisations must assess whether AI is the most appropriate tool to meet their needs. - Identify and document business requirements. - Evaluate if AI is necessary and how it will effectively meet your requirements. - Follow a 'Secure by Design' (SbD) methodology.	NIST CSF v2.0: PR.PS-06 CIS V8: 16.1	Business requirements document SbD process documentation SbD project artefacts Secure Development Lifecycle
1.2	A business impact assessment must be conducted at the outset to establish a foundation for evaluating both potential benefits and risks of AI adoption, ensuring alignment with the organisation's objectives and security standards.	NIST CSF V2.0: ID.RA-04 CIS V8: 7.3 7.4	Completed Business Impact Assessments
1.3	Establish clear guidelines and documentation for prohibited AI use cases. Ensure end-users understand limitations and restrictions.	NIST CSF V2.0: PR.AT.01 CIS V8: 14.1 14.2	Evidence of approved and prohibited AI use-cases and end-user guidelines.
2	Governance		
2.1	Organisations must exercise caution when using AI technologies, particularly for law enforcement purposes, to mitigate risks related to data confidentiality, integrity and authenticity while supporting responsible decision-making. Mandate a risk assessment process for all AI components, covering governance, known risks, and the use of personal data. Require that all components be sourced from trusted and approved	NIST CSF V2.0: GV.OC-02 GV.SC-01 GV.SC-03 GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 GV.SC-10 ID.RA-09 ID.RA-10 GV.OC-03 CIS V8: 15.1 15.2 15.3	Risk assessment, including details of how any risks are addressed and mitigated. Risk owners identified and assigned.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	providers, with documented provenance details. - Ensure that change management processes are in place that trigger risk assessment reviews. Triggers includes expansion or changes of use. - Identify risk owners and a formal approval from the appropriate risk owner.	15.4 15.5	
2.2	A high level review of data used within the proposed solution must be carried out to review whether AI can be used safely and in compliance with data processing requirements.	NIST CSF V2.0: ID.AM-07 ID.AM-08 PR.DS-10 CIS V8: 3.2 3.3	Evidence of data review, identified data sets which are suitable for AI usage and justification for use.
2.3	An AI governance framework should be established, or AI should be incorporated into a suitable existing governance framework. This governance structure must clearly define AI governance principles, document acceptable AI usage policies, and oversee all AI risk management activities.	NIST CSF V2.0: GV.OC-01 GV.RM-02 GV.RM-03 IM.IM-03 CIS V8: 15.1 17.4 17.5	Governance ToRs Governance Meeting minutes, actions and decision logs
2.4	Identity governance processes must be established to manage identities used by Agentic AI solutions.	NIST CSF V2.0: PR.AA-01 PR.AA-04 PR.AA-05 CIS V8: 5.1 5.2 6.3	Evidence of governance processes to support and manage Agentic AI.
2.5	Threat modelling must be: - used at each stage of AI development, deployment and integration of new components. - used to identify and inform users of all known harmful	NIST CSF V2.0: PR.AT-01 CIS V8: 14.1 14.2	Evidence of conducted threat modelling, identified security issues and proposal of appropriate security controls to reduce the risk

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	states, unmitigated risks, and misuse scenarios.		
3	Data Security		
3.1	Ensure data and technical readiness prior to AI adoption, including: - the use of high-quality, well-governed data. - appropriate security controls. - integration capabilities to support and manage AI technologies effectively.	NIST CSF V2.0: ID.AM-05 CIS V8: 1.1 2.1	Evidence of technical and data readiness for AI adoption
3.2	Ensure AI technologies comply with data protection policies. See ICO toolkit. Consider factors such as data classification, risks associated with personal data and appropriate mitigations. This includes ensuring organisational data is not input into an AI-based technologies without explicit consent.	NIST CSF V2.0: GV.OC-03 PR.DS-01 PR.DS-02 ID.AM-05 CIS V8: 3.2 3.6 3.10 3.11	Evidence that the following has been completed: - DPIA - DPO Engagement Third Party Assurance Reports
3.3	AI technologies must comply with data residency requirements, ensuring that all data, including training datasets, operational data and outputs, are stored and processed within approved jurisdictions.	NIST CSF V2.0: GV.SC-05 ID.AM-08 CIS V8: 15.1 15.2 15.3	Evidence that the following has been completed: - DPIA - Documented encryption and security controls Low Level designs
3.4	Access controls must be enforced to restrict who can read, input or amend data in the AI technologies to prevent unauthorised and unexpected changes.	NIST CSF V2.0: PR.AA-05 CIS V8: 5.2 6.3	Details of access controls used to manage AI and data access.
3.5	Access permissions to reference data should be regularly reviewed to ensure compliance with the principle of least privilege.	NIST CSF V2.0: PR.AA-05 CIS V8: 5.2 6.3	Evidence of: - Regular access review - Review of access privileges

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
4 Training and Reference Data			
4.1	Ensure AI technologies are limited to access only the data necessary for the specific tasks they are designed to perform. Unnecessary access increases security risks and exposes sensitive information.	NIST CSF V2.0: PR.AA-01 PR.AC-02 PR.AA-05 PR.PS-01 CIS V8:	Design decision logs Low Level Designs RBAC and Access Control policies
4.2	Ensure that the intended usage of an AI technology: - Is appropriate to the sensitivity of the data it was trained on or uses as reference data - has the controls required to ensure the security of the data.	NIST CSF V2.0: CIS V8:	Evidence with each AI solution which explains intended use, data it will access and the controls used to manage.
4.3	Data used in the AI solution should be ideally sourced from trusted, ethical and verified sources, and any third-party data sets must meet the same standards for integrity and compliance.	NIST CSF V2.0: ID.AM-03 ID.AM-07 PR.DS-10 CIS V8: 3.2 3.3	Details of data used with AI, where it is sourced from and how it meets requirements for integrity and compliance.
4.4	Data sources must be reviewed and evaluated to ensure they comply with security, privacy, and ethical requirements.	NIST CSF V2.0: ID.AM-08 PR.PS-01 PR.PS-06 CIS V8: 16.1 3.1 3.2	Evidence that the following has been completed: - Risk Assessments - Risk registers - DPIA Third Party Assurance Reports
4.5	Datasets must account for the geographical, contextual, behavioural and functional elements relevant to the specific environment or setting in which the AI technology will be deployed.	NIST CSF V2.0: ID.AM-08 PR.PS-01 PR.PS-06 CIS V8: 16.1 3.1 3.2	Details of data used with AI, where it is sourced from and how it meets requirements for the specific environments.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
4.6	When developing AI solutions, ensure AI training and reference data is securely protected to maintain the integrity and reliability of AI technologies.	NIST CSF V2.0: PR.DS-01 PR.DS-02 CIS V8: 3.1 3.2 3.3 3.6	Third Party Assurance Reports Evidence of supply chain analysis, Pen tests, ITHC reports Independent assessments/certifications (e.g., SOC 2 Type II, ISO/IEC 42001)
4.7	Implement integrity checks and validation techniques to prevent poisoning attacks.	NIST CSF V2.0: PR.DS-01, PR.DS-02, PR.DS-03, DE.CM-09 CIS V8: 3.1 3.2 3.6	Evidence of integrity checks and processes used to support the checks.
4.8	Verify that data used by the AI solution is encrypted both at rest and in transit, with encryption protocols that meet industry standards.	NIST CSF V2.0: PR.DS-01, PR.DS-02, PR.DS-03, DE.CM-09 CIS V8: 3.1 3.2 3.6	Evidence of where data encryption is used to protect data in transit and at rest.
4.9	Prioritise the use of approved LLMs and only pioneer with new LLMs when absolutely necessary. When using new pre-trained models (e.g., open source LLMs), commonly used as a base for AI technologies, ensure assurance is carried out to assess for security risks.	NIST CSF V2.0: ID.RA-01 ID.RA-03 ID.RA-04 CIS V8: 7.1 7.2 7.3 7.4 18.2	Evidence of assurance for pre-trained models.
4.10	Consider the risks of using sensitive and private data as reference data, and ensure that any sensitive data aspects have been obfuscated to minimise the impact of data leakage or making it harder to reconstruct	NIST CSF V2.0: PR.DS-01 PR.DS-02 CIS V8: 3.1 3.2 3.3 3.6	Evidence of processes and guidance with the use of reference data.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
5	People and Process Security		
5.1	Ensure you have the requisite skilled people, process and technology elements required to set up, run and maintain operational AI technologies and cope with their outputs.	NIST CSF V2.0: DE.CM-01 DE.CM-03 DE.CM-06 DE.CM-09 ID.AM-03 GV.OC-02; GV.RR-02 ID.AM-06 ID.AM-08 PR.AT-02 PR.IR-01 PR.IR-04 PR.PS-01 PR.PS-03 CIS V8: 8.1 8.2 8.5 8.11 1.1 2.1 3.2 3.3 14.1 14.2 4.1 4.2 4.4 11.1 11.4	Evidence that the following have been completed: • Project Plans • Target Operating Models • Role Descriptions/Skills Matrices • Process Documentation and Technology
5.2	Clearly define roles and responsibilities for AI oversight, ensuring that accountability for AI-driven decisions is explicitly assigned to individuals or teams.	NIST CSF V2.0: GV.OC-02 GV.RR-02 CIS V8: 17.5	RACI Matrix Decision Review and Escalation processes Governance meetings minutes
5.3	AI technologies must support, not replace, human oversight, ensuring informed decision-making and accountability. Keep human operators responsible for AI-assisted decisions, ensuring final	NIST CSF V2.0: PR.AT-02 CIS V8: 14.1 14.2	Audit reports, decision-making processes and governance policies Risk assessments and ownership Evidence of defined roles and responsibilities

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	accountability remains with personnel overseeing AI use.		
5.4	Establish clear human checkpoints within AI-driven processes to review critical decisions, challenge AI outputs and apply human expertise where necessary.	NIST CSF V2.0: PR.AT-02 CIS V8: 14.1 14.2	Evidence of processes used to review and validate AI outputs and decisions.
5.5	Educate all staff and officers about the risks of AI technologies and the appropriate use of these tools, including where and how they are authorised to use them. Cybersecurity training must incorporate AI security content, which should be reviewed and updated regularly to address new threats.	NIST CSF V2.0: GV.RR-01 PR.AT-01 PR.AT-02 PR.AA-01 PR.AA-02 CIS V8: 5.2 6.3 14.1 14.2 17.5	Education and Awareness (E&A) materials relating to the use of AI and E&A tracking and measurement of effectiveness of E&A campaign
6	Technical Security		
6.1	Continuously update and maintain threat intelligence feeds to monitor for attacks or compromises of AI technologies.	NIST CSF V2.0: ID.RA-02 ID.RA-03 ID.RA-08 DE.AE-02 RS.CO-03 RS.CO-05 RS.AN-05 CIS V8: 7.1 7.2 7.3 7.4	Evidence of threat intelligence feeds and their integration into monitoring solutions.
6.2	Ensure evolving AI technology and security threats are reviewed, and security policies and standards are regularly reviewed and updated to reflect changes.	NIST CSF V2.0: ID.AM-03 PR.DS-01 PR.DS-05 GV.PO-01 RS.CO-02 PR.DS-05 GV.PO-01 RS.CO-02 CIS V8:	Evidence of up-to-date cyber security policy and standards relating to AI

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
		3.1 3.2 3.7 3.8	
6.3	When using bespoke or local LLM solutions, technologies must be kept up to date with regular security patches, updates, and performance improvements. Organisations must ensure that all components, including models, software, and underlying infrastructure, receive timely updates to protect against emerging threats and vulnerabilities.	NIST CSF V2.0: RS.AN-05 RS.MI-03 CIS V8: 7.1 7.2 7.3 7.4 4.8	Patch management processes Low level designs
6.4	Vulnerability management tools and techniques must be used to regularly monitor and address vulnerabilities within AI technologies. These tools should scan AI environments to identify at risk components and harden them against potential security breaches or compromise. Vulnerabilities must be patched promptly, and security controls should be strengthened to minimise risks.	NIST CSF V2.0: DE.CM-08 ID.RA-01 ID.RA-02 PR.DS-06 PR.PS-02 PR.PS-01 RS.AN-05 RS.MI-03 CIS V8: 7.1 7.2 7.3 7.4 4.1 4.2 4.11	Establish VM processes and policies VM tool reports VM remediation logs
6.5	AI technologies must be protected against adversarial attacks that can manipulate their behaviour, leading to incorrect or malicious outputs. To mitigate these risks: Regularly perform penetration tests to identify and address vulnerabilities in the AI technologies. For third-party systems, obtain	NIST CSF V2.0: DE.AE-02 DE.AE-03 DE.AE-04 DE.AE-05 DE.CM-06 DE.DP-05 CIS V8: 8.1 8.2 8.5 13.1	Low Level Designs Evidence of supply chain analysis, ITHC/Pen Test reports

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	and review recent test results. • Validate and sanitise all incoming data to prevent malicious input from compromising the model. • Continuously monitor outputs for signs of unexpected or manipulative behaviour, with alerts for suspicious activity.	13.6 18.2	
6.6	Controls and monitoring must be deployed to actively monitor, detect and prevent prohibited AI platforms and services.	NIST CSF V2.0: DE.CM-01 DE.CM-06 DE.CM-09 CIS V8: 8.1 8.2 8.5	Evidence of controls and their alignment to AI platforms and services.
7	Managing AI Content		
7.1	Data used by the AI solution must be regularly tested to identify and mitigate bias.	NIST CSF V2.0: DE.CM-09 CIS V8: 8.5 18.2	Evidence of processes used to review AI platforms and test for bias.
7.2	Ensure that suppliers are able to provide proof of bias testing, along with evidence of mitigation measures and quality control processes.	NIST CSF V2.0: GV.SC-02 GV.SC-05 CIS V8: 15.1 15.2	Evidence from suppliers for how they manage bias in their solutions.
7.3	AI generated content must be independently verified to confirm accuracy and prevent misinformation. To ensure outputs are reliable, policing organisations should: • Implement automated validation mechanisms. • Should consider non-deterministic AI outputs • Cross-check AI outputs against reliable sources.	NIST CSF V2.0: PR.AT-02 PR.PS-06 CIS V8: 14.1 14.2 16.1	Evidence of Test reports, Audit Logs, Low level design documents Third Party Assurance reports

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	Establish human oversight processes to ensure that AI decisions are accurate and fair.		
7.4	AI technologies must provide clear, understandable and consistent explanations for decisions made. Therefore, policing organisations should: - Evaluate AI technologies against diverse user groups to identify potential biases and disparities in decision-making. - AI technologies must allow users to challenge, review and override AI outputs.	NIST CSF V2.0: ID.IM-03 CIS V8: 17.1 17.2	Evidence from suppliers for how their solution meets requirements.
7.5	Establish internal procedures to regularly review and audit decision making to ensure output remains accurate.	NIST CSF V2.0: PR.AT-02 CIS V8: 14.1 14.2	Evidence of procedures to manage AI content.
7.6	AI data sources, including external supplier feeds, live operational data and datasets, can be actively targeted to undermine model reliability. To mitigate these risks, organisations should: - Regularly audit and validate AI data sources to spot anomalies, inconsistencies or suspicious patterns. - Establish alerts for unexpected data types or results. - Ensure clear records of data origin and any modifications to track and verify AI data integrity.	NIST CSF V2.0: DE.AE-02 DE.CM-01 DE.CM-03 DE.CM-06 DE.CM-09 ID.AM-03 PR.AT-03 RS.AN-01 CIS V8: 8.1 8.2 8.5 8.11 3.2 14.3	Evidence of data sampling Low Level Designs Evidence of supply chain analysis, ITHC/Pen Test reports

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	- Scrutinise the supply chain for gaps that attackers can insert themselves into. - Regularly assure connections with external suppliers for any opportunities for model loss, damage, or pollution of the dataset.		
8	Agentic AI		
8.1	Identities used by AI must be fully governed in line with any other operational identities. The agents will have access to business platforms and will carry out activities aligned to standard business processes and will therefore need to be governed appropriately. Agent identities: - must be managed/ sponsored by a suitable responsible owner. - must be subject to regular access reviews to ensure permissions are appropriate and align with least privilege	NIST CSF V2.0: PR.AA-01 PR.AA-04 PR.AA-05 CIS V8: 5.1 5.2 6.3	Evidence of governance activities and processes to manage agentic identities appropriately.
8.2	Agentic processes and permissions must consider separation of duties and create separate agents and identities to maintain this and limit the impact of agent misuse or compromise.	NIST CSF V2.0: PR.AA-01 PR.AA-05 CIS V8: 5.1 5.2 6.3	Evidence of governance and processes to manage alignment to separation of duties and avoidance of excessive permissions
8.3	Agentic AI solutions must have a clearly identified business purpose and identified benefits which support the use of AI.	NIST CSF V2.0: GV.RM-07 ID.IM-03 CIS V8: 15.2 15.5 17.1 17.2	Evidence of business purpose, benefits case and use-cases.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
		17.4 17.5	
8.4	The Total Cost of Ownership (TCO) must be considered and detailed to ensure that the use of AI is cost-effective both in the short-term and also as part of its ongoing use. Considerations should include: • Licensing model. • Support costs. • Bundled compute costs or allowances. • Additional compute or 'token' costs. • Onboarding and ongoing management.	NIST CSF V2.0: GV.RM-07 GV.SC-06 GV.SC-07 ID.IM-03 IR.RA-10 CIS V8: 15.1 15.2 15.4 15.5 17.1 17.2 17.4 17.5	Evidence of TCO review and the considerations.
8.5	Agentic solutions should include business process maps to show the role(s) carried out by the agents and interactions with business systems and end-users. Business process flows will aid with assurance processes and provide transparency for where agentic AI is used.	NIST CSF V2.0: GV.RM-07 ID.IM-03 ID.AM-04 ID.AM-08 CIS V8: 15.1 15.2 15.3 15.5 3.2 3.3 17.1 17.2	Business process maps and process flows, showing the usage of agentic AI and interactions.
9	Business Continuity		
9.1	Clarify business impact of AI disruption (e.g. outages) and design for the appropriate levels of uptime and resiliency. Ensure the robustness of the platforms supporting the AI technology meet the levels of uptime and resiliency required.	NIST CSF V2.0: PR.AC-04 GV.OC-04 PR.IR-03 CIS V8: 6.1 6.2 6.3 11.1 11.2 11.4	Evidence that the following has been completed: • Business Impact Assessment • Risk Register • Service Level Agreements • Low Level Designs Design Decision Logs
92	Update disaster recovery plans to address AI-specific risks,	NIST CSF V2.0: ID.IM-04	Business continuity plans ensure that personnel can

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	including adversarial attacks and data poisoning, ensuring readiness for prompt recovery.	CIS V8: 11.1 11.2 11.4 17.4 17.6	perform tasks that were automated by AI. Evidence of updated and tested Disaster Recovery and Business Continuity plans
9.3	Ensure Business Continuity Plans consider the ability of personnel to perform tasks undertaken by AI during a system outage or disruption.	NIST CSF V2.0: ID.IM-04 CIS V8: 11.1 11.2 11.4 17.4 17.6	
9.4	To minimise service disruption and reduce AI technology exposure to threats, policing organisations should: - Implement appropriate levels of network segregation and access control to prevent lateral movement and direct access. - Deploy AI models in isolated environments. Only use or deploy a dedicated tenant-specific model, avoiding the use of shared models that may be trained on data from multiple tenants. - Use red team testing techniques to explore possible attack patterns. - Limit AI model exposure to internal systems and manage how much information the AI model returns to queries to limit the ability of threat actors to gather useful attack data. - Gain documented assurances from AI vendors regarding blocking the use	NIST CSF V2.0: DE.AE-02 DE.AE-03 DE.CM-01 DE.CM-03 DE.CM-06 DE.CM-09 ID.RA-02 ID.RA-03 PR.AC-01 PR.AC-04 PR.AC-05 PR.AC-06 PR.IR-01 PR.PS-01 RS.AN-05 RS.CO-05 RS.IM-01 RS.RP-01 CIS V8: 8.1 8.2 8.5 8.11 6.1 6.2 6.3 7.1 7.2 7.3 7.4 4.1	Demonstration of alignment to the NCSP Network Security Standard. Evidence that the following has been completed: - Business Impact Assessment - Low Level Designs - Design Decision Logs - Threat Intelligence Reports Red Team Reports

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
	of organisation data (prompts, responses etc) to train models.	4.2 4.11 17.4 17.5 11.1 11.4	
9.5	Conduct resilience testing and implement mitigation strategies accordingly.		
10	Monitoring		
10.1	All AI decisions and interactions must be securely logged with tamper resistant storage to prevent unauthorised modifications. Logs must be encrypted, stored immutably, and reviewed regularly for anomalies. Suppliers must confirm that their logs are comprehensive, include critical interactions and meet encryption and tamper-resistance standards. Access to logs must be carefully controlled through role-based access, ensuring that only authorised personnel can view log data, based on their role and the level of data sensitivity.	NIST CSF V2.0: PR.DS-01 PR.DS-02 PR.DS-10 CIS V8: 3.1 3.2 3.3 3.6	Evidence of alignment to the NCSP logging and monitoring standard. Design decision log Low Level Design documents Security Operations Centre Playbooks
10.2	Develop tailored logging and monitoring requirements for each AI instance, ensuring continuous tracking of AI behaviour and data inputs/outputs. This includes monitoring for anomalous behaviour, signs of manipulation and potential security breaches. Logs should be reviewed regularly for suspicious activities and real time alerts should be set up for immediate responses to any incidents.	NIST CSF V2.0: DE.AE-03 DE.CM-01 DE.CM-03 DE.CM-06 DE.CM-09 DE.DP-02 DE.DP-04 ID.AM-03 PR.IR-02 RS.AN-01 RS.CO-02 CIS V8: 8.1 8.2 8.5	Evidence of alignment to the NCSP logging and monitoring standard. Design decision logs Low Level Designs Security Operations Centre Playbooks AI monitoring use cases

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
		8.11 3.2 13.1 13.6	
11	Incident Response		
11.1	Establish AI-specific incident response plan to address the unique challenges of AI-related attacks. This plan should include clearly defined workflows for incident management, including rollback and shutdown procedures, as well as rapid deactivation protocols for swift risk mitigation. To ensure effectiveness, incident response strategies should be regularly tested through tabletop exercises and AI-specific playbooks should be developed to guide incident handling.	NIST CSF V2.0: DE.AE-02 DE.AE-03 DE.AE-04 DE.AE-05 DE.CM-01 DE.CM-03 DE.CM-06 DE.CM-09 DE.DP-02 DE.DP-04 DE.DP-05 PR.PS-01 PR.PS-02 RS.AN-01 CIS V8: 8.1 8.2 8.5 8.11 13.1 13.6 4.1 4.2 7.1 7.2 7.3 7.4 17.4	Evidence of alignment to the NCSP Cyber Incident Response standard. Evidence of an AI-specific Incident Response Plan (IRP) IRP Test Plans IRP Test results IRP improvement plans Security Operations Centre Playbooks
11.2	All AI security incidents must be logged, reported, and analysed. Suppliers must confirm that detailed incident analysis reports are generated and shared with affected customers.	NIST CSF V2.0: PR.PS-04 RC.CO-03 RC.IM-01 RC.IM-02 CIS V8: 17.4 17.6 11.1 11.2 11.4	Evidence of alignment to the NCSP Cyber Incident Response standard. Incident response plans reports, post incident reviews and improvement plans
12	Asset Management		

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
12.1	All AI-related assets must be recorded in the relevant asset register. Each record should clearly specify the asset's approved use, ownership, and any other information relevant to the AI technology and its usage. The asset register should be regularly reviewed and updated to ensure accuracy and completeness.	NIST CSF V2.0: ID.AM-01 ID.AM-02 CIS V8: 1.1 1.2 1.4 2.1 2.2 2.4	Evidence of alignment to the NCSP Application Management standard. Technology Asset Inventory
14	Supplier Security		
14.1	Verify that suppliers and open-source components meet existing security standards on identity management, access control and authentication.	NIST CSF V2.0: GV.SC-07 ID.RA-10 CIS V8: 15.1 15.2 15.3 15.4 15.5	Evidence that the following has been completed: • Third Party Assurance Reports • Legal/Regulatory guidance obtained. • Business Impact Assessments Identified risks have been recorded in relevant risk registers.
14.2	Ensure supply chain assurance extends to software and service providers involved in the development and maintenance of AI technologies.	NIST CSF V2.0: GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 CIS V8: 15.1 15.2 15.3 15.4 15.5	Evidence that the following has been completed: • Third Party Assurance Reports • Legal/Regulatory guidance obtained. • Business Impact Assessments Identified risks have been recorded in relevant risk registers.
14.3	Ensure third-party suppliers are assessed for any embedded use of AI within their services.	NIST CSF V2.0: GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 CIS V8: 15.1 15.2 15.3 15.4 15.5	Evidence from suppliers for their use of AI within their provided services.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Ref.	Minimum requirement	Control reference	Compliance Metric
14.4	Ensure suppliers provide transparency regarding data provenance and handling.	NIST CSF V2.0: GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 CIS V8: 15.1 15.2 15.3 15.4 15.5	Evidence from suppliers for their data handling procedures and data provenance within their AI solutions.
14.5	Ensure suppliers provide full visibility of AI usage in their products, including: • What data is processed by AI • Where the data is processed • Where the data is stored	NIST CSF V2.0: GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 CIS V8: 15.1 15.2 15.3 15.4 15.5	Assertions from suppliers showing what data is processed and locations used to process and store data.
14.6	Review ability to enable and disable AI functionality within products and services, and favour suppliers who can provide granular controls over AI usage within their products.	NIST CSF V2.0: GV.SC-04 GV.SC-06 GV.SC-07 GV.SC-09 CIS V8: 15.1 15.2 15.3 15.4 15.5	Evidence of AI functionality within selected products and how the AI functionality can be managed and disabled if not appropriate.

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Communication approach

This document will be communicated as follows:

- Internal peer review by the members of the National Cyber Policy & Standards Working Group (NCPSWG), which includes PDS and representatives from participating forces.
- Presentation to the National Cyber Policy & Standards Board (NCPSB) for approval.
- Formal publication and external distribution to PDS community, police forces and associated bodies.

Measurables generated by adopting this standard can also form part of regular cyber management reporting.

This standard should be distributed within IT and project teams to help complete an initial gap analysis which can inform any implementation plan. This implementation plan can be shared with force SIROs / Security Management Forum / Information Management. Consideration should also be given to raising awareness amongst force personnel of the implementation of this standard where it may affect them.

This standard should be mapped to a project lifecycle and internal governance prior to adoption. Following this, it should be provided to the Information Assurance communities and PMO's and should also be shared with procurement & commercial leads to ensure this is built into procurement activities.

Measurables generated by adopting this standard can also form part of regular Cyber management reporting and audit evidencing.

Document Compliance Requirements

(Adapt according to local policy needs.)

Equality Impact Assessment

(Adapt according to local policy needs.)

Appendices

PLEASE DELETE IF NO APPENDICES

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Review

This document will be reviewed at least annually (from the date of publication) and following any major change to Information Assurance (IA) strategy, membership of the community, or an identified major change to the cyber threat landscape. This ensures IA requirements are reviewed and that the policy continues to meet the objectives and strategies of the police service.

Related documents

DOCUMENT NAME	VERSION AND DATE
CIS Controls	8.1, June 2024
NIST Cyber Security Framework	2.0, February 2024
Covenant for Using AI in Policing	V1.1
AI Playbook for the UK Government - GOV.UK	February 2025
Code of Practice for the Cyber Security of AI - GOV.UK	January 2025
Principles for security of Machine learning ML - NCSC.GOV.UK	V2.0, May 2024
Guidelines for secure AI system development - NCSC.GOV.UK	V1.0, November 2023
OWASP Top 10 for Large Language Model Applications OWASP Foundation	V2025, November 2024
Why AI data quality is the key to AI success	

Table 1 - Related Documents

Vendor compliance and transparency – Microsoft Copilot

DOCUMENT NAME	VERSION AND DATE
Transparency and Control in Consumer Data Use	
Data, Privacy, and Security for Microsoft 365 Copilot	
Responsible AI Principles and Approach	
Microsoft 2025 Responsible AI Transparency Report	
Transparency Note for Azure OpenAI in Microsoft Foundry Models	
Transparency Note for Microsoft Copilot	

Table 2 - Vendor Compliance and Transparency

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE

Document Information

Document Location:

National Policing Policies & Standards Platform -

<https://knowledgehub.group/web/national-standards>

Version History

VERSION	AUTHOR	REVISION	DATE
1.0	PDS Cyber	Initial version	07/05/26
1.1	PDS Cyber	Annual review amendments & new template. Title changed from AI risk assessment to AI Secure Implementation standard at NCPSB 28/05/26	27/04/26

Approvals

VERSION	BODY	TITLE	DATE
1.0	NCPSB	National Cyber Policy & Standards Board	22/05/26
1.1	NCPSB	National Cyber Policy & Standards Board	28/05/26

DATE:	27/04/2026
VERSION:	1.1
DOCUMENT REFERENCE:	PDS-CSP-STD-AIR

COPYRIGHT:	2026
NUMBER OF PAGES:	26
CLASSIFICATION:	OFFICIAL FOR PUBLIC RELEASE