



Generative AI software and implications for learning, teaching and assessment

April 2023 update for staff

Professor Martin Hendry and Professor Moira Fischbacher-Smith

1. Introduction

The purpose of this memo is to give an update on our UofG response to the (rapidly evolving) impact of generative AI software tools¹ on our policies and practices around learning, teaching and (particularly) assessment. It follows an initial [email circulated to staff in mid-February 2023](#), outlining the general approach proposed for the remainder of session 2022-23 and thereafter, and an [open LTC meeting held in early March 2023](#) which began to explore some of the issues in more detail.

The key message of this update is “**don’t panic**”. In our mid-February email, we sought to reassure staff that we were not requiring Schools to make any modifications to their remaining 2022-23 exams and assessments. The core reasons for eschewing such a knee-jerk reaction are still true, and we remain keen to avoid disproportionate responses which (given the fast-changing nature of this AI field) may yet prove to be of limited, and only temporary, effect.

We also provide below some more specific information and guidance relevant to ChatGPT for the Spring 2023 diet and we give a brief update about our plans for session 2023-24 outlining what action we would encourage and discourage with respect to the use and detection of AI, particularly with respect to the student code of conduct.

2. ChatGPT and Turnitin

You may have read about recent developments involving the similarity checking software Turnitin. In January 2023 Turnitin indicated their intention to develop AI detection capabilities, and in March 2023 they unilaterally announced a plan to release their AI detection functionality on 4 April 2023, in advance of Universities’ Spring exam diet, but without full testing and with the AI detection report being automatically generated by the Turnitin software. This news was met with considerable unease across the sector, due to concerns about the rate of false positives (i.e. content which is *not* AI-generated but which is wrongly suspected to be as such) that Turnitin would produce and Universities being given no time to plan implementation and create appropriate underpinning policies.

After lobbying by a number of expert groups from across UK universities, on 29 March 2023 Turnitin changed their approach and made it possible for institutions to opt out of the AI detection functionality. Along with most other Russell Group universities, UofG has decided to take up this opt out, at least for the moment. We will, however, follow developments closely, including in-house testing of Turnitin’s AI detection software, in the coming months. Should it prove sufficiently robust and reliable, we believe that Turnitin may be a useful component of our approach to dealing with AI-generated content in future assessments and exams – but not for the Spring 2023 diet.

3. ChatGPT and the Spring 2023 exam diet

In our mid-February email, we encouraged staff to take note of the emerging guidance on how to minimise the impact of ChatGPT in assessment. Where exam and assessment questions had not yet already been finalised, we invited staff to consider making small changes to them – e.g. by shifting their focus towards higher-level reasoning and synthesis, and/or more current data or scenarios – to

¹ ChatGPT has become recognised as the archetype of such tools. We will refer to ChatGPT throughout this memo as a representative example of the wider range of generative AI software that is now available.

reduce their vulnerability to ChatGPT.

To assist with this, in February 2023 we began curating general guidance and resources on a [dedicated SharePoint page](#) and we have set up a [dedicated email address](#) for sharing links. We trust that colleagues have found these resources useful, although we recognise that the impact of ChatGPT will differ a great deal according to discipline.

As attention now turns to marking and moderating Spring 2023 exams and assessments, however, the key problem we face concerns whether it is possible to detect the presence of AI-generated text in students' submitted work. **Unfortunately, there is currently no “magic bullet” to solve this problem and to date we have instead focussed efforts on deterrence and awareness raising.** In a series of emails to students (beginning in early February and augmented by multiple further messages planned for April) we are emphasising that the “inappropriate collaboration with others” referred to in the University's [plagiarism policy](#) includes inappropriate interaction with any website or software that generates assessment responses, such as ChatGPT.

Are there *any* means of detecting AI-generated content that are currently available? As noted above, Turnitin is not (yet) reliable enough to do this systematically. Several other software tools and websites for detecting such content exist, but all are imperfect and subject to *false negatives* (i.e. AI-generated content that is *not* identified as such) as well as false positives. Examples of these software tools include:

- GPTZeroX: <http://gptzero.me/>
- DetectGPT: <https://detectgpt.ericmitchell.ai/>
- Writer AI: <https://writer.com/ai-content-detector/>
- Content at Scale: <https://contentatscale.ai/ai-content-detector/>

Colleagues may wish to experiment with these (and other) websites, using text generated by ChatGPT. However, consistent with our decision to opt out of Turnitin's AI detection capabilities, **we explicitly ask that colleagues do *not* upload student exam or assessment responses to any of these software tools – all of which we regard as currently too unreliable to indicate whether the responses have been AI-generated.**

As we seek to better understand the capabilities and limitations of ChatGPT, however, the principles that underpin some of the above AI-detection software tools may still be instructive. In the words of statistician George Box (and as quoted [here](#)): “All models are wrong, but some are useful”.

For example, the GPTZeroX software, developed by Edward Tian (Princeton University), analyses text in terms of its “burstiness” – which in this context is (loosely speaking) a measure of the tendency for long, complex sentences to be followed by shorter, simpler ones. Humans tend to write “burstier” text than is currently produced by software like ChatGPT. There are some further general trends and broad characteristics of AI-generated text that may also be instructive:

1. The level of sophistication of AI-generated text may be poorly matched to that normally expected for students on a given course or at a given level of study. Such text may appear too complex, or indeed too basic, compared with what students would typically produce.
2. AI-generated text may include examples of technical or specialised vocabulary or concepts that lie beyond the scope of the course syllabus or are not normally used by students. Similarly, AI-generated content may fail to reference subject-specific material that was covered in the course or may include spurious or fictitious references in its bibliography.
3. AI-generated text may lack coherence and fail to address the question asked in a meaningful way or may contain factual errors or inconsistencies that can be easily identified.
4. AI-generated text may contain repetitive phrases or sentences that are used to pad out the answer or may contain syntactic errors or convoluted sentences that are not typical of human writing.

5. The writing style of AI-generated text may also be atypical of that produced by students, and/or may be inconsistent *within* a piece of work. There might also be unusual repetition of phrases or ideas, or essentially identical responses to multiple similar questions.

Of course, the challenge with at least some of these suggested characteristics is that they may frequently be displayed by students too!

4. Following up on suspected AI-generated content?

Notwithstanding the above general trends, when considering an individual student's response to an exam or assessment question there may simply be no obvious "tell" for AI-generated text – particularly for e.g. essays where students' answers will be highly individualised. **Consequently, and given the limitations of existing detection tools already outlined above, we believe there is not yet available to us any means to detect AI-generated text that is reliable enough to form the basis of a student misconduct case.**

As things stand, therefore, suspected misuse of ChatGPT to generate an exam or assessment answer cannot be treated in the same way as a conventional plagiarism case. The latter would normally involve a reference document (e.g. an external source, textbook or webpage) against which the student's answer is compared, but in the case of suspected ChatGPT use no such reference document would exist. **Consequently, we believe that it is not appropriate at this time to pursue a misconduct case based solely on an unsubstantiated suspicion that a student's response is AI-generated, even where a detection tool might appear to confirm that suspicion.**

The two clear exceptions to this would be:

- where two or more students have used ChatGPT to answer a specific (typically short) question and their responses are sufficiently similar that they raise conventional suspicions of plagiarism.
- where a student's submitted work contains a number of fictitious references; this, similarly, would represent clear, *prima facie* evidence, worthy of further investigation, that the submitted work was not the student's own.

Should colleagues encounter either of the above examples - implausibly high similarity between students' answers or a fictitious list of references in a student's submitted work - the key guidance we wish to emphasise is to **follow the system**: i.e. make full and proper use of our existing misconduct procedures to investigate these cases. Do **not** introduce any new, informal or ad hoc approaches to investigate suspected use of ChatGPT.

Further details of the procedures to be followed when investigating cases of suspected plagiarism can be found within the University's [2022-23 plagiarism statement](#). Note that undergraduate non- Honours cases that are first offences may be investigated by the Head of School or their local designate, while all other cases must be investigated by the Senate Assessors for student conduct.

5. Looking ahead to session 2023-24

While the principal focus of this memo is to discuss implications for the Spring 2023 exam diet, we wanted also to briefly update colleagues on the longer-term direction of travel. As noted in our February email and discussed in more detail at the March Open LTC meeting, our core belief is that Generative AI tools are now firmly part of the landscape. We must continue to hold a firm line against their misuse by students in their assessments. At the same time, however, we should explore creative and constructive ways in which these tools can be incorporated responsibly into our learning and teaching and, where appropriate, in our assessment.

To this end, the dedicated [SharePoint page](#) continues to be updated regularly with new content, much of it focussed on innovative ways in which AI tools can be incorporated into future learning and

teaching – and on how assessment design can be made less vulnerable to misuse of such tools. A short-life working group has also been established, with membership drawn in part from the Assessment & Feedback strategy workstream, to create guidance for staff on the impact of the rapidly evolving capabilities of generative AI software for learning, teaching and assessment design in Session 2023-24 and beyond. The guidance produced by this working group is expected to take the form of a concise document that will contain practical advice and examples for use in designing courses and assessments. It is hoped that a preliminary version of this guidance will be available to staff by mid-June 2023 and further updates on this will be provided in due course. Moreover, Student Learning Development (SLD) and Research & Innovation (R&I) are also working to produce a range of guidance for taught students and research students, alongside the materials being developed for staff, and these will be available very soon. This guidance will include short lists of ‘dos and don’ts with AI’ but will also highlight a wider framework of guidance and information for students who are researching and writing using a variety of digital tools.

6. Concluding remarks

The aim of this short memo was to provide colleagues with some guidance on the potential impact of ChatGPT for Spring 2023 exams and coursework assessments. As stated previously, we have sought to adopt a measured, cautious and realistic response to the short-term challenges of ChatGPT – and above all to avoid a knee jerk reaction that would place an unreasonable burden on our teaching staff. Our proposed approach will likely not be perfect. What is essential, however, is that we work together to address as best we can any issues of academic integrity that arise – both in the short term, as we deal with the spring 2023 diet, and in the longer term as we seek to develop more robust approaches to course and assessment design that are better able to meet the challenges and opportunities that generative AI software present.

Please do contact us (directly, or via our [dedicated email address](#)) to share your questions, concerns, reflections and experiences about ChatGPT and the “Brave New World” it is opening for learning, teaching and assessment.

Martin Hendry and Moira Fischbacher-Smith
April 2023