

Threat Model

Attacker Knowledge

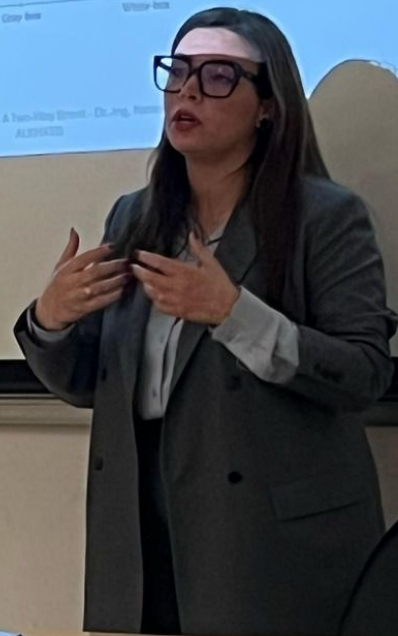
How much knowledge does the attacker have about the AI system?
Even with zero knowledge in the beginning an attacker will often acquire some system details over time.

Obscure Knowledge about		Partial Knowledge about	Full Knowledge about
Observable AI Decisions	Original Distribution	Architecture, Training Setup	Parameters, Architecture, Training Setup
Black box	Grey box	White box	

Computer Security & AI - A Two-Way Street - Dr. Jing, Nanyang Technological University

4/18/2023

28



AI HAS NO BOUNDARIES!
WHETHER YOU'RE AN ENGINEERING
STUDENT OR JUST PASSIONATE ABOUT
TECHNOLOGY, YOUR FIELD IS DEFINITELY
WITHIN OUR REACH



FROM IDEAS TO INTELLIGENCE
JOIN THE AI CLUB!

DON'T HESITATE TO REACH OUT

Instagram icon ULFG1.AI

LinkedIn icon ULFG1 AI



Email icon AI CLUB.ULFG1@GMAIL.COM



AI CLUB
ULFG1

Overview of AI Systems

- Study of the model alignment of the risks and potential unintended consequences by structure
- No known general structure
- Disruptive areas focus, e.g., atomic reactors, of autonomously directed entities, e.g., for a narrow purpose.
- Using governance for autonomously learned models, the following risk factors must be considered:
- The system must not be generally allowing the interface to generating higher-level actions.
- The system must be highly resistant to being hijacked by bad actors.
- The system must be using a secure architecture.
- The system must be using a secure network architecture.

AI HAS NO BOUNDARIES!
WE'RE NOT IN THE FUTURE... WE'RE IN THE PRESENT!
THE ONLY DIFFERENCE IS THAT WE'RE NOT YET THERE!
FROM IDEAS TO INTELLIGENCE
DON'T WAIT UNTIL THE AI CLUB!
@AIUBCLUB
AIUBCLUB.AI
AI CLUB CHANGING THE WORLD
AI CLUB CHANGING.COM



