Deepfake

Łukasz Migniewicz

Materiały źródłowe

Źródło A

Deepfake to coraz powszechniejsze zjawisko w Internecie, które wykorzystywane może być na całym świecie. [...] Słowo pochodzi ze zbitki dwóch zwrotów zaczerpniętych z języka angielskiego: *deep learning* (głębokie uczenie) oraz *fake* (fałszywka, podróbka). Jest to więc taka obróbka dźwięku i obrazu, która ma na celu utworzenie fałszywego wideo, wykorzystując obszerną technikę sztucznej inteligencji. Jej głównym założeniem jest stworzenie niemożliwych do odróżnienia od prawdziwych materiałów obrazowo-dźwiękowych, które zawierają rzeczywiste wypowiedzi, wystąpienia i dokonane czyny. *Deepfake* nie tworzy całkiem nowych filmów, jednak bazuje na materiale, który już pojawił się w sieci i jest odpowiednio przerabiany. Komputer wykrywa twarz, odnajduje na niej dziesiątki kluczowych punktów i daje możliwość połączenia ich w sieć, która umożliwia modyfikacje ich w ruchu. [...] pojawiło się wideo, w którym były prezydent Stanów Zjednoczonych, Barack Obama, nazywa obecnego – Donalda Trumpa – głupkiem. [...]

Deepfake w Internecie – co to jest?, 18.06.2020 r. [dostęp 10.12.2020]. Dostępne w Puffa.pl: http://www.puffa.pl/2020/06/18/deepfake-w-internecie-co-to-jest/.

Źródło B

W ostatnich latach technologia [...] stała się wszechobecna, umożliwiając ludziom na całym świecie otrzymywanie natychmiastowych zdjęć i filmów. Odzwierciedleniem tego [...] jest zdolność nawet stosunkowo niewykwalifikowanych użytkowników do manipulowania i zniekształcania przekazu mediów wizualnych. Podczas gdy wiele manipulacji jest wykonywanych dla zabawy lub dla wartości artystycznej, inne służą celom takim jak propaganda lub tworzenie kampanii dezinformacyjnych. Ta manipulacja multimediami wizualnymi jest możliwa dzięki szerokiej dostępności zaawanso-

wanych aplikacji do edycji obrazu i wideo, a także zautomatyzowanych algorytmów manipulacji, które umożliwiają edycję w sposób bardzo trudny do wykrycia wizualnie lub za pomocą aktualnej analizy obrazu i narzędzi do analizy wizualnej mediów. [...] Materiały wideo tego rodzaju są potencjalnym zagrożeniem dla bezpieczeństwa wewnętrznego każdego państwa, a także mogą stać się narzędziem wpływu na wyniki wyborów. Kolejny sfabrykowany wideoskandal może zagrozić bezpieczeństwu narodowemu lub wpłynąć na opinię publiczną, co stanowi pole działania dla oszustów chcących ingerować np. w nastroje polityczne w społeczeństwie, a także staje się nową bronią w wojnie informacyjnej. Technologia ta będzie naturalnym narzędziem wykorzystywanym przez państwa celem manipulowania opinią publiczną i przeprowadzania kampanii dezinformujących, a także podkopywania wiary w obecnie istniejące instytucje. Z wyłaniającymi się i przerażająco zaawansowanymi metodami śledzenia twarzy i wideo manipulacji, nadchodzi zatem nowa era dezinformacji.
W czasach, w których publiczne zaufanie do mediów i polityki jest już zagrożone, możliwość, że wszystko, co oglądamy w Internecie, może być przekonującą formą oszustwa wymyśloną przez jakąś osobę posiadającą nowoczesny komputer osobisty, może jeszcze bardziej zagrozić wierze w demokrację [...]

Olga Wasiuta, Sergiusz Wasiuta, *FakeApp jako nowe zagrożenie bezpieczeństwa politycznego i informacyjnego*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9(3), s. 136 [dostęp 10.12.2020]. Dostępne w Studia de Securitate: https://studiadesecuritate.up.krakow.pl/wp-content/uploads/sites/43/2019/10/9-1.pdf.

Źródło C

Korzystając ze sztucznie wygenerowanego głosu prezesa dużej firmy, przestępcy zadzwonili do dyrektora jednej z jej spółek zależnych i poprosili o wykonanie przelewu. Udało im się wyłudzić w ten sposób 243 tys. dolarów. [...] Według eksperta, przestępcy wykorzystali oprogramowanie sztucznie generujące głos. [...] przestępcy mogli zmontować fragmenty autentycznych nagrań głosu prezesa firmy. Wykorzystanie tej techniki zostało zaprezentowane m.in. na ubiegłorocznej konferencji Black Hat. Naukowcy pokazali wówczas, że już dostępne narzędzia dają możliwość zrekonstruowania głosu dowolnej osoby, by przełamać zabezpieczenie identyfikacji głosowej wykorzystywanej np. przez banki. Ich zdaniem dla takich zabezpieczeń jakość

dźwięku nie musi być idealna, by je oszukać. Udowodnili, że są w stanie takiego oszustwa dokonać w 10 minut.

Odtworzyli głos szefa firmy. Wyłudzili 243 tys. dolarów, 06.09.2019 r. [dostęp 10.12.2020]. Dostępne w Konkret24: https://konkret24.tvn24.pl/tech,116/odtworzyli-glos-szefa-firmy-wyludzili-243-tys-dolarow,967072.html.

Źródło D

Lądowanie Apollo 11 na Księżycu 20 lipca 1969 roku było przełomowym momentem w historii kosmosu. Jednak co, gdyby nie odniosło skutku? Nowy projekt MIT pokazuje siłę *deepfake*'u [...] – Los zarządził, że ludzie, którzy udali się na Księżyc w pokoju, pozostaną na Księżycu, aby spoczywać w pokoju – mówi Nixon w sfałszowanym filmie, odnoszącym się do astronautów Neila Armstronga, Buzza Aldrina i Michaela Collinsa [...]. Ekspertom MIT przygotowanie przekonującego 7-minutowego nagrania zajęło około pół roku. Połączono w nim nagrania NASA z fałszywym wystąpieniem Nixona. Wykorzystano technologie „deep-learning" sztucznej inteligencji, by głos i ruchy prezydenta były najbliższe rzeczywistym. To wideo jest pierwszym w ramach projektu Event of Moon Disaster, którego celem jest pokazanie ludziom, jak niebezpieczny wpływ mogą mieć zmanipulowane materiały wideo. [...]

Klaudia Stawska, *Lądowanie na Księżycu to kłamstwo? Wystarczy posłuchać prezydenta USA*, 21.07.2020 r. [dostęp 10.12.2020]. Dostępne w Tech.wp.pl: https://tech.wp.pl/ladowanie-na-ksiezycu-to-klamstwo-wystarczy-posluchac-prezydenta-usa-6534416849188481a.

[...] Projekt In Event of Moon Disaster zakładał jednak stworzenie alternatywnego przebiegu wydarzeń. Naukowcy wykorzystali w tym celu nie tylko oryginalne nagranie z Nixonem, ale również materiały archiwalne z tamtych wydarzeń. Urzeczywistnieniem fikcji zajęła się później sztuczna inteligencja. Ten DeepFake to prawdziwa perełka. Zachwyca i jednocześnie przeraża, jak bardzo zaawansowane są już algorytmy DeepFake. Inżynierowie z MIT nie ukrywają, że projekt ma na celu ostrzeżenie, w jakim kierunku może rozwinąć się ta technologia i jak niewyobrażalne spustoszenie może siać w mediach internetowych. Historycy najbardziej obawiają się sytuacji, w której fałszywe materiały filmowe, dotyczące kluczowych wydarzeń historycznych lub aktualnie trwających, będą publikowane np. na YouTube, gdzie zmanipulują

opinię publiczną, co doprowadzi później do niepokojów społecznych.

Sfałszowana katastrofa misji Apollo 11. Zobaczcie ten niesamowity DeepFake, 22.07.2020 r. [dostęp 10.12.2020]. Dostępne w Geekweek.pl: https://www.geekweek.pl/news/2020-07-22/sfalszowana-katastrofa-misji-apollo-11-zobaczcie-ten-niesamowity-deepfake-film/.

Źródło E

Technologia *deepfake* ma pozytywny potencjał edukacyjny. Dzięki interaktywności może zrewolucjonizować nasze lekcje historii. Pozwoli także przechować opowieści i pomoże przyciągnąć uwagę. Na przykład w 2018 r. Illinois Holocaust Museum and Education Centre skonstruowało hologramy – postaci udzielające wywiadów. Goście mogli więc rozmawiać z ocalałymi z Holokaustu i wchodzić w interakcje z nimi. Mogli zadawać pytania i wysłuchać ich historii. Wraz z postępem technologii *deepfake* tworzenie tego rodzaju wirtualnych historii może stać się osiągalne na znacznie szerszą skalę. [...] Technologia *deepfake* może pełnić funkcję CGI, pozwalając na odtworzenie postaci podobnych do już nieżyjących lub dawnych aktorów. Bohater więc nie musi umrzeć wraz z aktorem. Przykładem może być odtworzenie postaci wielkiego moffa Wilhuffa Tarkina, granej przez zmarłego w 1994 r. Petera Cushinga w filmie Gwiezdne Wojny: Łotr 1 z 2017 r. [...] Technologia SI może pomóc nam w tworzeniu wirtualnych muzeów. [...], a *deepfake* – umożliwić wskrzeszanie zmarłych artystów, na przykład Salvadora Dalego w poświęconym mu muzeum na Florydzie.

Yes, positive deepfake examples exist [pol.: Tak, pozytywne przykłady zastosowania *deepfake* istnieją] [dostęp 10.12.2020]. Dostępne w ThinkAutomation: https://www.thinkautomation.com/bots-and-ai/yes-positive-deepfake-examples-exist [tłum. Ł.M.].