JUNWEI DENG

junweid2@illinois.edu — 217-974-6501 — theaperdeng.github.io — Google Scholar

EXPERIENCE

University of Illinois Urbana-Champaign, United States
Ph.D. Candidate in Information Sciences *2023.8 - now*
Advisor: Prof. Jiaqi Ma

Intel, P.R. China
Machine Learning Engineer *2021.5 - 2023.7*

University of Michigan, United States
M.S. in Information *2019.9 - 2021.4*

Shanghai Jiao Tong University, P.R. China
B.S.E. in Electrical and Computer Engineering *2016.9 - 2020.8*

RESEARCH INTERESTS

Data-centric AI
Data attribution, data curation/selection, and data compensation for LLMs, recommender systems, time series modeling, and etc.

AI Toolkit Development
Deliver the innovation and facilitate the industrial and research communities.

SELECTED PAPERS

J. Deng, W. Tang, J. Ma. *Versatile Influence Function for Data Attribution with Non-Decomposable Loss.* **ICML 2025**.

J. Deng*, T. Li*, S. Zhang, S. Liu, Y. Pan, H. Huang, X. Wang, P. Hu, X. Zhang, J. Ma. *`dattri`: A Library for Efficient Data Attribution.* **NeurIPS 2024 (Spotlight)**.

J. Deng, S. Zhang, J. Ma. *Computational Copyright: Towards A Royalty Model for Music Generative AI.* **DPFM Workshop at ICLR 2024 (Best Paper Award)**.

J. Ma*, J. Deng*, Q. Mei. *Subgroup Generalization and Fairness of Graph Neural Networks.* **NeurIPS 2021 (Spotlight)**.

J. Ma*, J. Deng*, Q. Mei. *Adversarial Attack on Graph Neural Networks as An Influence Maximization Problem.* **WSDM 2022**.

SELECTED PREPRINTS

J. Deng*, T. Li*, S. Zhang, J. Ma. *Efficient Ensembles Improve Training Data Attribution.* **In Submission**.

WORK EXPERIENCE

Microsoft Research Asia, P.R. China
Research Intern *2025.5-2025.8*
- Propose and lead projects on time series foundation models and data curation.

Intel, P.R. China
Machine Learning Engineer *2021.5-2023.7*
- Design, research and lead a team to implement *"bigdl-llm"*, *"bigdl-nano"*, and *"bigdl-Chronos"* which collected more than 8K stars.
- More than 20000 lines open-source project contribution, over 300 pull requests' code review; More than 50 API design; More than 10 promotion/tech talks.

SELECTED PROJECTS

`dattri`: is a PyTorch library for developing, benchmarking, and deploying efficient and modern data attribution algorithms.

`bigdl-llm`: is a library for running LLMs on Intel hardware using INT4/FP4/INT8/FP8.

`bigdl-nano`: is a Python package to transparently accelerate PyTorch and TensorFlow applications on Intel hardware.

`bigdl-chronos`: is an application framework for building a fast, accurate and scalable time series analysis application.

TEACHING ASSISTANT

- **Introduction to Machine Learning**
  IS 327, University of Illinois Urbana-Champaign, Teaching Assistant.
- **Data Mining**
  SIADS 532, University of Michigan, Teaching Assistant.
- **Methods and Tools for Big Data**
  VE 472, Shanghai Jiao Tong University, Teaching assistant.

AWARDS & SCHOLARSHIP

- **Explorer Scholarship (2020)**, for outstanding students who went aboard for their graduate study provided by Shanghai Jiao Tong University.
- **Shanghai Outstanding College Graduate (2020)**, for outstanding students graduated in Shanghai.
- **National Scholarship (2018)**, the highest award provided to outstanding university students.

SKILL

- **Language**: Python, C++, Git, Shell, LaTeX
- **Framework**: Pytorch, Tensorflow, CUDA
- **English**: TOEFL 107/120 (Speaking: 25)

COMMUNITY SERVICE

- **Reviewer**: NeurIPS 2024/2025, ICLR 2025, ICML 2025, AISTATS 2025, AAAI 2026
- **Student Organizer**: Regulatable ML @ NeurIPS 2023/2024/2025