

Omar Khattab

✉ okhattab@mit.edu
🌐 omarkhattab.com

Education

- Sep 2019 – **Stanford University**, Stanford, CA
Jan 2025 **Ph.D. in Computer Science**, Advised by Christopher Potts & Matei Zaharia
Thesis: *Building more reliable and scalable AI systems with foundation model programming*
- Aug 2015 – **Carnegie Mellon University** in Qatar (CMU-Q), Doha, Qatar
May 2019 **B.S. in Computer Science** (Minor: Math), Advised by Mohammad Hammoud

Employment

- Jul 2025 – **Assistant Professor at MIT EECS.**
Affiliated with CSAIL and EECS AI+D & CS.
- Jun 2024 – **Research Scientist at Databricks.**
Jun 2025 Working on NLP, IR, and ML Systems.
- Jun–Sep 2022 **Research Ph.D. Intern at Apple** via a “PhD Scholar in AI/ML” fellowship.
Advised by Dr. Ruoming Pang, Distinguished Engineer at Apple AI/ML.

Selected Awards & Honors

- 2025–28 TIBCO Founders Career Development Professor, MIT School of Engineering.
2025 SIGIR 2025 “Best Paper Award”.
2025 “2025 Tartans on the Rise” Honoree, Carnegie Mellon.
2025 Mayfield | Divot AI List Honoree.
- 2022–24 “Apple PhD Scholar in AI/ML” Fellowship.
2021 “Top-5% of Stanford CS Course Assistants”.
- 2019–20 The Eltoukhy Family Graduate Fellowship. Stanford CS Department.
2019 “Alumni Award for Undergraduate Excellence” for Senior Thesis. CMU-Pittsburgh.
2019 “Outstanding Academic Achievement in Computer Science”. CMU-Qatar.
2019 “Best Project”, Annual Meeting of the Minds”. CMU-Qatar.
2018 “Qatar Campus Scholar”. CMU-Qatar.
- 2016–19 Full-Tuition Merit Scholarship. Qatar Foundation.
2015 “Top in the World” Outstanding Learner. Cambridge International Exams.

Papers & Manuscripts

* indicates co-first author

- ArXiv 2026 OBLIQ-Bench: Exposing Overlooked Bottlenecks in Modern Retrievers with Latent and Implicit Queries.
D Tchuindjo, D Shah, **O Khattab**.
- ArXiv 2026 Meta-Harness: End-to-End Optimization of Model Harnesses.
Y Lee, R Nair, Q Zhang, K Lee, **O Khattab**, C Finn.
- ArXiv 2025 Recursive Language Models.
A Zhang, T Kraska, **O Khattab**.
- ICLR 2026 GEPA: Reflective Prompt Evolution Can Outperform Reinforcement Learning (Oral).
LA Agrawal, S Tan, D Soylu, N Ziemis, R Khare, K Opsahl-Ong, A Singhvi, H Shandilya, MJ Ryan, M Jiang, C Potts, K Sen, AG Dimakis, I Stoica, D Klein, M Zaharia, **O Khattab**.
- ACM CAIS Reasoning-Intensive Regression.
2026 D Tchuindjo, **O Khattab**.
- ACM CAIS Composing Policy Gradients and Prompt Optimization for Language Model Programs.
2026 N Ziemis, D Soylu, LA Agrawal, I Miller, L Lai, C Qian, K Song, M Jiang, D Klein, M Zaharia, K D'Oosterlinck, C Potts, **O Khattab**.
- SIGIR 2025 WARP: An Efficient Engine for Multi-Vector Retrieval (Best Paper Award).
JL Scheerer, M Zaharia, C Potts, G Alonso, **O Khattab**.
- NeurIPS 2025 FreshStack: Building Realistic Benchmarks for Evaluating Retrieval on Technical Documents.
N Thakur, J Lin, S Havens, M Carbin, **O Khattab**, A Drozdov.
- EMNLP LangProBe: a Language Programs Benchmark.
- Findings 2025 S Tan, LA Agrawal, A Singhvi, L Lai, MJ Ryan, D Klein, **O Khattab**, K Sen, M Zaharia.
- ReNeuIR Drowning in Documents: Consequences of Scaling Reranker Inference.
2025 M Jacob, E Lindgren, M Zaharia, M Carbin, **O Khattab**, A Drozdov.
- ICLR 2025 Grounding by Trying: LLMs with Reinforcement Learning-Enhanced Retrieval.
S Hsu, **O Khattab**, C Finn, A Sharma.
- NAACL 2025 PAPILLON: PrivAcy Preservation from Internet-based and Local Language MOdel ENsembles.
L Siyan, VC Raghuram, **O Khattab**, J Hirschberg, Z Yu.

- ECIR 2025 ColBERT-serve: Efficient Multi-Stage Memory-Mapped Scoring.
K Huang, T Venkatesh, U Dingankar, A Mallia, D Campos, J Jiao, C Potts, M Zaharia, K Boahen, **O Khattab**, S Sarup, K Santhanam.
- ArXiv 2024 Prompts as Auto-Optimized Training Hyperparameters: Training Best-in-Class IR Models from Scratch with 10 Gold Labels.
J Xian, S Samuel, F Khoubssirat, R Pradeep, MA Sultan, R Florian, S Roukos, A Sil, C Potts, **O Khattab**.
- EMNLP 2024 Fine-tuning and prompt optimization: Two great steps that work better together.
D Soylu, C Potts, **O Khattab**.
- EMNLP 2024 Optimizing instructions and demonstrations for multi-stage language model programs.
K Opsahl-Ong, M Ryan, J Purtell, D Broman, C Potts, M Zaharia, **O Khattab**.
- EMNLP 2024 Problem-Oriented Segmentation and Retrieval: Case Study on Tutoring Conversations.
Findings R E Wang, P Wirawarn, K Lam, **O Khattab**, D Demszky.
- NAACL 2024 Assisting in Writing Wikipedia-like Articles From Scratch with Large Language Models.
Y Shao, Y Jiang, T Kanell, P Xu, **O Khattab**, M Lam.
- NAACL 2024 ARES: An Automated Evaluation Framework for Retrieval-Augmented Generation Systems.
J Saad-Falcon, **O Khattab**, C Potts, M Zaharia.
- EACL 2024 Backtracing: Retrieving the Cause of the Query.
Findings R Wang, P Wirawarn, **O Khattab**, N Goodman, D Demszky.
- ICLR 2024 DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines.
O Khattab, A Singhvi, P Maheshwari, Z Zhang, K Santhanam, S Vardhamanan, S Haq, A Sharma, T Joshi, H Moazam, H Miller, M Zaharia, C Potts. Spotlight paper: top 5% of submissions.
- Digital Image and Data Mining in Reticular Chemistry Using GPT-4V.
Discovery 2024 Z Zheng, Z He, **O Khattab**, N Rampal, M Zaharia, C Borgs, J T Chayes, O M Yaghi.
- ArXiv 2024 In-Context Learning for Extreme Multi-Label Classification.
K D'Oosterlinck, **O Khattab**, F Remy, T Demeester, C Develder, C Potts.
- KB Systems Building Efficient and Effective OpenQA Systems for Low-Resource Languages.
2024 E Budur, R Özçelik, D Soylu, **O Khattab**, T Güngör, C Potts.
- ArXiv 2023 DSPy Assertions: Computational Constraints for Self-Refining Language Model Pipelines.
A Singhvi, M Shetty, S Tan, C Potts, K Sen, M Zaharia, **O Khattab**.

- EMNLP 2023 UDAPDR: Unsupervised Domain Adaptation via LLM Prompting and Distillation of Rerankers.
J Saad-Falcon, **O Khattab**, K Santhanam, R Florian, M Franz, S Roukos, A Sil, M Sultan, C Potts.
- REML 2023 Resources and Evaluations for Multi-Distribution Dense Information Retrieval.
S Chatterjee, **O Khattab**, S Arora.
- ACL 2023 Moving Beyond Downstream Task Accuracy for Information Retrieval Benchmarking.
Findings K Santhanam, J Saad-Falcon, M Franz, **O Khattab**, A Sil, R Florian, S Roukos, A Sil, M Sultan, M Zaharia, C Potts.
- TMLR 2023 Holistic evaluation of language models.
P Liang, R Bommasani, T Lee, D Tsipras, D Soylu, ..., **O Khattab**, ..., Y Zhang, Y Koreeda. Transactions on Machine Learning Research (TMLR) 2023. This is a multi-component, 50-author project. I directed the Information Retrieval evaluation.
- ArXiv 2022 Demonstrate-Search-Predict: Composing retrieval and language models for knowledge-intensive NLP.
O Khattab, K Santhanam, XL Li, D Hall, P Liang, C Potts, M Zaharia. ArXiv preprint arXiv:2212.14024.
- CIKM 2022 PLAID: An Efficient Engine for Late Interaction Retrieval.
K Santhanam*, **O Khattab***, C Potts, M Zaharia. Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM).
- CIKM 2022 Introducing Neural Bag of Whole-Words with ColBERTer: Contextualized Late Interactions using Enhanced Reduction.
S Hofstätter, **O Khattab**, S Althammer, M Sertkan, A Hanbury. Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM).
- NAACL 2022 ColBERTv2: Effective and Efficient Retrieval via Lightweight Late Interaction.
K Santhanam*, **O Khattab***, J Saad-Falcon, C Potts, M Zaharia. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).
- ICLR 2022 Hindsight: Posterior-guided training of retrievers for improved open-ended generation.
A Paranjape, **O Khattab**, C Potts, M Zaharia, CD Manning. Proceedings of The Tenth International Conference on Learning Representations (ICLR).
- NeurIPS 2021 Baleen: Robust Multi-Hop Reasoning at Scale via Condensed Retrieval.
O Khattab, C Potts, M Zaharia. Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS'21). Spotlight paper: top 3% of submissions.

- ArXiv 2021 On the opportunities and risks of foundation models.
R Bommasani, DA Hudson, E Adeli, R Altman, S Arora, S von Arx, ..., **Omar Khattab**, ..., P Liang. ArXiv preprint arXiv:2108.07258. This is a multi-component, 114-author project. I co-authored the sections on Modeling and on Systems.
- SIGIR 2021 Learning passage impacts for inverted indexes.
A Mallia, **O Khattab**, T Suel, N Tonello. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Short paper.
- TACL 2021 Relevance-guided Supervision for OpenQA with ColBERT.
O Khattab, C Potts, M Zaharia.
- SIGIR 2020 ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT.
O Khattab, M Zaharia. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.
- SIGIR 2020 Finding the best of both worlds: Faster and more robust top-k document retrieval.
O Khattab, M Hammoud, T Elsayed. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.
- CIKM 2018 PolyHJ: A polymorphic main-memory hash join paradigm for multi-core machines.
O Khattab, M Hammoud, O Shekfeh. Proceedings of the 27th ACM International Conference on Information and Knowledge Management.
- VLDB 2018 LA3: A scalable link-and locality-aware linear algebra-based graph analytics system.
Y Ahmad, **O Khattab**, A Malik, A Musleh, M Hammoud, M Kutlu, M Shehata, T Elsayed. Proceedings of the VLDB Endowment 11 (8), 920-933.

Technical Blog Posts

- Blog 2026 Pedagogical RL: Teaching Models to Teach Themselves from Privileged Information.
S Chakraborty, N Ziems, F Huang, M Jiang, AS Bedi, **O Khattab**.
- Blog 2025 Recursive Language Models.
AL Zhang, **O Khattab**.
- Blog 2024 On Impactful AI Research.
O Khattab.
- BAIR 2024 The Shift from Models to Compound AI Systems.
M Zaharia, **O Khattab**, L Chen, JQ Davis, H Miller, C Potts, J Zou, M Carbin, J Frankle,
N Rao, A Ghodsi. Berkeley Artificial Intelligence Research blog.
- Two Sigma 2024 A Guide to Large Language Model Abstractions.
PY Zhong, H He, **O Khattab**, C Potts, M Zaharia, H Miller. Two Sigma Articles.
- SAIL 2021 Building Scalable, Explainable, and Adaptive NLP Models with Retrieval.
O Khattab, C Potts, M Zaharia. Stanford AI Lab (SAIL) blog.
- HAI 2021 A Moderate Proposal for Radically Better AI-powered Web Search.
O Khattab, C Potts, M Zaharia. Stanford HAI blog.

Selected Keynotes, Invited Talks, and Lectures

- 2026 Apr **ECIR LIR Workshop.** *Keynote*, First Workshop on Late Interaction and Multi-Vector Retrieval.
- 2025 Oct **University of Notre Dame.** CSE seminar, "Principles and Algorithms for Reliable AI Systems."
- 2025–26 **Carnegie Mellon University.** Guest lectures in Spring 2025, Fall 2025, and Spring 2026 on large language models and compound AI systems, 11-785 Introduction to Deep Learning.
- 2025 Fall **Carnegie Mellon University and MIT.** Guest lectures on models vs. systems, AI inference, and LLM programming.
- 2025 Jun **AI Engineer World's Fair.** "On Engineering AI Systems that Endure The Bitter Lesson."
- 2025 Jun **Databricks Data + AI Summit.** "DSPy 3.0 — and DSPy at Databricks," with Krista Opsahl-Ong.
- 2024 Oct **UC Berkeley.** Guest lecture, CS294/194-196 LLM Agents, "Compound AI Systems & the DSPy Framework."
- 2024 Aug **Princeton Language and Intelligence.** Speaker, Workshop on Useful and Reliable AI Agents.
- 2024 Jul **SIGIR 2024 LLM Day.** "DSPy: Self-Improving Language Programs."
- 2024 Apr **University of Waterloo, Purdue University, and Georgia Tech.** "Building More Reliable and Scalable AI Systems with Language Model Programming."
- 2024 Mar **CMU-Qatar, MIT, Caltech, and University of Maryland.** "Building More Reliable and Scalable AI Systems with Language Model Programming."

- 2024 Jan–Feb **Carnegie Mellon University and Yale.** “Building More Reliable and Scalable AI Systems with Language Model Programming.”
- 2023 Jul–Nov **Google X, Apple, UC Berkeley, Qatar University, ScaleByTheBay, Avey, and SIGIR REML.** Invited talks and lectures on DSPy.
- 2023 Jul **SIGIR ReNeuIR.** *Keynote*, “From BM25 to (Col)BERT to LLMs: Navigating a Maturing Efficiency-Effectiveness Tradeoff Space in Neural IR.”
- 2023 Jan–Apr **UC Berkeley, Databricks, Meta, Neeva, Netflix ML, Vectara, Google, Oracle Labs, Apple, and Samaya.** Invited research talks on Demonstrate–Search–Predict.
- 2022–2023 **UC Berkeley, CMU-Qatar, Apple, and Etsy.** Invited talks and guest lectures on ColBERT and ColBERTv2.
- 2020–2021 **Google Research IR Group, Google Research N2Formal, IBM Research, Stanford NLP, and University of Glasgow.** Invited talks on ColBERT, ColBERT-QA, and Baleen.

Selected Invited Podcasts and other Engagements

- 2026 Jan **a16z Podcasts.** Discussion with Martin Casado, “How Foundation Models Evolved: A PhD Journey Through AI’s Breakthrough Era.”
- 2025 Fall **ACM Queue Editorial Board.** Invited discussion leader on AI specification and learning.
- 2024–25 **DSPy community meetups.** Including “DSPy End-to-End”, “DSPy Boston”, and Bay Area DSPy meetups.
- 2022–24 **Neural Search Talks / Zeta Alpha, MLOps Community, Vertex, DataBrew, and Stanford CS224U Guest Podcasts.** Podcast episodes and interviews on DSPy, retrieval, and reliable AI systems.
- 2023 **SIGIR GenIR and LangChain.** Panels on language-model behavior and LM application development.
- 2023 **UC Berkeley SkyCamp, Cohere4AI, LlamaIndex, DemandBase, and MLOps Learners.** Tutorials and webinars on DSPy, retrieval, and language model programming.

Teaching

- Spring 2026 MIT 6.8610: Advanced Natural Language Processing (with Jacob Andreas & Chris Tanner)
- Fall 2025 MIT 6.7960: Deep Learning (with Sara Beery & Kaiming He)
- Spring 2022 Stanford CS224u: Natural Language Understanding (TA; Prof. Christopher Potts)
- Spring 2021 Stanford CS224u: Natural Language Understanding (TA; Prof. Christopher Potts)
- Spring 2019 CMU 15-210: Parallel & Sequential Data Structures & Algorithms (TA; Prof. Kemal Oflazer)
- Fall 2018 CMU 11-785: Introduction to Deep Learning (TA; Prof. Bhiksha Raj)
- Fall 2018 CMU 15-451: Algorithms Design & Analysis (TA; Prof. Christos Kapoutsis)
- Spring 2018 CMU 15-210: Parallel & Sequential Data Structures & Algorithms (TA; Prof. Kemal Oflazer)

Professional Service

- Organization 2026: ACM CAIS Program Co-Chair and Member of Steering Committee
2026: Organizer, 1st Late Interaction Workshop (LIR) at ECIR
- Review 2025: NAACL Senior Area Chair; ACL/ARR Area Chair; ICLR reviewer
2024: ACL ARR Senior Area Chair; ACL, ICLR, and ICML reviewer
2023: NeurIPS, ICLR, ACL ARR, and other venues
2021-22: TOIS, IP&M, and other venues
- Institutional 2026: MIT EECS Graduate Admissions Committee, Sub-Area Chair for AI Systems
2025: MIT EECS Graduate Admissions Committee
2023: Stanford CS PhD admissions in NLP
2022: CSLI Summer Internships
2021: CSLI Summer Internships
2020: Student-Applicant Support Program