A large language model-powered digital assistant for up-to-date COVID-19 outpatient management

Marco Montagna, Filippo Chiabrando, Federico Baldini, Davide Ficocelli, Giuseppe Faraci, Patrizia Rovere Querini

Submitted to: Journal of Medical Internet Research
on: October 07, 2024

Table of Contents

A large language model-powered digital assistant for up-to-date COVID-19 outpatient management

Marco Montagna[1*]; Filippo Chiabrando[1*]; Federico Baldini[2]; Davide Ficocelli[2]; Giuseppe Faraci[2]; Patrizia Rovere Querini[1]

[1]Vita-Salute San Raffaele University Milan IT
[2]Run2Cloud srl Milan IT
[*]these authors contributed equally

Corresponding Author:
Marco Montagna
Vita-Salute San Raffaele University
Via Olgettina 58
Milan
IT

Abstract

The medical community is both excited and worried by the advent of artificial intelligence in healthcare, whereby two disciplines meet for the benefit of patients. This is especially true when considering large language models (LLM) for generative functions. This letter briefly describes a successful collaboration between physicians and informaticians to develop a LLM-powered chatbot for outpatient management of COVID-19.

(JMIR Preprints 07/10/2024:67172)
DOI: https://doi.org/10.2196/preprints.67172

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A large language model-powered digital assistant for up-to-date COVID-19 outpatient management

Montagna M.[1,2], Chiabrando F.[1], Baldini F.[3], Ficocelli D.[3], Faraci G.[3], Rovere Querini P[1,2]

1. Vita-Salute San Raffaele University, School of Medicine, Milan, Italy
2. Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan
3. Run2Cloud Srl

F. Chiabrando is an internal medicine resident, Vita-Salute San Raffaele University, Milan, Italy
M. Montagna is a PhD student at Vita-Salute San Raffaele University, Milan, Italy and an internal medicine consultant at the Unit of Medical Specialties and Healthcare Continuity, IRCCS San Raffaele Scientific Institute, Milan, Italy
F. Baldini is Project Manager at Run2Cloud Srl for this research
D. Ficocelli is a developer at Run2Cloud Srl for this research
G. Faraci is CEO at Run2Cloud Srl
P. Rovere Querini is associate professor of internal medicine, Vita-Salute San Raffaele University, Milan, Italy, and director of the Unit of Medical Specialties and Healthcare Continuity, IRCCS San Raffaele Scientific Institute, Milan, Italy

Corresponding Author:
Marco Montagna
montagna.marco@hsr.it
Vita-Salute San Raffaele University, Via Olgettina 58, 20132, Milan, Italy

Keywords
Artificial Intelligence; Chatbot; Large Language Models; Clinical Decision Support Systems; COVID19

Introduction

Effectively delivering treatment for acute SARS-CoV-2 infection has been a challenging quest. The virus is highly mutagenic, generating variants resistant to previously effective therapies. The therapeutic window is narrow, forcing timely precision in COVID-19 management [1]. Moreover, large epidemic waves call the whole medical workforce – not just Infectious Disease specialists– to deal directly with patients and develop confidence in prescribing targeted treatment [2].
It is therefore essential for healthcare professionals to quickly access the latest clinical recommendations, to recommend appropriate treatment and/or refer the patient to the most appropriate level of care.
While clinical guidelines are regularly published, their format is fragmented and difficult to approach. Translating such a body of material into summarized, flexible, interactive quick-start guides would foster medical education and boost physicians' confidence, ultimately favoring proper patient management [3,4].
To this aim, we developed a large language model (LLM)-based chatbot fine-tuned on guidelines for outpatient management of COVID-19, tailored to the local Italian context.

Materials and methods

We put together a team of two physicians and two informaticians.

The physicians evaluated the end-users' requirements for the chatbot, that was expected to use proficient medical language, to strictly attain to clinical guidelines, and to collect all the necessary information before producing the final recommendation. Such information (age, comorbidities, time elapsed since symptom onset and since swab testing, peripheral oxygen saturation, renal and liver function tests) was instrumental for producing personalized patient advice. They also provided the medical background knowledge by selecting clinical guidelines from the Italian Drug Agency and from UpToDate, and by curating the master prompt to set the tone, structure and required elements for the conversation [5]. Lastly, physicians served as testers throughout the development process.

The informaticians provided the architectural framework to be tested, in five subsequent versions, each troubleshooting specific issues encountered. The first version was built on two workflows. Firstly, DialogFlowCX was employed to retrieve and store the personal clinical information of the patient inputted by the physician-user [6]. Secondly, such information was to be passed to FlowiseAI, a low code drag&drop environment, which would then operate on the collected data and provide the output, in the form of chat dialogue. To do so, FlowiseAI employed two modules: one managing the *upsert*, meant for document upload and reading through the use of four "nodes" (text splitter, document loader, vectorial database, and embeddings); the other managing the interaction through OpenAI's ChatGPT 3.5 and a Conversational Retrieval QA Chain [7]. The second version solved the intra-conversational recall and memory loss thanks to an upgrade of node architecture, with an OpenAI function agent connected to a memory buffer. The third version tackled the tendency of the chatbot to hallucinate, inventing random information while ignoring the provided documental knowledge, through deeper fine-tuning of the master prompt. In the fourth version, informaticians finally decided to get rid of DialogFlowCX as input system, merging all functions into FlowiseAI, and letting ChatGPT 4 – a slower-reacting yet more intelligent model – incorporate the functions of information retrieval, dialogue interaction and output formation. The fifth and final version perfected the inconsistency of the chatbot to adhere strictly to the provided medical background by reformatting the uploaded documentation in plain text format.

Results

As it is now, the chatbot provides well written, tailored and solid suggestions for COVID-19 outpatient treatment and monitoring. Such suggestions replicate with high fidelity the current guidelines. The output of the Chatbot accurately resembles a report from a consultancy given by a medical specialist.

Discussion

The COVID-19 pandemic highlighted the need of providing a fast, comprehensive and engaging access to medical knowledge to physicians. LLM-based chatbots are being investigated for their ability to ingest, summarize and supply such knowledge, in a disease-oriented fashion [3,8,9]. In this report we provide a preliminary experience of the design and testing of an LLM-based chatbot fine-tuned on guidelines for outpatient management of COVID-19.

The challenges we encountered were neither trivial nor irrelevant. To be able to collect information, store it, recall it and interlacing it with other data is an essential feature of the medical profession and cannot be oversimplified. To maintain strict adherence to guidelines, while allowing personalization of indications to the individual patient is also critical. Such competences can and should be stressed and monitored in the development of a chatbot. Before a chatbot can be deployed as a clinical decision support system, reliability and robustness in these domains must be assured, while the local and international regulatory framework evolves to encompass the use of such tools in the medical profession, addressing issues such as privacy and accountability [10]. Our preliminary work suggests that there is great potential for the use of chatbots in the medical field, but that there are still challenges that need to be overcome. We plan to do further research and test a new chatbot in the context of another disease.

Aknowledgements

This research was supported by a grant from Pfizer Inc.

Conflicts of interest

The authors declare no conflicts of interest

Abbreviations

COVID-19	Coronavirus	Disease	19
GPT	General	Pretrained	Transformer
LLM	Large Language Model		

Author contribution statement (CRediT)

FC, MM: conceptualization, methodology, writing: original draft; FB, DF, GF: methodology; PRQ: conceptualization, supervision.

BIBLIOGRAPHY

1. De Lorenzo R, Montagna M, Bossi E, Vitali G, Taino A, Cilla M, Pata G, Lazorova L, Pesenti R, Pomaranzi C, Bussolari C, Martinenghi S, Bordonaro N, Di Napoli D, Rizzardini G, Cogliati C, Morici N, Rovere-Querini P. A Pilot Study of the Efficacy and Economical Sustainability of Acute Coronavirus Disease 2019 Patient Management in an Outpatient Setting. Front Med (Lausanne) Frontiers; 2022;9:1191. doi: 10.3389/fmed.2022.892962

2. de Sutter A, Llor C, Maier M, Mallen C, Tatsioni A, van Weert H, Windak A, Stoffers J. Family medicine in times of 'COVID-19': A generalists' voice. European Journal of General Practice. Taylor & Francis; 2020. p. 58–60. PMID:32349550

3. Beck S, Kuhner M, Haar M, Daubmann A, Semmann M, Kluge S. Evaluating the accuracy and reliability of AI chatbots in disseminating the content of current resuscitation guidelines: a comparative analysis between the ERC 2021 guidelines and both ChatGPTs 3.5 and 4. Scand J Trauma Resusc Emerg Med Scand J Trauma Resusc Emerg Med; 2024 Dec 1;32(1). PMID:39327587

4. Guerra-Farfan E, Garcia-Sanchez Y, Jornet-Gibert M, Nuñez JH, Balaguer-Castro M, Madden K. Clinical practice guidelines: The good, the bad, and the ugly. Injury Elsevier; 2023 May 1;54:S26–S29. PMID:35135686

5. Farmaci utilizzabili per il trattamento della malattia COVID-19. AIFA. 2021. Available from: https://www.aifa.gov.it/aggiornamento-sui-farmaci-utilizzabili-per-il-trattamento-della-malattia-covid19

6. Flow basics | Dialogflow CX | Google Cloud. Available from: https://cloud.google.com/dialogflow/cx/docs/basics [accessed Oct 4, 2024]

7. Introducing ChatGPT. Available from: https://openai.com/blog/chatgpt [accessed Jul 19, 2023]

8. Montagna S, Aguzzi G, Ferretti S, Pengo MF, Klopfenstein LC, Ungolo M, Magnini M. LLM-based Solutions for Healthcare Chatbots: a Comparative Analysis. 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2024 Institute of Electrical and Electronics Engineers Inc.; 2024;346–351. doi: 10.1109/PERCOMWORKSHOPS59983.2024.10503257

9. Andrew A. Potential applications and implications of large language models in primary care. Fam Med Community Health BMJ Publishing Group; 2024 Jan 30;12(Suppl 1):2602. PMID:38290759

10. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 6;6(1):120. doi: 10.1038/s41746-023-00873-0