Shengjie Li

732-485-7674 | Plano, TX | shengjie.li.cs@gmail.com | linkedin.com/in/shengjie-li | shengjie-li.com

EDUCATION

University of Texas at Dallas — GPA: 3.87/4.0
Ph.D. in Computer Science — *Aug. 2020 – May 2025 (Expected)*

Rutgers University — GPA: 3.85/4.0
M.S. in Computer Science — *Sep. 2018 – May 2020*

South China Agricultural University — GPA: 86.3/100
B.Eng. in Software Engineering — *Sep. 2013 – June 2017*

INDUSTRY RESEARCH EXPERIENCE

Amazon — Seattle, Washington
Applied Scientist Intern — *May 2023 – Aug. 2023*

- Researched and developed an LLM-based workflow to auto-extract customer-centric product metadata for 5 types of products on amazon.com, poised to significantly automate labor-intensive tasks.
- Augmented LLMs with external knowledge sources, ensuring evidence-grounded results and reducing the risk of inaccurate model outputs (hallucinations).

ACADEMIC RESEARCH EXPERIENCE

Human Language Technology Research Institute at UT Dallas — Richardson, Texas
Research Assistant, Advisor: Vincent Ng — *May 2021 – Present*

- Built a pipelined model for entity coreference resolution, achieving the best performance in the CODI-CRAC 2022 shared task which is 1.6x as good as the baseline and 1.2x as good as the second-ranked team.
- Developed a top-performing end-to-end model for the discourse deixis track in the CODI-CRAC 2021 shared task, utilizing resolution constraints to achieve a performance of 1.8x as good as the second-ranked team.
- Developed an end-to-end model for discourse deixis resolution that leveraged task-specific characteristics and outperformed previous state-of-the-art by 27%, resulting in a first-author publication in EMNLP 2022.
- Collected and analyzed data for identifying propaganda content in Spanish magazines during World War II. Identified key challenges for developing a deep learning model. Published findings in AAAI 2023.

NLP Group at Rutgers University — Piscataway, New Jersey
Research Assistant, Advisor: Matthew Stone — *June 2019 – May 2020*

- Developed a multimodal classification model to predict the coherence relation between an image and its caption, aiding the development of controllable caption generation and leading to an ACL 2020 publication.

SELECTED PUBLICATIONS (3 OUT OF 6)

Ng, V., & **Li, S.**. Multimodal Propaganda Processing. (AAAI 2023)

Li, S., & Ng, V.. End-to-End Neural Discourse Deixis Resolution in Dialogue. (EMNLP 2022)

Alikhani, M., Sharma, P., **Li, S.**, Soricut, R., & Stone, M. Cross-modal Coherence Modeling for Caption Generation. (ACL 2020)

SKILLS

Programming Languages: C++, C#, Java, Python, SQL, Shell, LaTeX

Technologies: PyTorch, TensorFlow, Flask, Vue.js, MySQL, Oracle DBMS, PostgreSQL, Docker, LangChain

HONORS & AWARDS

Louis Beecherl, Jr. Graduate Fellowship at UT Dallas — 2022-2024

Silver Medal in the 2015 ACM-ICPC (International Collegiate Programming Contest) Asia Shenyang Regional Contest — Oct. 2015