

## Ethan Goh, MD, MS

Senior Research Engineer, Computational Medicine

### Publications

---

#### PUBLICATIONS

- **Asking the Right Questions: Benchmarking Large Language Models in the Development of Clinical Consultation Templates.** *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*  
McCoy, L. G., Wu, D., Khemani, S., Maharaj, S. K., Pahwa, A., Rosengaus, L., Giang, L., Jee, O., Goh, E., Haredasht, F. N., Chopra, K., Wu, D. J., Conteh, et al  
2026; 31: 400-416
- **Automated Evaluation of Large Language Model Response Concordance with Human Specialist Responses on Physician-to-Physician eConsult Cases.** *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*  
Wu, D. J., Haredasht, F. N., Wu, D., Ravi, V., McCoy, L. G., Weng, Y., Chopra, K., Everett, S. S., Nageeb, G., Chen, W., Ma, S. P., Maharaj, S. K., Tran, et al  
2026; 31: 372-387
- **"I Double Checked It with My Own Knowledge:" Physician Perspectives on the Use of AI Chatbots for Clinical Decision-Making.** *Journal of general internal medicine*  
Kerman, H., Siden, R., Cool, J. A., Hom, J., Goh, E., Ahuja, N., Shieh, L., Heidenreich, P., Yang, D., Rodman, A., Chen, J. H., Holdsworth, L. M.  
2026
- **Holistic evaluation of large language models for medical tasks with MedHELM.** *Nature medicine*  
Bedi, S., Cui, H., Fuentes, M., Unell, A., Wornow, M., Banda, J. M., Kotecha, N., Keyes, T., Mai, Y., Oez, M., Qiu, H., Jain, S., Schettini, et al  
2026
- **Asking the Right Questions: Benchmarking Large Language Models in the Development of Clinical Consultation Templates** *Pacific Symposium of Biocomputing*  
Liam, M. G., David, W., Sarita, K., Saloni, M. K., et al  
2026: 400-416
- **A typology of physician input approaches to using AI chatbots for clinical decision-making.** *NPJ digital medicine*  
Siden, R., Kerman, H., Gallo, R. J., Cool, J. A., Hom, J., Goh, E., Ahuja, N., Heidenreich, P., Shieh, L., Yang, D., Chen, J. H., Rodman, A., Holdsworth, et al  
2025
- **A typology of physician input approaches to using AI chatbots for clinical decision-making: a mixed methods study.** *medRxiv : the preprint server for health sciences*  
Siden, R., Kerman, H., Gallo, R. J., Cool, J. A., Hom, J., Goh, E., Ahuja, N., Heidenreich, P., Shieh, L., Yang, D., Chen, J. H., Rodman, A., Holdsworth, et al  
2025
- **Systematic Exploration of Hospital Cost Variability: A Conformal Prediction-Based Outlier Detection Method for Electronic Health Records.** *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*  
Grolleau, F., Goh, E., Ma, S. P., Masterson, J., Ross, T., Milstein, A., Chen, J. H.  
2025; 2025: 187-195
- **From Tool to Teammate: A Randomized Controlled Trial of Clinician-AI Collaborative Workflows for Diagnosis.** *medRxiv : the preprint server for health sciences*  
Everett, S. S., Bunning, B. J., Jain, P., Lopez, I., Agarwal, A., Desai, M., Gallo, R., Goh, E., Kadiyala, V. B., Kanjee, Z., Koshy, J. M., Olson, A., Rodman, et al  
2025

- **Physician clinical decision modification and bias assessment in a randomized controlled trial of AI assistance.** *Communications medicine*  
Goh, E., Bunning, B., Khoong, E. C., Gallo, R. J., Milstein, A., Centola, D., Chen, J. H.  
2025; 5 (1): 59
- **GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial.** *Nature medicine*  
Goh, E., Gallo, R. J., Strong, E., Weng, Y., Kerman, H., Freed, J. A., Cool, J. A., Kanjee, Z., Lane, K. P., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, et al  
2025
- **Systematic Exploration of Hospital Cost Variability: A Conformal Prediction-Based Outlier Detection Method for Electronic Health Records.** *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*  
Grolleau, F., Goh, E., Ma, S. P., Masterson, J., Ross, T., Milstein, A., Chen, J. H.  
2025; 2025: 187-195
- **Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial.** *JAMA network open*  
Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, et al  
2024; 7 (10): e2440969
- **Large Language Model Influence on Management Reasoning: A Randomized Controlled Trial.** *medRxiv : the preprint server for health sciences*  
Goh, E., Gallo, R., Strong, E., Weng, Y., Kerman, H., Freed, J., Cool, J. A., Kanjee, Z., Lane, K. P., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, et al  
2024
- **PRELIMINARY DATA: A RANDOMIZED CONTROLLED TRIAL OF LARGE LANGUAGE MODEL ASSISTANCE WITH COMPLEX DIAGNOSTIC REASONING**  
Kerman, H., Rodman, A., Goh, E., Cool, J., Gallo, R., Hom, J., Strong, E., Chen, J.  
SPRINGER.2024: S828-S829
- **Right Patient, Right Specialist, Right Time: Retrieval Augmented Generation for Specialty Referral Routing.** *AMIA ... Annual Symposium proceedings. AMIA Symposium*  
Haredasht, F. N., Goh, E., Ravi, V., Ashtari, P., Jiang, Y., Yuldashev, N., Grolleau, F., Gallo, R. J., Shah, A., Hur, E., Chopra, K., Jee, O., Lee, et al  
2024; 2024: 443-450
- **MedAlign: A Clinician-Generated Dataset for Instruction Following with Electronic Medical Records.** *Proceedings of the ... AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence*  
Fleming, S. L., Lozano, A., Haberkorn, W. J., Jindal, J. A., Reis, E., Thapa, R., Blankemeier, L., Genkins, J. Z., Steinberg, E., Nayak, A., Patel, B., Chiang, C. C., Callahan, et al  
2024; 38 (20): 22021-22030
- **Influence of a Large Language Model on Diagnostic Reasoning: A Randomized Clinical Vignette Study.** *medRxiv : the preprint server for health sciences*  
Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, et al  
2024
- **MEDALIGN: A Clinician-Generated Dataset for Instruction Following with Electronic Medical Records**  
Fleming, S. L., Lozano, A., Haberkorn, W. J., Jindal, J. A., Reis, E., Thapa, R., Blankemeier, L., Genkins, J. Z., Steinberg, E., Nayak, A., Patel, B., Chiang, C., Callahan, et al  
edited by Wooldridge, M., Dy, J., Natarajan, S.  
ASSOC ADVANCEMENT ARTIFICIAL INTELLIGENCE.2024: 22021-22030
- **ChatGPT Influence on Medical Decision-Making, Bias, and Equity: A Randomized Study of Clinicians Evaluating Clinical Vignettes.** *medRxiv : the preprint server for health sciences*  
Goh, E., Bunning, B., Khoong, E., Gallo, R., Milstein, A., Centola, D., Chen, J. H.  
2023
- **Remote evaluation of NAVIFY Oncology Hub using clinical simulation**  
Halligan, J., Goh, E., Lo, E. N., Chabut, D., Prime, M., Ghafur, S.  
LIPPINCOTT WILLIAMS & WILKINS.2023