

# Deepfake technologie

Karel Škubal

Faculty of Mechanical Engineering, Brno University of Technology  
Institute of Automation and Computer Science  
Technicka 2896/2, Brno 616 69, Czech Republic  
192822@vutbr.cz

*Abstract: Tato práce se zabývá stručným seznámením s technologií deepfake. V první části je přiblížen tento vcelku nový pojem a jeho krátká historie. Ve druhé části je na příkladech popsán princip.*

*Keywords: deepfake, hluboké učení, AI, neuronová síť*

## 1 Úvod

Hluboké učení je využíváno k řešení mnohých komplexních problémů - od analýzy velkých objemů dat po počítacové vidění a autonomní ovládání robotů. Velké pokroky v technologii, jež vedly k rychlému rozvoji hlubokého učení a umělé inteligence, však zapříčinily i vznik programů, které mohou představovat podstatné bezpečnostní hrozby. Jednou z nejnovějších možných hrozeb je vznik takzvaných deepfake médií. [5]

## 2 Co je deepfake

Pojem vznikl spojením anglických slov "deep learning" a "fake". Jedná se o využití hlubokého učení a umělé neuronové sítě k analýze velkých objemů dat, pomocí kterých se následně učí napodobovat obličejeové výrazy, neverbální komunikaci, hlas, nebo i samotnou skladbu jazyka, což umožňuje poté efektivně vytvářet hyperrealistické fotografie, videa a záznamy, které v některých případech téměř není možné lidským okem poznat od skutečnosti. [5] Přestože použití AI výrazně zrychlilo tento proces, pro dosažení uvěřitelného výsledku je stále třeba manuálně poupravovat mnoho faktorů natrénovaného programu pro potlačení různých artefaktů a dalších drobných nesouvislostí. [1]

## 3 Rozvoj deepfake technologií

Úpravy médií sahají až do 19. století. Tehdy se jednalo pouze o jednoduché úpravy fotografií. Jako konkrétní případ lze uvést fotografiu bývalého amerického prezidenta Abrahama Lincolna přibližně z roku 1865. Tato fotografie byla vytvořena překrytím hlavy na portrétu Johna C. Calhouна. [2, 3]

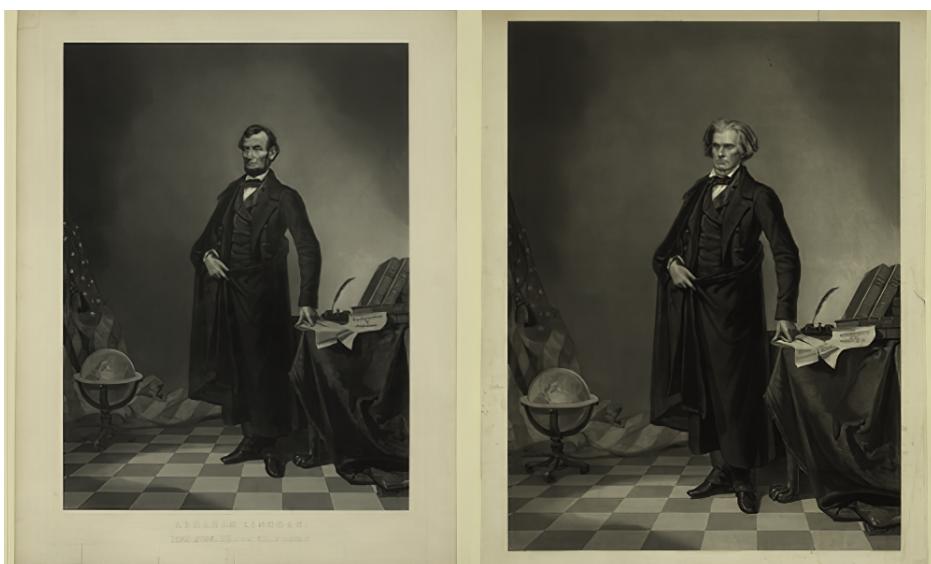


Figure 1: Upravená fotografie Abrahama Lincolna [3]

V kinematografii je již dlouho využíváno postprodukce k vytváření velmi reálně vypadajících situací. Ve velké většině případů je však použito počítačem generované grafiky (CGI). Tyto úpravy jsou však velmi časově náročné a pracné. Pro jejich vytváření je potřeba drahý software, hardware a trénovaný personál s mnohaletými zkušenostmi. [4]

Prvním projektem, který by se dal nazvat deepfakem, byl Video Rewrite vytvořený v devadesátých letech 20. století. Umožňoval ze stávajícího videa a nově vložené zvukové stopy, podle které poté byly sesynchronizovány osoby, posléze vytvořit nové video. [7, 8]

K velkému zlomu v popularitě došlo v roce 2017, kdy se deepfake videa začala objevovat různě po internetu. Začala se vynořovat například videa rozhovorů důležitých osob, ve kterých byla jejich řeč sesynchronizována na jinou zvukovou stopu. Nejrozšířenějším zdrojem těchto videí byla internetová stránka reddit, na které uživatel subredditu r/deepfakes využíval AI převážně pro vytváření deepfake pornografie, kde byl v původním videu obličej herce nahrazen jinou celebritou. [4, 10] Toto využití deepfake technologie bylo označeno jako nedobrovolná pornografie a subreddit byl zablokován. [6]

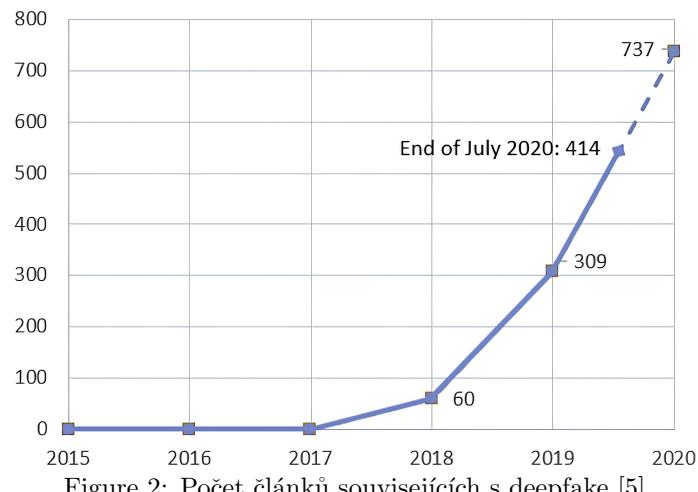


Figure 2: Počet článků souvisejících s deepfake [5]

## 4 Princip deepfake

V dnešní době se deepfake software dostává ke stále více uživatelům. Je to zapříčiněno neustále se zjednodušujícím použitím a vysokou výslednou kvalitou upravených videí s nevelkým úsilím.

Jednou z nejpoužívanějších variant vytváření deepfake obsahu využívá autoencoder. Při této metodě autoencoder extrahuje rysy z obrazů obličeje, které jsou poté pomocí dekodéru použity k rekonstrukci obličeje. K prohození dvou obličejů mezi zdrojem a cílem je potřeba vždy dvou páru kodér-dekodér. Tato metoda umožňuje najít a porovnat podobnosti mezi dvěma obličeji. [5]

Konkrétním příkladem je následující obrázek, kde byly pomocí deepfake zaměněny obličeje známých celebrit Alison Brie a Jim Carrey. [4]



Originální scéna ukazující Alison Brie

Deepfake náhrada Jima Carrey za Brie

Figure 3: Deepfake celebrit Alison Brie a Jim Carrey [4]

Pro vytvoření tohoto deepfake videa bylo potřeba tří kroků. V prvním kroku byl extrahován obraz originálního obličeje (nalevo). Ten byl poté použit jako vstup v hluboké neuronové síti, následně bylo použito strojové učení k automatické generaci patřičného výstupu s druhým obličejem. V posledním kroku byl tento vygenerovaný obličej vložen do originálního snímku pro vytvoření deepfake. [4]

### Tři kroky potřebné k vytvoření deepfake

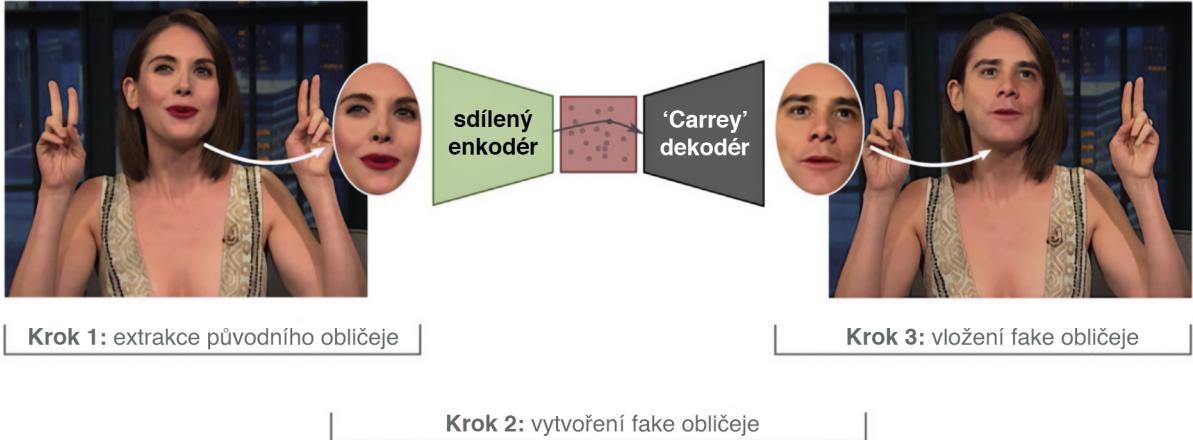


Figure 4: Deepfake celebrit Alison Brie a Jim Carrey [4]

## 4.1 Hluboké učení

K tomu, aby byl program schopný rozlišit různé rysy a výrazy obličeje, je použito hluboké učení - technika strojového učení, která může být použita k trénování hluboké neuronové sítě. Jak již název napovídá, neuronové sítě jsou tvořeny velkým množstvím umělých neuronů, také zvaných jako jednotky. Stejně jako v lidském mozku, každá tato samostatná jednotka provádí poměrně jednoduchý výpočet, avšak propojením všech těchto neuronů dohromady je možné řešit i složité nelineární problémy, jako je rozpoznaní konkrétní osoby z fotografie či videa.

Netrénovaná umělá neuronová síť obsahuje pouze náhodná spojení mezi jednotkami, z čehož plyne náhodný tok dat a tím i náhodný výsledek, tudíž jejím použitím by správný výsledek mohl být dosažen pouze čirou náhodou. Naopak, vycvičená neuronová síť má zesílené spoje mezi určitými jednotkami, tudíž je s ní možné dosáhnout kýzeného výsledku. Cílem hlubokého učení je tedy posílení spojů mezi umělými neurony, což postupně vede k minimalizaci výstupní chyby. [4]

## 4.2 Auto-enkodér

Proces učení auto-enkodéru spočívá v rozeznávání poměrně malého množství obličejobyých rysů ve vstupních datech. Jeho hlavním cílem je vzít vstupní obraz a zredukovat jej na právě tato klíčová data, jako je barva a tvar očí, barva kůže, emoční výrazy, a podobně. Pro následnou rekonstrukci je nutné, aby byl na vstupní snímky obou měněných osob použit stejný enkodér, jelikož u odlišných neuronových sítí mohou být jinak ohodnoceny dané obličejobyé rysy, tudíž by bylo dosaženo jiných hodnot v latentním prostoru.

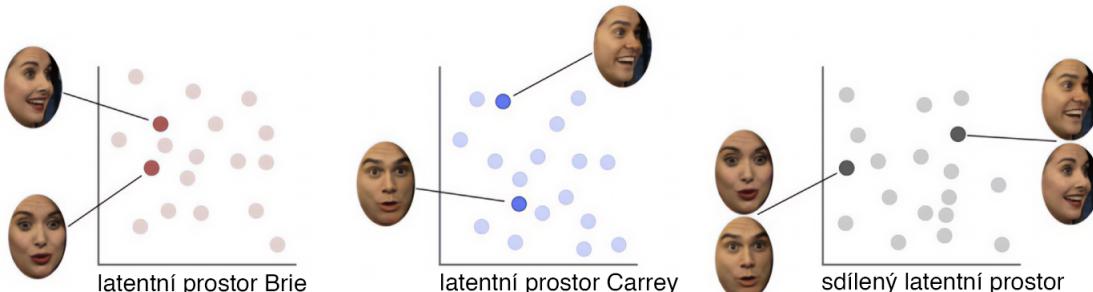


Figure 5: Latentní prostor [4]

## 5 Dopad deepfake

Možnost na počkání vytvořit snímek videa, který je téměř nerozpoznatelný od skutečnosti, s sebou přináší nespoleč nových možností. Od filmového průmyslu, přes vzdělání, až po zábavu. Deepfake přináší nové možnosti tvorby obsahu. Často je však debatován záporný dopad na společnost a možná negativní zneužití této technologie. [1]

Jako příklad pozitivního využití deepfake lze uvést třeba exponát v Muzeu Holokaustu v Illinois, kde je použit interaktivní hologram, jenž umožňuje návštěvníkům vyslechnout si příběh pamětníků holokaustu. [9]



Figure 6: Interaktivní hologram pamětníka holokaustu [9]

Další výhodou je, že pomocí deepfake animace lze překonat jazykovou bariéru, a zjednodušit tak přístup k médiím. Konkrétním příkladem tohoto využití je kampaň, která uváděla Davida Beckhama, jak mluví devíti světovými jazyky. Toto konkrétní využití by mohlo být přínosné i pro budoucí dabing filmů, kdy bude dosaženo perfektní synchronizace rtů. [9]

Šedou zónou deepfake je použití k zábavě. Na jednu stranu je to velmi silný nástroj pro tvorbu reálně vypadajících komedických scén. Na druhé straně vznikají kontroverzní videa politiků, která mohou časem ohrozit i národní bezpečnost. Proti tomuto typu škodlivému využití jsou již v některých amerických státech zavedeny zákony, které kriminalizují vytváření a distribuci již zmínované deepfake pornografie.

Zároveň je snaha o tvorbu a rozvoj aplikací, které zvládnou odhalit a případně označit deepfake videa například vodoznakem, čímž by se dalo zabránit případné dezinformaci. [1]

## Závěr

Úpravy digitálního obsahu nejsou žádná novinka. Většinou byly ale takovéto úpravy závislé jak na komplexnosti používaného programu, tak na zkušenostech člověka, který úpravy prováděl, aby bylo dosaženo reálně vypadajícího výsledku.

Přínos hlubokého učení je ten, že dříve nebo později se tyto technologie dostanou k běžným uživatelům, kteří mimo poskytnutí vstupních dat do neuronové sítě budou moci bez námahy upravovat celá videa. Spolu s prakticky nemožnou rozeznatelností od neupraveného obsahu by byl obsah na internetu stále méně důvěryhodný a téměř nerozeznatelný od pravdy. Toto je neutralizováno výzkumem do aplikací, které odhalují upravený obsah. S dalším rozvojem je možné, že každé médium bude muset obsahovat přesně daná zakódovaná data o místě a datu pořízení pro snadnou ověřitelnost pravosti.

## References

- [1] ADEE, S. What are deepfakes and how are they created. *Deepfakes*, 2 (2020).
- [2] AGARWAL, S., NORMAN, J., SEHGAL, V., AND THAKKAR, N. Photo tampering throughout history. *Photo tampering*, 1 (2021).
- [3] CENTER, B. D. Altered images. *Photo tampering*, 1 (2021).
- [4] KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., AND KIETZMANN, T. C. Deepfakes: Trick or treat? *Business Horizons* 63, 2 (2020), 135–146.
- [5] NGUYEN, T. T., NGUYEN, C. M., NGUYEN, D. T., NGUYEN, D. T., AND NAHAVANDI, S. Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573* (2019).
- [6] ROBERTSON, A. Reddit bans ‘deepfakes’ ai porn communities. *redditban*, 1 (2018).
- [7] SHAWN. The rise of deepfake technology: Where does it end? *Deep Learning 2020*, 1 (2020).
- [8] SONG, D. A short history of deepfakes. *Deepfakes*, 1 (2019).
- [9] THINKAUTOMATION. Yes, positive deepfake examples exist. *Deepfakes*, 1 (2020).
- [10] WESTERLUND, M. The emergence of deepfake technology: A review. *Technology Innovation Management Review* 9, 11 (2019).