

Improving Physician Communication Effectiveness Using Tailored LLM Prompts

Onofrio CAPPUCCI^a, Clara DE GENNARO^a,
Maura PUGLIATTI^{b,c}, Paola CROCIANI^a and Francesco GIULIANI^{a,1}

^aFondazione Casa Sollievo della Sofferenza Research Hospital

^bDepartment of Neuroscience and Rehabilitation, University of Ferrara, Ferrara, Italy

^cInterdepartmental Center of Research for Multiple Sclerosis and Neuro-inflammatory and Degenerative Diseases, University of Ferrara, Ferrara, Italy

ORCID ID: Onofrio CAPPUCCI <https://orcid.org/0009-0006-4032-1513>,

ORCID ID: Clara DE GENNARO <https://orcid.org/0000-0003-4803-8641>,

ORCID ID: Maura PUGLIATTI <https://orcid.org/0000-0001-9396-8683>,

ORCID ID: Francesco GIULIANI <https://orcid.org/0000-0002-5707-2158>

Abstract. We present the application of the COSTAR method as a tool to design LLM prompts that can effectively support the dialogue of neurologists with a pwMS

Keywords. Large Language Models, Multiple Sclerosis

1. Introduction

Communication between physicians and patients is key for ensuring quality of care and improving health outcomes. To establish effective communication, the contents must be tailored to the patient's profile and this is most relevant when physicians communicate diagnosis with potential important implications on patient's quality of life. We present a novel approach applicable to Large Language Models (LLMs) [1] to let them be more valuable tools to support physicians in their dialogue with patients.

2. Methods

We performed a literature scan and a physician informed selection of the most common questions posed by patient with Multiple Sclerosis (pwMS) to their doctors and prompted ChatGPT (GPT-4 model) with those questions, having this common structure "How would you reply to a patient with Multiple Sclerosis asking [question]?" E.g., one [question] can be [question] = "what is the cause of my disease?"

The replies produced by the LLM were generally very understandable yet lacked depth. We then redesigned the prompts introducing information about patient.

We specified the patient profile into the prompt introducing a description of patient emotional status built according to the Big Five Personality Test (B5PT) [2] and the

¹ Corresponding Author: Francesco Giuliani; E-mail: f.giuliani@operapadrepio.it.

Hamilton Anxiety Scale (HAS) [3]. To our surprise, in both cases the replies obtained by the LLM did not take into account the prompted values at all.

We then moved to the adoption of the COSTAR [4] methodology for prompt formulation, which provides a structured approach that must include the following elements: 1) Context, 2) Objectives, 3) Structure, 4) Tone, 5) Audience, and 6) Response. In a subsequent step, we also fed into the LLM the Italian guidelines on Multiple Sclerosis [5] operating the conversion from their original format to an optimally structured format according to the reformatting methods reported in [6].

3. Results, Discussion and Conclusions

The main result of our study relies on the efficacy of the COSTAR methodology in engineering prompts that, together with the appropriate feeding of medical guidelines and a natural language description of the patient profile, can effectively support physicians in their dialogue with pwMS.

A sample excerpt of a COSTAR-compliant prompt we used is reported the following.

I am a neurologist...one of my tasks is to communicate the diagnosis [or other questions] to patients affected by multiple sclerosis....Your task is to help me communicate [question] to the patient, taking into account their psychological profile, gender, age, education, and guidelines....The guidelines are contained in the file [reformatted guidelines]....The patient is a 25-year-old woman, she holds a master's degree in agriculture....she loves to travel....The response must be empathetic and detailed.

The replies we obtained with prompts designed according to this structure provide answers that are complete both in clinical terms and in adaptation to the characteristic of the specific patient. Starting from these preliminary results, a cohort of neurologists will assess the LLM replies to these COSTAR formatted prompts according to the dimensions of clinical correctness, empathy, and readability. A future development can be the assessment of the actual benefits of this tool in clinical communication. Note: Study funded by the European Union –Next Generation EU – NRRP M6C2 – Investment 2.1 Enhancement and strengthening of biomedical research in the Italian NHS.

References

- [1] Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. *Front Med (Lausanne)*. 2024 Oct 29;11:1477898.
- [2] John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big five trait taxonomy. In: *Handbook of personality: Theory and research*. 3rd ed. New York: Guilford Press; 2008. p. 114–58.
- [3] Maier W, Buller R, Philipp M, Heuser I. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *J Affect Disord*. 1988;14(1):61–8.
- [4] Teo S. How I won Singapore's GPT-4 Prompt Engineering Competition: A deep dive into the CO-STAR framework [Internet]. Singapore: Government Technology Agency of Singapore (GovTech); 2023 [cited 2025 May 4]. Available from: <https://vivas.ai>
- [5] Società Italiana di Neurologia, Diagnosi e terapia della Sclerosi Multipla nell'Adulto, 2022, Sistema Nazionale delle Linee Guida – ISS - <https://www.iss.it/-/snlg-sclerosi-multipla-adulto>
- [6] Kresevic S, Giuffrè M, Ajcevic M, Accardo A, Crocè LS, Shung DL. Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. *Digit Med*. 2024 Apr 23;7(1):102. doi: 10.1038/s41746-024-01091-y. PMID: 38654102; PMCID: PMC11039454.