Lang Cao

(+1) 217-621-0532 | windszzlang@gmail.com
Webpage: windszzlang.github.io | GitHub: github.com/windszzlang

Education

University of Illinois at Urbana-Champaign (UIUC) **Urbana, USA**
Master of Science in Computer Science (Research-based Program) *Aug. 2022 - May 2024*
- Advised by Prof. Jimeng Sun
- Research Area: AI for Healthcare; Natural Language Processing
- GPA: 3.89/4.0

Wuhan University of Technology (WHUT) **Wuhan, China**
Bachelor of Engineering in Software Engineering *Sept. 2018 - June 2022*
- GPA: 4.351/5.0 (3.94/4.0); Rank: 1st/79

Publications

PILOT: Legal Case Outcome Prediction with Case Law [Paper] [Code]
- **Lang Cao**, Zifeng Wang, Cao Xiao, Jimeng Sun.
- 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2024 Main Conference.

AutoAM: An End-To-End Neural Model for Automatic and Universal Argument Mining [Paper][Code]
- **Lang Cao** (Independent Research).
- 19th anniversary of the International Conference on Advanced Data Mining and Applications, ADMA 2023.

CBCP: A Method of Causality Extraction from Unstructured Financial Text [Paper][Code]
- **Lang Cao**, Shihua Zhang, Juxing Chen.
- 2021 5th International Conference on Natural Language Processing and Information Retrieval, NLPIR 2021.

Clustering of Functionally Related Genes Using Machine Learning Techniques [Paper]
- Yujing Xue and **Lang Cao**.
- 2021 5th International Conference on Compute and Data Analysis, ICCDA 2021.

Intelligent Cross-sensing Sensor Based on Deep Learning [Paper]
- Lingfei Xu, Jiaming Zhang, **Lang Cao**, Xinyu Hu.
- 2021 6th IEEE International Conference on Signal and Image Processing, ICSIP2021.

DiagGPT: An LLM-based Chatbot with Automatic Topic Management for Task-Oriented Dialogue [Paper][Code]
- **Lang Cao** (Independent Research).
- Under Review (ACL 2024).

Enhancing Reasoning Capabilities of Large Language Models: A Graph-Based Verification Approach [Paper][Code]
- **Lang Cao** (Independent Research).
- Under Review (NeurIPS 2024).

Learn to Refuse: Making Large Language Models More Controllable and Reliable through Knowledge Scope Limitation and Refusal Mechanism [Paper][Code]
- **Lang Cao** (Independent Research).
- Under Review (ACL ARR).

AutoRD: An Automatic and End-to-end Rare Disease Knowledge Graph Construction System Based on Ontologies-enhanced Large Language Models [Paper][Code]
- **Lang Cao**, Adam Cross, Jimeng Sun.
- Under Review (JAMIA).

Large Language Models Synthesize Real-World Evidence from Medical Literature
- Zifeng Wang, **Lang Cao**, Benjamin Danek, Yichi Zhang, Qiao Jin, Zhiyong Lu, Jimeng Sun.
- Under Review (Nature Medicine).

Experiences

Sunlab, UIUC **Urbana, US**
Research Assistant *Jan. 2023 - Now*
- Research Focus: Natural Language Processing for Applications in Healthcare and Legal.
- Projects:
  - Legal Case Outcome Prediction with Case Law
  - Large Language Models for Rare Disease Information Extraction and Potential Rare Disease Prediction
- Advisor: Jimeng Sun, Danica Xiao

Keiji (Healthcare AI Startup) **Urbana, US**
Research Internship *Jan. 2024 – Now*
- Developed eligibility criteria generator demo, which can automatically generate patient eligibility criteria for clinical trials based on predefined parameters, thereby streamlining the trial setup process.
- Developed meta-analysis demo, which can automatically aggregate and analyze data from multiple studies, providing insights and trends that help in making informed decisions.

LegalNow, LegalDAO (Legal AI Startup) **Beijing, China**
Research Internship *June 2023 – Sept. 2023*
- Individually completed the construction of an AI legal chatbot in the demo version of the product, which can assist and guide users in achieving multiple functions related to contracts drafting.
- Provided technical support and developed the overall AI framework for the first launching product.

Bioinformatics Innovation Lab (Text Group), WHUT **Wuhan, China**
Research Assistant *Sept. 2020 - Jan. 2021*
- Research Focus: Information Extraction and Text Mining.
- Projects: Named Entity Extraction from Biomedical Literature
- Advisor: Jing Peng

iFLYTEK CO. LTD. **Hefei, China**
NLP Algorithm Engineer at Smart Car Technology R&D Division *June 2021 - Aug. 2021*
- Maintained and developed an automatic data iteration algorithm for training data of smart car AI system.
- Advisor: Shen'an Li

Software

PyHealth [Github] [PyHeath-GPT]
- A Deep Learning Python Toolkit for Healthcare Applications.
- Work Content: developed an AI chat assistant to help new users understand and learn how to use PyHealth.

Honors & Awards

- Silver Medal, top 5% in Kaggle Common Lit Readability Prize (2021.8)
- Top 2% in Alibaba Tianchi NLP Chinese Pre-training Model Generalization Ability Challenge (2021.1)
- 2nd Prize in National College Computer Ability Challenge - Artificial Intelligence Application Contest (2021.1)
- Silver Award in China College Students' "Internet Plus" Innovation and Entrepreneurship Contest (2020.9)
- National Scholarship (1%), WHUT (2020); Merit Student Model Honor (5‰), WHUT (2020); First-class Scholarship, WHUT (2021); Merit Student Honor, WHUT (2021); Outstanding Graduate, WHUT (2022.6); Outstanding Thesis, WHUT (2022.6)
- The National Champion of the FIRST LEGO League in China (2014.6); Gold Award at the Asia-Pacific Championship of the FIRST LEGO League (2016.7)

Skills

- **Programming:** Python, C/C++, Java, JavaScript/TypeScript, Shell
- **Machine Learning Techniques:** PyTorch, TensorFlow, Keras, Huggingface Transformers, LangChain, DGL, Scikit-learn, NumPy, Pandas
- **Software Development Techniques:** Vue.js, React.js, Node.js, Django, Flask, Express, Hadoop, SQL, MFC, QT
- **Others:** LaTeX, Markdown, Git, Linux