Lang Yu

1118 7Th Ave NW, 222, Issaquah, WA 98027
(773) 219 - 5437   yulang94@gmail.com

EDUCATION

The University of Chicago   *9/2015 – 06/2021*   *Chicago, IL*
  Ph.D. in Computer Science
  Research Focus: Machine Learning, Natural Language Processing, Computational Linguistics,
  Cumulative GPA: 3.92 / 4.0

Beijing University of Aeronautics and Astronautics   *9/2011 – 8/2015*   *Beijing, China*
  B.S. in Computer Science and Engineering
  Overall GPA: 3.91 / 4.0

PUBLICATIONS

- Li, Jiaxuan, **Lang Yu**, and Allyson Ettinger. "Counterfactual reasoning: Testing language models' understanding of hypothetical scenarios." arXiv e-prints (2023): arXiv-2305.
- Kim, Sanghee J., **Lang Yu**, and Allyson Ettinger. ""No, They Did Not": Dialogue Response Dynamics in Pre-trained Language Models." Proceedings of the 29th International Conference on Computational Linguistics. 2022.
- Li, Jiaxuan, **Lang Yu**, and Allyson Ettinger. "Counterfactual reasoning: Do Language Models need world knowledge for causal inference?." NeurIPS 2022 Workshop on Neuro Causal and Symbolic AI (nCSI).
- **Yu, Lang.** Analyzing and Improving Compositionality in Neural Language Models. Diss. The University of Chicago, 2021.
- **Yu, Lang,** and Allyson Ettinger. "On the Interplay Between Fine-tuning and Composition in Transformers." Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021.
- **Yu, Lang**, and Allyson Ettinger. "Assessing Phrasal Representation and Composition in Transformers." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.
- (Manuscript) **Yu, Lang,** John Goldsmith. "Unsupervised Vocabulary Expansion with Hierarchical Signature Structure", 2019.
- **Yu, Lang**, Zhongzhi Luan, Xiangzheng Sun, Zhe Wang, and Hailong Yang. "VinaSC: Scalable Autodock Vina with fine-grained scheduling on heterogeneous platform." In Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on, pp. 790-793. IEEE, 2016.
- Meng, You, **Lang Yu**, Zhongzhi Luan, Depei Qian, Ming Xie, and Zhigao Du. "A Black-Box Approach for Detecting the Failure Traces." In International Conference on Trustworthy Computing and Services, pp. 252-259. Springer Berlin Heidelberg, 2013.

RESEARCH EXPERIENCE & EMPLOYMENT

Meta

Senior Research Scientist   *08/2023 – Present*   *Bellevue, WA*
  I am a senior research scientist at Ads Monetization Ranking AI (formerly Core ML) team. I work on exploring state-of-the-art machine learning techniques and apply them on Reels short-form video surface.

Research Scientist   *09/2021 – 08/2023*   *Bellevue, WA*
  I was a research scientist at Aggregated Event Processing team. I built privacy-preserving machine learning systems for Ads Signal across all Meta products. I developed end-to-end

prototypes for modeling advertisement conversions triggered by Meta Ads without using strong user-identifiable information. Specifically, I was focusing on "Signal Loss" problem, where advanced deep learning models are built for ads measurement and calibration without relying on information with potential privacy concerns of user tracking. I was responsible for developing data processing pipelines, improving crucial machine learning models, running analysis, and productizing the project, together with ads delivery and measurement teams.

University of Chicago   *07/2018 –06/2021*   *Chicago, IL*
Ph.D. Researcher

- **Analyzing Composition in Deep Neural Networks**
  My research generally focused on the interdisciplinary between computer science and linguistic. I worked on studying compositions in deep neural networks. I investigated semantic and syntactic composition in a recurrent neural language model RNNG. Additionally, I worked on probing compositionality in deep transformer models. I proposed a set of probing tasks with careful control of dataset artifacts, teasing apart sophisticated compositional signal and lexical effects. I also investigated the impact of fine-tuning language models on compositionality in these models. My research also explored on counterfactual reasoning and dialogue response in LLMs.

Teaching Assistant

      I have rich experience in teaching both undergraduate and graduate level classes. My TA experiences ranges from CS intro to specialized courses including Computer Architecture, Database and Machine Learning. As my research interests evolve, I then taught advanced ML courses, cross-listed in CS and Statistics departments.

Facebook   *06/2020 – 09/2020*   *Seattle, WA*
Ph.D. Intern

- **Public Figures to Follow (PFTF)**
  I was on Public Connections team, working on "Public Figures to Follow" (PFTF). I built end-to-end pipelines to generate samples and trained multiple deep neural models for user recommendations. Multiple approaches have been experimented targeting at improving user follow through rate.

Facebook   *06/2019 – 9/2019*   *Seattle, WA*
Ph.D. Intern

- **Email Pages You Might Like (ePYML)**
  I worked with Page Connection Team on building neural networks for user-specific page recommendations. I built data pipelines to process user and page data and generate corresponding features on a daily basis. To support fast online evaluation and minimize memory usage for large-scale data, I implemented and optimized Two-Tower Sparse Network (TTSN) to learn user embeddings and page embeddings. And during online evaluation, the compressed embeddings are retrieved from memory, and a fast dot product is used to generate user-page affinity scores. In addition to model implementation and tuning, I ran experiments both offline and online, demonstrating notable improvement on user conversion rate over previous models.

- **Linguistica Project**   *01/2017 – 06/2018*   *Chicago, IL*
  I worked Linguistica Project (http://linguistica.uchicago.edu/). The goal of this project is to propose a language-independent morphological learning model based on Minimum Description Length (MDL). Specifically, the model learns and generates morphological

analysis of the corpus without human supervision. I am focusing on the morphological analysis of out-of-vocabulary (OOV) words and signature statistics modeling. To improve the accuracy of analysis, I am exploiting the evolution of signatures while feeding new workloads to the model. More specifically, the pattern of new suffixes and stems occurrences is modeled by several Pitman-Yor processes. In addition, the temporal knowledge makes the model able to evolve while reading more words that have not been observed before. Having modeled signature evolution, finite-state automata are constructed for the morphological analysis.

Large-scale System Group   *9/2015 – 12/2016*   *Chicago, IL*
Research Assistant

- **DNA Alignment with Leven UAP Approach**
  We proposed a Levenshtein-automata-based approach to perform DNA short reads alignment. Multiple optimizations, including short read packing, automata transition sharing and redundant states removal, are implemented to further accelerate the alignment process. With the help of UAP, which is a specialized in-memory computing unit that performs automata operations efficiently, a speedup over 8X can be obtained compared with existing applications. In this project, I built everything from draft and implemented several optimizations in the UAP compiler that significantly compress the size of target binary image.

- **Accelerating Data Analytics (ADA)**
  In this project, we focused on large-scale data streaming system, targeting at improving performance for in-memory data analytics systems. Starting from Spark, we collected detailed data transformation information (compression/decompression, serialization/deserialization) by instrumenting Spark and Snappy compression and evaluated it with Big Data Benchmark workloads.

Intel Ltd.   *12/2014 – 7/2015*   *Beijing, China*
Internship in Software & Services Group

- **Porting and In-depth Optimization of Autodock Vina on MIC**
  With the help of my supervisor, I proposed and implemented a dynamic heterogeneous scheduling framework from scratch, which allows Autodock Vina (a molecular docking software) to be executed in multi-host concurrent pattern on clusters with CPU + MIC architecture. Large-scale experiments were performed in Tianhe-2 cluster. This work has been published in the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Shanghai Center for Bioinformatics Technology   *7/2014 – 8/2014*   *Shanghai, China*
Research Assistant

  I was mainly responsible for the maintenance and management of high-performance computing infrastructure in the center. Besides, I had a rudimentary knowledge of their research work, especially the large data set analysis, protein-protein interaction network analysis.

Sino-German Joint Software Institute   *10/2013 – 7/2015*   *Beijing, China*
Research Assistant

- **The Event Chain Discovery and Fault Prediction in Distributed System**
  This project mainly focused on the discovery and learning of abnormal event trains in order to make reliable prediction of possible faults in the distributed system. Our research findings are presented in the paper "A Black-box Approach for Detecting the Failure Traces."

TEACHING ASSISTANT EXPERIENCE

- **CMSC 25400 Machine Learning (Spring 2021)**

- **STAT 37710 Machine Learning (Winter 2021) Head TA**
- **CMSC 25300 / CMSC 35300 / Statistics 27700 Mathematical Foundations of Machine Learning (Fall 2020) Head TA**
- **STAT 37710 Machine Learning (Spring 2020)**
- **CMSC 15100 Introduction to Computer Science (Winter 2019)**
- **CMSC 25300 / CMSC 35300 / Statistics 27700 Mathematical Foundations of Machine Learning (Fall 2019)**
- **CMSC 25020 / CMSC 35020 Computational Linguistics (Spring 2018)**
- **CMSC 12200 Computer Science with Applications 2 (Winter 2018)**
- **CMSC 15100 Introduction to Computer Science I (Autumn 2017)**
- **CMSC 12300 Computer Science with Applications 3 (Spring 2017)**
- **CMSC 23500 / CMSC 33550 Introduction to Database (Winter 2017)**
- **CMSC 12100 Computer Science with Applications 1 (Autumn 2016)**
- **CMSC 22200 / CMSC 32200 Computer Architecture (Fall 2015)**

OTHER PROJECTS

- **Design & Implementation of a MIPS Multi-Cycle Processor**
  It is a Verilog implementation of MIPS Multi-cycle. The processor supports 53 MIPS instructions, which are most commonly used. The accomplished processor is a generally fledged multi-cycle processor, able to accomplish the complete process of interruption entrance, service and return.
- **Design & Implementation of Linux-like Kernel**
  The accomplished kernel incorporates functional modules including: file system, memory management, process scheduler, shell etc. The process-scheduling module is improved, replacing the scheduler using basic Round-robin algorithm with a more flexible scheduler, which can be configured to use dynamic priority queue scheduling, lottery scheduling and other five scheduling algorithms. The kernel can accomplish all the core functions of a modern operating system.
- **Design & Implementation of PL/0 Compiler**
  The compiler is able to compile PL/0 programs and generate executable binary code on x86; with error localization, accurate error messages could be produced. Moreover, several optimizations commonly used in modern compiler are also implemented, including common sub-expression elimination, active variable analysis, peephole optimization, read-write optimization, register (local and global) optimization etc.

PROFESSIONAL SKILLS

- Proficient with Pytorch and deep transformer models
- Proficient with parallel programming
- Proficient in programming on Intel Many Integrated Core architecture
- Proficient in C/C++ & Python programming
- Experienced in Java programming

AWARDS & HONORS

- National Outstanding Graduates Awards, 2015, Beijing University of Aeronautics and Astronautics
- China National Scholarship, 2015, Beijing University of Aeronautics and Astronautics
- Honorable Mention in Mathematical Contest in Modeling (MCM), 2014
- Outstanding Undergraduate Awards, 2014, Beijing University of Aeronautics and Astronautics