

Summary

NLP Research Engineer focused on improving the reliability and capabilities of language model systems, particularly through retrieval and evaluation. First author of a Stanford-affiliated EMNLP 2025 paper co-authored with Christopher D. Manning and Nelson F. Liu, demonstrating that strong RAG baselines can outperform more complex pipelines and highlighting the importance of retrieval recall.

Interested in advancing how models access and reason over information, including embedding-based retrieval, evaluation design, and their interaction with training and post-training in modern language models.

Research

“Stronger Baselines for Retrieval-Augmented Generation with Long-Context Language Models”

EMNLP 2025 / Stanford University

FIRST AUTHOR

- Co-authored with Christopher D. Manning (Stanford) and Nelson F. Liu (OpenAI / Stanford), published at EMNLP 2025 Main Conference.
- Showed that strong RAG baselines can match or outperform complex multi-step pipelines, highlighting the importance of retrieval recall.
- Paper in ACL Anthology | Code on GitHub

Stanford University

Stanford, US / Cologne, Germany

NLP RESEARCH COLLABORATION

- Conducted research on retrieval-augmented generation and long-context QA under supervision of Christopher D. Manning and Nelson F. Liu.
- Benchmarked multi-step RAG approaches (e.g., ReadAgent, RAPTOR) against strong baselines across long-context QA datasets.
- Identified retrieval recall as a dominant factor in performance, with implications for RAG system design and embedding-based retrieval.
- Developed evaluation approaches for measuring RAG performance in evolving language model systems.

“Expanding Horizons in RAG: Exploring and Extending the Limits of RAPTOR”

Stanford University

RESEARCH PROJECT (STANFORD CS224N)

- Analyzed and redesigned RAPTOR's retrieval-tree construction, improving retrieval effectiveness for long-context QA.
- Final Report on web.stanford.edu

Experience

Blicity Inc.

US / Germany

FOUNDER & FULL-STACK ENGINEER

- Built and deployed a collaborative multi-user system (iOS client, Elixir backend) with real-time synchronization and scalable infrastructure.

PwC (cundus AG), IBM

US / Germany / China

BUSINESS INTELLIGENCE CONSULTANT & IT PROJECT SUPPORT

- Consulted finance clients with IBM TM1 BI solutions and delivered training sessions for financial planning teams.
- Contributed to international infrastructure projects and coordinated cross-functional teams across multiple regions.

Education

Stanford University — Artificial Intelligence Programs

Stanford, US

GRADUATE & PROFESSIONAL COURSEWORK IN AI / NLP

- CS224N (Graduate, A): NLP with Deep Learning — transformers, attention, language model training.
- XCS221 (Professional): AI principles & Techniques — machine learning, core AI paradigms, state-based models, variable-based models.

University of Applied Sciences Bonn-Rhein-Sieg

Sankt Augustin, Germany

M.Sc. IN COMPUTER SCIENCE — FOCUS: COMPLEX SOFTWARE SYSTEMS

Berlin School of Economics and Law / IBM Germany

Berlin, Germany

B.Sc. IN BUSINESS INFORMATICS (DUAL STUDY PROGRAM)

- Bachelor's thesis on the automated analysis of news articles using natural language processing techniques.