Thesis Defense

Computer Science Master's Program

"When Agents Spend Watts: Measuring Performance-Energy Trade-offs in Agentic AI Systems with AgenticBench"

By **Logan Barker**

Abstract:

Agentic AI systems, which extend large language models with iterative reasoning and tool use, are effective for solving multi-step problems but introduce substantial inference-time overhead and significant energy costs due to repeated model calls and orchestration logic. The current evaluation landscape for agentic AI is fragmented, focusing on task performance or correctness, and lacks integrated energy measurement. This gap prevents researchers and practitioners from practically comparing agentic systems on both performance and energy efficiency.

This thesis introduces AgenticBench, a modular benchmark harness designed to evaluate agentic AI systems while collecting latency, token, tool-use, and energy signals. Its systems-oriented goal is to implement a reproducible framework that combines benchmark execution, energy profiling, and comparative analysis. Empirically, the framework is used to rigorously measure performance-energy trade-offs across various agentic architectures, models, and benchmarks.

To validate AgenticBench, 84 experiments were conducted on Windows and Linux workstations, testing four models (gpt-oss, qwen3, qwen2.5:3b, llama3.2:3b) and three frameworks (LangChain, AutoGen, CrewAI) across six benchmarks (e.g., GSM8K, MATH, HotpotQA). Energy was measured using a hybrid approach, combining hardware-accurate GPU power measurements with estimated CPU energy. Key findings demonstrate significant trade-offs: (1) Framework choice alone introduces a 2–10× energy overhead, with AutoGen being the most expensive due to its async message-passing. (2) Large API models (gpt-oss, qwen3) provide superior accuracy but incur 10–100× the energy cost of local 3B models. (3) qwen3's extended chain-of-thought reasoning led to the highest energy consumption. (4) LangChain paired with gpt-oss achieved the best overall performance-energy balance. These results highlight trade-offs invisible in accuracy-only reporting and offer concrete guidance for choosing agent configurations.

Date: Tuesday, June 9th, 2026
Time: 1:00 PM – 3:00 PM
Location: 14-238b
Zoom: https://calpoly.zoom.us/j/9196734099
Committee: Dr. Franz Kurfess, Dr. Alex Bisberg, Dr. Maria Pantoja