

Guanghui Qin

🌐 <https://gqin.me/>

✉ hiaoxui@gmail.com

🎓 Google Scholar

Education

Johns Hopkins University

Ph.D. in Computer Science

2019 – 2024

Peking University

B.S. in Physics & Computer Science

2015 – 2019

Experience

Microsoft

Washington, USA

- *Senior Researcher at Microsoft AI (Manager: Tristan Naumann)*

2026 – Present

Training **large multimodal models** for healthcare, partnering with Mayo Clinic.

- *Senior Researcher at Microsoft Research (Manager: Hoifung Poon)*

2024 – 2026

Researched in **multimodal** methods for **AI in biomed/science**, including: (1) a **vision model** to predict neurological outcomes from **brain MRI**; (2) a **reinforcement learning** framework that achieved state-of-the-art performance in chest X-ray **report generation**; and (3) a benchmark that evaluates **agents'** performance in healthcare with CLI.

- *Research Intern at Microsoft Research (Mentor: Corby Rosset)*

2023

Conducted research on **efficient methods** for long-context **LLMs**. Proposed a method to **compress** the context of LLaMA by up to 20× with minimal performance tradeoffs. The method improved **retrieval-augmented generation (RAG)**.

- *Research Intern at Microsoft Research (Mentor: Anthony Platanios)*

2022

Studied a new research problem in **user action prediction**. Built a dataset from GitHub with 20 million actions and implemented **GNN models** to predict interactions among users and repositories (e.g., mentions, pull requests, and issues).

Johns Hopkins University

Maryland, USA

- *Ph.D. Candidate (Advisor: Benjamin Van Durme)*

2019 – 2024

Conducted research in **natural language processing** and **machine learning**, focusing on **efficient methods for large language models**, **prompt tuning**, and **semantic parsing**.

- *Visiting Researcher (Mentor: Jason Eisner)*

2018

Studied **temporal event stream** modeling. Proposed a particle smoothing method for sampling events from a neural **Hawkes process**. The **stochastic process** could interact with a deductive temporal database such as **Datalog**.

Microsoft Research Asia

Beijing, China

- *Research Intern (Mentors: Jin-Ge Yao & Chin-Yew Lin)*

2017 – 2018

Proposed a Semi-HMM-based model for grounding natural language to **structured data** for use in **data-to-text generation**.

Skills

- Programming languages: Python, Java, and C/C++. Other languages: Shell, $\text{\LaTeX}$, and SQL.
- Natural Language Processing (PhD topic): Transformers, LLMs, information retrieval, and semantic parsing.
- Computer Vision: CNNs, (3D) ViTs, and medical image processing.
- Supervised fine-tuning and reinforcement learning for multimodal LLMs (verl, MS-SWIFT, etc.).
- Distributed training and performance optimization (runtime and memory).
- Experience processing real-world biomedical data, including DICOM/NIfTI files and luminescence images.
- Physics training as undergraduate armed me with critical and scientific thinking.
- Networking/Web/Databases: Founded WallessPKU and have been funding and maintaining it since 2017. This proxy service has helped more than 65,000 Chinese students circumvent internet censorship of the Great Firewall.

Academic Service

Served as an *area chair* for NeurIPS (2026). Served as a *reviewer* for NeurIPS (2019–2023 and 2025), ICLR (2019–2021 and 2023–2026), ICML (2020, 2021, and 2026), CVPR (2026), ACL (2021 and 2025), EMNLP (2019–2022 and 2026), NAACL (2024), AACL (2021), and AKBC (2020).

Selected Publications

- [GigaBrain: multimodal AI for modeling heterogeneous MRI sequences in precision neuro-oncology](#). **Guanghui Qin**, Kristina H. Young, Prakash Ambady, Jiachen Tu, Cliff Wong, Jeya Maria Jose Valanarasu, Yiyan Lin, Yongkang Li, Xiaodong Liu, Sheng Zhang, Tristan Naumann, Carlo Bifulco[†], Sheng Wang[†], Hoifung Poon[†]. *Under review; manuscript available upon request*. 2026.
- [Evaluating the robustness and readiness of large frontier models in health AI applications](#). Yu Gu^{*†}, Jingjing Fu^{*}, Xiaodong Liu, Jeya Maria Jose Valanarasu, Noel C.F. Codella, Reuben Tan, Qianchu Liu, Ying Jin, Sheng Zhang, Jinyu Wang, Rui Wang, Lei Song, **Guanghui Qin**, Naoto Usuyama, Cliff Wong, Hao Cheng, HoHin Lee, Praneeth Sanapathi, Sarah Hilado, Tristan Naumann, Javier Alvarez-Valle, Jiang Bian, Mu Wei, Khalil Malik, Lidong Zhou, Jianfeng Gao, Eric Horvitz, Matthew P. Lungren, Doug Burger, Eric Topol[†], Hoifung Poon[†], and Paul Vozila. In *Nature Medicine*. 2026.
- [Masked-Diffusion Autoencoders for 3D Medical Vision Representation Learning](#). Jiachen Tu^{*}, **Guanghui Qin**^{*}, Theodore Zhao, Jeya Maria Jose Valanarasu, Sheng Zhang, Tristan Naumann, Fan Lam, Sheng Wang, and Hoifung Poon. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2026.
- [X-Reasoner: Towards Generalizable Reasoning Across Modalities and Domains](#). Sheng Zhang^{*}, Qianchu Liu^{*}, **Guanghui Qin**^{*}, Timothy Ossowski, Aiden Gu, Ying Jin, Sid Kiblawi, Sam Preston, Mu Wei, Paul Vozila, Tristan Naumann, and Hoifung Poon. In *arXiv, technical report*. 2025.
- [Scaling medical imaging report generation with multimodal reinforcement learning](#). Qianchu Liu^{*}, Sheng Zhang^{*}, **Guanghui Qin**^{*}, Yu Gu, Ying Jin, Sam Preston, Yanbo Xu, Sid Kiblawi, Wen-wai Yim, Tim Ossowski, Tristan Naumann, Mu Wei[†], and Hoifung Poon[†]. In *arXiv, under review*. 2026.
- [Dodo: Dynamic Contextual Compression for Decoder-only LMs](#). **Guanghui Qin**, Corby Rosset, Ethan Chau, Nikhil Rao, and Benjamin Van Durme. In *Annual Meeting of the Association for Computational Linguistics (ACL, oral)*. 2024.
- [Ras suppression potentiates rear actomyosin contractility-driven cell polarization and migration](#). Yiyan Lin^{*}, Dhiman S Pal^{*†}, Parijat Banerjee, Tatsat Banerjee, **Guanghui Qin**, Yu Deng, Jane Borleis, Pablo A. Iglesias, and Peter N. Devreotes[†]. In *Nature Cell Biology*. 2024.
- [Nugget: Neural Agglomerative Embeddings of Text](#). **Guanghui Qin** and Benjamin Van Durme. In *International Conference on Machine Learning (ICML)*. 2023.
- [The NLP Task Effectiveness of Long-Range Transformers](#). **Guanghui Qin**, Yukun Feng, and Benjamin Van Durme. In *European Chapter of the Association for Computational Linguistics (EACL, oral)*. 2023.
- [Learning How to Ask: Querying LMs with Mixtures of Soft Prompts](#). **Guanghui Qin** and Jason Eisner. In *North American Chapter of the Association for Computational Linguistics (NAACL)*. 2021. **Best Short Paper Award**
- [LOME: Large Ontology Multilingual Extraction](#). Patrick Xia^{*}, **Guanghui Qin**^{*}, Siddharth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron S. White, and Benjamin Van Durme. In *European Chapter of the Association for Computational Linguistics (EACL)*. 2021.
- [Neural Datalog Through Time: Informed Temporal Modeling via Logical Specification](#). Hongyuan Mei, **Guanghui Qin**, Minjie Xu, and Jason Eisner. In *International Conference on Machine Learning (ICML, oral)*. 2020.
- [Imputing Missing Events in Continuous-Time Event Streams](#). Hongyuan Mei, **Guanghui Qin**, and Jason Eisner. In *International Conference on Machine Learning (ICML, oral)*. 2019.
- [Learning Latent Semantic Annotations for Grounding Natural Language to Structured Data](#). **Guanghui Qin**, Jin-Ge Yao, Xuening Wang, Jinpeng Wang, and Chin-Yew Lin. In *Empirical Methods in Natural Language Processing (EMNLP, oral)*. 2018.

* indicates equal contribution; † indicates corresponding author.

Last updated on August 17, 2026.