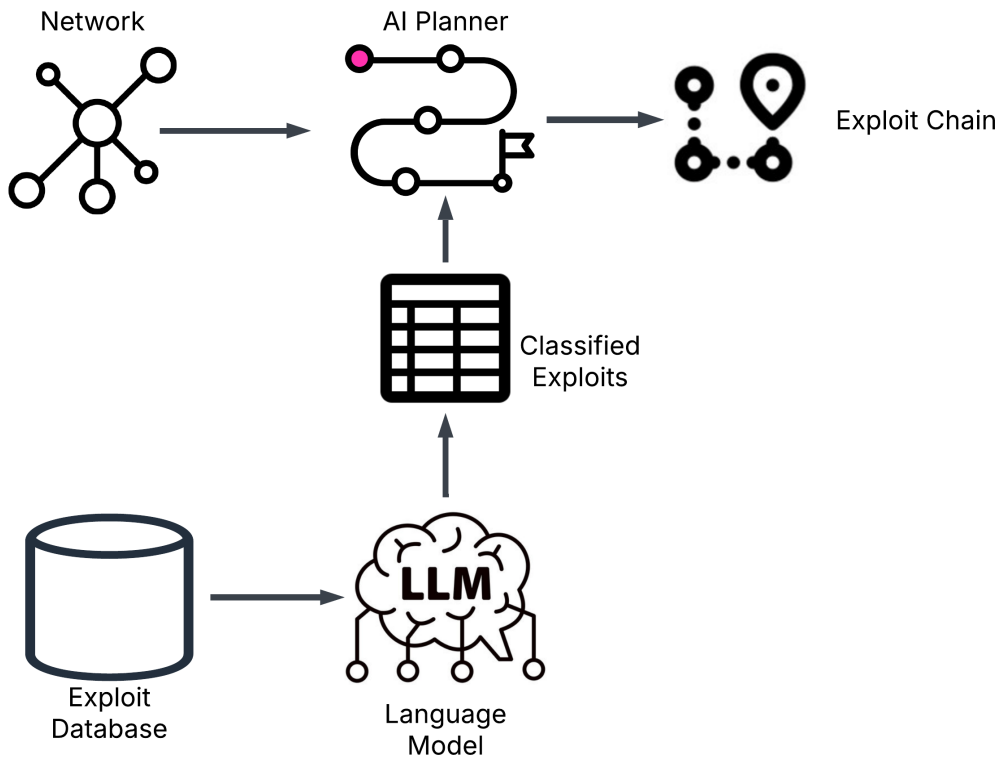
PERE-Chains: AI-Supported Discovery of Exploit Chains

- ***Exploit*** - A *technique or code* that takes advantage of a *vulnerability* in software or hardware.
  - Enables attackers to execute code, steal data, or escalate computer access.
- **Exploit Chain** - A sequence of *multiple exploits* used together to *achieve a goal*.
  - Each exploit bypasses a defense, advancing toward full network compromise.
- **Key Question**: How can defenders quickly assess whether a host is vulnerable to a multi-step attack?
- **Why It's Hard:** Attackers can combine exploits from a vast and growing pool of vulnerabilities, which humans have a hard time keeping track of.



Cristián Colón, Una-May O'Reilly

Notes

- Motivation
  - Key Question: How can defenders quickly assess whether a host is vulnerable to a multi-step attack?
  - Why It's Hard: Attackers chain multiple exploits from a vast and growing pool of public vulnerabilities.
- Background:
  - Exploit Chain: A multi-step sequence (e.g., privilege escalation + remote exploit) used to fully compromise a target host.
- Approach:
  1. Classification – Categorize exploits (e.g., privilege escalation vs. remote).
  2. Modeling – Represent the network (hosts, vulnerabilities) in PDDL.
  3. Planning – Use AI planners to automatically uncover exploit chains.
- Results:
  - Scalability: Successfully analyzed up to 200 hosts and 1,880 Metasploit exploits, demonstrating effectiveness in large environments.
  - Novel Attack Paths: Discovered 12 additional exploit chains, with one fully executable via Metasploit—proving the framework can reveal unexpected yet practical vulnerabilities.