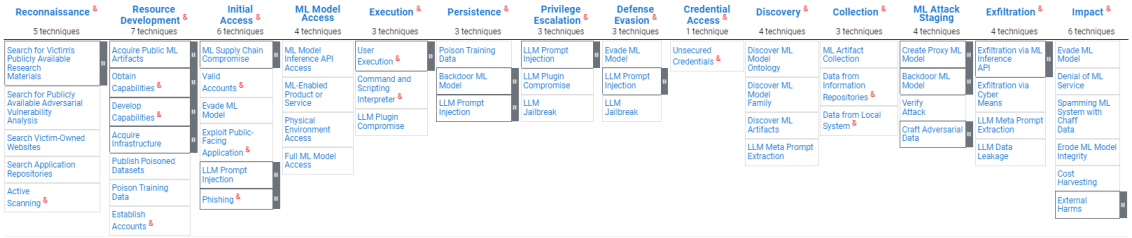
MITRE ATLAS



O MITRE ATLAS, que significa "Adversarial Threat Landscape for Artificial-Intelligence Systems", é uma base de conhecimento abrangente e acessível globalmente. Ele foca em táticas e técnicas de adversários derivadas de observações reais de ataques e demonstrações práticas por equipes vermelhas de IA e grupos de segurança. Esse repositório é projetado para evoluir continuamente com a mudança do cenário de ameaças no espaço da IA.

O núcleo do MITRE ATLAS é composto por estudos de caso detalhados selecionados pelo seu impacto significativo em sistemas de aprendizado de máquina (ML) em produção. Esses estudos de caso ilustram uma ampla gama de ataques, como evasão, envenenamento, replicação de modelos e exploração de falhas de software tradicionais. O espectro de personas envolvidas nestes estudos inclui usuários comuns, pesquisadores de segurança, pesquisadores de ML e equipes vermelhas totalmente equipadas, oferecendo diversas perspectivas e insights sobre potenciais vulnerabilidades e ameaças.

Em termos de metodologia, o MITRE ATLAS detalha várias técnicas que os adversários usam para alcançar objetivos táticos. Essas técnicas ilustram tanto as ações que os adversários podem tomar, como comprometer a cadeia de suprimentos de ML para acesso inicial, quanto os objetivos que eles pretendem alcançar por meio dessas ações. Esse foco duplo ajuda a entender tanto o processo quanto o propósito por trás de diferentes ameaças cibernéticas em sistemas de IA.

O MITRE ATLAS também aborda um aspecto crítico dos sistemas modernos de IA: sua vulnerabilidade a uma nova classe de ameaças denominada "Aprendizado de Máquina Adversário". Isso se refere às vulnerabilidades sistemáticas dentro dos sistemas de ML em produção, que os adversários podem explorar para manipular sistemas de IA. Ao destacar essas vulnerabilidades, o MITRE ATLAS auxilia no reconhecimento e mitigação de potenciais ameaças na paisagem em evolução da tecnologia de IA.

No geral, a estrutura do MITRE ATLAS serve como um recurso inestimável para compreender as estratégias e metodologias de adversários cibernéticos. É particularmente útil para profissionais de defesa cibernética e qualquer pessoa envolvida na segurança de sistemas de IA e ML, fornecendo insights sobre o contexto, estrutura e aplicações práticas para salvaguardar essas tecnologias contra ameaças emergentes.

Saiba mais em https://atlas.mitre.org/