

Task-oriented chatbots have increasingly adopted neural language models, but their application in sensitive domains like healthcare remains limited due to safety concerns. Nutritional counselling, i.e., offering tailored dietary advice, represents a promising yet under-explored use case for these technologies. Annotated datasets for nutritional counselling have demonstrated that language models can produce mostly safe outputs in this domain. However, the effect of AI-delivered nutritional counselling using this technology is largely unknown. This thesis evaluates the integration of a rule-based chatbot with a fine-tuned large language model (LLM) for nutritional counselling and a prompted LLM for rephrasing templated chatbot outputs to improve user engagement. The approach was assessed intrinsically using automatic metrics and human assessments, and extrinsically through a user trial. Results showed no consistent dietary improvements or emotional benefits for participants receiving nutritional counselling, and minimal influence of rephrased responses on user behaviour. The findings emphasise the need for further research on evaluating AI-driven healthcare.