Junheng Hao

Contact Information	Email: junhenghao@microsoft.com Email (Personal): haojh.ucla@gmail.com Mobile: +1 (424)355-5950	Homepage LinkedIn Google Scholar	
Current Employment	Researcher / Member of Technical Staff, Microsoft Microsoft GenAI Research		
RESEARCH INTERESTS	 Large language model (Phi-3, GPT-40) Training: Pre-training, Post-training, RLHF, etc Systematic data strategies for LLM: Data selection, Synthetic data generation Customized LLM development: Domain-specific LLM, Reasoning/Coding LLM + Knowledge Graph (KG) LLM Benchmarking/Evaluation 		
Education	University of California Los Angeles (UCLA), CA, USA Ph.D. in Computer Science Thesis: Incorporating ontological information in knowledge gray Advisors: Yizhou Sun, Wei Wang	Sep 2022 oh learning and applications	
	Tsinghua University , Beijing, China B. Eng. in School of Information Science and Technology B. Sc (Econ, Minor) in School of Economics and Management	May 2017	
Professional Experience (Industry)	Researcher at Microsoft GenAI, Redmond, WAManagers: Young Jin Kim, Weizhu Chen	Oct 2022 - Current	
	Research Intern at Microsoft Research (MSR), Redmond, WAMentors: Chieh-Han Wu, Zhihong (Iris) Shen, Ye-Yi Wang, Jen	Jun 2021 - Sep 2021 nnifer Neville	
	PhD Research Intern at IBM Research AI, San Jose, CAMentor: Chuan Lei, Berthold Reinwald, Fatma Ozcan	Jun 2020 - Sep 2020	
	Applied Scientist Intern/Student Researcher at Amazon, Seattle,Mentors: Tong Zhao, Luna Xin Dong, Christos Faloutsos	WA Jun 2019 - Dec 2019	
	Research Intern at NEC Lab America, Princeton, NJMentors: Lu-An Tang, Zhichun Li, Haifeng Chen	Jun 2018 - Sep 2018	
SELECTED PUBLICATIONS & TECHNICAL REPORTS	 Phi-4-mini Technical Report: Compact yet Powerful Multimodal Language Models via Mixture-of-LoRAs Microsoft GenAI team. Microsoft Blog: One year of Phi: Small language models making big leaps in AI 		
	 [2] Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone Microsoft GenAI team. Microsoft Blog: Discover the New Multi-Lingual, High-Quality Phi-3.5 SLMs 		
	[3] SciAgent: A Tool-augmented LLM for Scientific Reasoning Yubo Ma, Junheng Hao, Ruochen Xu, Shuohang Wang, Zhibin Gou, Liangming Pan, Yujiu Yang, Yixin Cao, Aixin Sun, Hany Hassan Awadalla, Weizhu Chen. EMNLP 2024.		
	[4] Language Models can be Logical Solvers Jiazhan Feng, Ruochen Xu, Junheng Hao, Hiteshi Sharma, Dongyan Zhao. NAACL 2024		

	[5] Multi-source Inductive Graph Knowledge Transfer Junheng Hao, Lu-An Tang, Yizhou Sun, Zhengzhang Chen, Haifeng Chen, Junghwan Rhee, Zhichun Li and Wei Wang. ECML-PKDD 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. The Web Conference (WWW) 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. KDD 2021 (Applied Data Science Track)	
	[8] P-Companion: Framework for Diversified Complementary Product Recommendation Junheng Hao, Tong Zhao, Jin Li, Luna Xin Dong, Christos Faloutsos, Yizhou Sun, Wei Wang. CIKM 2020 (Applied Research Track)	
	[9] Bio-JOIE: Joint Representation Learning of Biological Knowledge Bases Junheng Hao, Chelsea JT. Ju, Muhao Chen, Yizhou Sun, Carlo Zaniolo, Wei Wang. ACM BCB 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. KDD 2019 (Research Track) 	
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 (Web Conference), ICLR, IJCAI, EMNLP, 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 (Google DeepMind) Invited tech talk: <i>Recommendations on Documents, Products and More? A Knowledge</i> <i>Graph Approach</i>	
	• Dec 2021: Coupang, Ranking, Discovery and Personalization. Invited tech talk: <i>Knowledge Graphs Meets Product Recommendation: One Deep Learning Solution</i> .	
	• July 2019: Amazon, Product Graph. Invited talk: <i>Representation Learning on Knowledge Graphs: Embedding, Logic Rules and Graph Neural Networks</i> (with Prof. Yizhou Sun).	
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++ Language: Mandarin (Native), English (Proficient) 	