

Tushar Khot



736 N 83rd St.
Seattle, WA 98103

T 608.692.2174

tushark@allenai.org
<http://allenai.org/team/tushark/>

Profile

I am a Lead Research Scientist at AI2, specializing in machine reasoning problems such as question answering and decision-making with LLMs. My research spans various aspects of solving complex reasoning tasks — building benchmarks, baselines and general-purpose solutions (incl. symbolic, neural and agent-based approaches). I am looking for positions where I can leverage my 10+ years of research experience to have impact on real users.

Experience

- Lead Research Scientist, Allen Institute for AI, Seattle, WA 2023-Present
- 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
 - Collaborated with other researchers and engineers on the development of AI2's first instruction-tuned model: Tulu and base pre-trained LLM: OLMo
 - 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
- Software Intern, Amazon, Bangalore, India Summer 2005
- Fraud detection for A9 Search using weblog analysis

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)

Full Publication List

Google Scholar: <https://scholar.google.com/citations?user=8mkljgAAAAJ>
Semantic Scholar: <https://www.semanticscholar.org/author/Tushar-Khot/2236429>

Selected Publications

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?

Yao Fu, Hao Peng, Litu Ou, Ashish Sabharwal, **Tushar Khot**
Specializing Smaller Language Models towards Multi-Step Reasoning. In **ICML 2023.**
Research Question: Can we teach smaller models (e.g. Flan-T5-11B) to mimic complex reasoning from larger LLMs (e.g. GPT3) and at what cost?

Tushar Khot, Harsh Trivedi, Matthew Finlayson, Yao Fu, Kyle Richardson, Peter Clark, Ashish Sabharwal
Decomposed prompting: A Modular Approach for Solving Complex Tasks. In **ICLR 2023.**
Research Question: How can we compose smaller models and tools to solve complex tasks that can't be solved by a single model/tool?

Nikil Roashan Selvam, Sunipa Dev, Daniel Khashabi, **Tushar Khot**, Kai-Wei Chang
The Tail Wagging the Dog: Dataset Construction Biases of Social Bias Benchmarks. In **ACL 2023.** ✨**Outstanding Paper Award** ✨
Research Question: Do bias benchmarks have other non-social biases that can make them unreliable indicators of true model bias?

Service

Served as Senior AE for ACL Rolling Review, AC for AACL, SPC for AACL and IJCAI
Served as reviewer and meta-reviewer for AI and NLP conferences (AACL, IJCAI, ICML, NeurIPS, UAI, AISTATS, ACL, NAACL, EMNLP).

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