Zifeng Ding

Tel: +49-(0)1751934698
Email: zd320@cam.ac.uk
Personal Site: https://zifengding.github.io/

Education

Ludwig Maximilian University of Munich, Ph.D. in Computer Science — Jun 2021 – Feb 2025

Technical University of Munich, MSc in Electrical and Computer Engineering — Oct 2018 - Mar 2021

East China Normal University, BEng in Electrical Engineering — Sept 2014 - Jun 2018

Professional & Academic Experience

University of Cambridge, Research Associate (Advisor: Prof. Andreas Vlachos) — Oct 2024 – present

Working at Cambridge NLIP Group on three main topics: (1) Automated fact-checking to counter hallucination and misinformation, including the creation of a real-world dataset of 1,297 image-text claims annotated with question-answer（QA) pairs and web-based evidence, developed with human annotators. (2) Temporal reasoning with large language models (LLMs), proposing benchmarks to evaluate LLMs' capabilities in this area. The first benchmark, FOReCAst, is based on 2,256 real-world forecasting tasks from Metaculus and evaluates LLMs' forecasting ability by considering both prediction accuracy and confidence. The second benchmark, TCP, is synthetically generated to evaluate planning under interdependent temporal constraints and is built on a data generation pipeline that combines code and LLMs to produce correctly labeled, distribution-controlled data, while remaining easily extensible to larger scales. (3) Temporal graph reasoning with LLMs by fine-tuning them through reinforcement learning (RL) with customized outcome-based reward to capture temporal evolution patterns, resulting in an explainable graph forecaster that adapts to diverse temporal graphs, including those unseen during training. The fine-tuned 4B model shows around 45% relative improvement compared with GPT-5 Mini, delivering accurate prediction and coherent reasoning traces. Additionally proposed an LLM-as-a-Judge system, validated by human annotators, which measures reasoning quality and quantifies hallucinations of LLMs in temporal graph reasoning.

CAMEL-AI.org & Eigent AI, Visiting Research Scientist — Mar 2025 - present

(1) Leading the Loong project (open-sourced at: https://github.com/camel-ai/loong), focused on building an agentic synthetic data generation environment for LLM post-training using Python-based verifiers and LLM agents. Created a high-quality seed dataset of 8,729 examples across 12 domains (e.g., logic, graph theory), each paired with executable solution code and semantically verified answers. Developed a few-shot prompting pipeline to generate diverse, verifiable QA pairs based on seed data. Currently fine-tuning LLMs with RL on the generated data to showcase the potential of RL with verifier feedback. (2) Designed a multi-agent framework to solve math problems and generate synthetic reasoning traces in the form of long, purely language-based chains of thought. Open-sourced the framework (https://github.com/camel-ai/magenta) and fine-tuned LLMs on the generated traces, resulting in improved performance on competition-level math benchmarks.

University of Oxford, Visiting Researcher (Supervisor: Prof. Michael Bronstein) — Apr 2024 – Oct 2024

Worked on temporal graph representation learning using state space models; proposed a method that achieved state-of-the-art performance while reducing inference time by up to 40% and GPU memory usage by up to 45% compared to prior approaches.

Ludwig Maximilian University of Munich, Ph.D. Student (Supervisor: Prof. Volker Tresp) — Jun 2021 – Feb 2025

Worked on temporal knowledge graphs (TKGs) and natural language processing, including: (1) initiating the first studies on inductive knowledge representation learning for TKGs, leveraging meta-learning and LLMs; and (2) developing the first question answering framework for forecasting over TKGs by combining language models with graph representation learning to answer natural language questions about future events.

Siemens AG, Ph.D. Student — Jun 2021 – Aug 2024

Responsible for applying graph machine learning and LLMs to industrial use cases, with a focus on identifying critical supply chains in supply chain knowledge graphs (KGs). Designed a retrieval-augmented generation (RAG) system that engages in multi-turn dialogue to retrieve critical information from the KG and generate answers based on the retrieved content. The system has been successfully integrated into the company's internal workflow.

Selected First Author Publications

Full Publication List: https://scholar.google.com/citations?user=8RapuD4AAAAJ&hl=en

- Ding, Z., Huang, S., Cao, Z., Kondrup, E., Yang, Z., Huang, X., Sui, Y., Yuan, Z., Zhu, Y., Hu, X., Poursafaei, F., Bronstein, M., Vlachos, A., **Self-Exploring Language Models for Explainable Link Forecasting on Temporal Graphs via Reinforcement Learning**, under review (2025).
- Wang, W. *, Ding, Z.*, Gu, J., Cao, R., Meinel, C., de Melo, G., Yang, H, **Image Tokens Matter: Mitigating Hallucination in Discrete Tokenizer-based Large Vision-Language Models via Latent Editing**, NeurIPS (2025). *Equal contribution.
- Huang, X. *, Hu, X. *, Ding, Z.*, He, Y., Alzarooni, W., Ye, Z., Fan, W., He, B., Bo, H., Hu, C., Li, G., **Distilling Tool Knowledge into Language Models via Back-Translated Traces**, ICML 2025 Workshop on Multi-Agent Systems (2025), *Equal contribution.
- Ding, Z., Yan, S., Yuan, Z., Hu, X., Lin, F., Vlachos, A., **TCP: a Benchmark for Temporal Constraint-Based Planning**, EMNLP (2025).
- He, Y. *, He, B. *, Ding, Z.*, Lupidi, A., Zhu, Y., Chen, S., Zhang, C., Chen, J., Ma, Y., Tresp, V., Horrocks, I., **Supposedly Equivalent Facts That Aren't? Entity Frequency in Pre-training Induces Asymmetry in LLMs**, COLM (2025), *Equal contribution.
- Ding, Z., Li, Y., He, Y., Norelli, A., Wu, J., Tresp, V., Bronstein, M., Ma, Y., **DyGMamba: Efficiently Modeling Long-Term Temporal Dependency on Continuous-Time Dynamic Graphs with State Space Models**, TMLR (2025).
- Ding, Z., Wu, J., Wu, J., Xia, Y., Xiong, B., Tresp, V., **Temporal Fact Reasoning over Hyper-Relational Knowledge Graphs**, EMNLP (2024).
- Ding, Z., Cai, H., Wu, J., Ma, Y., Liao, R., Xiong, B., Tresp, V., **zrLLM: Zero-Shot Relational Learning on Temporal Knowledge Graphs with Large Language Models**, NAACL (2024) **Oral**.
- Ding, Z., Qi, R., Li, Z., He, B., Wu, J., Ma, Y., Meng, Z., Chen, S., Liao, R., Han, Z., Tresp, V., **ForecastTKGQuestions: A Benchmark for Temporal Question Answering and Forecasting over Temporal Knowledge Graph**s, ISWC (2023).
- Han, Z.*, Ding, Z.*, Ma, Y., Gu, Y., Tresp, V., **Learning Neural Ordinary Equations for Forecasting Future Links on Temporal Knowledge Graphs**, EMNLP (2021), *Equal contribution.

Skills

Technical skills: proficient in Python, PyTorch, Linux; working knowledge of C/C++ and git.

Language skills: Chinese (Native), English (Full Professional), German (Limited Working).

Honors & Awards

European Network of AI Excellence Centres (ELISE) Scholarship Apr 2024

Honorable Mention of Automated Knowledge Base Construction (AKBC) 2022 Nov 2022

Community Services

Organizer of the 8th and 9th Fact Extraction and VERification (FEVER) workshop @ ACL 2025 and EACL 2026.
Area Chair of ACL Rollong Review Oct, 2025, the Temporal Graph Learning Workshop @ KDD 2025, Scaling Environments for Agents (SEA) Workshop @ NeuIPS 2025.
Reviewer of ACL Rolling Review, NeurIPS 2023-2025, ICLR 2025, COLM 2025, ICML 2023.