ZEFAN CAI

✉ zefncai@gmail.com · ⚲ Google Scholar · ⚲ Main Page

EDUCATION

Peking University - School of Software & Microelectronics   2022 – Present
M.Sc. majoring in Computer Science Technology; GPA: 3.71/4.0 (top 5%)
Supervisor: Prof. **Baobao Chang** from Institute of Computational Linguistics (ICL) from Peking University

Beijing Jiaotong University - College of Computer Science and Technology   2018 – 2022
B.Eng. majoring in Computer Science; GPA: 3.61/4.0 (Rank 2/31)

PUBLICATIONS, SUBMISSIONS AND PREPRINTS

My previous research mainly include **Large Language Models (LLMs)** [1][3][7][11], **Language Grounding with Vision** [6][8][9][12], **AI4Science** [5] and **Traditional NLP** [2][4][10][13].

1. MMICL: Empowering VLM With Multi-Modal In-Context Learning   [PDF] [Code] [Data] [Model]
   Haozhe Zhao*, **Zefan Cai***, ..., Zixuan Liu Sheng Wang, Wenjuan Han, Baobao Chang   **ICLR 2024, Poster**

2. Compositional Task Representations for Large Language Models   [PDF] [Code]
   Nan Shao*, **Zefan Cai***, Chonghua Liao, Yanan Zheng, and Zhilin Yang   **ICLR 2023 Poster**

3. SANTA: Separate Strategies for Inaccurate and Incomplete Annotation Noise in DS-NER   [PDF] [Code]
   Shuzheng Si*, **Zefan Cai***, Shuang Zeng, Guoqiang Feng, Jiaxing Lin, Baobao Chang   **ACL 2023**

4. Human-In-The-Loop through Chain-of-Thought   [PDF]
   Zefan Cai, Baobao Chang, Wenjuan Han   **ACL ARR 2023 October, Under Review**

5. DialogVCS: Robust Natural Language Understanding in Dialogue System Upgrade   [PDF]
   Zefan Cai*, Xin Zheng*, Tianyu Liu*, ..., Baobao Chang, Yunbo Cao   **ACL ARR 2023 October, Under Review**

6. SynGPT: Synthesizing high-order biological data points using GPT   [Code]
   Zixuan Liu∗, **Zefan Cai**∗, ..., Baobao Chang, Hoifung Poon, Sheng Wang   **RECOMB 2024, Under Review**

7. ML-Bench: LLMs Leverage Open-source Libraries for Machine Learning Tasks   [PDF] [Demo] [Code]
   Yuliang Liu*, Xiangru Tang*, **Zefan Cai***, ..., Baobao Chang, ..., Arman Cohan, Mark Gerstein   **Preprint**

8. DiffCap: Exploring Continuous Diffusion on Image Captioning   [PDF] [Code]
   Yufeng He*, **Zefan Cai***, Xu Gan, Baobao Chang   **Preprint**

9. VeCAF: VLM-empowered Collaborative Active Finetuning with Training Objective Awareness
   Rongyu Zhang*, **Zefan Cai***, ..., Kurt Keutzer, Baobao Chang, ..., Shanghang Zhang   **CVPR 2024, Under Review**

10. Mitigating Language Performance Disparity in mPLMs via Teacher Language Selection and Cross-lingual Distillation
    Haozhe Zhao*, **Zefan Cai***, Shuzheng Si, Liang Chen, ..., Baobao Chang   **ACL ARR 2023 October, Under Review**

11. Large Language Models are not Fair Evaluators   [PDF] [Code]
    Peiyi Wang, Lei Li, Liang Chen, **Zefan Cai**, ..., Qi Liu, Tianyu Liu, Zhifang Sui   **AAAI 2024, Under Review**

12. Towards End-to-End Embodied Decision Making with Multi-modal Large Language Model   [PDF] [Code]
    Liang Chen, ..., **Zefan Cai**, ..., Baobao Chang   **NeurIPS 2023 Foundation Models for Decision Making Workshop**

13. CENSOR: Distantly-Supervised Named Entity Recognition with Uncertainty-aware Teacher Learning and Student-student Collaborative Learning   [PDF]
    Helan Hu, Shuzheng Si, Haozhe Zhao, Shuang Zeng, Kaikai An, **Zefan Cai**, Baobao Chang   **Preprint**

RESEARCH EXPERIENCE

Peking University - Supervisor: Baobao Chang   Sep. 2022 – Present
MMICL: Empowering VLM With Multi-Modal In-Context Learning [6]

- Build novel architecture of Vision-Language Model (VLM) capable of **integrating image-text interleved data**.
- Construct MIC dataset containing **image-text interleved data and ICL data** based on existing datasets.
- Fine-tune MMICL VLM on MIC dataset and evaluate on MME bench, MM bench and other VQA benchmarks.

ML-Bench: LLMs Leverage Open-source Libraries for Machine Learning Tasks [7]

- Propose **ML-BENCH** to evaluate LLMs in **leveraging existing functions in open-source libraries**.
- Design data labeling principles, lead lablers to construct datasets, and conduct automatic execution examination.

- Design settings to accommodate different LLMs (i.e., LLMs and agents) and provide RAG setting.
- Design ML-Agent to select GiitHub, locate README file based on instruction, navigate codebase and output code.
- Fine-tune & evaluate open-source code LLMs, evaluate close-source LLMs and agents in ML-BENCH.

Human-In-The-Loop through Chain-of-Thought [3]

- Design human correction method for incorrect decomposed sub-question based on **question decomposition** by CoT.
- Design **Diversity-Entropy** metrics to locate hard decomposed sub-question for LLMs, conduct manual correction.
- Design **Cost-Utility Analysis Framework** to analyze utility and cost in the human-in-the-loop system.

Large language models are not fair evaluators [11]

- Conduct GPT-3.5&4 evaluation on LLM-generated responses and discover order bias of LLMs-as-evaluators.
- Implement CoT-based, majority-voting-based and order-based methods to perform aligned evaluation with human.
- Introduce **Diversity-Entropy** [3] to measure difficulty of each example in evaluation and seeks human assistance.

VeCAF: VLM-empowered Collaborative Active Finetuning with Training Objective Awareness [9]

- Optimize a parametric data selection model and incorporate training objective (i.e. loss) of models being tuned.
- Utilize **text embedding** of VLM to **augment image features of pre-trained vision model (PVM)** for data selection
- Implement data selection algorithm, cross-modal embedding distillation and fine-tune PVM.

Zero-Shot Event Extraction (EE) by following definition - Supervisor: Baobao Chang, Nanyun Peng

- Explore whether LLM-based Event Extraction models can **generalize to unseen events** by given event definitions.
- Evalaute **scaling law on EE** by training LLM in scales (i.e. # event type , # sample and # event definition).
- Fine-tune LLM with LLM-generated event types & event definitions & samples and evaluate in unseen event types.

SANTA: Separate Strategies for Inaccurate and Incomplete Annotation Noise in DS-NER [2]

- We propose the SANTA to handle two types of noise: inaccurate and incomplete annotation separately.
- Fine-tune span-based NER model and achieve a new SOTA on five public datasets.

DialogVCS: Robust Natural Language Understanding in Dialogue System Upgrade [4]

- Introduce new intent detection benchmark based on annotations in existing datasets.
- Implement focal-loss, negative sampling and multi-label cross entropy methods as baselines of the benchmark.

University of Washinton - Supervisor: Sheng Wang   March 2023 – Dec. 2023

SynGPT: Synthesizing high-order biological data points using GPT [5]

- Employ **in-context learning** to **synthesize high-order biomedical data** (i.e., combinatorial data of **drug combinations** and **trigenic interactions**) from GPT-4 and introduce **Diversity-Entropy** [3] to sample reliable data point.
- Develop **learning from the noisy label** approach by to exclude GPT-synthesized data that are different from real data.
- This work is under review in **RECOMB 2024** and the following work is prepared for **Nature Research journals**.

Tsinghua University - Supervisor: Zhilin Yang   May 2022 – Sep. 2022

Compositional Task Representations in Large Language Models [1]

- Fine-tune a T5 model connected with VQ-VAE architecture from scratch to implement **compositional generalization**.
- Conduct case studies and analysis to show the interpretability and controllability of the learned codebook.

INTERNSHIP

Microsoft - AZURE AI Team at Seattlle - Supervisor: Wen Xiao   Nov. 2023 – Present

- Construct dataset in **long-context image-text interleaved scenarios with multiple images** based on ARXIV paper.
- Conduct evaluation on VLMs in table understanding and diagram understanding

HONORS AND AWARDS

Merit Student (Top 10%), Peking University   Sept. 2023
Merit Student (Top 10%), Beijing Jiaotong University   Sept. 2020

ACADEMIC SERVICE

- Reviewer: EMNLP, ACL Rolling Review