Lang Cao

(+1) 217-621-0532 | lgcao2000@163.com
Personal Page: windszzlang.github.io

Education

University of Illinois at Urbana-Champaign (UIUC) **Urbana, USA**
Master of Science in Computer Science (Research-based Program) *Aug. 2022 - May 2024 (expected)*
- 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

AutoAM: An End-To-End Neural Model for Automatic and Universal Argument Mining [Paper][Code]
- **Lang Cao** (Independent Research).
- In 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.
- In 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**.
- In 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.
- In 2021 6th IEEE International Conference on Signal and Image Processing, ICSIP2021.

PILOT: Legal Case Outcome Prediction with Case Law
- **Lang Cao**, Zifeng Wang, Cao Xiao, Jimeng Sun.
- Under Review.

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

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

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.

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

LAMB: Language Spectrum as a Memory Base
- Ke Yang, **Lang Cao**.

- Preprint.

Experiences

Sunlab, UIUC **Urbana, US**
Research Assistant *Jan. 2023 - Now*
- Research Focus: Natural Language Processing for Applications in Healthcare and Legal.
- Advisor: Jimeng Sun, Danica Xiao

LegalNow, LegalDAO (Legal AI Startup) **Beijing, China**
Cofounder & AI Tech Leader *June 2023 - Now*
- 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 *Jan. 2021 - May. 2022*
- Research Focus: Information Extraction and Text Mining.
- 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

Course: 20Spring Algorithm Design and Analysis, WHUT **Wuhan, China**
Teaching Assistant *Feb. 2020 - June 2020*

English Teaching Volunteer in Thailand **Chiangmai, Thailand**
Educational volunteer *Jan. 2020*
- Provided voluntary English teaching for 60 primary school students (35h).

Software

PyHealth [Github] [PyHeathChat]
- 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.

LegalNow AI Lawyer [Homepage]
- A smart and friendly Lawyer-Grade AI for users to create or review legal documents.
- Work Content: developed the overall AI framework. Designed and improved prompts iteratively.

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)
- 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, Shell
- **Techniques:** PyTorch, TensorFlow, Keras, Huggingface Transformers, LangChain, DGL, Scikit-learn, NumPy, Pandas, Django, Flask
- **Others:** LaTeX, Markdown, Git, SQL, Linux