

# Junheng Hao

---

CONTACT INFORMATION	Microsoft Research, Generative AI <i>E-mail:</i> junhenghao@microsoft.com <i>Mobile:</i> +1 (424)355-5950 <i>Office:</i> Microsoft Building #99, Office 2911, 14850 NE 36th St, Redmond, WA 98052	<a href="#">Homepage</a> <a href="#">LinkedIn</a> <a href="#">Google Scholar</a> <a href="#">Semantic Scholar</a> <a href="#">DBLP Profile</a>
CURRENT EMPLOYMENT	<b>Researcher, Microsoft</b> Microsoft Research, Generative AI (Redmond, WA)	Oct 2022 - Current
KEY RESEARCH AREAS	Large language model (LLM): data synthesis, training and customization, natural language processing	
EDUCATION	<b>University of California Los Angeles (UCLA), CA, USA</b> Ph.D. in Computer Science <b>Thesis:</b> Incorporating ontological information in knowledge graph learning and applications <b>Advisors:</b> <a href="#">Yizhou Sun</a> , <a href="#">Wei Wang</a>  <b>Tsinghua University</b> , Beijing, China B. Eng. in School of Information Science and Technology B. Sc (Econ, Minor) in School of Economics and Management	Sept 2022     May 2017
PROFESSIONAL EXPERIENCE (INDUSTRY)	<b>Research Intern</b> at Microsoft Research (MSR), Redmond, WA • <b>Mentors:</b> Chieh-Han Wu, Zhihong (Iris) Shen, <a href="#">Ye-Yi Wang</a> , <a href="#">Jennifer Neville</a>  <b>PhD Research Intern</b> at IBM Research AI, San Jose, CA • <b>Mentor:</b> Chuan Lei, Berthold Reinwald, <a href="#">Fatma Ozcan</a>  <b>Applied Scientist Intern/Student Researcher</b> at Amazon, Seattle, WA • <b>Mentors:</b> Tong Zhao, <a href="#">Luna Xin Dong</a> , <a href="#">Christos Faloutsos</a>  <b>Research Intern</b> at <b>NEC Lab America</b> , Princeton, NJ • <b>Mentors:</b> Lu-An Tang, Zhichun Li, <a href="#">Haifeng Chen</a>	June 2021 - Sept 2021  June 2020 - Sept 2020  June 2019 - Dec 2019  June 2018 - Sept 2018
SELECTED PUBLICATIONS & TECHNICAL REPORTS	<ul style="list-style-type: none"><li>[1] <i>Phi-4-mini Technical Report: Compact yet Powerful Multimodal Language Models via Mixture-of-LoRAs</i> Microsoft GenAI team. <a href="#">Microsoft Blog: One year of Phi: Small language models making big leaps in AI</a></li><li>[2] <i>Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone</i> Microsoft GenAI team. <a href="#">Microsoft Blog: Discover the New Multi-Lingual, High-Quality Phi-3.5 SLMs</a></li><li>[3] <i>SciAgent: A Tool-augmented LLM for Scientific Reasoning</i> Yubo Ma, <b>Junheng Hao</b>, Ruochen Xu, Shuohang Wang, Zhibin Gou, Liangming Pan, Yujiu Yang, Yixin Cao, Aixin Sun, Hany Hassan Awadalla, Weizhu Chen The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)</li><li>[4] <i>Language Models can be Logical Solvers</i> Jiazhan Feng, Ruochen Xu, <b>Junheng Hao</b>, Hiteshi Sharma, Dongyan Zhao 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)</li><li>[5] <i>Multi-source Inductive Graph Knowledge Transfer</i> <b>Junheng Hao</b>, Lu-An Tang, Yizhou Sun, Zhengzhang Chen, Haifeng Chen, Junghwan</li></ul>	

Rhee, Zhichun Li and Wei Wang. Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD). September 2022.

- [6] *Metadata-Induced Contrastive Learning for Zero-Shot Multi-Label Text Classification*  
Yu Zhang, Zhihong Shen, Chieh-Han Wu, Boya Xie, **Junheng Hao**, Ye-Yi Wang, Kuansan Wang and Jiawei Han. Proceedings of The Web Conference (WWW) 2022. April 2022.
- [7] *MEDTO: Medical Data to Ontology Matching using Hybrid Graph Neural Networks*  
**Junheng Hao**, Chuan Lei, Abdul Quamar, Vasilis Efthymiou, Fatma Ozcan, Yizhou Sun, Wei Wang. Proceedings of 27th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD, Applied Data Science Track). August 2021.
- [8] *P-Companion: Framework for Diversified Complementary Product Recommendation*  
**Junheng Hao**, Tong Zhao, Jin Li, Luna Xin Dong, Christos Faloutsos, Yizhou Sun, Wei Wang. Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM), Applied Research Track. October 2020. [\[Amazon Blog\]](#)
- [9] *Bio-JOIE: Joint Representation Learning of Biological Knowledge Bases*  
**Junheng Hao**, Chelsea J.-T. Ju, Muhao Chen, Yizhou Sun, Carlo Zaniolo, Wei Wang. Proceedings of The 11th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB 2020), September 2020. **Best Student Paper Award.**
- [10] *Universal Representation Learning of Knowledge Bases by Jointly Embedding Instances and Ontological Concepts*  
**Junheng Hao**, Muhao Chen, Wenchao Yu, Yizhou Sun, Wei Wang. Proceedings of 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD, Research Track). August 2019.

PATENT  
APPLICATIONS

- [11] *OntoGNN: Hybrid Graph Neural Networks for Ontology Matching.*  
Chuan Lei, **Junheng Hao**, Vasilis Efthymiou, Fatma Ozcan, Abdul Quamar. U.S. Patent Application (Sept. 2021)

ACADEMIC  
SERVICES

- **Associate Editor (Editorial Board):** JMIR
- **Conference Area Chair / Senior Program Committee:** ECML-PKDD, CIKM
- **Conference Program Committee & Reviewer:** NeurIPS, KDD, ICML, AAAI, WWW, ICLR, IJCAI, EMNLP, ACL Rolling Review, ICDE, SDM, WSDM, ECML-PKDD, CIKM.
- **Journal Reviewer:** TPAMI, TBD, TIST, TKDD
- **Conference Volunteer:** ICLR, KDD, EMNLP, NeurIPS.

INVITED TALKS

- May 2023: Knowledge Graph Conference  
Invited presentation: “*Ontology-aware Knowledge Graphs and Empowered Multidisciplinary Applications in Industry*”
- July 2022: Google Brain.  
Invited tech talk: *Recommendations on Documents, Products and More? A Knowledge Graph Approach*
- June 2022: Microsoft, Azure Cognitive Services Research Group.  
Invited tech talk: *Knowledge Augmented Applications: NLP, Bioinformatics and Recommendation*
- Dec 2021: Coupang, Ranking, Discovery and Personalization. Invited tech talk: *Knowledge Graphs Meets Product Recommendation: One Deep Learning Solution.*
- Aug 2021: Microsoft Search, Assistant and Intelligence (MSAI). Graph learning session: *Knowledge Graph with Ontology Learning and Applications.*

- July 2019: Amazon, Product Graph. Invited talk: *Representation Learning on Knowledge Graphs: Embedding, Logic Rules and Graph Neural Networks* (with Yizhou Sun).

#### HONORS AND AWARDS

- **Best Student Paper Award** (ACM BCB) 2020
- NAACL D&I Subsidy Award, ACL DEI Grant 2022
- SIGIR Student Travel Grant (CIKM) 2020, 2021
- Student Travel Award (KDD) 2019, 2020

#### TEACHING

- CSM146: Introduction to Machine Learning (Winter 2021)
- CS145: Introduction to Data Mining (Fall 2020, Fall 2018)
- CS32: Introduction to Computer Science II, Data Structures (Spring 2019, Winter 2019)

#### SKILLS

- **Programming:** Python (PyTorch, TensorFlow), C/C++, L<sup>A</sup>T<sub>E</sub>X, MATLAB, SQL
- **Language:** Mandarin (Native), English (Proficient), Spanish (Basic)