# AIM-FM: Advancements In Medical Foundation Models: Explainability, Robustness, Security, and Beyond

## Abstract

Towards next-generation medical analysis: Unlock the potential of medical foundation models for more explainable, robust, secure diagnosis solutions

Website: `https://wymancv.github.io/AIM-FM-NeurIPS-Workshop/`

## Main Proposal

**Introduction** There have been notable advancements in large foundation models (FMs) [9, 8, 10, 14, 15, 12], which exhibit generalizable language understanding, visual recognition, and audio comprehension capabilities. These advancements highlight the potential of personalized AI assistants in efficiently assisting with daily tasks, ultimately enhancing human life.

Healthcare is one of the most crucial industries touching every individual. Yet, due to large populations and limited medical professionals, it faces significant challenges, including the high cost and low doctor-to-population ratio [1, 3, 2, 5]. This shortage is more pronounced in rural and developing regions, where access to qualified doctors is severely limited, exacerbating health disparities and preventing timely treatment for common and complex conditions alike. *Hence, there is a critical need to develop effective, affordable, and professional AI-driven medical assistants.*

Despite the great success in general domains [6, 4, 16], FMs struggle in specific domains requiring strict professional qualifications, such as healthcare [13, 11, 7], which has high sensitivity and security risk. In light of the growing healthcare demands, this workshop aims to explore the potential of Medical Foundation Models (MFMs) in smart medical assistance, thereby improving patient outcomes and streamlining clinical workflows. Considering the primary clinical needs, we emphasize the explainability, robustness, and security of the large-scale multimodal medical assistant, pushing forward its reliability and trustworthiness. *By bringing together expertise in diverse fields, we hope to bridge the gap between industry and academia regarding precision medicine, highlighting clinical requirements, inherent concerns, and AI solutions.* Through this cooperative endeavor, we aim to unlock the potential of MFMs, striving for groundbreaking advancements in healthcare.

**Objectives** This workshop aims to unleash the significant potential of cutting-edge MFMs in enhancing medical procedures and healthcare, which can be detailed in the following aspects:

- To promote the exploration of multimodal models and data for boosting automatic medical applications and clinical decisions.
- To provide a platform for interdisciplinary collaboration among computer scientists, clinical physicians, and industry professionals, bridging the gap between industry and academia.
- To explore real-world clinical demands, ethical concerns, and advanced technology progress, lightening future research directions in the field of intelligent medicine.
- To showcase the potential of cutting-edge research, development, and adaptation of MFMs to a wide range of medical applications, promoting the fusion of technology and applications.
- To deepen the understanding of how large models can facilitate intuitive control and collaboration for clinical purposes.

Preprint. Under review.

- To encourage broad discussions on the current challenges and potential solutions in integrating AI in the healthcare sector.
- To facilitate the fairness and security of the research on medical AI and intelligent healthcare.
- To improve the efficiency and effectiveness of workflows on medical applications and clinical scenarios.

With the support of the NeurIPS community, we hope to make a significant step forward in next-generation intelligent healthcare. By unlocking the capabilities of MFMs in medical field, the accessibility and effectiveness of healthcare can be enhanced, thereby improving surgical and treatment outcomes. This effort attempts to revolutionize healthcare globally, delivering insights into patient-friendly healthcare and doctor-friendly diagnostic procedures.

**Encouraged Topics.** This workshop offers an interdisciplinary platform for exceptional researchers to present their latest progress in MFMs regarding healthcare support. Attendees will discuss technological advancements and delve into MFMs-driven diagnostic systems in the healthcare sector. Key topics of interest for the workshop may cover, but are not limited to, the following:

- **Medical MFMs.** Developing large-scale medical foundation models for medical applications, including diagnosis, prognosis, treatment, surgical assistance, etc.
- **Explainable MFMs.** Attempting to open the black box of MFMs in medical decision-making, ensuring transparency and interpretability in diagnostic outcomes.
- **Robust Diagnosis.** Exploring and enhancing the robustness of MFMs in diverse and challenging medical scenarios, including scarcity/misalignment of medical data, parameter-efficient tuning, validation techniques, etc.
- **Patient Privacy.** Strategies for data/model privacy in tuning and testing MFMs, including federated learning, data encryption, machine unlearning, uncustomization, etc.
- **Human-AI Interaction.** Studying the interaction dynamics to enhance the collaboration between healthcare professionals/patients and AI, e.g., prompting engineering, feedback refining, system designing, etc.
- **Multimodal Learning.** Research on effectively using heterogeneous medical data for training and addressing multi-modal challenges, such as modality misalignment and missing.
- **Generative Model for Healthcare.** Aiming to develop generative models for producing multimodal data for healthcare, including the generation of medical images, videos, reports, biology structures, etc.
- **Efficient MFMs.** Research on efficient learning of MFMs in medical assistants, e.g., data efficiency, annotation efficiency, small foundation model, etc.
- **Agent for Healthcare.** Towards the applications of AI agent systems in healthcare, including diagnosis, prognosis, surgical assistance, telehealth, etc.
- **Fairness in MFMs.** Targeting to develop fair multimodal models in healthcare and addressing potential bias from data, model, annotation, evaluation, etc.

**Attendence** Considering that NeurIPS has seen a significant increase in general attendance, the expected number of attendees and other related events are detailed as follows. (1) 150-200 in-person attendees. (2) 200-500 online attendees. (3) 80-100 paper submissions. (4) 20-25 posters.

**Tentative List of Invited Speakers** The workshop will invite the following in-person keynote speakers with profound experience in medical AI and multimodal learning.

- **Prof. Tianming Liu** (University of Georgia) (Confirmed)
  Email: lei@stanford.edu
- **Prof. James Zou** (Stanford University) (Confirmed)
  Email: jamesz@stanford.edu
- **Prof. Keane Pearse** (University College London) (Confirmed)
  Email: p.keane@ucl.ac.uk

- **Prof. Faisal Mahmood** (Harvard University) (Confirmed)
  Email: faisalmahmood@bwh.harvard.edu
- **Prof. Sheng Wang** (University of Washington) (Confirmed)
  Email: swang@cs.washington.edu

**Diversity Statement** We have made a great effort to ensure the diversity of organizers and keynote speakers, encompassing various aspects such as race, gender, expertise, and research areas. We are resolute in promoting inclusivity by actively inviting participants from various backgrounds, striving for gender parity, cultural richness, and representation of underrepresented communities among our speakers, committee, and workshop contributors. More details are listed as follows.

***Organizers.*** We have assembled a diverse team to ensure a broad range of perspectives and experiences. **(a) Racial Diversity.** Our organizers come from different racial and ethnic backgrounds, including Asian, European, and North American. This diversity in race and ethnicity brings a richness of cultural viewpoints and experiences to the workshop organization. **(b) Gender Diversity.** Our organizer team comprises an equal balance of male and female researchers. We believe this gender parity fosters a more inclusive and well-rounded approach to addressing the challenges in medical foundation models. **(c) Area Diversity.** Our team comprises biologists with expertise in biomedicine and computer scientists focused on developing general artificial intelligence. In addition, we have research specialists dedicated to advancing AI-assisted medical applications and clinical practice. **(d) Expertise Diversity.** Our organizers cover a range of research fields, such as clinical diagnostics, pathology analysis, medical imaging analysis, genomics, drug discovery, and robotic surgery. This diverse expertise guarantees that the workshop addresses various medical needs and concerns.

***Keynote Speakers.*** We have curated a diverse lineup of keynote speakers to ensure a broad representation of perspectives and experiences. **(a) Racial Diversity.** Our speakers come from various racial and ethnic backgrounds, including Asian, American, European, and Indian, reflecting the global nature of the medical research community. **(b) Expertise Diversity.** Our speakers represent a wide range of research areas within the medical domain, including clinical medicine, computational pathology, computational biology, medical image analysis, biomarker discovery, etc. This diversity of expertise ensures that the workshop covers a comprehensive range of topics and applications related to medical foundation models. **(c) Institution Diversity.** In addition to researchers from academic institutions, we have invited speakers from leading industry companies, such as Google, to provide insights into the practical applications and challenges of deploying medical foundation models in real-world settings. This can help connect theoretical research with practical application, ensuring a thorough grasp of the progress made and challenges faced in this cutting-edge field.

By fostering diversity among our organizers and speakers, we aim to create an inclusive and enriching environment that encourages diverse perspectives, promotes cross-cultural understanding, and drives innovation in the field of medical foundation models.

**Earlier Versions** As one of the emerging topics, no prior workshops have specifically addressed the topic of large-scale medical foundation models, emphasizing the novelty and growing interest in this interdisciplinary field. As a pioneering area, it holds enormous potential in revolutionizing medical practices, particularly with the rapid progress in large-scale multimodal learning.

**Technical Needs** No other special technical needs.

**Other Comments** This workshop aims to delve into the profound impact of large-scale multimodal learning on healthcare, particularly in the context of medical assistance. Our primary goal is to foster the development of intelligent, adaptive, and collaborative medical assistants that can learn from vast amounts of multimodal data. By bringing together diverse researchers worldwide, we hope to enhance the precision, safety, and efficiency of MFMs-based medical diagnosis. We emphasize the critical aspects of explainability, robustness, and security in medical AI assistance, as they are crucial for gaining trust and ensuring the ethical use of AI in healthcare.

## Organization Information

**Organizer List** The organizer list is as follows.

- **Prof. Yixuan Yuan** (The Chinese University of Hong Kong)
  Email:yxyuan@ee.cuhk.edu.hk
  Yuan's research interests include medical image analysis, deep learning in healthcare, abnormality detection, and multimodal foundation models, which cover the contexts of the brain, endoscopy, dermatoscopy, ophthalmoscope, and dentistry. She has published over 200 papers in top journals and conferences in medical AI fields.
- **Prof. Qin Yao** (University of California, Santa Barbara)
  Email: yaoqin@ucsb.edu
  Yao's research focuses on the robustness of machine learning, including adversarial robustness and out-of-distribution generalization. She is also interested in developing general ML algorithms and applying them to computer vision, NLP, and healthcare applications, especially diabetes.
- **Prof. Xiang Li** (Massachusetts General Hospital & Harvard Medical School)
  Email: xiangli.shaun@gmail.com
  Li's primary research interest is medical data analysis, focusing on clinical data streamlining, big data frameworks, multimodal multi-scale image fusion, and foundation models.
- **Prof. Ying Wei** (Nanyang Technological University)
  Email: ying.wei@ntu.edu.sg
  Wei's research focuses on developing algorithms for general machine intelligence through knowledge transfer and compositionality. This enables the adaptation of previous learning in large models to new tasks with minimal supervision quickly and the application of techniques to real-world problems with limited data, such as drug discovery.
- **Prof. Bulat Ibragimov** (University of Copenhagen)
  Email: bulat@di.ku.dk
  Ibragimov's research interests include machine learning in medicine, computer-aided diagnosis, medical image analysis, and human-AI interaction. He received various awards, including the Novo Nordisk Award for Young Data Science Investigator.
- **Prof. Linda Petzold** (University of California, Santa Barbara)
  Email: petzold@engineering.ucsb.edu
  Her research is focused on modeling, analysis, simulation and software, applied to multiscale, networked systems in biology, materials and social networks. Her research group has been developing advanced algorithms for discrete stochastic simulation of systems where the fate of a few key molecules can make a big difference to important outcomes.

**Related Experience** The organizers have extensive experience hosting related events, including:

- EARTH: Embodied AI and Robotics for HealTHcare **WORKSHOP** at Medical Image Computing and Computer-Assisted Intervention (**MICCAI**) conference 2024
- The Critical View of Safety **CHALLENGE** at **MICCAI** 2024
- Women in Machine Learning **WORKSHOP** at International Conference on Machine Learning (**ICML**) 2021-2024
- Advancements in Foundation Models for Medical Imaging **SPECIAL ISSUE** on IEEE Transactions on Medical Imaging (**TMI**) 2024
- Advancements in Foundation Models **SPECIAL ISSUE** on IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**) 2024
- R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Foundation Models **WORKSHOP** at Neural Information Processing Systems (**NeurIPS**) 2023
- Medical Image Learning with Limited and Noisy Data **WORKSHOP** at **MICCAI** 2023
- Machine Learning in Medical Imaging **WORKSHOP** at **MICCAI** 2023

- Medical Large Models **SYMPOSIUM** at International Joint Conference on Artificial Intelligence (**IJCAI**) 2023
- Mining and Learning from Time Series (MILETS) **WORKSHOP** at ACM Special Interest Group on Knowledge Discovery and Data Mining (**SIGKDD**) 2022
- Pre-training: Perspectives, Pitfalls, and Paths Forward **WORKSHOP** at **ICML** 2022
- Vision Meets Drones **CHALLENGE** at International Conference on Computer Vision (**ICCV**) 2021
- Meta-Learning **WORKSHOP** at **NeurIPS** 2021
- Deep Generative Models **WORKSHOP** at **MICCAI** 2021

Given the shared expertise and focused interests among NeurIPS attendees, we are optimistic about attracting sufficient participants to engage in thought-provoking discussions of this research field.

**Program Committee**

- **Dr. Zhen Chen** (Centre for Artificial Intelligence and Robotics; zhen.chen@cair-cas.org.hk)
- **Dr. Xiaoqing Guo** (University of Oxford; xiaoqing.guo@eng.ox.ac.uk)
- **Dr. Xiaohan Xing** (Stanford University; xhxing@stanford.edu)
- **Dr. Wuyang Li** (The Chinese University of Hong Kong; wuyangli@cuhk.edu.hk)
- **Mr. Qiushi Yang** (City University of Hong Kong; qsyang2-c@my.cityu.edu.hk)
- **Dr. Yu Jiang** (The Chinese University of Hong Kong; yujiang@cuhk.edu.hk)
- **Dr. Chen Yang** (The Chinese University of Hong Kong; chenyang001@cuhk.edu.hk)
- **Dr. Pengyu Wang** (The Chinese University of Hong Kong; pengyuwang@cuhk.edu.hk)
- **Dr. Zhihao Peng** (The Chinese University of Hong Kong; zhihaopeng@cuhk.edu.hk)
- **Mr. Fan Bai** (The Chinese University of Hong Kong; fanbai@link.cuhk.edu.hk)
- **Prof. Bo Zhao** (The University of Edinburgh; bozhaonanjing@gmail.com)
- **Mr. Yuxin Du** (Beijing Academy of Artificial Intelligence; yuxindu444@gmail.com)

## References

[1] Z. Huang, F. Bianchi, M. Yuksekgonul, T. J. Montine, and J. Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023.

[2] C. Li, C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon, and J. Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems*, 36, 2024.

[3] Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, and Y. Zhang. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. *Cureus*, 15(6), 2023.

[4] H. Liu, C. Li, Q. Wu, and Y. J. Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024.

[5] J. Liu, Y. Zhang, J.-N. Chen, J. Xiao, Y. Lu, B. A Landman, Y. Yuan, A. Yuille, Y. Tang, and Z. Zhou. Clip-driven universal model for organ segmentation and tumor detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21152–21164, 2023.

[6] S. Liu, H. Cheng, H. Liu, H. Zhang, F. Li, T. Ren, X. Zou, J. Yang, H. Su, J. Zhu, et al. Llava-plus: Learning to use tools for creating multimodal agents. *arXiv preprint arXiv:2311.05437*, 2023.

[7] Z. Liu, Y. Li, P. Shu, A. Zhong, L. Yang, C. Ju, et al. Radiology-llama2: best-in-class large language model for radiology. arxiv. *Preprint posted online*, 29, 2023.

[8] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

[9] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023.

[10] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

[11] G. Wang, G. Yang, Z. Du, L. Fan, and X. Li. Clinicalgpt: large language models finetuned with diverse medical data and comprehensive evaluation. *arXiv preprint arXiv:2306.09968*, 2023.

[12] J. Wang, Z. Liu, L. Zhao, Z. Wu, C. Ma, S. Yu, H. Dai, Q. Yang, Y. Liu, S. Zhang, et al. Review of large vision models and visual prompt engineering. *Meta-Radiology*, page 100047, 2023.

[13] Z. Wang, Z. Wu, D. Agarwal, and J. Sun. Medclip: Contrastive learning from unpaired medical images and text. *arXiv preprint arXiv:2210.10163*, 2022.

[14] C. Wu, W. Lin, X. Zhang, Y. Zhang, W. Xie, and Y. Wang. Pmc-llama: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association*, page ocae045, 2024.

[15] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, et al. A foundation model for generalizable disease detection from retinal images. *Nature*, 622(7981):156–163, 2023.

[16] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023.