# AI for Offensive Security
## Autonomous Red Teaming

Dr. Emre Süren

Royal Hacking Lab
lab.cybercampus.se

CYBERCAMPUS SVERIGE

KTH VETENSKAP OCH KONST

## Contact

emsuren@kth.se
www.kth.se/profile/emsuren/
lab.cybercampus.se

## Experience

LLM security
Cyber Threat Intelligence
Vulnerability research

## Organization type

Public University

## Software

Open-source prompt injection tool
Novel prompt-injection method/tool
Autonomous pentesting agents

## Grants

GPU access (for LLM research)
Cyber Threat Intelligence for Sweden

## Human Resources

1 Researcher (myself)
10+ Thesis Students

## Existing Collaborators

AICell Lab – LLM red teaming

## Lab Infrastructure

IoT Devices

## Topic/s that are of our interest

Cybersecure tools, tech, and services relaying on AI (CYBER-09-CYBERAI)

## Project Idea

Agentic Red Teaming

## Research problem

Red teaming
Complex skillset
Time-consuming task

## What we own to solve the problem

Autonomous pentesting agent
Novel 2-phase guardian agent evasion method

## Role

Partner

# Dr. Emre Süren

Royal Hacking Lab

CYBERCAMPUS SVERIGE

KTH VETENSKAP OCH KONST