

<b>Topic</b>	<b>Dataset Creation for Fine-Tuning Language Models to Evaluate Document Relevance in Response to User Queries</b>
<b>Overview</b>	<p>Large Language Models (LLMs) like GPT and BERT have transformed how we interact with AI by understanding and generating human-like text. Initially trained on extensive datasets to grasp a wide range of language nuances, LLMs often require fine-tuning to adapt to specific tasks or behaviors in targeted domains. This fine-tuning process involves training the model on a smaller, specialized dataset that closely represents the scenarios it will face, enhancing its ability to perform designated tasks accurately and interact in expected ways.</p> <p>Fine-tuning aligns the model's responses with user intents and desired outcomes, ensuring the model's interactions are both relevant and engaging. This process is critical in making LLMs effective and reliable in practical, real-world settings. The performance of a fully trained model is highly dependent on the quality of the used data.</p> <p>This thesis aims to create a robust dataset that will enable the fine-tuning of a large language model (LLM) to assess whether a document or text excerpt is suitable for answering specific user queries. The goal is to develop a methodology for dataset compilation and annotation that accurately reflects document relevance, enhancing the model's ability to facilitate effective information retrieval.</p> <p><b>Objectives:</b></p> <ul style="list-style-type: none"><li>• <b>Literature Review:</b> Conduct a comprehensive review of existing approaches in dataset creation for language model training, focusing on relevance determination in the context of question answering.</li><li>• <b>Dataset Design:</b> Develop criteria for selecting and annotating texts that determine their relevance to specific queries. This involves defining relevance metrics and establishing a scalable annotation process.</li><li>• <b>Model Fine-Tuning:</b> Utilize the created dataset to fine-tune an LLM in cooperation with the supervisor.</li><li>• <b>Evaluation:</b> Design and implement a series of experiments to evaluate the model's accuracy in identifying relevant</li></ul>

	documents. Compare the performance with existing models and baseline systems.
<b>Language</b>	German or English
<b>Exemplary literature</b>	<p>Abdullin, Y., Molla, D., Ofoghi, B., Yearwood, J., &amp; Li, Q. (2023). Synthetic Dialogue Dataset Generation using LLM Agents. In S. Gehrmann, A. Wang, J. Sedoc, E. Clark, K. Dhole, K. R. Chandu, E. Santus, &amp; H. Sedghamiz (Eds.), Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM) (pp. 181–191).</p> <p>Mukherjee, S., Mitra, A., Jawahar, G., Agarwal, S., Palangi, H., &amp; Awadallah, A. (2023). Orca: Progressive Learning from Complex Explanation Traces of GPT-4 (Version 1). arXiv.</p>
<b>Additional information</b>	<p>Type of thesis: Bachelor or master thesis</p> <p>Start: As soon as possible</p> <p>Requirements:</p> <ul style="list-style-type: none"> <li>- Programming skills, preferably in Python</li> </ul> <p>Publication: Excellent work can be submitted and published as a conference paper or at TU Prints</p>
<b>Contact</b>	<p>If you are interested, please contact Nils Schönfeld (<a href="mailto:nilslucas.schoenfeld@bwl.tu-darmstadt.de">nilslucas.schoenfeld@bwl.tu-darmstadt.de</a>)</p> <p>Prof. Dr. Dr. Ruth Stock-Homburg</p>