Arkadiusz Popławski, Kamil Pawlak

Deepfake- współczesne zagrożenie dla polityki i mediów.

Czym jest Deepfake?

Deepfake to zaawansowana technologia oparta na złożonych algorytmach sztucznej inteligencji. Ma na celu manipulację rzeczywistością, prezentując osoby czy też wydarzenia w sposób nieprawdziwy, ale jak najbardziej realistyczny. Technologia ta umożliwia użytkownikowi tworzenie nie tylko autentycznie wyglądających, ale także fałszywych obrazów, filmów lub dźwięków, co komplikuje rozróżnienie, czy materiał jest autentyczny. Nazwa pochodzi od połączenia dwóch terminów: *deep learning*[1] (zaawansowana technika uczenia maszynowego, która wykorzystuje głębokie sieci neuronowe do analizy, syntezy i manipulowania danymi multimedialnymi) oraz *fake* (z ang. „fałsz").

Technika deepfake wykorzystuje algorytmy uczenia maszynowego oraz sieci neuronowe do analizy cech charakterystycznych, ruchu, mimiki, gestów, tonu głosu i sposobu mówienia. Dzięki temu program jest w stanie generować coraz dokładniejsze dane wyjściowe, lepiej odwzorowujące oryginał. Im bardziej wytrenujemy algorytmy, tym lepszy uzyskamy efekt. Aby stworzyć realistycznego deepfake'a potrzeba wielu prób i błędów, aby doszło do wytrenowania algorytmu i uzyskania pożądanego efektu.

Jak powstaje Deepfake?

Proces tworzenia deepfake'ów zazwyczaj obejmuje dwa główne komponenty: generator i dyskryminator. Są one częścią algorytmu zwanego Generative Adversarial Networks (GAN)[2]. Generator wyszkolony na wielu przykładach tworzy fałszywą treść imitującą oryginał. Z kolei dyskryminator stara się rozróżnić obrazy fałszywe od prawdziwych, ucząc się na danych rzeczywistych i wygenerowanych. W miarę postępu szkolenia oba komponenty są kolejno ulepszane, co skutkuje bardziej realistycznymi deepfake. Technologia może być jednak wykorzystywana do celów oszustw i dezinformacji, dlatego ważne jest opracowanie skutecznych sposobów wykrywania i powstrzymywania jej rozprzestrzeniania się.

---

[1] Taweh Beysolow II, *Introduction to deep learning using R*, San Francisco 2017, s. 1-2.

[2] *Generative Adversarial Networks GAN,* BFirst.Tech, 20.06.2019, https://bfirst.tech/generative-adversarial-networks-gan[29.05.2024].

Skąd się wziął Deepfake?

Termin deepfake powstał w 2017 roku. W tym czasie użytkownicy Reddit o pseudonimie „deepfakes" zaczęli udostępniać filmy przygotowane przy użyciu zaawansowanych technik głębokiego uczenia się, aby nakładać twarze gwiazd na ciała aktorów[3].

Deepfake może być wykorzystywany zarówno do celów rozrywkowych, jak i marketingowych, na przykład w kampaniach reklamowych. Stanowi to świetną opcję do kreatywnego tworzenia nowych treści. Niestety, stwarza również poważne zagrożenia związane z manipulacją wizerunkiem i pozyskiwaniem danych. Oszuści wykorzystujący oprogramowanie deepfake mogą wyłudzać dane, podając się za członka rodziny lub firmę, w której ofiara posiada daną usługę. Potencjalny poszkodowany nieświadomie może ulec manipulacji i doprowadzić do wycieku swoich poufnych danych. Innym poważnym zagrożeniem technologii deepfake jest jej zastosowanie w polityce. Może zostać wykorzystana do wywierania wpływów lub ośmieszania publicznych przedstawicieli.

Jakie zagrożenia niesie ze sobą Deepfake?

Deepfake może być wykorzystywany do manipulacji różnego rodzaju informacjami a także do szerzenia dezinformacji, co niestety stanowi potencjalne ryzyko dla społeczeństwa i demokracji. Przykładowo, deepfake może być używany do tworzenia fałszywych materiałów kompromitujących znane osoby publiczne lub polityków, co z kolei może powodować wpływanie na opinię publiczną, i tym samym szerzenie bądź powielanie fałszywego przekazu którego autorem nie jest ani dana osoba publiczna ani żaden z polityków, to jedno z poważniejszych zagrożeń jeśli chodzi o przekaz informacji publicznej za pomocą takich materiałów (zdjęć czy wideo)[4].

Jak można rozpoznać Deepfake?

Rozpoznanie materiału stworzonego technologią deepfake to naprawdę nie lada wyzwanie, ale istnieją pewne różnice odróżniające od autentycznego materiału, na które warto zwrócić uwagę, są to min:

---

[3] Mateusz Mróz, *Co to jest deepfake, jak go stworzyć i czy stanowi niebezpieczeństwo*, BOTLAND.com.pl, 15.07.2023, https://botland.com.pl/blog/co-to-jest-deepfake-jak-go-stworzyc-i-czy-stanowi-niebezpieczenstwo/ [dostęp: 29.05.2024].

[4] *Deepfake – zagrożenie dla rzeczywistości* https://antywirus.com/deepfake-zagrozenie-dla-rzeczywistosci/[dostęp: 29.05.2024]

- Niespójna mimika (Odbiegająca znacząco od naturalnej)
- Nienaturalne zachowanie (Bowiem deepfake może wydawać się nienaturalne w zachowaniu, zwłaszcza jeśli przedstawia osobę, która wykonuje nietypowe czynności oraz ruchy)
- Podejrzane źródło (Zawsze warto sprawdzić źródło treści i zweryfikować, czy jest to zaufane i autentyczne źródło informacji, na ogół źródło z którego pochodzi deepfake takim właśnie nie jest)
- Manipulacje dźwięku (Np.: głos różniący się nieco od autentycznego głosu danej postaci)

O czym warto pamiętać?

Przede wszystkim o tym że deepfake jest nowym zagrożeniem w erze sztucznej inteligencji, które znacząco może wpływać na percepcję faktów i rzeczywistości, co z kolei może prowadzić do poważnych konsekwencji dla społeczeństwa i szerokiej opinii publicznej. Jednym z głównych skutków takiej manipulacji jest spadek zaufania do mediów, albowiem kiedy deepfake staje się coraz bardziej powszechny, ludzie mogą zacząć wątpić w autentyczność informacji, które widzą i słyszą w mediach, a to z kolei będzie wpływało na obniżony poziom zaufania do takiej informacji i niechętnie sięganie po nią. Ponadto za pomocą takich spreparowanych materiałów można wpływać na relacje międzyludzkie, kontakty czy nawet stosunki dyplomatyczne, co może ze sobą nieść zagrożenia w świecie polityki i szeroko pojmowanej dyplomacji międzynarodowej. Warto również pamiętać że w środowiskach hejterskich, czy na „farmach trolli'' technologii deepfake używa się również do manipulowania i tworzenia konfliktów oraz podsycania nienawiści[5].

Dlatego podsumowując, trzeba szczególnie pamiętać o mądrym korzystaniu z treści (zdjęć i wideo) publikowanych przez osoby publiczne oraz polityków, dokładne weryfikowanie źródeł, porównywanie materiałów z materiałami z wcześniejszych lat, a także korzystaniu z wielu innych narzędzi które mają za zadanie przeciwdziałać deepfake'om.

[5] *Czym jest Deepfake?* https://chronpesel.pl/wyludzenia-i-kradzieze/czym-jest-deepfake[dostęp 29.05.2024]

BIBLIOGRAFIA

Publikacje naukowe:

1. Beysolow TawehII, *Introduction to deep learning using R*, San Francisco 2017.
2. Foster David, *Generative deep learning teaching machines to paint write compose and play*, O'Reilly Media, 28.06.2022.
3. Generative Adversarial Networks GAN, BFirst.Tech, 20.06.2019.
4. Goodfellow Ian, Bengio Yoshua, Courville Aaron, *Deep Learning (Adaptive Computation and Machine Learning series)*, The MiT Press, 18.11.2016.

Źródła internetowe:

5. *Czym jest Deepfake?* https://chronpesel.pl/wyludzenia-i-kradzieze/czym-jest-deepfake[dostęp 29.05.2024]
6. Deepfake – zagrożenie dla rzeczywistości https://antywirus.com/deepfake-zagrozenie-dla-rzeczywistosci/[dostęp: 29.05.2024]
7. *Generative Adversarial Networks GAN,* BFirst.Tech, 20.06.2019 https://bfirst.tech/generative-adversarial-networks-gan[dostęp: 29.05.2024]
8. *Mateusz Mróz, Co to jest deepfake, jak go stworzyć i czy stanowi niebezpieczeństwo, BOTLAND.com.pl, 15.07.2023,* https://botland.com.pl/blog/co-to-jest-deepfake-jak-go-stworzyc-i-czy-stanowi-niebezpieczenstwo/ [dostęp: 29.05.2024].