Ni Lao

Homepage: http://www.cs.cmu.edu/~nlao
LinkedIn: https://www.linkedin.com/in/ni-lao
Google Scholar: link
(412) 651-5467 **Email**: noon99@gmail.com

I enjoy the process of understanding intelligence by recreating it.

EXPERIENCE

Google, Mountain View

Nov 2021 - now — Research Scientist

- MUM foundation model Platform
  - This is a widely used foundation model platform and have won many awards (e.g. Google Tech Impact 2023, Search 2022 Tech Impact)
  - Tech lead for grounding and agent solutions. Evaluation and launch of various architectures such as dual encoder, heavy encoder, MoE
  - Solve dual encoder cross-lingual quality issues with a novel pre-training objective (also provide modeling consulting)
  - Authored user tutorial and guideline for dual encoder, robustness & safety, RL training for text generation, and agent technology
- Modeling consultant for various product launches
  - Step by step word problem solving in google.com
  - Semantic short video retrieval (also help to setup eval pipeline)
  - Grounded summarization for assistant answers
  - Shopping/Search subintent generation
- Research (google scholar)
  - LLM/reasoning senior area chair for ACL, Neurips and EMNLP in 2023
  - Modeling consulting for spatially aware model architectures
  - Efficient/diffusion text decoding

Apple, Cupertino

May 2020 - Oct 2021 — Research Scientist

- KG answers and Web answers
  - Define and develop passage answer metrics, which are critical for day-to-day development and launch decisions. Improve short answer metric efficiency using parsing techniques such as Duckling.
  - Unify the evaluation of different answer domains with infra teams
  - Launched neural semantic parsing in search and siri.
- Research
  - Organize Apple's participation of the EfficientQA competition
  - Co-organize the NLP internship program, and co-host summer interns
  - Serve as one of the technical sponsors for Apple's NLU fundings

Mosaix.ai, Palo Alto

Feb 2018 - Apr 2020 — Co-founder and Chief Scientist

- Built and ran a team of NLP/ML engineers/researchers from the ground up. Managed the development and scaling of NLP/ML/quality infrastructures as the foundation for a voice AI platform, which serves millions of users.
- Research (collaborations, talks/interviews, reviews/chairs, ICLR workshop organizer) and publications at top AI conferences with innovations.

Google, Mountain View

Jul 2012 - Feb 2018 — Research Scientist

- Deep online/offline question-answering models. I led the first deep sequence scoring model launches in Google's Web QA system.
- Large scale pre-computation of factoid and non-factoid question-answer pairs from query logs and Web documents.
- Flexibly structured (html) answers in Google's answer box.
- KG schema and semantic parser induction from logs, KG, and Web QA system.
- Model-based knowledge graph confidence estimation, and error detection.
- Knowledge graph construction from Web documents.

Carnegie Mellon University, Pittsburg

Jul 2006 - Jun 2012 — Research Assistant

- Efficient randomized reasoning approach for IR, NLP and recommendation
- Relational CRFs structure learning, and hidden variable induction
- Architecture the CMU's cross-lingual question answering system (JAVELIN) which was a precursor to the IBM Watson system

Microsoft Research Asia, Beijing

Jul 2003 - Jun 2006 — Research Assistant

- Large scale clustering and classification of online products using product image, text and other metadata
- Personalized search ranking by simulating user experience from search log
- An innovative learning to rank method that fit piecewise linear curves
- Automatic operating system troubleshooting based on text descriptions, system config changes, and system call events

State Key Lab of Intelligent Technology and Systems, Beijing

Feb 2001 - Jun 2003 — Research Assistant

- I was one of the main developers for the Tsinghua Aeolus soccer system, which was the world champion of RoboCup Simulation League in 2001 & 2002.
- I work on modeling (e.g., dynamic programming and geometry computation)
- I also developed the debugger visualizing the players' world models, and the bandwidth constrained communication scheme between players.

EDUCATION

Carnegie Mellon University, Pittsburg

July 2006 - June 2012 — PhD in Language Technology

Thesis: Efficient Random Walk Inference with Knowledge Bases.

Tsinghua University, Beijing

July 2003 - June 2006 — Master in Computer Science

Thesis: Data Mining Problems in Automatic Computer Diagnosis.

July 1999 - June 2003 — Bachelor in Electronic Engineering

Thesis: Mining Spatial-Temporal Data Using Constructive Induction.