Justyna Wilczyńska-Baraniak[1]

Małgorzata Wilczyńska[2]

Państwowa Akademia Nauk Stosowanych im. ks. Bronisława Markiewicza w Jarosławiu

https://doi.org/10.29316/9788368103359_12

DEEPFAKE JAKO STRUKTURALNE ZAGROŻENIE DLA BEZPIECZEŃSTWA INFORMACJI

DEEPFAKES AS A STRUCTURAL THREAT TO INFORMATION SECURITY

Streszczenie

Rozdział analizuje technologię deepfake jako strukturalne zagrożenie dla bezpieczeństwa informacji, koncentrując się na konsekwencjach jej demokratyzacji. Punktem wyjścia jest obserwacja, że rozwój narzędzi open source, usług chmurowych oraz intuicyjnych interfejsów doprowadził do sytuacji, w której tworzenie deepfake-'ów nie wymaga dziś specjalistycznej wiedzy technicznej. Obniżenie progu wejścia umożliwia masową produkcję syntetycznych treści audiowizualnych, co znacząco zwiększa skalę nadużyć, w tym oszustw finansowych, manipulacji politycznej oraz naruszeń dóbr osobistych. W części empirycznej przedstawiono liczne przypadki podszywania się w czasie rzeczywistym, m.in. w Hongkongu, Wielkiej Brytanii,

[1] https://orcid.org/0009-0005-3148-8197, e-mail: justaw@o2.pl

[2] https://orcid.org/0000-0002-1969-3498, e-mail: malgorzata.wilczynska@pansjar.edu.pl

Singapurze i państwach UE. Przykłady te pokazują, że tradycyjne metody weryfikacji tożsamości oparte na obrazie i głosie tracą swoją wiarygodność, co prowadzi do poważnych strat finansowych i podważa zaufanie do komunikacji cyfrowej. Analiza wskazuje, że deepfake staje się narzędziem umożliwiającym obchodzenie systemów uwierzytelniania oraz realizację oszustw na masową skalę. W części regulacyjnej omówiono odpowiedź Unii Europejskiej, w szczególności AI Act, który wprowadza legalną definicję deepfake oraz obowiązek transparentności. W rozdziale podkreślono, że obowiązujące ramy prawne, w tym RODO, nie nadążają za dynamiką rozwoju technologii i nie zapewniają ochrony przed systemowymi skutkami dezinformacji. Wnioski wskazują na konieczność łączenia regulacji prawnych, rozwiązań technicznych oraz edukacji cyfrowej jako warunku skutecznego przeciwdziałania zagrożeniom wynikającym z demokratyzacji technologii deepfake. Celem rozdziału jest analiza mechanizmów obniżenia progu wejścia do tworzenia deepfake'ów oraz omówienie konsekwencji tego procesu dla bezpieczeństwa informacji, zaufania do mediów i skuteczności istniejących ram prawnych. Analiza demokratyzacji technologii deepfake pozwala lepiej zrozumieć, dlaczego współczesne zagrożenia informacyjne mają charakter masowy, trudny do kontrolowania i wykraczający poza tradycyjne ramy regulacyjne. Treści prezentowane w rozdziale opracowano z wykorzystaniem teoretycznych metod badawczych, wiodącą z nich była analiza krytyczna literatury przedmiotu, analiza kwerendy źródeł internetowych, jak również metoda analizy, syntezy, uogólnienia oraz wnioskowania

Słowa kluczowe: deepfake, bezpieczeństwo informacji, demokratyzacja technologii, treści audiowizualne, oszustwa finansowe, podszywanie się (impersonation), weryfikacja tożsamości, dezinformacja, zaufanie do komunikacji cyfrowej, regulacja sztucznej inteligencji (AI Act)

Abstract

This chapter examines deepfake technology as a structural threat to information security, with particular emphasis on the consequences of its rapid democratization. The analysis begins with the observation that the widespread availability of opensource tools, cloudbased services, and intuitive user interfaces has fundamentally lowered the technical threshold for producing synthetic audiovisual content. As the chapter notes, "creating deepfakes no longer requires specialized technical knowledge", which has enabled largescale misuse by individuals lacking advanced expertise. This shift has transformed deepfakes from isolated incidents into a systemic risk affecting both public and private institutions. The empirical section presents documented cases of realtime impersonation in Hong Kong, the United Kingdom, Singapore, and

EU member states. These incidents demonstrate that traditional identityverification methods based on facial and voice recognition “are losing their reliability”, resulting in substantial financial losses and undermining trust in digital communication. The cases illustrate how synthetic audio and video can be weaponized to bypass authentication mechanisms and facilitate fraud at scale. The regulatory analysis focuses on the European Union’s AI Act, which introduces a legal definition of deepfake and establishes transparency obligations for synthetic media. The chapter argues that existing legal frameworks, including the GDPR, do not adequately address the authenticity of digital content and fail to mitigate systemic risks such as disinformation and the liar’s dividend. The conclusion highlights the need for a combined strategy integrating legal regulation, technical safeguards, and AI literacy as essential components of an effective response to the challenges posed by the democratization of deepfake technology.

Keywords: deepfake, information security, democratization of technology, audiovisual content, financial fraud, impersonation, identity verification, disinformation, trust in digital communication, AI regulation (AI Act)

Wstęp

Deepfake to technologia oparta na metodach sztucznej inteligencji, wykorzystująca zaawansowane modele generatywne, w szczególności generatywne sieci przeciwstawne (GAN) oraz autoenkodery wariacyjne (VAE), do syntezy realistycznych treści audiowizualnych, takich jak obrazy, nagrania wideo oraz dźwięk. Termin *deepfake* wywodzi się z języka angielskiego i stanowi połączenie słów *deep*, odnoszącego się do technik głębokiego uczenia maszynowego (*deep learning*), oraz *fake*, oznaczającego treści fałszywe lub celowo zmanipulowane (NASK, 2025). Termin *deepfake* pojawił się w literaturze i dyskursie publicznym pod koniec 2017 roku, początkowo jako określenie nowej klasy syntetycznych treści audiowizualnych generowanych przy użyciu metod uczenia maszynowego. Geneza terminu *deepfake* wiąże się z działalnością anonimowego użytkownika platformy Reddit, posługującego się pseudonimem „deepfakes”, który właśnie pod koniec 2017 roku zapoczątkował społeczność internetową poświęconą tworzeniu i udostępnianiu syntetycznych materiałów audiowizualnych (Regan, 2025). To od jego nazwy użytkownika wywodzi się określenie, które z czasem zostało zaadaptowane do opisu całej klasy technologii generatywnych.

W niniejszym rozdziale szczególną uwagę poświęcono zagadnieniu demokratyzacji technologii deepfake jako jednemu z kluczowych wyzwań dla bezpieczeństwa informacji. Wybór tego obszaru analizy wynika z faktu, że powszechna dostępność narzędzi do generowania deepfake'ów, w tym rozwiązań open source oraz usług chmurowych, prowadzi do jakościowej zmiany skali i charakteru potencjalnych nadużyć. Demokratyzacja ta, choć sprzyja innowacjom, twórczości oraz zastosowaniom edukacyjnym, jednocześnie umożliwia masową produkcję treści manipulacyjnych przez osoby nieposiadające specjalistycznej wiedzy technicznej. Celem rozdziału jest zatem analiza mechanizmów obniżenia progu wejścia do tworzenia deepfake'ów oraz omówienie konsekwencji tego procesu dla bezpieczeństwa informacji, zaufania do mediów i skuteczności istniejących ram prawnych. Analiza demokratyzacji technologii deepfake pozwala lepiej zrozumieć, dlaczego współczesne zagrożenia informacyjne mają charakter masowy, trudny do kontrolowania i wykraczający poza tradycyjne ramy regulacyjne. Rozdział opiera się na analizie literatury naukowej, raportów branżowych oraz udokumentowanych przypadków nadużyć z lat 2024-2025. Dobór case studies został dokonany według dwóch kryteriów: (1) potwierdzona przez organy ścigania lub media wysokiej wiarygodności autentyczność zdarzenia oraz (2) reprezentatywność dla zjawiska demokratyzacji technologii deepfake, rozumianej jako obniżenie progu wejścia do generowania syntetycznych treści. Analiza obejmuje zarówno aspekty techniczne, jak i regulacyjne, co pozwala na uchwycenie strukturalnego charakteru zagrożeń dla bezpieczeństwa informacji. Treści prezentowane w rozdziale opracowano z wykorzystaniem teoretycznych metod badawczych, wiodącą z nich była analiza krytyczna literatury przedmiotu, analiza kwerendy źródeł internetowych, jak również metoda analizy, syntezy, uogólnienia oraz wnioskowania.

Geneza i rozwój technologii deepfake

W początkowym okresie deepfake polegał głównie na manipulacji wizerunkiem poprzez zastępowanie twarzy osób występujących w istniejących nagraniach wizerunkami osób publicznych. Mimo kontrowersyjnego charakteru tych praktyk, działalność ta przyczyniła się do gwałtownego wzrostu zainteresowania technologiami syntezy obrazu i wideo oraz zapoczątkowała szeroką debatę naukową, prawną i etyczną nad konsekwencjami rozwoju technologii deepfake (Etienne, 2021). Według raportu Deeptrace

technologia rozwija się bardzo szybko (Ajder, Patrini, Cavalli, Cullen, 2019). Na początku 2019 roku w Internecie zaprezentowano 7964 filmów deepfake, dziesięć miesięcy później ich liczba wynosiła 14678. (Toews, 2022). Szerokie zainteresowanie środowiska naukowego i opinii publicznej technologią deepfake zostało zapoczątkowane przez dwa przełomowe projekty badawcze: „*Face2Face"* (Thies i in., 2016). oraz „*Synthesizing Obama"* (Suwajanakorn, Seitz, KemelmacherShlizerman, 2017), które po raz pierwszy w sposób systematyczny zademonstrowały możliwość realistycznej syntezy mimiki i synchronizacji mowy, ujawniając potencjalne konsekwencje tej technologii dla wiarygodności treści audiowizualnych. Projekt Face2Face autorstwa zaprezentował możliwość interaktywnego przenoszenia mimiki twarzy z osoby źródłowej na osobę widoczną w nagraniu docelowym. System analizował ruchy twarzy mówiącego takie jak uśmiech, zmarszczenie brwi czy ruch ust, a następnie odwzorowywał je na twarzy osoby w nagraniu, generując wideo wyglądające jak autentyczne. Co istotne, technologia działała w czasie zbliżonym do rzeczywistego, co unaoczniło, że manipulacja obrazem wideo może być nie tylko realistyczna, lecz także szybka i trudna do wykrycia. Projekt ten stał się jednym z pierwszych dowodów na to, że możliwe jest tworzenie przekonujących fałszywych reakcji emocjonalnych lub wypowiedzi osób publicznych. Projekt *Synthesizing Obama* pokazał, że możliwe jest syntetyczne generowanie całych wypowiedzi osób publicznych (Suwajanakorn i in., 2017). Badacze wykorzystali setki godzin nagrań Baracka Obamy, aby nauczyć model charakterystycznych ruchów ust, mimiki i gestów byłego prezydenta USA. Następnie system potrafił zsynchronizować ruchy ust z dowolnym nagraniem audio, tworząc wideo wyglądające jak autentyczne przemówienie. Projekt ten wywołał szeroką debatę, ponieważ po raz pierwszy zaprezentowano technologię umożliwiającą przypisywanie osobom publicznym wypowiedzi, których nigdy nie wygłosiły. Od czasu tych dwóch projektów technologia deepfake znalazła zastosowanie w szerokim spektrum obszarów: od internetowych memów, poprzez sztukę, aż po przemysł filmowy (Chesney, Citron, 2019). Największe zainteresowanie opinii publicznej wzbudziły jednak zastosowania związane z treściami politycznymi, takie jak fałszywe nagrania przemówień Borisa Johnsona i Jeremy-'ego Corbyna opublikowane przez think tank Future Advocacy w kontekście wyborów powszechnych w Wielkiej Brytanii w 2019 roku (Future Advocacy, b.d.). W 2020 roku grupa aktywistów Extinction Rebellion udostępniła fałszywe nagranie premier Belgii, Sophie Wilmès, sugerujące możliwy związek

między wylesianiem a pandemią COVID19 (Galindo, 2020). Te szeroko komentowane przykłady, w połączeniu z szybkim postępem technologicznym w zakresie jakości deepfake'ów, wywołały poważne obawy w środowisku zajmującym się etyką sztucznej inteligencji. Eksperci zwracają uwagę, że największym zagrożeniem jest niemożność odróżnienia deepfake od prawdziwych treści (Etienne, 2021).

Deepfake jako zagrożenie dla bezpieczeństwa informacji

Technologie oparte na sztucznej inteligencji rozwijają się w zawrotnym tempie, wspierając niemal każdy aspekt naszego codziennego życia, od rozrywki i pracy po edukację i bezpieczeństwo. Jednak wraz z postępem rośnie także skala zagrożeń. Jednym z najbardziej niepokojących przykładów złośliwego wykorzystania sztucznej inteligencji przez cyberprzestępców jest technologia deepfake, generowanie realistycznych, lecz fałszywych treści wizualnych i dźwiękowych. Treści te są następnie wykorzystywane do dezinformacji, niszczenia reputacji i oszustw na niespotykaną dotąd skalę (Zegarow, 2025). Dynamiczny rozwój sztucznej inteligencji oprócz szeregu szans rozwojowych niesie ze sobą także zagrożenia dla praw i wolności osób fizycznych, w szczególności w postaci zjawiska tzw. deepfake'ów, w których generowane przez AI fałszywe obrazy, dźwięki i nagrania wideo, wydają się coraz bardziej wiarygodne, co zwiększa przestrzeń do nadużyć i odziaływania negatywnych konsekwencji tego typu treści dla osób fizycznych, których dane osobowe mogą być bezprawnie wykorzystywane (Zontek, 2025). Niemożność jednoznacznego odróżnienia deepfake od materiałów autentycznych stanowi jedno z kluczowych zagrożeń związanych z rozwojem technologii deepfake. W raporcie Deeptrace, wskazano, że jakość treści generowanych przy użyciu tej technologii osiągnęła poziom, na którym zarówno dla użytkowników indywidualnych, jak i dla systemów automatycznej detekcji staje się coraz trudniejsze wiarygodne rozróżnienie pomiędzy mediami autentycznymi a syntetycznymi *(„the quality of deepfake content has reached a level where it is increasingly difficult for both humans and automated systems to reliably distinguish between real and synthetic media")* (Ajder i in., 2019). Inne zagrożenia wynikające z technologii deepfake to: zjawisko „liar's dividend", dezinformacja polityczna, naruszenie dóbr osobistych, oszustwa finansowe i ataki socjotechniczne. Upowszechnienie deepfake'ów prowadzi do zjawiska określanego mianem „*liar's dividend"*, polegającego na możliwości

kwestionowania nawet autentycznych materiałów jako fałszywych (Chesney, Citron, 2019). W dłuższej perspektywie zjawisko to może osłabić zaufanie społeczne do mediów, instytucji publicznych oraz dowodów wykorzystywanych w postępowaniach sądowych. Deepfake'i stanowią szczególnie groźne narzędzie dezinformacji, zwłaszcza w kontekście politycznym. Mogą być wykorzystywane do generowania fałszywych wypowiedzi osób publicznych, co sprzyja manipulacji opinią publiczną i ingerencji w procesy demokratyczne. Jak wskazuje literatura, deepfake'i umożliwiają masową dystrybucję nieprawdziwych, lecz wysoce realistycznych treści audiowizualnych, co w konsekwencji prowadzi do osłabienia jakości dyskursu *demokratycznego („deepfakes have the potential to undermine democratic discourse by enabling the large-scale dissemination of false but highly convincing audiovisual information"* (Ajder i in., 2019). Deepfake w kontekście naruszeń dóbr osobistych - brak zgody na wykorzystywanie wizerunku, w szczególności w materiałach pornograficznych to obecnie kolejne najpowszechniejsze i jednocześnie najbardziej szkodliwe zastosowanie tej technologii. Skutki tego rodzaju nadużyć obejmują poważne konsekwencje psychologiczne, społeczne oraz reputacyjne dla osób, których wizerunek został wykorzystany bez ich zgody („*non-consensual deepfake pornography represents the most prevalent and harmful use of deepfake technology to date*") (Ajder i in., 2019). Deepfake jako narzędzie oszustw finansowych i ataków socjotechnicznych – Deepfake'i stanowią także istotne zagrożenie dla bezpieczeństwa informacji oraz ochrony tożsamości. Technologia ta jest coraz częściej wykorzystywana w oszustwach finansowych oraz atakach socjotechnicznych, w których syntetyczne nagrania głosu i obrazu służą do podszywania się pod konkretne osoby i omijania mechanizmów uwierzytelniania. Od zarania cyberbezpieczeństwa, najsłabszym ogniwem w każdym systemie obronnym był, jest i prawdopodobnie zawsze będzie człowiek. Atakujący doskonale o tym wiedzą, dlatego sztuka manipulacji, perswazji i oszustwa – czyli socjotechnika – od zawsze była ich najskuteczniejszą bronią (Gil, 2026). Jak wskazują analitycy Sensity AI, syntetyczne materiały audio i wideo mogą zostać wykorzystane jako narzędzie umożliwiające obchodzenie systemów weryfikacji tożsamości oraz realizację oszustw na masową skalę („*synthetic audio and video can be weaponized to bypass identity verification systems and facilitate fraud at scale*") (Sensity AI, 2021). Prawo często nie nadąża za szybkim rozwojem technologii. Istniejące regulacje dotyczące ochrony danych osobowych, prawa autorskiego czy zniesławienia nie zawsze obejmują specyfikę technologii deepfake. Zróżnicowane

podejścia legislacyjne utrudniają skuteczne egzekwowanie odpowiedzialności i zwalczanie nadużyć, zwłaszcza w kontekście transgranicznym (Afshari, Mohammadi, 2025). Obniżenie progu wejścia do tworzenia deepfake'ów, dzięki narzędziom open source i usługom chmurowym, stanowi paradoksalne wyzwanie. Z jednej strony sprzyja innowacjom, twórczości i zastosowaniom edukacyjnym, z drugiej umożliwia masową produkcję treści manipulacyjnych przez osoby bez specjalistycznej wiedzy (Ajder i in., 2019).

Tabela 1. Deepfake jako zagrożenie dla bezpieczeństwa informacji

Typ zagrożenia	Mechanizm deepfake	Przykład	Adekwatne środki prawne/ techniczne
Podszywanie się (impersonation)	Realtime deepfake: obraz + głos	Hongkong 2024: „wszyscy uczestnicy rozmowy, poza ofiarą, byli wygenerowani jako deepfake"	Weryfikacja poza kanałem, procedury potwierdzania tożsamości
Oszustwa finansowe	Klonowanie głosu, manipulacja wideo	Arup 2024; Singapur 2025	Zasada podwójnej autoryzacji, limity przelewów, szkolenia pracowników
Dezinformacja polityczna	Fałszywe wypowiedzi osób publicznych	Przykłady z UK i Belgii	Oznaczanie treści (AI Act art. 50), fact-checking, edukacja medialna
Naruszenia dóbr osobistych	Deepfake pornograficzny	Raport Deeptrace: „non-consensual deepfake pornography represents the most prevalent and harmful use"	Deepfake naruszający dobra osobiste może rodzić odpowiedzialność cywilną i karną. Kluczowe znaczenie ma jednak szybka reakcja: zgłoszenie treści do usunięcia, wykorzystanie procedur platform oraz bieżący monitoring internetu. W praktyce skuteczna ochrona zależy bardziej od natychmiastowego ograniczenia szkody niż od późniejszego dochodzenia roszczeń.

Źródło: opracowanie na podstawie: R. Chesney, D.K. Citron, „Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", California Law Review 2019, t. 107, s. 1753-1820; N. Afshari, A. Mohammadi, „The Legal Implications of Deepfake Technology: Privacy, Defamation, and the Challenge of Regulating Synthetic Media", Legal Studies in Digital Age 2023, t. 2, nr 2, s. 13-23; South China Morning Post, 4.02.2024; The Guardian, 17.05.2024; Singapore Police Force, 7.04.2025; Deeptrace, The State of Deepfakes: Landscape, Threats, and Impact, 2019.

Deepfake jako zagrożenie dla bezpieczeństwa informacji nie może być neutralizowany przy użyciu pojedynczego instrumentu. Skuteczna odpowiedź wymaga modelu trójfilarowego, łączącego regulacje prawne, odporność organizacyjnotechniczną oraz rozwój kompetencji w zakresie sztucznej inteligencji. Dopiero współdziałanie tych filarów pozwala ograniczyć strukturalne skutki demokratyzacji technologii deepfake i chronić zaufanie do komunikacji cyfrowej.

Demokratyzacja technologii deepfake

W erze, kiedy sztuczna inteligencja (AI) staje się coraz bardziej powszechna, istnieje rosnące zapotrzebowanie na demokratyzację dostępu do generatywnej sztucznej inteligencji (GAI). Jednak wraz z tą demokratyzacją pojawia się również potrzeba świadomości i odpowiedzialnego korzystania z tej potężnej technologii. Demokratyzacja oznacza, że coraz więcej osób ma dostęp do narzędzi opartych na sztucznej inteligencji, co z kolei rodzi wyzwania związane z etyką, prywatnością i bezpieczeństwem danych. Dlatego kluczowym elementem tego procesu jest edukacja społeczeństwa na temat zasad funkcjonowania i potencjalnych skutków AI (Integral Solution, b.d.). Demokratyzacja technologii *deepfake* oznacza, że tworzenie realistycznych, syntetycznych treści audiowizualnych przestało być domeną wąskiej grupy specjalistów. Jeszcze kilka lat temu generowanie takich materiałów wymagało zaawansowanej wiedzy technicznej, dostępu do specjalistycznego sprzętu oraz znacznych nakładów finansowych. Obecnie bariery te uległy istotnemu obniżeniu, co ma bezpośredni wpływ na skalę i charakter nadużyć związanych z wykorzystaniem deepfake'ów (Chesney, Citron, 2019). Jednym z kluczowych czynników tego procesu jest powszechna dostępność narzędzi typu open source, które można bezpłatnie pobrać i wykorzystać. Oprogramowanie to jest często wyposażone w gotowe instrukcje oraz uproszczone procedury, co sprawia, że jego obsługa nie wymaga specjalistycznego przygotowania informatycznego. W praktyce oznacza to, że każda osoba posiadająca podstawowe umiejętności cyfrowe może wygenerować materiał przypominający autentyczne nagranie wideo lub audio, bez konieczności rozumienia mechanizmów działania algorytmów sztucznej inteligencji. Z perspektywy bezpieczeństwa informacji istotne jest to, że otwarty charakter takich narzędzi utrudnia kontrolę nad ich wykorzystaniem,

a także sprzyja szybkiemu rozpowszechnianiu rozwiązań umożliwiających tworzenie coraz bardziej realistycznych treści manipulacyjnych (Ajder i in., 2019, s. 3-6). Kolejnym elementem obniżającym próg wejścia są usługi chmurowe oraz modele oferowane w formule „as a service". W tym modelu użytkownik nie musi posiadać własnej infrastruktury technicznej ani instalować specjalistycznego oprogramowania. Wystarczy dostęp do internetu oraz podstawowe dane wejściowe (np. nagranie głosu lub zdjęcia), aby w krótkim czasie wygenerować gotowy materiał. Z punktu widzenia prawa i regulacji szczególnie istotne jest to, że takie usługi działają globalnie, często poza jurysdykcją państwa, w którym dochodzi do nadużycia. Ułatwia to zarówno masową produkcję treści manipulacyjnych, jak i zdecydowanie utrudnioną identyfikację sprawcy oraz przypisania mu odpowiedzialności prawnej, zwłaszcza w przypadkach o charakterze transgranicznym (Ajder i in., 2019; Chesney, Citron, 2019). Najważniejszą konsekwencją opisanych zmian jest jednak fakt, że tworzenie deepfake'ów nie wymaga dziś specjalistycznej wiedzy technicznej. Nowoczesne narzędzia są projektowane w taki sposób, aby użytkownik mógł posługiwać się nimi intuicyjnie – poprzez wybór gotowych opcji, wgrywanie plików i zatwierdzanie rezultatów. W efekcie technologia deepfake staje się dostępna dla bardzo szerokiego kręgu użytkowników, w tym także dla osób działających w złej wierze. Prowadzi to do jakościowej zmiany zagrożeń: zamiast pojedynczych, incydentalnych przypadków manipulacji pojawia się zjawisko masowe, trudne do kontrolowania i jeszcze trudniejsze do skutecznego sankcjonowania w ramach istniejących instrumentów prawnych (Ajder i in., 2019; Chesney, Citron, 2019). Opisane obniżenie progu wejścia stanowi punkt wyjścia do dalszych rozważań nad eskalacją nadużyć związanych z deepfake'ami. Pokazuje ono, że problem nie leży wyłącznie w samej technologii, lecz w jej powszechnej dostępności, która wyprzedza zdolność systemów prawnych i instytucjonalnych do reagowania. Z tego względu demokratyzacja technologii deepfake powinna być traktowana jako czynnik strukturalny, determinujący zarówno skalę zagrożeń, jak i utrudnioną ochronę prawną (Chesney, Citron, 2019).

Studia przypadków (Case studies) nadużyć związanych z demokratyzacją technologii deepfake

Podszywanie się w rozmowach wideo w czasie rzeczywistym (*real-time deepfake*)

Jednym z najbardziej problematycznych przykładów nadużyć związanych z demokratyzacją technologii deepfake są przypadki podszywania się pod osoby funkcyjne podczas rozmów wideo prowadzonych na żywo, wykorzystywane w szczególności do wyłudzeń finansowych. W tego rodzaju sytuacjach sprawca jest w stanie w czasie rzeczywistym wygenerować obraz i głos osoby pełniącej określoną funkcję w organizacji, takiej jak dyrektor, członek zarządu czy bezpośredni przełożony, co znacząco zwiększa wiarygodność fałszywej komunikacji.

Z punktu widzenia praktyki organizacyjnej istotne jest, że ataki te nie polegają na przesyłaniu wcześniej przygotowanego nagrania, lecz na prowadzeniu interaktywnej rozmowy, która z perspektywy odbiorcy nie różni się od autentycznego spotkania online. Pozwala to sprawcy na bieżąco reagować na pytania rozmówcy i dostosowywać przekaz do sytuacji, co dodatkowo utrudnia wykrycie nadużycia. Skuteczność tego rodzaju podszywania się znacząco wzrasta w organizacjach, które opierają weryfikację tożsamości wyłącznie na cechach behawioralnych lub wizualnych, takich jak rozpoznanie twarzy, głosu czy charakterystycznego sposobu wypowiadania się, bez stosowania dodatkowych procedur potwierdzających. W konsekwencji decyzje o charakterze finansowym lub organizacyjnym mogą zostać podjęte na podstawie komunikacji, która sprawia wrażenie w pełni autentycznej, mimo że w rzeczywistości została wygenerowana syntetycznie.

Z perspektywy prawa i bezpieczeństwa informacji przypadki te pokazują, że biometria behawioralna przestaje być wystarczającym środkiem uwierzytelnienia, a zaufanie do form komunikacji audiowizualnej jako dowodu lub podstawy decyzji ulega istotnemu osłabieniu. Zjawisko to unaocznia również, w jaki sposób powszechna dostępność narzędzi deepfake prowadzi do jakościowej zmiany zagrożeń – od sporadycznych incydentów do realnego ryzyka systemowego dla funkcjonowania instytucji i obrotu gospodarczego.

Hongkong – fałszywa wideokonferencja z „CFO" (2024)

Jednym z najbardziej nagłośnionych przypadków podszywania się przy użyciu technologii deepfake była sprawa ujawniona przez policję w Hongkongu na początku 2024 r. Pracownik działu finansowego międzynarodowej spółki został zaproszony na wideokonferencję, w której uczestniczyły osoby wyglądające i brzmiące jak znani mu współpracownicy, w tym dyrektor finansowy spółki. Podczas spotkania polecono mu dokonanie pilnych przelewów na wskazane rachunki bankowe.

Pracownik początkowo podejrzewał próbę oszustwa, jednak po dołączeniu do rozmowy wideo i rozpoznaniu twarzy oraz głosów przełożonych uznał polecenie za autentyczne. W rezultacie dokonał 15 przelewów na łączną kwotę około 25 milionów dolarów amerykańskich. Dopiero późniejsza weryfikacja wykazała, że wszyscy uczestnicy rozmowy – poza ofiarą – byli wygenerowani jako deepfake, a wizerunki i głosy powstały na podstawie publicznie dostępnych materiałów. Sprawa została szeroko opisana w prasie azjatyckiej i międzynarodowej (Harvey Kong, 2024).

Przypadek ten pokazuje, że obecność rozmówcy w czasie rzeczywistym, a nawet udział wielu „uczestników", nie może być już traktowana jako wystarczające potwierdzenie tożsamości. Z perspektywy odpowiedzialności organizacyjnej istotne jest, że decyzje finansowe zostały podjęte na podstawie komunikacji, która wizualnie i głosowo spełniała wszystkie dotychczasowe kryteria autentyczności.

Arup – potwierdzona ofiara deepfake'ów (Wielka Brytania / Hongkong, 2024)

Kolejnym istotnym przykładem jest sprawa brytyjskiej firmy inżynieryjnej Arup, która publicznie potwierdziła, że padła ofiarą oszustwa wykorzystującego technologię deepfake. Pracownik spółki w Hongkongu został nakłoniony do przekazania około 20 milionów funtów brytyjskich po rozmowie wideo z osobami podszywającymi się pod kadrę kierowniczą. Spółka wskazała, że w oszustwie wykorzystano fałszywe obrazy i głosy, a komunikacja miała charakter przekonującego spotkania służbowego. Co istotne, Arup podkreślił, że nie doszło do naruszenia systemów informatycznych, a mechanizm oszustwa polegał wyłącznie na przełamaniu zaufania pracownika do komunikacji audiowizualnej. Sprawa Arup ma szczególne znaczenie,

ponieważ dotyczy dużej, profesjonalnej organizacji, posiadającej rozbudowane procedury bezpieczeństwa. Pokazuje to, że deepfake'i nie są zagrożeniem marginalnym ani ograniczonym do słabo zabezpieczonych podmiotów, lecz stanowią ryzyko systemowe dla obrotu gospodarczego (The Guardian, 2024).

Singapur – „narada kierownictwa" i przelew 499 tys. USD (2025)

W 2025 r. opisano również przypadek z Singapuru, w którym dyrektor finansowy spółki uczestniczył w wideokonferencji przypominającej naradę kierownictwa. W trakcie spotkania osoby wyglądające jak znani menedżerowie przedstawiły pilną potrzebę dokonania przelewu w związku z poufną transakcją. Dyrektor, działając w przekonaniu, że uczestniczy w autentycznym spotkaniu, zatwierdził przelew na kwotę około 499 tysięcy dolarów amerykańskich. Późniejsze ustalenia wykazały, że cała rozmowa została wygenerowana za pomocą technologii deepfake, a sprawcy posłużyli się wizerunkami i głosami opartymi na materiałach publicznych. Przypadek ten potwierdza, że technologia deepfake umożliwia skalowalne oszustwa, obejmujące zarówno bardzo wysokie, jak i „średnie" kwoty, a więc potencjalnie dotykające znacznie szerszego kręgu podmiotów (Tookitaki, 2025).

Studium przypadku nadużyć z wykorzystaniem deepfake'ów w państwach Unii Europejskiej

Zjawisko podszywania się przy użyciu technologii deepfake nie ogranicza się do państw trzecich ani do pojedynczych, incydentalnych zdarzeń, lecz znajduje swoje potwierdzenie również w praktyce państw członkowskich Unii Europejskiej. Przypadki opisywane w prasie europejskiej pokazują, że technologia ta jest wykorzystywana zarówno w obrocie gospodarczym, jak i w sferze publicznej, prowadząc do poważnych strat finansowych oraz podważenia zaufania do komunikacji cyfrowej (Reuters, 2025).

Jednym z najbardziej znaczących przykładów jest sprawa ujawniona we Włoszech w 2025 r., w której sprawcy posłużyli się wygenerowanym za pomocą technologii deepfake głosem ministra obrony w celu wyłudzenia środków finansowych od znanych przedsiębiorców. W rozmowach telefonicznych podszywano się pod przedstawicieli administracji publicznej,

powołując się na rzekome działania państwowe wymagające pilnego wsparcia finansowego. Co najmniej jeden z adresatów tych rozmów dokonał przelewu na kwotę około jednego miliona euro, pozostając w przekonaniu, że działa na podstawie autentycznego polecenia organu państwowego. Sprawa ta została nagłośniona w mediach międzynarodowych i stała się przedmiotem postępowania organów ścigania, którym udało się zablokować przekazane środki.

Przypadek ten pokazuje, że deepfake'i mogą być wykorzystywane do podszywania się nie tylko pod osoby prywatne lub kadrę zarządzającą przedsiębiorstw, lecz także pod przedstawicieli władzy publicznej. W konsekwencji zagrożone zostaje nie tylko bezpieczeństwo obrotu gospodarczego, lecz również zaufanie obywateli do instytucji państwowych i oficjalnych kanałów komunikacji.

W jednym z pierwszych szeroko opisywanych przypadków w Europie dyrektor spółki zależnej w UE otrzymał telefon od osoby, która brzmiała dokładnie jak prezes spółkimatki z Niemiec. Na podstawie rozmowy telefonicznej nakazano mu dokonanie pilnego przelewu ok. 220 tys. EUR. Dopiero później ustalono, że rozmowa została przeprowadzona przy użyciu technologii klonowania głosu. Sprawa ta jest często wskazywana w literaturze jako *początek fali deepfake'ów w oszustwach korporacyjnych w Europie,* pokazując, że problem istniał jeszcze przed masowym upowszechnieniem technologii deepfake (ZDNet, 2019).

Opisane przypadki pokazują, że podszywanie się w rozmowach wideo przy użyciu technologii deepfake nie ma charakteru hipotetycznego, lecz stanowi realne, powtarzalne zagrożenie, prowadzące do poważnych strat finansowych. Kluczowe jest to, że tradycyjne kryteria weryfikacji tożsamości oparte na obrazie i głosie tracą swoją wiarygodność, co wymusza zmianę standardów należytej staranności oraz uzasadnia interwencję regulacyjną, w tym rozwiązania prawne przyjęte w odniesieniu do deepfake w UE – Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Artificial Intelligence Act), Dz. U. UE L 2024/1689, art. 3 pkt 60, art. 4, art. 50 ust. 2 (EUR-Lex, b.d.).

Powyższe dane empiryczne stanowią podstawę do interpretacji, że demokratyzacja technologii deepfake prowadzi do jakościowej zmiany zagrożeń.

Deepfake a prawo – wyzwania regulacyjne

Regulacja deepfake'ów w Rozporządzeniu o sztucznej inteligencji (AI Act) – Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 w sprawie sztucznej inteligencji (AI Act) stanowi pierwszy kompleksowy akt prawa unijnego, który wprost odnosi się do zjawiska deepfake'ów jako systemowego zagrożenia dla zaufania do komunikacji cyfrowej. Ustawodawca unijny przyjął założenie, że problemem nie jest wyłącznie sama technologia, lecz powszechność jej stosowania oraz ryzyko wprowadzania odbiorców w błąd co do autentyczności treści audiowizualnych.

Definicja deepfake w AI Act (art. 3 pkt 60 AI Act) – AI Act wprowadza legalną definicję „deepfake", w której „deepfake" oznacza wygenerowane przez AI lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmiot (art. 3 pkt 60 AI Act). Definicja ta ma kluczowe znaczenie interpretacyjne, ponieważ pozwala objąć regulacją nie tylko treści tworzone w całości jako technologia deepfake, lecz także materiały częściowo zmodyfikowane, jeżeli mogą one wprowadzać odbiorcę w błąd co do ich rzeczywistego charakteru.

Obowiązek ujawniania treści deepfake (transparentność) – Centralnym mechanizmem regulacyjnym AI Act w odniesieniu do deepfake'ów jest obowiązek transparentności. Zgodnie z art. 50 AI Act, podmioty udostępniające lub stosujące systemy AI generujące lub modyfikujące treści audiowizualne są zobowiązane do poinformowania odbiorców, że dana treść została wygenerowana lub zmanipulowana przez AI, jeżeli mogłaby ona sprawiać wrażenie autentycznej - Motywy Rozporządzenia (UE) 2024/1689 dotyczące transparentności treści generowanych przez AI oraz ochrony odbiorców przed wprowadzaniem w błąd. Celem tego rozwiązania nie jest zakazanie deepfake'ów jako takich, lecz przywrócenie równowagi informacyjnej po stronie odbiorcy. Ustawodawca unijny uznał, że brak wiedzy o syntetycznym charakterze treści stanowi zasadnicze źródło ryzyka prawnego i społecznego, w tym ryzyka manipulacji, oszustw oraz naruszenia dóbr osobistych – Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Artificial Intelligence Act), art. 50 ust. 2, Dz. U. UE L 2024/1689 (EUR-Lex, b.d.).

Obowiązek kompetencji (AI literacy) – Istotnym uzupełnieniem obowiązku ujawniania jest wprowadzony w art. 4 AI Act obowiązek zapewnienia

odpowiedniego poziomu kompetencji w zakresie sztucznej inteligencji po stronie dostawców oraz podmiotów stosujących systemy AI. Przepis ten ma charakter horyzontalny i obejmuje również podmioty korzystające z narzędzi generujących treści deepfake.

Z perspektywy prawa oznacza to, że brak wiedzy technicznej lub nieświadomość ryzyk związanych z deepfake'ami nie może stanowić usprawiedliwienia dla podmiotów profesjonalnych. AI Act przesuwa ciężar odpowiedzialności z indywidualnego odbiorcy na organizacje, które decydują się wykorzystywać systemy AI w działalności gospodarczej lub instytucjonalnej.

Czas obowiązywania przepisów AI Act – AI Act został opublikowany w Dzienniku Urzędowym UE w 2024 r. i wszedł w życie 20 dni po publikacji. Z uwagi na zakres regulacji przewidziano jednak stopniowe stosowanie przepisów. Przepisy ogólne, w tym definicje oraz obowiązek AI literacy, zaczęły obowiązywać wcześniej. Obowiązki dotyczące transparentności deepfake'ów, w tym art. 50, stosuje się od 2 sierpnia 2026 r. Oznacza to, że podmioty działające na rynku unijnym mają określony czas na dostosowanie procedur organizacyjnych i komunikacyjnych do nowych wymogów**.**

Deregulacja („Omnibus") a regulacja deepfake w AI Act

W debacie publicznej oraz eksperckiej towarzyszącej wdrażaniu Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 w sprawie sztucznej inteligencji (AI Act) pojawiły się pytania o wpływ równolegle zapowiadanych inicjatyw deregulacyjnych Komisji Europejskiej, określanych zbiorczo mianem pakietu „Omnibus", na zakres i trwałość obowiązków przewidzianych w tym akcie. W szczególności dotyczy to regulacji odnoszących się do treści typu deepfake, które już na etapie prac legislacyjnych zostały uznane za istotne zagrożenie dla zaufania do komunikacji cyfrowej.

Pod pojęciem „Omnibus deregulacyjny" rozumie się zapowiadany w latach 2024-2025 pakiet inicjatyw Komisji Europejskiej mających na celu ograniczenie nadmiernych obciążeń administracyjnych, uproszczenie obowiązków sprawozdawczych oraz poprawę konkurencyjności przedsiębiorstw działających na rynku wewnętrznym Unii Europejskiej. Należy podkreślić, że Omnibus nie stanowi jednego, spójnego aktu prawnego, lecz zbiór propozycji o zróżnicowanym charakterze normatywnym, obejmujących m.in. zapowiedzi legislacyjne, przeglądy istniejących regulacji oraz instrumenty

typu soft law. Zakres tych inicjatyw koncentruje się przede wszystkim na obszarach takich jak raportowanie niefinansowe, regulacje ESG czy wybrane obowiązki compliance.

Na obecnym etapie brak jest jednak jakichkolwiek podstaw do twierdzenia, że inicjatywy deregulacyjne Omnibus wpływają na regulację deepfake'ów zawartą w AI Act. Rozporządzenie to zostało przyjęte jako akt o charakterze horyzontalnym i systemowym, którego celem jest ochrona interesów publicznych o fundamentalnym znaczeniu, w tym zaufania do rynku, ochrony konsumentów oraz praw podstawowych. Przepisy odnoszące się do deepfake'ów, w szczególności legalna definicja zawarta w art. 3 pkt 60 AI Act oraz obowiązek transparentności przewidziany w art. 50 AI Act, nie zostały objęte żadnym projektem uchylenia, zawieszenia ani modyfikacji w ramach zapowiadanych działań deregulacyjnych. Nie istnieje również akt prawa wtórnego ani projekt legislacyjny, który zmieniałby obowiązek ujawniania syntetycznego charakteru treści audiowizualnych generowanych lub modyfikowanych przez systemy AI.

Trwałość regulacji deepfake w AI Act znajduje silne uzasadnienie aksjologiczne i funkcjonalne. Deepfake'i zostały bowiem zakwalifikowane przez unijnego prawodawcę nie jako zwykłe obciążenie administracyjne dla przedsiębiorców, lecz jako zjawisko stwarzające ryzyko systemowe dla zaufania publicznego. Obowiązek transparentności przewidziany w art. 50 AI Act służy ochronie odbiorcy przed wprowadzaniem w błąd i odnosi się bezpośrednio do takich wartości jak integralność rynku wewnętrznego, bezpieczeństwo obrotu gospodarczego, ochrona konsumentów oraz poszanowanie praw podstawowych, w tym prawa do informacji i autonomii decyzyjnej. Są to obszary, które Komisja Europejska konsekwentnie wyłącza z zakresu deregulacji, uznając je za nienegocjowalne elementy porządku prawnego Unii.

W konsekwencji inicjatywy deregulacyjne mogą wywierać jedynie pośredni wpływ na praktykę stosowania przepisów dotyczących deepfake-'ów, nie zaś na ich obowiązywanie normatywne. Możliwy wpływ Omnibusa może przejawiać się na poziomie interpretacyjnym, w szczególności poprzez akcentowanie zasady proporcjonalności w sposobie realizacji obowiązków transparentności, większą rolę wytycznych Komisji Europejskiej oraz organów nadzorczych, a także promowanie technicznych form samoregulacji, takich jak standardy branżowe czy rozwiązania typu watermarking.

Nie oznacza to jednak zniesienia ani osłabienia obowiązku ujawniania treści deepfake, lecz jedynie potencjalne zróżnicowanie form jego wykonania.

Powyższe prowadzi do wniosku, że regulacja deepfake'ów w AI Act stanowi trwały element unijnego porządku prawnego, odporny na bieżące inicjatywy deregulacyjne. Deepfake nie jest bowiem traktowany jako koszt regulacyjny, lecz jako ryzyko strukturalne dla bezpieczeństwa informacji i zaufania do komunikacji cyfrowej, co uzasadnia utrzymanie rygorystycznych obowiązków transparentności niezależnie od zmian w innych obszarach prawa unijnego.

Deepfake a ochrona danych osobowych (GDPR)

Rozwój technologii deepfake rodzi istotne pytania dotyczące zakresu i skuteczności ochrony przewidzianej w Rozporządzeniu Parlamentu Europejskiego i Rady (UE) 2016/679 (GDPR) (EUR-Lex, b.d.).

Mogłoby się wydawać, że regulacja ochrony danych osobowych stanowi wystarczający instrument przeciwdziałania nadużyciom związanym z wykorzystywaniem wizerunku, głosu lub innych cech osoby fizycznej w technologii deepfake. Pogłębiona analiza prowadzi jednak do wniosku, że GDPR jedynie częściowo obejmuje zjawisko deepfake i nie odpowiada na jego kluczowe ryzyka systemowe. GDPR chroni dane osobowe, ale nie chroni prawdziwości ani autentyczności treści, co stanowi lukę w kontekście deepfake. (European Data Protection Board, 2024). GDPR nie przeciwdziała skutecznie dezinformacji, manipulacji ani erozji zaufania, ponieważ koncentruje się na relacji administrator – dane, a nie na skutkach społecznych mediów syntetycznych. (European Parliament Tackling deepfakes in European policy, 2021).

Analiza prowadzi jednak do wniosku, że GDPR jedynie częściowo obejmuje zjawisko deepfake i nie odpowiada na jego kluczowe ryzyka systemowe.

Po pierwsze, choć wizerunek twarzy, nagranie głosu oraz cechy biometryczne wykorzystywane w technologii deepfake co do zasady mieszczą się w definicji danych osobowych, a niekiedy również danych biometrycznych w rozumieniu GDPR, regulacja ta koncentruje się na legalności przetwarzania danych, a nie na skutkach społecznych i informacyjnych generowanych treści syntetycznych. GDPR chroni dane osobowe jako takie, lecz nie zapewnia

ochrony prawdziwości ani autentyczności treści, co stanowi zasadniczą lukę w kontekście deepfake'ów, których istotą jest wytworzenie realistycznej, lecz fałszywej reprezentacji osoby mogącej wprowadzać odbiorców w błąd.

Po drugie, GDPR nie przeciwdziała skutecznie dezinformacji, manipulacji ani erozji zaufania do komunikacji audiowizualnej. Mechanizmy ochrony danych osobowych nie są projektowane z myślą o zapobieganiu zjawiskom takim jak masowa dystrybucja deep fake, manipulacja opinią publiczną czy tzw. *liar's dividend*, polegający na możliwości kwestionowania autentycznych materiałów jako rzekomo wygenerowanych przez sztuczną inteligencję. Ochrona danych osobowych nie jest bowiem instrumentem służącym ochronie integralności dyskursu publicznego ani zaufania do informacji jako takiej.

Po trzecie, model odpowiedzialności przyjęty w GDPR opiera się na możliwości identyfikacji administratora danych oraz skutecznego egzekwowania wobec niego obowiązków. W praktyce deepfake'i są jednak często tworzone i rozpowszechniane w środowisku transgranicznym, anonimowo lub przy wykorzystaniu zdecentralizowanych platform, co znacząco ogranicza realną skuteczność instrumentów takich jak prawo do usunięcia danych czy prawo do sprzeciwu. Usunięcie danych źródłowych nie prowadzi ponadto do usunięcia skutków ich wcześniejszego wykorzystania, zwłaszcza gdy modele generatywne zostały już wytrenowane, a treści weszły do obiegu informacyjnego.

Powyższe ograniczenia tłumaczą, dlaczego GDPR nie może być traktowane jako wystarczający instrument reagowania na zagrożenia związane z deepfake'ami. Ochrona przed skutkami tej technologii wymaga bowiem nie tylko kontroli nad danymi osobowymi, lecz także regulacji ukierunkowanych na transparentność treści, ochronę odbiorcy przed wprowadzaniem w błąd oraz przeciwdziałanie ryzykom systemowym dla zaufania do komunikacji cyfrowej. Z tego względu unijny prawodawca zdecydował się na przyjęcie odrębnych regulacji w AI Act, które uzupełniają, a nie zastępują, mechanizmy ochrony danych osobowych przewidziane w GDPR. (European Data Protection Board, 2024; European Parliament – STOA, 2021; Levitt, 2024).

Na podstawie przedstawionych regulacji można stwierdzić, że istniejące ramy prawne nie zapewniają ochrony przed systemowymi skutkami dezinformacji – co potwierdza obserwacja, że RODO nie nadąża za dynamiką rozwoju technologii i nie zapewnia ochrony przed systemowymi skutkami dezinformacji.

Od ochrony danych osobowych do ochrony zaufania informacyjnego: ewolucja podejścia Unii Europejskiej.

Analiza ograniczeń GDPR w kontekście technologii deepfake prowadzi do szerszego wniosku o charakterze systemowym. Regulacja ochrony danych osobowych została zaprojektowana w odmiennym paradygmacie ryzyk. GDPR koncentruje się na relacji administrator-dane oraz na ochronie jednostki przed nieuprawnionym przetwarzaniem informacji dotyczących jej osoby. Tymczasem technologia deepfake generuje zagrożenia, które wykraczają poza sferę indywidualnej prywatności i obejmują bezpieczeństwo informacyjne, wiarygodność komunikacji oraz zaufanie do treści audiowizualnych jako podstawy podejmowania decyzji.

W tym sensie deepfake ujawnia granice paradygmatu ochrony danych osobowych jako podstawowego instrumentu reagowania na nowe zagrożenia informacyjne. Nawet pełna zgodność z wymogami GDPR nie zapobiega bowiem erozji zaufania do komunikacji cyfrowej, masowej dezinformacji ani możliwości kwestionowania autentycznych treści jako rzekomo wygenerowanych przez sztuczną inteligencję. Ochrona danych osobowych nie jest narzędziem zaprojektowanym do ochrony prawdziwości, autentyczności ani integralności informacji jako takiej.

Na tym tle Rozporządzenie o sztucznej inteligencji (AI Act) należy postrzegać nie jako konkurencyjną wobec GDPR regulację, lecz jako kolejny etap ewolucji unijnego podejścia do ryzyk technologicznych. AI Act przesuwa punkt ciężkości z ochrony danych osobowych sensu stricto na ochronę odbiorcy informacji oraz stabilności środowiska informacyjnego. Wprowadzenie legalnej definicji deepfake oraz obowiązku transparentności treści technologii deepfake stanowi odpowiedź na ryzyka, których GDPR nie był w stanie objąć ze względu na swój zakres przedmiotowy i aksjologiczny.

Na podstawie dokonanej analizy można stwierdzić stopniowe przejście od regulacji skoncentrowanej na ochronie jednostki przed nadużyciami w przetwarzaniu danych do regulacji ukierunkowanej na ochronę zaufania publicznego, integralności rynku oraz autonomii decyzyjnej odbiorców treści cyfrowych. AI Act nie zastępuje GDPR, lecz je uzupełnia, odpowiadając na jakościowo nowe zagrożenia związane z generatywną sztuczną inteligencją, w tym z masową produkcją treści, które mogą sprawiać wrażenie autentycznych, mimo że zostały wygenerowane za pomocą technologii deepfake.

Deepfake w kontekście prawa autorskiego

Rozwój technologii deepfake rodzi istotne wyzwania na gruncie prawa autorskiego (Ustawa o prawie autorskim i prawach pokrewnych, 1994, art. 1, art. 2, art. 16, art. 17), zarówno w zakresie legalności wykorzystywania materiałów źródłowych, jak i statusu prawnego treści wygenerowanych syntetycznie. Problematyka ta dotyczy w szczególności relacji pomiędzy twórcą pierwotnego utworu, osobą, której wizerunek lub głos są wykorzystywane, a podmiotem generującym deepfake (European Copyright Society, 2023).

W pierwszej kolejności należy zwrócić uwagę, że proces tworzenia deepfake'ów niemal zawsze opiera się na wykorzystaniu istniejących utworów, takich jak nagrania wideo, materiały audio, fotografie czy wystąpienia publiczne (WIPO, 2019-2023). Materiały te, o ile spełniają przesłanki twórczości, podlegają ochronie na gruncie prawa autorskiego. Ich pobieranie, kopiowanie oraz przetwarzanie w celu trenowania modeli generatywnych może stanowić zwielokrotnianie i opracowanie utworu, wymagające zgody uprawnionego, chyba że znajduje zastosowanie jedna z ustawowych licencji lub wyjątków.

W praktyce problematyczne jest w szczególności trenowanie modeli AI na publicznie dostępnych materiałach audiowizualnych, bez zgody autorów. Choć w debacie pojawiają się argumenty odwołujące się do dozwolonego użytku lub analiz tekstów i danych (text and data mining), zakres tych wyjątków jest ograniczony i nie zawsze obejmuje wykorzystanie utworów w celu generowania realistycznych, konkurencyjnych treści audiowizualnych. W konsekwencji deepfake'i mogą prowadzić do naruszenia majątkowych praw autorskich, a w niektórych przypadkach również autorskich praw osobistych, zwłaszcza gdy wygenerowana treść zniekształca sens, kontekst lub wymowę pierwotnego utworu. Odrębną kwestią pozostaje status prawny samego deepfake'a jako utworu (Gervais, 2019). Zgodnie z dominującym poglądem, treści wygenerowane w pełni autonomicznie przez system AI nie korzystają z ochrony prawa autorskiego, ponieważ brak jest elementu twórczości człowieka. W praktyce jednak wiele deepfake'ów powstaje przy istotnym udziale użytkownika, który dobiera materiały źródłowe, parametry generowania oraz dokonuje selekcji efektów. W takich przypadkach możliwe jest uznanie deepfake'a za utwór zależny lub utwór inspirowany, co dodatkowo komplikuje relacje prawnoautorskie i utrudnia jednoznaczne przypisanie odpowiedzialności (Barta, Markiewicz, 2023; Samuelson, 2017).

Z perspektywy bezpieczeństwa informacji istotne jest, że prawo autorskie nie zostało zaprojektowane jako instrument ochrony przed dezinformacją. Nawet skuteczna ochrona praw twórców nie zapobiega bowiem rozpowszechnianiu treści syntetycznych, które formalnie nie naruszają praw autorskich, a jednocześnie mogą wprowadzać odbiorców w błąd co do autentyczności przekazu. Oznacza to, że prawo autorskie pełni w kontekście deepfake'ów funkcję jedynie częściowo ochronną i nie odpowiada na kluczowe ryzyka systemowe.

Deepfake a zniesławienie i ochrona dóbr osobistych

Technologia deepfake stanowi szczególnie poważne wyzwanie w obszarze zniesławienia oraz naruszenia dóbr osobistych, ponieważ umożliwia przypisywanie osobom fizycznym wypowiedzi, zachowań lub cech, których w rzeczywistości nigdy nie przejawiały. W przeciwieństwie do tradycyjnych form fałszu, deepfake'y charakteryzują się wysokim stopniem realizmu, co znacząco zwiększa ich siłę oddziaływania oraz potencjał szkodliwy: Kodeks cywilny – art. 23-24 (ochrona dóbr osobistych); Kodeks karny – art. 212 (zniesławienie), art. 216 (zniewaga). Europejska Konwencja Praw Człowieka – art. 8 i art. 10 (prawo do prywatności vs. wolność wypowiedzi).

Na gruncie prawa cywilnego deepfake może prowadzić do naruszenia takich dóbr osobistych jak cześć, dobre imię, reputacja, wizerunek czy prawo do tożsamości. Wygenerowanie i rozpowszechnienie realistycznego nagrania wideo lub audio, sugerującego określone wypowiedzi lub działania danej osoby, może wywołać trwałe skutki reputacyjne, niezależnie od późniejszego sprostowania. Szczególnie niebezpieczne są sytuacje, w których deepfake funkcjonuje w obiegu informacyjnym przez krótki czas, lecz zdąży wywrzeć znaczący wpływ na opinię publiczną.

W kontekście prawa karnego deepfake'y mogą realizować znamiona zniesławienia lub zniewagi, zwłaszcza gdy przypisują osobie cechy lub działania mogące ją poniżyć w opinii publicznej albo narazić na utratę zaufania potrzebnego do wykonywania określonej funkcji. Charakterystyczną cechą deepfake'ów jest jednak utrudniona identyfikacja sprawcy, często działającego anonimowo, transgranicznie lub przy użyciu zdecentralizowanych platform. W praktyce znacząco ogranicza to skuteczność klasycznych instrumentów odpowiedzialności karnej.

Szczególną kategorię stanowią deepfake'i naruszające dobra osobiste o charakterze intymnym, tworzone i rozpowszechniane bez zgody osoby, której wizerunek lub głos zostały wykorzystane. Skutki takich działań wykraczają poza sferę reputacyjną i obejmują poważne konsekwencje psychologiczne, społeczne oraz zawodowe. W tych przypadkach deepfake staje się narzędziem przemocy informacyjnej, a jego realizm potęguje doznaną krzywdę.

Z perspektywy systemowej kluczowe jest to, że zniesławienie przy użyciu deepfake różni się jakościowo od tradycyjnych form pomówienia. Odbiorca ma bowiem do czynienia nie z opinią lub relacją, lecz z pozornie bezpośrednim „dowodem" w postaci obrazu lub głosu. W rezultacie dochodzi do podważenia zaufania nie tylko do konkretnej osoby, lecz także do komunikacji audiowizualnej jako takiej. Zjawisko to wpisuje się w szerszy problem erozji zaufania informacyjnego, który nie może być skutecznie rozwiązany wyłącznie za pomocą klasycznych instytucji prawa cywilnego i karnego.

W konsekwencji deepfake ujawnia ograniczenia tradycyjnych konstrukcji zniesławienia, opartych na założeniu, że fałszywość przekazu da się stosunkowo łatwo wykazać, a jego autor zidentyfikować. W warunkach masowej produkcji realistycznych treści syntetycznych założenia te przestają być aktualne, co uzasadnia potrzebę uzupełnienia ochrony dóbr osobistych o rozwiązania systemowe, takie jak obowiązek transparentności treści, podnoszenie kompetencji cyfrowych oraz mechanizmy prewencji organizacyjnej (Ajder i in., 2019; WEF, 2024).

Podsumowanie

Analiza technologii deepfake przeprowadzona w niniejszym opracowaniu prowadzi do wniosku, że demokratyzacja tej technologii stanowi czynnik strukturalny współczesnych zagrożeń informacyjnych, a nie jedynie uboczny efekt postępu technologicznego. Obniżenie progu wejścia do tworzenia realistycznych treści poprzez dostępność narzędzi open source, usług chmurowych oraz intuicyjnych interfejsów spowodowało jakościową zmianę charakteru ryzyka. Deepfake przestał być technologią niszową, wymagającą specjalistycznej wiedzy i zaplecza technicznego, a stał się

narzędziem powszechnie dostępnym, zdolnym do masowego oddziaływania na środowisko informacyjne.

Demokratyzacja technologii deepfake oznacza, że zagrożenia związane z manipulacją treściami audiowizualnymi nie mają już charakteru incydentalnego, lecz przybierają formę zjawiska systemowego. Skala potencjalnych nadużyć wykracza poza możliwości reagowania opartych wyłącznie na odpowiedzialności ex post, takich jak dochodzenie roszczeń cywilnych czy ściganie karne sprawców. W warunkach masowej produkcji i dystrybucji treści syntetycznych tradycyjne instrumenty prawne okazują się niewystarczające, a sama identyfikacja sprawcy często staje się niemożliwa. W konsekwencji demokratyzacja deepfake'ów należy postrzegać jako czynnik strukturalnie zmieniający architekturę bezpieczeństwa informacji, a nie jedynie jako nowe narzędzie wykorzystywane w ramach istniejących schematów zagrożeń. W tym kontekście kluczowym wnioskiem jest konieczność łączenia trzech komplementarnych obszarów działań: regulacji prawnych, edukacji w zakresie kompetencji cyfrowych i sztucznej inteligencji (AI literacy) oraz rozwiązań technicznych. Żaden z tych elementów, rozpatrywany w oderwaniu od pozostałych, nie jest w stanie skutecznie przeciwdziałać zagrożeniom wynikającym z masowego wykorzystania deepfake'ów.

Po pierwsze, regulacje prawne, w szczególności przyjęte na poziomie Unii Europejskiej Rozporządzenie o sztucznej inteligencji (AI Act), stanowią niezbędne ramy normatywne dla reagowania na ryzyka związane z deepfake'ami. Wprowadzenie legalnej definicji deepfake oraz obowiązku transparentności należy postrzegać jako próbę przywrócenia równowagi informacyjnej po stronie odbiorcy. Regulacja ta nie zmierza do zakazania technologii jako takiej, lecz do ograniczenia jej potencjału manipulacyjnego poprzez wymóg ujawniania charakteru jej treści. Jednocześnie analiza wykazuje, że prawo nawet w swojej najbardziej nowoczesnej postaci nie jest w stanie samodzielnie zagwarantować bezpieczeństwa informacji w środowisku cyfrowym, w którym treści mogą być generowane i rozpowszechniane z ogromną szybkością.

Po drugie, edukacja i rozwój kompetencji w zakresie AI (AI literacy) stanowią kluczowy element prewencji. Obowiązek zapewnienia odpowiedniego poziomu kompetencji, przewidziany w AI Act, sygnalizuje przesunięcie odpowiedzialności z indywidualnego odbiorcy na organizacje i instytucje korzystające z systemów AI. W praktyce oznacza to

konieczność systematycznego budowania świadomości ryzyk związanych z deepfake'ami, zarówno wśród pracowników sektora publicznego i prywatnego, jak i w społeczeństwie jako całości. Brak wiedzy technicznej lub nieumiejętność krytycznej oceny treści audiowizualnych przestają być kwestią indywidualnej ostrożności, a stają się czynnikiem ryzyka organizacyjnego i systemowego.

Po trzecie, nieodzowne są rozwiązania techniczne, obejmujące zarówno systemy wykrywania deepfake, jak i mechanizmy oznaczania, uwierzytelniania oraz śledzenia pochodzenia treści. W warunkach, w których generowanie deepfake'ów odbywa się w skali masowej i w czasie rzeczywistym, wykrywanie nie może opierać się wyłącznie na ludzkiej percepcji czy manualnej weryfikacji. Rozwiązania techniczne powinny być traktowane nie jako narzędzie opcjonalne, lecz jako integralny element infrastruktury bezpieczeństwa informacji, porównywalny do systemów ochrony przed złośliwym oprogramowaniem czy filtrów antyspamowych.

Zestawienie tych trzech obszarów prowadzi do kluczowego wniosku o konieczności podejścia systemowego do bezpieczeństwa informacji. Deepfake nie jest bowiem wyłącznie problemem technologicznym, prawnym czy etycznym, lecz zjawiskiem, które podważa fundamenty zaufania do komunikacji audiowizualnej jako takiej. Erozja tego zaufania ma bezpośrednie konsekwencje dla funkcjonowania rynku, instytucji publicznych, procesów decyzyjnych oraz demokracji. Odpowiedzią na te wyzwania nie może być fragmentaryczna interwencja, lecz skoordynowany model łączący regulacje, edukację i technologię.

Podsumowując, analiza demokratyzacji technologii deepfake pozwala lepiej zrozumieć, dlaczego współczesne zagrożenia informacyjne mają charakter masowy, trudny do kontrolowania i wykraczający poza tradycyjne ramy regulacyjne. Deepfake nie jest jedynie kolejnym narzędziem manipulacji, lecz symptomem głębszej transformacji środowiska informacyjnego, w którym autentyczność przestaje być oczywista, a zaufanie wymaga aktywnej ochrony. W tym sensie wyzwania związane z deepfake'ami stanowią jeden z kluczowych testów zdolności systemów prawnych, organizacyjnych i technologicznych do adaptacji w erze generatywnej sztucznej inteligencji.

Literatura

1. Afshari, N., Mohammadi, A. (2023). *The Legal Implications of Deepfake Technology: Privacy, Defamation, and the Challenge of Regulating Synthetic Media*, Legal Studies in Digital Age, t. 2, nr 2, s. 13-23; s. 16-19. Pobrano z: https://docslib.org/doc/12559428/the-state-of-deepfakes-landscape-threats-and-impact-henry-ajder-giorgio-patrini-francesco-cavalli-and-laurence-cullen-september-2019 (dostęp: 15.02.2026).
2. Ajder, H., Patrini, G., Cavalli, F., Cullen, L., (2019). *The State of Deepfakes: Landscape, Threats, and Impact*, Deeptrace Labs., Amsterdam. Pobrano z: regmedia.co.uk.; https://regmedia.co.uk/2019/10/08/deepfake_report.pdf (dostęp: 15.02.2026).
3. Barta, J., Markiewicz, R. (2023). *Prawo autorskie*, Wolters Kluwer, Warszawa.
4. Chesney, R., Citron, D. (2019). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, California Law Review, t. 107, nr 6, s. 1753-1820. Publisher University of California Berkeley School of Law. Pobrano z: https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security (dostęp: 10.03.2026; Pobrano z: https://scholarship.law.bu.edu/faculty_scholarship/640/ (dostęp: 15.02.2026).
5. European Commission. (2020). *European Commission – Joint Research Centre (JRC) Artificial Intelligence and Disinformation: Assessment of Risks and Countermeasures (2020).* Pobrano z: https://publications.jrc.ec.europa.eu/repository/handle/JRC122150, (dostęp: 15.02.2026).
6. European Copyright Society. (2023). *Opinion on AI Training and Copyright*, Copyright and Generative AI: Opinion of the European Copyright Society (online) (dostęp: 15.02.2026). dokument został przyjęty merytorycznie w 2023 r., natomiast oficjalnie opublikowany w 2025 r. przez European Copyright Society.
7. European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*. (przyjęta: 17 grudnia 2024 r., opublikowana: 18 grudnia 2024 r.). Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models | European Data Protection Board (online), (dostęp: 15.02.2026).

8. European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models* Adopted on 17 December 2024. Pobrano z: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf (dostęp: 15.02.2026); EDPB (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models. 18 December 2024.* Pobrano z: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en (dostęp: 15.02.2026).
9. EUR-Lex. (b.d.). Document 32024R1689; Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji). Pobrano z: https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:32024R1689 (dostęp: 15.02.2026); https://sip.lex.pl/akty-prawne/dzienniki-UE/rozporzadzenie-2024-1689-w-sprawie-ustanowienia-zharmonizowanych-przepisow-72359064 (dostęp: 15.02.2026); ELI: http://data.europa.eu/eli/reg/2024/1689/oj (dostęp: 15.02.2026); https://eur-lex.europa.eu/eli/reg/2024/1689/oj (dostęp: 15.02.2026).
10. EUR-Lex. (b.d.). Document 32016R0679, Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych oraz w sprawie swobodnego przepływu takich danych (ogólne rozporządzenie o ochronie danych) – Dz.U. UE L 119 z 4 maja 2016, s. 1-88. Pobrano z: https://eur-lex.europa.eu/eli/reg/2016/679/oj (dostęp: 15.02.2026).
11. EUR-Lex. (b.d.). Document 32024R1689; ELI: http://data.europa.eu/eli/reg/2024/1689/oj; Pobrano z: https://eur-lex.europa.eu/eli/reg/2024/1689/oj (dostęp: 15.02.2026).

12. European Parliament – STOA. (2021). *Tackling deepfakes in European policy*. Study, Scientific Foresight Unit (STOA), DirectorateGeneral for Parliamentary Research Services (EPRS), Brussels. Tackling deepfakes in European policy | Think Tank | European Parliament (online), (dostęp: 15.02.2026).
13. European Parliament Tackling deepfakes in European policy. (2021). Pobrano z: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_BRIE%282021%29690039%28ANN%29_EN.pdf (dostęp: 15.02.2026).
14. Etienne, H. (2021). The future of online trust (and why Deepfake is advancing it). W: J. MacIntyre, L. Medsker (red.), *AI and Ethics* (s. 553-562). Springer. https://doi.org/10.1007/s43681-021-00072-1
15. Future Advocacy. (2019). *Deepfakes*, Deepfakes - Future Advocacy, Pobrano z: https://futureadvocacy.com/case-studies/deepfakes/ (dostęp: 15.02.2026).
16. Galindo, G. (2020). *XR Belgium posts deepfake of Belgian premier linking Covid19 with climate crisis*. The Brussels Times. Pobrano z: https://www.brusselstimes.com/106320/xr-belgium-posts-deepfake-of-belgian-premier-linking-covid-19-with-climate-crisis (dostęp: 15.02.2026).
17. Gervais, D. (2019), *The Machine as Author*, Iowa Law Review, t. 105.
18. Gil, Ł. (2026). *Socjotechnika 2.0 – Nowe trendy w atakach na czynnik ludzki*, Pobrano z: https://nflo.pl/baza-wiedzy/socjotechnika-nowe-trendy-ataki/ (dostęp: 15.02.2026).
19. Harvey Kong. (2024). Hong Kong Police / South China Morning Post (2024), *'Everyone looked real': multinational firm's Hong Kong office loses HK$200 million after scammers stage deepfake video meeting*. South China Morning Post „ '*Everyone looked real': multinational firm's Hong Kong office loses HK$200 million after scammers stage deepfake video meeting",* 4 February 2024. Pobrano z: https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage (dostęp: 15.02.2026).
20. Integral Solution (b.d.). *Demokratyzacja Generatywnej Sztucznej Inteligencji (GAI),* https://integralsolutions.pl/demokratyzacja-generatywnej-ai-gai/ (dostęp: 15.02.2026).

21. Lumen, C. (2025). *Deepfakes and the Limits of Law: A Comparative Analysis of Regulatory Frameworks in the U.S., EU, and China*.
22. Levitt, A. (2024). *Navigating Privacy in the Age of AI: Evaluating the Adequacy of the GDPR and CCPA to Combat Deepfake Technology*, NHSJS.
23. NASK. (2025). *Deepfaki – prawdziwy problem z fałszywą rzeczywistością*. Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy. Pobrano z: https://www.nask.pl/magazyn/deepfaki-prawdziwy-problem-z-falszywa-rzeczywistoscia (dostęp: 15.02.2026).
24. Regan, G. (2025). A Brief History of Deepfakes — Reality Defender. Pobrano z: https://www.realitydefender.com/insights/history-of-deepfakes (dostęp: 15.02.2026).
25. Reuters. (2025). *Scammer defrauds Italian businessman of €1m using AI voice*. Pobrano z: https://www.reuters.com/world/europe/scammer-defrauds-italian-businessman-1m-using-ai-voice-2025-02-13/ (dostęp: 15.02.2026).
26. Rmero Moreno, F. (2024). *Generative AI and deepfakes: a human rights approach to tackling harmful content*, International Review of Law, Computers & Technology.
27. European Commission. (2020). *Artificial Intelligence and Disinformation (White Paper on AI)*
28. Samuelson, P. (2017). *Allocating Ownership Rights in ComputerGenerated Works*, University of Pittsburgh Law Review.
29. Sensity AI. (2019-2024). *Deepfakes and Identity Fraud*.
30. Suwajanakorn, S., Seitz, S. M., Kemelmacher-Shlizerman, I. (2017). *Synthesizing Obama: Learning Lip Sync from Audio*, W: Proceedings of the ACM Transactions on Graphics (TOG), t. 36, nr 4, 2017, art. 95. Publisher: Association for Computing Machinery, New York United States.
31. Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M., (2016). *Face2Face: RealTime Face Capture and Reenactment of RGB Videos*, W: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (s. 2387-2395). Publisher: IEEE Institute of Electrical and Electronics Engineers. Pobrano z: https://ieeexplore.ieee.org/document/7780631 (dostęp: 10.03.2026).

32. The Guardian. (2024). „*UK engineering firm Arup falls victim to £20m deepfake scam”.* Pobrano z: https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video (dostęp: 15.02.2026).
33. Tookitaki. (2025). *The CEO Wasn't Real: Inside Singapore's $499K. Deepfake Video Scam*. Pobrano z: https://www.tookitaki.com/blog/deepfake-ceo-scam-singapore-2025.
34. Toews, R. (2022). *Deepfakes Are Going to Wreak Havoc on Society. We Are Not Prepared,* Forbes. Pobrano z: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/ (dostęp: 15.02.2026).
35. Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Dz. U. z 2025 r. poz. 24. Pobrano z: https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU19940240083/U/D19940083Lj.pdf (dostęp: 15.02.2026).
36. WEF. (2024). *Global Risks Report.* Pobrano z: https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf?_gl=1*jr4w0n*_up*MQ..*_gs*MQ..&gclid=EAIaIQobChMItuLJqrqXkwMVFBiiAx2mbTr3EAAYAiAAEgJXU_D_BwE&gbraid=0AAAAAoVy5F7Kuvrf-kac-tATd1bVSjEwgp (dostęp: 15.02.2026).
37. WIPO. (2019). *Artificial Intelligence and Intellectual Property*, World Intellectual Property Organization Geneva, Switzerland. Pobrano z: https://www.wipo.int/en/web/technology-trends/artificial_intelligence/story; https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf (dostęp: 11.03.2026).
38. ZDNet. (2019). „Forget email: scammers use CEO voice 'deepfakes' to con workers into wiring cash". Pobrano z: https://www.zdnet.com/article/forget-email-scammers-use-ceo-voice-deepfakes-to-con-workers-into-wiring-cash/ (dostęp: 15.02.2026); *Wall Street Journal* ((cyt. za ZDNet).
39. Zegarow, P. (2025). *Deepfaki – prawdziwy problem z fałszywą rzeczywistością,* NASK. Pobrano z: https://www.nask.pl/magazyn/deepfaki-prawdziwy-problem-z-falszywa-rzeczywistoscia (dostęp: 15.02.2026).
40. Zontek, K. (2025). *Deepfake – zagrożenia i środki ochrony,* AURACO. Pobrano z: https://auraco.pl/blog/deepfake-zagrozenia-i-srodki-ochrony/ (dostęp: 15.02.2026).