Wen Cheng

wcheng@smail.nju.edu.cn · Nanjing University, Nanjing, China ·

Education Background

Nanjing University 2022 – Now

Master Computer Science *Admission with Exam Exemption*

Hefei University of Technology 2018 – 2022

Bachelor Computer Science and Technology *Rank #1 for Academical Recommendation*

Research Experience

USee: Ultrasound-based Device-free Eye Movement Sensing Under Review

***Wen Cheng**, Mingzhi Pang, Haoran Wan, Shichen Dong, Dongxu Liu, Wei Wang*

- Proposing USee, an innovative device-free system that utilizes ultrasound to sense subtle eye movements, specifically saccades.
- Introducing a novel signal processing pipeline that uncovers the relationship between micro-movements and signal decomposition residuals, a typically overlooked component, enabling direct extraction of eye movements.
- Validating the effectiveness of USee through comprehensive experiments conducted on COTS devices.

Security Attack on LLM-based Code Completion Tools arXiv preprint

***Wen Cheng**, Ke Sun, Xinyu Zhang, Wei Wang*

- Investigating the long-overlooked inherent security risks posed by LLM-based Code Completion Tools (LCCTs).
- Designing customized attack methodologies tailored to the unique workflows of LCCTs, achieving a 99.6% success rate in attacks on GitHub Copilot, a platform with over a million users, and successfully extracting sensitive user information, including 54 physical addresses and 314 email addresses. On Amazon Q, another popular LCCT, the attack success rate reaches 46.3%.
- Demonstrating that code-based attacks present severe threats to general-purpose LLMs, with high success rates against models in the GPT series, including the advanced GPT-4o.

QAQ: Quality Adaptive Quantization for LLM KV Cache arXiv preprint

*Shichen Dong, **Wen Cheng (co-first)**, Jiayu Qin, Wei Wang*

- Observing, for the first time, the differential sensitivity of Key cache and Value cache to quantization in LLMs, and proposing the core insight that outliers in the KV cache should be treated separately.
- Based on these insights, designing a dynamic adaptive quantization method that achieves a leading compression ratio.

W2KPE: Keyphrase Extraction with Word-Word Relation ICASSP 2023

***Wen Cheng**, Shichen Dong, Wei Wang*

- Developing a novel keyphrase extraction approach leveraging word-word relations to enhance extraction accuracy, and introducing techniques such as sentence fusion, keyphrase encoding, and a combined loss function.
- Achieving first place in the ICASSP 2023 MUG Challenge with this innovative methodology.

Work Experience

Microsoft Research Asia - Shanghai 2024.8 - Now

Research Internship Wireless Group

ByteDance Data Platform 2022.1 - 2022.5

Internship Software Engineer

Nanjing University 2023 Spring

Teaching Assistant Computer Architecture

Selected Awards

Nanjing University Distinguished Graduate Student	2023
Graduation with Honor: Excellent College Graduate of Anhui Province	2022
Undergraduate President Award **(Top 30 of 8000)**	2022
Undergraduate China National Scholarship **(Top 1%)**	2020, 2021
Provincial Second Prize, China Collegiate Programming Contest (CCPC)	2020, 2021
Meritorious Winner, MCM/ICM	2020

Others

- Personal website: https://sensente.github.io/