

# Mr. Shitou ZHANG

D.O.B. Sep. 20, 2000 | Tel: (+86) 15110268753 / (+852) 67443439

[stzhang@link.cuhk.edu.hk](mailto:stzhang@link.cuhk.edu.hk) | <https://stzhang-patrick.github.io/>

## EDUCATION

---

**The Chinese University of Hong Kong (CUHK)**

*Sep. 2022 – Jun. 2023*

*Master of Science in Information Engineering*

**Wuhan University (WHU)**

*Sep. 2018 – Jun. 2022*

*Bachelor of Science in Archival Science (Information Management)*

## WORK EXPERIENCE

---

**Research Intern, DeepLang Technology Co. Ltd.**

*Mentor: Dr. Gang Chen*

*Sep. 2023 – Present*

**Duties included:** Mixture of Experts (MoE) pretraining

**Technical Consultant, FiduciaEdge Technologies Co. Ltd., Hong Kong**

*Supervisor: Prof. John K. Zao*

*Jul. 2023 – Aug. 2023*

**Duties included:** Federated learning system implementation and deployment

**Research Assistant, Key Laboratory of Archival Intelligent Development and Service, NAAC**

*Supervisor: Prof. Ping Wang*

*Jun. 2023 – Jul. 2023*

**Duties included:** Domain-oriented LLM development and deployment

**Research Assistant, Wuhan University**

*Supervisor: Dr. Zuchao Li*

*Mar. 2023 – Jun. 2023*

**Duties included:** LLM pre-training

## PROJECT EXPERIENCE

---

### ● LingoWhale

*Sep. 2023 – Present*

**Description:** Developed [LingoWhale](#), a series of general backbone LLMs with a focus on Chinese language processing; Achieved leading results in LLM benchmarks, including [MMLU](#) and [C-eval](#); Open-sourced LingoWhale-8B, available on [HuggingFace](#).

### ● BATGPT

*Mar. 2023 – Present*

**Description:** Pre-trained [BATGPT](#), a bilingual LLM available in various sizes: 15B, 30B, and 70B; Deployed as the foundation model of [Powthink](#)'s services in production environment; Released the 15B v1.0 version for public use, available on [HuggingFace](#).

### ● ArcGPT

*Jun. 2023 – Sep. 2023*

**Description:** Led the development of ArcGPT, a domain-specific LLM optimized for archival applications.

### ● FL on fECP

*Jul. 2023 – Aug. 2023*

**Description:** Re-implemented and optimized [ClusterFL](#) using Flower; Deployed on Vecow Vhub (arm-based edge AI computing platform); Integrated FL systems with fECP (Trusted Execution Environment).

## PUBLICATIONS

---

- [1] [Shitou Zhang](#), Ping Wang\*, Zuchao Li\*, Jingrui Hou, Qibiao Hu. “**Confidence-based Syntax Encoding Network for Better Ancient Chinese Understanding**”. Accepted by Information Processing & Management (IPM), to appear.

- [2] Shitou Zhang, Zuchao Li, Xingshen Liu, Liming Yang, Ping Wang. “**ArcMMLU: A Library and Information Science Benchmark for Large Language Models**”. [2311.18658.pdf \(arxiv.org\)](#)
- [3] Shitou Zhang, Siyuan Peng, Ping Wang, Jingrui Hou. “**Archives Meet GPT: A Pilot Study on Enhancing Archival Workflows with Large Language Models**”. Accepted by iConference 2024, to appear.
- [4] Shitou Zhang, Jingrui Hou, Siyuan Peng, Zuchao Li, Qibiao Hu, Ping Wang. “**ArcGPT: A Large Language Model Tailored for Real-world Archival Applications**”. [2307.14852.pdf \(arxiv.org\)](#)
- [5] Zuchao Li, Shitou Zhang, Hai Zhao\*, Yifei Yang, Dongjie Yang. “**BATGPT: A Bidirectional Autoregressive Talker from Generative Pre-trained Transformer**”. [2307.00360.pdf \(arxiv.org\)](#)
- [6] Shitou Zhang, Ping Wang\*, Zuchao Li\*, Jingrui Hou. “**Enhancing Ancient Chinese Understanding with Derived Noisy Syntax Trees**”. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop). [2023.acl-srw.15.pdf \(aclanthology.org\)](#)

## MANUSCRIPTS

---

- [1] Jiajia Li, Letian Peng, Shitou Zhang, Ping Wang, Zuchao Li, Hai Zhao. “**Semantics-Preserved Distortion for Personal Privacy Protection in Information Management**”. *Expert Systems with Applications (ESWA)*, Under review.

*Latest update on 9<sup>th</sup> Dec. 2023*