# ZHANLI LI (李展利)

(+86) 156-2350-1965 ◇ lizhanli@stu.zuel.edu.cn ◇ Zhanli-Li.github.io ◇ Zhanli-Li (150+ stars)

## EDUCATION

**Zhongnan University of Economics and Law** *September 2023 - June 2027*
Wenlan School of Business (Honors College), B.Sc. in Digital Economy
**Weighted Avg: 93.73/100 Rank: 2/80 2025 National Scholarship**
**Research Interests:** LLMs, Agentic Training, Deep Learning, Causal Inference

## PUBLICATIONS

**Zhanli Li**, Huiwen Tian, Lvzhou Luo, Yixuan Cao, Ping Luo. *DeepRead: Document Structure-Aware Reasoning to Enhance Agentic Search. KDD 2026* (CCF-A), **Under Review**. [**First Author**]

**Zhanli Li**, Yixuan Cao, Lvzhou Luo, Ping Luo. *Navigating Large-Scale Document Collections: MuDABench for Multi-Document Analytical QA. ACL 2026 Findings* (CCF-A). [**First Author**]

**Zhanli Li**, Zichao Yang. *ESG Rating Disagreement and Corporate Total Factor Productivity: Inference and Prediction. Finance Research Letters*, 2025. (CAS Q1 Top, JCR Q1, IF: 6.9) [**First Author**]

## RESEARCH EXPERIENCE

**Vision-Informed Agentic Training for Data Science**‡ *March 2026 – Present*
**Role:** Project Leader
To address the demands of **multimodal** data analysis in real-world scenarios, we decompose data science agent tasks into four hierarchical levels: **descriptive**, **diagnostic**, **predictive**, and **prescriptive**. A progressive training environment is built around curriculum learning, enabling models to transition smoothly from basic statistical cognition to complex reasoning and decision generation. The project focuses on **process supervision** strategies, particularly the suppression of **overlap** between visual and textual information to reduce redundant perception and alleviate ineffective reasoning paths, thereby fostering a "**less is more**" learning paradigm in multimodal data science.
▷ Selected for the **Zhongguancun Academy Shenlan Program**; intended submission to ICLR 2026.

**DeepRead: Document Structure-Aware Reasoning to Enhance Agentic Search**† *October 2025 – February 2026*
**Role:** Project Leader
Existing agentic search frameworks often treat long documents as flat chunk collections, ignoring native **hierarchical structures** and **sequential logic**. We propose **DeepRead**, a structure-aware document reasoning agent that preserves layout fidelity via OCR, constructs **paragraph-level coordinates**, and equips the LLM with two synergistic tools: **Retrieve** and **ReadSection**. This induces an emergent "locate-then-read" reasoning paradigm, significantly mitigating context fragmentation issues inherent in conventional retrieval. Across four benchmarks covering diverse document types, DeepRead outperforms **Search-o1** style agentic search baselines by an average of **10.3%**.
▷ **Submitted to** *KDD 2026 (CCF-A)*. The preprint was reported by the prominent tech media **New Wisdom (新智元)** and received over 130 saves on Xiaohongshu.

**MuDABench: Multi-Document Analytical QA Benchmark and Multi-Agent Framework**† *April 2025 – December 2025*
**Role:** Project Leader
**Large-scale** document collections contain rich knowledge, yet scalable multi-document analytical QA across **thousands** of long documents remains challenging. We constructed one of the **largest** multi-document analytical QA benchmarks to date (**80k+** pages, **332** analytical questions, **4,964** structured intermediate facts) using **distant supervision** and expert verification, enabling systematic evaluation of multi-document retrieval, cross-document aggregation, and numerical reasoning. Experiments reveal that standard retrieval

services struggle in this setting: OpenAI's **File Search** achieves a best accuracy of only **13.68%**. Our proposed **Multi-Agent** workflow raises this to **26.51%**. Further error analysis indicates that the primary bottleneck lies in the stability and high-precision localization of **single-document information extraction**.
▷ **Published in** *ACL 2026 Findings (CCF-A).*

**Causal Inference Framework via SHapley Additive exPlanations in Corporate Finance**[*] *September 2024 – January 2025*
**Role:** Project Leader
Traditional panel data models rely heavily on linear assumptions and struggle to capture complex nonlinear dynamics. We address this by encoding panel structure to enhance model fitting capacity while preserving **causal interpretability**. We developed an interpretable modeling framework integrated with **Optuna** for automatic hyperparameter tuning, achieving an $R^2$ of 0.76 in TFP prediction—substantially outperforming linear baselines. We further employed **SHapley Additive exPlanations (SHAP)** to trace mechanisms within the panel data, quantifying and decomposing the **nonlinear marginal contributions** of ESG rating disagreement, thereby uncovering latent causal pathways.
▷ **Published in** *Finance Research Letters* (CAS Q1 Top, JCR Q1, IF: 6.9). Selected as **Outstanding Paper at the Tsinghua University Causal Inference Seminar (Top 3%)**.

[†] *Conducted under the supervision of Associate Professors Yixuan Cao and Ping Luo, Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences*

[‡] *Conducted under the supervision of Assistant Professor Wentao Zhang, Peking University & Zhongguancun Academy*

[*] *Conducted under the supervision of Assistant Professor Zichao Yang, Wenlan School of Business, Zhongnan University of Economics and Law*

## INTERNSHIP EXPERIENCE

**Beijing Paoding Technology Co., Ltd. —Research Dept.** *June 2025 – February 2026*
**Role:** AI Research Intern **Mentors:** Ping Luo, Yixuan Cao
**Project:** Improving Context Retrieval in Production-Grade RAG Systems

Leveraged the company's proprietary **document parsing** model PDFlux and integrated its structured outputs (e.g., layout-aware paragraphs/sections) into the commercial RAG product ChatDOC, focusing on mitigating **context-missing** issues when using rerankers. The system has been deployed in production environments of several top-tier financial institutions, markedly enhancing operational efficiency.

## SELECTED AWARDS

**China National Scholarship** *November 2025*
Highest academic honor for undergraduates in China (Top 0.2%).

**China Undergraduate Mathematical Contest in Modeling** *November 2025*
First Prize (Team Leader, Top 3%).

**Hunan Province Outstanding Student (High School)** *April 2023*
Awarded to the top 0.1% of high school students in Hunan Province.

## TECHNICAL SKILLS & SERVICE

**Languages:** Chinese (native), English (CET-6: 508)
**Programming Languages:** Python, C/C++, LaTeX, Markdown
**Tools:** Docker, SSH, tmux, Ubuntu
**Libraries & Frameworks:** ms-swift, vllm, verl, llamaindex, transformers, torch, sklearn, pandas, numpy
**Academic Service:** Independent Reviewer for *Finance Research Letters*