

AI Usage in Theses at IfIS

AI-supported tools for text generation and assistance in text creation in various forms have reached market maturity for some time now. Naturally, such tools can be a great help, especially when writing theses. On the other hand, the use of such tools can easily give the impression of deception, especially if the rules are unclear. According to § 11 (4) sentence 3 APO, "[t]he examiner shall announce at the start of an examination what aids are authorised and the handling of unauthorised aids"¹. This document does exactly that, i.e., **the following rules for the use of AI-supported tools when writing your Bachelor's, project or Master's thesis at IfIS are binding.**

We distinguish between **tools for writing support**, e.g. grammar checking and slight rephrasing (e.g. Grammarly), and **tools for text generation**, in particular language models such as ChatGPT. Tools for writing support are not subject to any further rules at IfIS, as the text creation is still carried out by the student. The following rules apply to text generation:

1. Generated texts **must** only be used as **direct quotations** (never as indirect quotations), stating the language model used, the prompt used for generation and the date of generation! The information must be placed as a footnote after the quotation.

Example: „Information retrieval (IR) is a branch of computer and information science that deals with finding and extracting relevant information from a large amount of unstructured data. Typically, it involves searching for documents or data sets that fulfill certain criteria, such as relevance to a search query. IR includes various processes such as indexing texts, performing search queries and evaluating the relevance of results. Methods from the fields of machine learning, natural language processing and statistical models are often used to improve the efficiency and accuracy of searches. Well-known applications include search engines such as Google, document management systems and digital libraries. The challenge lies in interpreting users' search queries precisely and finding suitable answers in the vast amounts of data, whereby the timeliness and relevance of the information provided are crucial.“²

2. All common rules for working with citations apply. Especially, changes in the generated text (i.e., changes, adding phrases and delete parts of the text) must be marked by [].

Example:

- *Shorten the text:* „Information retrieval (IR) is a branch of computer and information science that deals with finding and extracting relevant information from a large amount of unstructured data. [...] Well-known applications include search engines such as Google, document management systems and digital libraries. [...]“²
- *Insert new information:* „Information retrieval (IR) is a branch of computer and information science that deals with finding, [...] extracting, [and presenting] relevant information from a large amount of unstructured data. [...]“²

¹ Cited by the official English translation of the APO (<https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Stabsstellen/SPK/ordnungen-leitlinien-fakten/lehre/allgemeine-pruefungsordnung-en.pdf>)

² Generated by ChatGPT 4o (2024-06-18). Prompt: “Write a short summary concerning Information Retrieval.”

- *Changing the text:* „Information retrieval (IR) is a branch of computer and information science that deals with [discovering] and extracting relevant information from a large amount of unstructured data. [...]“²
3. The student is **responsible in all respects** for compliance with the **usual quality criteria**, in particular the **correctness** of generated passages. In other words: Even if the generated passages are marked as direct quotations, they do not have the quality of a direct quotation from, e.g., a peer-reviewed paper. Non-obvious statements must therefore be substantiated with additional sources. Or in other words: the statement "it says so because ChatGPT wrote it that way" is no excuse if an error was found in a passage during marking the thesis.

Examples:

- In the generated text from before, the statement "Well-known applications include search engines such as Google, document management systems and digital libraries." is generated. Typically, one would expect examples of specific IR systems here, e.g. in digital libraries.
- Google is rather a web search system which is a special case of classic IR and differs significantly from IR systems in digital libraries (e.g., library catalogs). This fact does not necessarily have to be explicitly emphasized, but a follow-up question can be asked in the defense of the thesis. In this case, the student must be aware of this, i.e., it is not advisable to use the language models on "autopilot".
- Another pitfall lies in the generation of references. A model often generates fictitious references, i.e., papers or books that appear convincing at first glance but are freely invented. Students themselves are responsible for checking **whether the cited sources exist and whether they also contain the relevant information.**

In summary, we are not critical of the use of AI-supported tools as an aid, as long as they are not used as a substitute for own considerations or critical reflection on the thesis.