

Tushar Khot



🏠 8017 128th Ave. NE
Kirkland, WA 98033

☎ 608.692.2174

tushar.v.khot@gmail.com
<https://tusharkhot.github.io/>

Profile

I am a researcher that specializes in building models and benchmarks for question answering, coding and multi-modal reasoning problems. I have published about 100 articles with total 22K citations as part of my research career and been involved in two product and model releases in my industry career.

Experience

Member of Technical Staff, Microsoft AI, Redmond, WA 2025-Present
Worked and led efforts spanning pre-training for VLM grounding, imbuing agentic capabilities for Microsoft Excel and building a helpful coding assistant (beyond just code editing). Contributed to the final MAI-Thinking-1 model and technical report.

Staff Research Scientist, Google DeepMind, Seattle, WA 2024-2025
Developing next generation of agents for Gemini that can use tools and perform multi-modal reasoning for browser use agents.

Lead Research Scientist, Allen Institute for AI, Seattle, WA 2023-2024

- Leading and advising projects on building agents that can assist researchers by automating end-to-end experiment execution and data-driven hypothesis discovery
- Directed projects on improving reasoning abilities of LLMs (e.g. adapting to model failures) and the impact of personalization on model bias
- Advised and mentored researchers on projects involving development of a tool-use benchmark, model distillation and instruction tuning with LLMs

Senior Research Scientist, Allen Institute for AI, Seattle, WA 2019-2022

- Developed new benchmarks for multi-hop reasoning and models to solve these problems
- Advised and mentored research interns and external collaborators across multiple projects including bias in NLP models and challenging multi-hop QA problems

Research Scientist, Allen Institute for AI, Seattle, WA 2014-2018
Developed structured approaches, in collaboration with other researchers and engineers, for question answering using *probabilistic models (MLNs)* and *Integer Linear Programming (ILP)*

Software Engineer-II/III, Google R&D Center, Bangalore, India 2006-2008
Launched Google Local Search for India (<http://local.google.co.in>)

Research Intern, Microsoft Research, Redmond, WA Fall 2010
Predicting user behavior based on their profiles

Education

University of Wisconsin-Madison, USA – MS, PhD 2008-2014
Computer Science, 2008-2014 (GPA: 3.96/4.00) **Minor:** Statistics and Management
Thesis: *Efficient Learning of Statistical Relational Models*

Advisor: Prof. Jude Shavlik **Co-Advisor:** Prof. Sriraam Natarajan

Worked on developing efficient structure learning methods for probabilistic relational models and information extraction from documents.

National Institute Of Technology (NIT), Tiruchirappalli, India – B.Tech. 2002-2006

Computer Science, 2008-2014 (GPA: 9.67/10.00; Gold Medalist)

Full Publication List

Google Scholar: <https://scholar.google.com/citations?user=8mkljgAAAAJ>

Semantic Scholar: <https://www.semanticscholar.org/author/Tushar-Khot/2236429>

Selected Publications

Harsh Trivedi, **Tushar Khot**, Mareike Hartmann, Ruskin Manku, Vinty Dong, Edward Li, Shashank Gupta, Ashish Sabharwal, Niranjan Balasubramanian

AppWorld: A Controllable World of Apps and People for Benchmarking Interactive Coding Agents. In **ACL 2024**. ✨**Best Resource Paper Award** ✨

Research Question: How can we benchmark agents for complex day-to-day tasks (e.g. *Buy the item in my partner's email*) in a rigorous & reproducible manner?

Ben Bogin, Kejuan Yang, Shashank Gupta, Kyle Richardson, Erin Bransom, Peter Clark, Ashish Sabharwal, **Tushar Khot**

SUPER: Evaluating Agents on Setting Up and Executing Tasks from Research Repositories. In

EMNLP 2024. ✨**Outstanding Paper Award** ✨

Research Question: Can LLMs be used to autonomously set up and reproduce results from research repositories?

Archiki Prasad, Alexander Koller, Mareike Hartmann, Peter Clark, Ashish Sabharwal, Mohit Bansal, **Tushar Khot**

ADaPT: As-Needed Decomposition and Planning with Language Models. In **NAACL (Findings) 2024**.

Research Question: Can we build LLMs-as-agents that can adapt to LLM capabilities or task complexity via recursive decomposition?

Skills

Languages: Python, Scala, C/C++, Java, Perl, MATLAB, Shell Scripting, Prolog, HTML

Technology: PyTorch, Transformers (Huggingface), OpenAI API, AllenNLP, DyNet, Microsoft Scope, Condor Grid Computing, Stanford NLP Toolkit, MapReduce, Big Table

LLMs: Pre-training, Post-training (SFT and RL), RL for agents and helpfulness, agent orchestration

Service

Served as Senior AE for ACL Rolling Review, AC for AAI, SPC for AAI and IJCAI

Served as reviewer and meta-reviewer for AI and NLP conferences (AAAI, IJCAI, ICML, NeurIPS, UAI, AISTATS, ACL, NAACL, EMNLP).