

# Guanghui Qin

🌐 <https://gqin.me>    ✉ [gqin2@jhu.edu](mailto:gqin2@jhu.edu)

## EDUCATION

---

### Johns Hopkins University

*Ph.D. in Computer Science* (Advisor: Benjamin Van Durme)

Maryland, US

Aug 2019 – Summer 2024 (expected)

### Peking University

*B.S. in Physics & Computer Science*

Beijing, China

Sept 2015 – Jun 2019

## RESEARCH EXPERIENCE

---

### Microsoft Research Lab (MSR)

*Research Intern* (Mentor: Corby Rosset)

Washington, US

May 2023 – Aug 2023

- o Proposed Nugget2D, a *context compression* method for *large language models*.
- o Applied Nugget2D to downstream tasks, achieving a *20x compression rate* with minimal performance tradeoff.

### Semantic Machines, Microsoft

*Research Intern* (Mentor: Anthony Platanios)

Remote, US

May 2022 – Aug 2022

- o Studied a new research problem for *user action predictions* and built a dataset from GitHub.
- o Implemented *graph neural networks* baseline to learn the references between texts and entities.

### Center for Language and Speech Processing, Johns Hopkins Univ.

*Visiting Researcher* (Mentor: Hongyuan Mei and Jason Eisner)

Maryland, US

Jun 2018 – Oct 2018

- o Worked on a *particle smoothing* solution for *neural Hawkes process*, a method for temporal event sequence modeling.
- o A framework that enables *neural Hawkes process* to interact with the database through *Datalog*.

### Microsoft Research, Asia (MSRA)

*Research Intern* (Mentor: Jin-Ge Yao and Chin-Yew Lin)

Beijing, China

Nov 2017 – Jun 2018

- o Proposed a *Semi-HMMs* based model for grounding natural language to structured data.
- o Implemented a demo to induce templates from the corpus for *data-to-text generation*.

### Institute of Computer Science and Technology, Peking Univ.

*Research Intern* (Mentor: Xiaojun Wan)

Beijing, China

Feb 2017 – Oct 2017

- o Built a dataset for *humorous dialog* with a baseline model for the generation of the joke punch lines.

## AWARDS & SCHOLARSHIPS

---

- o Best Short Paper Awardee in NAACL 2021      Association for Computational Linguistics (ACL), 2021
- o Outstanding Reviewer      Association for Computational Linguistics (ACL), 2019
- o May 4th Scholarship      Peking University, 2016 and 2018
- o Silver medalist      Chinese Physics Olympiad (CPhO), 2014

## SKILLS & NON-RESEARCH EXPERIENCE

---

- o Programming languages: Python, JAVA, Rust, and C/C++. Other languages: Tex, Shell script, SQL.
- o Machine learning framework: PyTorch, Lightning AI, and DeepSpeed.
- o Network engineering: Since 2017, I've been developing and funding WallessPKU, a non-profit anti-censorship project with more than 37k users, to help students to circumvent The Great Firewall in China.

## ACADEMIC SERVICE

---

I serve as a reviewer for the conferences of NeurIPS (2019 and 2020 as secondary; 2021 to 2023), ICLR (2019 and 2020 as secondary; 2021, 2023, and 2024), ICML (2020 and 2021), ACL (2021), EMNLP (2019 to 2022; *outstanding reviewer award* in 2019), AAAI (2021), and AKBC (2020 as secondary).

## PUBLICATIONS

---

- **Nugget2D: Dynamic Contextual Compression for Scaling Decoder-only Language Models.**  
Guanghui Qin, Corby Rosset, Ethan C. Chau, Nikhil Rao, Benjamin Van Durme. In *arXiv*. 2023.
- **Nugget: Neural Agglomerative Embeddings of Text.**  
Guanghui Qin and Benjamin Van Durme. In *Proceedings of the Conference on International Conference on Machine Learning (ICML)*. 2023.
- **Ras-mediated homeostatic control of front-back signaling dictates cell polarity.**  
Y Lin\*, DS Pal\*, P Banerjee, T Banerjee, **G Qin**, Y Deng, . . . , and P Devreotes. In *bioRxiv*. 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)*. 2023.
- **Learning How to Ask: Querying LMs with Mixtures of Soft Prompts.**  
Guanghui Qin and Jason Eisner. In *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. 2021. **Best Short Paper Award**
- **Everything Is All It Takes: A Multipronged Strategy for Zero-Shot Cross-Lingual IE.**  
M. Yarmohammadi, S. Wu, M. Marone, H. Xu, S. Ebner, **G. Qin**, Y. Chen, J. Guo, C. Harman, K. Murray, A.S. White, M. Dredze, and B. Van Durme. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2021.
- **LOME: Large Ontology Multilingual Extraction.**  
Patrick Xia\*, **Guanghui Qin\***, S. Vashishtha, Y. Chen, T. Chen, C. May, C. Harman, K. Rawlins, A.S. White, and B. Van Durme. In *Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL) (Demo)*. 2021.
- **Iterative Paraphrastic Augmentation with Discriminative Span-based Alignment.**  
Ryan Culkin, J Edward Hu, Elias Stengel-Eskin, **Guanghui Qin**, and Benjamin Van Durme. In *Transactions of the Association for Computational Linguistics (TACL)*. 2021.
- **Neural Datalog Through Time: Informed Temporal Modeling via Logical Specification.**  
Hongyuan Mei, **Guanghui Qin**, Minjie Xu, and Jason Eisner. In *Proceedings of the Conference on International Conference on Machine Learning (ICML)*. 2020.
- **CopyNext: Explicit Span Copying and Alignment in Sequence to Sequence Models.**  
Singh Abhinav, Patrick Xia, **Guanghui Qin**, Mahsa Yarmohammadi, and Benjamin Van Durme. In *Proceedings of Workshop on Structured Prediction for NLP*. 2020.
- **Imputing Missing Events in Continuous-Time Event Streams.**  
Hongyuan Mei, **Guanghui Qin**, and Jason Eisner. In *Proceedings of the Conference on International Conference on Machine Learning (ICML)*. 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 *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018.
- **Data2Text Studio: Automated Text Generation from Structured Data.**  
Longxu Dou, **Guanghui Qin**, Jinpeng Wang, Jin-Ge Yao, and Chin-Yew Lin. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018.

(\* indicates equal contribution)