Interim Guidance for Staff on the Responsible Use of Artificial Intelligence (AI) Large Language Models (LLMs) to Protect Personal and Corporate Information (e.g. Bing Chat Enterprise[1], ChatGPT, GPT-4, Bard, etc.)

Background

Generative and publicly available AI models are creating excellent opportunities to enhance education and research, and opportunities to improve service delivery, effectiveness and efficiency. These also bring challenges in terms of ethics, bias, academic integrity, privacy and the protection of corporate information.

The following is presented as initial guidance to help ensure colleagues use this technology responsibly and mitigate the risks of disclosing personal data or business-sensitive information while using AI models, ensuring we preserve the integrity and confidentiality of our corporate information and meet our legal obligations under the UK GDPR and the Data Protection Act 2018.

Scope

This guidance applies to all employees and contractors who process University information. It will continue to be reviewed and updated as the University's overarching policy and strategy for the use of AI across its activities is developed.

Key Points

Do not input personal data into an AI model.

Do not input the University's corporate information (i.e. proprietary or confidential information which is not publicly available) into an AI model.

Do not use AI models to make decisions about individuals. Under the UK GDPR individuals have the right not to be subject to a decision based solely on automatic processing.

Recommended Practices

Interacting with an AI Large Language Model typically requires the submission of text, which may be a question or statement, which is then used by the model to generate text or other media output. Data in prompts submitted to models for processing may be stored and used for other purposes, including the training of AI models, by the organisations hosting the LLMs. It is therefore vital to craft prompts carefully by providing sufficient context to enable the model to generate a helpful response without disclosing personal or corporate information.

Here are some guidelines to bear in mind while constructing prompts:

Anonymise information: Ensure that data within prompts is anonymised. In other words, do not use any direct identifiers or details which could indirectly identify individuals within prompts. An indirect identifier is information which could identify an individual if combined with other

[1] Microsoft has implemented privacy and security features that make Bing Chat Enterprise a relatively lower risk option than some other web-based services (e.g. there is no chat history, and neither prompts nor responses will be used to train models).

information reasonably available to the processor of the data, including information obtainable from other sources (e.g. published online, or from another organisation).

Generalise prompts: Frame prompts in a general way, where possible, without specifying particular details. If, for example, you are looking for inspiration relating to a strategy for "Edinburgh Napier University", you could ask for strategies which might be useful for a "UK Higher Education Institution". As with hypothetical scenarios, this reduces the risk of disclosing identifiable or otherwise sensitive information.

Keep prompts hypothetical: Rather than submitting genuine scenarios, where possible use hypothetical scenarios with the same or a similar context. For example, if as a lecturer you were using an LLM to assist with writing a business proposal to develop a specialised new micro credential course, you could avoid mentioning the name of the university, the title of the proposed course, and the target market. Additionally, instead of submitting something like "*I am developing a course along the following lines…*", consider how the issues you are interested in can be explored more obliquely. For example, you could try something like "*If a course with features X were to be developed at a UK University, targeting a market with features Y…*", taking care not to over-specify X or Y. Drafting your question separately, before pasting into a chat interface, may help, and you may also find it helpful to consult with colleagues before submitting the prompt. With a careful approach like this you may receive excellent LLM output, applicable to the real-world scenario in question, without divulging potentially identifying or otherwise sensitive information.

Contextualise with care: The conversational nature of services like ChatGPT and Google Bard means that contextual information held by hosting organisations, and thus potentially available for other uses like training AI models, may "snowball" over the course of one or more chats. This can increase the risk of disclosing indirect identifiers, even if direct identifiers are avoided, and can also increase the risk of disclosing sensitive corporate information. Keep track of what has been provided to an AI model, taking this into account before submitting new prompts.

Limit prompt detail: As a general rule, limit the detail within prompts submitted to an LLM. As with contextual information, please take care not to over-specify what is requested from the model as this may inadvertently disclose personal or otherwise sensitive data.

Do not use the technology to make decisions about individuals: Article 22 of the UK GDPR states that a "*data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".* Colleagues should also avoid using the technology to *contribute* to decisions about individuals unless a Privacy Impact Assessment has been completed and there is a robust process in place approved by the appropriate Dean/Director.

Verify outputs: However plausible any LLM outputs are, always question their accuracy and truthfulness. Check this output independently and ensure that you are able to defend the veracity of any information relied on with reference to more authoritative sources.

Information Governance Team

October 2023