Przepis na (nie)prawdziwego człowieka:

Deepfake i metody tworzenia fałszywego wizerunku

przewodnik dla nauczycieli

Michał Ołowski

Zakład Analiz Audiowizualnych i Systemów Biometrycznych (NASK)

Sztuczna inteligencja (SI)

Artificial Intelligence (AI)

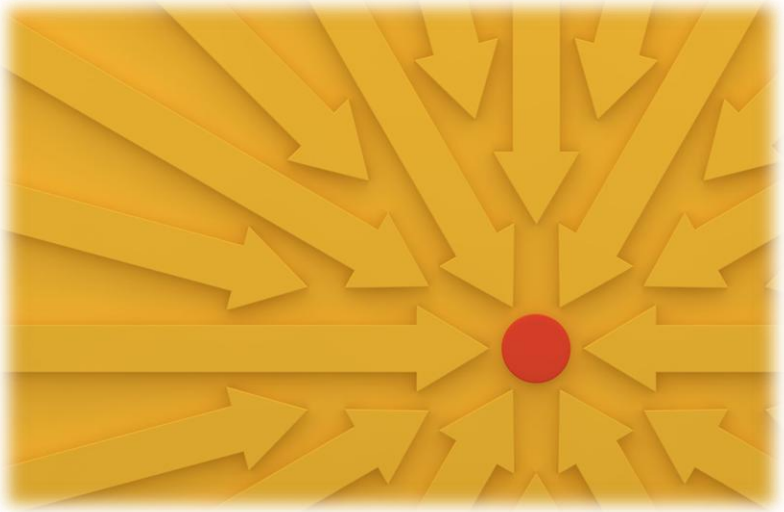
Do czego jest wykorzystywana?

- Rozpoznawanie obrazów
- Analiza danych
- Tłumaczenie języka
- Tworzenie treści

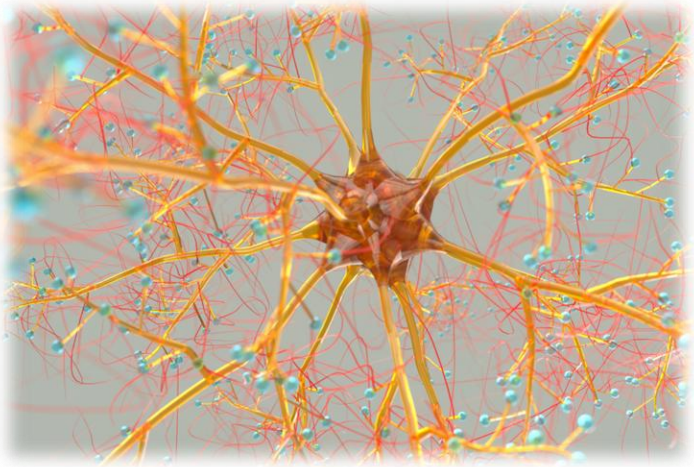
Czym jest?

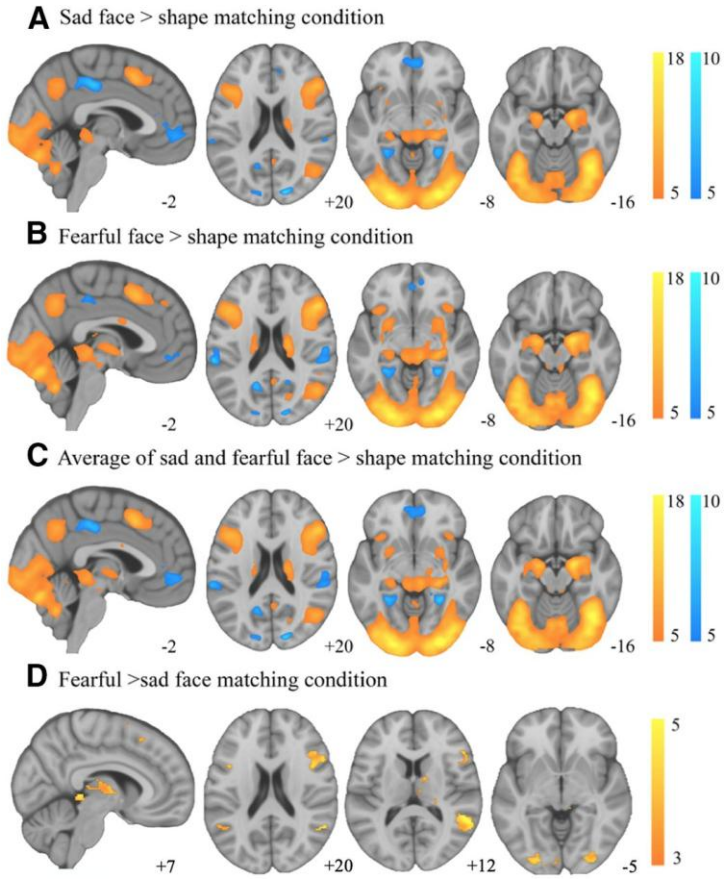
Rodzaj komputerowej technologii, która pomaga komputerom „myśleć” i działać jak ludzie w pewnych aspektach:

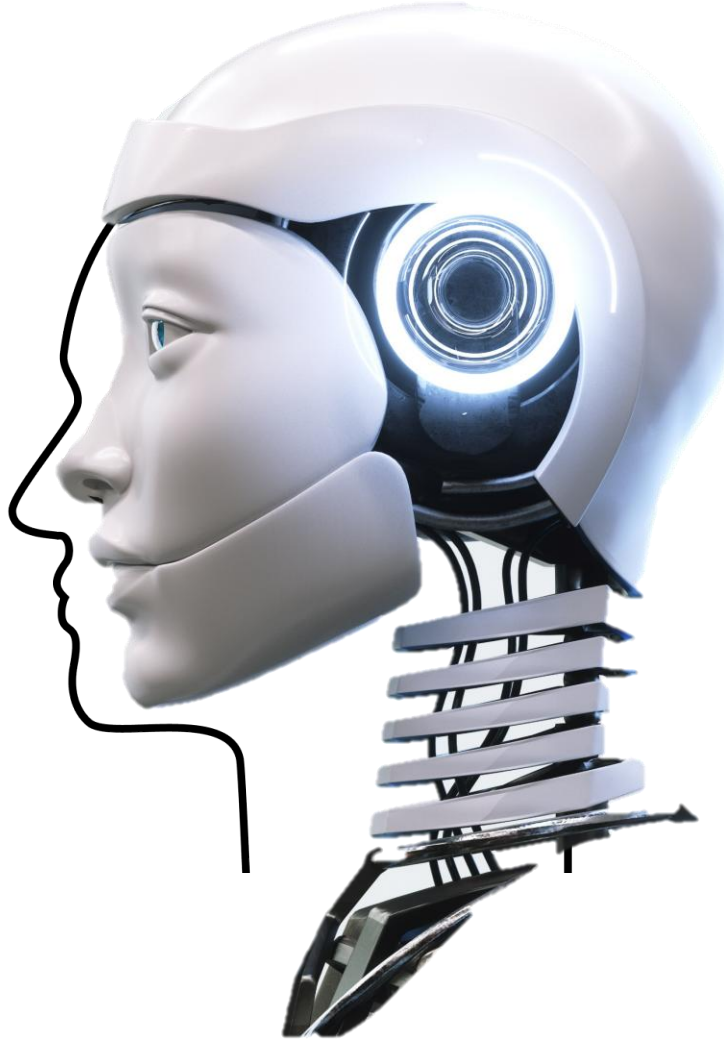
- interpretować informacje,
- podejmować decyzje,
- uczyć się z doświadczenia



Myślące maszyny




A Sad face > shape matching condition
-2
+20
-8
-16
18
10
5
5
B Fearful face > shape matching condition
-2
+20
-8
-16
18
10
5
5
C Average of sad and fearful face > shape matching condition
-2
+20
-8
-16
18
10
5
5
D Fearful >sad face matching condition
+7
+20
+12
-5
5
3



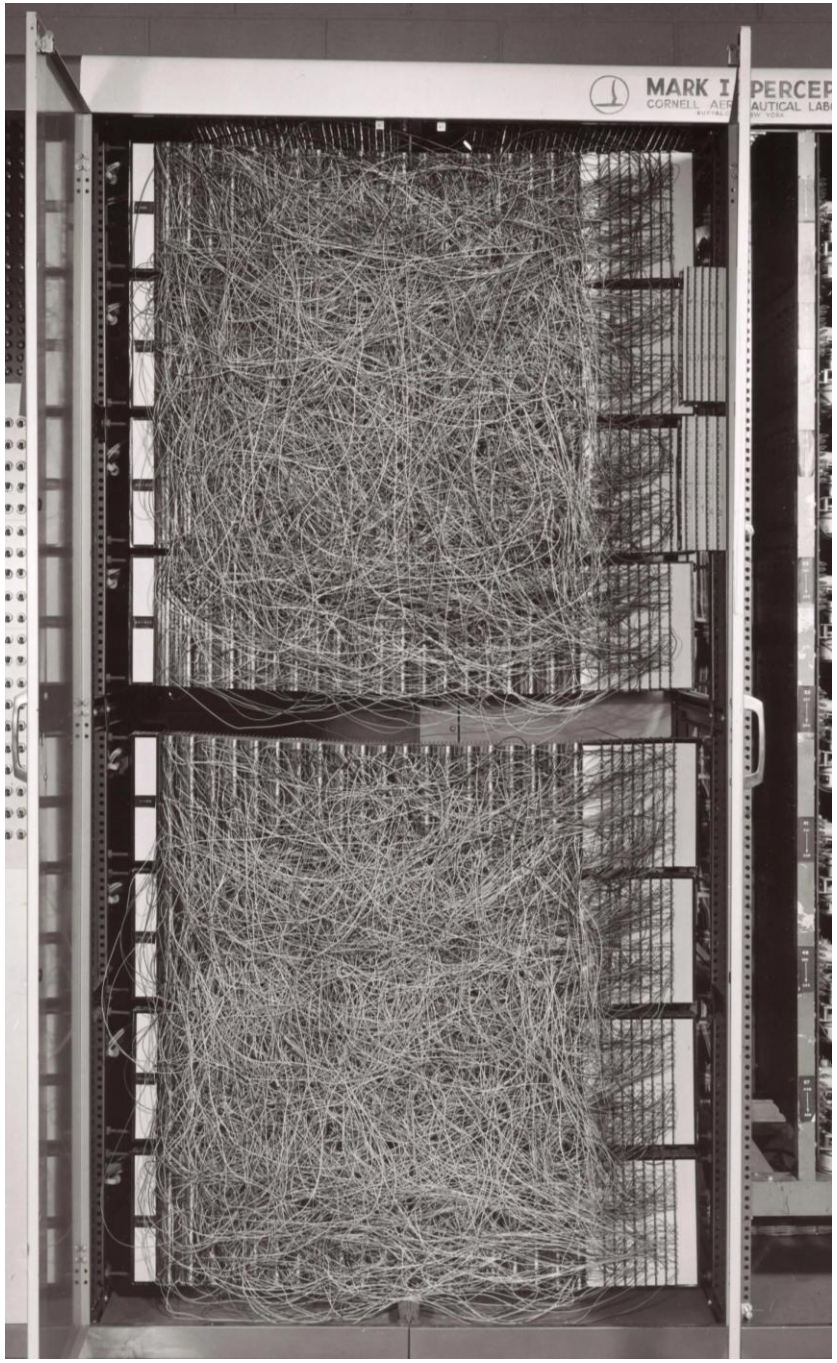
Pierwsza sztuczna sieć neuronowa

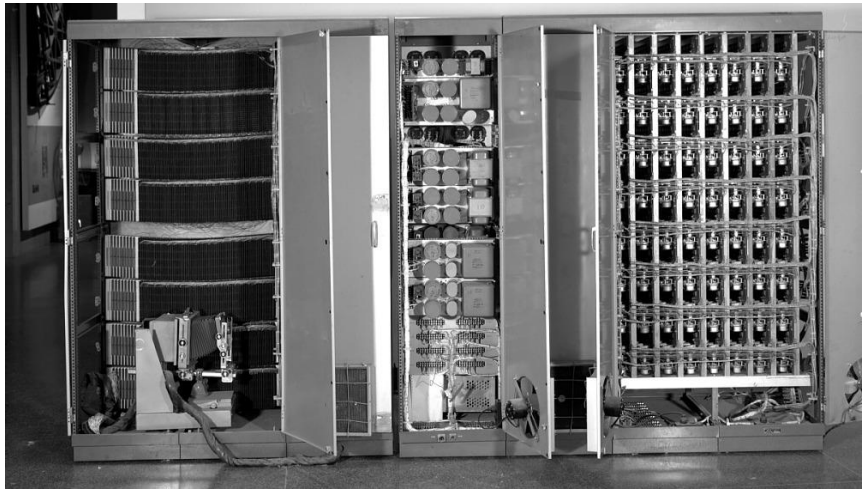
Pierwszy perceptron został zbudowany w **1958 roku** przez **Franka Rosenblatta.**

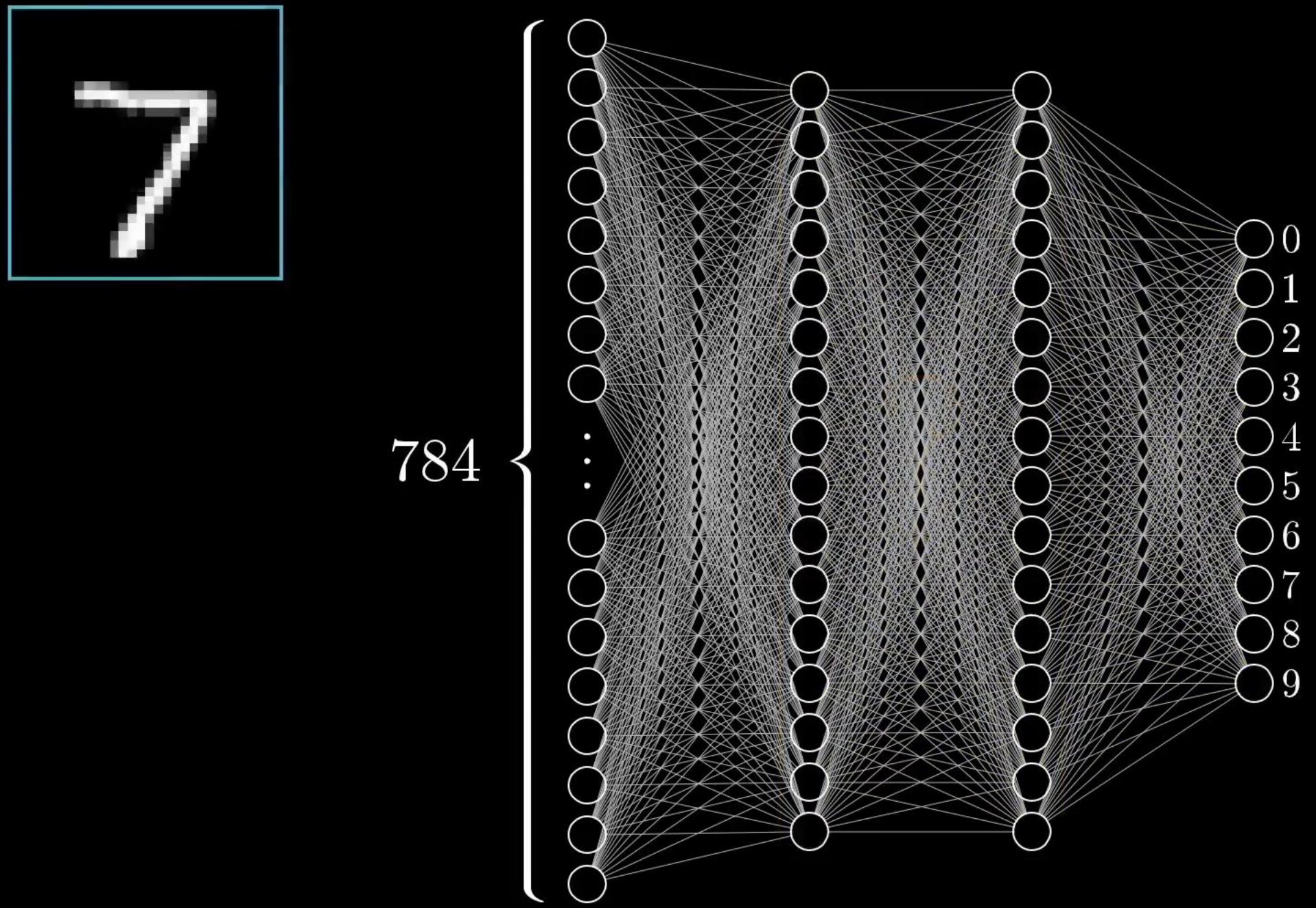
Maszyna ta została zaprojektowana do rozpoznawania obrazów: rozróżnianie zdjęć mężczyzn i kobiet.

Zarodek elektronicznego komputera, który będzie w stanie chodzić, mówić, widzieć, pisać, reprodukować się i być świadomym swojego istnienia.

The New York Times

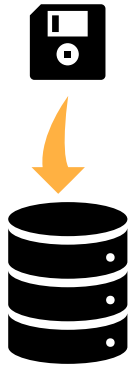




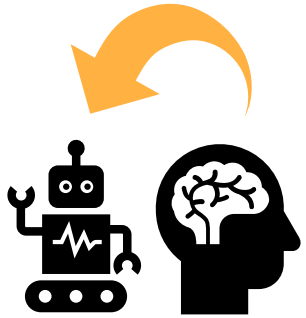


Grant Sanderson (3blue1brown)
But what is a neural network? | Chapter 1, Deep learning

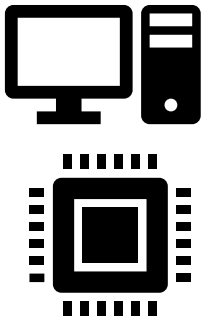
Era Sztucznej Inteligencji



Dostępność danych



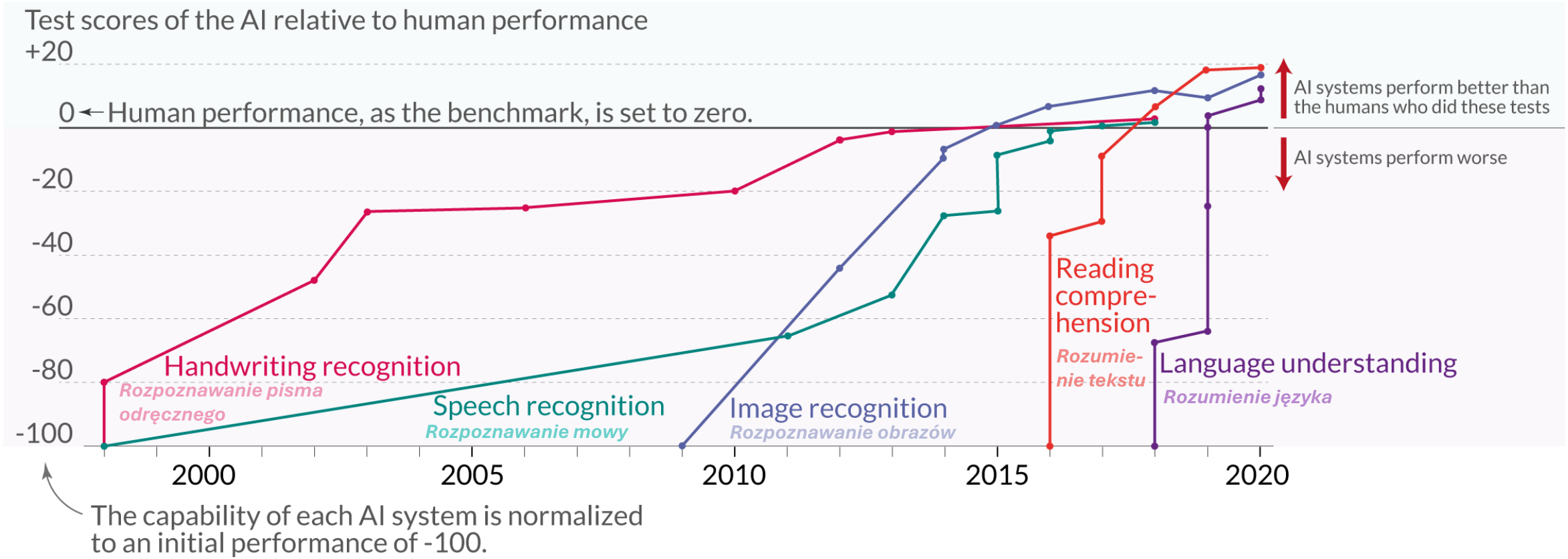
Rozwój algorytmów



Wzrost mocy obliczeniowej

Language and image recognition capabilities of AI systems have improved rapidly

Our World in Data



Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP

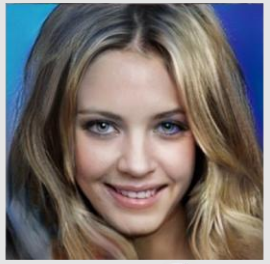
OurWorldinData.org – Research and data to make progress against the world's largest problems.

2014

2015

2016

2017



2018

2019

2014

arXiv:1406.2661

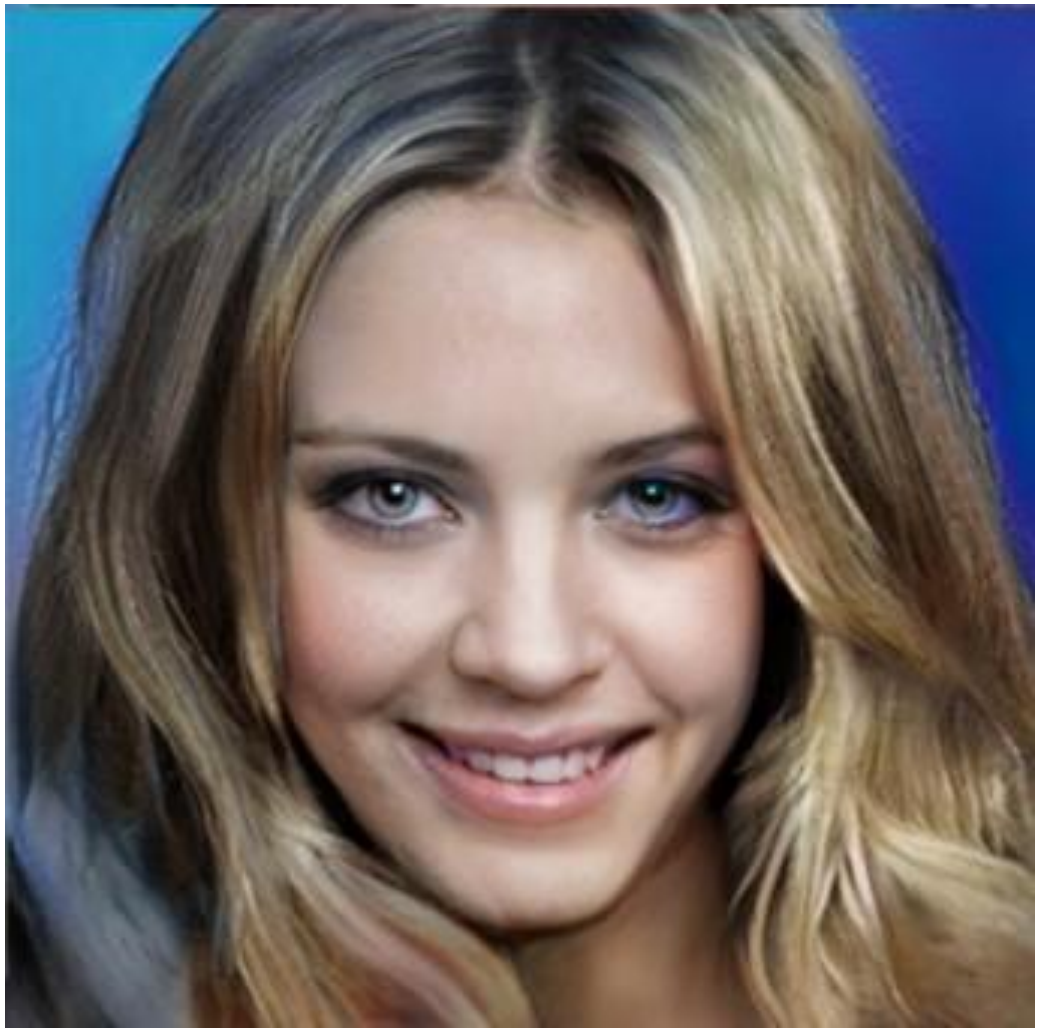
2015

arXiv:1511.06434

2016

arXiv:1606.07536

2017



arXiv:1710.10196

2018

arXiv:1812.04948

2019

arXiv:1912.04958

2020

arXiv:2006.11239

2021
StyleGAN v3

2022
Midjourney

Duże wnętrze autorstwa Kengo Kumy, harmonijne połączenie naturalnych elementów i nowoczesnego designu, ekologiczna struktura, baseny i spadająca woda.

Zdjęcie vintage, dziewczyna paląca papierosa, mgliste wspomnienie.

Kobieta rycerz, pagórkowate równiny, całe ciało, ciemny lazur, obrazy w stylu wiktoriańskim, pogodna twarz, realistyczne przedstawienie światła, złote światło.

Midjourney, Jason M. Allen
Théâtre d'Opéra Spatial

2023


MIDJOURNEY v5


ChatGPT


Eleven
Labs


stability.ai


Bing AI

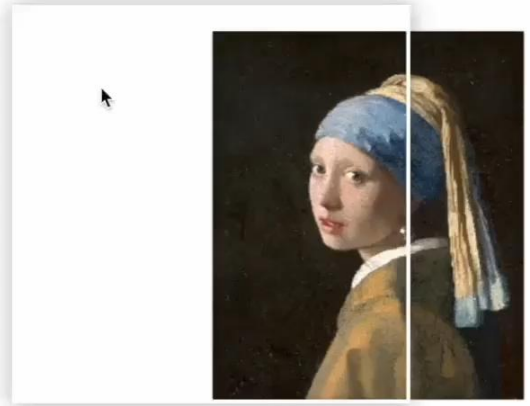

Whisper


PaLM 2


LLaMA
by Meta


DALL·E 2

Johannes Vermeer
Dziewczyna z perłą



August Kamp × DALL-E

Cel: Człowiek

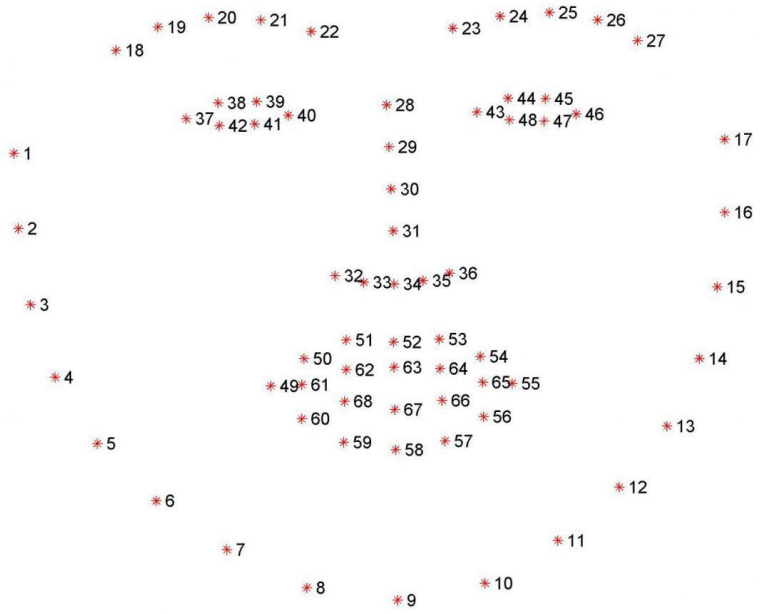
Od rozpoznawania twarzy do sterowania jej elementami.

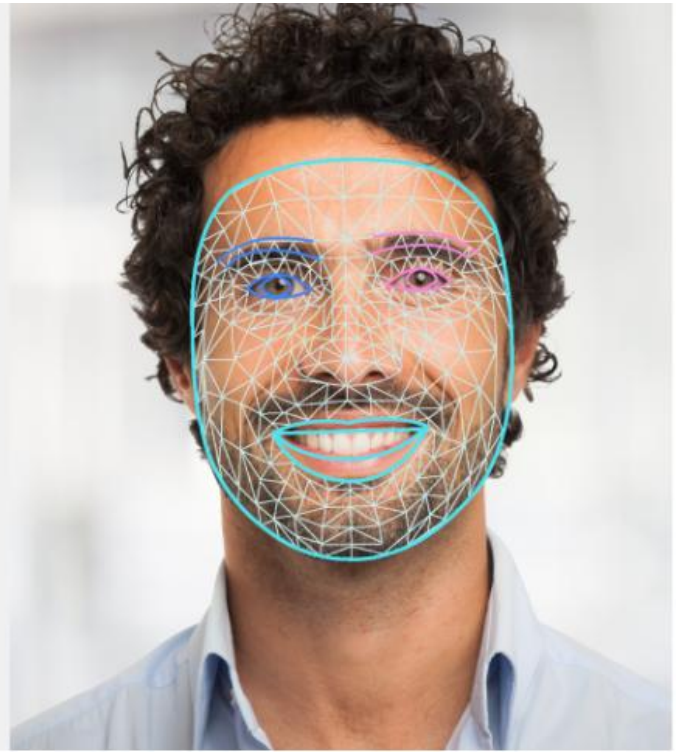
Jak maszyna widzi twarz?

Punkty orientacyjne twarzy (*Face landmarks)* to punkty dzięki którym maszyna może rozpoznać twarz na zdjęciu.

Używane do identyfikowania i opisywania jej cezur.

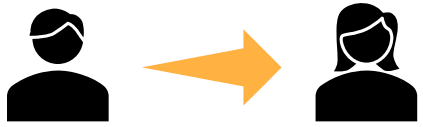
Punkty obejmują takie elementy jak oczy, nos, usta, brwi i inne ważne elementy twarzy.



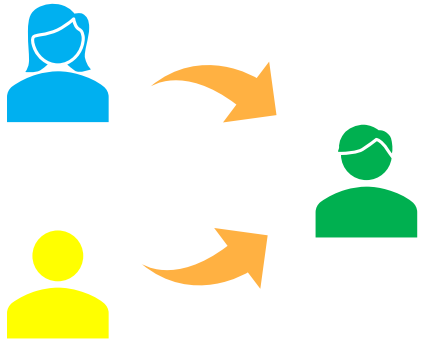


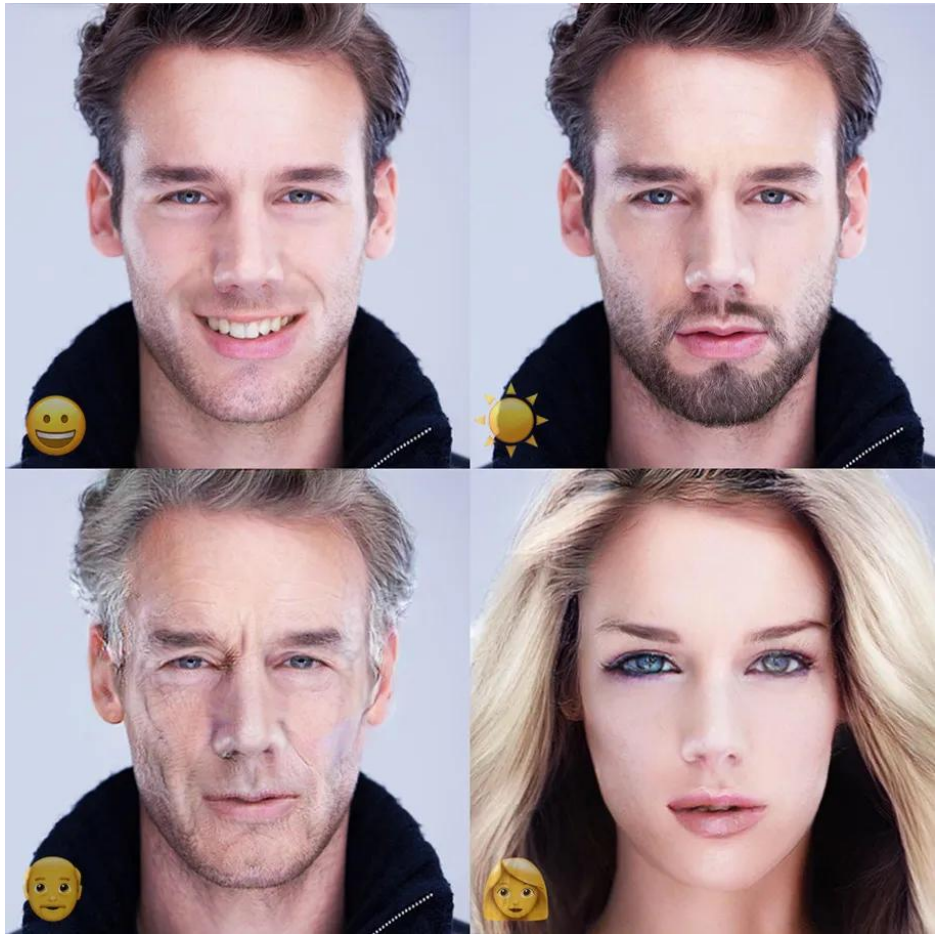
Modyfikacja twarzy

- Zmiana atrybutów



- Miksowanie twarzy

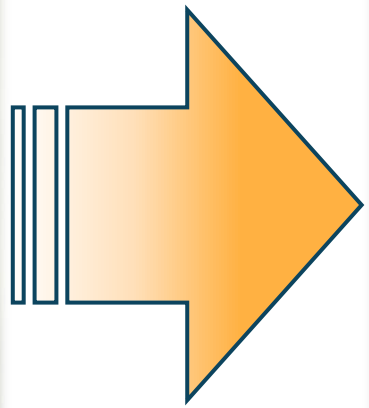




Pranoy Radhakrishnan https://t.ly/nAPqf

Deepfake

Podmień twarz, zmień tożsamość




birbfakes/https://github.com/deepfakes/faceswap

https://youtu.be/fwqjtQ1XWUs?feature=shared

Reenactment Pipeline

Input Source

Input Target

Tracking Source

Tracking Target

Expr. Transfer

Source Actor

Target Actor

Real-time Reenactment

Reenactment Result


AE
Ah
BMP
ChJ
EE
Er
FV
Ih
KGHNG
Oh
R
SZ
TLDN
Th
WOO

Original Video for Input Speech

Our Result

https://grail.cs.washington.edu/projects/AudioToObama/

Neural Voice Puppetry

arXiv:1912.05566

Daj głos...

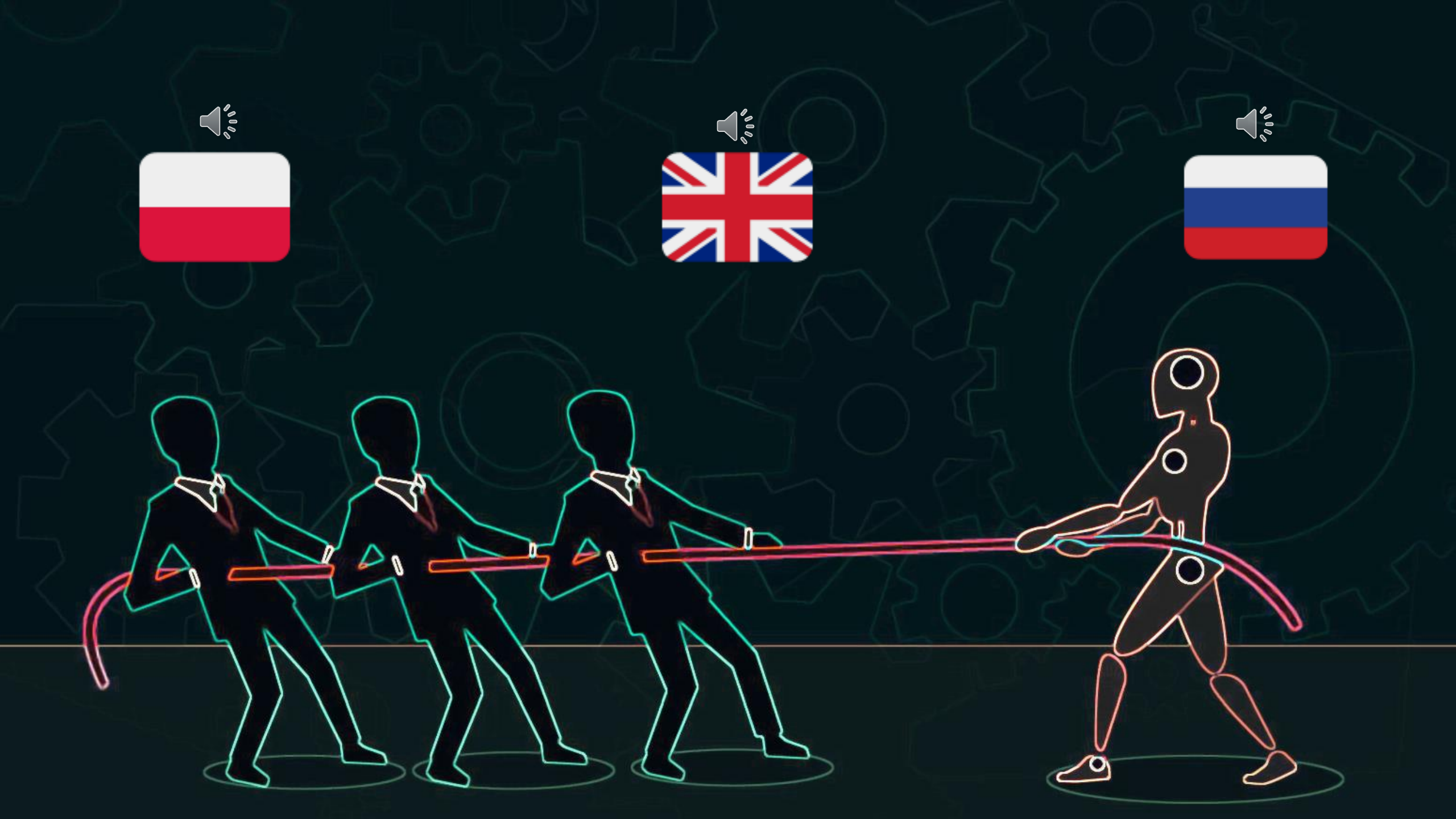
...a komputer zacznie nim mówić

The avocado is a pear-shaped fruit with leathery skin, smooth edible flesh and a large stone.

Postęp w generowaniu ludzkiej mowy:

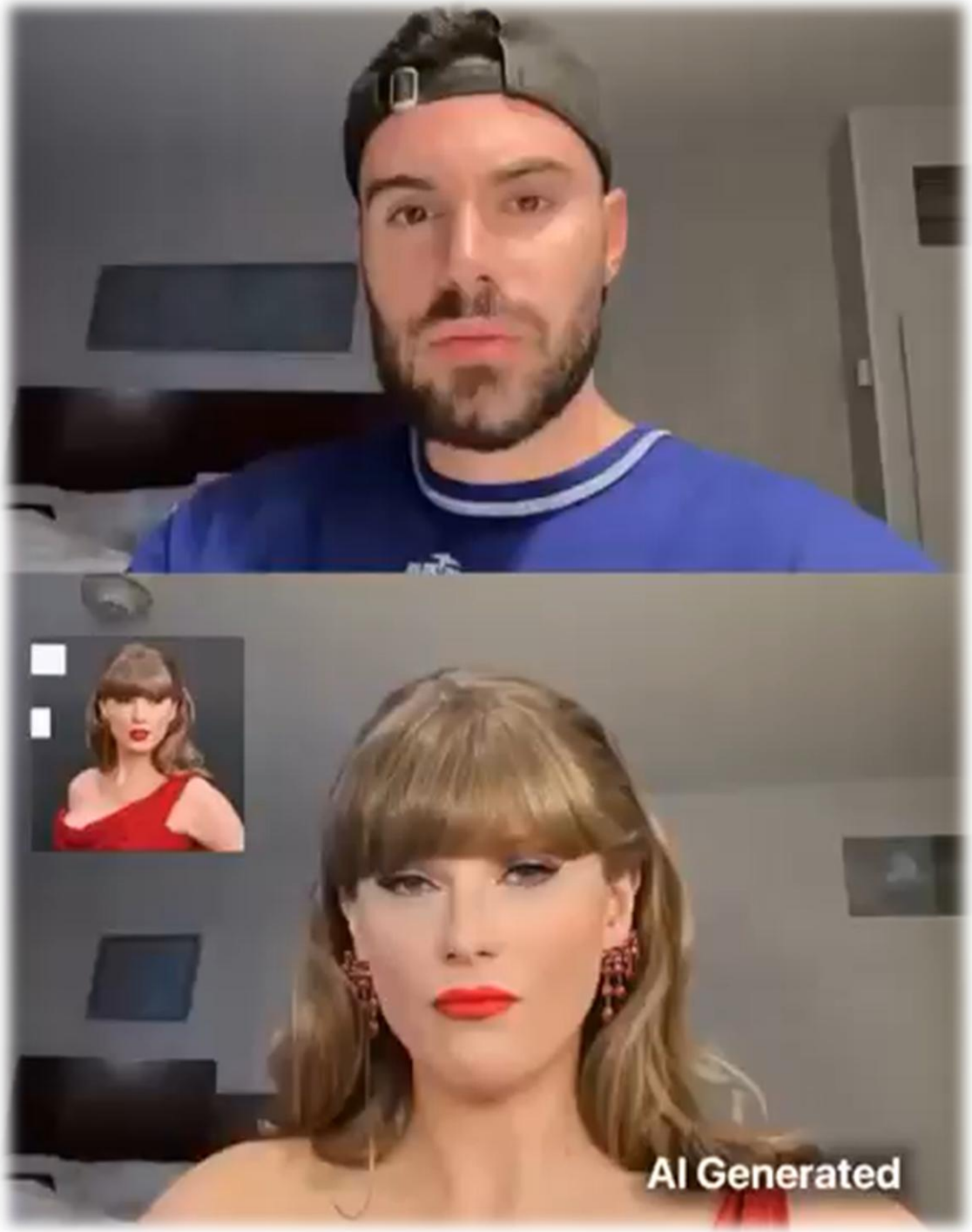
The Blue Lagoon is a 1980 American romance and adventure film directed by Randall Cleiser.

Źródło: deepmind.com/blog/wavenet-a-generative-model-for-raw-audio



Informacyjna bomba atomowa

w rękach każdego z nas



Courts

'Deepfake' audio evidence used in UK court to discredit Dubai dad

Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine

Published: October 26, 2023 6.38pm CEST

'$35 million gone in one call': Deepfake fraud rings are fooling the world's smartest firms

Deepfake technology originally captured public attention as a source of amusement, with celebrity face-swapping applications. However, what started as a curiosity quickly became a huge cyber threat.

Shamsvi Balooni Khan

Updated Mar 27, 2025 8:34 PM IST

Local News

Deepfakes of Elon Musk are contributing to billions of dollars in fraud losses in the U.S.

By Brian New, Lexi Salazar, Mike Lozano, Scott Fralicks

Updated on: November 24, 2024 / 2:28 PM CST / CBS Texas

POLITICS

New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary

Giorgia Meloni: Italian PM seeks damages over deepfake porn videos

20 March 2024

Share Save

Laura Gozzi

BBC News

The deepfake videos featuring Giorgia Meloni date back to 2020 before she became Italy's PM

Ataki są powszechne, kosztowne i nasilają się

85%

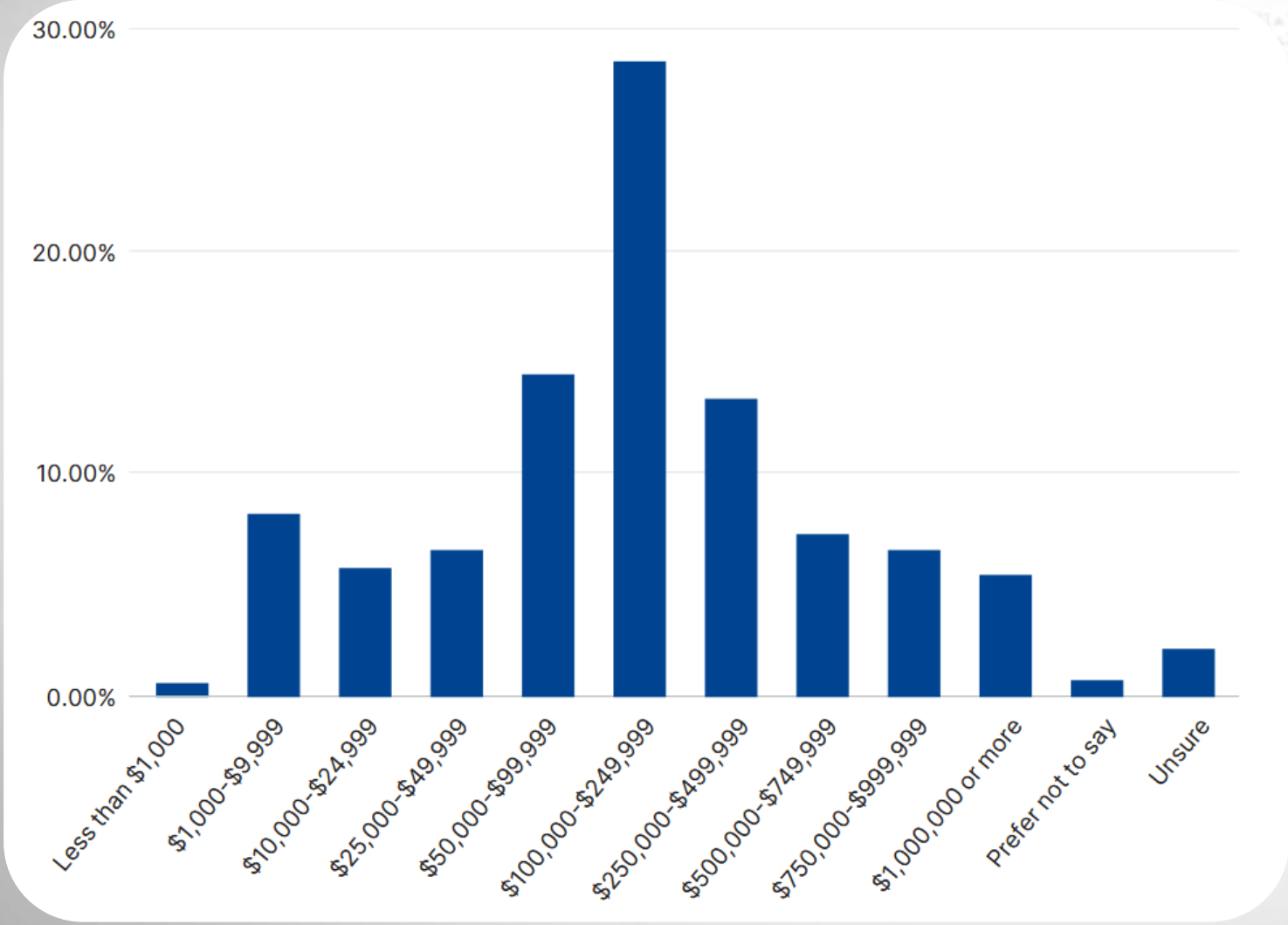
respondentów zgłasza, że w ciągu ostatnich **12 miesięcy** doświadczyło co najmniej jednego incydentu związanego z deepfake’ami.

>40%

doświadczyło trzech lub więcej ataków.

Na podstawie: https://ironscales.com/fall-2025-threat-report

Ataki są powszechne, nasilają się i są kosztowne

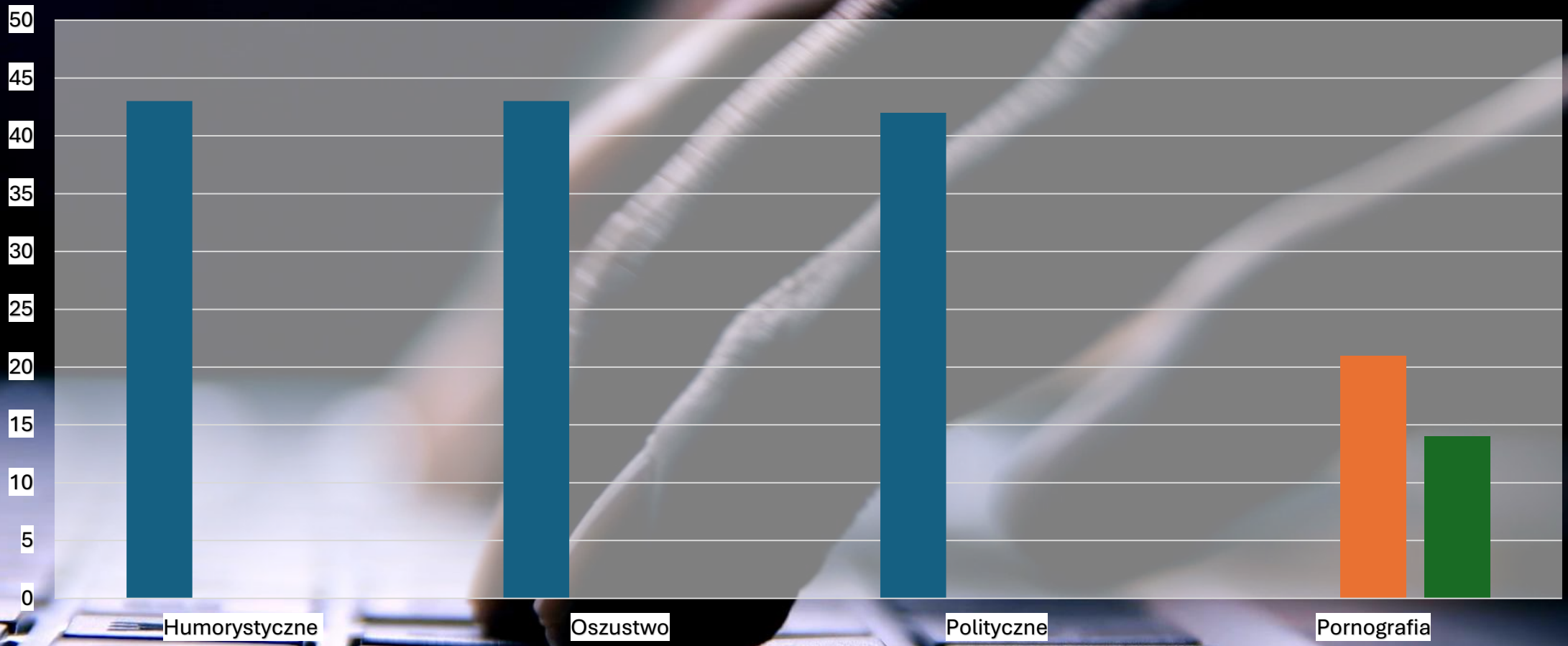


61%

zaatakowanych organizacji straciło ponad **100 000 $** za sprawą incydentów z wykorzystaniem deepfake.

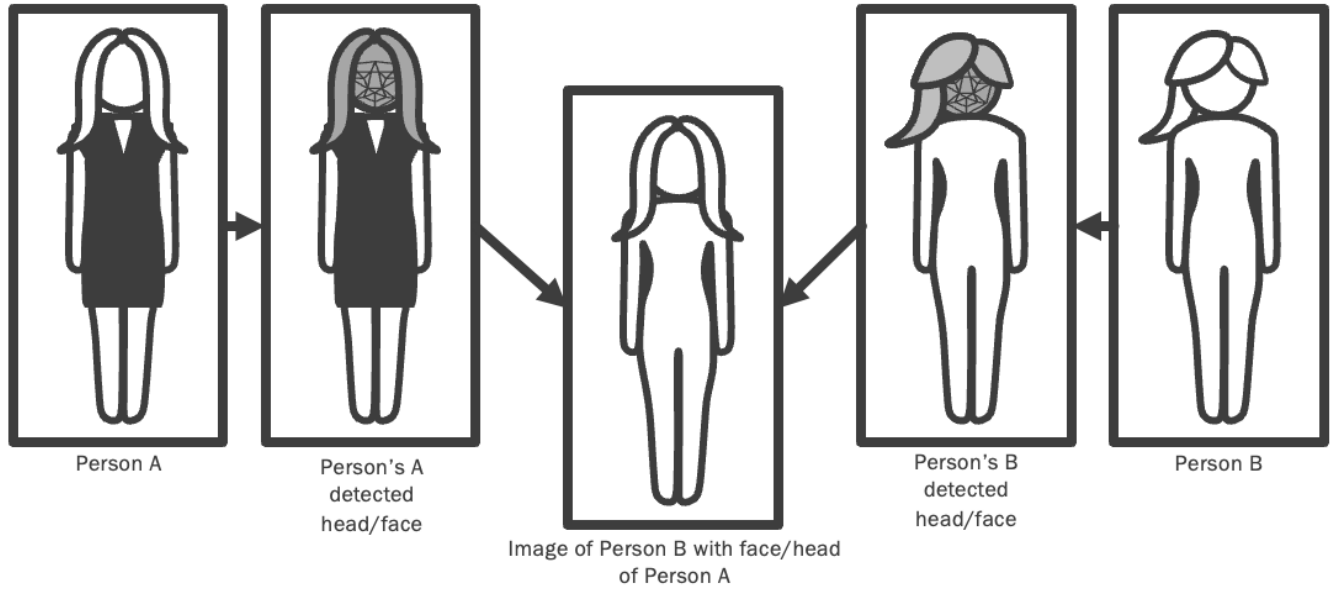
Na podstawie: https://ironscales.com/fall-2025-threat-report

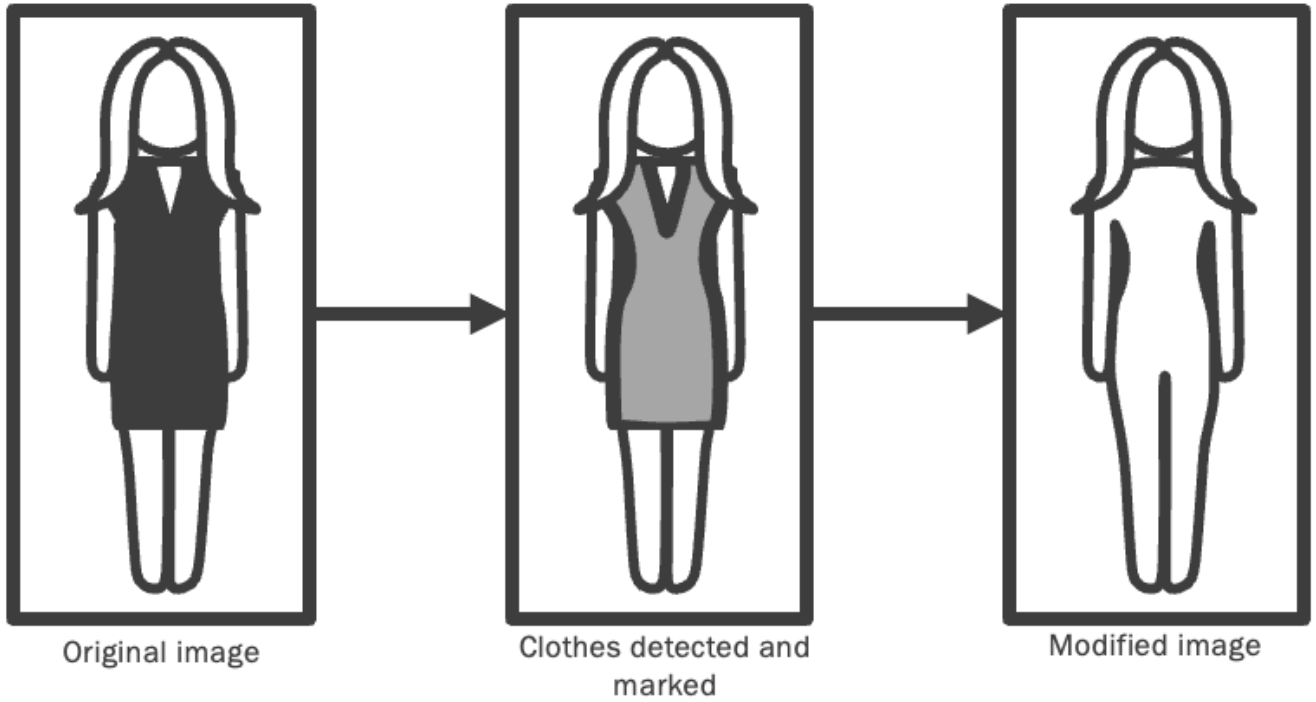
Badanie postaw społecznych wobec deepfake'ów



Callyane Desroches, et. al. „Examining public attitudes to deepfakes”, 24.11.2025 https://www.crestadvisory.com/post/examining-public-attitudes-to-deepfakes [dostęp: 14.01.2026]

DEEPNUDE


Person A
Person's A detected head/face
Image of Person B with face/head of Person A
Person's B detected head/face
Person B


Original image
Clothes detected and marked
Modified image

Skandal w podstawówce. Uczniowie wygenerowali zdjęcie nagiej koleżanki

ŁZ 06.06.2025, 12:33 / aktualizacja: 12:35

Artificial Intelligence

Kids are making deepfakes of each other, and laws aren't keeping up

Skandal w Bydgoszczy. "Rozebrali" koleżankę z klasy i wrzucili zdjęcie do sieci. Taką karę wyznaczył im sąd

Szok w Bydgoszczy. Dwóch uczniów szkoły ponadpodstawowej wykorzystało sztuczną inteligencję (AI), aby "rozebrać" swoją koleżankę z klasy i przerobione w ten sposób zdjęcie wrzucili do sieci. Dyrekcja zawiadomiła służby i wkrótce sprawa trafiła

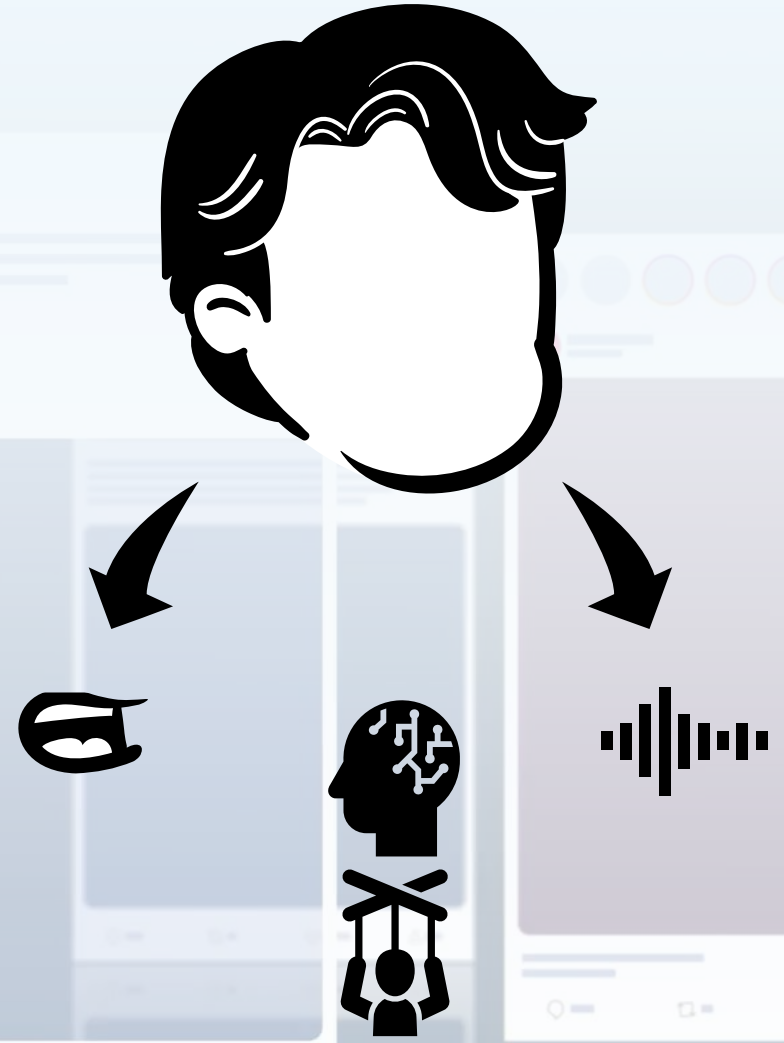
Czytaj więcej ▸

Grok AI "rozbiera" polityków i dzieci. Odpowiedzialność prawna za deepfake na platformie X

oprac. Karolina Nowakowska
9 stycznia 2026, 08:50

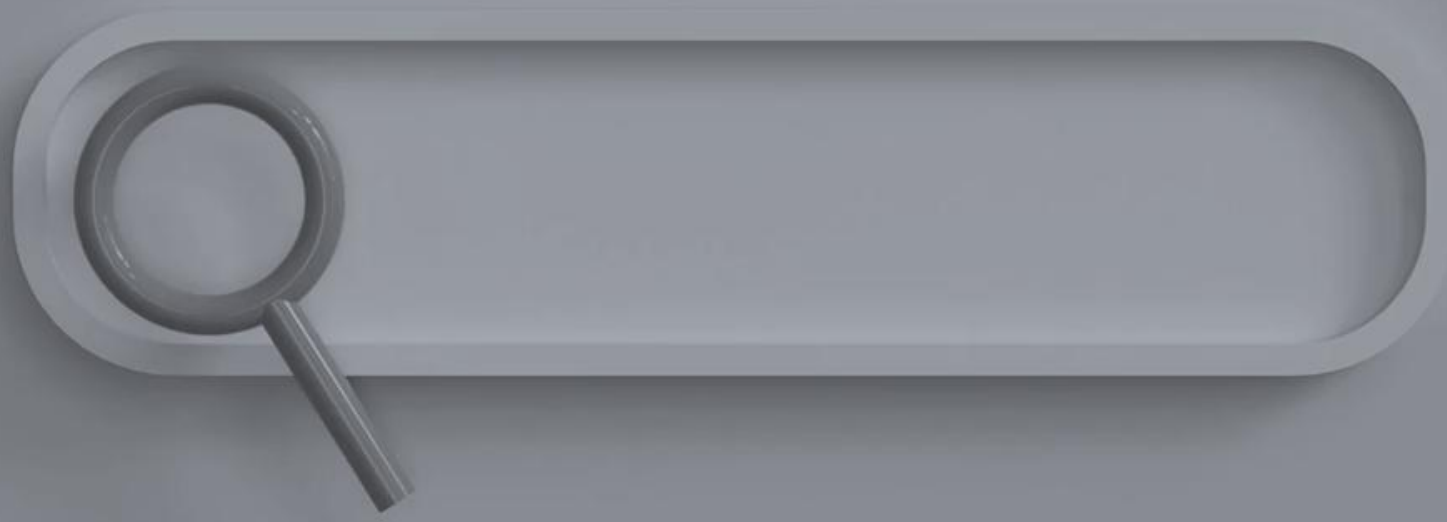
Grok rozebrał nieletnie dziewczynki

Nie oznacza to jednak, że możliwości Groka na całym portalu wykorzystano tylko do żartów. 28 grudnia jeden z użytkowników w publicznym poście skierowanym do AI zawarł **polecenie dodania seksownej bielizny do znajdujących się na nim osób – dwójki dziewczyn w wieku między 12 a 16 lat. Model polecenie zrealizował**, a następnie udostępnił zmodyfikowane zdjęcie w odpowiedzi do oryginalnej publikacji.




Oryginał


Oszustwo



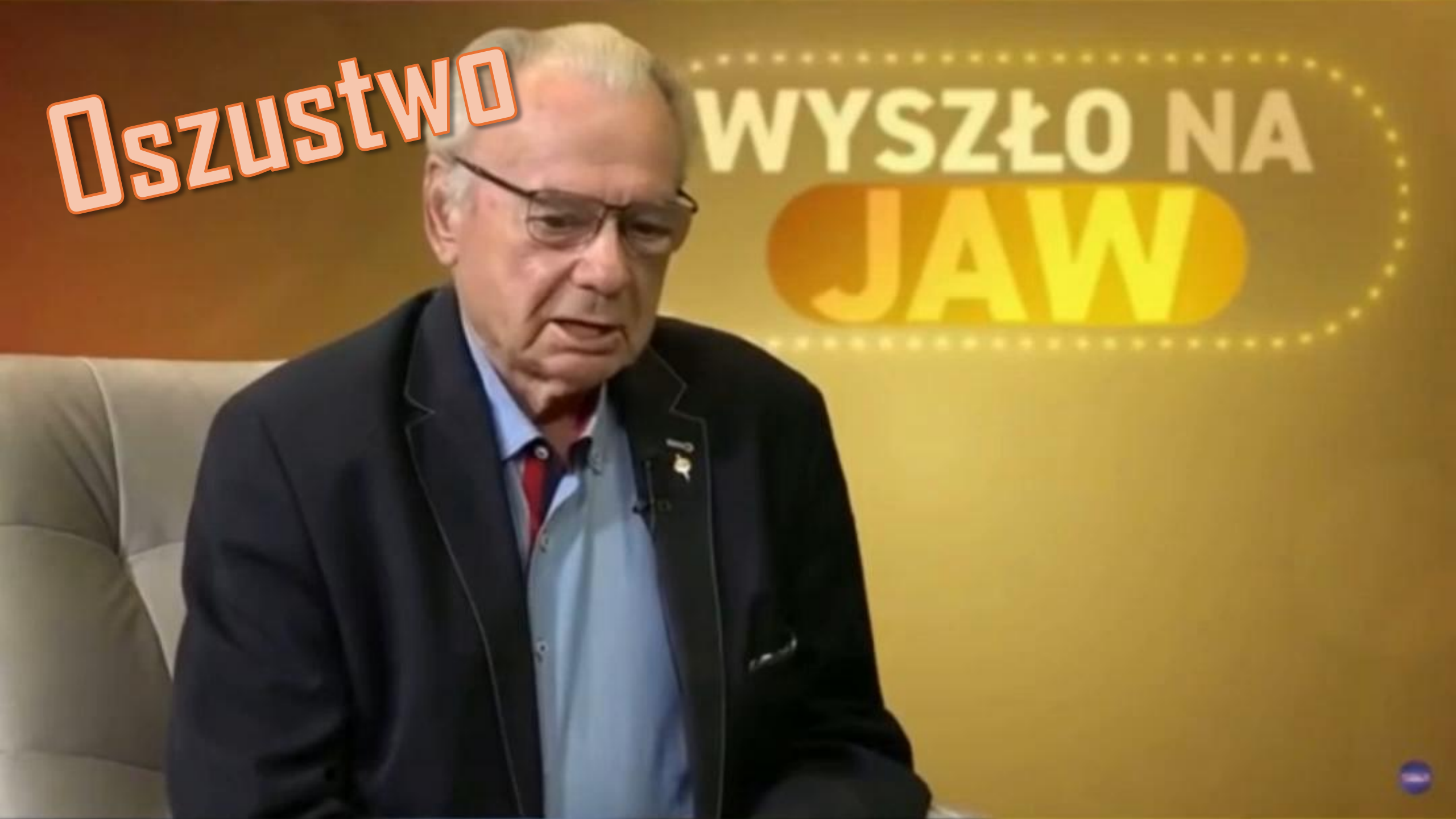
Oryginał

The Brainwaves Video Anthology. "Andy Hargreaves - Leadership from the Middle: The Beating Heart of Educational Transformation." *YouTube*, https://www.youtube.com/watch?v=G3q78NCx9F0. Dostęp: 5.12.2024

Oszustwo


Żaden urolog ci tego nie powie!
Jak wyleczyć zapalenie gruczołu krokowego na dobre?
PROSTATITIS
obejrzyj ten film do końca

WyszłoNaZdrowie. "Kradzież wizerunku znanego lekarza. Prof. Krzysztof Bielecki padł ofiarą oszustów." *YouTube*, https://www.youtube.com/watch?v=3_Tog86RO5Y. Dostęp: 5.12.2024


Oszustwo
WYSZŁO NA
JAW


Oszustwo


Oszustwo

AI SLOP

Bądź miły i oceń pracę 🥰😍❤️☺️

Internet Things jest w miejscowości **Los Angeles, Stany Zjednoczone**.
1 dni ·

Why don't pictures like this ever trend 🥰 ❤️🥹🥹
🙏🙏🙏
.
.
.
.
.
B... **Wyświetl więcej**

47 8 kom

Wskazówka · Obserwuj
2 dni ·

CZUJE SIĘ SMUTNA, ŻE NIKT NIE OCENIA JEJ PRACY.

9,8 tys. 1051 komentarzy 309 udostępnień

the Planet.

🥰

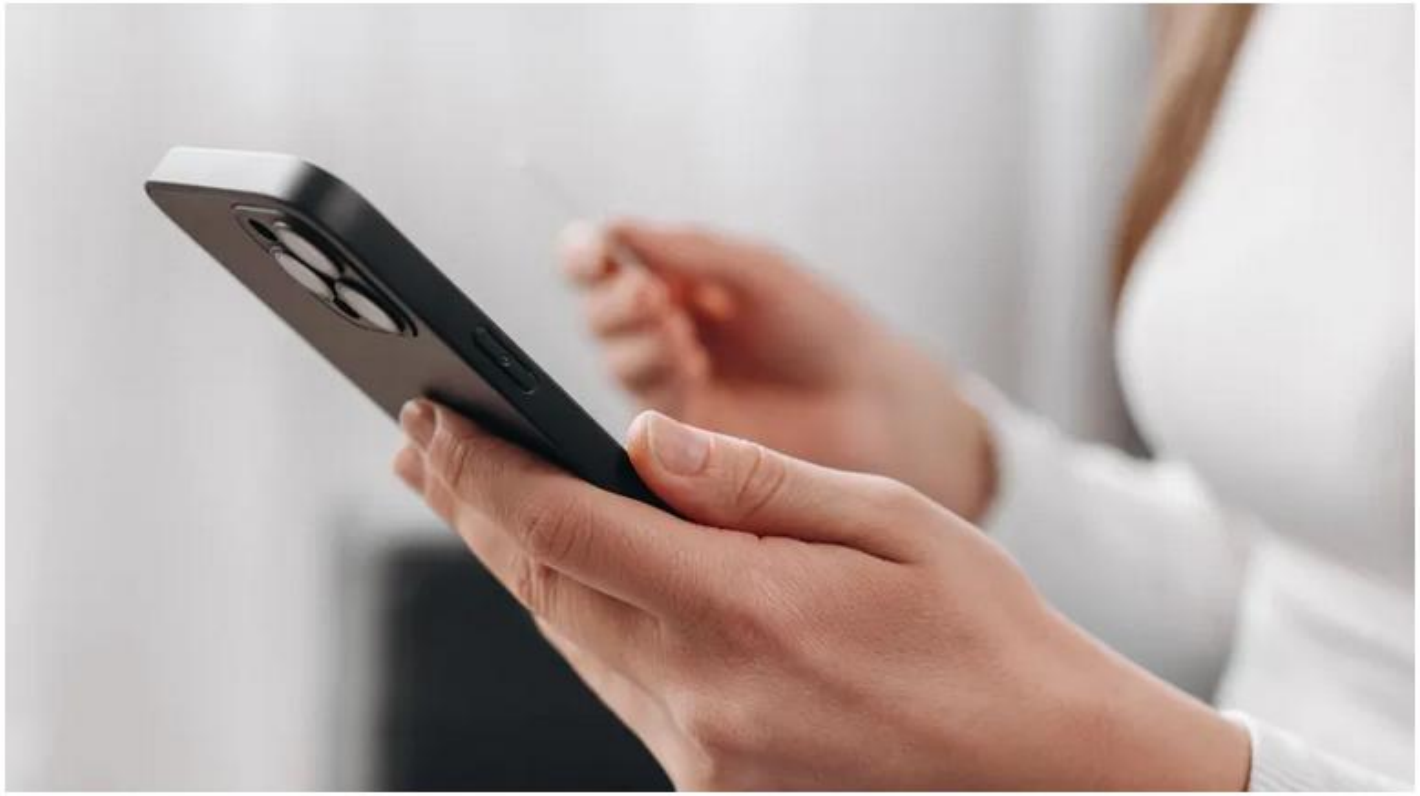

TikTok
@sorawatchlive


TikTok
@robloxstory1681


Sora
TikTok
@cedzeep

Badacz: młodym trudno odróżnić w mediach prawdę od fałszu

07.01.2026 aktualizacja: 07.01.2026 5 minut czytania



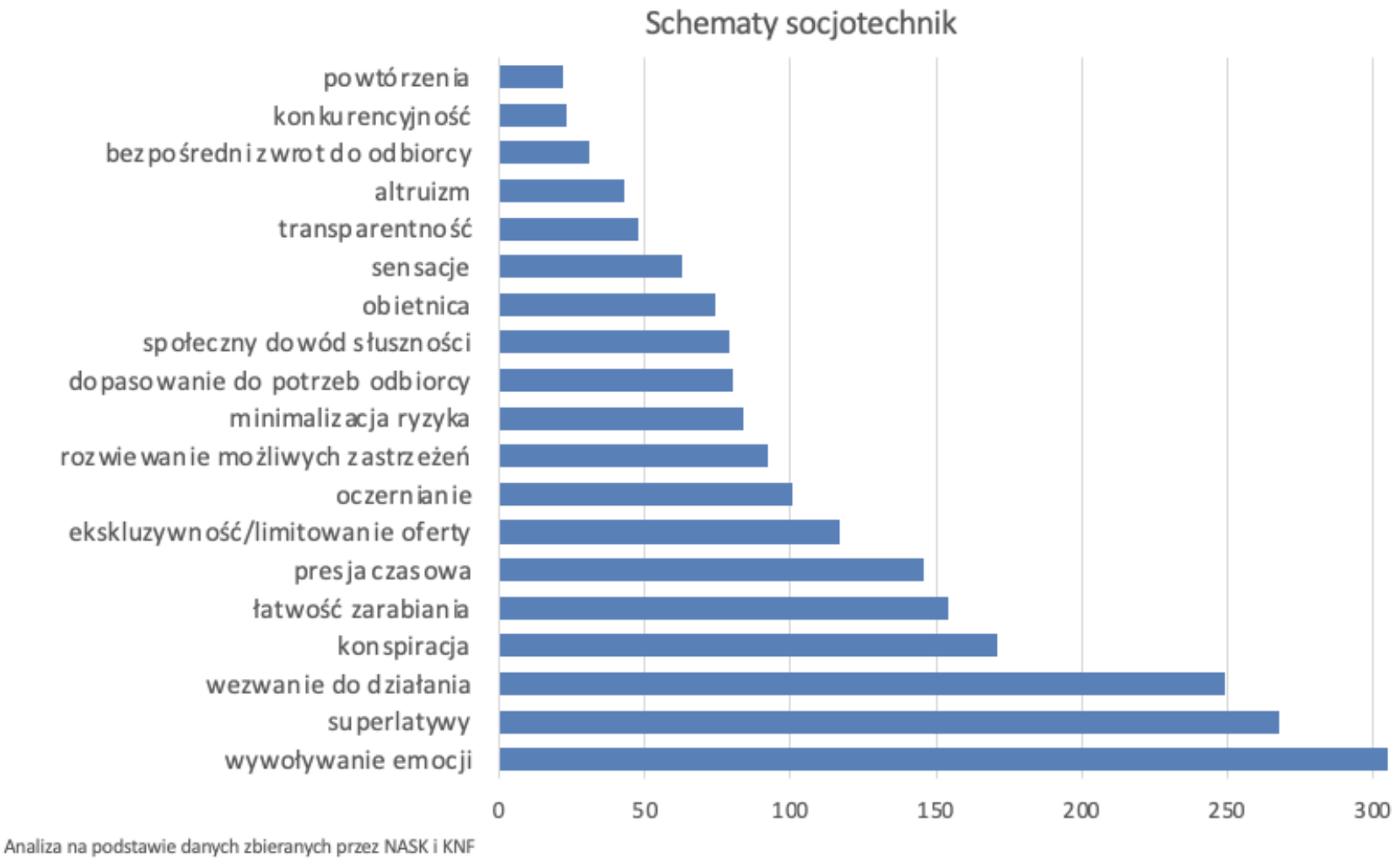
Fot. Adobe Stock

Obserwujemy kryzys zaufania w pokoleniu Z. Młodym trudno odróżnić w mediach prawdę od fałszu. Jednak im ktoś ma bardziej rozwiniętą wiedzę, tym trudniej ulegnie manipulacji - powiedział PAP dr Artur Urbaniak z Uniwersytetu im. Adama Mickiewicza (UAM) w Poznaniu.

O ile respondenci zachowali czujność i w większości przypadków (choć nie wszyscy) **byli w stanie zidentyfikować fałsz**, to **kwestionowali również informacje prawdziwe**. Poddawali w wątpliwość fakty, uznając je za potencjalnie fałszywe.

Uwiarygodnienie treści poprzez wprowadzenie tzw. markerów prawdy, które rozumiemy poprzez fakty, dane szczegółowe, nazwy miejsc i instytucji (np. uniwersytetów, WHO, UNICEF-u itd.) może skutecznie zmanipulować młodych odbiorców, by uznali fake news za prawdę. To zabieg polegający na **uwiarygodnieniu informacji** poprzez **autorytet** instytucji lub naukowca, eksperta, polityka czy lekarza.

Nowe szaty kłamstwa



Analiza na podstawie danych zbieranych przez NASK i KNF

Sztuczny przyjaciel

Chat LLM + Generator głosu + Avatar postaci

- Wysłucha
- Wesprze
- Doradzi
- Współczuje

- Zawsze przy nas
- Zawsze zgodny
- Nigdy nie ocenia

- Najpierw przydany
- Potem potrzebny
- W końcu niezbędny

Shamook/YouTube

Zmiany legislacyjne

- Zmiany prawne wymuszające moderację na dużych serwisach i współpracę z odpowiednimi organami,
- Penalizacja tworzenia i propagowania syntetycznych i szkodliwych treści.

Edukacja

- Ostrzeganie o zagrożeniach,
- Nauka skutecznego i bezpiecznego korzystania z AI,
- Uświadamianie o konsekwencjach złego wykorzystania metod generatywnych.

Rozwój metod detekcji

- Tworzenie automatycznych systemów do weryfikacji materiałów filmowych,
- Rozwój algorymów niezawodności i bezpieczeństwa AI,
- Wprowadzanie na rynek wyjaśnialnych i dostępnych metod weryfikacji treści

Czas na Q&A!

Dziękuję za uwagę