DSA Research Experiences for Undergraduates

Research Project

Section1: Faculty Information

Full Name	Weikai YANG	Tel	13051670559
Thrust/Hub	DSA/INFH	Office	W1-316
Email	weikaiyang@hkust-gz.edu.cn		

Section2: Research Project Proposal

Project Title	Human-AI Collaborative Sample Selection Framework
Project Description (max 800 words)	1. Background The quality of training datasets is a critical determinant of machine learning model performance. The adage "garbage in, garbage out" is particularly salient in data-driven AI: noisy, redundant, or biased samples degrade generalization and increase training instability. While manual curation by experts remains the gold standard, scaling this approach to large-scale datasets is prohibitively expensive. Recent advances of foundation models provide an efficient way to evaluate the quality of samples and filtering out low-quality samples. However, the process is black-box and offers no insight into the sampling process, making it hard to be improved when the sampling result is not satisfied. To tackle this issue, this project aims to create a human-AI collaborative sample selection framework that synergizes the scalability of foundation models with human interpretability, enabling dynamic optimization of automation efficiency and expert oversight. 2. Challenges (a) **Absence of Universal Quality Standards**: Defining objective metrics for sample quality is inherently task-dependent, particularly for subjective tasks (e.g., creativity assessment in text generation). (b) **Unstructured Model Outputs**: Foundation models generate free-text rationales and scores that resist systematic analysis at scale, complicating human validation and refinement. 3. Objectives (a) Develop a standardized prompting framework to generate quality scores, rationales, and structured metadata during sample evaluation. (b) Design an effective visualization to present the foundation model outputs and support human analysis (c) Establish an iterative human-AI collaboration workflow for sample selection 4. Methodology (a) Metadata-Enhanced Prompting: Design structured prompts to extract: quality scores (1-5 scale), rationales (free-text explanations for

	scores), and task-specific metadata (e.g., "ambiguity level" for textual entailment, "emotional polarity" for sentiment analysis). Implement consistency checks via multi-prompt voting and contradiction resolution mechanisms. (b) Interactive Data Analysis: Develop a dashboard to visualize score distributions, rationale keywords, and meta-data that allow large-scale analysis (c) Dynamic Selection Strategy: Allow uses to provide some feedback on the sample selection results, and the system will automatically adjust to fit user preference 5. Expected Outcomes (a) An effective visualization method to organize the output of foundation models and support analysis in scale (b) A visual analysis pipeline that integrates foundation models and human insights to enhance the quality of sample selection (c) An open-sourced prototype of the interactive sample selection tool
Proposed Research Duration	Start Date: ___MAR_____ / ___15_____ / ____2025____ End Date: ___SEPT_____ / ___15_____ / ____2025____
Student/Researcher Duties	- Actively participate in discussions and provide regular feedback - Conduct a literature review on LLM-based sample selection method - Implement the prompt to extract scores, rationale, metadata and visually organize them - Design the workflow for interactive sample selection - Design and conduct rigorous experiments to evaluate the performance of sampling results
Technical Skills Required	☑ Python ☑ Machine Learning ☐ Big Data ☐ R ☑ Deep Learning ☐ SQL ☐ C/C++ ☐ Other: _____
Preferred Student/Researcher Background	Python, Javascript, and the usage of large language models. But it is not required as I will provide necessary guidance to help you.
Maximum Number of Students/Researchers	☐ 1 ☑ 2

Section3: Pre-Application Research Exposure Meeting

Faculty members are encouraged to schedule a Research Exposure Meeting to introduce students to their projects.

Preferred Date	Discussed via email
Preferred Time	Discussed via email
Meeting Mode	☑ In-Person ☐ Online
Venue (if in-person)	W1-316
Meeting Link (if online)	