

Northumbria University AI Principles

Background

Northumbria University is committed to unlocking potential, transforming lives, and creating new knowledge that benefits society; we are committed to using artificial intelligence critically and ethically to achieve these strategic ambitions. Generative AI tools can support experimentation, enhance productivity, and facilitate new ways of working, but they may produce inaccurate or biased outputs, and must not be relied upon as a replacement for professional judgement, academic rigour, or subject expertise. AI is not a replacement for thinking, but it can support student learning and the professional and academic work of staff.

These principles apply to all staff and students at Northumbria University, including those based at partner institutions and on transnational education programmes. They set out the values and expectations that underpin our approach to the use of artificial intelligence across all aspects of our work. They form the foundation of the Northumbria AI Policy Framework, underpinning all related policies and guidance.

Artificial intelligence is an umbrella term for a range of technologies that enable computer systems to perform tasks that would typically require human intelligence. For the purposes of the Northumbria AI Policy Framework, the most relevant category is **generative AI**, systems that can produce text, images, code, audio, video, and other content in response to user prompts. Tools such as Claude, ChatGPT, and Microsoft Copilot are examples of generative AI, and specifically of **large language models (LLMs)**, systems trained on vast quantities of text that can generate fluent, contextually relevant responses across almost any topic.

It is important to understand that generative AI does not think, reason, or understand in the way humans do. It generates outputs by predicting what text (or other content) is most likely to follow a given input, based on patterns in its training data. This means AI outputs can be fluent and plausible while also being factually wrong, biased, or misleading. Critical evaluation of AI-generated content is always necessary. AI supports thinking, but it cannot replace it.

AI and education strategy

AI can actively support two of Northumbria's core educational commitments: Universal Design for Learning and Experiential Learning. AI tools offer flexible, personalised support that can reduce barriers for students with disabilities, neurodiverse students, and those for whom English is not a first language. It is not a substitute for reasonable adjustments, but an additional means of engaging with learning. AI is most valuable as a study assistant, for testing ideas, exploring problems, seeking feedback, and rehearsing professional practice.

AI and risk

Not all AI use carries the same level of risk, but one principle applies in every case: a human being is always accountable for any output, decision, or action that involves AI. Wherever AI is used, the user must understand what the AI has produced, be able to evaluate it critically, and be prepared to take full responsibility for it. It is also important to be aware that over-reliance on the uncritical use of AI can reduce the capacity for independent thinking, judgement, and the development of expertise.

Low-risk use involves tasks where human review is straightforward, and the consequences of error are limited. Examples include using AI to brainstorm ideas, draft routine communications, summarise notes, or check grammar and formatting. In these cases, a human remains clearly in control, the output is easily verified, and any mistakes are unlikely to cause significant harm.

Medium-risk use involves tasks where outputs require careful human scrutiny before use, and where errors could have meaningful consequences. Examples include using AI to support research, generate teaching materials, analyse data, or draft policy documents. Human expertise is essential to evaluate the quality and accuracy of the output, and responsibility for the final product always rests with the person using it.

High-risk use involves tasks where errors or bias in AI outputs could have serious consequences for individuals or communities, or where there is a risk that AI use could undermine trust, fairness, or accountability. Examples include using AI to inform decisions about students' academic progress, generating content that will be shared publicly without independent verification, or using AI as the basis for medical, legal, or financial advice. High risk use is strongly discouraged, and in many cases is specifically prohibited.

Approved AI tools

Northumbria University has currently approved two large language models for general use by staff and students: Claude by Anthropic and Microsoft Copilot. Both tools have been procured and configured specifically for institutional use, which means that information entered into them is not used to train the models or accessible outside of the institution. Staff and students can use these tools with confidence that their data are handled in accordance with the University's data protection obligations. Use of other AI tools is not prohibited, but users should be aware that they may not offer the same data protections and should exercise caution about what information they share with them.

Guidance on approved tools and support for their use is available in the [AI hub](#).

Prohibited use

Certain uses of AI are prohibited at Northumbria University. This applies to all uses of AI, including institutionally approved systems and other tools. These include any use that violates [Anthropic's acceptable usage policies](#), which set out baseline restrictions on harmful, deceptive, and illegal content. Northumbria's own prohibitions build on this foundation and reflect our specific institutional responsibilities.

The following uses are prohibited:

- Using AI to generate content that is discriminatory, abusive, or intended to harass or harm any individual or group.
- Using AI to produce, distribute, or assist in the production of misinformation or deliberately misleading content.
- Using AI to make or substantially inform decisions about students' academic progress, outcomes, or welfare without appropriate human review and accountability. In line with UK GDPR, AI must not be used for solely automated decision-making producing legal or similarly significant effects.
- Using AI to mark, grade, or assess student work.
- Using AI to directly generate feedback on student work from an assessment submission. (AI may, however, be used to develop or refine notes or voice memos into written feedback, provided the judgements expressed are the assessor's own.)
- Uploading copyright material into a non-university sanctioned tool without the permission of the copyright holder.

- Entering personal, sensitive, or confidential data into AI tools without explicit institutional approval. Entering special category data (racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data (where used for identification), health data, and data concerning sex life or sexual orientation) is prohibited unless subject to Data Protection Impact Assessment and Data Protection Officer approval.
- Using AI to generate content that infringes copyright or misrepresents the origin or authorship of work.
- Using AI in ways that breach Northumbria's existing policies on academic misconduct, digital use, data protection, or professional conduct.

This list is not exhaustive. Where there is uncertainty about whether a particular use is permitted, staff and students should seek guidance before proceeding.

Northumbria University AI Principles

1. AI should support human judgement and never replace it. AI can enhance human decision-making but cannot replace it. Academic judgement, professional expertise, and human responsibility remain central to everything we do. AI must never be used to make decisions that require human accountability, including the assessment and marking of student work.

2. AI is a tool for everybody, not a source of new inequality. Northumbria is committed to ensuring that all students and staff have equitable access to high-quality AI tools as unequal access to AI creates and entrenches disadvantage. Our institutional provision of sector-leading AI tools reflects this commitment.

3. Embed critical AI literacy throughout our programmes and professional practice. Engaging with AI requires more than technical competence. Staff and students are expected to develop critical AI literacy: an understanding of how AI systems work, their limitations and biases, and their ethical, social, environmental, and political implications. Using AI responsibly, or choosing when not to use it, means making thoughtful and ethical judgements.

4. Be transparent about how we use AI. Use of AI should be open and declared. Where AI is used to generate outputs or content, this must be transparently disclosed with appropriate attribution. Staff and students are expected to be honest about when and how AI has contributed to their work. Undisclosed AI use that misrepresents the origin of work is a breach of academic and professional integrity.

5. Use AI responsibly and ethically. AI must be used responsibly, appreciating that AI systems can produce outputs that are inaccurate, biased, or discriminatory and they reproduce dominant and privileged forms of knowledge. Users are responsible for critically evaluating AI-generated content, identifying errors, and not presenting AI outputs as authoritative without verification, with due regard for data protection, intellectual property, copyright, and the privacy of others. Content that is harmful, discriminatory, or misleading must never be generated or shared. Use of AI that could cause harm to individuals or communities is prohibited.

6. Appreciate the costs of AI and minimise our impact. We recognise that AI systems carry significant environmental and human costs, including the energy demands of training and running large language models, water consumption in data centre cooling, and the labour conditions involved in AI development and content moderation. We have a commitment to minimising these impacts, and we will continue to scrutinise the environmental and ethical credentials of the AI tools

we adopt and recommend. Staff and students are encouraged to be mindful of the footprint of their AI use and to avoid unnecessary or resource-intensive use where alternatives exist.

7. Support informed choices on AI use. We are committed to ensuring that staff and students have access to the information and support they need to make genuinely informed decisions about their AI use. Where AI use is essential as part of a programme or work activity, we respect the range of opinions on AI use and endeavour to make reasonable adjustments where possible.

8. Be aware of the impact of AI on wellbeing. AI use should support, not undermine, the wellbeing of staff and students. Concerns about AI-related anxiety, dependency, deskilling, or professional identity are legitimate and will be taken seriously. Northumbria is committed to ensuring that AI is introduced and embedded in ways that are supportive and productive.

9. Base integrity decisions on judgement not automated detection. AI detection tools can be unreliable and must not be used as the basis for misconduct allegations. The scores from AI detection tools should be considered advisory only and must never be the sole grounds for initiating misconduct proceedings. Individual judgement, not automated detection, is the foundation of our approach to academic and professional integrity.

10. Ensure active governance and review. Given the pace of change in AI development, this framework will be reviewed annually to ensure it remains current, relevant, and fit for purpose.

AI Use Disclosure

This document was developed with the assistance of Claude by Anthropic, which was used throughout the drafting process to generate, refine, and redraft text in response to detailed human direction, to sense-check language and policy framing, and to support the synthesis of consultation feedback. All substantive decisions, including the policy positions, principles, and values expressed in these documents, were made by the human author. The final content reflects human authorship and judgement, and responsibility for it rests entirely with the author.