

## KARL STRATOS

A seasoned AI researcher with deep expertise in machine learning and natural language processing ([technical notes](#)). I have extensive experience in both industry (big tech and startup) and academia (faculty at premier AI institutes) with a track record of publications ([Google Scholar](#), h-index 23). I have worked on a wide range of problems, which involved **pre-training and post-training large language models**, **retrieval-augmented generation**, and **unsupervised representation learning**. I am eager to contribute to transformative AI research and products that create real, positive impact.

### Employment

*Senior Member of Technical Staff*, **Essential AI**, San Francisco, CA (September 2024—)

- Research Lead at a full-stack AI startup founded by [Ashish Vaswani](#)
- Primary contributor to developing second-order optimizers that improved data efficiency for LLM pretraining by 10-15% over AdamW ([paper](#), [blog](#))
- Core contributor to developing an LLM agent that performed multi-step research on private documents to answer complex financial questions with citations

*Assistant Professor*, **Rutgers University**, New Brunswick, NJ (September 2019—August 2024)

- Tenure-track faculty at Rutgers, one of the highest-ranked public research universities
- Directed a group of ~10 PhD and master's students, developed state-of-the-art neural models for document retrieval, question answering, entity linking, coreference resolution.
- Published 15 papers at top-tier NLP/ML conferences, won the ICLR 2022 Spotlight, the [ACL Test-of-Time Paper Award \(2022\)](#), received the [Google Faculty Research Award \(2020\)](#).
- Taught courses on ML foundations and advanced NLP taken by 100s of undergraduate and graduate students each year.

*Consultant*, **NAVER U.Hub Inc.**, Remote (January 2022—August 2024)

- Advised a team of developers at Naver (the largest search portal in South Korea) in deploying AI services that faced > 20 million daily users.

*Consultant*, **Bloomberg L.P.**, Remote (June—December, 2020)

- Fixed a gradient bug in the existing implementations of Word2Vec ([paper](#)).
- Open-sourced a highly scalable and performant implementation ([code](#), processes 750G of text in a few days on a single 16-core machine).

*Research Assistant Professor*, **Toyota Technological Institute at Chicago**, Chicago, IL (June 2017—August 2019)

- Conducted cutting-edge research on representation learning that resulted in novel algorithms.
- Developed a new [theory](#) on fundamental limitations of estimating mutual information (joint work with [David McAllester](#), > 300 citations as of January 2025).
- Received the [Bloomberg Data Science Research Grant \(2018\)](#), NVIDIA GPU Grant (2017).

*Adjunct Assistant Professor*, **Columbia University**, New York, NY (January—May, 2017)

- Instructor for a graduate-level course on NLP (~100 students)

*Senior Research Scientist*, **Bloomberg L.P.**, New York, NY (August 2016—May 2017)

- Founding member of the Bloomberg AI team
- Primary contributor to the team's transition from statistical models to neural networks in providing company-wide NLP services.
- Proposed, trained, and deployed a character-level LSTM for named-entity recognition with an approximately 5–10% improvement in F1 over the existing feature-based CRF model.

*Research Assistant*, Department of Computer Science, **Columbia University**, New York, NY (September 2011—June 2016)

*Research Intern*, **Google Inc.**, New York, NY (June–August, 2014)

- Mentor: [Slav Petrov](#). Project: spectral methods for parsing

*Research Intern*, **Microsoft Research**, Boston, MA (June–August, 2013)

– Mentor: [Sham Kakade](#). Project: deep learning for sequence labeling

*Summer Workshop Participant*, **Johns Hopkins University Center for Language and Speech Processing**, Baltimore, MD (July–August, 2011)

– Mentors: [Alexander Berg](#) and [Tamara Berg](#). Project: generating image descriptions

*Research Intern*, **University of Rochester**, Rochester, NY (June–August, 2010)

– Mentor: [Lenhart Schubert](#). Project: logic-based inference engine

## Education

*Columbia University, New York, NY (September 2011—June 2016)*

**Ph.D. in Computer Science** (Advisor: [Michael Collins](#))

Thesis: [Spectral Methods for Natural Language Processing](#)

*University of Rochester, Rochester, NY (August 2008—May 2011)*

**Bachelor of Science in Computer Science**

**Bachelor of Arts in Mathematics**

**Minor in History**

## Selected Publications

Ishaan Shah\*, Anthony M. Polloreno\*, Karl Stratos\*, Philip Monk, Ashish Vaswani, and others (\* equal contribution). [Practical Efficiency of Muon for Pretraining](#). Essential AI 2025

Wenzheng Zhang, Wenyue Hua, and Karl Stratos. [EntQA: Entity Linking as Question Answering](#). ICLR 2022 (**Spotlight**)

Wenzheng Zhang and Karl Stratos. [Understanding Hard Negatives in Noise Contrastive Estimation](#). NAACL 2021

Karl Stratos and Sam Wiseman. [Learning Discrete Structured Representations by Adversarially Maximizing Mutual Information](#). ICML 2020

David McAllester and Karl Stratos. [Formal Limitations on the Measurement of Mutual Information](#). AISTATS 2020

Karl Stratos. [A Sub-Character Architecture for Korean Language Processing](#). EMNLP 2017

Karl Stratos, Michael Collins, and Daniel Hsu. [Unsupervised Part-Of-Speech Tagging with Anchor Hidden Markov Models](#). TACL 2016

Karl Stratos, Do-kyum Kim, Michael Collins, and Daniel Hsu. [A Spectral Algorithm for Learning Class-Based n-gram Models of Natural Language](#). UAI 2014

Margaret Mitchell, Xufeng Han, Jesse Dodge, Alyssa Mensch, Amit Goyal, Alex Berg, Kota Yamaguchi, Tamara Berg, Karl Stratos, and Hal Daumé III. [Midge: Generating image descriptions from computer vision detections](#). EACL 2012 (**ACL Test-of-Time Paper Award, 2022**)

## (Co-)Advisees

[Wenzheng Zhang](#) (PhD in progress at Rutgers), [Rajarshi Bhowmik](#) (PhD in CS at Rutgers), [Mingda Chen](#) (PhD at TTIC), [Zewei Chu](#) (PhD in CS at UChicago), [Daniel Edmiston](#) (PhD in linguistics at UChicago)

## Teaching

Graduate Course on Natural Language Processing, Rutgers University (2020–2024) [[Lectures](#)], Columbia University (2017)

Undergraduate Course on Machine Learning, Rutgers University (2020–2023)

[Machine Learning Summer School](#), Toyota Technological Institute at Chicago (2018)

## Professional Services

Area Chair at top-tier ML and NLP conferences (ICML, NeurIPS, ACL, EMNLP, AAAI) and Standing Reviewer for journals (TACL, JMLR) since 2019.

## Personal Projects

[minDPR](#): distributed data parallel implementation of noise contrastive learning (PyTorch)

[AMMI](#): structured representation learning by adversarial training (PyTorch)

[SimpleNet](#): general deep learning library (C++, Eigen)

[Singular](#): canonical correlation analysis for word embeddings (C++, Eigen)