GIG CYMRU NHS WALES
Addysg a Gwella Iechyd Cymru (AaGIC)
Health Education and Improvement Wales (HEIW)

User Guide
Safeguard AI Practices: Guardrails for Safe AI Use within HEIW

Status: **Approved**

Version: 1.0

Publish Date: 03/04/2025

Author: *Clinical Lead for Artificial Intelligence*

Approver: *Assistant Director of Data & Analytics*

Owner: **HEIW Workforce & Data Analytics**

Review Date: **03/04/2026**

Table of Contents

1 Document Control

1.1 Version History

Amended By	Version	Status	Date	Summary of Changes
Clinical Lead For Artificial Intelligence	0.1	Published	03/04/2025	Final document put into correct template
		Choose an item.	Click or tap to enter a date.	
		Choose an item.	Click or tap to enter a date.	
		Choose an item.	Click or tap to enter a date.	

1.2 Distribution

This document has been distributed to:

Name	Date	Version	Inform/ Review/ Approve
Assistant Director of Data and Analytics	03/04/2025	1.0	Approved
Head of Digital Transformation	03/04/2025	1.0	Approved
	Click or tap to enter a date.		Choose an item.
	Click or tap to enter a date.		Choose an item.
	Click or tap to enter a date.		Choose an item.

1.3 Document Location

The master copy of this document is held at the following SharePoint location:

Digital, Data and Engagement - User Guides - All Documents

2 Safeguard AI Practices: Guardrails for Safe AI Use within HEIW

AI is a powerful tool but must be used responsibly. The principles shown here will help ensure that we protect data, comply with legal and ethical standards, and harness AI's potential effectively and safely. Always seek guidance from your information governance (IG) office if in doubt.

Data Sensitivity and Privacy

Principle: Never input sensitive, confidential, or personal data into any AI system, regardless of whether it is internal or external to the organisation.

Why: AI systems may store or use data for training purposes or inadvertently expose it to unauthorised individuals and lead to unintended exposure or data breaches.

Understand Data Sharing Risks

Principle: Microsoft Copilot operates within a single-tenant framework. This means that it thinks HEIW, HBS/Trusts and PHW and NHS executive are all one organisation. Be mindful that data shared could be visible to others in the tenant / NHS.

Also: remember by using LLMs, your inputs (what you write / say / upload) and the LLM outputs (what the model says, writes or creates) may then be used to further train the model. The option to decline the model using your data for training is available and should be considered. **This does NOT mitigate the risk of data breaches.**

Why: Prevent unintended visibility or breaches of data confidentiality.

Comply with GDPR and Other Regulations

Principle: Always ensure AI use complies with data protection laws such as GDPR. If in doubt check with your information governance team.

Why: Avoid legal repercussions and protect individual rights.

Validate AI Outputs

Principle: Treat AI-generated content as a starting point, not a final answer. Always validate outputs for accuracy and appropriateness.

Why: AI may generate plausible but incorrect or biased results.

Avoid Overreliance

Principle: Use AI to assist, not replace, human judgment or expertise.

Why: Preserve accountability and ensure critical thinking remains central.

Transparency in AI Use

Principle: Clearly communicate when AI tools are being used, especially in client or public-facing contexts.

Why: Build trust and avoid potential ethical concerns.

Avoid Plagiarism

Principle: Do not copy and use AI-generated content verbatim without proper review and attribution.

Why: Uphold intellectual property rights and originality.

Regular Training and Awareness

Principle: Ensure all users receive training on safe AI use, potential risks, and best practices especially regarding prompting if using Large language models (LLMs). HEIW content is being produced and once approved will feature here.

Why: Reduce the risk of misuse and enhance informed usage.

Do Not Input Passwords or Security Details

Principle: Never input login credentials, PINs, or any security-related information into AI tools. When registering for AI tools such as ChatGPT with a work email address, it is mandatory to use a unique and strong password. The password must be different from those used for other systems.

Why: Prevent unauthorised access or security breaches.

Respect Intellectual Property

Principle: Avoid using AI to generate or manipulate copyrighted material without proper permissions.

Why: Stay compliant with intellectual property laws.

Report Risk / Misuse concerns

Principle: Report any suspected misuse of AI tools or breaches to the designated data and / or information governance officer.

Why: Ensure accountability and maintain organisational integrity.

Monitor Ethical Implications

Principle: Be mindful of AI's impact on fairness, equity, and inclusion. AI models are trained on what may be biased data therefore outputs can be biased. AI can also make things up or hallucinate.

Why: Avoid perpetuating biases or causing harm.

Keep Up with Updates

Principle: Stay informed about updates or changes in the organisation's AI tools or policies.

Why: Ensure ongoing compliance and effective use.

Test AI Responsibly

Principle: Experiment with AI in controlled environments using non-sensitive data.

Why: Prevent accidental breaches during testing or exploration.

Ensure Data Anonymisation

Principle: Anonymise data before inputting it into AI tools. Remove any personal or identifying information from data before using it in AI tools to keep it private and secure.

Why: Enhance data privacy and protection.

Be a Critical Thinker

Principle: Approach AI recommendations critically, questioning their relevance and accuracy. Remember AI models can make things up and be biased.

Why: Maintain human oversight and accountability.

The following pneumonic can be used as an aid memoir to use AI safely.

Mnemonic: "SAFEGUARD AI PRACTICES"

S: Secure Data and Systems

- **S**ensitive data must not be shared.
- **A**ccess controls should be strictly maintained.
- **F**ollow GDPR and data protection laws.
- **E**nsure anonymisation of data where possible.

G: Guard Against Misuse

- **G**overn AI use through training and awareness.
- **U**se only approved AI tools.
- **A**void plagiarism and respect intellectual property.
- **R**eport misuse or breaches immediately.
- **D**o not input passwords or security details.

A: Assess AI Outputs

- **A**lways validate AI outputs for accuracy.
- **I**nform others when AI is used to ensure transparency.

P: Protect Ethical Standards

- **P**revent overreliance by preserving human oversight.
- **R**espect fairness and inclusion to avoid bias.
- **A**lign AI use with organisational goals.
- **C**autiously test AI with non-sensitive data.
- **T**hink critically about AI recommendations.

I: Implement Responsible Usage

- **I**dentify risks of external AI tools.
- **C**ommunicate the risks of data sharing.
- **E**ducate yourself on updates to AI policies.
- **S**tay vigilant about ethical and legal implications; report concerns.