ontrat doctoral – ED Galilée

Titre du sujet :

Unité de recherche : Laboratoire de recherche en informatique pour la santé–LIMICS-UMRS 1142

- Discipline : Informatique
- Direction de thèse : Karima Sedki. Co-directeur : Jean-Baptiste Lamy
- Contact : karima.sedki@univ-paris13.fr
- Domaine de recherche : Intelligence artificielle, Argumentation formelle, guides de bonnes pratiques
- Mots clés : LLMs, extraction et formalisation de connaissances médicales, Argumentation formelle, raisonnement, explicabilité

Medical knowledge [1,2] is available in different forms: textual (e.g., clinical practice guidelines (CPGs), PubMed articles, and systematics reviews), structured (e.g., ontologies, knowledge graphs), and knowledge from experience (e.g., clinical practice, discussions). This knowledge is important for clinicians that use it for diagnosis, treatment, researchers for building models like ontologies and AI systems, etc. However, take advantage of this knowledge in daily clinical practice faces several problems [3]: i) volume and complexity (guidelines are long textual documents, making their interpretation difficult in clinical practice), ii) contextual gaps (guidelines do not give recommendations for all patients), iii) contradictions and inconsistency between guidelines, and iv) lack of comprehension (guidelines are not always well understood by clinicians, posing a risk of non-adoption).

Recent advances have shown the growing use of Large Language Models (LLMs) in the medical domain to perform various tasks such as automating information extraction from long textual documents, suggesting treatments, and answer questions by the application of various prompting techniques [4]. This makes them promising tools for use in decision making, but they suffer from hallucinations, logical inconsistencies, and lack of contestability and explainability, particularly when faced with highly complex clinical reasoning (LLMs often make unfaithful reasoning, i.e., the derived output does not follow the generated reasoning steps [5,6]. In [7,8], conducted studies have revealed that LLMs are currently not ready for autonomous clinical decision-making and highlighted the need for continued research and collaboration between healthcare professionals and computer scientists to ensure safe, ethical, and effective deployment of LLMs in clinical practice.

The aim of this PhD thesis is to is to develop methods based on LLMs and formal argumentation to (1) extract, structure, and formalize knowledge of CPGs in the form of argumentation system and (2) use the obtained system for making decisions. The concept of argumentation [9] draws inspiration from the argumentative nature of human reasoning (in guidelines, recommendations are typically supported or justified by arguments). Argumentation offers a structured, transparent, and explainable framework for handling conflicting knowledge and evidence through the explicit modelling of supporting and attacking arguments. By combining the natural language understanding capabilities of LLMs with the rigor of argumentation theory, the approach aims to generate explainable, robust, logically grounded recommendations for clinicians. This synergy enables not only the identification of relevant arguments and counterarguments within CPGs but also supports transparent decision-making aligned with clinical evidence and reasoning principles.
The key objectives are to address the following challenges:

1) **Knowledge Extraction and Formalization from CPGs:** The aim is to design methods leveraging LLMs to extract, structure, and formalize medical knowledge from CPGs. The main challenges are: i) Extracting clinically relevant knowledge, including recommendations, arguments, and preferences from unstructured and heterogeneous textual CPGs, ii) Capturing logical dependencies, contextual conditions, preferences, and argument structures that underpin these recommendations, and iii) Translating the natural language outputs of LLMs into formal argumentation graphs. The methodology will focus on: i) Selecting LLM architectures with strong performance in medical tasks (e.g , BioGPT, Mistral/Mixtral) and developing methods suitable for the targeted extraction of recommendation components, ii) Using symbolic methods such as semantic parsing and ontology mapping to transform LLMs outputs into structured

knowledge suitable for formal reasoning and iii) Developing hybrid representations that explicitly link unstructured textual evidence from guidelines to argumentation framework that allow formal reasoning and explainability. Our hypothesis is that integrating the descriptions of argument schemes [10] in the prompts may help LLMs in identifying key knowledge.
Given the diversity of existing argumentation frameworks, which differ in expressive power and evaluation strategies, it is necessary to identify the most appropriate formalisms for clinical applications

2) Formal Argumentative Reasoning: The objective is to implement the reasoning process of the generated argumentation framework capable of evaluating, comparing, identifying potential conflicts and ensuring logical consistency with established medical knowledge. Our objective here is to implement semantics that allows quantifying arguments and recommendations in the spirit of gradual semantics of QuAD systems [11]. This allows rank order the set of recommendations which makes explanations complete and justifiable.

3) Generating and verbalizing explanations: The objective is to produce and verbalize explanations in natural language. The main challenges here are: i) Generating faithful explanations that reflect the reasoning steps of the argumentation reasoning and ii) Adapting the level of explanation following the user profile and needs. Our methodological approach consists in: i) Designing an explanation model including symbolic explanations derived from argumentation framework, and LLM-based natural language verbalization, ii) Training or prompting LLMs using structured reasoning traces as input to ensure alignment with the underlying logic, and iii) Evaluating explanation quality using human-in-the-loop protocols (expert review, trust and usability assessments).

4) Clinical evaluation: The aim is to evaluate the proposed system, assessing its accuracy, interpretability, usability and trustworthiness in a simulated clinical contexts. This includes two evaluations: i) Comparison between the expert-produced test datasets (including medical questions, arguments and decisions) with argument graph produced by the system, ii) Simulation trial for evaluating the proposed system with physicians, on the basis of fictitious clinical cases.

Bibliographie

[1] M. Solomon. Making Medical Knowledge. Oxford University Press, 04 2015.
[2] S. H. Woolf, R. Grol, A. Hutchinson, M. Eccles, and J. Grimshaw. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ Clinical research ed., 7182(318) :527–530, 1999.
[3] M. D. Cabana, C. S. Rand, N. R. Powe, A. W. Wu, M. H. Wilson, P.-A. C. Abboud, and H. R. Rubin. Why don't physicians follow clinical practice guidelines? a framework for improvement. JAMA, 15(282) :1458–65, 1999.
[4] D. Chen, S. A. Alnassar, K. E. Avison, R. S. Huang, and S. Raman. Large language model applications for health information extraction in oncology: Scoping review. JMIR Cancer, 11, 2025.
[5] G. Freedman, A. Dejl, D. Gorur, X. Yin, A. Rago, and F. Toni. Argumentative large language models for explainable and contestable claim verification. In AAAI-25, pages 14930–14939, 2025.
[6] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald. On faithfulness and factuality in abstractive summarization, 2020.
[7] Evaluating large language models and agents in healthcare: key challenges in clinical applications. Intelligent Medicine, 5(2) :151–163, 2025.
[8] P. Hager, F. Jungmann, R. Holland, and et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med, 30 :2613–2622, 2024.
[9] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artif. Intell., 77(2) :321–358, 1995.
[10] Douglas Walton, Chris Reed, and Fabrizio Macagno. Argumentation Schemes. Cambridge University Press, New York, 2008
[11] A. Rago, F. Toni, M. Aurisicchio, and P. Baroni. Discontinuity-free decision support with quantitative argumentation debates. In KR, pages 63–73, 2016.

Karima Sedki