

Jingheng Ye (叶劲亨)

Email: yejh22@mails.tsinghua.edu.cn | Website: [Homepage](#), [Google Scholar](#), [Github](#)

Address: Shenzhen, Guangdong Province, China

EDUCATION

South China University of Technology

Bachelor of Computer Science and Technology

Guangzhou, China

September 2018 – June 2022

- GPA: 3.95 / 4.00 Rank: 1 / 160

Tsinghua University

Graduate Student of Computer Science

Shenzhen, China

September 2022 – June 2025 (expected)

- GPA: 3.93 / 4.00 Advisor: Prof. Hai-tao Zheng
- **Research Area:** Grammatical Error Correction (GEC), Education, Retrieval-Augmented Generation (RAG), LLM-based Agent

EXPERIENCE

Research Internship on Alibaba

Retrieval-Augmented Generation & LLM-based Agent

Institute for Intelligent Computing, Alibaba Group

December 2023 - Present

- Developing an e-commercial agent as a product buying assistant that helps to pinpoint users' demands.
- Investigating the research area of LLM-based Agent, including memory, reasoning, memory and tool using, etc.

Teaching Assistant of Knowledge Engineering

3-credit course

Tsinghua University

February 2024 - Present

- Hold weekly office hours for a graduate class on knowledge engineering.
- Grade assignments and a final project, answer questions, meet regularly with the professor.

Development Internship on Tencent

Database System

Database R&D Center Department, Tencent

July 2021 - September 2021

- Familiar with tencent database system TDSQL.
- Develop a tool to automatically test database compatibility.

PUBLICATIONS

(First Author) [CLEME: Debiasing Multi-reference Evaluation for Grammatical Error Correction](#) [EMNLP 2023]

- We point out that there are biases in the existing automatic evaluation metrics for grammar error correction (GEC) in multi-reference evaluation settings, and propose a multi-reference unbiased evaluation metric **CLEME**, which significantly improves the effectiveness of grammatical correction evaluation.

(First Author) [MixEdit: Revisiting Data Augmentation and Beyond for Grammatical Error Correction](#) [EMNLP 2023]

- We identifies two factors that affect the effectiveness of GEC data augmentation - Affinity and Diversity, and propose **MixEdit**, a data augmentation strategy that can significantly improve correction performance.

(Co-Author) [A Frustratingly Easy Plug-and-Play Detection-and-Reasoning Module for Chinese Spelling Check](#) [EMNLP 2023]

- We propose a plug and play "**detection and inference**" model for Chinese spelling correction (CSC) tasks, which decomposes the task into three progressive steps: detection, inference, and correction.

(First Author) [ProductAgent: Benchmarking Conversational Product Search Agent with Asking Clarification Questions](#) [Under Review]

- We propose **ProductAgent**, a conversational information seeking agent equipped with abilities of strategic clarification question generation and dynamic product retrieval. Furthermore, we propose the benchmark called PROCLARE to evaluate the agent's performance both automatically and qualitatively with the aid of a LLM-driven user simulator.

(First Author) [EXCGEC: A Benchmark of Edit-wise Explainable Chinese Grammatical Error Correction](#) [Under Review]

- We introduce the task of **EXplainable GEC (EXGEC)** focusing on the integral role of both correction and explanation tasks, and propose EXCGEC, a tailored benchmark for Chinese EXGEC consisting of 8,216 explanation-augmented samples.

(First Author) [CLEME2.0: Towards More Interpretable Evaluation by Disentangling Edits for Grammatical Error Correction](#) [Under Review]

- We propose **CLEME2.0**, a reference-based evaluation strategy that can describe four elementary dimensions of GEC systems, namely hit-correction, error-correction, under-correction, and over-correction.

(First Author) [Corrections Meet Explanations: A Unified Framework for Explainable Grammatical Error Correction](#) [Under Review]

- We introduce a unified explainable GEC framework that jointly perform explanation and correction tasks in a sequence-to-sequence generation manner, hypothesizing both tasks would benefit each other.

RECENT SCHOLASTIC AWARDS

CCL23-Eval CLTC Track-1, Second Prize

August, 2023

Chinese Learner Text Correction, CLTC

- CCL2023-Eval Task7 CLTC (Chinese Learner Text Correction) Track 1: Multidimensional Chinese Learner Text Correction. The resultant technical report is published in China National Conference on Computational Linguistics (CCL) venue, [available here](#).
- As the **team captain**, I am responsible for data curation, model and algorithm design, model training, report rewriting, etc.

Geneva International Exhibition of Inventions, Silver Prize

April, 2023

SmartCorrect: Multilingual and Multi-Scenario Intelligent Text Error Correction Platform

- As the second author (my mentor as the first author), I am mainly responsible for algorithm implementation, platform design, and work scheduling. I learned to effectively cooperate with team member to build up a comprehensive correction platform.

CCL22-Eval CLTC Task-4, First Prize

October, 2022

Chinese Learner Text Correction, CLTC

- CCL2022-Eval Task CLTC (Chinese Learner Text Correction) Track 4: Multi-reference Multi-source Chinese Learner Text Correction. The technical report is published in CCL venue, [available here](#).
- As the **team captain**, I am responsible for data curation, model and algorithm design, model training, report rewriting, etc.

The Mathematical Contest in Modeling (MCM), Meritorious Winner

February, 2021

National College Student Mathematics Competition, Guangdong Division, Third Prize

December, 2020

RECOGNITIONS

National Scholarship

Sep, 2019

- The highest honor that students in higher education institutions can receive in China, with an award rate of only 0.2%.

National Encouragement Scholarship

Sep, 2020

Merit Student of South China University of Technology

Sep, 2020

Fist class scholarship at the school level

Sep, 2021

First class special scholarship at the college level

Sep, 2023