Contribution ID: **2** Type: **not specified**

Keynote: On New Security and Safety Challenges Posed by LLMs and How to Evaluate Them

Thursday, June 6, 2024 4:30 PM (45 minutes)

Large Language Models (LLMs) are integrated into many widely used and real-world applications and use-case scenarios. With their capabilities and agentic-like adoption, they open new frontiers to assist in various tasks. However, they also bring new security and safety risks. Unlike previous models with static generation, LLMs' nature of dynamic, multi-turn, and flexible functionality makes them notoriously hard to robustly evaluate and control. This talk will cover some of these new potential risks imposed by LLMs, how to evaluate them, and the challenges of mitigations.

THIS KEYNOTE CAN ALSO BE ATTENDED ONLINE! Registration is possible here.

Presenter: ABDELNABI, Sahar S. (Microsoft)