

Bosheng Ding (Singapore Permanent Resident)

Phone: +65-97288446

Website: boshengding.com

Email: bosheng001@ntu.edu.sg

Research Interests **Natural Language Processing (NLP): Large Language Models, Human-AI Interactions, Multilingual NLP, Data Efficient Methods**

My research interests are deeply rooted in the exploration and advancement of **Natural Language Processing (NLP)**, with a particular focus on the development and application of **Large Language Models (LLMs)**. I am fascinated by the transformative potential of LLMs in understanding and generating human language, aiming to push the boundaries of how these models facilitate more intuitive and meaningful **Human-AI interactions**. Recognizing the global nature of communication, I am also committed to expanding the capabilities of **Multilingual NLP**, striving to create systems that can seamlessly operate across different languages and cultural contexts. Furthermore, acknowledging the limitations of data availability in many scenarios, I am dedicated to pioneering **Data Efficient Methods** that enable these sophisticated models to learn effectively from limited datasets. Through my work, I aspire to not only advance the technical excellence of NLP technologies but also to ensure they are accessible, equitable, and beneficial across diverse global communities.

Education **Nanyang Technological University (NTU)** Jan 2020 - Jan 2024

Ph.D. Candidate in Computer Science, with NTU-Alibaba Joint PhD Full Scholarship

Supervisor: **Prof. Shafiq Joty & Prof. Luu Anh Tuan**

Thesis: “Advancing Low Resource Information Extraction and Dialogue System using Data Efficient Methods”

National University of Singapore (NUS) Jan 2021 - June 2021

PhD Exchange Programme

Nanyang Technological University (NTU) Aug 2012 - June 2016

*B.Eng. in Electrical and Electronic Engineering (**Specialization: Computer Engineering**), with Singapore Government Full Scholarship*

Publications

Published Papers¹

1. Is GPT-3 a Good Data Annotator? (ACL 2023)
Bosheng Ding*, Chengwei Qin*, Linlin Liu, Yew Ken Chia, Boyang Li, Shafiq Joty, and Lidong Bing.
2. GlobalWoZ: Globalizing MultiWoZ to Develop Multilingual Task-Oriented Dialogue Systems. (ACL 2022)
Bosheng Ding, Junjie Hu, Lidong Bing, Mahani Aljunied, Shafiq Joty, Luo Si, and Chunyan Miao.
3. MulDA: A Multilingual Data Augmentation Framework for Low-Resource Cross-Lingual NER. (ACL 2021)
Linlin Liu*, **Bosheng Ding***, Lidong Bing, Shafiq Joty, Luo Si, and Chunyan Miao.
4. DAGA: Data Augmentation with a Generation Approach for Low-resource Tagging Tasks. (EMNLP 2020)
Bosheng Ding*, Linlin Liu*, Lidong Bing, Canasai Kruengkrai, Thien Hai Nguyen, Shafiq Joty, Luo Si, and Chunyan Miao.
5. Chain of Knowledge: A Framework for Grounding Large Language Models with Structured Knowledge Bases. (ICLR 2024)

¹The superscript * indicates co-first authors.

Xingxuan Li, Ruochen Zhao, Yew Ken Chia, **Bosheng Ding**, Lidong Bing, Shafiq Joty, Soujanya Poria

6. Retrieving Multimodal Information for Augmented Generation: A Survey. (EMNLP 2023)
Ruochen Zhao, Hailin Chen, Weishi Wang, Fangkai Jiao, Do Long, Chengwei Qin, **Bosheng Ding**, Xiaobao Guo, Minzhi Li, Xingxuan Li, and Shafiq Joty.
7. On the Effectiveness of Adapter-based Tuning for Pretrained Language Model Adaptation. (ACL 2021)
Ruidan He, Linlin Liu, Hai Ye, Qingyu Tan, **Bosheng Ding**, Liying Cheng, Jiawei Low, Lidong Bing, and Luo Si.
8. Unravelling the landscape of Large Language Models: A systematic review and future perspectives. (Journal of Electronic Business & Digital Economics, forthcoming.)
Qinxu Ding, Ding Ding, Yue Wang, Chong Guan, **Bosheng Ding**

Preprint

1. LogicLLM: Exploring Self-supervised Logic-enhanced Training for Large Language Models. (Under review)
Fangkai Jiao, Zhiyang Teng, Shafiq Joty, **Bosheng Ding**, Aixin Sun, Zhengyuan Liu, Nancy F Chen
2. Panda LLM: Training Data and Evaluation for Open-Sourced Chinese Instruction-Following Large Language Models. (Technical Report)
Fangkai Jiao*, **Bosheng Ding***, Tianze Luo*, Zhanfeng Mo*

Industry Experiences

Alibaba Group (Singapore) Jan 2020 - Jan 2024
NTU-Alibaba Joint Ph.D Programme

- Conducted research on algorithms with the aim of enhancing **Named Entity Recognition (NER)** efficiency within the realm of **E-Commerce operations**, applicable to both single and multiple language contexts. This resulted in an average improvement of 1-3% in the F1-score, particularly in newly emerging product categories.
- Pioneered the exploration of **Large Language Models** within our research team, innovating new methods for utilizing these models in the annotation of multi-class data tasks.
- Led the initiative to create the first extensive **multilingual task-oriented dialogue dataset**, designed for global application. This dataset was developed with special attention to local context and entities in the context of globalization, ensuring real-world applicability across 20 languages.
- Mentored junior doctoral candidates in scholarly research while establishing collaborative relationships with eminent scholars from leading research universities, including Carnegie Mellon University, Tsinghua University, University of Michigan and University of Wisconsin-Madison, among others.

PwC Consulting (Singapore) Nov 2017 - Nov 2019
Consultant

- Developed sophisticated **financial risk** and **economic projection models**, providing key analytical perspectives to prominent banks and financial entities across Singapore, the Philippines, Malaysia, Thailand, Indonesia, Australia, and South Africa.

- Independently engineered complex multi-variable regression models and performed rigorous statistical analyses to validate the findings.
- Built solid finance domain knowledge and Engaged in a collaborative effort with economists, data scientists, and finance professionals to offer consultative support on risk management compliant with the IFRS 9 framework.

Phillip Securities Pte Ltd (Singapore)
Investment Analyst

Aug 2016 - Nov 2017

- Independently secured new clients, successfully accumulating over SGD 1.3 million in Assets Under Management (AUM) within an 8-month period.
- Strategically gathered and analyzed market data and intelligence from diverse sources and channels.
- Diligently monitored client portfolios and maintained close follow-up with clients to discuss performance outcomes.

Product Experiences

CapGo (Singapore)
Key contributor

Feb 2023 - Dec 2023

- Pioneered the development of groundbreaking algorithms that leverage Large Language Models (LLMs) and Retrieval Augmented Generation (RAG), enhancing the efficiency of market research processes in Google Sheets by 10 times.
- Conceptualized and architected the entire product framework, leading to a strategic pivot from an investment banking tool to a sophisticated market research platform.
- Recognized as the "Product of the Day" on Product Hunt. For more details, refer to [this link](#).
- For additional information, please visit our website at capgo.ai.

Open Source Project

PandaLLM (Open-sourced Large Language Models for Chinese)
Project Leader

May 2023 - Present

- Pioneered the development and open-sourcing of a comprehensive pipeline for Large Language Models, encompassing training, fine-tuning, and inference processes.
- Our team is a trailblazer in the international arena for the continuous pre-training and fine-tuning of the LLaMA series, specifically for Chinese, Legal, and Code Generation domains. Notable projects include *Panda-LLaMA2-13B-Chat*, *Legal-Panda-13B-Chat*, and *Code-Panda-13B-Python*.
- Achieved over 1000 stars on GitHub. Explore our repository at [here](#).

Teaching & Mentoring Experiences

Singapore University of Social Sciences
Associate Lecturer

CET175 Introduction to Generative AI

June 2023 - Present

MKT365 Social Media Metrics & Analytics

Mar 2023 - Present

NTU School of Computer Science and Engineering

Teaching Assistant

AI6127 Deep Neural Networks for NLP

Jan 2022 - May 2022

CE7455 Deep Learning for Natural Language Processing

Jan 2022 - May 2022

MH6812 Advanced NLP with Deep Learning

Jan 2022 - May 2022

Mentored Students: Chengwei Qin (PhD student), Xinze Li (PhD student), Zhanfeng Mo (PhD student), Jiayi Liu (Msc student)

Honors & Awards **NTU-Alibaba Joint PhD Scholarship** (2020 - 2023), Singapore Economic Development Board & Alibaba Group & Nanyang Technological University

2014 NTU President Research Scholar, Nanyang Technological University

Singapore Government Full Scholarship (2011 - 2016), Singapore Ministry of Education.

Professional Service

Reviewer for Conferences

Empirical Methods in Natural Language Processing (EMNLP)

Annual Meeting of the Association for Computational Linguistics (ACL)

ACM International Conference on Web Search and Data Mining (WSDM)

ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)

North American Chapter of the Association for Computational Linguistics (NAACL)

Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)

Programming Languages

Python (Transformers, Pytorch), SQL, Latex, Bootstrap

Languages

- **Mandarin, Teochew:** Native
- **English, Cantonese:** Bilingual proficiency

References

Shafiq Joty

Research Director

Salesforce AI Research

Associate Professor (On leave)

School of Computer Science and Engineering, Nanyang Technological University

sjoty@salesforce.com

Junjie Hu

Assistant Professor

Department of Biostatistics, Department of Computer Science and Data Science Institute, University of Wisconsin-Madison

junjie.hu@wisc.edu

Luu Anh Tuan

Assistant Professor

School of Computer Science and Engineering, Nanyang Technological University

anhluan.luu@ntu.edu.sg

Hobbies & Interests

Embarking on a global journey of flavors, melodies, and underwater adventures, my life is a vibrant tapestry woven from diverse interests and passions. A conversationalist at heart, I thrive on the rich tapestry of human stories, finding joy in the unique narratives each person brings to the table. With stamps from 22 countries in my passport, my love for travel is more than a hobby; it's a quest for cultural immersion and culinary exploration. Speaking of which, my prowess in whipping up authentic Chinese cuisine is matched only by my enthusiasm for discovering the world's most intriguing cocktail bars, savoring creations that are as much art as they are beverages. My musical tastes are an eclectic mix, resonating with the legendary anthems of Queen, the soul-stirring melodies of Coldplay, and the distinctive rhythms of Eason Chan, creating a soundtrack to my life's journey. The melodious strings of the Erhu have been my companions since secondary school, adding a touch of traditional elegance to my repertoire. As an avid reader, I dive into the depths of literature with the same fervor I exhibit as an Advanced Open Water Diver, exploring the mysteries beneath the waves. My life is an ongoing adventure, a blend of flavors, sounds, and sights, each experience enriching my soul and broadening my horizon.