MATHEMATICS DEPARTMENT SEMINAR

Summer Research in Artificial Intelligence: From Secure to Local LLMs

Hiding LLM Fingerprints from GPU Side Channels

Tom Regpala
CV Path - 2025 Summer Research Program

Abstract: Transformer models can be fingerprinted by monitoring GPU activity. We test two system-level defenses. Jittering injects short, randomized bursts, dropping a trained identifier's accuracy from about 95% to 61% at a 33% runtime cost. Constant-load padding keeps the GPU busy to flatten signatures, cutting accuracy to 9% while making inference about 6.6 times slower. The choice reflects a security versus performance trade-off for fine-tuned or proprietary models.

CSUB-GPT: Open-Source Local LLMs and for Campus Use

Juan Rodriguez Aguilar and Christian Rodriguez
Chevron 2025 SURE Summer Research Program and ELEVATE Grant

Abstract: We introduce a local assistant built on open models like Mistral 7B and DeepSeek 7B/14B, exploring whether small, retrainable LLMs can rival a Custom GPT in academia. The system ingests documents, builds embeddings in ChromaDB, and serves a browser chatbot via Gradio on an RTX A4000 workstation, the CSU Nautilus cluster, and Google Colab. Results highlight offline access, on-premise privacy, CSUB-specific customization, and zero API costs, with plans for broader course coverage, advising, and campus-wide deployment.

2:10 PM TO 3:00 PM

Wednesday, November 19, 2025

Science III, Room 240