Fanyou Wu | Applied Scientist II @ Amazon | 3YOE

+1 (206) 593-7565 | fanyou.wu@outlook.com | github.com/wufanyou | linkedin.com/in/wufanyou | wufanyou.com

Technical Experience

Applied Scientist II Jan 2023 — Present
Amazon, *PXT Central Science* *Seattle, WA*
- Develop a transformer-based job recommendation system using click stream data.
- Apply retrieval-augmented generation (RAG) and fine-tune Large Language Model (LLM) to develop an internal application.
- Conduct research projects regarding LLMs, e.g., dialog generation, hallucination, fairness and role playing.

Postdoc Researcher Jan 2022 — Dec 2022
Purdue University, *Department of Forestry and Natural Resources* *West Lafayette, IN*
- Develop iOS App with ARKit and CoreML to estimate tree diameter, height, and species.
- Use the semantic segmentation method with PyTorch to estimate tree growth ring edges images with extremely noisy.

Machine Learning Engineer Aug 2017 — Dec 2017
Envision Digital *Shanghai, China*
- Use LightGBM to solve time series problems to better predict user demand for electricity.
- Learn to apply machine learning models on IoT platforms based on Hadoop and Hive.

Education

Ph.D in Forestry, *Purdue University* Jan 2018 — Dec 2021
M.S. in Wood Material, *University of Eastern Finland* Sept 2015 — Apr 2017
B.E in Wood Material, *Nanjing Forestry University* Sept 2011 — Jun 2015
Exchange Student in Wood Products Processing, *University of British Columbia* Aug 2013 — May 2014

Selected Competition Experience (Full List at wufanyou.com)

Winners for Two Sub-tasks in Meta KDD CUP 2024 **$1000**
• Develop retrieval augmented generation (RAG) solution. [link]
Second Place in Amazon KDD CUP 2022 Task 2 and Task 3 **$4000**
• Develop natural language processing (NLP) models to solve the e-commerce search problem. [link]
Second Place in CVRP 2021 The 2nd Agriculture-Vision Prize Challenge **$4200**
• Develop semantic segmentation models for farmland prediction, involving distributed training and model ensemble. [link]
Second Place in NeurIPS 2020 Traffic4cast Competition **$5000**
• Use PyTorch to develop deep time spatial models for traffic map movies prediction. [link]
First Place in KDD CUP 2020 Reinforcement Learning Competition Track (Reposition) **$8000**
• Use PyTorch and a self-developed simulator (in Julia) to solve vehicle repositioning problems based on Deep-Q Learning. [link]
First Place in IJCAI-19 Alibaba Adversarial AI Challenge Competition (Traget Attack Track) **$5000**
• Use TensorFlow to perform adversarial attack based on a modification of Projection Gradients Descents. [link]

Selected Recent Publications (Full List at Google Scholar)

- Zhang, Z., Xu, W., **Wu, F.**, & Reddy, CK. (2025). Falsereject: A resource for improving contextual safety and mitigating over-refusals in llms via structured reasoning. COLM 25′ main. [link]
- Liu, Z., Maharjan, S., **Wu, F.**, Parikh, R., Bayar, B., Sengamedu, SH. & Jiang, M. (2025). Disentangling Biased Knowledge from Reasoning in Large Language Models via Machine Unlearning. ACL 25′ main. [link]
- **Wu, F.**, Xu, W., Reddy, CK., & Sengamedu, SH. (2024). Synthesizing Conversations from Unlabeled Documents using Automatic Response Segmentation. ACL 24′ findings. [link]
- Liu, Y., **Wu, F.**, Liu, Z., Wang, K., Wang, F. & Qu, X. (2023). Can language models be used for real-world urban-delivery route optimization? The Innovation. [JCR: Q1] [link]
- Xu, W., Hu, W., **Wu, F.** & Sengamedu, SH. (2023). DeTiME: Diffusion-Enhanced Topic Modeling using Encoder-decoder based LLM. EMNLP 23′ findings. [link]
- Liu, Y., **Wu, F.**, Lyu, C., Liu, X., Li, S., Ye, J. & Qu, X. (2022). Deep Dispatching: A Deep Reinforcement Learning Approach for Vehicle Dispatching on Online Ride-hailing Platform. Transportation Research Part E: Logistics and Transportation Review. [JCR: Q1][link]
- **Wu, F.**, Gazo, R., Benes, B., & Haviarova, E. (2021). Deep BarkID - A Portable Tree Bark Identification System by Knowledge Distillation. European Journal of Forest Research. [JCR: Q1] [link]
- **Wu, F.**, Gazo, R., Haviarova, E., & Benes, B. (2021). Wood identification based on longitudinal section images by using deep learning. Wood Science and Technology. [JCR: Q1] [link]
- Liu, Y., **Wu, F.**, Lyu, C., Liu, X., & Liu, Z. (2021). Behavior2vector: Embedding Users' Personalized Travel Behavior to Vector. IEEE Transactions on Intelligent Transportation Systems. [JCR: Q1] [link]

Last Updated: 2025-07-21