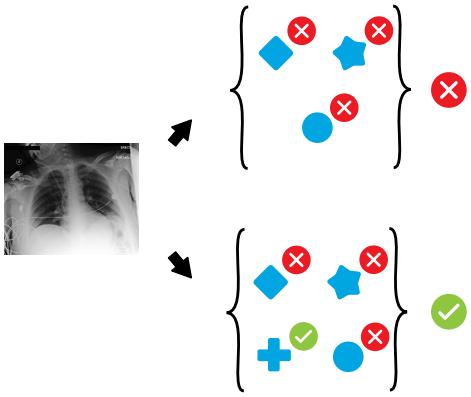
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS
Learning and Dynamical Systems Group (Dr. Michael Muehlebach)

Master Thesis Project

Risk Controlling Large Language Models for Radiology

Large language models (LLMs) hold significant promise in the medical domain. A prime opportunity lies in the generation of text reports based on x-ray images, which is highly time-consuming if performed by human radiologists. However, a remaining limitation of LLMs lies in the phenomenon of "confabulations" (or "hallucinations") – LLM responses that lack factual support, potentially leading to substantial harm if trusted.



A rapidly evolving field in machine learning, called *risk control,* aims to address this limitation. These methods calibrate machine learning model outputs such that rigorous probabilistic guarantees on their admissibility are provided. While the establishment of such methods constitutes a first step toward deployable systems, they have thus far only been calibrated and evaluated using automated admissibility assessments that rely on machine learning or n-gram based heuristics. In this, they fail to capture relevant nuances in textual LLM responses that can only be reliably assessed by human domain experts.

The goal of this project is to bridge the gap between methodology and application. Concretely, the thesis incorporates

1. Collaborating with medical professionals to collect human expert assessments of LLM-generated radiology reports.
2. Constructing a high-quality dataset that accurately reflects the complex criteria for "admissibility" in radiology text, setting a new benchmark for future research.
3. Applying existing methods to the newly curated dataset and compare their performance to automated assessments, highlighting strengths and limitations in real clinical contexts.

The project will be a collaboration between the Max Planck Institute for Intelligent Systems and the Universitätsklinikum Tübingen.

Prerequisites

1. Passion for the medical domain and real-world impact.
2. Familiarity with machine learning, probability theory and statistics.
3. Strong programming skills in Python and PyTorch. Prior experience with language models and Hugging Face libraries is a plus.

Contact

If you have any questions do not hesitate to contact us. When applying for a project, please include your CV, bachelor's and master's transcripts, and a one-page letter of motivation describing your research interests and educational background.

Klaus-Rudolf Kladny, kkladny@tuebingen.mpg.de