LLMs for Modern Clinical Evidence Summarization

Weng Lab @ Dept. of Biomedical Informatics
Columbia University Irving Medical Center
2023.10.24

Evidence Based Medicine (EBM)

Levels of Evidence Meta analysis / Systematic Reviews **RCTs** Quality **Observational Studies** Case Series and Reports **Background and Expert Opinions** Volume

~133K trials registered at <u>ClinicalTrials.gov</u> since 2020¹.

67.3 weeks² to publish a review.

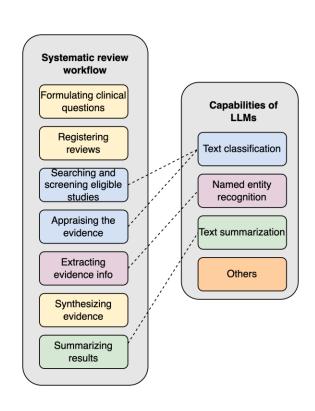
Clinical data as a source of evidence³.

!! Need for scalable and trustworthy approaches!!

- 1. ClinicalTrials.gov, accessed on Oct 9th, 2023.
- 2. Borah, Rohit, et al. "Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry." BMJ open 7.2 (2017): e012545.
- 3. Gershman, Boris, David P. Guo, and Issa J. Dahabreh. "Using observational data for personalized medicine when clinical trial evidence is limited." Fertility and Sterility 109.6 (2018): 946-951.

Future directions

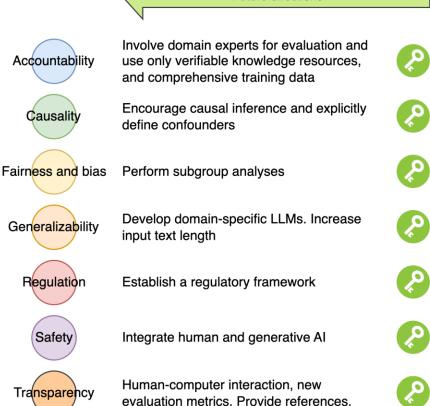
Challenges in Leveraging LLMs for EBM



Challenges Intrinsic or extrinsic hallucinations, synthesis of contradicting sources Biased assessment of treatment effect Data bias, knowledge bias, exposure bias, disparity censorship Generalist or specialist, Robustness to distribution shift Lack of policies or laws Safe utilization of LLM-generated summaries Model complexity, uncertain model

behaviors, proprietary technology, diverse

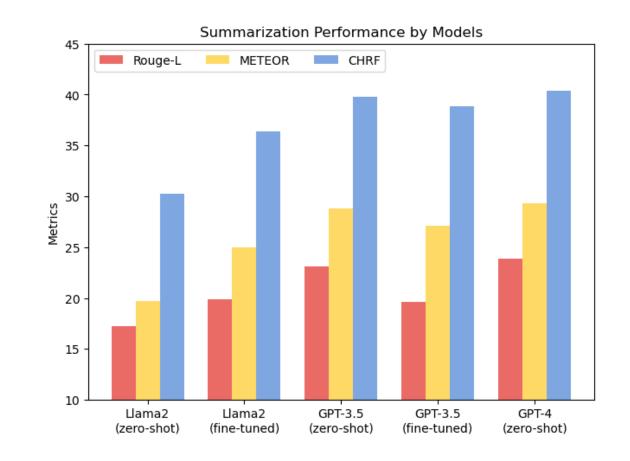
stakeholders, and up-to-date evidence.



How may academia research complement industry?

Case Study: Zero-shot vs. Fine-tuned LLMs for Medical Evidence Summarization

- **Task:** LLMs for medical evidence summarization¹.
- **Data:** over 8,000 systematic reviews.
- **Methods:** fine-tuned and evaluated LLMs.

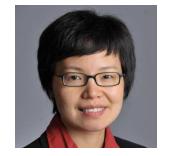


^{1.} Tang, Liyan, et al. "Evaluating large language models on medical evidence summarization." npj Digital Medicine 6.1 (2023): 158.

Related Publications

- Tang, L., Sun, Z., Idnay, B., Nestor, J.G., Soroush, A., Elias, P.A., Xu, Z., Ding, Y., Durrett, G., Rousseau, J.F., Weng, C., and Peng, Y., 2023. Evaluating large language models on medical evidence summarization. *Nature Partner Journals Digital Medicine*, 6(1), p.158.
- **Peng, Y.**, Rousseau, J.F., Shortliffe, E.H. and **Weng, C.**, 2023. AI-generated text may have a role in evidence-based medicine. *Nature Medicine*, pp.1-2.
- Liu, H., Peng, Y. and Weng, C., 2023. How Good Is ChatGPT for Medication Evidence Synthesis?. *Stud Health Technol Inform*. 2023, p.1062.
- Kang, T., Sun, Y., Kim, J.H., Ta, C., Perotte, A., Schiffer, K., Wu, M., Zhao, Y., Moustafa-Fahmy, N., **Peng, Y.** and **Weng, C.**, 2023. EvidenceMap: a three-level knowledge representation for medical evidence computation and comprehension. *Journal of the American Medical Informatics Association*, 30(6), pp.1022-1031.
- Turfah, A., Liu, H., Stewart, L.A., Kang, T. and Weng, C., 2022, June. Extending PICO with Observation Normalization for Evidence Computing. In *MEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation: Proceedings of the 18th World Congress on Medical and Health Informatics* (Vol. 290, p. 268). IOS Press.
- Chen, Z., Liu, H., Liao, S., Bernard, M., Kang, T., Stewart, L.A. and **Weng, C.**, 2022. Representation and Normalization of Complex Interventions for Evidence Computing. *Studies in Health Technology and Informatics*, 290, pp.592-596.
- Kang, T., Turfah, A., Kim, J., Perotte, A. and **Weng, C.**, 2021. A neuro-symbolic method for understanding free-text medical evidence. *Journal of the American Medical Informatics Association*, 28(8), pp.1703-1711.
- Sahin, A.N., Goldstein, A. and Weng, C., 2019. Analysis of evidence appraisals for interventional studies in family medicine using an informatics approach. *Primary Health Care Research & Development*, 20, p.e123.

Our Team



Chunhua Weng (PI)



Yifan Peng (PI, Collaborator)



Zhiyong Lu (Collaborator)



Gongbo Zhang



Qiao Jin (Collaborator)



Betina R. Idnay

Collaborators:















Acknowledgement:

Our projects were supported by the National Library of Medicine, National Institutes of Health, through grants numbered R01LM014344, R01LM009886 and R01HG012655.