# Yifei Li

PH.D. STUDENT · UNIVERSITY OF PENNSYLVANIA

*Legal name: Yifei Li; Publish as: Li S. Yifei*

liyifei@seas.upenn.edu | realliyifei.github.io | realliyifei | realliyifei | @realliyifei | Google Scholar

***Domains**: Large Language Model, Multimodal, Expert-Level NLP, and Scalable AI*

## Education

**University of Pennsylvania (UPenn)** — *Philadelphia, PA, USA*

Ph.D. in Computer and Information Science (Earned M.S.E. in Data Science); Advised by Prof. Mark Yatskar — *GPA: 3.97/4.0 | 2028**

**Oklahoma State University (OSU)** — *Stillwater, OK, USA*

B.S. in Computer Science, B.S. in Mathematics, and B.S.B.A. in Management — *GPA: 3.8/4.0 | 2019*

**Sun Yat-Sen University (SYSU)** — *Guangzhou, China*

B.M. in Management *(Notes: International joint-degree program associated with OSU)* — *GPA: 3.8/4.0 | 2019*

## Publications

Conceptor-Aided Debiasing of Large Language Models
Li S. Yifei, Lyle Ungar, João Sedoc
*Empirical Methods in Natural Language Processing (EMNLP)*
URL: `https://arxiv.org/abs/2211.11087` 2023

## Skills

| | |
|---|---|
| **Coding Languages** | Python, Java, MATLAB, R, C++, SQL, JavaScript, HTML/CSS |
| **Libraries and Tools** | PyTorch, Hugging Face, OpenAI, Faiss, Spark, Shell, Git, LaTeX |
| **AI Models** | GPT family, Llama, CLIP, BERT family, Stable Diffusion, MoCo, YOLO, SOLO, GAN family |

## Research Projects

**Survey QA: Mining Natural Long-Form Expert-Level QA from Survey Papers at Scale** — *UPenn*

Advised by Prof. Mark Yatskar — *Feb. 2024 - Present*

- Develop an advanced long-form scientific QA system across 23 expert domains, surpassing reading comprehension difficulty, with multi-document support and inline citations automatically extracted from research papers by an LLM to scale expert-level data for pretraining.

**Conceptor-Aided Debiasing of Large Language Models** — *UPenn*

Advised by Prof. João Sedoc and Prof. Lyle Ungar - Paper — *Jan. 2022 - Jul. 2022*

- Use conceptors–a soft projection method–to identify and remove the bias subspace in contextual embeddings of LLMs by post-processing subspace projection and continued training; Mitigate intersectional bias by manipulating the bias subspace by logical operators of conceptors.

**Neuro-Symbolic Dual-System on Task-Oriented Dialogue Generation** — *UPenn*

Advised by Prof. Chris Callison-Burch and Dr. Lara Martin (Research Course Project) - Report — *Mar. 2022 - May. 2022*

- Adapt neuro-symbolic dual-system to improve the consistency and coherence in task-oriented dialog generation: Upon a user belief state as the grounding of dialog history, using two GPTs to generate utterance and symbolically verify the consistency with belief states.

**Improving Text-to-Image Diffusion Generation via Large Language Models** — *UPenn*

Advised by Prof. Chris Callison-Burch and Prof. Mark Yatskar (Master Thesis) - Report — *Aug. 2022 - May. 2023*

- Propose (1) an imagine-then-verbalize approach using LLMs to enhance image generation with richer context; (2) a sketch-then-draw method leveraging LLM coding to improve diffusion generation numerical consistency.

**Probing CLIP Zero-Shot Ability** — *UPenn*

Advised by Prof. Mark Yatskar (Independent Study) — *Jan. 2022 - May. 2022*

- Evaluate the zero-shot ability of CLIP on fine-grained datasets e.g. iNaturalist; Try to build a language utility to help users ask the right questions to CLIP: leverages the sentence-BERT to cluster different types of web texts, then compares and ranks their CLIP similarity scores.

## Professional Activities

2022-24 **EMNLP*2 / ACL / COLM / ARR*3,** Reviewer

2021-23 **CIS 530 NLP (*2), CIS 522 DL, CIS 520 ML,** Teaching Assistant — *UPenn*

## Awards & Honors

| | | |
|---|---|---|
| 2023 | **Outstanding Teaching Award,** Significant contributions as teaching assistant | *UPenn* |
| 2016-19 | **President's Honor Roll,** Maintain Excellent GPA | *OSU* |
| 2018 | **Emeritus Math Faculty Scholarship,** Mathematics Department, 1-2 Student(s) Each Year | *OSU* |
| | **SST Scholarship,** Computer Science Department | *OSU* |

## Miscellaneous

- Yifei worked in financial and technology consulting positions at several financial and accounting firms when he focused on business, then during his master's at UPenn, he developed a great passion for artificial intelligence and made a significant shift to AI academia :)