ST MARY'S UNIVERSTITY, TWICKENHAM GUIDANCE IN RELATION TO THE SAFE USE OF GENERATIVE AI

BACKGROUND

The purpose of this guidance is to **encourage the informed and responsible use** of Generative AI to ensure that we maximize the benefits whilst minimizing any risks of concerns.

The use of Generative AI is transforming the way individuals are working. Informed and responsible use of Generative AI has the potential to increase efficiency in the workplace, improve decision making and foster innovation and so is encouraged by the University. However, with these benefits come potential risks, including data protection breaches, copyright issues, the protection of confidential information, ethical considerations, and compliance with wider legal obligations.

Whether users should use Generative AI in the performance of their job for the University is dependent on the specific use case which Users are using Generative AI. The issues raised with the use of Generative AI to assist with menial tasks (e.g. what is the traffic like today on the A3) are much different than skilled tasks (e.g. drafting outward facing University publications). Accordingly, Users should exercise good and sound judgment consistent with these guidelines.

SCOPE

These guidelines apply to all employees of St Mary's University, Twickenham (the "University") and those performing work and/or services for the University.

These guidelines cover the use of Generative AI for University purposes. They do not apply to the use of Generative AI purely for personal reasons.

Separate policies exist for the use of Generative AI in Teaching and by Students. These policies can be found at Annex A and Annex B.

TRANSPARENCY, ETHICAL AND RESPONSIBLE USE OF AI PRINCIPLES

Transparent, ethical and responsible use of AI underpins this guidance and is essential to the safe and effective use of AI.

Generative AI should not replace or override the decisions or insights of our human experts. It is a supplementary tool, not a primary decision maker.

- Generative AI can produce an output that is inaccurate, incorrect, misleading, biased, inappropriate or otherwise offensive or which violates copyright or other legal requirements. This means that critical thought must be applied to all outputs.

- All work product created using Generative AI must be reviewed by the User for accuracy and be consistent with these guidelines.

The use of Generative AI should be transparent.

- Information about the use of Generative AI should be disclosed by Users. Such disclosure would include the nature of the Generative AI system being used, the purpose for which it is being used, and any potential impact on individuals' data and privacy.

- For example, if creating a document, data, and/or information using Generative AI, its use should be disclosed and made clear on the document.

- The following is illustrative language that can be used for disclosure purposes:

  "The author generated this text in part with [what Generative AI is used]. Upon generating draft language, the author reviewed, edited and revised the language and takes responsibility for the content of this publication".

  Always use Generative AI ethically and responsibly.

- Users must not generate content to impersonate, bully or harass another person, or to generate explicit or offensive content. In addition to transparency, consider the following ethical principles:

  - *Fairness:* Generative AI should not be used to create unfair or inequitable conditions.

  - *Accountability:* Users who deploy Generative AI will be responsible for acting with professionalism and be accountable for any intentional and potential harm or misuse.

  - *Respect for Privacy:* The use of Generative AI should respect the privacy of Users and should not be used to collect, store, or access their personal data without their consent.

  - *Inclusiveness:* Generative AI systems and their output should empower everyone, serving to help combat the possibility for discrimination.

GUIDELINES

1. Definitions.

Terminology used in relation to Generative AI can be confusing. We set out below some common terms used when describing AI and what they mean together with definitions used in these guidelines.

1.1 "Generative AI" this refers to a type of artificial intelligence which can be used to create new content (for example, text, code, images, videos or music) (referred to as the *output*). The AI uses machine learning algorithms to analyse large data sets.
Some specific current examples of such GenAI systems include ChatGPT, Bing AI, Bard, Jasper and Synthesia.

1.2 "Closed Generative AI this means an AI tool purchased by the University. This closed platform can then be configured to protect sensitive data from being exposed to people outside of its intended use.

1.3 "Open Generative AI" means a publicly available (often free or low cost) AI tool such as ChatGPT. The Prompts and data entered into the Generative AI application can often then be used by the platform and could be exposed to the general public.

1.4 “Large language models (LLMs)”. LLMs are a type of Generative AI that can generate human like text in response to a prompt. They use deep learning techniques and massive data volumes to generate a response.

1.5 “IT” means information technology.

1.6 “Network” means a group of computer systems and other computing hardware devices that are connected through communication channels such as the internet, to facilitate communication and resource sharing.

1.7 “System” means all IT equipment, including personal and University owned, connected to the Network, or accessing Applications. This includes, but is not limited to, desktop computers, laptops, smartphones, tablets, printers, data and voice Networks, networked devices, software, electronically stored data, portable data storage devices, third party networking systems, and telephone handsets.

1.8 “Application” means a software program that runs on a system.

1.9 “Student” means a person enrolled at St Mary’s University.

1.10 “Users” means persons who have access to the any System. This includes employees, contractors, contingent workers, agents, consultants, vendors, service providers, suppliers, and other third parties. For the avoidance of doubt it does not apply to Students for which the following policy appended at Annex A applies.

1.11 “GPT” this is short for "generative pre-trained transformer", which is a type of LLM that uses deep learning to produce natural language texts based on information requested in the input. ChatGPT is an example of a GPT model which can be used to generate text.

1.12 “Hallucination”. LLMs can produce outputs which may initially appear to be believable but are in fact highly inaccurate or fabricated. This is known as a hallucination**.**

1.13 “Prompts” these are the inputs or queries that a User provides to the Generative AI application to receive the required output. Prompts can be used by the Generative AI application to further train the LLM.

1.14 “Personal Data” means all information relating to an identified or identifiable individual.

2. Processing Activities

2.1 The Use of **Closed Generative AI** through an AI tool purchased by the University is approved for internal University purposes. The Use of **Closed Generative AI** for internal University purposes carries less risk but the principles of transparency, ethical and responsible use outlined above in this guidance should still be adhered to. If a User has any concerns about their use of Closed Generative AI they should contact the Chief Information Officer kevin.braim@stmarys.ac.uk, in the first instance.

2.1 Use of **Open Generative AI** in support of University activities can be higher risk than the use of Closed Generative AI, as the data provided (the Prompts) to the LLM can often then be publicly available to other Users which could have legal and reputational consequences for the University. Accordingly use of Open Generative AI in support of University activities is divided into the following three categories: (1) Prohibited, (2) High-Risk, (3) Low Risk.

2.2 **Prohibited Processing Activities** Users should not engage in the following activities with **Open Generative AI**.

a) *No Non-Publicly Available Personal Data*. Users must not enter non-publicly available Personal Data into Generative AI (e.g. National Insurance Number, medical records, financial information). For the avoidance of doubt Personal Data is any information that can be used to confirm a person's identity.

b) *No Student Information.* Users must not enter student information into Generative AI.

c) *No University Confidential Information*. Users must not enter University confidential information into Generative AI. This includes items such as meeting notes, proprietary information, financial records or analysis, images, audio, video and non-public data and information. This would also include, confidential, sensitive or proprietary employer or third-party customer, supplier or employee related data.

d) *Human Resources*. Users must not enter the following HR information into Generative AI: hiring (including job posting), promotion, discipline or termination of employees.

e) *Legal and Compliance*. Drafting legal documents (including contracts), compliance reports or other legal or regulatory activities with potential legal implications.

2.3 **High Risk Processing Activities**. The following activities are considered high risk. Accordingly, the **use of Open Generative AI** for any of these activities must have human intervention, review and/or approval before being used or relied upon by the User. In addition, the User should carefully consider the factors outlined in 2.4(a) to (g) below.

a) *Public Documents.* Preparation of publicly-facing University statements, press releases, advertisements, promotions or similar written material.

b) *Decision Making.* Generating insights or recommendations that directly influence crucial decisions, such as strategic planning, financial investment or operational changes.

c) *Student Enquiries.* Automatically generating responses to student enquires, prompts or complaints.

d) *Predictive Modelling.* Predicting or anticipating future events, trends, or behaviours based on data analysis.

e) *Product Development.* Creating, enhancing, or diversifying products or services by generating concepts, designs and/or solutions.

2.4 **Factors for consideration when undertaking High Risk Processing Activities**

a) *Data Privacy and Security.* The use of Generative AI must comply with all privacy, data protection and University IT and Communications Systems policies.

b) *Bias and Discrimination.* The use of Generative AI must nor result in bias and/or discrimination against any student, employee and/or other individual.

c) *Plagiarism.* The use of Generative AI must not result in plagiarism. For further, guidance please refer to the Universities policies on Academic Integrity and ensure compliance with the Advertising Standards Authority and the Competition and Markets Authority guidance.

d) *Copyright Infringement.* The use of Generative AI must not result in copyright infringement. Be aware of any intellectual property rights owned by third parties such as copyright, database rights or trademark rights. Abide by any relevant licensing conditions regarding intellectual property rights in the AI Application's terms of use and ensure that third party proprietary data or material is not entered into the Application or Prompt without the third party's permission.

e) *Misinformation.* Generative AI can produce inaccurate or misleading information. The use of Generative AI must not result in the University producing a public document that contains incorrect, inaccurate or misleading information.

f) *Confidentiality.* Search queries entered into Generative AI are capable of being reverse engineered. The use of Generative AI must not result in the University breaching a duty of confidentiality.

g) *Reputational damage:* The use of Generative AI must not result in reputational damage to the University.

2.5. The exercise by the User of sound judgement and adherence to the guidance contained at 2.4 (a) to (g) is paramount in deciding whether or not to proceed with a High risk processing activity.

2.6 **Low Risk Processing Activities.** The following activities are low risk activities for which no described review is required.

a) The use of Generative AI for personal use that does not involve any of the activities or risks in Sections 2.2 to 2.3 above.

b) Personal risk is considered a Low Risk Processing Activity for the University. This risk analysis does not take into account any possible risks to the individual using Generative AI for personal purposes. Activities and the types of inputs for Personal Use are subject to an individual's own discretion.

c) The use of Generative AI for instructional purposes that does not involve any of the activities or risks described in Sections 2.2 to 2.3 above does not violate existing University policies on confidentiality, data protection, academic integrity or any other policy.

d) For the purposes of Clause 2.6 Low Risk Processing Activities would include but are not limited to:

   i) *Productivity Enhancements:* Generative AI can be utilised for general-purpose tasks like drafting emails, reminders, setting up meetings, summarising text or other administrative tasks, provided such tasks would not result in the disclosure of sensitive or confidential information belonging to the University.

   ii) *Search Replacement:* Generative AI can be used for general how-to questions that you would previously have used a search engine to answer, such as: How many states are there in America?

   iii) *Brainstorming and Idea Generation:* Generative AI can be used as a brainstorming tool to gather a wide range of ideas, which should be further reviewed and refined by human experts. For initial research into a topic, idea or to gain an overview for example "what are the main ethical concerns for students when using Generative AI tools? Or identifying/summarising core concepts or viewpoints in a particular disciplinary area for example "what were the prevalent influences on 18th century painters?"

iv) *Code Development / Review / CoPilot*: Generative AI is excellent for code review, writing basic functions, generating test cases etc. However, be careful not to include any proprietary algorithms or datasets.

v) *Drafting internal memoranda and presentations:* Subject always to these documents remaining internal to the University.

3. Exceptions

3.1 Exceptions to these Guidelines may be considered in very limited circumstances when potential risk and harm to the University and can be mitigated and should be directed to the Chief Information Officer kevin.braim@stmarys.ac.uk in advance.

4. Training, Awareness and Technical Support

4.1 For technical support in accessing and using Generative AI, please contact the Chief Information Officer kevin.braim@stmarys.ac.uk.

4.2 Any use of Generative AI by a User is subject to the User selecting the opt-out option before first use. This will prevent the data the User enters into the prompt being used by the LLM to train itself. If the opt out selection is unclear, or is not available on the Generative AI application, please contact the Chief Information Officer kevin.braim@stmarys.ac.uk for further clarification.

4.3 Be aware that third parties may build a service on top of generative AI applications. Avoid inputting any information or data into these add-ons. If in doubt about how add-ons may be operating in relation to a Generative AI application you are using, speak with kevin.braim@stmarys.ac.uk in the first instance.

4.4 Users must apply the same security measures we apply to all our IT applications and comply at all times with our IT and Communications Systems policy. This includes using strong passwords, updating applications as required and not installing software from external sources without authorisation from the Chief Information Officer.

4.5 If you have any concerns that you may not have adhered to these guidelines this should be immediately reported by email to the Chief Information Officer kevin.braim@stmarys.ac.uk.

5. Monitoring and Review

5.1 These guidelines have been approved by Senior Leadership Team and will be implemented by the schools and professional services departments, working closely with support services.

52 This guidance will be reviewed at least annually, or more frequently, if necessary, to ensure it remains relevant and up to date with technological advancements and legislative changes. Recommendations for updates or improvements to the guidance may be submitted to kevin.braim@stmarys.ac.uk or legal@stmarys.ac.uk.

6. Related Policies and Guidelines

6.1 These guidelines form part of the information security management system (ISMS) at the University. These guidelines should be read in conjunction with all other University information management policies, which are reviewed and updated as necessary to maintain an effective Information Security Management System to meet the University's business needs and legal obligations.

6.2 These guidelines supplement and should be read in conjunction with our other policies and produces in force from time to time including but without limitation our:

a) Policy on AI for teaching and learning and accompanying staff and student guidance;
b) Equality and Diversity Policy Statement and Code of Practice;
c) Employee Privacy Notice;
d) Data Protection Policy;
e) IT and Communications Systems Policy.

Version Control: Version 1.1 Created 18th July 2024

Approved by: kevin Braim 18th July 2024

Next review date: 18th July 2025

This guidance is available on Staffnet