Precision Health in the Age of Large Language Models

Hoifung Poon
hoifung@microsoft.com
Microsoft Research
Redmond, WA, USA

Tristan Naumann
tristan@microsoft.com
Microsoft Research
Redmond, WA, USA

Sheng Zhang
zhang.sheng@microsoft.com
Microsoft Research
Redmond, WA, USA

Javier González Hernández
gonzalez.javier@microsoft.com
Microsoft Research
Cambridge, UK

ABSTRACT

Medicine today is imprecise. Among the top 20 drugs in the U.S., up to 80% of patients are non-responders. The goal of precision health is to provide the right intervention for the right people at the right time. The key to realize this dream is to develop a data-driven, learning system that can instantly incorporate new health information to optimize care delivery and accelerate biomedical discovery. In reality, however, the health ecosystem is mired in overwhelming unstructured data and excruciating manual processing. For example, in cancer, standard of care often fails, and clinical trials are the last hope. Yet less than 3% of patients can find a matching trial, whereas 40% of trial failures simply stem from insufficient recruitment. Discovery is painfully slow as a new drug may take billions of dollars and over a decade to develop.

In this tutorial, we will explore how large language models (LLMs) can serve as a universal structuring tool to democratize biomedical knowledge work and usher in an intelligence revolution in precision health. We first review background for precision health and give a broad overview of the AI revolution that culminated in the development of large language models, highlighting key technical innovations and prominent trends such as consolidation of AI methods across modalities. We then give an in-depth review of biomedical LLMs and precision health applications, with a particular focus on scaling real-world evidence generation and drug discovery. To conclude, we discuss key technical challenges (e.g., bias, hallucination, cost), societal ramifications (e.g., privacy, regulation), as well as exciting research frontiers such as prompt programming, knowledge distillation, multi-modal learning, causal discovery.

ACM Reference Format:

Hoifung Poon, Tristan Naumann, Sheng Zhang, and Javier González Hernández. 2023. Precision Health in the Age of Large Language Models. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), August 6–10, 2023, Long Beach, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3580305.3599568

KDD '23, August 6–10, 2023, Long Beach, CA, USA

ACM ISBN 979-8-4007-0103-0/23/08.
https://doi.org/10.1145/3580305.3599568

1 TARGET AUDIENCE AND PREREQUISITES

This tutorial is intended for computer science researchers and biomedical practitioners interested in applying advanced AI methods to precision health. While audience with background in the above areas would benefit most, this tutorial is designed to present a picture of precision health in the age of large language models (LLMs) such that no specific background knowledge is assumed.

2 TUTORS

Hoifung Poon is General Manager at Health Futures of Microsoft Research and an affiliated professor at the University of Washington Medical School. He leads biomedical AI research and incubation, with the overarching goal of structuring medical data to accelerate discovery and improve delivery for precision health. He has given tutorials on this topic at top conferences such as ACL and AAAI. His research spans a wide range of problems in machine learning (ML) and NLP, and prior work has been recognized with Best Paper Awards from premier venues such as NAACL, EMNLP and UAI.

Tristan Naumann is a Principal Researcher at Microsoft Research's Health Futures working on problems related to clinical NLP and machine reading. His research focuses on exploring relationships in complex, unstructured healthcare data using NLP and unsupervised learning techniques. He is currently serving as a General Chair of NeurIPS and has previously served in a variety of Organizing Chair roles for NeurIPS, AHLI CHIL, ML4H, and Clinical NLP.

Sheng Zhang is a Senior Researcher at Health Futures, Microsoft Research. His research focuses on self-supervised NLP. He has over 30 patents and research articles in NLP, acted on the review panel of numerous leading journals (such as Transactions on ACL, Computational Linguistics and BMC Bioinformatics) and conferences (such as ACL, AAAI, EMNLP, and NAACL). He also served as area chair at top NLP conferences such as EMNLP and NAACL and have the experience of organizing workshops in these premier conferences.

Javier González Hernández is a Principal Researcher at Health Futures of Microsoft Research. His research focuses on the development of probabilistic ML methods for uncertainty quantification and data-efficient sequential decision making. He works on the challenges arising when uncertainty of different types (the loss of precision induced by numerical calculations, data errors, model miss-calibration, etc.) need to be be propagated, controlled and reduced in complex pipelines. He is also interested on how causal inference can be used to leverage decision making methods and to improve the understanding of complex systems and processes.

3 TUTORIAL OUTLINE

- Precision Health (30 min)
  - The dream
  - Why now
  - Why progress is slow
  - How AI can help
- The Intelligence Revolution (30 min)
  - A brief history of NLP and deep learning
  - The great consolidation in AI
  - Large language models (LLMs)
- LLMs for Precision Health (45 min)
  - Biomedical LLMs
  - GPT-4 in medicine
  - General vs specialized LLMs
  - LLMs for real-world evidence
  - LLMs for drug discovery
- Application challenges (45 min)
  - Bias
  - Hallucincation
  - Privacy
  - Regulation
  - Domain-specific adaptation
- Research Frontiers (30 min)
  - Prompt programming
  - Knowledge distillation
  - Multi-modal learning
  - Causal discovery

4 RELATED TUTORIALS

To our knowledge, there has not been a similar tutorial. The following tutorials are most related and we discuss the main differences.

[ACL-17/AAAI-18] Machine Reading for Precision Medicine One tutor, HP, co-presented this tutorial in 2017/2018. Since then, NLP has undergone rapid progress, with the emergence of LLMs and consolidation of AI methods. This tutorial will cover these exciting advances. It will also go beyond NLP and machine reading to review advances in multi-modal and causal learning for health.

[KDD 2021] Artificial Intelligence for Drug Discovery by Profs. Jian Tang, Fei Wang and Feixiong Chen covers very related topics regarding AI applications in drug discovery. However, since then, there have been major advances in self-supervised learning in general, and in LLMs. We would like to introduce such recent progress to the participants. Our proposed tutorial will also go beyond drug discovery and review the entire drug development life cycle, including clinical trials and post-market studies.

[NeurIPS 2022] Advances in NLP and their Applications to Healthcare] by Prof. Ndapa Nakashole focuses on the use of external memory to facilitate reasoning in Transformer architectures, drawing on applications in the healthcare domain for example. By contrast we orient around the broader scope of precision health including research and discovery.

[ISMB 2022] Towards Precision Medicine with Graph Representation Learning [1] by Michelle M. Li and Prof. Marinka Zitnik also orients around the area of precision health, but focuses on methods in graph neural networks (GNN) rather than LLMs.

5 IMPORTANT REFERENCES

We plan to cover key prior work, including: LLMs [2, 3, 8, 12], biomedical applications [5, 7, 9, 10, 14], and research frontiers [4, 6, 11, 13, 15, 16].

6 POTENTIAL SOCIETAL IMPACT

Our expectation is that after attending our tutorial, all participants will be familiar with the recent progress in LLMs and their applications in precision health. Computer scientists will be equipped with knowledge about precision health, which can empower them to help advance AI methods for precision health. Biomedical practitioners will learn about the power and growth opportunities in cutting-edge AI methods such as large language models, which can empower them to develop meaningful applications. Moreover, we hope the tutorial can help bridge the two communities and foster interdisciplinary collaboration.

By supercharging progress in precision health, we can attain great societal impact with LLMs. For care delivery, instead of following the traditional "one-size-fits-all" paradigm, patients can be prescribed with much more effective treatment tailored made for their conditions. For biomedical discovery, LLMs can empower researchers to assimilate a vast amount of biomedical knowledge and real-world data, thus greatly accelerating hypothesis generation and exploration. For health equity and access, by democratizing health knowledge work, LLMs can empower clinical practitioners to scale sophisticated decision making with humans in the loop, thus vastly broadening the access of high-quality health care. We also discuss significant application challenges such as bias, hallucination, cost, privacy, regulation.

REFERENCES

[1] Michelle M Li, Kexin Huang, and Marinka Zitnik. 2022. Graph Representation Learning in Biomedicine and Healthcare. *Nature Biomedical Engineering* (2022).
[2] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
[3] Bubeck *et al.* 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL]
[4] Grégoire Mialon *et al.* 2023. Augmented Language Models: a Survey. arXiv:2302.07842 [cs.CL]
[5] Harsha Nori *et al.* 2023. Capabilities of GPT-4 on Medical Challenge Problems. (March 2023). https://www.microsoft.com/en-us/research/publication/capabilities-of-gpt-4-on-medical-challenge-problems/
[6] Jason Wei *et al.* 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903 [cs.CL]
[7] Karan Singhal *et al.* 2022. Large Language Models Encode Clinical Knowledge. arXiv:2212.13138 [cs.CL]
[8] Long Ouyang *et al.* 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL]
[9] Renqian Lu *et al.* 2022. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics* 23, 6 (09 2022). https://doi.org/10.1093/bib/bbac409 arXiv:https://academic.oup.com/bib/article-pdf/23/6/bbac409/47144271/bbac409.pdf bbac409.
[10] Ross Taylor *et al.* 2022. Galactica: A Large Language Model for Science. arXiv:2211.09085 [cs.CL]
[11] Sheng Zhang *et al.* 2023. Large-Scale Domain-Specific Pretraining for Biomedical Vision-Language Processing. arXiv:2303.00915 [cs.CV]
[12] Tom B. Brown *et al.* 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]
[13] Xuezhi Wang *et al.* 2023. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv:2203.11171 [cs.CL]
[14] Yu Gu *et al.* 2021. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthcare* 3, 1, Article 2 (oct 2021), 23 pages. https://doi.org/10.1145/3458754
[15] Yizhong Wang *et al.* 2022. Self-Instruct: Aligning Language Model with Self Generated Instructions. arXiv:2212.10560 [cs.CL]
[16] Yongchao Zhou *et al.* 2023. Large Language Models Are Human-Level Prompt Engineers. arXiv:2211.01910 [cs.LG]