

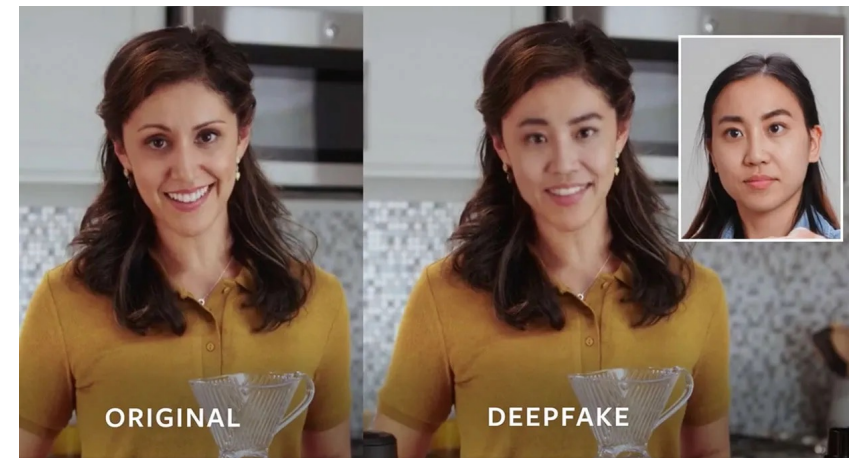
Creation and detection of malicious synthetic media - a preliminary survey on deepfakes

Anton Firc, Kamil Malinka a Petr Hanáček

FIT VUT

Definice deepfakes

- Deeplearning + fake = deepfake
- Podmnožina syntetických médií
 - Zobrazují události, co se nikdy nestaly
- Vytvořené pomocí hlubokých neuronových sítí

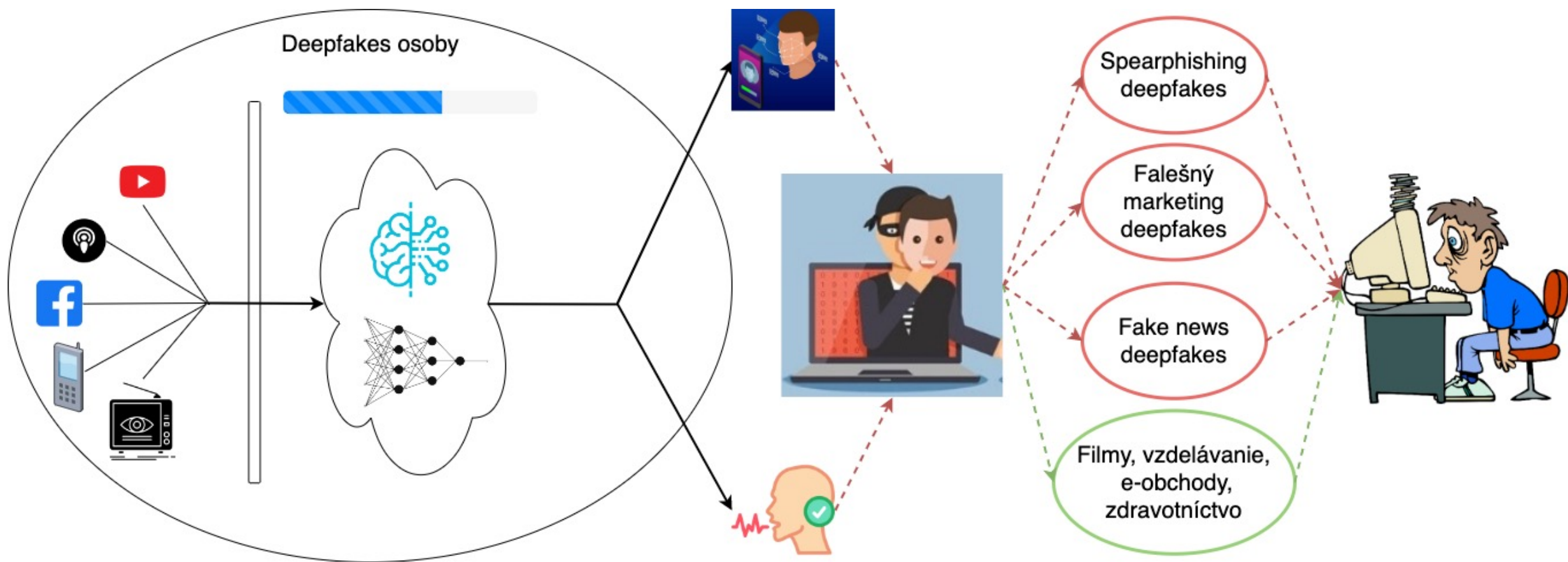


<https://spectrum.ieee.org/facebook-ai-launches-its-deepfake-detection-challenge>

Aktuální příklady z médií



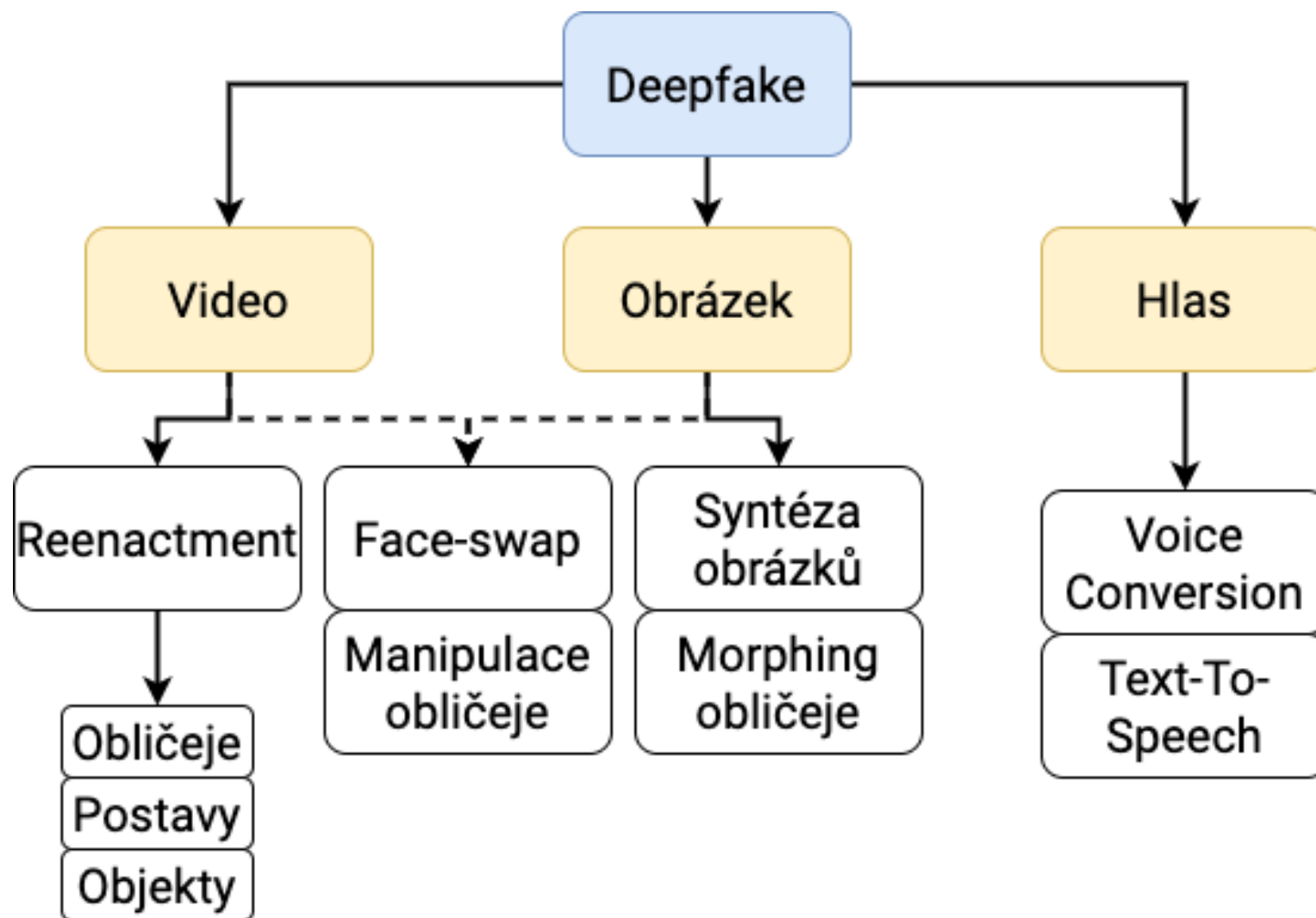
Model útočníka



Vektory útoku

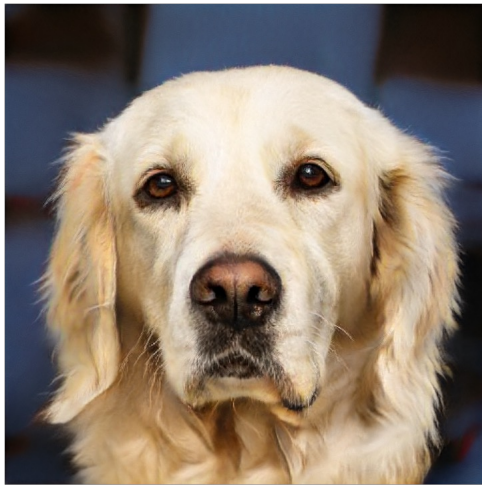
- Biometrické systémy
 - Hlas
 - Obličej
- Sociální inženýrství
 - Phishing / Spearphishing
 - Vhishing
- Fake news
- Fake marketing
- Pomluva
- Manipulace důkazů
- Boti na sociálních sítích
- KnowYourCustomer proces

Typy deepfakes



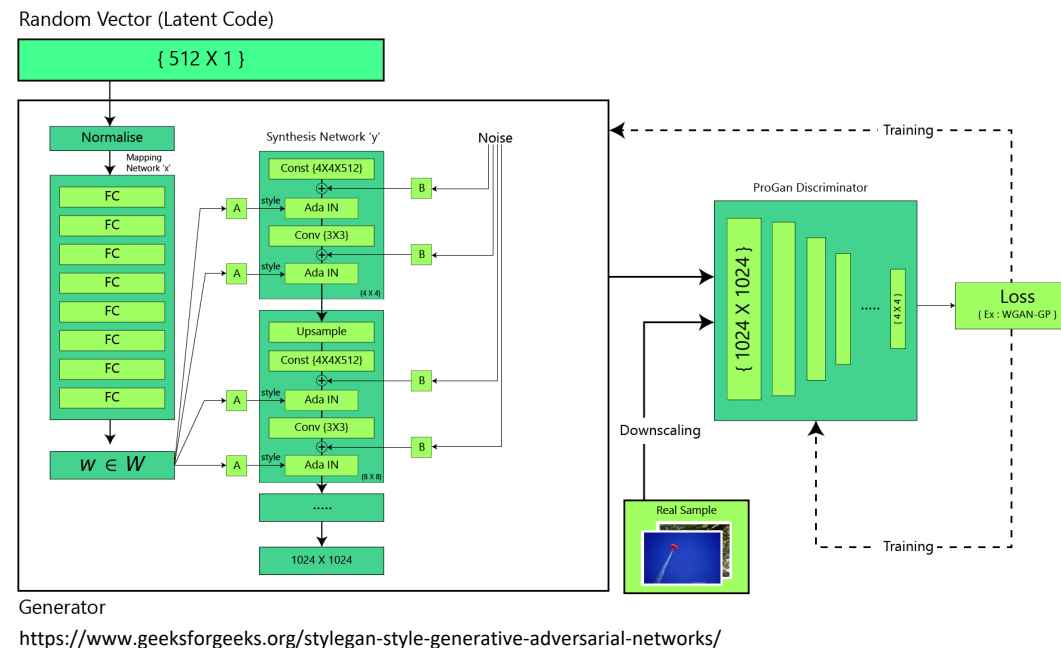
Syntéza obrázků

- Syntéza ne-existujících objektů
 - Lidi, zvířata, scenérie, auta...
- Trénink nad velkým množstvím obrázků stejné kategorie
- Náhodné generování z naučeného latentního prostoru



Syntéza obrázků - technologie

- GANy a jejich variace
- Naučení reprezentace obrázků v latentním prostoru
- Generace nových obrázků z naučeného latentního prostoru



Syntéza obrázků – útočník

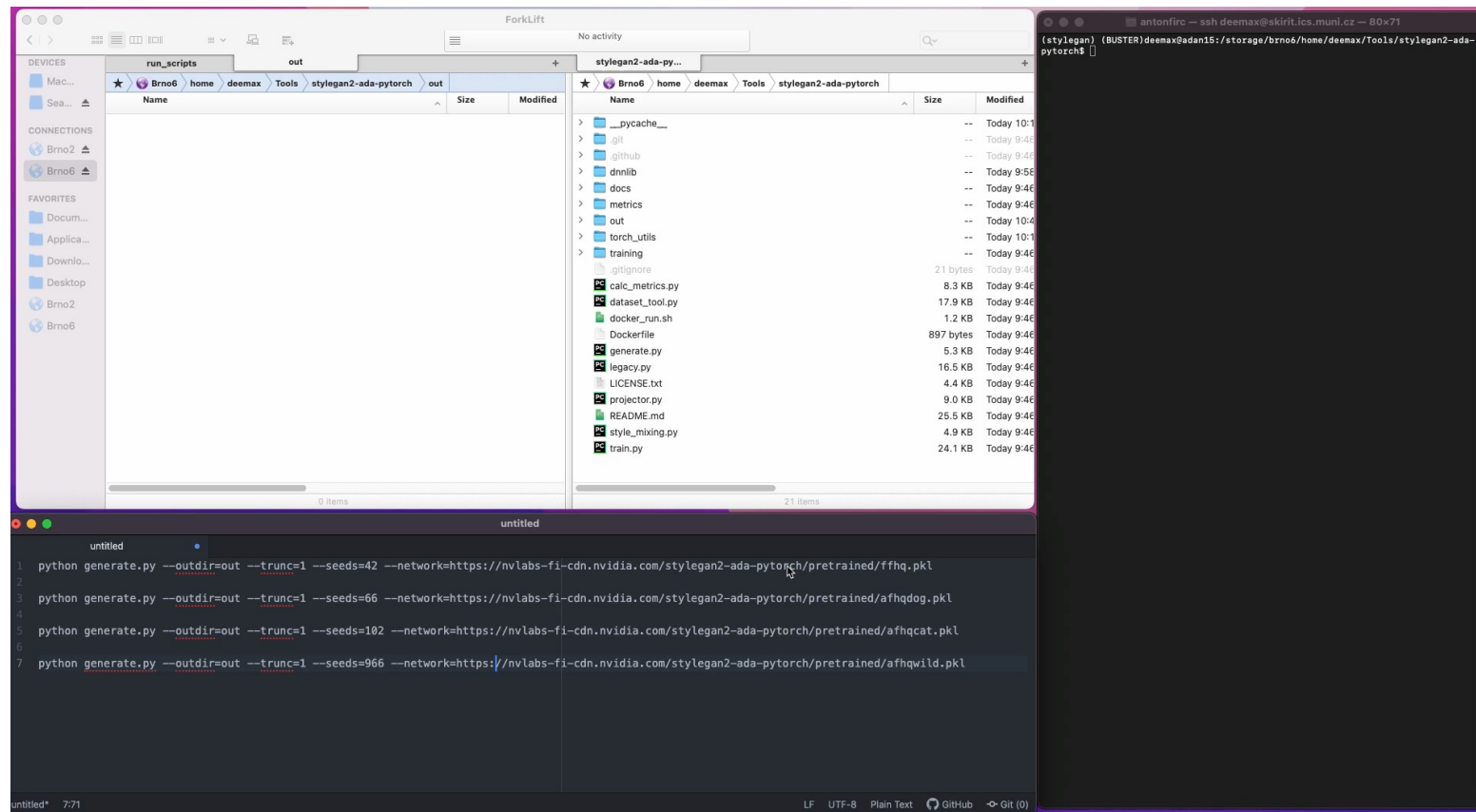
- Boti na sociálních sítích
 - Syntetické profilové obrázky = náročná detekce
- Negativní tvrzení o identitě
 - “nemám přístup do systému, ale měl bych”



<https://plus.rozhlas.cz/socialni-davky-zneuzivaji-bohati-7844454>

Syntéza obrázků – ukázka

<https://github.com/NVlabs/stylegan2-ada-pytorch>



Morphing obličeje

- Morphing = efekt změny z jednoho obrázku na druhý
- 2 obličeje -> jeden podobný oběma
- Nebezpečí pro automatizované hraniční kontroly identity



<https://mappingignorance.org/2018/02/21/triangulations-face-morphing/>

Morphing obličejů - postup

1. Oba obrázky jsou umístěné na samostatné vrstvy taky aby se prolínali oči.
2. Označí se důležité body na obličejích.
3. Automaticky se vygeneruje přechod z jednoho obličejů na druhý.
4. Výsledný obrázek je vybrán na základě největší podobnosti.

Morphing obličejů - technologie

- Založené na důležitých bodech
 - Deformace obrazu
 - Free-form deformation, Delaunay triangulation, mass-spring based deformation, ...
 - vyžadují následný post-processing
- Založené na hlubokém učení
 - GANy
 - Interpolácia v latentnom priestore
 - Autoenkodéry

Morphing obličeje – útočník

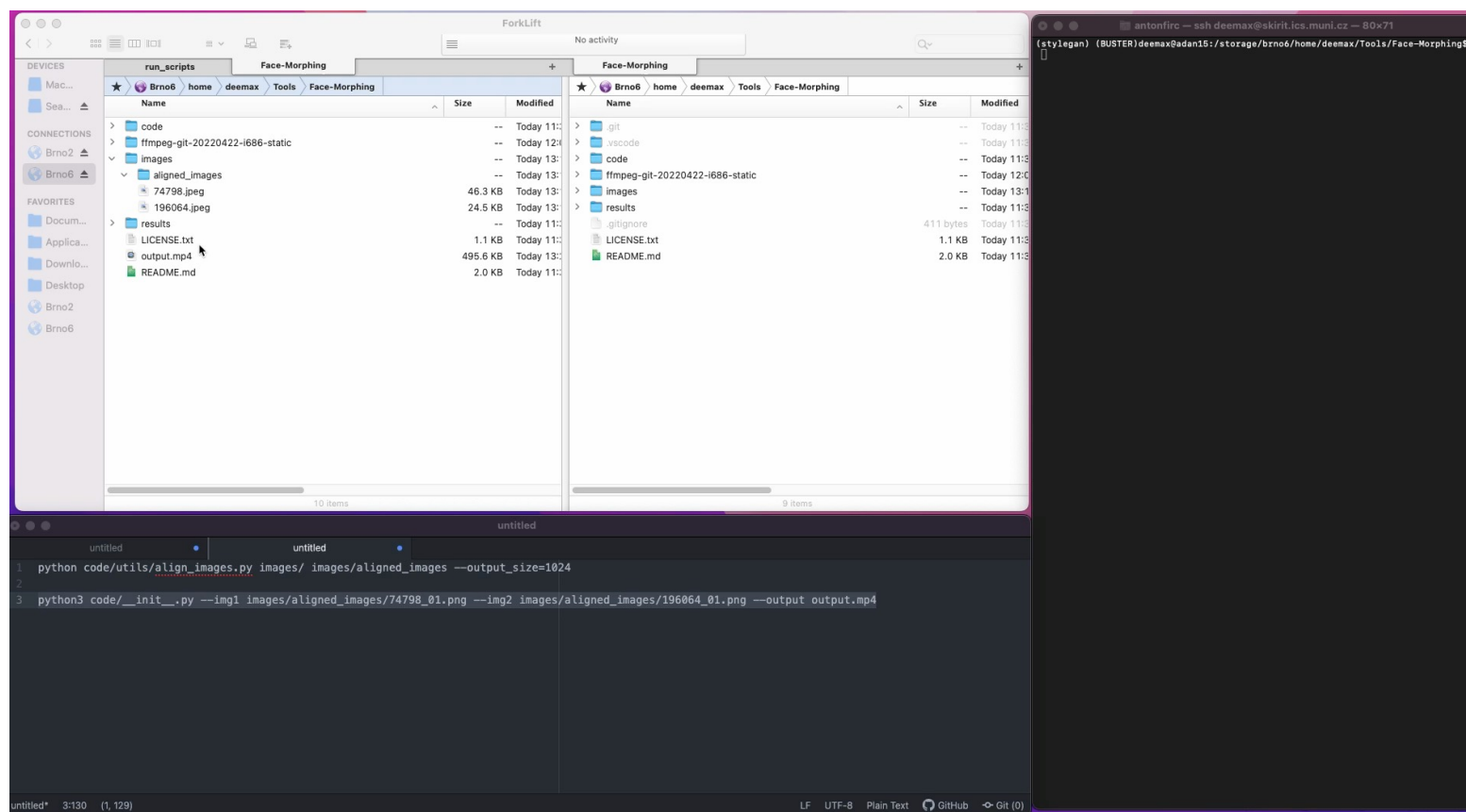
- Problém pre obličejové biometrie
 - Automatické hraniční kontroly
- Dva lidi, jeden občanský průkaz (pas)



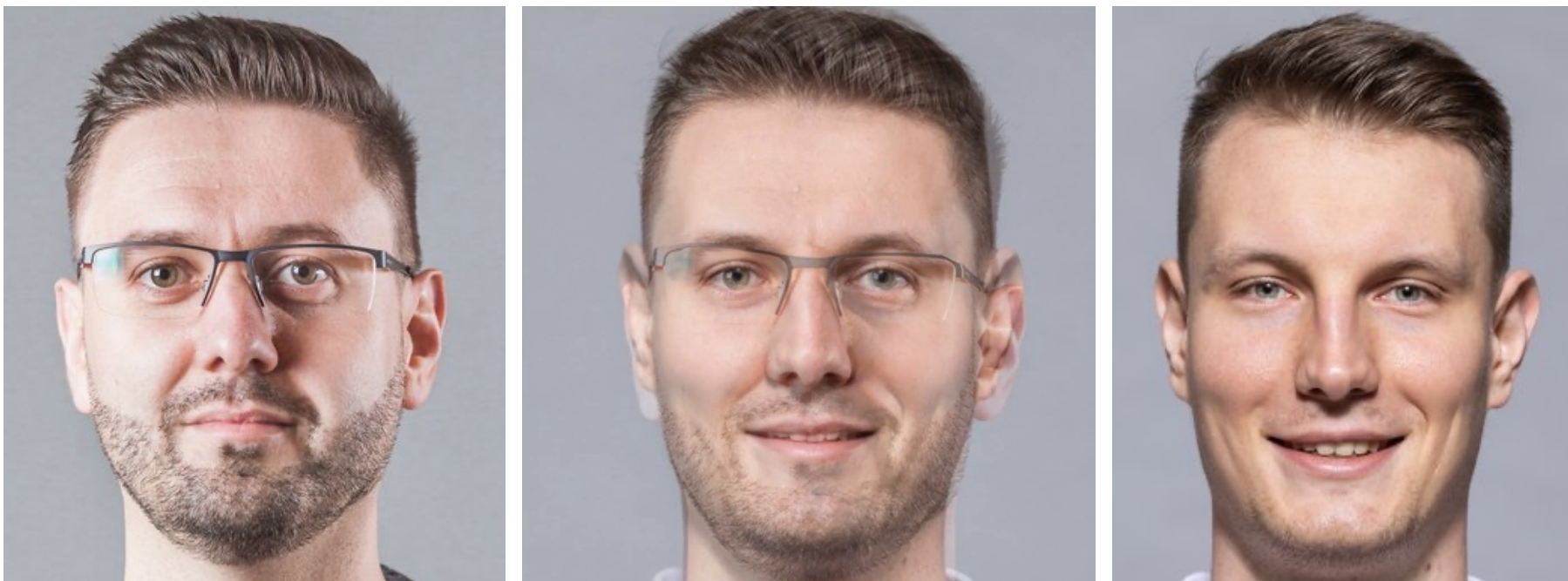
<https://www.schengenvisa.info/news/estonia-implements-new-automated-border-control-gates-for-travellers/>

Morphing obličejů – ukázka

<https://github.com/Azmarie/Face-Morphing>

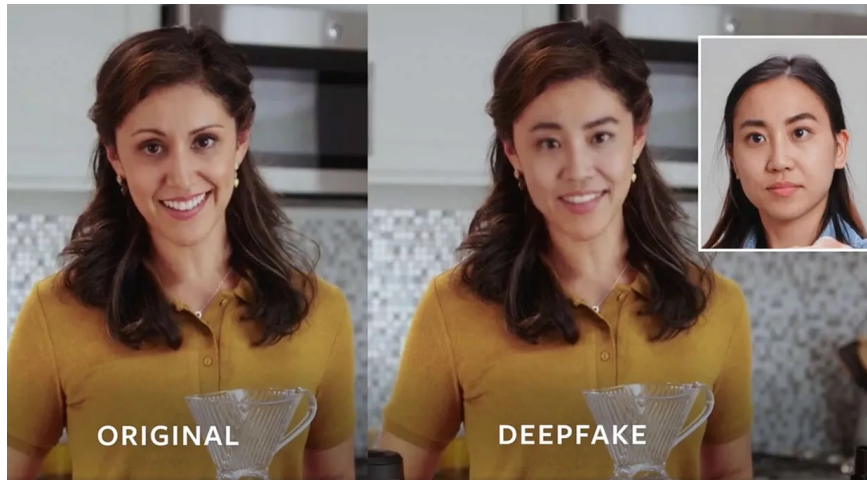


Morphing obličeje – ukázka



Výměna obličeje (Face swap)

- Obličej ze zdrojové fotografie je přenesen na hlavu v cílové fotografii
- Obrázky, ale i video
- Extrakce obličeje ze zdrojové osoby -> lokalizace obličeje cílové osoby -> transpozice -> smíchaní pixelů na okrajích



<https://spectrum.ieee.org/facebook-ai-launches-its-deepfake-detection-challenge>



<https://clippingpanda.com/photoshop-face-swap/>

Face swap - technologie

- Metody hlubokého učení
 - GAN
 - VAE
 - CNN
- Proces:
 1. Extrakce obličejů, jejich zarovnání a segmentace
 2. Naučení funkce přenosu obličeje z hlavy na druhou hlavu
 3. Přeložení obličeje pomocí natrénované funkce a retušování výsledku

Face swap – útočník

- Útok na biometrické systémy
- Krádež identity
- Pomluva
- Manipulace s důkazy



<https://sumsub.com/blog/learn-how-fraudsters-can-bypass-your-facial-biometrics/>

Face swap – ukázka

<https://github.com/iperov/DeepFaceLab>



”Rozpohybování” (Reenactment)

- Foto-realistická re-animace obrázku podle zdrojového videa
- Generování vizuální informace a synchronizace audia a videa

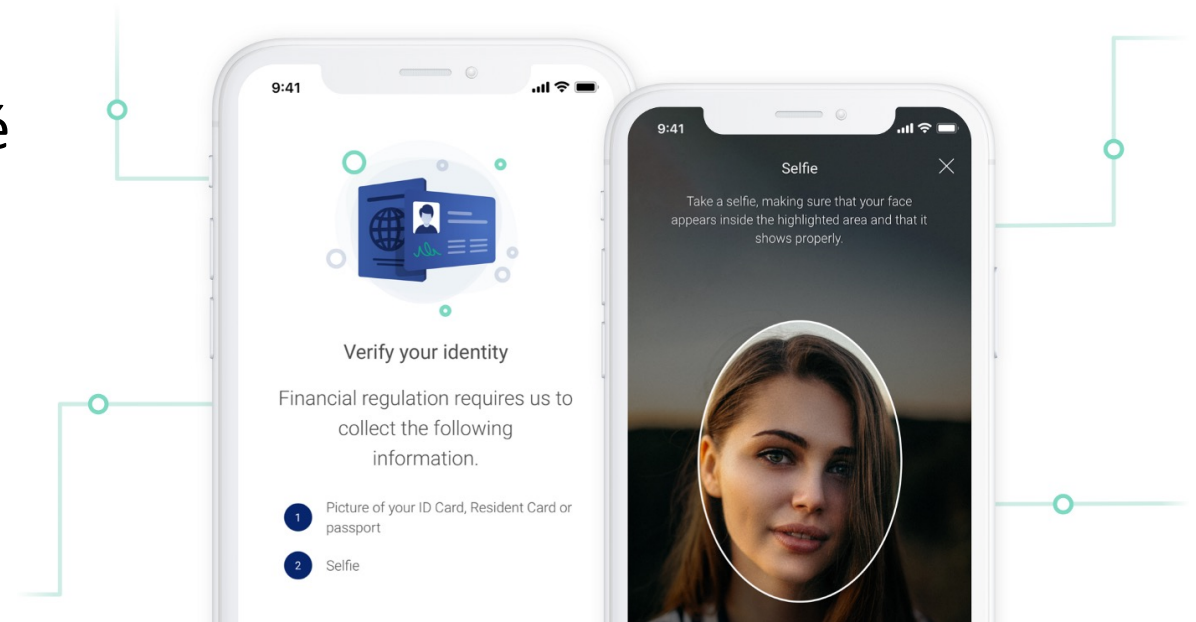


Reenactment – technologie

- Video-driven reenactment
 - Parametrický model obličeje
 - Úprava cílového obličeje parametry zdrojového obličeje
- Audio-driven animation
 - Predikce pohybu obličeje dle řeči
 - LSTM, mapování hlásek na parametry obličejů
- Audio-driven reenactment
 - Generuje pohybující se obličej dle zdrojové řeči
 - Primárně GANy

Reenactment – útočník

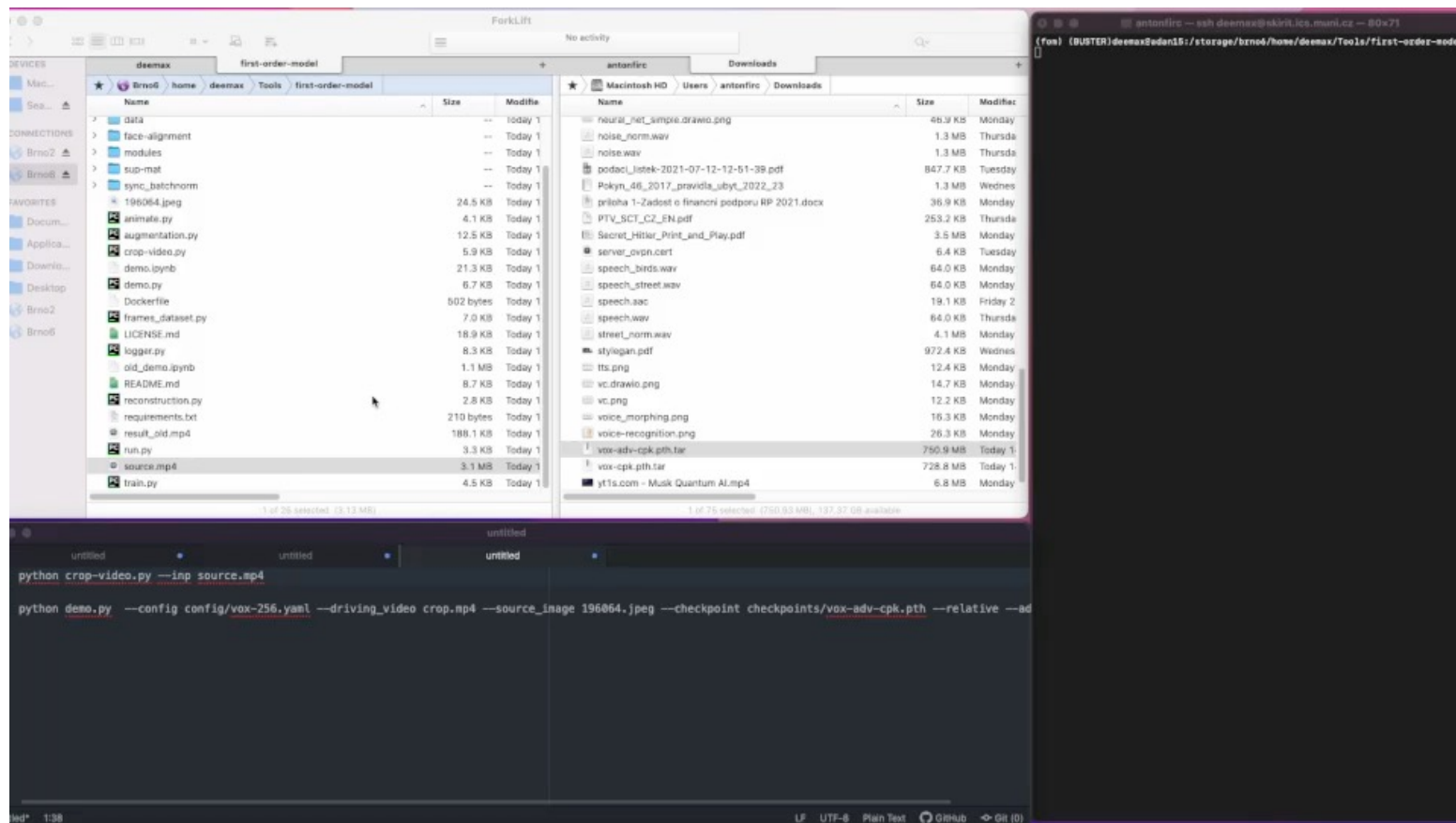
- Krádež identity
- Falešné zprávy
- Know Your Customer proces
 - Ověření identity v aplikaci na základě fotky dokladu a nasnímání obličeje
 - Možnost oklamat test živosti



<https://medium.com/2gether/this-is-why-the-kyc-know-your-customer-is-so-important-for-you-df3f5f1f4e29>

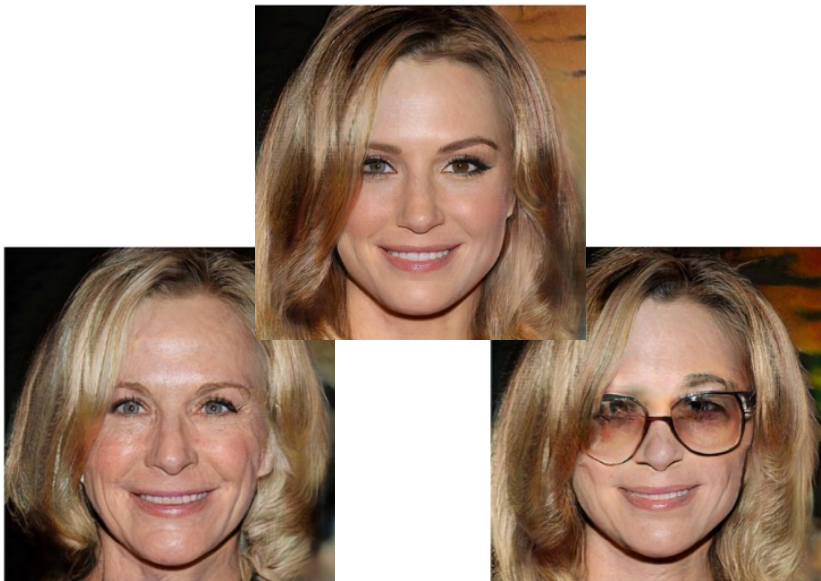
Reenactment – ukázka

<https://github.com/AliaksandrSiarohin/first-order-model>



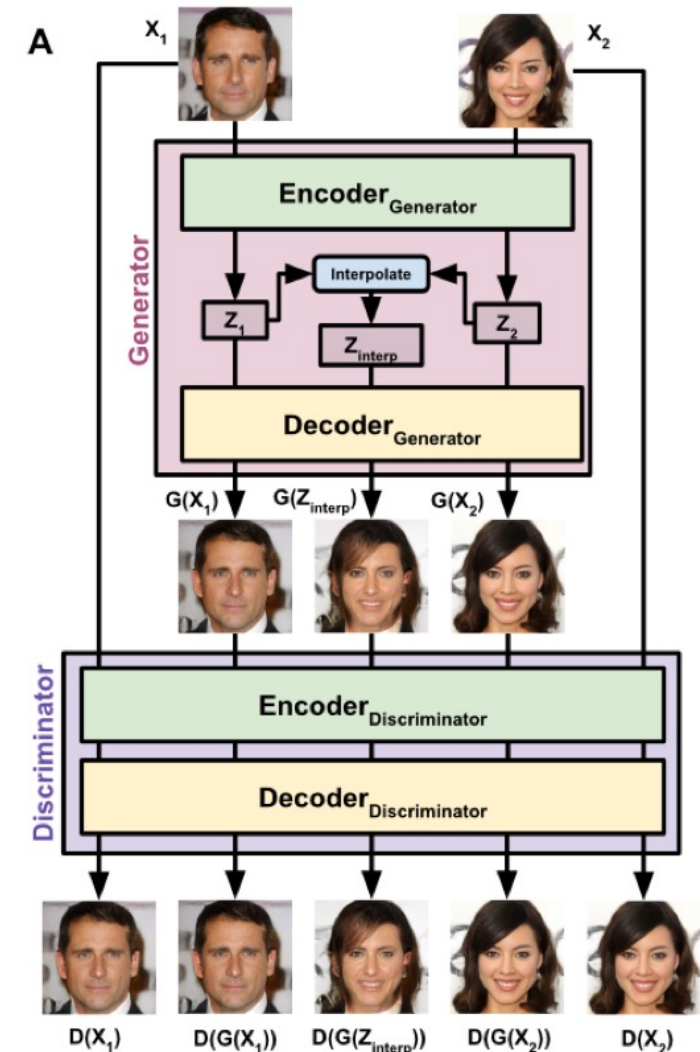
Manipulace obličeje

- Technika úpravy vybrané části obličeje v obrázku nebo na videu
- Přidání nebo odstranění vlasů, brýlí nebo tetování
- Úprava atributů v latentním prostoru a zpětná generace obličeje



Manipulace obličeje - technologie

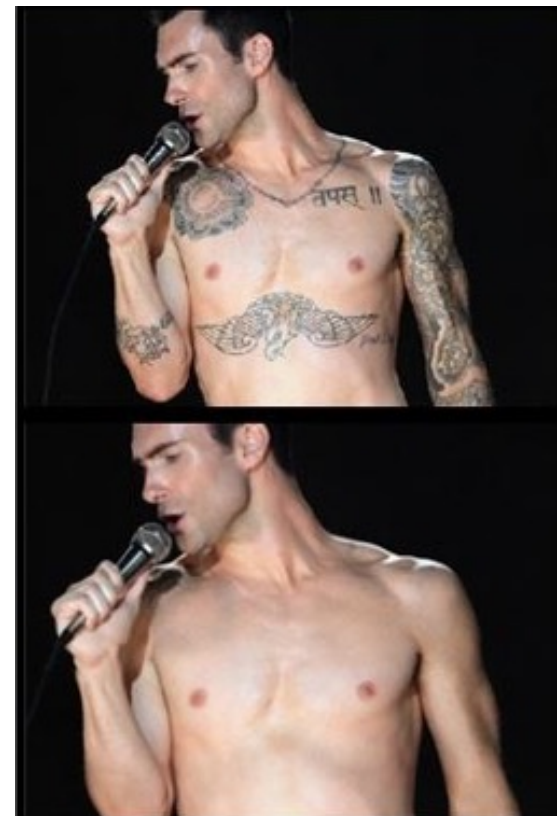
- Primárně GAN a VAE
- Velká podobnost se syntézou obličejů
- Úpravy latentního kódu obličejů



<https://timsainburg.com/assets/PDF/GAIA.pdf>

Manipulace obličeje – útočník

- Úprava atributů obličeje tak, aby osoba nebyla rozpoznána biometrickým systémem
- Manipulace důkazů u soudu



<https://github.com/vijishmadhavan/SkinDeep>

Manipulace obličej - ukázka

<https://github.com/genforce/interfacegan>

▶ Fetch Codebase and Models

[] ↪ 1 cell hidden

▶ Define Utility Functions

[] ↪ 1 cell hidden

▼ Select a Model

12s

▶

model_name: stylegan_celebahq

latent_space_type: W

Show code

20s

▶

```
[2022-05-20 14:08:01,341][INFO] Loading pytorch model from `models/pretrain/stylegan_celebahq.pth`.
[2022-05-20 14:08:01,621][INFO] Successfully loaded!
[2022-05-20 14:08:01,624][INFO] `lod` of the loaded model is 0.0.
```

▼ Sample latent codes

1s

16s

[4]

num_samples:

noise_seed:

Známé filmy, noví herci

Sám doma?



Iron Man?



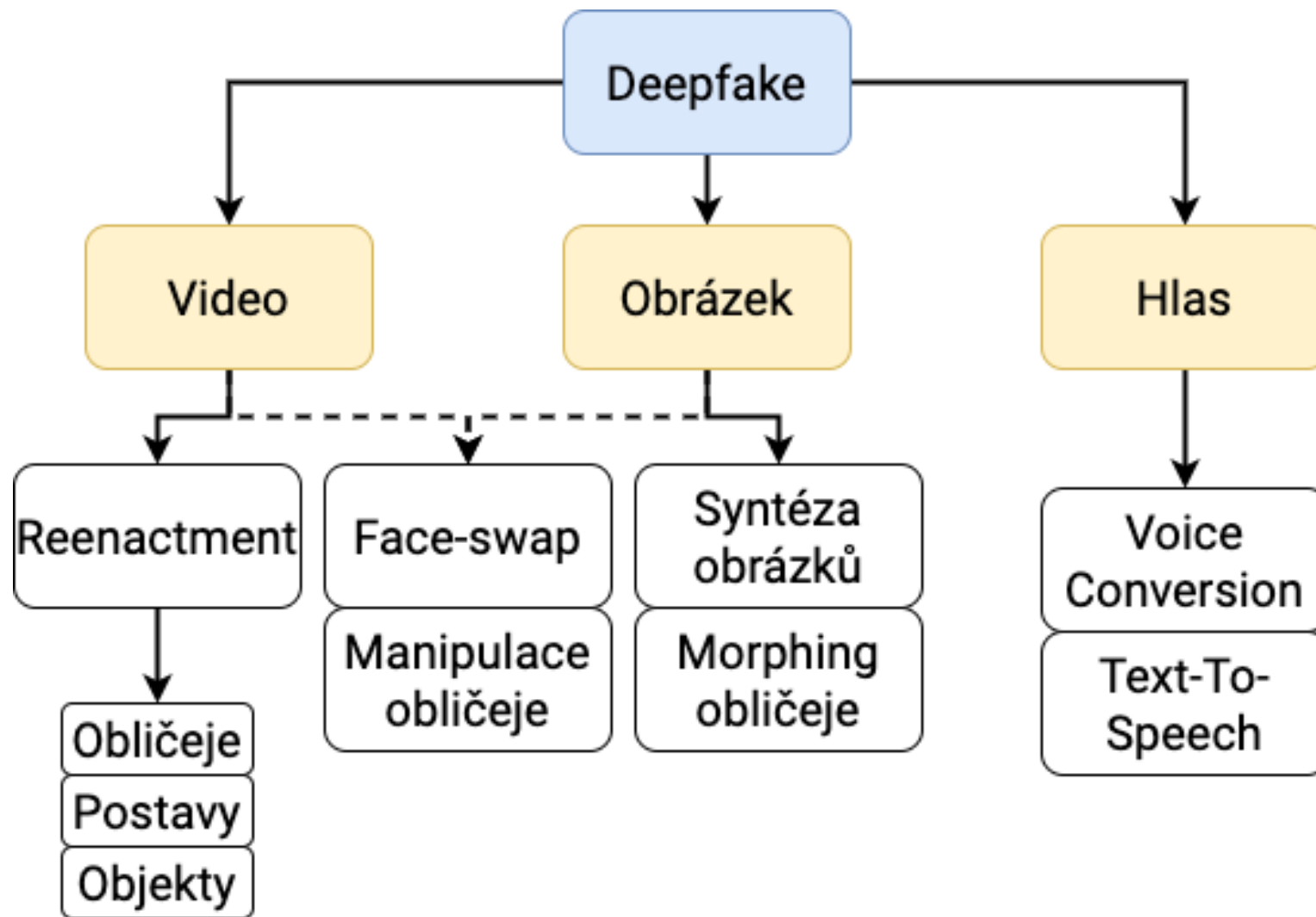
Nástroje na tvorbu deepfakes - obličej

Nástroj	Syntéza obličej	Morphing obličej	Face swap	Reenactment	Úprava obličej
StyleGAN	✓	x	x	x	x
MMGeneration	✓	x	x	x	x
Face-Morphing	x	✓	x	x	x
DeepFaceLab	x	x	✓	x	x
FaceShifter	x	x	✓	x	x
Face2Face	x	x	x	✓	x
NeuralVoicePuppetry	x	x	x	✓	x
SkinDeep	x	x	x	x	✓
InterFaceGAN	x	x	x	x	✓

Datasety - obličej

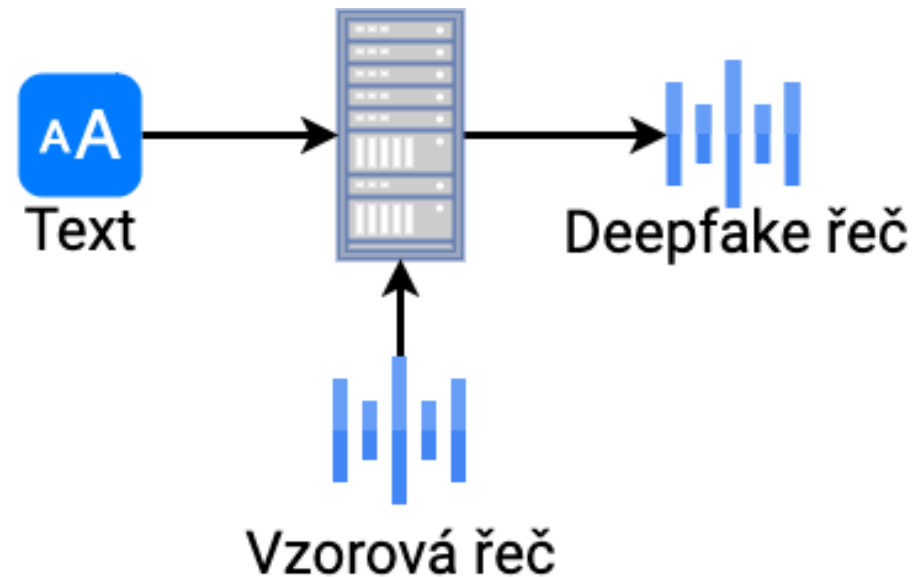
Dataset	Syntéza obličej	Morphing obličej	Face swap	Reenactment	Úprava obličej
PGGAN	✓	x	x	x	x
iFakeFaceDB	✓	x	x	x	x
FaceForensics	x	x	✓	✓	✓
FaceForensics++	x	x	✓	✓	✓
Celeb-DF	x	x	✓	x	x
DeeperForensics-1.0	x	x	✓	x	x

Rozdělení deepfakes



Syntéza: převod textu na řeč (Text-to-speech)

- Převod textu na řeč v hlase ze vzorové nahrávky

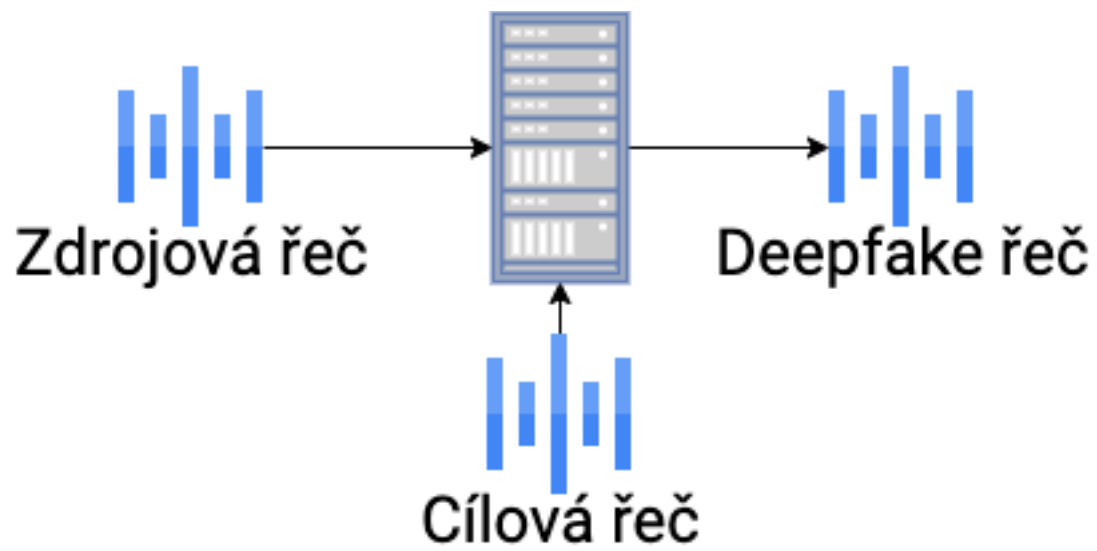


TTS syntéza - technologie

- Formant synthesis
 - Přímě modeluje frekvence signálu řeči
 - Formant = rezonanční frekvence hlasivek
- Articulatory synthesis
 - Modeluje chování hlasivek
- Concatenative synthesis
 - Spájení krátkých segmentů původní řeči do nové podoby
 - GMM, HMM -> neuronové sítě (transformer modely)

Změna hlasu (Voice Conversion)

- Změna hlasu v zdrojové nahrávce na hlas z cílové nahrávky při zachování obsahu zdrojové nahrávky



Voice cloning - variace

- Paralelní / Ne-paralelní
 - Vyžaduje páry stejných nahrávek pro trénink?
- Text-dependent / Text-independent
 - Je potřeba textový přepis nahrávek?
- Language-dependent / Language-independent
 - Funguje jen v jazyku v kterém proběhl trénink?

Voice cloning - technologie

- Paralelní
 - GMM, HMM
 - Bilineární modely, lineární regrese
 - Neuronové sítě
- Ne-paralelní
 - GAN
 - AE
 - VAE

Hlasové deepfakes - útočník

- Vhishing – voice phishing
- Krádež identity
- Pomluva



<https://whatismyipaddress.com/vishing>

Nástroje na tvorbu deepfakes - hlas

Nástroj	Text-to-speech	Voice Conversion
Real-Time-Voice-Cloning	✓	✗
MozillaTTS	✓	✗
CoquiTTS	✓	✗
YourTTS	✓	✓
FragmentVC	✗	✓
StarGAN	✗	✓
AdaptiveVC	✗	✓

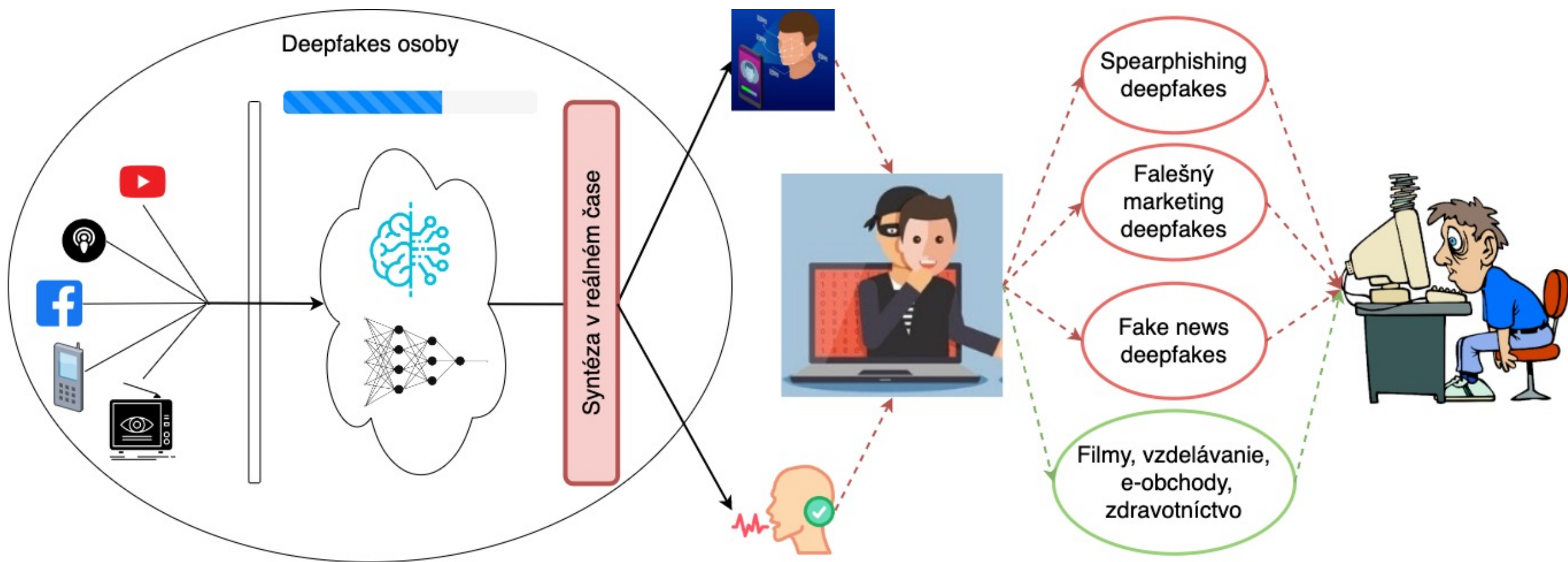
Datasety - hlas

Dataset	Text-to-speech	Voice Conversion
ASVspoof	✓	✓
WaveFake	✓	x
FoR	✓	x
SYNSPEECHDB	✓	✓
FMFCC-A	✓	✓

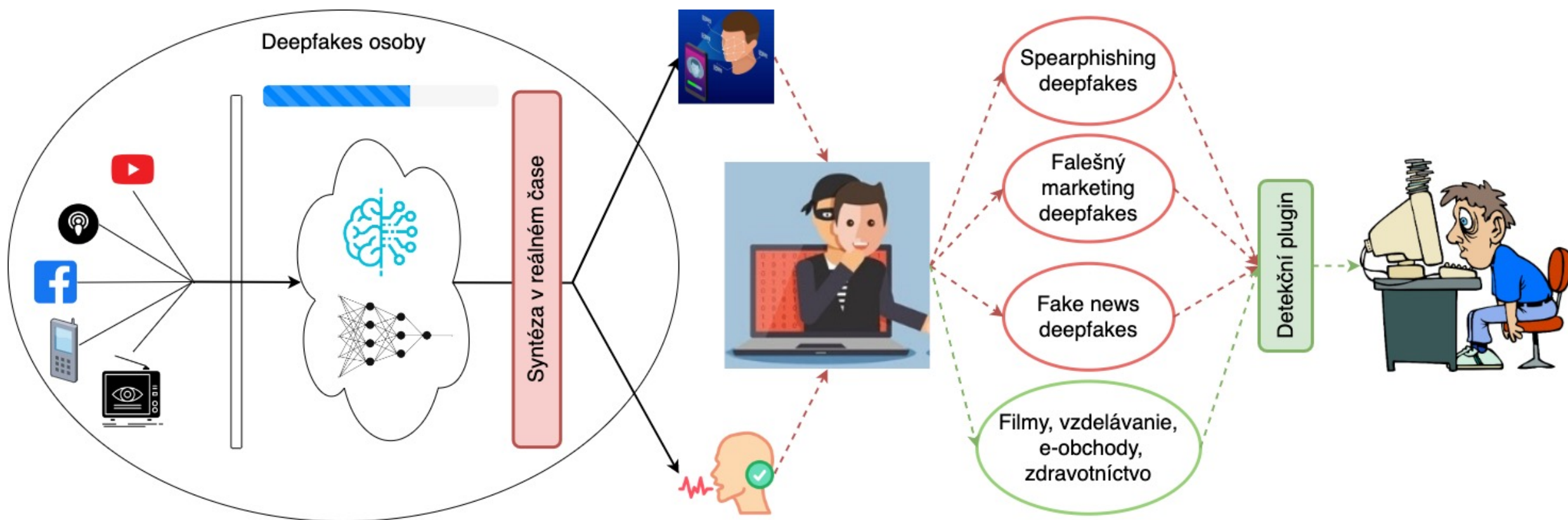
Závěr?

- Deepfakes mohou být nebezpečné
- Je potřeba věnovat se vývoji kvalitní detekce a zvyšovat lidské povědomí
- Je možné, že v dohledné době nebudeme schopni rozlišit pravé a deepfake média...

Je před námi temná budoucnost?



Pracujeme na řešení :)



Ať nekončíme tak negativně... :)

