Segurança da Informação

Especificação do Mini-projeto

Grupo 3
Filipa Lopes
Inês Pereira
Sara Esteves

Possíveis títulos:

- *Deepfakes*: como criar, detetar e combater vídeos manipulados
- *Deepfakes*: uma nova ameaça à segurança da informação. Como criar, detetar e combater.

Dentro do tema "informação falsa *versus* informação precisa ou verdadeira", decidimos aprofundar o tema dos vídeos manipulados, mais concretamente o caso dos *deepfakes.*

Deepfakes são ficheiros de vídeo, imagem ou voz manipulados usando Inteligência Artificial (IA), que pretendem fazer o seu conteúdo falso passar por autêntico. O termo resulta da junção das palavras *fake* (falsificação) e *deep learning* (aprendizagem profunda), um tipo de IA que imita o funcionamento do cérebro humano, aprendendo com grandes quantidades de dados e criando conteúdo novo. O nosso trabalho centrar-se-á apenas nos *deepfakes* de vídeo.

Principal objetivo:

Explorar o tema dos vídeos *deepfakes* no contexto da segurança da informação.

Objetivos específicos:

- Compreender como os desenvolvimentos na área da Inteligência Artificial (IA) possibilitaram a criação de *deepfakes* cada vez mais realistas;
- Explorar como criar vídeos *deepfakes* utilizando ferramentas disponíveis em linha;
- Identificar aplicações positivas e negativas de *deepfakes*, aprofundando como podem ser uma ameaça à segurança da informação;

- Investigar estratégias que podem ser adoptadas para a deteção de vídeos *deepfakes*, incluindo soluções baseadas em tecnologia de IA;
- Se o tempo para a feitura do trabalho o permitir, tentar realizar uma recolha e seleção de fontes que possam ser usadas em ações de sensibilização e de literacia informacional e digital neste âmbito.

Estrutura provisória:

1. Introdução
2. Inteligência Artificial (IA) e vídeos manipulados: o que são os *deepfakes*?
3. Como manipular um vídeo com IA e criar um *deepfake*
4. Aplicações positivas e negativas dos *deepfakes*
5. Estratégias de deteção de *deepfakes* (incluindo tecnologias de IA)
6. Lista de fontes de informação para ser usada em ações de sensibilização de literacia informacional e digital neste âmbito
7. Conclusões
8. Referências bibliográficas