Universiteit Leiden

To	Examiners Faculty of Humanities
From	Vice-deans Faculty of Humanities
C.c.	Boards of Examiners Faculty of Humanities
Date	12-06-2023
Phone number	2094 (Anne Zwetsloot)
Topic	Fraud due to chatbot use (including ChatGPT)

Introduction

This memo aims to explain the application of existing regulations regarding the use of chatbots (including ChatGPT). Currently, many examiners are concerned about possible fraud using chatbots, especially now, during the exam and thesis period. This memo explains how to recognize chatbot fraud and what to do when fraud is suspected.

ChatGPT was launched at the end of November 2022, accelerating the discussion about chatbots and AI in education. ChatGPT was not created in a vacuum but parallel to various other AI tools, such as Deepl for translation and Quillbot for paraphrasing. ChatGPT is currently attracting the most attention because it is free, it can be used by anyone via a browser, and it has received much attention in the media. It is likely that more free AIs that can perform similar tasks will become available in time.

ChatGPT is an AI-based chatbot. It generates text in response to a 'prompt' (question or command from a user) based on a vast dataset of text material (including from the internet). The user can specify the form and length of the text. The program generates these texts by making a statistical prediction of what the most likely next word in the text would be. That prediction is based on previously used phrases or sentences. Chatbots do not understand what they write: the quality of the answer is directly related to the quality of the question (the prompt).

Chatbots like ChatGPT have all kinds of consequences for education, both positive and negative. More information about chatbots in education can be found on the websites mentioned at the end of this memo.

1. When is it fraud?

When a student uses AI (to a large extent) for a writing assignment, and the examiner cannot distinguish between the parts written by the student or the AI, this impedes the judgment of knowledge and skills, which is fraud.
Based on Article 7.12b of the Higher Education Act (WHW), fraud is understood to mean any act or omission on the part of a student aimed at making it wholly or partially impossible to form a correct assessment of their knowledge, insight, and skills. This includes, among other things, having others make (parts of) a text.

2. How do you recognize the use of chatbots?

It can be challenging to determine whether chatbots have been used, but there are some clues:

- **Writing style**: If the writing style in a text differs (significantly) from the usual writing style of the student. Moreover, according to ChatGPT itself, it can be recognized by the following elements:
  - Repetitive or long-winded formulations
  - Illogical or inconsistent answers
  - Excessive use of specific words or phrases
  - Excessive use of adjectives and adjectival clauses
  - Lack of personal flair or individuality
  - Limited variation in sentence structure
  - Less use of informal or colloquial language
  - Unusual or unnatural choice of words
- **Content errors**: A chatbot is a language model that cannot think for itself. The chatbot can state untruths with a strong conviction. The software has, after all, been trained on vast amounts of data from the internet, where lots of incorrect information can be found.
- **Outdated data**: The current version of ChatGPT (January 2023) is only trained on data up to the end of September 2021. This means that the chatbot is unable to use more recent data. This restriction is likely short-lived, given the speed of developments. In addition, a student can enter current data in a prompt, which ChatGPT can then use. ChatGPT can also be forced to display current data: it will then make something up based on older texts.
- **Math errors**: A chatbot is a language model, not a calculator: the text is generated in response to mathematical questions, but the outcome can be a random number based on the human preference for numbers.
- **Incorrect source citation**: Chatbots use probability, based on numerous sources, to generate answers. They do so when compiling pieces of text, but also when producing a simple factual answer to a question. As a result, you can never exactly know what sources have been used for the output. This means that the given answers are not transparent and difficult to verify, making accurate and complete source citations impossible.
- **Non-existent sources**: If you ask ChatGPT to provide the sources used to generate an answer, the chatbot will often fabricate non-existent sources. When you ask for existing and/or peer-reviewed sources, you will generally get appropriate sources, however, these sources will not contain all the data on which the chatbot has based the answer. The use of non-existent sources can be seen in Turnitin's plagiarism checker: usually, sources are marked yellow because they are existing text, but Turnitin will consider fabricated sources to be original text and, therefore, not mark it.

3. Turnitin: not (yet) reliable and will be removed

The chatbot's expressions are unique; the texts cannot be found on the internet. Therefore, the standard plagiarism checkers do not work. Leiden University currently does not use software that can detect whether a chatbot has been used because the results are not (yet) reliable enough; this also applies to the tool in Turnitin. Turnitin's plagiarism checker is just as reliable as ever. This percentage can be used.

As of June 16, the AI Writing Detection Tool will be disabled in Turnitin. Since the beginning of April, Turnitin has integrated an AI Writing Detection tool into FeedbackStudio. While Turnitin itself claims that the chance is <1% that this tool marks a text as "written by AI" when it is not, we cannot verify or reproduce this claim. It is also impossible to determine exactly how the Turnitin tool works and what it bases its conclusions on. That is why it has been decided at a central level to turn off this functionality in Turnitin. The functionality will no longer be visible; this applies to both existing and future submissions. Furthermore, the reports can also no longer be requested.

4. What to do if you suspect chatbot use?

- **Report your suspicion:** Examiners are obligated to report (a suspicion of) fraud to the Board of Examiners.
    - Send an email to the Board of Examiners with the evidence in the attachment
    - The Board of Examiners needs: 1) screenshots of all fragments where you suspect AI use; 2) the Turnitin plagiarism report; 3) the original submission
    - Do not argue with the student, and do not share the screenshots. Inform the student that the case will be forwarded to the Board of Examiners and that they will answer all questions. It is possible to include a comment on Brightspace that the paper is not yet graded pending the decision of the Board of Examiners on whether it can be graded and make this comment visible to the student.
    - Give the Board of Examiners some guidance in your email: describe whether the assignment meets the requirements and what makes you think it might be AI-generated/supported.
    - Before submitting the case, verify whether the references exist. Please note: this may also include sources ascribed to relevant, existing and well-known authors of which no articles or studies exist under such a title. So they did not write the publication. Please indicate in your email whether you have checked the references and whether they are genuine.
- **Do not give a grade or assessment (yet)**: If fraud is suspected, an exam (assignment/report) will only be assessed and graded by the examiner when the Board of Examiners has determined whether or not fraud has occurred and whether or not to impose sanctions.

- **Do not impose sanctions**: The law stipulates that sanctions (such as declaring the assignment invalid or excluding a student from the examination) may only be imposed by the Board of Examiners.

5. What does the Examination Board do?

The Board of Examiners will treat cases of fraud using AI text generators the same as other cases of fraud and plagiarism. The Board of Examiners will determine whether or not fraud has been committed, possibly after a meeting with the student. In its decision, the Board of Examiners must clearly state, in addition to the imposed sanction, what the violation consisted of and what the violated regulation was. A clear reference must also be made to, for example, the Rules and Guidelines or OER. If the Board of Examiners concludes that the student has committed fraud, they will impose sanctions. These sanctions are in proportion to the nature and seriousness of the fraud committed.

6. Closing remarks

The use of AI in education remains a complex subject, and it is important to discuss the subject of academic integrity with students, also specifically in relation to AI. The university does not prohibit the use of chatbots, but it is included in, among other things, the manual for Boards of Examiners that the student's text may not have been written by someone else, and that includes a chatbot. Pretending that an AI-generated text is your original text is against the academic integrity guidelines.

We are currently hearing a lot of signals about an increase in fraud cases due to the use of ChatGPT. We understand that this is a considerable burden for examiners and Boards of Examiners. It is important that we inform students well about the use of AI in order to prevent this as much as possible, but that does not remove the extra burden at the moment. We are concerned about this, which is why we monitor developments within the faculty and university as well as developments in the field of AI as closely as possible.

Contact information for support and questions:

- ECOLe (ecole@hum.leidenuniv.nl)) for questions about Turnitin
- O&K (oenk@hum.leidenuniv.nl) for questions about this memo

More information:

Would you like more information about AI in education?

- https://www.staff.universiteitleiden.nl/education/it-and-education/ai-in-education?cf=humanities&cd=fgw-board-office
- The Faculty of Humanities ChatGPT reading suggestions: https://www.staff.universiteitleiden.nl/vr/humanities/fgw-learning/tips-and-tricks/chatgpt-reading-suggestions-and-sources?cf=humanities&cd=fgw-board-office
- https://teachingsupport.universiteitleiden.nl/getting-started/on-chatgpt
- https://www.leidenlearninginnovation.org/stories/the-possibilities-and-problems-of-ai-and-its-detection-in-education/

You can practice with ChatGPT via LlinC:

- https://www.leidenlearninginnovation.org/stories/chatgpt-exercise/

Or take the course on ChatGPT during the teaching fair on June 30, 2023:

- https://www.staff.universiteitleiden.nl/events/2023/06/teaching-fair-2023?cf=humanities&cd=fgw-board-office

On Machine Translation:

- Machine Translation Literacy (google.com)
- Machine Translation Literacy - Teaching Resources (google.com)

Colleagues at the Faculty of Social and Behavioural Sciences have written a blog about the AI Detection tool. Read it here:

- https://www.leidenanthropologyblog.nl/articles/when-all-positives-are-false-the-emergence-of-an-academic-whodunnit

More information about the Turnitin AI Writing Detection tool can be found here:

- https://universiteitleiden.screenstepslive.com/a/1666201-ai-writing-detection.

About AI/ChatGPT in general:

- https://lifearchitect.ai/