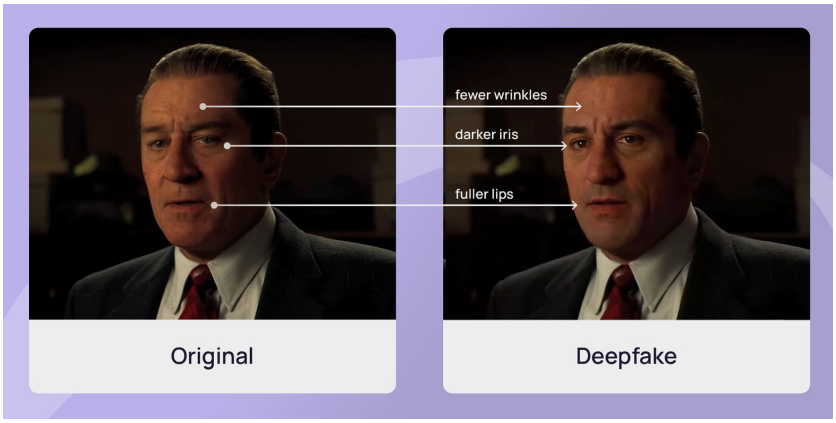
FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

ZÁPADOČESKÁ
UNIVERZITA
V PLZNI

Deepfake technologie



Západočeská Univerzita V Plzni
Katedra Kybernetiky
Semestrální práce - HKUI

Stanislav Rauch
1. semestr
31. prosince 2024

1 Úvod

Deepfake technologie představuje rychle rozvíjející se oblast, která využívá pokročilé algoritmy strojového učení, zejména hlubokých neuronových sítí, k vytváření realistických, avšak falešných multimediálních obsahů. Tyto technologie umožňují manipulovat s obrazem, zvukem či videem tak, že výsledný produkt je jen obtížně rozeznatelný od původního autentického obsahu.

Cílem této práce je poskytnout přehled o historii deepfake technologií, principech jejich fungování, metodách jejich rozpoznávání a etických otázkách spojených s jejich používáním.

2 Vznik Deepfake technologií

Původ deepfake technologií sahá do 90. let, kdy výzkumníci využívali počítačem generované obrazy (CGI) k vytváření realistických zobrazení lidských postav. Technologie zažila významný rozvoj v 10. letech 21. století, který byl poháněný vznikem rozsáhlých datových sad, pokrokem v technikách strojového učení a zdokonalenými výpočetními schopnostmi, což společně posunulo tento obor vpřed.

Zásadním momentem ve vývoji deepfake technologií byl pokrok v hlubokém učení z roku 2014, který představil Ian Goodfellow a jeho kolegové. Goodfellowova práce na generativních adversariálních sítí (GANs) znamenala zásadní průlom v oblasti strojového učení, což nakonec umožnilo vznik další generace vysoce pokročilých deepfake obrazů, videí a audia.

3 Principy fungování deepfake technologií

Základem deepfake technologií jsou již zmíněné GANs, které se skládají ze dvou neuronových sítí: generátoru a diskriminátoru. Generátor vytváří syntetická data, tedy obrázky, video nebo audio. Diskriminátor vyhodnocuje výstup generátoru a snaží se rozlišit generovaná data od skutečných dat. Tento proces se neustále opakuje a při každém opakování dochází k postupnému zlepšování kvality generovaných dat, dokud diskriminátor nevyhodnotí vygenerovaná data jako téměř nerozeznatelná od reálných.

Kromě GANs se využívají také autoenkodéry a další architektury hlubokého učení pro specifické úkoly, jako je převod hlasu či synchronizace pohybu rtů s daným audiem.

4 Detekce a prevence

Vzhledem k potenciálním rizikům spojeným s deepfake technologiemi je výzkum zaměřen také na vývoj nástrojů pro jejich detekci. Metody zahrnují detekci nesrovnalostí v biologických signálech (např. mrkání) či využití neuronových sítí trénovaných na rozpoznávání syntetických médií. Důležité je rovněž vzdělávání veřejnosti o existenci a rizicích deepfake technologií. Například společnosti jako Microsoft vyvíjejí nástroje pro detekci deepfake videí, které analyzují vizuální a audio stopy za účelem identifikace manipulací.

Při rozlišování pravých a uměle vytvořených videí se lidé mohou řídit několika poznávacími znaky, mezi něž patří například tyto:

- Pravidelnost mrkání
- Nekonzistentní audio
- Nepřirozené pohyby těla
- Nekonzistentní odrazy světla v brýlích
- Nekonzistentní vzhled pokožky
- Nepřirozený pohyb rty nebo nepřirozené okolí úst

5 Etické a právní aspekty

S rozvojem deepfake technologií bohužel vyvstávají i závažné etické a právní otázky:

- **Dezinformace:** Možnost vytváření falešných zpráv či projevů veřejných osobností může vést k manipulaci veřejného mínění.
- **Porušení soukromí:** Neoprávněné použití podoby či hlasu jednotlivců bez jejich souhlasu představuje zásah do jejich osobnostních práv.
- **Právní regulace:** Nedostatečná legislativa v oblasti deepfake technologií komplikuje tresty za jejich zneužití. V některých zemích již byly přijaty zákony zakazující tvorbu a šíření deepfake materiálů bez souhlasu dotčených osob, zatím ale bez celosvětové regulace.
- **Politická manipulace:** Například šíření deepfake videí s falešnými prohlášeními politiků nebo prezidentů.
- **Podvody:** Například použití syntetického hlasu generálního ředitele pro falešné telefonáty a podvody.

Nakonec největším rizikem je narušení důvěry, které by mohlo v lidech vzbudit pochybnosti o pravosti zpráv na internetu.

Seznam použité literatury

- SoSafe. (2024, March 18). *How to spot a deepfake - SoSafe*. SoSafe. https://sosafe-awareness.com/blog/how-to-spot-a-deepfake/
- SEON. (2024, December 20). *What is a deepfake - How do deepfakes work? — SEON*. https://seon.io/resources/dictionary/deepfakes/
- Reality Defender — Enterprise-Grade Deepfake Detection. (n.d.-b). https://www.realitydefender.com/blog/history-of-deepfakes/