XI NINGYUAN

derbachbuaa@gmail.com · (+86) 15365261570

EDUCATION BACKGROUND

Beihang University (BUAA) School of Computer Science and Engineering *2021.09 - 2025.06*

MAJOR: *Computer Science and Technology* *GPA: 89.37 / 100*

INTERNSHIP PROJECTS

AI ALGORITHM DEVELOPMENT INTERN, GEELY HOLDING GROUP *2024.01 - Present*

LLaMA-3 Chinese Ability Enhancement *2024.05 - 2024.09*

Aim to explore optimal strategies for enhancing LLMs' capabilities in unfamiliar languages through continual pre-training and supervised fine-tuning and improve their performances across domains.

Primary Responsibilities

- Led the continual pre-training (CPT) of Llama-3 models (8B and 70B) to enhance Chinese language capabilities and improve performance in areas such as mathematics, programming, and emotional intelligence, and fine-tuned the Additional Language Mix Ratio (ALMR) to balance language proficiency with task-specific performance.
- Examined the correlation between ALMR and Learning Rate (LR), and demonstrated that an optimal mix ratio not only improved task-specific performance but also accelerated loss convergence.
- Implemented Supervised Fine-Tuning (SFT) after CPT, and confirmed that the combined CPT-SFT approach improved emotional intelligence, outperforming the 70B model in both task accuracy and emotional understanding.

Research Output

https://arxiv.org/abs/2409.06624

- Deployed the 70B Llama-3 model in a chatbot system, and improved user interactions through advanced emotional intelligence to provide more nuanced and context-aware dialogues in Chinese.
- Achieved notable progress in Chinese benchmarks (C-Eval, LCSTS) and domain-specific tasks (GSM8K, HumanEval), and derived the optimal combination of ALMR and LR.

Brain-Inspired Human Thinking Modeling Using Dual-Layer SFT *2024.03 - Present*

Aim to develop a novel model architecture that integrates simultaneous reasoning and response generation to mimic human cognitive processes, enhancing performance across various NLP tasks.

Primary Responsibilities

- Developed the Thinking and Expression (TaS) model architecture, integrating simultaneous reasoning and response generation to enhance performance across various NLP tasks.
- Applied a dual-layer fine-tuning mechanism by employing SFT on both the intermediate reasoning layer and the speaking layer, enabling the model to learn intermediate reasoning steps and corresponding final responses.
- Designed a two-stage decoding process during inference, improving performance in tasks requiring multi-step reasoning, such as Theory of Mind (ToM) and open-domain dialogues.
- Trained the model using diverse datasets, including GPT-4 auto-generated data, rule-based annotations, and human annotations for thought content, enhancing the model's robust reasoning capabilities.

Research Output

https://arxiv.org/abs/2409.12059

- Achieved state-of-the-art results in Theory-of-Mind benchmarks (TOMI and BIGTOM) with a 98.7% score.
- Deployed the TaS model in a real-world conversational AI system, providing more contextually accurate and emotionally nuanced responses, showcasing its ability to handle complex, multi-step dialogue tasks.

Alleviating Hallucinations in Large Language Models with Skepticism Modeling *2024.06 - Present*

Aim to develop a Skepticism Modeling approach to mitigate hallucinations in LLMs by integrating uncertainty estimation and skeptical reasoning, enhancing accuracy and reliability.

Primary Responsibilities

- Developed the Skepticism Modeling (SM) approach to mitigate hallucinations in LLMs by integrating uncertainty estimation and skeptical reasoning.
- Enhanced LLMs' self-assessment capabilities through continual pre-training (CPT) with skepticism tokens, followed by supervised fine-tuning (SFT), resulting in improved accuracy and reliability.
- Implemented a self-evaluation mechanism, allowing models to express calibrated uncertainty, increasing trustworthiness in fields requiring high precision, such as healthcare and finance.
- Conducted experiments demonstrating SM's effectiveness in reducing hallucinations.

Research Output

https://arxiv.org/abs/2409.06601

- Achieved state-of-the-art results in various question-answering benchmarks.
- Published findings and released project code to support further research and development in skepticism-enhanced LLMs.

RESEARCH EXPERIENCES

Large Language Model-based Email Security Gateway *2023.10 - 2024.02*

BUAA National Key Laboratory for Software Development Environment *Supervisor: Prof. Zhang Hui*
Leader

- Developed an email security gateway within BUAA, and integrated a large language model to detect and block phishing and spam emails, thereby reducing telecom fraud and improving email security.
- Expanded the Llama-2 Chinese vocabulary by removing duplicate tokens, and attained a final token size of 49,953 for improved model performance in Chinese email detection.
- Applied LoRA for efficient fine-tunin, increased the number of trainable parameters, and utilized the original Stanford Alpaca templates for prompt design and prediction tasks. Achieved a three-fold increase in accuracy over the base model on a custom binary classification dataset.

BUAA-Haidian Public Security Bureau AI Application *2023.10 - 2024.01*

LLM Fine-tuning

- Trained the Baichuan-13B model for various public security scenarios, empowering future applications in law enforcement, safety promotion, and consultation services.
- Adopted prompt-based learning, and designed systematic prompt templates to enable efficient task execution.
- Leveraged Low-Rank Adaptation (LoRA) for parameter-efficient fine-tuning, and achieved approximately 30% improvement in accuracy compared to the base model. Enhanced model accuracy using Retrieval-Augmented Generation (RAG) to address hallucination issues and ensure knowledge freshness, and integrated ChatGPT-Embedding API for sentence-level vectorization and query-fused prompt generation, enhancing topic control such as case summary requests.

SysY to MIPS Compiler *2023.09- 2023.12*

Individual Project

- Designed a compiler translating SysY to LLVM IR and then to MIPS with more than 15,000 lines of code.
- Completed six iterations of development, utilized Git for version control, and used scripts for debugging in Ubuntu.

Fine-Grained Vehicle Recognition with Attention Mechanism *2022.11-2023.04*

Algorithm Developer

- Integrated the Convolutional Block Attention Module (CBAM) into EfficientNet, and applied dual-channel attention to extract key information from vehicle images more effectively.
- Expanded the HyperVID dataset threefold using a web crawler to increase the data volume. Achieved a Top-1 accuracy of 84.21% and Top-5 accuracy of 98.47%, representing improvements of 1.11% and 1.91% respectively compared to the baseline network.

♡ AWARDS & HONORS

First Prize, Higher Education Club Cup National Undergraduate Mathematical Modeling Contest *2023.12*
Provincial Second Prize, The 15th Chinese Mathematics Competitions *2023.11*
Second-Class Scholarship for Academic Excellence *2022.10*
Second-Class Scholarship for Discipline Competition *2022.10*
Second-Class Scholarship for Student Work *2022.10*

ADDITIONAL INFORMATION

- **Student Work:** Member of the Rights Department, BUAA Student Union; Staff, Student Union of the School of Computer Science and Engineering; Class Monitor;
- **Social Practice:** Volunteered in Longnan, Jiangxi Province for rural education (Second Prize Outstanding Summer Social Practice Team of BUAA).
- **Language Proficiency:** English (IELTS: 7.0 ; CET-4 & CET-6)
- **Coding Skills:** Python, Java, C, PyTorch, Git, Vue, React