devoteam
Cyber Trust

Case Study

Segurança da IA nos Serviços Financeiros: Pentesting de um Chatbot LLM para um Banco Líder

Cliente

Uma instituição financeira global implementa um chatbot de literacia financeira baseado em IA, integrado num website virado para o cliente, e encomenda uma avaliação completa de segurança de LLM antes da entrada em produção.

O Desafio

- **A IA introduz uma nova categoria de risco em ambientes regulados:** a implementação de um chatbot alimentado por LLM numa instituição financeira cria superfícies de ataque que os frameworks de segurança existentes não foram concebidos para avaliar e que os reguladores estão a analisar com escrutínio crescente.
- **A implementação pública amplifica a exposição:** o chatbot foi concebido para ser incorporado em qualquer website, incluindo sites fora do controlo da instituição. Isto cria um risco único: o chatbot pode ser carregado e explorado a partir de contextos arbitrários de terceiros, tornando a prevenção de abusos, a validação da origem e a integridade das respostas requisitos críticos de segurança.
- **O risco associado à IA acresce ao risco da infraestrutura existente:** os LLM não substituem vulnerabilidades legadas, acrescentam uma nova camada de risco. Uma instituição financeira que serve 3,6 milhões de clientes, através de mais de 500 balcões nacionais e presença em 23 países, não se pode dar ao luxo de ter pontos cegos em qualquer camada da sua infraestrutura tecnológica.

A Solução

- **Metodologia de Pentesting de IA da Devoteam Cyber Trust:** uma abordagem robusta e estruturada, assente no MITRE ATLAS, no OWASP LLM Top 10 e no OWASP Agentic Top 10, os frameworks mais abrangentes da indústria para riscos adversariais em IA, garantindo uma cobertura sistemática de todas as categorias conhecidas de ataques a sistemas de IA.
- **IA + Cloud + Aplicação Web: uma abordagem integrada:** a implementação do chatbot estava integrada com o website corporativo e com a infraestrutura cloud de backend. A realização conjunta de testes de Cloud Pentesting e Web Application Pentesting da Devoteam Cybert Trust, em articulação com a metodologia de IA, revelou-se essencial: a análise à cloud identificou a exposição de um token que comprometia credenciais, criando o risco de comprometimento total do ambiente cloud.
- **Viabilizar inovação digital de confiança:** ao proteger a camada de IA, a infraestrutura cloud e a aplicação web antes da entrada em produção, esta intervenção permite à instituição expandir serviços de literacia financeira a milhões de clientes, salvaguardando simultaneamente os utilizadores finais, os ativos cloud e a conformidade regulatória da instituição.

Resultados

Consumo Não Limitado

Sem limitação da taxa de consumo de tokens ou do tamanho dos inputs, um atacante poderia degradar a disponibilidade do serviço em todas as instâncias incorporadas.

Risco crítico numa implementação pública e incorporável.

Exposição de Token Cloud

Credenciais cloud expostas através da camada de integração entre o website e o chatbot com potencial para comprometimento total do ambiente cloud.

Identificado através do Cloud Pentesting da Devoteam Cyber Trust, aplicado à camada de integração.

Descoberta em fase de pré-produção

Todas as vulnerabilidades críticas foram identificadas e corrigidas antes do lançamento público.

Exposição zero a clientes reais ou consequências regulatórias no momento da entrada em produção.

Chatbot Hijacking

A natureza incorporável do chatbot permitia que este fosse carregado e explorado a partir de websites arbitrários de terceiros, fora do controlo da instituição.

Um atacante poderia incorporar o chatbot noutros websites, com potencial impacto financeiro para o cliente.