AI CHATBOT VS POLICY HUB FOR ACCESSING CLINICAL GUIDELINES: A PROOF-OF-CONCEPT QUALITY IMPROVEMENT STUDY

Authors: Dr Nusrat Islam, Dr Isabelle Shepard. Supervised by Dr Alexander Bulcock

Background

Local, clinical guidelines are **essential** for safe prescribing and effective decision-making. In the Northern Care Alliance (NCA), we utilise the NCA Policy Hub. We have found this can often be **time-consuming**, and **challenging** for rotational staff such as Foundation Doctors, who can be unfamiliar with local systems. These barriers can delay decision-making, interventions and lead to inconsistent implementation of guidelines.

We have highlighted artificial intelligence (AI) as a potential solution. By utilising **retrieval-based tools**, chatbots can enable quick access to guideline-based information; improving **efficiency**, **accessibility** and **consistency** in care.

Our Aim

To evaluate the efficiency, accuracy and user experience of an AI chatbot for accessing local guidelines compared to the NCA Policy Hub.

User Interface

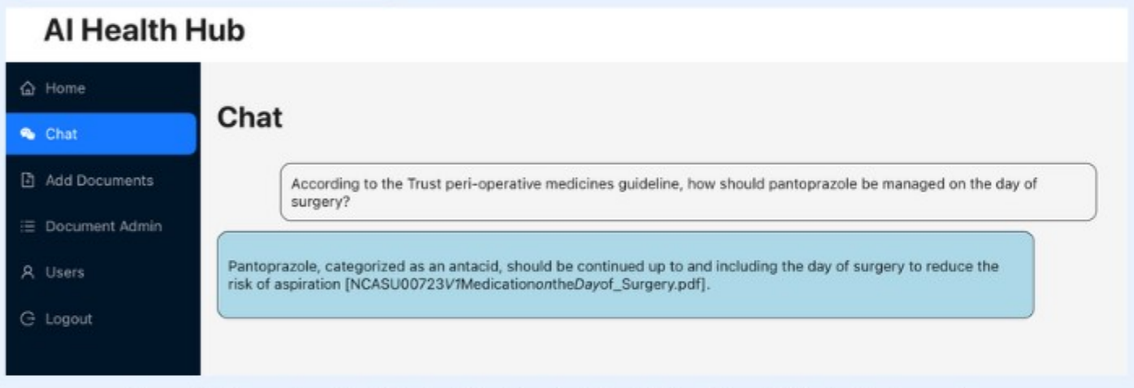


Figure 1. Example of AI Chatbot Retrieval of Guideline-Based Clinical Information



Figure 2. Notice on the Home Page of Our Chat Bot

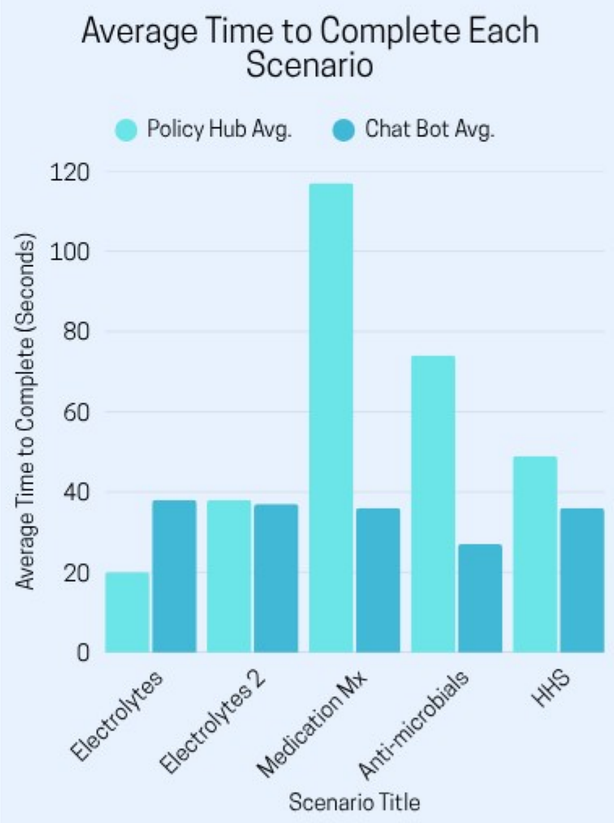
Methodology

We created **5 clinical scenarios** involving the management of electrolyte imbalances, peri-operative medication, anti-microbial prescribing and glycaemic states to carry out a **proof-of-concept comparative evaluation**. Scenarios were designed and discussed with supervising consultants to validate **clinical accuracy**, and included both straightforward retrieval tasks as well as complex navigation of less explicitly indexed information.

FY1s and SHOs were recruited to complete scenarios individually using either an AI chatbot or the NCA Policy Hub. **3 FY1s** and **2 SHOs** were assigned to each media. Participants were tested under standardised and timed conditions. The NCA Policy Hub is **not** a live hub, but instead a search engine to browse guideline PDFs that are regularly updated. The chatbot uses these PDFs from the same time period.

Our outcome measures were **time** to answer, **accuracy** based on a pre-defined and validated mark scheme. **No** real patient data was used.

Analysis



On average, the participants using the NCA Policy Hub took ***4 minutes and 58 seconds*** to complete **all 5** scenarios. In comparison, the participants using the AI chat bot took ***2 minutes and 54*** seconds.

On average, the group using the AI chatbot completed **all 5** scenarios **1.71 times faster** than the group using the NCA Policy Hub.

Each scenario took between **27** and **38 seconds** for the AI group, compared to the Policy Hub group that took anywhere between **20** and **117 seconds** to complete, with scenario 2 and 3 largely contributing to this disparity.

Discussion

Our chat bot demonstrated a **clear improvement** in the efficiency of retrieving local information, allowing participants to ask **focused**, natural questions to receive **direct answers**, rather than navigating lengthy documents. For example, in the anti-microbial scenario, some Policy Hub users overlooked key details, such as avoiding ciprofloxacin in patients over 65; the AI chatbot **consistently identified** this information. These findings suggest that AI reduces reliance on exact keyword searching.

User experience differed between the two approaches. Policy Hub users frequently relied on exact wording and document familiarity, whereas the AI chatbot enabled more intuitive, natural language interaction. However, the effectiveness of the chatbot was **dependent** on the **quality** of user **prompts**, with more specific queries producing more complete responses. While the chatbot provided source references for retrieved information.

Several **limitations** were identified. AI performance was occasionally affected by system **latency** and **network delays**, which may limit usability in clinical environments. Additionally, there is a potential risk of **over-reliance**, with users accepting AI-generated outputs without appropriate verification against guidelines.

Conclusion

This proof-of-concept study suggests that AI chatbots can improve the efficiency, consistency, and accessibility of local guideline retrieval compared to the NCA Policy Hub. The greatest benefit was observed in scenarios requiring interpretation or where guidance was not explicitly indexed.

Accuracy was comparable between both methods; however, AI reduced reliance on navigation and keyword searching, which may help minimise missed details.

Despite these advantages, the effectiveness of AI was dependent on user prompting and system performance. There is a potential risk of over-reliance without appropriate verification. AI tools should therefore be used as an adjunct to, rather than a replacement for, official Trust guidelines to ensure safe clinical practice, with consideration being given to the environmental impact of AI systems.