University of St Andrews

Use of University personal data and other information with Generative Artificial Intelligence (GenAI)

Document type	**Policy**
Scope (applies to)	Staff and students
Applicability date	07/02/2024
Review / Expiry date	05/02/2025
Approved date	07/02/2024
Approver	Chief Information Officer
Document owner	Associate Chief Information Officer (Cyber Security and Resilience)
School / unit	IT Services
Document status	Published
Information classification	Public
EDI review/Equality impact assessment	None
Key terms	Information governance and management/Information security
Purpose	To communicate the risks as presently understood to maintaining the confidentiality, integrity and availability of University information and personal data in the use of publicly available GenAI tools

Version number	Purpose / Changes	Document status	Author of changes, role and school / unit	Date
0.1	Draft for consultation	Draft	Dean Drew, DCIO, IT Services, Chris	18 September 2023

			Milne, Head of Information Assurance and Governance	
1.0	First published version		Dean Drew, DCIO, IT Services, Chris Milne, Head of Information Assurance and Governance	5/Feb/2024

Introduction

Generative AI is a broad term used to describe any type of artificial intelligence (AI) that can be used to create new text, images, video, audio, or code. Large Language Models (LLMs) are part of this family of AI tools, which are used to produce text outputs.

ChatGPT and Google's Bard are two examples of publicly available, web-based versions of generative AI, that produce text outputs on a given subject from previously inputted/uploaded material. These tools are often used to summarise articles/reports, or to generate answers to a question, or to produce code, however as others have identified, this comes with several risks, which this Policy outlines along with the mitigations believed to be necessary at this time, for the safe use of such tools, so that individuals and the University are not harmed.

This Policy is informed by guidelines provided by the UK Government to the Civil Service (September 2023) (1) and from risk analysis undertaken by the National Cyber Security Centre (the "NCSC") (March 2023) (2).

Purpose and scope

This Policy sets out how University staff and students should use Generative AI tools, to protect personal data and other data and information, so to protect individuals and the University from harm, and to meet legislative obligations such as upholding the data protection principles, and the protection of intellectual property rights.

Intended audience

This Policy applies to all staff, contractors, and students.

Where this Policy applies

This Policy applies where any GenAI tools are used by staff, contractors and students when conducting their work.

Legislative and regulatory framework

Use of GenAI tools is subject to the same legislative and regulatory framework set out in the University ICT Regulations, which includes:

- UK and European data protection laws for the protection of personal data.
- UK and international laws for the protection of intellectual property.
- Scottish law that seeks to prevent individuals from abusive behaviour and sexual harm.
- UK legislation that creates a duty on universities to have due regard for preventing people from being drawn into terrorism.
- The JANET Acceptable use policy.

Relationship with existing University policies and regulations

This Policy should be read in conjunction with:

- Regulations governing the use of University information and communications technology (ICT) facilities (the ICT Regulations).
- The Information Classification Policy.
- Student harassment and bullying Policy
- Dignity at Work Policy

The University is developing separate policy and guidelines for the use of Generative AI in the teaching, learning, assessment and research.

Risks from the use of Generative AI tools

The NCSC provide a useful summary of how Generative AI tools work, notably how they turn inputs into outputs, noting -

> *"LLMs are undoubtedly impressive for their ability to generate a huge range of convincing content in multiple human and computer languages. However, they're not magic, they're not artificial general intelligence, and contain some serious flaws, including:*
>
> > *"they can get things wrong and 'hallucinate' incorrect facts*
> >
> > *"they can be biased, are often gullible (in responding to leading questions, for example)*
> >
> > *"they require huge compute resources and vast data to train from scratch*
> >
> > *"they can be coaxed into creating toxic content and are prone to 'injection attacks.'"*

The following risks are noted –

- LLMs typically retain the information that is passed to them for machine learning. Therefore, where this happens, significant stores of information will build up. Significantly –

  > "*This could mean that the LLM provider (or its partners/contractors) are able to read queries, and may incorporate them in some way into future versions*.
  >
  > "*A question might be sensitive because of data included in the query, or because who is asking the question (and when). Examples of the latter might be if a CEO is discovered to have asked 'how best to lay off an employee?', or somebody asking revealing health or relationship questions*.
  >
  > "*Another risk, which increases as more organisations produce LLMs, is that queries stored online may be hacked, leaked, or more likely accidentally made publicly accessible. This could include potentially user-identifiable information* [i.e., personal data]
  >
  > "*A further risk is that the operator of the LLM is later acquired by an organisation with a different approach to privacy than was true when data was entered by users*."

Mitigations

1. Information that is classified as **internal**, **confidential**, or **strictly confidential** by the University Information Classification Policy must not be disclosed/input into publicly available Generative AI tools. As,

   a. In most instances those data will contain personal data. By placing such data into the hands of a third-party, the University will lose control, leaving it open for that party to make secondary use of those data (for future machine learning). Such use will be unfair and by extension unlawful, meaning the first data protection is breached and

by extension the University will have not upheld its obligations as prescribed by the UK General Data Protection Regulation.

b. The third data protection principle requires that personal data are accurate, and the accuracy is maintained. Many commentators have identified that Generative AI tools are presently prone to 'hallucination' – getting things wrong and biases. Thus, where materials linked to living individuals are passed to such tools, there is potential for inaccurate personal data to be created about those persons.

2. Individuals should not put information into publicly available Generative AI tools that, if compromised or lost, could have damaging consequences for other individuals, groups of individuals, or the University more generally.

    a. This could also include copyright protected materials, where the volume of material used in Generative AI tools exceeds the provisions for fair dealing, available to educational institutions and students, thus breaching copyright, unless a licence is secured to use protected materials for that purpose.

    b. Additionally, individuals should be aware that where materials returned from Generative AI are subject to Copyright law, that use should only be made within the obligations of the said legislation.

3. The UK Civil service guidelines for the use of Generative AI advocate the use of a 'three Hows' framework. This is recommended, and reads –

    a. *"**How your question will be used by the system**. These systems learn based on the information you enter. Just as you would not share work documents on social media sites, do not input such material into generative AI tools.*

    b. *"**How answers from generative AI can mislead**. These tools can produce credible looking output. They can also offer different responses to the same question if it is posed more than once, and they may derive their answers from sources you would not trust in other contexts. Therefore, be aware of the potential for misinformation from these systems. Always apply the high standards of rigour you would to anything you produce, and reference where you have sourced output from one of these tools.*

    c. *"**How generative AI operates**. A generative AI tool, such as a LLM, will answer your question by probabilistically choosing words from a series of options it classifies as plausible. These tools cannot understand context or bias. Always treat with caution the outputs these tools produce and challenge the outputs using your own judgement and knowledge*."

Use of 'approved' Generative AI tools

IT Services with the Information Assurance and Governance team will create and maintain a catalogue of University-approved Generative AI tools, outlining what levels of data and information can safely be used within those applications, and for what purposes. Their use should be made in-line with this Policy and the University ICT Regulations.

The selection and procurement of Generative AI tools should follow existing University policies and procedures for the procurement of goods and services.

Before personal data which the University is responsible for is used in approved Generative AI tools, it may be necessary to first contact a Data Protection Impact Assessment ("DPIA"). Please liaise with the University Data Protection Officer (email dataprot@st-andrews.ac.uk) to understand if use of personal data is conditional on a DPIA.

Reporting breaches

The University is obligated to protect personal data and to keep such secure. If personal data have been passed to a publicly available Generative AI tool, then email dataprot@st-andrews.ac.uk, with the details, as such may be a personal data breach. Where a personal data breach has occurred, the University is under a legal obligation to log such incidents and where appropriate to report those to the UK Information Commissioner. Reporting the suspected or actual loss of personal data timeously is of significant importance; this can aid the University to promptly recover and/or contain a loss, thereby preventing or minimising harm to others.

Policy review

Generative AI is a rapidly evolving technology and the University will closely monitor developments, including recommendations from regulators and practitioners. This Policy will be reviewed biannually to ensure it remains relevant and appropriate, however an earlier review may be triggered as a result of:

- any significant change to relevant legislation, to university policy or to procedures primarily concerned with data confidentiality, integrity and availability.
- any incident relevant to this policy.

Contacts and further information

Any questions on the use of such technologies should be directed to the University Chief Information Officer (cio@st-andrews.ac.uk) in the first instance, with questions on the use of personal data directed to the Head of Information Assurance and Governance (dataprot@st-andrews.ac.uk).

Questions on the use of this technology in teaching, learning and assessment should be directed to the Vice Principal Education (Proctor) (proctor@st-andrews.ac.uk).

References

(1) Cabinet Office, (September 2023), *Guidance to civil servants on use of generative AI*. Available online: Guidance to civil servants on use of generative AI - GOV.UK (www.gov.uk), accessed 24 October 2023.

(2) National Cyber Security Centre, (March 2023), *ChatGPT and large language models: what's the risk?* Available online: Guidance to civil servants on use of generative AI - GOV.UK (www.gov.uk), accessed 24 October 2023.

Version number	Purpose / Changes	Document status	Author of changes, role and school / unit	Date
0.1	Draft for consultation	Draft	Dean Drew, DCIO, IT Services, Chris	18 September 2023

			Milne, Head of Information Assurance and Governance	
1.0	First published version		Dean Drew, DCIO, IT Services, Chris Milne, Head of Information Assurance and Governance	5/Feb/2024