Wengran (Emma) Wang

wengran.emma.wang@gmail.com | emmableu.github.io

Objective

I have expertise in ML model development & large-scale inference, and user experience research. I look for roles that leverage modern ML to deepen human insight, and enable responsible AI innovation.

Professional Experience

- **Machine Learning Engineer @ Apple, Information Intelligence Infra** *02/2023 – Present*
  - End-To-End ML Modeling: Designed and productionized ML models for large-scale content analysis to enhance foundation model training:
    - Web Page Classification: I conducted qualitative and quantitative research to develop web page classification taxonomies, trained a BERT-based model to generate page topic signals, and used these signals to select pages more relevant for foundation model training.
    - Content Selection: I trained and fine-tuned BERT-based classifiers to identify documents with high educational value and broad knowledge coverage, evaluated them against multiple benchmarks, and selected content for foundation model training.
    - Body Extraction: I trained and fine-tuned a MarkupLM-based model to extract main webpage content, evaluated its performance, and deployed it in the data pipeline to provide high-quality data for foundation model training.
  - User Research: I conducted large-scale user studies to inform content quality and taxonomy strategy:
    - Survey Design: I designed, conducted and analyzed human rater surveys to understand perceptions of webpage content and topic clarity.
    - Content Quality Analysis: I applied qualitative and quantitative methods to evaluate content accuracy and quality; built topic taxonomy for web page signals.
- **Research Assistant @ NC State University, Center for Educational Informatics** *02/2020 – 02/2023*
  - Computing Education Technology: I designed, prototyped, and conducted human evaluation on 6+ tools to enhance engagement and accessibility.
    - Idea Builder: Design tool for building open-ended projects, converts visual drawings into working programs.
    - Pinpoint: Enables debugging and comprehension through visual execution traces.
    - Step Tutor: Automated generation of step-by-step examples.
  - **Impact:**
    - Published **12** first-authored, peer-reviewed conference and journal papers.
    - Earned **1** Best Paper Award at a major academic venue.
    - Mentored **4+** undergraduate and master's students toward academic and career achievements.
    - Developed tools adopted by **1,000+** students and integrated into college curricula.

Technical Qualifications

- **Programming Languages:** Python, C++, Java, JavaScript
- **ML Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn
- **Model Deployment & Serving:** ONNX, Triton Inference Server, FastAPI, Docker, Kubernetes
- **Data & Experimentation Tools:** Pandas, NumPy, Weights & Biases, Apache Spark, Airflow
- **Web & Other Technologies:** React, Node.js, REST APIs, SQL, Git

Education

- **North Carolina State University** *Aug. 2019 - Present*
  Ph.D. in Computer Science Raleigh, USA
- **North Carolina State University** *Aug. 2016 - May 2018*
  M.Sc. in Statistics Raleigh, USA
- **Zhejiang University** *Aug. 2013 - July 2016*
  B.Sc. in Environmental Science; Minor in French Language and Literature Hangzhou, China