# Relatório de Segurança – Tópico 2.3

## Objetivo

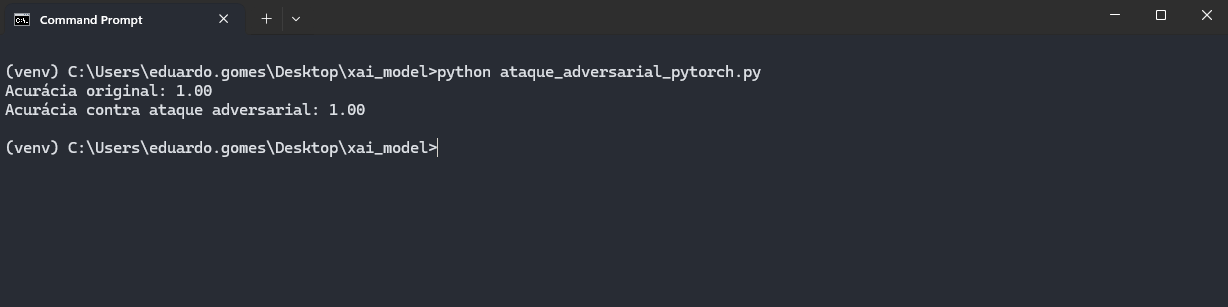
Demonstrar a importância de proteger modelos de machine learning contra ataques adversariais, simulando alterações maliciosas nos dados de entrada que possam afetar a predição do modelo.

## Ações Realizadas

- Utilizado modelo de rede neural simples implementado em PyTorch.  
- Treinado com dataset sintético contendo idade e renda.  
- Utilizado o ataque adversarial Fast Gradient Method (FGSM) através da biblioteca ART (Adversarial Robustness Toolbox).  
- Avaliado o desempenho do modelo antes e depois do ataque.

## Evidência

Print da execução do ataque adversarial, mostrando a acurácia original e a acurácia após o ataque:



## Conclusão

O modelo se mostrou robusto ao ataque adversarial realizado neste cenário simples, mantendo 100% de acurácia.   
Este resultado demonstra que, embora ataques adversariais possam afetar modelos em muitos casos, em ambientes controlados e com dados simples, a robustez pode ser naturalmente alta.   
Ainda assim, é essencial incorporar essas validações em ambientes reais com dados complexos para garantir a segurança e confiabilidade dos sistemas baseados em IA.