

# Yufei Li

COMPUTER SCIENCE · MACHINE LEARNING RESEARCHER

☎ (+1) 858-291-2678 | ✉ yxl190090@utdallas.edu | 🏠 yul091.github.io/liyufei.me/ | 📞 yul091 | 🌐 Yufei(Ethan) Li | 🐦 @CavMonkey

## Summary

I am a second-year computer science Ph.D. student from the University of Texas at Dallas (UTD). I conducted research at the intersection of natural language processing (NLP), Computer Vision (CV), and Software Engineering (SE). I also focused on the application of NLP in Data Mining and Recommender Systems, and investigated the explainable AI such as model uncertainty measurements.

## Education

### University of Texas at Dallas

Dallas, TX, U.S.

PH.D. IN COMPUTATIONAL SCIENCE

2020 - Present

Artificial Intelligence Group, CSE; Advisor: Prof. Wei Yang

### University of California, San Diego

San Diego, CA, U.S.

M.S. IN ELECTRICAL AND COMPUTER ENGINEERING

2018 - 2020

Thesis: Application of A\* and Genetic Algorithm in TSP Path Planning Problem; Advisor: Prof. Farinaz Koushanfar

### Xi'an Jiaotong University

Xi'an, Shaanxi, China

B.S. IN MECHANICAL ENGINEERING

2014 - 2018

Thesis: Aircraft fuel tank oil sloshing simulation based on SPH method and center of gravity distribution analysis; Advisor: Prof. Shuai Zheng

## Publications

### SHARE: a System for Hierarchical Assistive Recipe Editing.

LI, S., LI, Y., NI, J., McAULEY, J. | 2021 *arXiv:2105.08185 (CoRR)* [PDF]

- Proposed a task of controllable recipe editing: adapting a base recipe to satisfy a user-specified dietary constraint to assist home cooks.
- Tackled this problem with a hierarchical framework which performs simultaneous ingredient substitution via a Set Transformer, followed by conditional instruction generation using a Copy Attention network.

### Estimating Predictive Uncertainty Under Program Data Distribution Shift.

LI, Y., CHEN, S., YANG, W. | 2021 *arXiv:2107.10989 (CoRR)* [PDF]

- Defined three real-world program distribution shifts based on practical software development scenarios.
- Investigated the effectiveness of existing uncertainty measurements under program distribution shifts and proposed a large-scale benchmark for their performance.
- Analyzed each method's pros and cons from a logic design perspective to inspire a domain-specific uncertainty design.

### GLIB: Towards Automated Test Oracle for Graphically-Rich Applications.

CHEN, K., LI, Y., CHEN, Y., FAN, C., HU, Z., YANG, W. | 29th *Foundations of Software Engineering (FSE 2021)* [PDF]

- Proposed a code-based data augmentation approach for generating UI glitch images on game apps.
- Designed & built a CNN model for UI glitch detection and localized the glitch area using a saliency map to facilitate bug fixing.

## Work Experience

### NEC Laboratories America, Inc.

Princeton, NJ, U.S.

RESEARCH INTERN, PYTORCH

Jun 2021 - Aug 2021

- Annotated name entities and relations with regular expression rules on CVE texts for distant supervision.
- Incorporated the pre-trained GPT-2 backbone into a sequence labeling framework for joint entity & relation extraction.
- Proposed a bootstrap training procedure for denoising distant labels and selecting high-quality instances.

### SeekTruth Scientific and Technical Corporation

Beijing, China

RESEARCH INTERN, TENSORFLOW

Jul 2019 - Sep 2019

- Designed and boosted an adaptive discrimination definition model for objection detection.
- Designed a light-weight CNN model for identifying the direction of videos.
- Built a joint key point & pose recognition model for character detection.

# Projects

---

## GAET, Embedding Evaluation

*Dallas, TX, U.S.*

RESEARCH PROJECT, PYTORCH

*Jul 2020 - Sep 2020*

- Proposed a low-cost offline metric for evaluating the generalizability of code embedding in software engineering applications.
- Patched the existing pre-trained embedding based on the semantic metamorphic relationship in vector space to improve the generalizability.

## Negative Sampling with Bias Correlation, Data Mining

*San Diego, CA, U.S.*

RESEARCH PROJECT, PYTORCH

*Nov 2019 - Jun 2020*

- Explained the unsuccessful performance of the popularity-based sampling scheme in terms of sampling bias introduced from the non-uniform negative sampling process.
- Proposed two ways to correct the bias and design related negative sampling distributions to boost accuracy from tradition Bayesian Personalized Ranking (BPR).

## Pet Adoption Speed Prediction, Natural Language Processing

*San Diego, CA, U.S.*

RESEARCH PROJECT, TENSORFLOW

*Jan 2019 - Mar 2019*

- Parsed and tokenized the description text of pets, encoded the words with word embedding.
- Designed & built a machine learning framework to combine image features & natural language meanings for popularity prediction.

## Automatic Delivery Vehicle Design, Path Planning

*San Diego, CA, U.S.*

RESEARCH PROJECT, PYTHON & MATLAB

*Mar 2019 - Jun 2019*

- Incorporated the Courier and TSP travel agent problems in designing autonomous delivery vehicles.
- Designed & built a dynamic path planning model combining genetic algorithm and A\* algorithm that for path planning.

# Awards

---

## VEX Robotics International Competitions

TEAM COMPETITION, C LANGUAGE

- Excellent award and runner-up in 2017 VEX Robotics World Championship, Louisville, KY, U.S..
- Excellent award and runner-up in 2016 VEX Robotics Asia Open, Beijing, China.
- First class honor in 2016 VEX Robotics China Open, Xi'an, China.

## Scholarship Awards

PERSONAL

- National encouragement scholarship 2017, top10% of the 300+ applicants.
- National encouragement scholarship 2016, top10% of the 300+ applicants.
- National encouragement scholarship 2015, top10% of the 300+ applicants.