

Introduction

Artificial intelligence (AI) chatbots for primary care have yet to gain traction in Switzerland despite high-performing large language models (LLM) being preferred for accessing and retrieving complex information.

“In a Nutshell”, an online educational program for physicians, organises clinical information in a NUTE manner (near-real time, understandable, time-saving, and efficient) using “Ask Dr. Nuts”, a purpose-built chatbot for retrieving curated medical information.

This tool can be a co-pilot for navigating medical literature in daily care practice. We assessed the UTuCo (Utility, Trustworthiness and Correctness) of the chatbot’s information and its acceptance by physicians.

Methods

The chatbot was trained using Azure’s pre-processed model. Medical literature was stored in segmented units, converted into semantic embeddings, and used to generate the responses.

We analysed 1015 interactions for the date and time of interaction, the message type (user or chatbot), the document context, and response correctness.

A survey captured user experiences and expert confidence in the answers—the first analysed “Therapy” and “Diagnosis” classes with a battery of non-case-sensitive strings.

The experts evaluated high-rated responses across five medical fields.

Results

Physicians used chatbots during practice hours (Fig1 a) to support diagnosing and treating (b) the most common diseases aligned with general practitioners' roles (c).

Users rated the “In a Nutshell” answers (d, e, f) trustworthy and recommendable: 78.6% trusted the answers (g), 57.1 % would consult the chatbot again (h), and 53.6% would recommend it to colleagues (i).

Medical experts confirmed the chatbot’s accuracy (71%) (j), completeness (44%) (k), and applicability (81%) (l).

Conclusions

The potential of medical AI chatbots in primary care highlights challenges that must be addressed before broader implementation. Responses require greater transparency, traceability, accuracy, and rate. The chatbot must provide reliable information in all official Swiss languages.

Establishing agreements on training data, evaluation criteria, accuracy, and risk factors is crucial for quality assurance. Ensuring high medical content standards, comprehensively training AI, and benchmarking against other AI healthcare tools is necessary for improving patient care.

Dominique Froidevaux - Chief Executive Officer of Pro Medicus GmbH. The “In a Nutshell” content and the “Ask Dr. Nuts” chatbot are Pro Medicus GmbH products.

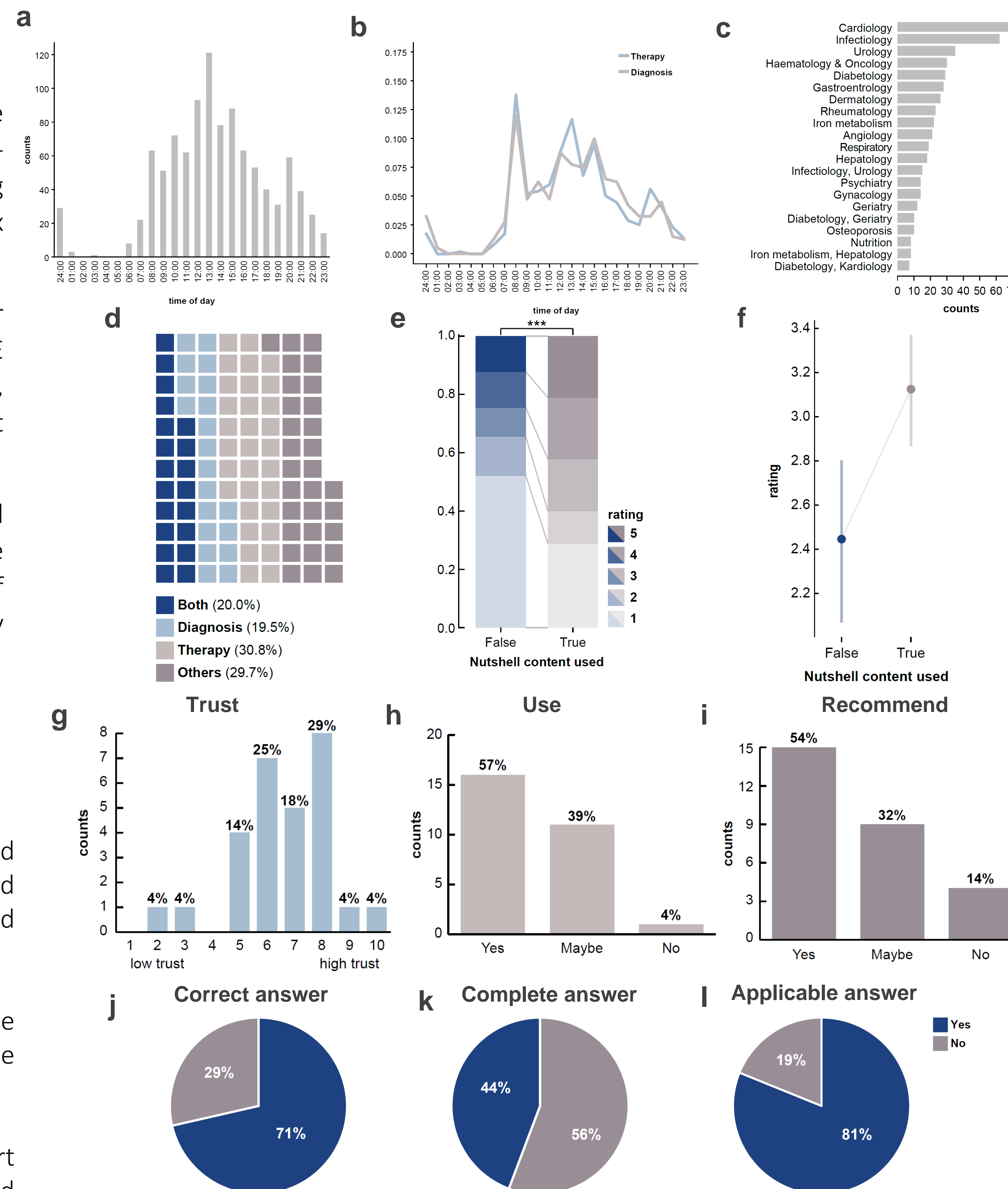


Figure 1. Histograms (a) and (b) show user-chatbot interactions over a day filtered by “Therapy” and “Diagnosis” classes, and bar chart (c) shows the medical topics accessed. (d) The waffle chart highlights the fraction of user-chatbot interactions by class. (e) The studded bar chart evidences the significant differences ($p < 0.001$ Chi-square test) in responses rated (rating from 1 to 5) using or not chatbot content. (f) “In a Nutshell” content use rates the answers—survey results. Bar graphs contain the counts per category and the fractional equivalent in percentages above each bar for chatbot (g) trust, (h) re-use, and (i) recommendation to a colleague. The pie chart graphs present the percentage of (j) correct, (k) complete and (l) applicable chatbot answers to daily medical practice.