Lang Cao

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

Education

University of Illinois at Urbana-Champaign (UIUC) **Urbana, USA**
*Master of Science in Computer Science (**Research-oriented Program**)* *Sept. 2022 - May 2024 (expected)*
- Advised by Prof. Jimeng Sun
- 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 and 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**, and 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.

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: 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).

Selected Research Projects

Exploring Knowledge Scope Limitation and Refusal Mechanism for LLMs **Urbana, USA**
Independent Research *Sep. 2023 – Nov. 2023*
- Large language models (LLMs) have demonstrated impressive language understanding and generation capabilities, enabling them to answer a wide range of questions across various domains. However, the hallucination render LLMs unreliable and even unusable in many scenarios. Instead of attempting to answer all questions, we explore a refusal mechanism that instructs LLMs to refuse to answer challenging questions to avoid errors.
- We propose a simple yet effective solution called Learn to Refuse (L2R), which incorporates the refusal mechanism to enable LLMs to recognize and refuse to answer questions that they find difficult to address. To achieve this, we utilize a structured knowledge base to represent all the LLM's understanding of the world, enabling it to provide traceable gold knowledge. This knowledge base is separate from the LLM and initially empty, and it is progressively expanded with validated knowledge. When an LLM encounters questions outside its domain, the system recognizes its knowledge scope and determines whether it can answer the question independently. Additionally, we introduce a method for automatically and efficiently expanding the knowledge base of LLMs. Through qualitative and quantitative analysis, we demonstrate that our approach enhances the controllability and reliability of LLMs.

Graph-Based Verification to Enhance Reasoning Capabilities of LLMs **Urbana, USA**
Independent Study *Oct. 2023 - Dec. 2023*
- Large Language Models (LLMs) have showcased impressive reasoning capabilities, particularly when guided by prompts in chain-of-thought reasoning tasks such as math word problems. However, there is still significant room for enhancing the reasoning abilities of LLMs. Some studies suggest that the integration of an LLM output verifier can boost reasoning accuracy without necessitating additional model training.
- We follow these studies and introduce a novel graph-based method to further augment the reasoning capabilities of LLMs. We posit that multiple solutions to a reasoning task, generated by an LLM, can be represented as a reasoning graph due to the logical connections between intermediate steps from different reasoning paths.
- We propose the Reasoning Graph Verifier (RGV) to analyze and verify the solutions generated by LLMs. By evaluating these graphs, models can yield more accurate and reliable results. Our experimental results show that our graph-based verification method not only significantly enhances the reasoning abilities of LLMs but also outperforms existing verifier methods in terms of improving these models' reasoning performance.

LLM-based Chatbot with Automatic Topic Management for Goal-Oriented Dialogue **Beijing, China**
Independent Study *June 2023 – Aug. 2023*
- Current LLMs have shown proficiency in answering general questions. However, basic question-answering dialogue often falls short in complex diagnostic scenarios, such as legal or medical consultations.
- These scenarios typically necessitate Goal-Oriented Dialogue, wherein an AI chat agent needs to proactively pose questions and guide users towards specific task completion. Previous fine-tuning models have underperformed in TOD, and current LLMs do not inherently possess this capability.
- We introduce DiagGPT (Dialogue in Diagnosis GPT), an innovative method that extends LLMs to TOD scenarios. Our experiments reveal that DiagGPT exhibits outstanding performance in conducting TOD with users, demonstrating its

potential for practical applications.

Legal Case Outcome Prediction with Case Law **Urbana, USA**

Advised by Danica Xiao and Jimeng Sun *June 2023 – Aug. 2023*

- Machine learning shows promise in predicting the outcome of legal cases, but most research has concentrated on civil law cases rather than case law systems. We identified two unique challenges in making legal case outcome predictions with case law. First, it is crucial to identify relevant precedent cases that serve as fundamental evidence for judges during decision-making. Second, it is necessary to consider the evolution of legal principles over time, as early cases may adhere to different legal contexts.
- We proposed a new model named PILOT (PredictIng Legal case OuTcome) for case outcome prediction. It comprises two modules for relevant case retrieval and temporal pattern handling, respectively.
- To benchmark the performance of existing legal case outcome prediction models, we curated a dataset from a large-scal e case law database. We demonstrate the importance of accurately identifying precedent cases and mitigating the temp oral shift when making predictions for case law, as our method shows a significant improvement over the prior methods that focus on civil law case outcome predictions.

Reading Paragraph by Paragraph to Make LLMs Accept Infinite Input **Urbana, USA**

Independent Research *Nov. 2023 – Now (Ongoing)*

- Large language models still struggle when dealing with very long or infinite input sequences. Our approach involves incorporating additional parameters to store temporary information from overlong input sequences. First, we segment the input sequence into chunks that adhere to token limitations of the model. we freeze the main model and exclusively train the additional parameters on these chunks. The new information will be stored in these new parameters. After that, we can employ the entire model for testing based on specific instructions. It is like an LLM reads a long story paragraph by paragraph; then it can memorize them and answer some questions. When the model encounters a new task or context, we reinitialize this parameter matrix and repeat this process.

Rare Disease Knowledge Graph Construction and Prediction with Ontologies-based LLMs **Urbana, USA**

Advised by Prof. Jimeng Sun *Sept. 2023 – Now (Ongoing)*

- The discovery and understanding of rare diseases are of great importance and urgency. Unfortunately, the lack of related data and the shortage of human efforts make it necessary to leverage machine learning methods to enhance this process. Large language models exhibit impressive zero-shot capabilities, but they often lack specialized structured knowledge. To address this, our approach involves incorporating existing medical ontologies to augment the knowledge of these models. We then employ prompts to guide large language models in extracting information about rare diseases and related phenotypes. Using this information, we construct a knowledge graph, allowing us to predict potential rare diseases in patients.

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, Huggingface Transformers, LangChain, DGL, Scikit-learn, NumPy, Pandas, Django, Flask
- **Others:** LaTeX, Markdown, Git, SQL, Linux