Student name: Cheng Xue Supervisor contact name: Adrian Clear

Student number: 24259661 Supervisor contact email: adrian.clear@universityofgalway.ie

RAG-based Chatbot for Domestic Abuse Support

Introduction

Domestic abuse remains a critical issue worldwide, with victims often struggling to access timely information and emotional support. Recent research emphasizes the potential of chatbots to offer guidance on recognizing abuse, making safety plans, and directing survivors to legal or social services [1]. In this project, we aim to develop a Retrieval-Augmented Generation (RAG) [2] chatbot to provide victims of domestic abuse with accurate, empathetic, and confidential assistance.

Background Research or Context

Existing support services often rely on fixed structure guidelines or live chat staffed by human operators. Interviews with experts and analysis of support websites reveal these common needs [1]: rapid risk assessment, accurate legal and practical guidance (e.g., housing, financial aid), and, most importantly, empathetic communication that assures users "You are not at fault" and "You are not alone." Research shows that a text-based conversational agent can offer immediate help while preserving privacy and reducing perceived judgment. However, purely AI-driven responses based on general large language model (LLM) like ChatGPT may introduce risks if the chatbot cannot handle sensitive situations. A hybrid RAG approach mitigates this by grounding responses in high-quality, up-to-date knowledge bases, ensuring reliability and safety.

Proposed Project

This project will focus on building a chatbot that uses Retrieval-Augmented Generation (RAG) to provide survivors (or potential survivors) of domestic abuse with accurate, contextually grounded, and empathetic support.

Retrieval-Augmented Generation (RAG) is an approach that enhances large language models (LLMs) by integrating external knowledge retrieval into the text generation process. Instead of relying solely on pre-trained model parameters, RAG dynamically retrieves relevant documents or data from an external source (e.g., a vector database or knowledge base) to improve response accuracy and relevance[3].

The RAG approach blends two core techniques: first, it retrieves reliable and up-to-date information, such as legal guidelines, local support resources, and best practices for safety planning, from an external knowledge base collected in collaboration with domestic abuse support organizations. Then, it uses generative methods to shape coherent, conversational responses that reflect the retrieved information. By grounding all outputs in real data rather than relying solely on a large language model, the chatbot reduces the risk of misinformation and maintains a trustworthy, up-to-date knowledge source, which is very important in our use case.

The project will collaborate with external organizations like Tusla, as well as social media, online forums, and relevant websites to compile references, from legal guidelines to safety-planning advice. These materials will then feed into our RAG pipeline[4], ensuring that all outputs are founded on credible, up-to-date data. Consultation with these organizations' staff will also help shape the chatbot's conversation flows and validate its effectiveness.

Overall, by grounding generative responses in a carefully maintained repository of expert content, the chatbot aims to offer confidential, and effective guidance to those experiencing or witnessing domestic abuse.

Timeline

Pre-Exam (before 9 May) Continue reading on RAG-based systems and domestic abuse use cases papers; collect initial datasets and references for legal and psychosocial support using some already open source dataset (not so up-to-date or not specialized for Ireland, but I can use it to test, for example, Reddit Mental Health Dataset and Bangladeshi Male Domestic Abuse Dataset); try to use these and some noted libraries to build and refine the RAG pipeline: implement the retriever and generator; In the meantime collaborate with external organizations (like Tusla

or online forums) to gather latest resources, mainly like testimonies of domestic abuse experiences; read other review papers about using chatbot to support people in demostic abuse but not have to be with LLMs.

Core Development (9 May – 19 June): Establish minimum viable product (MVP) with essential conversation flows (like recognizing abuse, offering next steps), prepare my formal PP.

PP Submission (20 June): Submit the required Project Proposal document.

Testing and Iteration (Late June – Mid-July): Test with domain experts or related users, validate accuracy, responsiveness, and empathy in the chatbot's interactions. Iterate on the chatbot's design based on feedback and test, ensuring that outputs remain reliable and empathetic.

Refinement and adjustment (Mid-July – Early August): if all previous steps are completed and are good enough: add more features.(e.g., "quick exit" options, delete data for privacy) and try to consider real production environment (deployment, GDPR, AI regulation law) with safety features.

Final Thesis Submission (Early August - 21 August): Document all outcomes, and limitations, and suggestions for future enhancements.

References:

[1] Saglam, Rahime Belen, Jason RC Nurse, and Lisa Sugiura. "Designing chatbots to support victims and survivors of domestic abuse." *arXiv preprint arXiv:2402.17393* (2024).

[2] Sun, Lipeipei, et al. "Persona-L has Entered the Chat: Leveraging LLM and Ability-based Framework for Personas of People with Complex Needs." *arXiv preprint arXiv:2409.15604* (2024).

[3] Zhao, Penghao, et al. "Retrieval-augmented generation for ai-generated content: A survey." *arXiv preprint arXiv:2402.19473* (2024).

[4] Neupane, Subash, et al. "From questions to insightful answers: Building an informed chatbot for university resources." *arXiv preprint arXiv:2405.08120* (2024).