PRÁVO VERSUS *DEEPFAKES*

TINA MIZEROVÁ*

Abstract: **Law Versus Deepfakes**
The text focuses on deepfakes, which is audiovisual content created by artificial intelligence that depicts people, things or events and which the audience perceive as authentic or realistic. The proliferation of deepfakes in cyberspace entails a wide risk of their misuse. The aim of the text is to introduce and describe deepfakes, their misuse and analyse regulation which targets deepfakes and their dissemination. The regulation includes mainly the European regulation of digital space, i.e., the AI Act, Digital Services Act, GDPR and other related regulation. In addition to the analysis of concrete regulation, the text addresses the question of whether regulation can prevent the dangerous misuse of deepfakes and respond to the negative consequences of such proliferation.

Keywords: deepfake; artificial inteligence; Artificial Inteligence Act; Digital Services Act; authenticity; transparency

Klíčová slova: *deepfake*; umělá inteligence; Akt o umělé inteligenci; nařízení o digitálních službách; autenticita; transparentnost

DOI: 10.14712/23366478.2025.394

ÚVOD

Na začátku roku 2024 se na sociálních sítích začala objevovat videa českých politiků nabízejících investice s nevídaným zhodnocením. Jednalo se ale o podvod, který způsobil škodu téměř tři miliony korun[1] a před kterým varoval i prezident Pavel, jehož podvodná kopie investice nabízela také.[2] Podvodníci využili *deepfakes*: videa vytvořená pomocí umělé inteligence, která působí natolik autenticky, že mohou oklamat své publikum. *Deepfakes* poprvé vstoupily do povědomí v roce 2017 v souvislosti

* Mgr. et Mgr. Tina Mizerová, Masarykova univerzita, Právnická fakulta, Ústav práva a technologií; e-mail: 471315@mail.muni.cz; ORCID: 0009-0000-6003-9562.

1 ŠVIHEL, P. Přišli jsme o všechno a stydíme se. Pár uvěřil, že investice zaštiťuje Petr Pavel. In: *Seznam Zprávy* [online]. 25. 4. 2023 [cit. 2024-05-28]. Dostupné na: https://www.seznamzpravy.cz/clanek/domaci-zivot-v-cesku-podvodnici-se-zastitili-prezidentovym-jmenem-seniorsky-par-prisel-o-vsechno-229909.

2 VRLÁK, M. Pavel varuje před deepfake. In: *ČT24 – Česká televize* [online]. 6. 7. 2024 [cit. 2024-07-12]. Dostupné na: https://ct24.ceskatelevize.cz/clanek/domaci/pavel-varuje-pred-deepfake-nikdo-z-ustavnich-cinitelu-by-nedelal-reklamu-pochybne-komercni-350902.

s šířením falešných pornografických fotografií a videí celebrit.[3] Od té doby se značně rozšířila kapacita a dostupnost generátorů *deepfakes*[4] a s jejich rostoucí dostupností zároveň roste i riziko zneužití *deepfakes*. Vzniká tak řada opatření s cílem reagovat na riziková šíření *deepfakes* a mezi tato opatření patří i právní regulace.

Cílem následujícího článku je představit *deepfakes* a jejich riziková využití, identifikovat právní úpravu, která na *deepfakes* dopadá, a zhodnotit, zda představuje efektivní nástroj pro předcházení těmto rizikovým využitím a omezení jejich negativních důsledků. První část článku věnuji obecnému představení *deepfakes*. V druhé části článku se soustředím na právní úpravu, která na *deepfakes* dopadá. Konkrétně jde o nařízení o ochraně fyzických osob v souvislosti se zpracováním osobních údajů,[5] Nařízení o digitálních službách,[6] dále Kodex zásad boje proti dezinformacím z roku 2022 a Směrnici o potírání násilí vůči ženám a domácího násilí.[7] Soustředím se především na Akt o umělé inteligenci,[8] který jako první obsahuje právní úpravu zaměřenou přímo na *deepfakes*.

Při analýze legislativy se zaměřím na následující výzkumné otázky, které reflektují cíle textu: (1) jak se vybrané předpisy vztahují na *deepfakes* a zda (2) mohou zamezit zneužití *deepfakes* nebo alespoň (3) omezit negativní důsledky takových zneužití. Pro zodpovězení první otázky využívám doktrinální metodu, v případě druhé a třetí otázky využiji sociálně-legální metodu, kdy budu pracovat s poznatky o *deepfakes* z oblasti jiných vědních oborů.

DEEPFAKES – AUTENTICKÝ AUDIOVIZUÁLNÍ OBSAH ZKRESLUJÍCÍ REALITU

KOŘENY FENOMÉNU *DEEPFAKES*

Slovo *deepfake* se poprvé objevilo v souvislosti se zmíněným šířením intimních videí a fotografií celebrit na sociální sítí Reddit, kdy tento obsah jako první zveřejnil uživatel vystupující pod pseudonymem Deepfake. Název je složený z anglických slov *„deep"*, odkazující na *„deep learning"*, což je proces, při kterém se nástroje strojového učení trénují v rozpoznávání a klasifikaci dat,[9] a slova *„fake"*, označující falešný obsah.[10] I přes pochybný původ se pojem *deepfake* vžil pro označení

[3] KIETZMANN, J. a kol. Deepfakes: Trick or treat? *Business Horizons*. 2020, Vol. 63, No. 2, s. 136.

[4] MILLER, M. Deepfakes: Real Threat. In: *KPMG* [online]. 2023 [cit. 2024-05-28]. Dostupné na: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/deepfakes-real-threat.pdf.

[5] Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů (dále „GDPR").

[6] Nařízení Evropského parlamentu a Rady (EU) 2022/2065 ze dne 19. října 2022 o jednotném trhu digitálních služeb (dále „DSA").

[7] Směrnice Evropského parlamentu a Rady (EU) 2024/1385 ze dne 14. května 2024 o potírání násilí na ženách a domácího násilí (dále „směrnice o potírání násilí na ženách").

[8] Nařízení Evropského parlamentu a Rady (EU) 2024/1689 ze dne 13. června 2024, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci (dále „AI Akt").

[9] BENGIO, Y. – HINTON, G. – LECUN, Y. Deep learning. *Nature*. 2015, Vol. 521, No. 7553, s. 436.

[10] BROOKS, T. a kol. The increasing threat of deepfake identitites. In: *U.S. Department of Homeland Security* [online]. 2019 [cit. 2024-05-06]. Dostupné na: https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

audiovizuálního obsahu vytvořeného pomocí generativní umělé inteligence.[11] *Deepfakes* se řadí do širší kategorie syntetických médií, kam spadá veškerý mediální obsah vytvořený nebo upravený pomocí umělé inteligence. Dále se můžeme setkat s pojmy *„cheapfakes"*,[12] označujícím manipulativní audiovizuální obsah, který ale nebyl upraven pomocí umělé inteligence, a *„shallowfakes"*, což jsou běžné úpravy a vylepšení vzhledu osob na fotografiích a videích sdílených online.[13] Příkladem *cheapfake* je video předsedkyně americké sněmovny reprezentantů Nancy Pelosi, které někdo sestříhal tak, že Pelosi působila, jako by byla pod vlivem omamných látek. Příkladem *shallowfake* může být využití filtru, který na fotografii či videu vylepší naši podobu.

GENERÁTORY A DETEKTORY *DEEPFAKES*

Deepfakes (a syntetická média obecně) se od manuální úpravy fotografií liší využitím umělé inteligence, jež tvorbu obsahu zjednodušuje a zrychluje. V případě manuálních úprav je pro vytvoření autenticky působícího obsahu nutné, aby tvůrce měl přístup k příslušným technologiím a uměl s nimi zručně zacházet.[14] Například k vytvoření videa, kde byl osobě vyměněn obličej, bylo třeba použít VFX technologie, jinak primárně využívané při produkci celovečerních filmů. Nástroje generativní umělé inteligence umožňují vytvářet realistické tváře, manipulovat rysy nebo vyměňovat identitu osobám na existujících fotografiích či videích,[15] kdy tyto nástroje jsou jednoduše dostupné i na běžně využívaných chytrých zařízeních. To znamená, že syntetická média a *deepfakes* může tvořit kdokoliv.[16]

Technologie sloužící k tvorbě *deepfakes*[17] jsou *generativní soupeřící sítě* (GAN),[18] které se skládají ze dvou neuronových sítí.[19] První je *generator*, jenž vytváří obsah, a druhou *discriminator*, který posuzuje autenticitu vytvořeného obsahu tak, že jej porovnává s reálnými daty, na kterých byl *generator* trénován.[20] Konkrétní generátory *deepfakes* využívají různé modifikace GAN. Důležitou roli při trénování GAN hrají velké datové soubory, protože čím větší a rozmanitější jsou data, na kterých se systém učil, tím přesněji mohou sítě generovat autenticky působící obsah.

---

[11] SOMMERS, M. Deepfakes, explained. In: *MIT Sloan* [online]. 2. 5. 2024 [cit. 2024-05-06]. Dostupné na: https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

[12] HAO, K. The biggest threat of deepfakes isn't the deepfakes themselves. In: *MIT Technology Review* [online]. 10. 10. 2019 [cit. 2024-05-13]. Dostupné na: https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/.

[13] ANTOGNINI, A. – WOODS, A. K. Shallow Fakes. *Penn State Law Review*. 2023, Vol. 128, No. 69, s. 78.

[14] KIETZMANN a kol., *c. d.*, s. 136.

[15] JUEFEI-XU, F. a kol. Countering Malicious DeepFakes: Survey, Battleground, and Horizon. *arXiv*. 2022, No. 2103.00218, s. 2.

[16] FIRC, A. – HANÁČEK, P. – MALINKA, K. Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. *Heliyon*. 2023, Vol. 9, No. 4, s. 20.

[17] Dále „generátory *deepfakes*".

[18] Anglicky Generative Adversarial Networks (GAN).

[19] Umělé neuronové sítě, které se po vzoru neuronů v lidském těle skládají ze spousty drobných a propojených uzlů poskládaných do vrstev.

[20] GOODFELLOW, I. a kol. Generative Adversarial Nets. In: GHAHRAMANI, Z. – WELLING, M. a kol. (eds.). *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. Montreal: Curran Associates, 2014, s. 2672.

Existují i nástroje pro detekci *deepfakes*, které také využívají umělou inteligenci. Detektory se soustředí na identifikaci nedostatků *deepfakes*, a to jak technických nedostatků vznikajících při jejich generování, tak různých chyb v přirozeném zobrazení osob. Mezi technické nedostatky patří například rozdíly v tzv. *spatial domain*, kde lze *deepfake* od reálného obsahu rozpoznat pomocí různých viditelných i neviditelných fragmentů, nebo chyby ve frekvenční oblasti, kde se *deepfake* od reálného obsahu liší změnou hodnoty jednotlivých pixelů. Dále se detektory zaměřují na nedostatky v přirozeném zobrazení osob, jako je nepřirozený pohyb rtů nebo neadekvátní srdeční rytmus.[21] Také *deepfakes*, které tvoří pouze zvuk, lze od autentických nahrávek rozlišit pomocí technických i biologických nedostatků.[22] Neexistuje ale žádný univerzálně spolehlivý detektor.

Možnosti detektorů se odvíjí od dostupných dat, proto za generátory *deepfakes* vždy zaostávají. Jsou spolehlivé především při práci se známými daty, v případě nových dat jejich spolehlivost klesá.[23] Generátory se navíc snaží přicházet se způsoby, jak detektory oklamat: například pomocí odstranění výše zmíněných technických nedostatků, využití pokročilých filtrů či adversariálních útoků, při kterých se *deepfakes* drobně změní tak, aby je detektory nerozpoznaly.[24]

RIZIKOVÁ VYUŽITÍ *DEEPFAKES*

Deepfakes představují riziko hlavně kvůli třem faktorům: své přesvědčivosti, atraktivitě a obtížné možnosti ověření. Přesvědčivost i atraktivita *deepfakes* souvisí s náročností kognitivních procesů. Zatímco čtení psaného textu je kognitivně náročnější proces, obrazy či videa stačí pasivně sledovat.[25] Proto má publikum tendenci preferovat audiovizuální obsah nad psaným textem. Roli zde hraje i tzv. *reality heuristic* – audiovizuální obsah vystihuje naši každodenní realitu lépe než psaný text, což dále zvyšuje naše tendence preferovat audiovizuální obsah. Navíc si informace získané prostřednictvím obrázků, zvuku a videí zapamatujeme a následně vybavíme lépe a přesněji než informace získané z textu.[26] Od přesvědčivosti audiovizuálního obsahu se pak odvíjí jeho důvěryhodnost, kdy publikum zpravidla považuje audiovizuální obsah za objektivní zobrazení reality.[27] Ten tak často doprovází psaný text jako důkaz posilující hodnověrnost informací v textu. Pokud *deepfake* neobsahuje obraz, ale jen zvuk, stále je pro publikum jednodušší vnímat jej než vnímat psané slovo a důvěryhodnost zvukových

---

21 JUEFEI-XU a kol., *c. d.*, s. 20–30.
22 FIRC – HANÁČEK – MALINKA, *c. d.*, s. 19–20.
23 Tamtéž, s. 21.
24 JUEFEI-XU a kol., *c. d*, s. 34–36.
25 Imagery Vs Text. Which does the brain prefer? In: *World of Learning* [online]. 7. 9. 2015 [cit. 2024-07-06]. Dostupné na: https://www.learnevents.com//learning-insights/imagery-vs-text-which-does-the-brain-prefer/.
26 GRABER, D. A. Seeing Is Remembering: How Visuals Contribute to Learning from Television News. *Journal of Communication*. 1990, Vol. 40, No. 3, s. 148.
27 GEISE, S. Visual Framing. In: RÖSSLER, P. (ed.). *The International Encyclopedia of Media Effects*. Chichester: Wiley Blackwell, 2017, s. 2054.

nahrávek posiluje například využití hovorové řeči či rétorických prvků navozujících důvěrnější vztah s posluchačem.[28]

Atraktivita *deepfakes* vychází z výše zmíněné preference publika, kterou dále podporují algoritmy sociálních sítí. Například na Twitteru (dnes X) příspěvky amerických prezidentských kandidátů obsahující fotografie či grafiky nasbíraly více reakcí uživatelů[29] než tweety bez nich.[30] K podobným výsledkům došel na stejné sociální síti i výzkum zaměřený na produktový marketing.[31] Algoritmy sociálních sítí proto audiovizuální obsah preferují a uživatelům jej ukazují přednostně.[32]

Horší možnost ověření souvisí předně s výše popsanou nedostupností univerzálně spolehlivého detektoru *deepfakes*. Roli zde ale hraje i znalost publika, které může *deepfakes* odhalit. Například *deepfake* video s ministrem vnitra Rakušanem nepřesvědčilo dvě třetiny respondentů[33] a *deepfake* video vyobrazující Baracka Obamu s jistotou odhalila polovina participantů.[34] V druhém případě ale shlédnutí *deepfake* videa vedlo ke zvýšení nedůvěry participantů k online audiovizuálnímu obsahu, což poškozuje veřejnou debatu jako celek.[35] Z odborné veřejnosti pak zaznívají hlasy, že se *deepfake* generátory neustále zlepšují a blížíme se do éry, kdy nebude možné jakkoliv rozpoznat *deepfake* od skutečného audiovizuálního obsahu.[36]

Konkrétní riziková využití *deepfakes* rozdělili Chesney a Citron na (1) riziko pro osoby soukromého práva a (2) riziko pro společnost jako celek. Je třeba mít na paměti, že se tyto dvě oblasti budou vždy prolínat, neboť konkrétní případy mohou mít negativní dopad na jednotlivce i společnost a zároveň se s *deepfakes* pojí výše zmíněná rostoucí nedůvěra v audiovizuální obsah jako celek, jež negativně ovlivňuje všechny.[37]

Riziko pro jednotlivce představuje využití *deepfakes* při vydírání nebo poškození reputace jednotlivců či právnických osob zveřejněním *deepfakes*, jejichž autenticitu je obtížné rozporovat. *Deepfakes* se využívají při podvodech, jako jsou v úvodu zmíněné

---

28 KISCHINHEVSKY, M. a kol. WhatsApp audios and the remediation of radio: Disinformation in Brazilian 2018 presidential election. *Radio Journal: International Studies in Broadcast & Audio Media*. Vol. 18, No. 2, s. 151.

29 Tzv. lajků (vyjádření podpory) a retweetů (sdílení příspěvku).

30 PANCER, E. – POOLE, M. The popularity and virality of political social media: hashtags, mentions, and links predict likes and retweets of 2016 U.S. presidential nominees' tweets. *Social Influence*. 2016, Vol. 11, No. 4, s. 264.

31 LI, Y. – XIE, Y. Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*. 2020, Vol. 57, No. 1, s. 17.

32 DHANESH, G. – DUTHLER, G. – LI, K. Social media engagement with organization-generated content: Role of visuals in enhancing public engagement with organizations on Facebook and Instagram. *Public Relations Review*. 2022, Vol. 48, No. 2, s. 6.

33 Většina české populace neví, co je „deepfake". In: *CEDMO* [online]. 14. 3. 2024 [cit. 2024-05-23]. Dostupné na: https://cedmohub.eu/cs/vetsina-ceske-populace-nevi-co-je-deepfake/.

34 CHADWICK, A. – VACCARI, C. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*. 2020, Vol. 6, No. 1, s. 6.

35 Tamtéž, s. 9.

36 Many AI researchers think fakes will become undetectable. In: *The Economist* [online]. 17. 1. 2024 [cit. 2024-10-20]. Dostupné na: https://www.economist.com/science-and-technology/2024/01/17/many-ai-researchers-think-fakes-will-become-undetectable.

37 VACCARI – CHADWICK, *c. d.*, s. 7.

investice, dále *deepfakes* zdokonalují vishingové útoky,[38] například při podobném útoku na společnost GymBeam podvodníci použili *deepfake* kopii ředitele společnosti při videohovoru se zaměstnancem společnosti.[39] *Deepfakes* zvládají oklamat biometrické ověření i ovládání virtuálních hlasových asistentů.[40] Značnou část existujících *deepfakes* představují falešné intimní snímky, někdy označované jako *deep porn*,[41] které vyobrazuje především ženy a kromě poškození reputace představuje i obrovský zásah do lidské důstojnosti, a svým obětem může způsobit psychickou újmu.[42] *Deep porn* se již dotklo řady známých osobností, například zpěvačky Taylor Swift[43] nebo italské premiérky Giorgii Meloni,[44] ale také mnohem méně známých českých influencerek[45] nebo nezletilých španělských školaček.[46]

Nebezpečí pro společnost *deepfakes* představují tehdy, když zkreslují diskusi o důležitých politických otázkách, podrývají důvěru v demokratické instituce, vyostřují společenské rozdíly, ovlivňují volební výsledky, narušují diplomatické vztahy, činnost ozbrojených nebo zpravodajských služeb či poškozují ekonomiku.[47] Příkladem takto nebezpečného *deepfake* je audio napodobující hlasy předsedy politické strany Progresivné Slovensko Michala Šimečky a novinářky Denníku N Moniky Tódové. *Deepfake*, ve kterém hlasy obou osob hovoří o falšování voleb ve prospěch Šimečkovy strany, se začal šířit těsně před začátkem voleb do Národné rady v roce 2023.[48] Není možné přesně zjistit, jak tento *deepfake* ovlivnil výsledky voleb, nicméně kombinace rychlého šíření nahrávky, dojmu její autenticity a omezené možnosti aktérů a médií na ni reagovat[49] přispěly ke zvýšení rizika, že nahrávka mohla do rozhodování voličů zasáhnout.

---

38 Za vishing se označují podvodné telefonáty, jejichž cílem je od oběti získat citlivé údaje nebo finanční prostředky.

39 BREJČÁK, P. Deepfake v byznysu: Kolega volal s mou kopií čtvrt hodiny, říká šéf miliardového GymBeamu. In: *CzechCrunch* [online]. 16. 8. 2023 [cit. 2024-05-06]. Dostupné na: https://cc.cz/deepfake-v-byznysu-kolega-volal-s-mou-kopii-ctvrt-hodiny-rika-sef-miliardoveho-gymbeamu/.

40 FIRC – HANÁČEK – MALINKA, *c. d.*, s. 23–24.

41 VAN DER SLOOT, B. – WAGENSVELD, Y. Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review*. 2022, Vol. 46, s. 13.

42 CITRON, D. K. – CHESNEY, R. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*. 2019, Vol. 107, No. 1753, s. 1773.

43 BALLOVÁ, D. Falešné snímky Taylor Swift zaplavily internet. Ženy a děti jsou stále častěji obětí deepfake. In: *Deník N* [online]. 7. 2. 2024 [cit. 2024-05-22]. Dostupné na: https://denikn.cz/1346164/falesne-snimky-taylor-swift-zaplavily-internet-zeny-a-deti-jsou-stale-casteji-obeti-deepfake/.

44 STARCEVIC, S. Italy's Giorgia Meloni called to testify in deepfake porn case. In: *POLITICO* [online]. 21. 3. 2024 [cit. 2024-05-22]. Dostupné na: https://www.politico.eu/article/italian-pm-giorgia-meloni-called-to-testify-in-deepfake-porn-case/.

45 BENÍŠKOVÁ, E. „Někdo hodil mou fotku do AI, aby mě svlékl." Tři české influencerky o své zkušenosti s novou formou sexuálního obtěžování. In: *HEY FOMO* [online]. 22. 10. 2023 [cit. 2024-05-06]. Dostupné na: https://heyfomo.cz/nekdo-hodil-mou-fotku-do-ai-aby-me-svlekl-tri-ceske-influencerky-o-sve-zkusenosti-s-novou-formou-sexualniho-obtezovani.

46 GUY, J. Outcry in Spain as artificial intelligence used to create fake naked images of underage girls. In: *CNN* [online]. 20. 9. 2023 [cit. 2024-05-22]. Dostupné na: https://www.cnn.com/2023/09/20/europe/spain-deepfake-images-investigation-scli-intl/index.html

47 CITRON – CHESNEY, *c. d.*, s. 1777.

48 SVOBODOVÁ, I. Jak snadno zmanipulovat volby. In: *Týdeník Respekt* [online]. 21. 10. 2023 [cit. 2024-05-22]. Dostupné na: https://www.respekt.cz/tydenik/2023/43/jak-snadno-zmanipulovat-volby.

49 Na Slovensku platilo moratorium, které médiím 48 hodin před dnem konání voleb znemožnilo o kandidujících subjektech informovat.

Kvůli popsaným rizikovým využitím zaznívají hlasy volající po důrazném omezení *deepfakes*.[50] Z druhé strany zaznívají námitky, že se jedná o neutrální technologii, jejíž riziko se odvíjí od kontextu, ve kterém se využívá.[51] *Deepfakes* lze využít i pro umělecké nebo edukativní účely, například *deepfake* malíře Salvátora Dalího vítá návštěvníky jeho muzea na Floridě.[52] *Deepfakes* vrací hlas těm, kteří jej ze zdravotních důvodů ztratili[53] a v Číně jsou populární *deepfakes* vyobrazující zesnulé osoby, jenž pomáhají vyrovnat se se ztrátou blízkých.[54] Využití *deepfakes* tak odpovídá Kranzbergovu prvnímu zákonu, dle kterého technologie není dobrá ani špatná, ale není neutrální, kdy stejná technologie vede v různých kontextech ke zcela odlišným výsledkům a zároveň mívá rozličné důsledky, které dalece přesahují bezprostřední účel technologie.[55]

PRÁVNÍ REGULACE *DEEPFAKES*

Mnohá riziková využití *deepfakes* představují potenciální hrozbu pro základní práva jednotlivců i pro zájmy společnosti jako celku. Z této hrozby vyplývá oprávněná potřeba reagovat na tohle riziko i prostřednictvím právní úpravy. Právo není jediným protiopatřením, které na riziko zneužití *deepfakes* reaguje. Soukromý i veřejný sektor se snaží dále vyvíjet a vylepšovat detektory *deepfakes*.[56] Další iniciativy pracují na vytvoření nástrojů zajišťujících autenticitu reálného audiovizuálního obsahu.[57] Pomoci má také zvýšení povědomí veřejnosti o *deepfakes*, například skrz výuku na školách[58] nebo školení zaměstnanců.[59]

Právní regulace *deepfakes* s sebou přináší možnosti, které žádné jiné protiopatření nenabízí. Za prvé, právo umožňuje vymezit, v jakých případech představuje rizikové využití *deepfakes* nezákonný obsah, čímž chrání jednotlivce i společnost před skutečně

---

50 VAN DER SLOOT – WAGENSVELD, *c. d.*, s. 13.

51 ŁABUZ, M. Regulating Deep Fakes in the Artificial Intelligence Act. *Applied Cybersecurity & Internet Governance*. 2023, Vol. 2, No. 1, s. 11.

52 AOUF, R. S. Museum creates deepfake Salvador Dalí to greet visitors. In: *Dezeen* [online]. 24. 5. 2019 [cit. 2024-05-20]. Dostupné na: https://www.dezeen.com/2019/05/24/salvador-dali-deepfake-dali-musuem-florida/.

53 O'BRIAN, M. Illness took away her voice: AI created a replica she carries in her phone. In: *AP News* [online] 13. 5. 2024 [cit. 2024-05-22]. Dostupné na: https://apnews.com/article/ai-recreating-lost-voice-illness-a6512c33481072c22182c116d2cbe419.

54 YANG, Z. Deepfakes of your dead loved ones are a booming Chinese business. In: *MIT Technology Review* [online]. 7. 5. 2024. [cit. 2024-10-13]. Dostupné na: https://www.technologyreview.com/2024/05/07/1092116/deepfakes-dead-chinese-business-grief/.

55 KRANZBERG, M. Technology and History: "Kranzberg's Laws". *Technology and Culture*. 1986, Vol. 27, No. 3, s. 544–545.

56 LAWSON, A. A Look at Global Deepfake Regulation Approaches. In: *Responsible AI* [online]. 24. 4. 2023 [cit. 2024-10-13]. Dostupné na: https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/.

57 FARID, H. Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety*. 2022, Vol. 1, No. 4, s. 22–23.

58 Doporučení pro využívání umělé inteligence na základních a středních školách. In: *Národní pedagogický institut České republiky* [online]. 2024 [cit. 2024-10-13]. Dostupné na: https://digitalizace.rvp.cz/files/ai-doporuceni-online-a4.pdf.

59 Stáhněte si manuál DEEPFAKE 2024: Obranná strategie pro české firmy. In: *Česká asociace umělé inteligence* [online]. 27. 12. 2023. [cit. 2024-05-31]. Dostupné na: https://asociace.ai/deepfake-2024/.

nebezpečnými *deepfakes* a zároveň poskytuje právní jistotu tvůrcům a šiřitelům legitimních *deepfakes*. Za druhé, právo nabízí obětem, jejichž práva byla dotčena nezákonnými *deepfakes*, nástroj k ochraně těchto práv a prostředek k dosažení zadostiučinění. A za třetí, prostřednictvím performativní regulace[60] lze motivovat relevantní aktéry k přijetí a podpoře dalších protiopatření reagujících na šíření nebezpečných *deepfakes*, skrz které lze dále čelit rizikům s ním spojeným.

LEGÁLNÍ DEFINICE *DEEPFAKES*

První zmínka o *deepfakes* v evropské legislativě se objevila v posíleném Kodexu zásad boje proti dezinformacím z roku 2022,[61] který *deepfakes* zmiňuje v souvislosti s manipulativním obsahem vytvořeným pomocí AI.[62] S konkrétní legální definicí přichází až AI akt, kde jsou *deepfakes* definovány následovně: *„Deep fake – obrazový, zvukový nebo video obsah vytvořený nebo manipulovaný umělou inteligencí, který se podobá existujícím osobám, objektům, místům subjektům či událostem a který by se dané osobě mohl nepravdivě jevit jako autentický nebo pravdivý.“*[63] Łabuz identifikuje čtyři aspekty této definice:[64]

1) Technologický: Musí se jednat o obsah vytvořený pomocí nástrojů umělé inteligence nebo strojového učení.
2) Typologický: Jde o audiovizuální obsah (tedy obraz, zvuk nebo video).
3) Subjektivní: Vyobrazuje existující osoby, objekty, místa či události.
4) Efektní: jejich vyobrazení se jeví jako autentické.

Díky této definici je možné odlišit *deepfakes* od syntetických médií. V obou případech se jedná o audiovizuální obsah vytvořený pomocí umělé inteligence, *deepfakes* se liší potenciálem oklamat své publikum, v němž dokáží vzbudit přesvědčení, že sleduje reálné a autentické vyobrazení osob, míst či událostí. Ačkoliv se *deepfakes* často spojují s dezinformacemi,[65] pro které je klíčový úmysl tvůrce nebo šiřitele dezinformace, definice *deepfakes* s úmyslem nepracuje a oproti dezinformacím klade důraz na běžné publikum a jeho vnímání daného obsahu. To je žádoucí, protože s potenciálem oklamat publikum se pojí největší nebezpečí *deepfakes*.[66] Přesto je potenciál klamat a efekt,

---

60 Performativní regulace spočívá v tom, že zákonodárce prostřednictvím právních norem ukládá povinnosti definičním autoritám, kdy konkrétní způsob dosažení těchto povinností je ponechán na definičních autoritách vzhledem k parametrům konkrétní sítě.

61 Kodex zásad boje proti dezinformacím je *soft-law* nástroj vytvořený Evropskou komisí, jehož cílem je snížit šíření dezinformací na online platformách. Tento kodex byl poprvé přijat v roce 2018 a aktualizován v roce 2022. Sdružuje signatáře, kterými jsou relevantní online platformy a obsahuje jednotlivá opatření, k jejichž dodržování se signatáři dobrovolně zavazují.

62 Závazek 15.1 Kodexu zásad boje proti dezinformacím z roku 2022.

63 Článek 3 odst. 60 AI aktu.

64 ŁABUZ, *c. d.*, s. 5–6.

65 Dezinformace se vyznačují právě úmyslem jejich šiřitele sdílet manipulativní obsah, viz např. definice, s níž pracuje Evropská unie: *„Dezinformacemi se rozumí falešný nebo zavádějící obsah, který je šířen s úmyslem klamat nebo zajistit hospodářský či politický prospěch a který může způsobit újmu veřejnosti.“*

66 BAILENSON, J. N. – HANCOCK, J. T. The Social Impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking*. 2021, Vol. 24, No. 3, s. 149.

který má konkrétní obsah na osoby, velmi subjektivní – klasifikace obsahu jako *deepfake* tak v praxi nebude jednoznačná.

Kromě AI aktu obsahuje regulaci zaměřenou na *deepfakes* i DSA, z něj vycházející *Pokyny pro poskytovatele velmi velkých online platforem a velmi velkých internetových vyhledávačů ke zmírňování systémových rizik pro volební procesy* (dále jen „Pokyny komise“) a směrnice o potírání násilí na ženách. DSA neobsahuje explicitní definici či zmínku o *deepfakes*. Poskytovatelům velmi velkých online platforem a vyhledávačů doporučuje (jako opatření ke zmírnění systémových rizik) zavést rozpoznatelná označení pro informace představující: *„Vytvořený nebo manipulovaný obrazový, zvukový nebo video obsah, který se znatelně podobá existujícím osobám, objektům, místům nebo jiným subjektům či událostem a který by se určité osobě mohl nepravdivě jevit jako autentický nebo pravdivý.“*[67] Popsané systémové riziko odpovídá definici *deepfakes* z AI aktu ve třech aspektech ze čtyř. Chybí technologický aspekt, tedy využití umělé inteligence. Uvedený příklad je tak oproti *deepfakes* širší a může se vztahovat i na *cheepfakes* a *shallowfakes*.

Pokyny komise vydala Evropská komise na základě čl. 35 odst. 3 DSA v dubnu 2024 v souvislosti s blížícími se volbami do Evropského parlamentu. Dokument doporučuje poskytovatelům velmi velkých online platforem a vyhledávačů zavést opatření reagující na riziko využití umělé inteligence pro vytváření klamavého, nepravdivého nebo manipulativního obsahu. *Deepfakes* jsou přímo zmíněny v bodu 40 písm. b) pokynů komise, který je definuje jako: *„Syntetické nebo manipulované obrázky, zvuky nebo videa, které se nápadně podobají existujícím osobám, předmětům, místům, subjektům, událostem nebo zobrazují události, které se nestaly, jako skutečné nebo je zkreslují a které by se určité osobě mohly nepravdivě jevit jako autentické nebo pravdivé.“*[68] Pokyny komise tak víc reflektují definici v AI aktu, opět ale chybí technologický aspekt, což ale kompenzuje zařazení mezi rizika související s využitím AI. Tato definice tak obsahuje všechny čtyři aspekty, které identifikoval Łabuz.

Směrnice o potírání násilí na ženách přímo zmiňuje *deepfakes* v souvislosti s výrobou a zpřístupněním nekonsensuální pornografie: *„K takové výrobě nebo manipulaci by se měla řadit i výroba deepfake, jestliže se materiál znatelně podobá existujícím osobám, předmětům, místům nebo jiným subjektům či událostem a znázorňuje sexuální praktiky jiné osoby a jiným osobám by se mohl nepravdivě jevit jako autentický nebo pravdivý.“*[69] Definice opět opomíjí technologický aspekt a neobsahuje odkaz na jakoukoliv jinou definici *deepfake*. Využití *deepfakes* zužuje pouze na falešný pornografický obsah.

Za pozornost stojí i fakt, že se jednotlivé předpisy neshodují v pravopisu slova *„deepfake“*, kdy AI akt a Pokyny komise se hovoří o *deep fakes* (s mezerou), směrnice o potírání násilí na ženách pracuje s pojmem *deepfakes* (bez mezery).[70] Rozdíl v psaní pojmu nemusí představovat větší problém, ačkoliv pro vytvoření právní regulace schopné zamezit rizikovému využití *deepfakes* by bylo žádoucí, aby se unijní definice

---

[67] Článek 35 odst. 1 písm. k) DSA.
[68] Bod 40 písm. b) Pokynů komise.
[69] Bod odůvodnění 19 směrnice o potírání násilí na ženách.
[70] ŁABUZ, *c. d.*, s. 9.

shodovaly, a ideálně, aby legální definice v AI aktu byla standardem, ze kterého bude další unijní a případná národní legislativa vycházet.[71]

STÁVAJÍCÍ LEGISLATIVA VERSUS *DEEPFAKES*

AI akt je prvním nařízením, které přímo reguluje *deepfakes*. Nepřímo na *deepfakes* dopadají i další unijní předpisy, jež přibližuje tato podkapitola. Prvním z nich je GDPR, protože jak při trénování GAN, tak při vytváření obsahu generátory *deepfakes* využívají fotografie a jiné obrazové či zvukové záznamy osob, které lze považovat za osobní údaje dle čl. 4 odst. 1 GDPR. Pokud *deepfake* zobrazuje konkrétní osobu, jedná se také o osobní údaj a je nutné mít k jeho zpracování zákonný důvod, kterým bude obvykle souhlas subjektu údajů[72] nebo oprávněný zájem příslušného správce.[73] Oprávněný zájem zde může existovat při využití *deepfakes* v rámci uměleckého nebo satirického díla, pokud přínos uměleckého díla převáží zásah do práv zobrazené osoby.[74] Souhlas osob, jejichž osobní údaje generátor zpracovává, i osob vyobrazených v konkrétních *deepfakes*, musí být informovaný, svobodně daný, konkrétní a jednoznačný. *Deepfakes* navíc mohou pracovat i s údaji, jako je rasa nebo etnicita, které je možné zpracovávat pouze na základě splnění některé z podmínek uvedené v čl. 9 odst. 2 GDPR.[75]

Při využití *deepfakes* pro legitimní účely, například v rámci marketingových kampaní, je třeba mít souhlas vyobrazených osob a zároveň respektovat zásady zpracování osobních údajů.[76] Většina rizikových využití *deepfakes* ale zákonný důvod pro zpracování osobních údajů mít nebude. Vyobrazené osoby se tak mohou domáhat práva na opravu či práva na výmaz. Problematická (a často nemožná) je ale samozřejmě identifikace a vyhledání osoby, která *deepfake* vytvořila, protože bude často anonymní a nedohledatelná.[77] GDPR a standardy ochrany osobních údajů tak mohou hrát roli hlavně při vývoji nástrojů generativní umělé inteligence (včetně generátorů *deepfakes* a syntetických médií) skrz princip *privacy by design*.

Dále na *deepfakes* dopadá DSA, a to prostřednictvím regulace systémových rizik, jež se vztahuje na velmi velké online platformy a vyhledávače.[78] Ty musí každoročně vyhodnotit systémová rizika a přijmout opatření k jejich zmírnění. K systémovým rizikům patří šíření nezákonného obsahu a obsahu, který by mohl mít nepříznivý vliv na některá základní práva uživatelů, na občanský diskurz, volební procesy či veřejnou bezpečnost.

---

71 Tamtéž, s. 27.

72 Článek 6 odst. 1 písm. a) GDPR.

73 Tamtéž, článek 6 odst. 1 písm. f).

74 ROMERO MORENO, F. Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers & Technology*. 2024, Vol. 38, No. 3, s. 297–326.

75 Tamtéž, s. 12–13.

76 Generative AI: The Data Protection Implications. In: *Confederation of European Data Protection Organizations (CEDPO)* [online]. 16. 10. 2023 [cit. 2024-10-13]. Dostupné na: https://cedpo.eu/wp-content/uploads/generative-ai-the-data-protection-implications-16-10-2023.pdf.

77 VAN HUIJSTEE, M. a kol. Tackling deepfakes in European policy. In: *Úřad pro publikace Evropské unie* [online]. 3. 2. 2022 [cit. 2024-10-13]. Dostupné na: https://data.europa.eu/doi/10.2861/325063.

78 Velmi velké platformy a vyhledávače jsou platformy a vyhledávače, které mají v zemích EU alespoň 45 milionů uživatelů a jež jako velmi velké platformy nebo vyhledávače určí evropská komise po konzultaci s členským státem nebo příslušným koordinátorem digitálních služeb (viz Článek 33 DSA, odst. 1 a 4).

Riziková využití *deepfakes* souvisí se všemi zmíněnými systémovými riziky, velmi velké platformy a vyhledávače je tak musí brát v potaz. Viditelné označování manipulativního audiovizuálního obsahu a zavedení funkce pro označování takového obsahu je dle DSA jedním z možných opatření pro zmírnění systémových rizik.[79] K podobnému kroku zavazuje své signatáře[80] i Kodex zásad boje proti dezinformacím z roku 2022.[81]

DSA je relevantní i v případech, kdy *deepfakes* šířené na hostingových platformách představují nezákonný obsah.[82] Poskytovatelé hostingových služeb mají povinnost zavést mechanismus umožňující nahlašování nezákonného obsahu a existující oznámení včasně, objektivně a s náležitou péčí vyřídit.[83] *Deepfakes* jsou nezákonným obsahem tehdy, když vyobrazují explicitně protiprávní obsah, typicky dětskou pornografii nebo obsah podporující terorismus, nebo v případech, kdy zasahují do autorskoprávní ochrany či osobnostních práv. V těchto případech bude klasifikace obtížná: u *deepfake* vyobrazující osoby bude třeba posoudit, zda se jedná o (1) nekonsenzuální vyobrazení, které (2) nepokrývá žádná ze zákonných licencí a (3) představuje zásah do osobnostních práv. Podobně u *deepfakes* využívajících autorské dílo je třeba zhodnotit, zda nejde o využití na základě některé ze zákonných licencí.[84]

TRANSPARENTNOST JAKO LÉK NA VŠE

AI akt rozlišuje systémy umělé inteligence podle rizika, které mohou představovat, a na jednotlivé kategorie se vztahuje odpovídající regulace. Těmito kategoriemi jsou (1) zakázané postupy v oblasti AI, jejichž uvádění na trh je zakázáno,[85] (2) vysoce rizikové systémy,[86] na které se vztahují speciální požadavky, (3) obecné modely AI.[87] Generátory *deepfake* jsou zařazeny mezi obecné systémy AI, tvoří ale speciální podkategorii, na niž se vztahuje požadavek na transparentnost obsažený v čl. 50 odst. 2 a 4 AI Aktu. Článek 50 odst. 2 se týká především vývoje a nasazení AI systémů a dopadá na poskytovatele AI,[88] kteří musí zajistit, že syntetický textový i audiovizuální obsah bude strojově označený a detekovatelný jako uměle vytvořený nebo jinak manipulovaný. Článek 50 odst. 4 dopadá na subjekty zavádějící AI[89] a vztahuje se pouze na *deepfakes*. Subjekty zavádějící AI musí při šíření *deepfakes* zároveň zveřejnit upozornění, že

---

79 Článek 35 odst. 1 písm. k) DSA.

80 Mezi signatáře tohoto závazku patří následující společnosti a platformy: Adobe, Google, TikTok, Meta, Microsoft a Seznam.

81 Závazek 15.1 Kodexu zásad boje proti dezinformacím z roku 2022.

82 Definice nezákonného obsahu viz článek 3 písm. h) DSA.

83 Článek 16 DSA.

84 VAN HUIJSTEE a kol., *c. d.*, s. 40.

85 Článek 5 AI aktu.

86 Tamtéž, článek 6.

87 Tamtéž, článek 53 a následující.

88 Poskytovatelé jsou fyzické nebo právnické osoby, veřejné orgány, agentury nebo jiné subjekty, které vyvíjí systém AI či obecný model AI nebo nechávají vyvíjet systém AI či obecný model AI a uvádějí je na trh nebo které uvádějí systém AI do provozu pod svým vlastním jménem, názvem nebo ochrannou známkou, ať už za úplatu, nebo zdarma (článek 3 odst. 3 AI Aktu).

89 Zavádějícím subjektem je fyzická nebo právnická osoba, veřejný orgán, agentura nebo jiný subjekt, který v rámci své pravomoci využívá systém AI, s výjimkou případů, kdy je systém AI využíván při osobní neprofesionální činnosti (článek 3 odst. 4 AI Aktu).

byl obsah uměle vytvořen nebo s ním bylo manipulováno. Tato povinnost se nevztahuje na ty *deepfakes*, jejichž použití zákon povoluje při předcházení, odhalování, vyšetřování a stíhání trestných činů. Dále existuje výjimka pro *deepfakes*, které jsou součástí uměleckého, tvůrčího, satirického, fiktivního či obdobného díla – stačí je označit způsobem, který nebrání zobrazení nebo užívání díla.

Vzhledem k rizikům popsaným v první části článku se zařazení *deepfakes* mezi obecné modely AI může zdát překvapivé, a tento krok se setkal s kritikou. Některá riziková využití *deepfakes* mohou naplňovat kritéria pro zařazení do kategorie zakázaných postupů v oblasti AI – například použití *deepfake* k vydírání osob, které zasahuje do autonomie jedince a nutí jej učinit rozhodnutí, která by jinak neučinil, nebo vede k závažným psychickým následkům.[90] Nejčastěji se kritické hlasy přiklání k zařazení *deepfakes* mezi vysoce rizikové systémy, a to kvůli široké škále rizikových využití. Dříve zmíněný příklad slovenských prezidentských voleb ukazuje, že se již nyní *deepfakes* využívají k ovlivnění veřejného mínění a volebních procesů. AI akt přitom systémy, které mohou ovlivnit volby, řadí mezi vysoce rizikové systémy.[91] Dalším argumentem pro jejich zařazení do přísnější kategorie jsou riziková využití představující nebezpečí pro jednotlivce a jeho základní práva, jako je *deep porn* a další manipulativní vyobrazení poškozující jedince, jeho pověst či důstojnost.[92] Nezařazení mezi vysoce rizikové systémy také opomíjí nebezpečí spočívající v přesvědčivosti a dlouhodobém působení na publikum, kdy vědomé i nevědomé vystavení *deepfakes* posiluje existující předsudky či falešné představy a prohlubuje společenské rozdíly.[93] Jako žádoucí se jeví zařazení konkrétních, často zmiňovaných rizikových využití *deepfakes* do kategorie vysoce rizikových systémů či jejich zákaz nebo klasifikace jako nezákonný obsah.[94]

V odůvodnění AI aktu zákonodárce přiznává, že některé systémy představují specifické riziko kvůli hrozícímu zneužití za účelem vydávání se za jiné osoby nebo k podvodům, a mohou prohlubovat nedůvěru v informační ekosystém jako celek.[95] Riziko dále zvyšuje rostoucí uživatelská dostupnost generativní AI a stálé zdokonalování jejich výstupů. Zákonodárce zde ale zvolil odlišný přístup, a to transparentnost. Cílem úpravy v AI aktu je podporovat zavádění důvěryhodné AI.[96] Úprava zde navazuje na Etické pokyny pro zajištění důvěryhodnosti UI[97] vytvořené expertní skupinou na AI, dle kterých je transparentnost stěžejní pro zajištění důvěryhodné AI a součástí požadavku na transparentnost je i informování uživatelů, že jednají s AI systémem.[98] Odpovědí na

---

90 ROMERO MORENO, *c. d*., s. 7.
91 Příloha III, odst. 8 písm. b) AI aktu.
92 ROMERO MORENO, *c. d.*, s. 8–9.
93 MESARČÍK, M. a kol. Stance on The Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence: Artificial Intelligence Act. In: *ResearchGate* [online]. 2022 [cit. 2024-10-20]. Dostupné na: https://www.researchgate.net/publication/359116983_Stance_on_The_Proposal_for_a_Regulation_Laying_Down_Harmonised_Rules_on_Artificial_Intelligence_-_Artificial_Intelligence_Act.
94 ŁABUZ, *c. d.*, s. 15.
95 Bod odůvodnění 132 AI aktu.
96 Tamtéž, článek 1.
97 SMUHA, N. a kol. Etické pokyny pro zajištění důvěryhodnosti UI. In: *Úřad pro publikace Evropské unie* [online]. 2019 [cit. 2024-10-20]. Dostupné na: https://op.europa.eu/cs/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.
98 Tamtéž, článek 78.

riziková využití *deepfakes* tak má být zavedení nástrojů garantujících autenticitu obsahu ruku v ruce s viditelným označováním syntetických médií a *deepfakes*.[99]

AI akt je platný od 1. 8. 2024 a účinný od 1. 8. 2026. Komise se snaží povinnost označovat *deepfakes* a syntetická média uspíšit a zároveň rozšířit, protože AI akt dopadá pouze na poskytovatele a subjekty zavádějící AI, kdy do této kategorie nepatří osoby používající AI při běžné, neprofesionální činnosti – tedy běžní uživatelé internetových platforem. K rozšíření povinnosti označovat *deepfakes* i na ně komise využívá již účinné DSA a související Pokyny komise a Kodex boje proti dezinformacím z roku 2022. Všechny tři tyto dokumenty volají po zavedení povinnosti označovat *deepfakes* (i syntetická média) a vytvoření uživatelsky přívětivých funkcí k jejich označení. Dále obsahují doporučení označovat i syntetický reklamní obsah,[100] upravit proces moderace obsahu na sítích tak, aby při něm bylo možné odhalit syntetická média a *deepfakes*, a podporovat inovace s cílem zajistit autenticitu audiovizuálního obsahu.[101] Adresáty těchto doporučení jsou ale pouze velmi velké online platformy a vyhledávače a některé už tato doporučení vyslyšely. Meta vyžaduje, aby uživatelé označovali obsah vytvořený a upravený pomocí AI,[102] TikTok doporučuje označovat syntetická média a vyžaduje označování *deepfakes*,[103] označit zásadní úpravy provedené pomocí AI musí i tvůrci na YouTube.[104] Povinnost označit obsah generovaný pomocí AI zde dopadá na všechny uživatele, bez ohledu na to, k jakému účelu síť využívají.

Mimo velmi velké online platformy a vyhledávače je situace složitější. Po nabytí účinnosti AI aktu budou neoznačené *deepfakes* představovat nezákonný obsah, který musí provozovatelé hostingových platforem odstranit, pokud se o něm dozví.[105] Povinnost označovat *deepfakes* ale dopadá pouze na subjekty zavádějící AI, mezi které nepatří osoby využívající AI systémy při osobní, neprofesionální činnosti. Při zhodnocení, zda je neoznačený *deepfake* nezákonným obsahem, musí provozovatelé hostingových platforem zhodnotit následující otázky:

1. Jedná se o *deepfake*?

Obsah musí odpovídat definici *deepfake* dle AI aktu, kdy obtížné je posouzení technologického a efektního aspektu. Technologický aspekt, tedy zda jde o obsah vytvořený pomocí AI, je obtížné posoudit kvůli neexistenci spolehlivého a univerzálního detektoru. Hrozí zde riziko falešně pozitivních i falešně negativních výsledků. Zhodnocení efektního aspektu je problematické proto, že obsah může mít na každou osobu rozdílný efekt a objektivně změřit potenciál klamat nejde.

---

[99] Body odůvodnění 132, 133 a 134 AI Aktu.
[100] Tamtéž, článek 40 písm. c).
[101] Tamtéž, článek 40 písm. d).
[102] Our Approach to Labeling AI-Generated Content and Manipulated Media. In: *Meta* [online]. 5. 4. 2024 [cit. 2024-08-02]. Dostupné na: https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/.
[103] About AI-generated content. In: *TikTok Help Center* [online]. 2024 [cit. 2024-10-06]. Dostupné na: https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content.
[104] Disclosing use of altered or synthetic content. In: *YouTube Help* [online]. 2024 [cit. 2024-10-06]. Dostupné na: https://support.google.com/youtube/answer/14328491#examples_as.
[105] Článek 6 odst. 1 písm. b) DSA.

2. Jde o *deepfake*, na který dopadá povinnost označit obsah dle čl. 50 odst. 4 AI aktu?

K posuzování druhé otázky je třeba přistoupit poté, co je možné obsah klasifikovat jako *deepfake*. Provozovatelé hostingových platforem musí zhodnotit, zda *deepfake* zveřejnil subjekt zavádějící AI, tedy jestli jde o *deepfake* zveřejněný v rámci osobní, neprofesionální činnosti či nikoliv. Podobně jako u efektního aspektu *deepfakes* zde neexistuje jasná hranice, která by rozdílná využití odlišila. Případně se zde nabízí otázka, zda nejde o obsah, na který by dopadala výjimka pro umělecká a tvůrčí využití *deepfakes*.

Pro provozovatele tak může být zajišťování souladu s požadavky kladenými AI aktem velmi obtížné. Řešením může být (po vzoru velmi velkých online platforem) po uživatelích vyžadovat označení veškerého audiovizuálního obsahu generovaného pomocí AI, bez ohledu na to, za jakým účelem jej uživatelé šíří a zda má obsah potenciál klamat své publikum. Provozovatelé se zpřísněním podmínek mohou vyhnout zodpovědnosti za případná neodstranění neoznačených *deepfakes*. Problematická bude i nadále neexistence spolehlivého detektoru *deepfakes*, protože narozdíl od velmi velkých online platforem a vyhledávačů menší platformy nemusí mít dostatečné kapacity k zavádění uživatelsky dostupných funkcí pro zajišťování autenticity online obsahu. Problémem může být také neochota povinnost označovat *deepfakes* vymáhat. Například zmíněný *deepfake* předsedy Progresivného Slovenska Martina Šimečky se začal šířit na platformě Telegram,[106] která je známá nízkou ochotou moderovat obsah a nedostatečnou kapacitou reagovat na stížnosti uživatelů, což z něj dělá ideální síť pro šíření nenávistného obsahu nebo dezinformací.[107] Ačkoliv nyní Telegram reaguje na šíření *deep porn* v Jižní Koreji zpřísněním uživatelských podmínek s cílem zamezit šíření *deep porn* a přijímá opatření k efektivní detekci *deepfakes* a nezákonného obsahu,[108] skepse ohledně případného zamezení šíření dalších manipulativních *deepfakes* je namístě.

OZNAČOVÁNÍ *DEEPFAKES* NEZAMEZÍ JEJICH RIZIKOVÝM VYUŽITÍM ANI NEGATIVNÍM NÁSLEDKŮM

Ačkoliv se z nadpisu podkapitoly může zdát, že označování *deepfakes* žádný pozitivní efekt nepřinese, realita není tak negativní. Z experimentů zaměřených na schopnost publika rozpoznat *deepfakes* vyplývá, že vědomí o jejich existenci zvyšuje šance, že uživatelé *deepfake* odhalí.[109] Také upozornění na fakt, že daný obsah je *deepfake* politickou parodií vedlo k tomu, že jí respondenti dokázali lépe porozumět.[110]

---

[106] SVOBODOVÁ, *c. d.*

[107] TKÁČOVÁ, N. Dezinformace na Telegramu v České republice: organizační a finanční pozadí. In: *Prague Security Studies Institute* [online]. 23. 1. 2024 [cit. 2024-10-06]. Dostupné na: https://www.pssi.cz/download//docs/10993_dezinformace-na-telegramu-v-ceske-republice-organizacni-a-financni-pozadi.pdf.

[108] NG, K. Telegram apologises for handling of deepfake porn material. In: *BBC* [online]. 4. 9. 2024 [cit. 2024-10-20]. Dostupné na: https://www.bbc.com/news/articles/cvg45kz47dno.

[109] APPEL, M. – PRIETZEL, F. The detection of political deepfakes. *Journal of Computer-Mediated Communication*. 2022, Vol. 27, No. 4, s. 9.

[110] HANG, L. – SHUPEI, J. "I know it's a deepfake": The role of AI disclaimers and comprehension in the processing of deepfake parodies. *Journal of Communication*. 2024, Vol. 74, No. 5, s. 11.

Rozšíření požadavku na transparentnost *deepfakes* tak může kultivovat online prostředí, zvýšit povědomí publika o existenci *deepfakes* a jeho schopnosti *deepfakes* odhalit. Dle dat ze srpna 2023 umělá inteligence denně vytvoří 38 milionů obrázků (a lze předpokládat, že číslo dále poroste),[111] zvýšení schopností uživatelů rozpoznat obsah generovaný umělou inteligencí je tak v každém případě žádoucí. Regulace zároveň klade důraz na to, aby nestavěla překážky vývoji v oblasti generativní umělé inteligence, a naopak se snaží relevantní subjekty motivovat ke spolupráci na vývoji nových technických řešení.[112]

Nelze ale očekávat, že transparentnost zabrání rizikovým využitím *deepfakes*. Představa, že by útočníci při vishingových útocích své potenciální oběti upozornili, že se jedná o *deepfake*, je absurdní. Státní i nestátní subjekty využívající *deepfakes* v rámci trestné činnosti, pro vytváření propagandy nebo k jiným nezákonným či rizikovým činnostem své *deepfakes* označovat nebudou.[113] Aktéři zneužívající *deepfakes* navíc mohou mít přístup ke generátorům vytvářejícím sofistikované *deepfakes*, což dále komplikuje už tak obtížnou detekci.

Označování *deepfake* obsahu také nemusí zabránit případným negativním účinkům. Dle stávajících výzkumů sice mohou upozornění na problematický obsah omezit jeho další šíření,[114] i zřetelně označené *deepfakes* ale vyvolávají negativní emoce. Lze je tak stále zneužít k šíření manipulativního obsahu nebo k poškození něčí pověsti. Také osoby nekonsenzuálně vyobrazené v *deep porn* transparentní upozornění na fakt, že jde o *deepfake*, neochrání před psychickou újmou nebo poškozením pověsti.[115]

TRESTNĚPRÁVNÍ POSTIH *DEEPFAKES*

Je proto žádoucí, že se EU nespoléhá pouze na transparentní označování, ale iniciuje kriminalizaci *deepfake* pornografie v rámci směrnice o potírání násilí na ženách.[116] Na iniciativu EU již reagoval i český zákonodárce, kdy Ministerstvo spravedlnosti v březnu 2024 předložilo návrh nového trestného činu zneužití identity k výrobě pornografie a její šíření. Návrh je v současné době v mezirezortním připomínkovém řízení a dle důvodové zprávy by měl reagovat především na vytváření *deepfake* pornografie.[117]

---

[111] SUKHANOVA, K. AI Image Generator Market Statistics – What Will 2024 Bring? In: *The Tech Report* [online]. 19. 1. 2024 [cit. 2024-07-01]. Dostupné na: https://techreport.com/statistics/software-web/ai-image-generator-market-statistics/.

[112] Článek 40, bod 2, písm. ii) Pokynů komise.

[113] ŁABUZ, *c. d.*, s. 19.

[114] MARTEL, C. – RAND, D. G. Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in Psychology*. 2023, Vol. 54, s. 4.

[115] TOPARLAK, R. T. The false promise of transparent deep fakes. In: *Hertie School Centre for Digital Governance* [online]. 9. 11. 2022 [cit. 2024-07-02]. Dostupné na: https://www.hertie-school.org/en/digital-governance/research/blog/detail/content/the-false-promise-of-transparent-deep-fakes-how-transparency-obligations-in-the-draft-ai-act-fail-to-deal-with-the-threat-of-disinformation-and-image-based-sexual-abuse.

[116] Článek 5 směrnice o potírání násilí na ženách.

[117] Zákon, kterým se mění zákon č. 40/2009 Sb., trestní zákoník, ve znění pozdějších předpisů, zákon č. 141/1961 Sb., o trestním řízení soudním (trestní řád), ve znění pozdějších předpisů, a další související

I jiná riziková využití *deepfakes*, například v úvodu zmíněné podvodné investice, jsou už nyní trestnými činy. *Deepfakes* tak lze vnímat jako další technologii, jíž lze využít v rámci *cyber-enabled*[118] kyberkriminality, a trestní právo jako další nástroj, s jehož pomocí lze reagovat na riziková využití *deepfakes*, a to buď prostřednictvím existujících skutkových podstat, nebo vytvořením nových. Například u často zmiňovaného šíření *deepfakes* s cílem ovlivnit volby se lze inspirovat v Kalifornii nebo Texasu, kde je vytváření a šíření *deepfakes* vyobrazujících kandidáty šedesát (resp. třicet) dní před volbami trestné.[119]

Orgány činné v trestním řízení musí počítat jak s problémy spojenými s běžnou kyberkriminalitou (jako je obtížná identifikace pachatele), tak se specifiky *deepfakes*, kam se opět řadí jejich problematická detekce. Použití *deepfake* při páchání trestné činnosti zvyšuje nároky na personální i technologické kapacity orgánů činných v trestním řízení. Dle Europolu je třeba nejen zvýšit kvalifikaci personálu či investovat do technického vybavení, ale také zavést postupy pro zajištění audiovizuálních důkazů za účelem zajištění jejich autenticity před soudy.[120]

Navíc detektory *deepfakes* využívané orgány činnými v trestním řízení AI akt řadí mezi vysoce rizikové systémy AI.[121] Jedná se o systémy, které jsou využívány donucovacími orgány nebo jejich jménem pro hodnocení spolehlivosti důkazů v průběhu vyšetřování nebo stíhání trestných činů – a bez softwaru, který dokáže posoudit, že se jedná o *deepfake*, si lze stíhání takových trestných činů obtížně představit. Pozitivem je, že zařazení do přísnější kategorie může zajistit integritu a důvěryhodnost používaných detektorů.

ZÁVĚR

Evropská regulace digitálního prostoru reaguje na šíření *deepfakes* zejména úpravou směřující na detektory *deepfakes* (GDPR a AI akt), šíření *deepfakes* na platformách (DSA a s ním související předpisy), a především označováním syntetického obsahu a *deepfakes* (AI akt). Je žádoucí, že AI akt nastavuje pravidla pro transparentní označování *deepfakes*, které může vést ke kultivaci online prostředí a zvýšení schopnosti uživatelů *deepfakes* rozpoznat. Jednoznačným pozitivem je fakt, že část provozovatelů velmi velkých online platforem začala vyžadovat označování syntetických médií a *deepfakes* již před nabytím účinnosti AI aktu. Důležité ale je, aby provozovatelé tuto povinnost efektivně vymáhali a zároveň šíření *deepfakes* dále reflektovali v rámci vyhodnocování systémových rizik.

---

zákony. In: *Portál informačního systému ODok Úřadu vlády České republiky* [online]. 26. 3. 2024 [cit. 2024-07-02]. Dostupné na: https://www.odok.cz/portal/veklep/material/KORND3QJZZZ3/.

[118] *Cyber-enabled* neboli kyberneticky umožněná kriminalita označuje tradiční trestné činy, jejichž možnosti rozšiřuje využití technologií.

[119] DELFINO, R. Deepfakes on Trial: A Call to Expand the Trial Judge'S Gatekeeping Role to Protect Legal Proceedings from Technological Fakery. *Hastings Law Journal*. 2022, Vol. 74, No. 2, s. 304.

[120] Facing reality?: Law enforcement and the challenge of deepfakes: An observatory report from the Europol innovation lab. In: *European Union: Publications Office of the European Union* [online]. 4. 3. 2024 [cit. 2024-10-06]. Dostupné na: https://data.europa.eu/doi/10.2813/158794.

[121] Příloha III, bod 6 písm. c) AI aktu.

Největší slabinu AI aktu vidím v problematickém určení, kdy se jedná o *deepfake*. Nařízení sice přichází s definicí *deepfake*, v praxi ale bude určení, kdy se jedná o *deepfake*, problematické. Absence spolehlivého detektoru znamená, že nelze odhalit obsah vytvořený pomocí AI (technologický aspekt), vyhodnocení potenciálů *deepfake* klamat je velmi subjektivní a vztažení povinnosti označovat *deepfakes* pouze na subjekty zavádějící AI znamená, že je navíc třeba zjišťovat, zda se *deepfake* šíří v rámci profesionální činnosti či nikoliv. Tyto problémy budou zatěžovat především menší hostingové platformy, jejichž kapacity pro moderaci obsahu a detekci *deepfakes* či syntetických médií jsou nižší.

Regulace *deepfakes* v AI aktu a dalších předpisech také nezabrání rizikovým využitím *deepfakes* a s odstraněním negativních důsledků jejich šíření pomůže jen částečně. Aktéři zneužívající *deepfakes* mohou stále šířit neoznačené *deepfakes*, a navíc využívat pokročilejší generátory, které mohou oklamat detektory. Dalším nástrojem dopadajícím na nezákonná šíření *deepfakes* je tak trestní právo. *Deepfakes* lze stíhat pomocí existujících skutkových podstat (jako podvodné investice zmíněné v úvodu) nebo nových (jako *deep porn*). K tomu, aby bylo trestní právo efektivní, je nutné navýšit kapacity orgánů činných v trestním řízení, aby byly schopné trestné činy využívající *deepfakes* vyšetřovat.

Zároveň je třeba počítat s faktem, že se v blízké budoucnosti budeme setkávat s neodhalitelnými *deepfakes*. Proto je třeba kromě kapacit k označování a odhalování *deepfakes* zavádět i nástroje, jež umožní spolehlivě ověřit autenticitu reálného obsahu.

Mgr. et Mgr. Tina Mizerová
Právnická fakulta Masarykovy univerzity
471315@mail.muni.cz
ORCID: 0009-0000-6003-9562