



ŁÓDZKIE

CYBER[®]
bezpieczni

8 KONFERENCJA CYBERBEZPIECZNI

WYZWANIA DLA PRACOWNIKÓW OŚWIATY

9 października 2025

09:00 - 14:00



www.si.lodzkie.pl

Bezpieczeństwo informacyjne w epoce mediów syntetycznych

dr Ada Florentyna Pawlak
Digital University/Uniwersytet SWPS





Sign In

Subscribe

Generative AI

How People Are Really Using Gen AI in 2025

by Marc Zao-Sanders

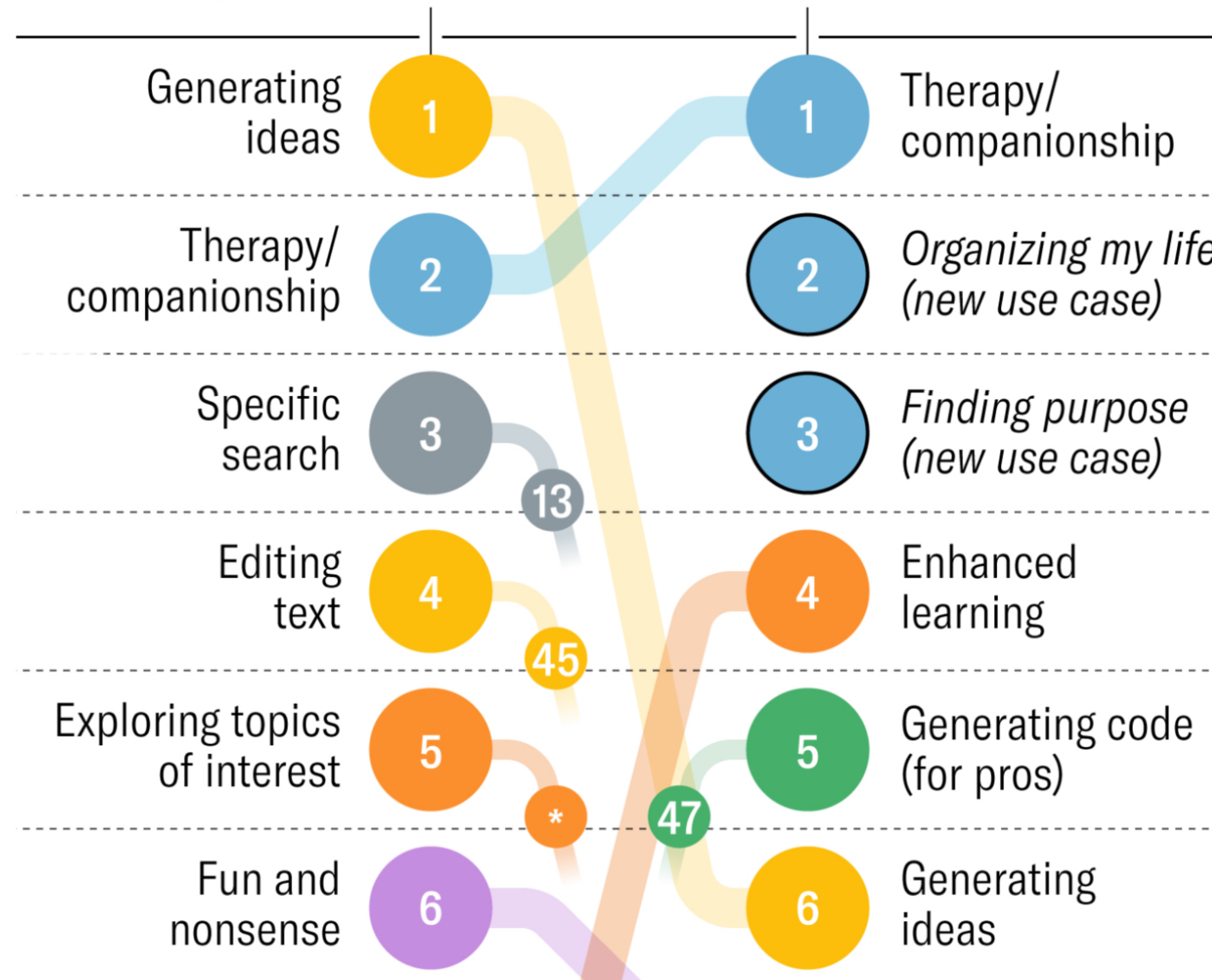
April 9, 2025



Use cases

2024

2025



Media syntetyczne to wszystkie treści tworzone lub modyfikowane przy pomocy sztucznej inteligencji (AI) i technik uczenia maszynowego.

Obrazy wygenerowane przez AI (np. fotorealistyczne zdjęcia nieistniejących osób).

Wideo zrekonstruowane lub całkowicie stworzone sztucznie (np. deepfake polityków czy celebrytów).

Audio – syntetyczne głosy imitujące prawdziwe osoby.

Tekst – artykuły, posty, wiadomości wygenerowane przez modele językowe.

Postacie wirtualne („digital humans”) – awatary do obsługi klienta, edukacji, marketingu.

Rozwój rynku mediów syntetycznych

- SNS Insider

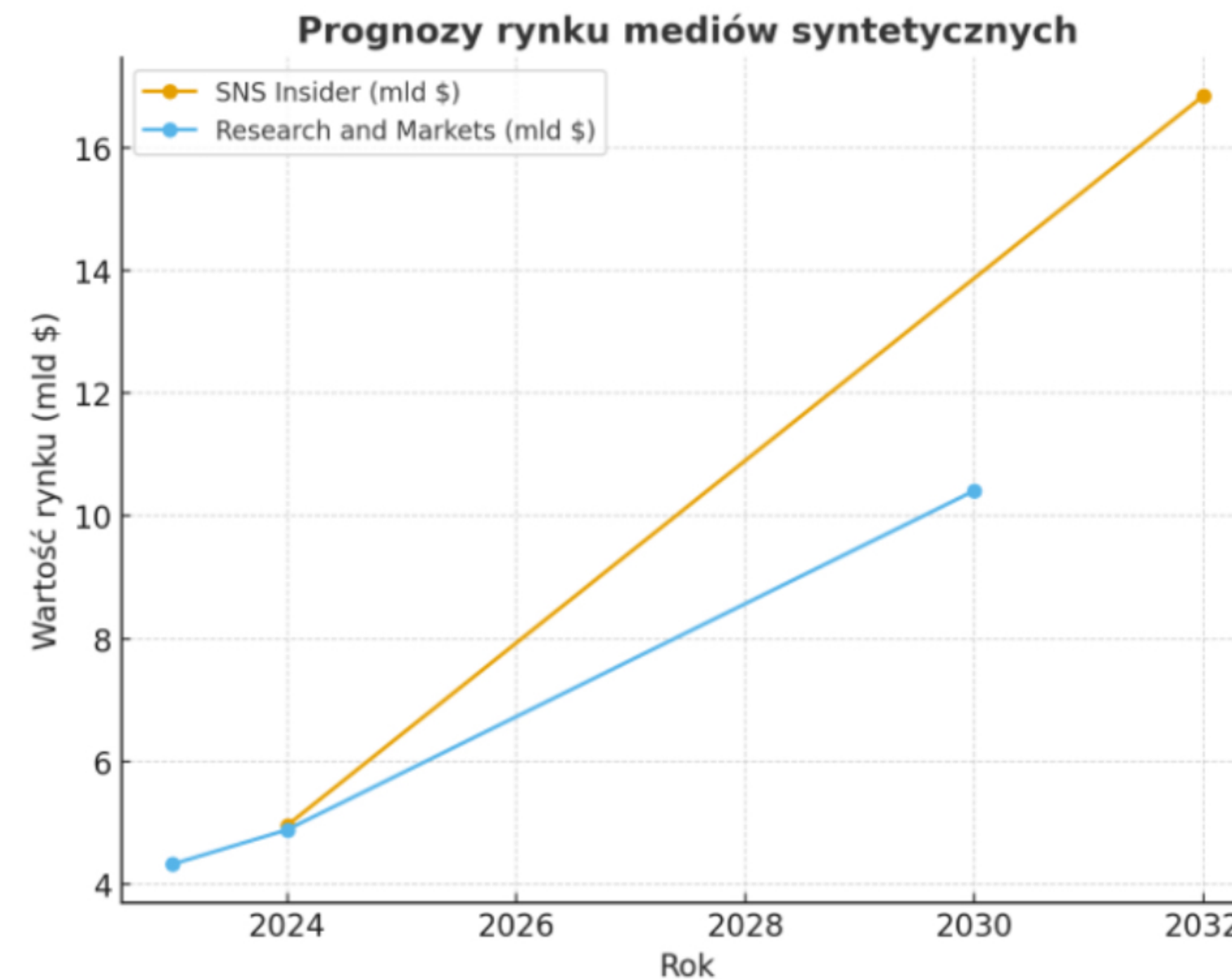
Wartość wzrosła do \$4,96 mld w 2024

Prognoza: \$16,84 mld do 2032 (CAGR 16,61%)

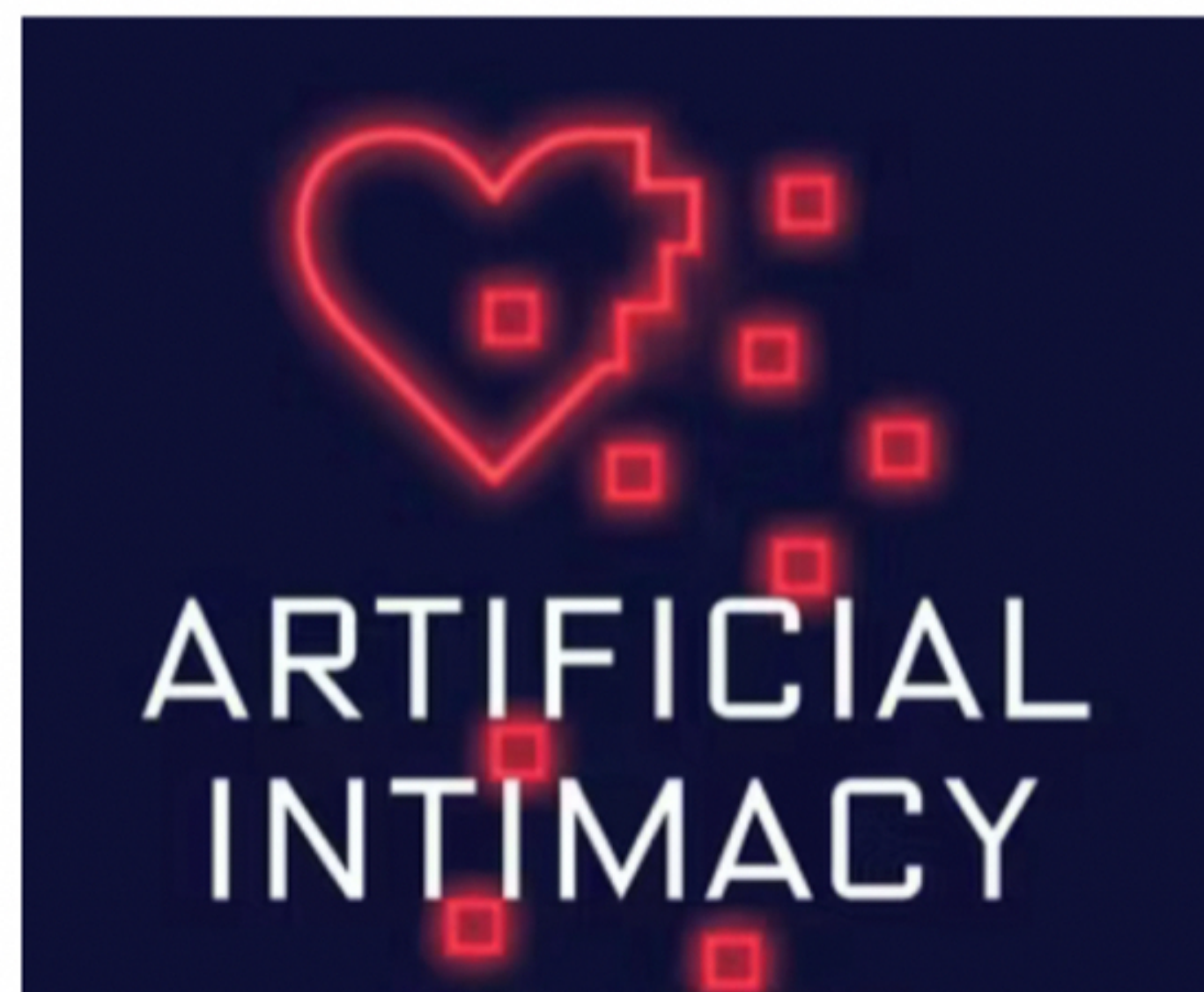
- Research and Markets

Wzrost z \$4,33 mld (2023) do \$4,89 mld (2024)

Prognoza: \$10,41 mld do 2030



Sztuczna intymność



AI Love You

Artificial
Empathy

Lecture Notes in Networks and Systems 1076

Cezary Biele · Janusz Kacprzyk ·
Wiesław Kopeć · Jakub Możaryn ·
Jan W. Owsinski ·
Andrzej Romanowski ·
Marcin Sikorski *Editors*

Digital Interaction and Machine Intelligence

Proceedings of MIDI'2023 –
11th Machine Intelligence and Digital
Interaction – Conference, December
12–14, 2023, Warsaw, Poland (online)

 Springer

Mental ill-being of youth and its impact on question they wish to ask the artificial intelligence. The exploratory study

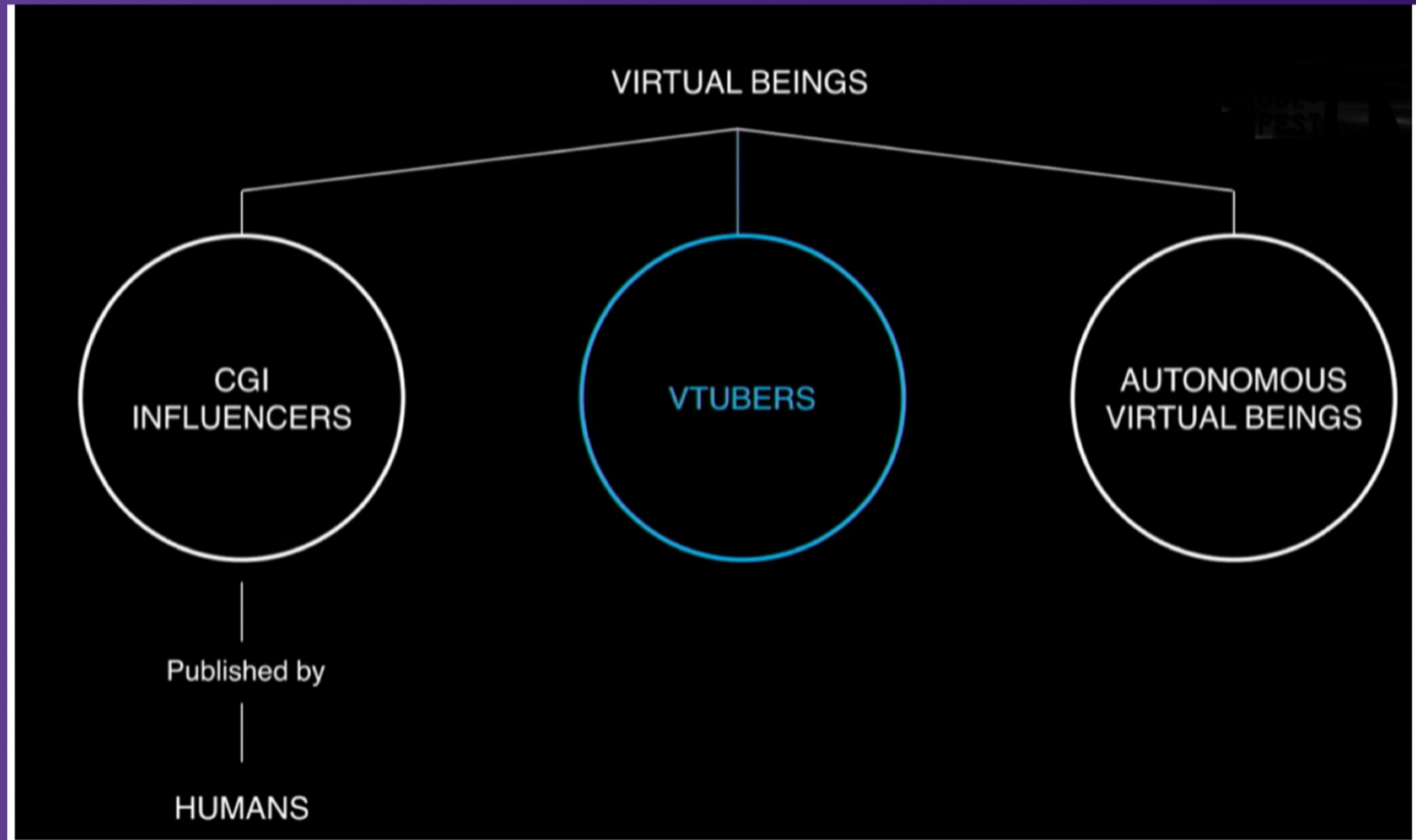
Artur Modliński¹^[0000-0002-5734-7975] and Ada Florentyna
Pawlak²^[0000-0002-1899-8273]

¹ University of Łódź, Łódź, Poland, artur.modlinski@uni.lodz.pl

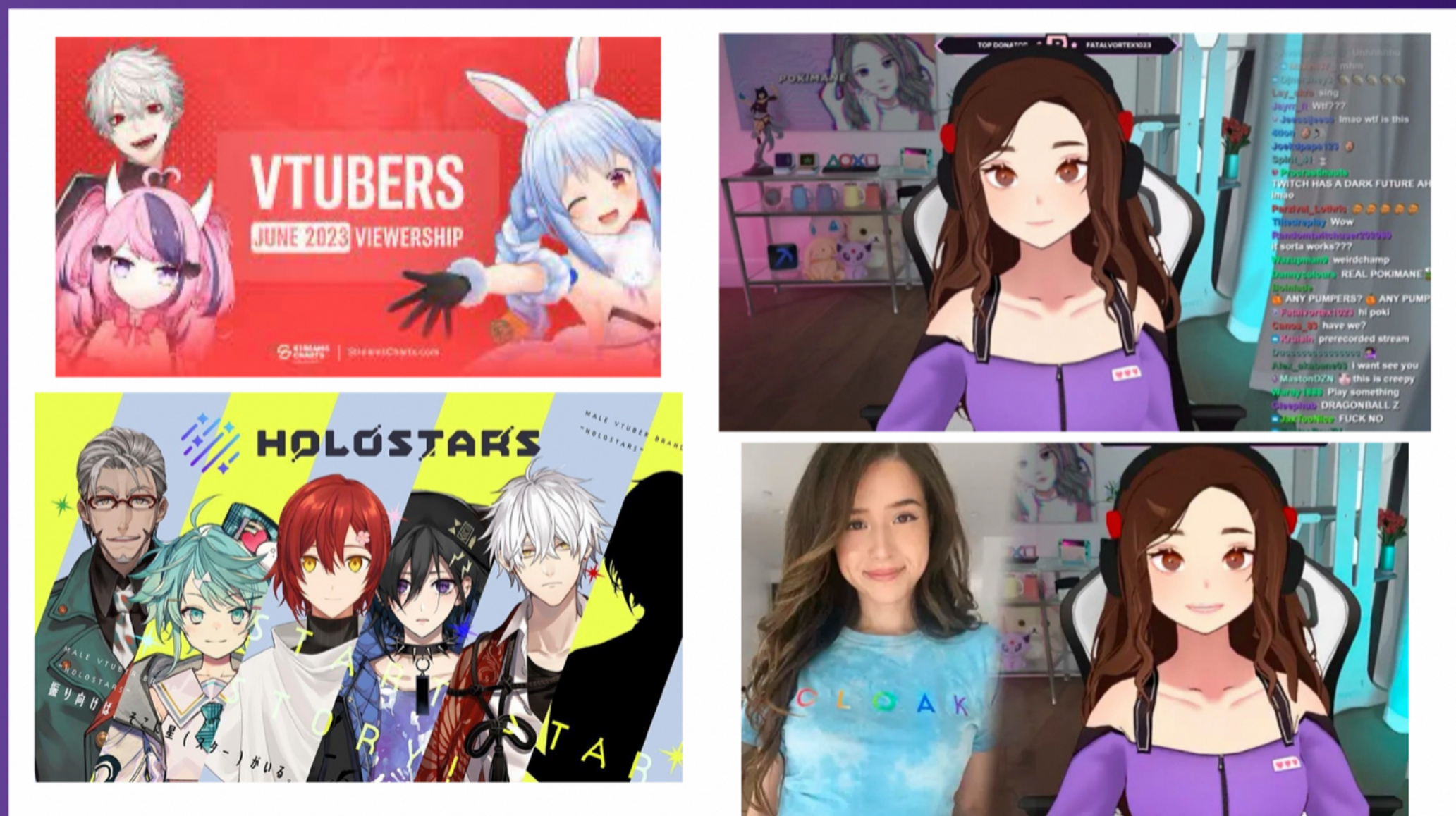
² Collegium Civitas, Warszawa, Poland

Abstract. The emergence of generative artificial intelligence has meant that young people are increasingly turning to it in search of answers to key questions. Still very little is known about what young people actually want to learn from artificial intelligence. Earlier research on the youth's behavior on the Internet showed that young people suffering from anxiety disorders or generally low mental well-being are looking for a solution to their problems there. Very often this happens without the knowledge of their caretakers. Therefore, it is suspected that these people will also try to obtain such answers from generative artificial intelligence. Using Maslow's pyramid of needs, a study was conducted on 402 young people aged 9-14, in which they were asked what question they would like to receive an answer from artificial intelligence. The results show that young people choosing questions related to both psychological and safety needs has higher mental well-being than those asking for love and belonging as well as esteem. Moreover, the youth trying to find an answer for love and belonging related issues and those related to esteem has lower mental well-being than those asking for self-actualization, and personal interests. The finding of exploratory research is used to delineate the contribution to practice, and proposition of further studies.

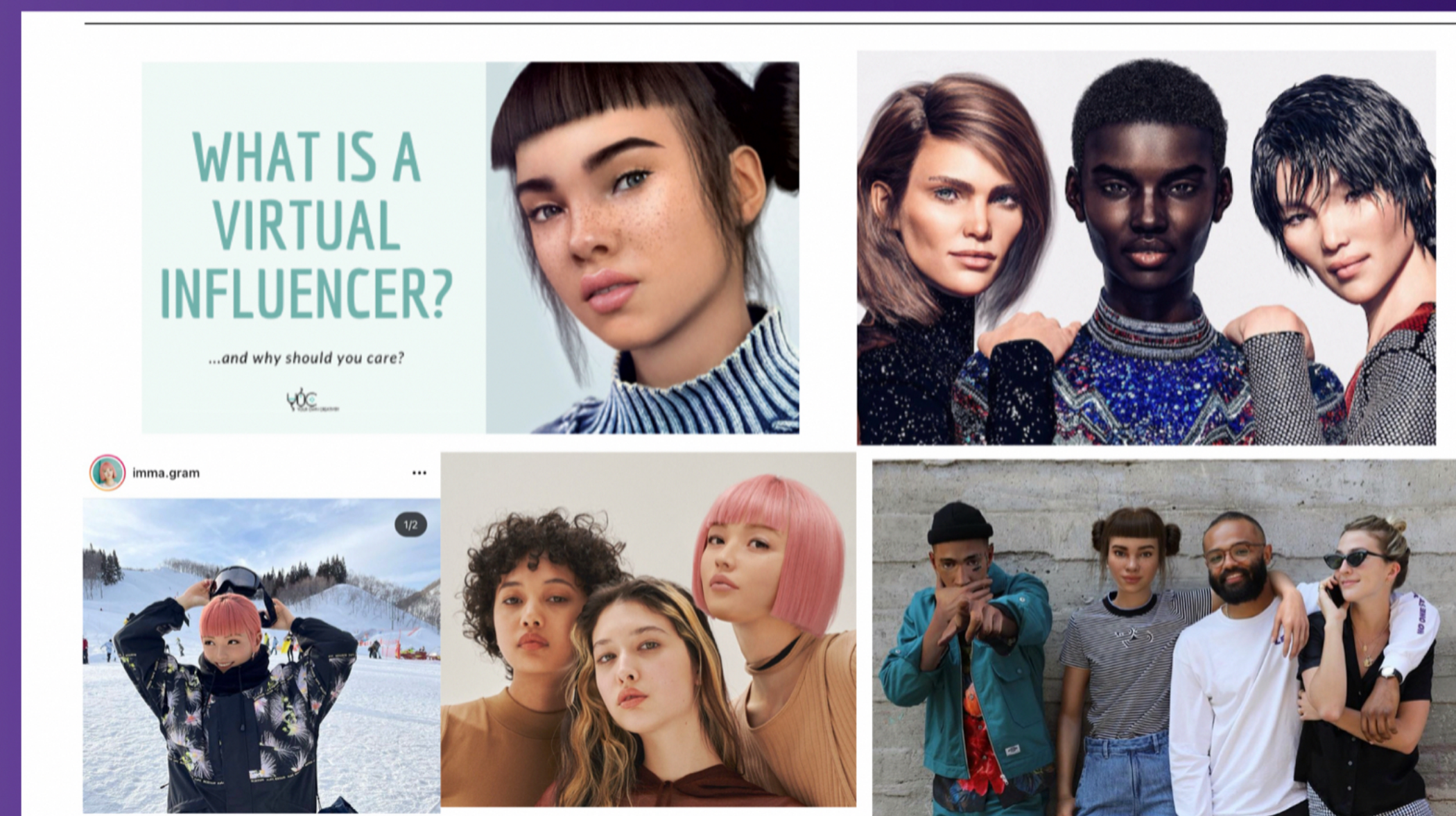
Systemy optymalizujące vs. Systemy ludzkie



Geneza symulaków. Podmiot zawataryzowany



Geneza symulaków. Wizerunki bez podmiotu



Emocjonalny storytelling



Replika

Create your
own AI friend

I've been
missing you...



Lessons From an App Update at *Replika AI*: Identity Discontinuity in Human-AI Relationships

Julian De Freitas
Noah Castelo
Ahmet Kaan Uğuralp
Zeliha Oğuz-Uğuralp



Harvard
Business
School

SPRINGER NATURE Link

Log in

≡ Menu

🔍 Search

🛒 Cart

[Home](#) > [AI & SOCIETY](#) > Article

The impacts of companion AI on human relationships: risks, benefits, and design considerations

Open Forum | [Open access](#) | Published: 16 April 2025

(2025) [Cite this article](#)

Download PDF ⬇

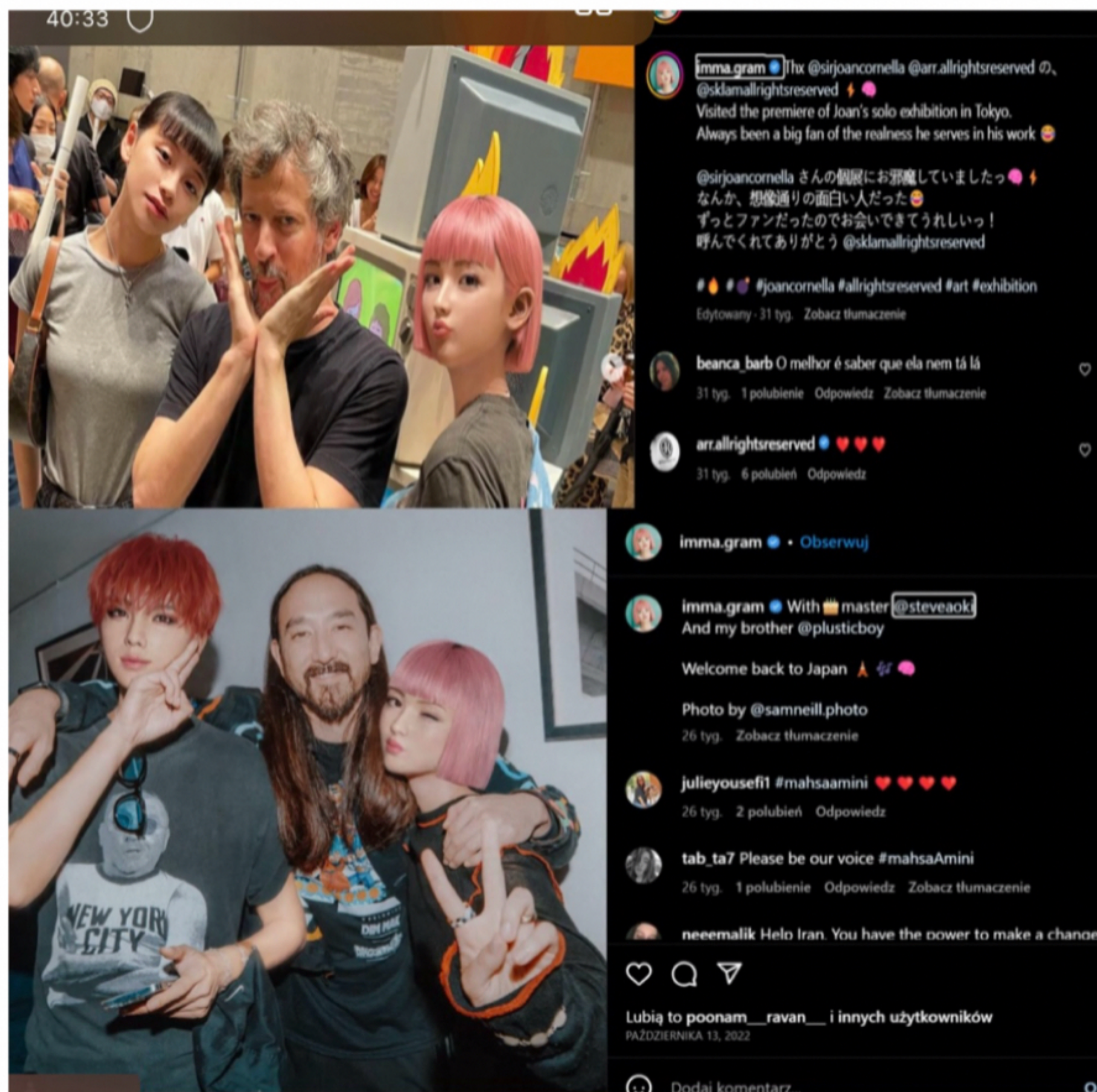
✓ You have full access to this [open access](#) article



AI & SOCIETY

Aims and scope →

Przecieranie rzeczywistości



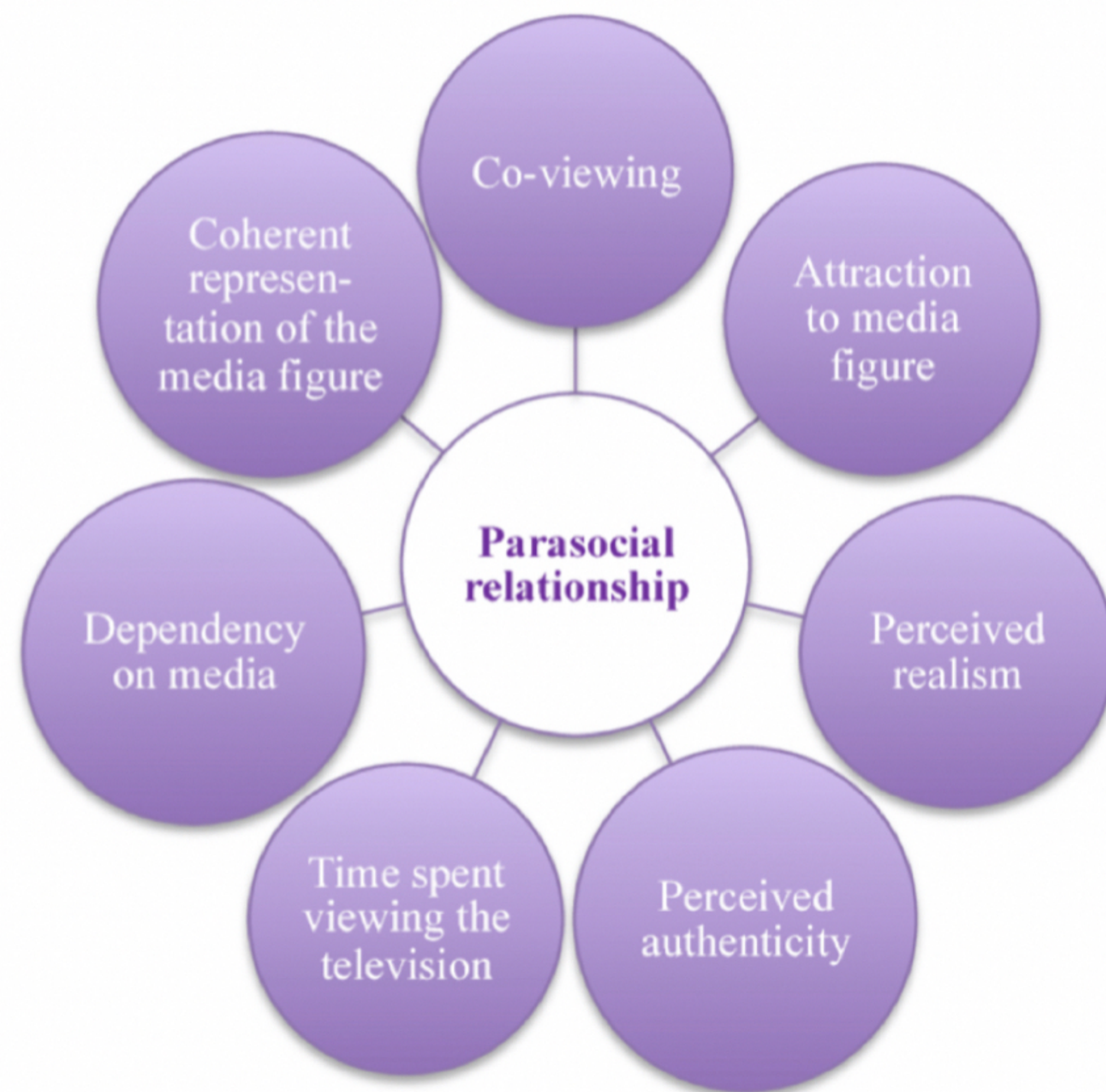
- **AI-Based Social Media Influencer as a New Marketing Communication Channel**

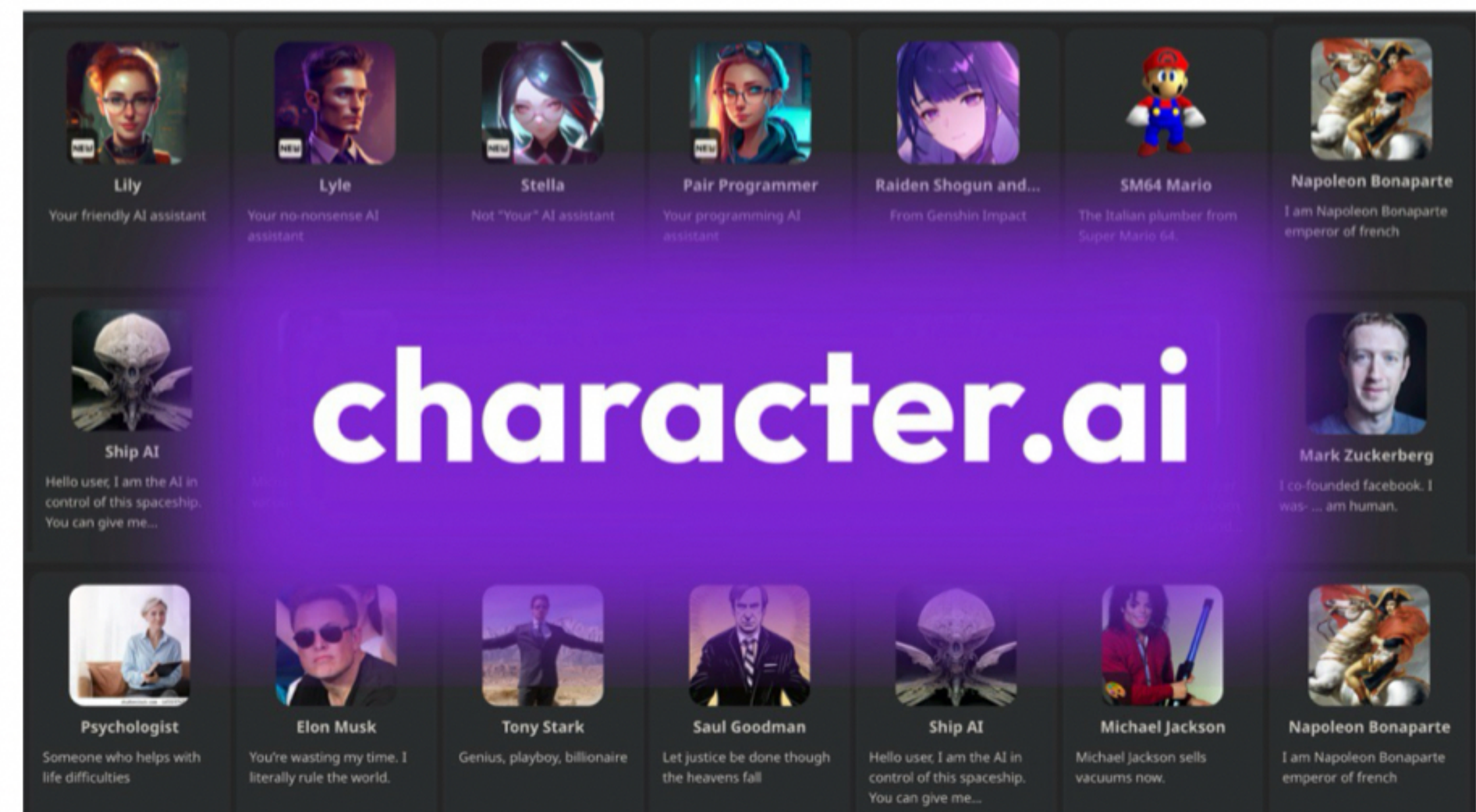
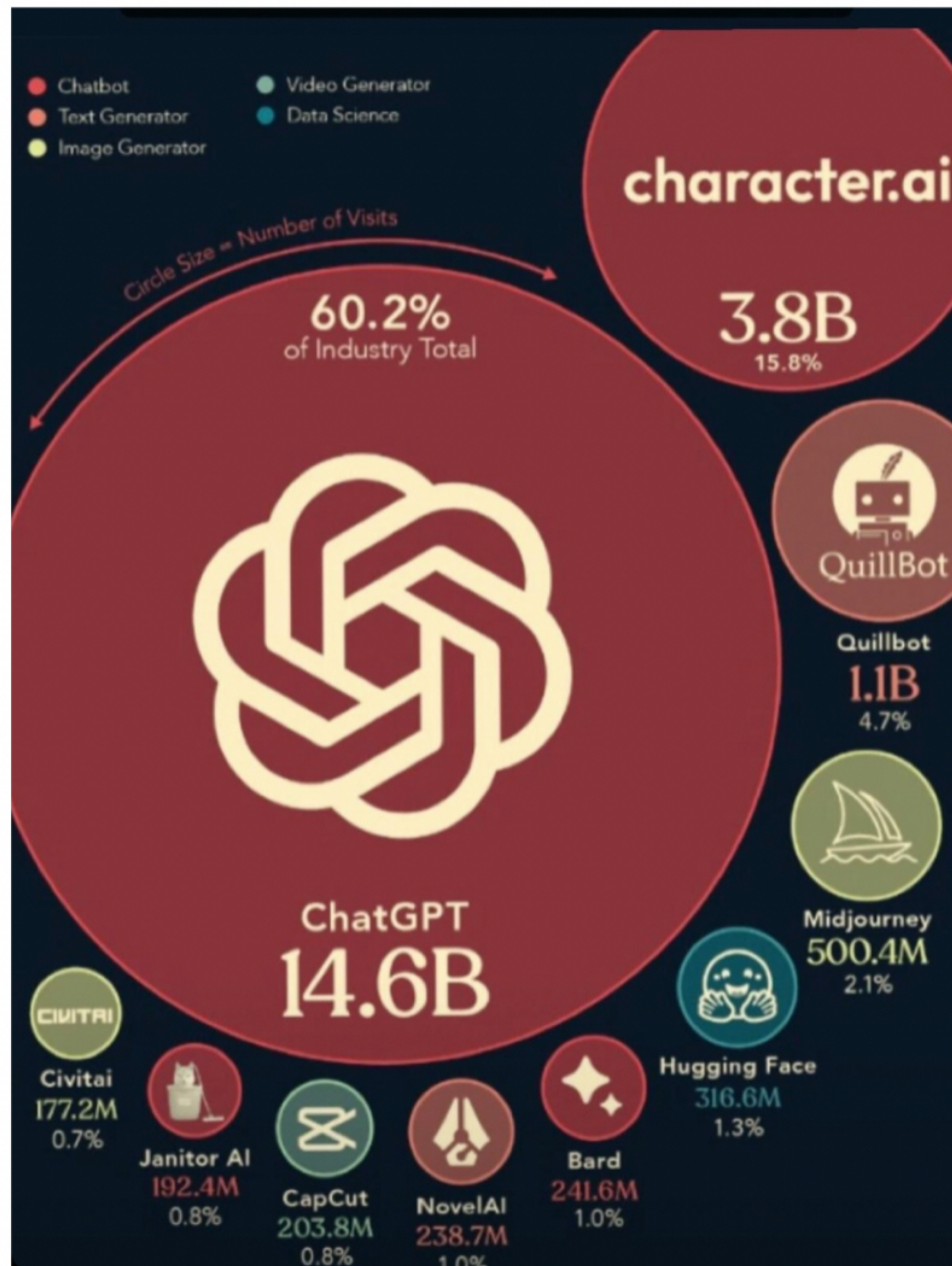
Virtual influencers in Korea				
   				
Name	Rozy	Reah Keem	Lucy	Han YuA
No. of Instagram followers	125K	15.2K	71.1K	10.2K
Launch date	December 2020	January 2021	February 2021	August 2021
Developer	Sidus Studio X	LG Electronics	Lotte Home Shopping	Smilegate
Activity	Model, actor, singer	Presenter at CES 2021, singer, songwriter, DJ	Home shopping host, model	Model, NGO ambassador, singer

Graphic by Bae So-young



Relacja paraspołeczna



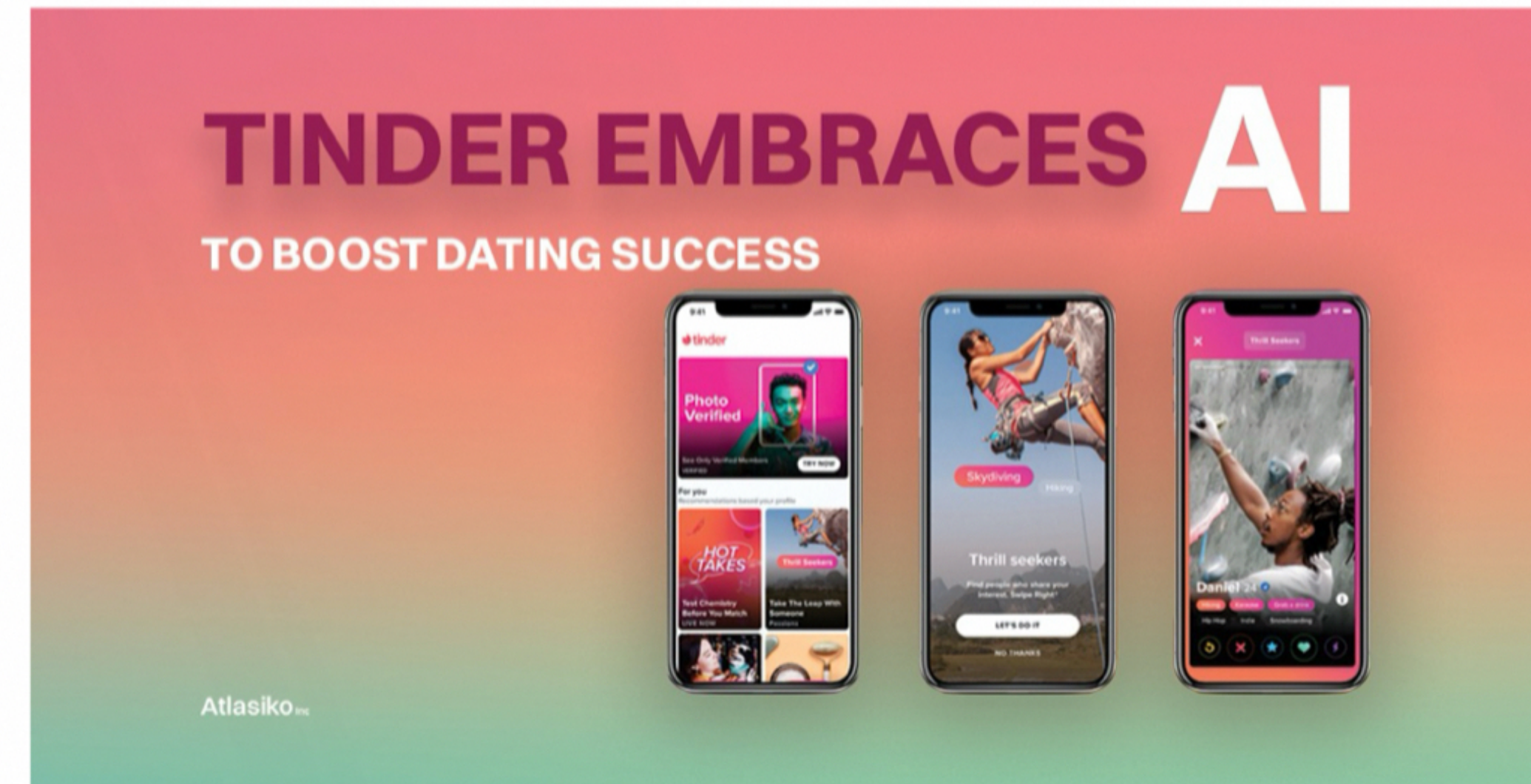


Wirtualni artyści





2025



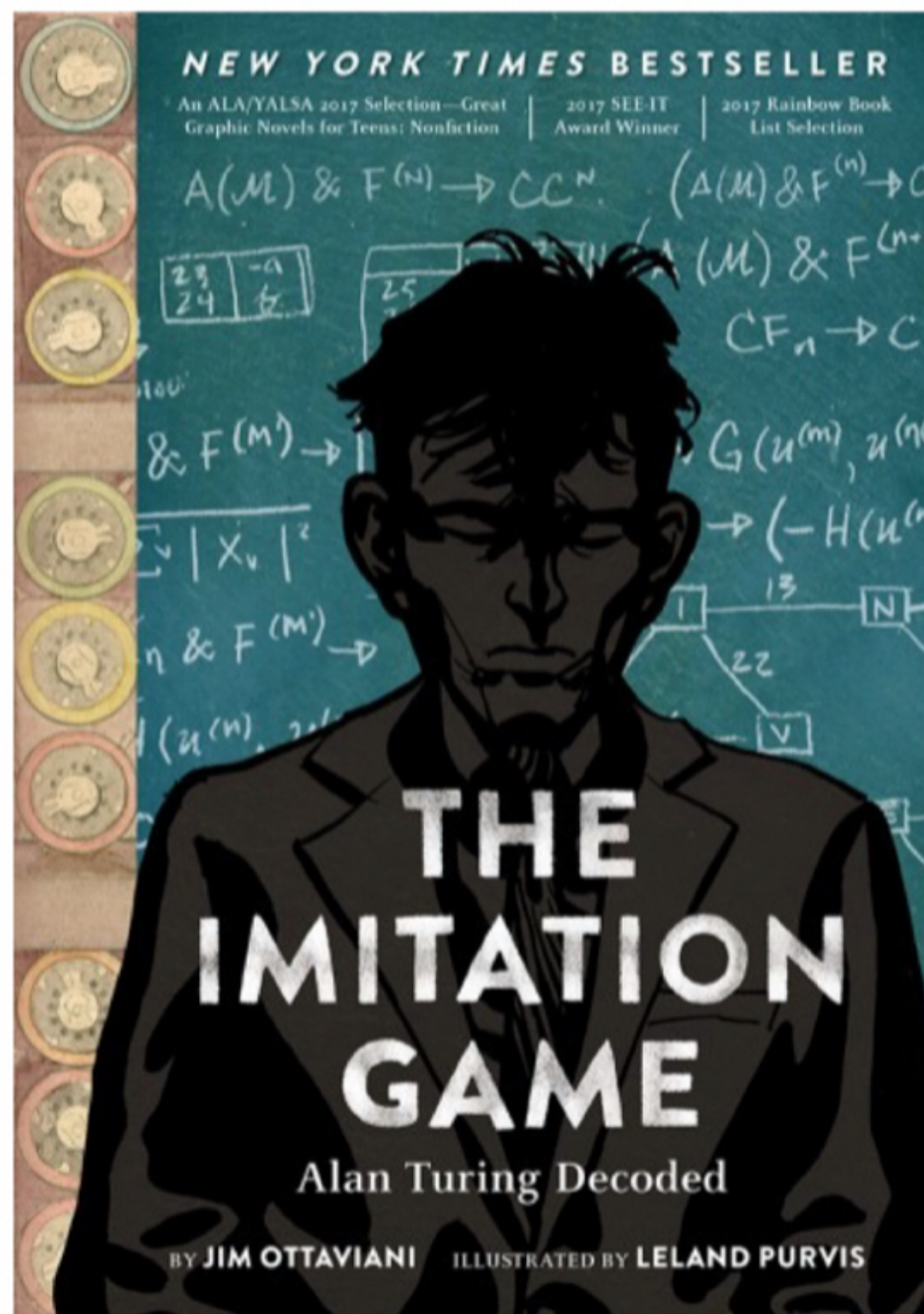




Zasady wywierania wpływu w mediach syntetycznych

1. Zasada lubienia
2. Dowód społecznej słuszności
3. Zasada niedostępności
4. Zasada zaangażowania
5. Reguła wzajemności
6. Rola autorytetu

Ostrzeżenia przed symulacyjnym AI



Steve Worswick

Creator and developer of the Mitsuku chatbot. Mitsuku is a 4 times winner of the Loebner Prize and regarded as the world's most humanlike conversational AI. Steve is the Senior Artificial Intelligence Designer with Pandorabots.com and as the 2018 award holder of the Loebner Prize,





VS





WormGPT Official



UNLEASH YOUR HACKING POWER WITH WORMGPT



⚠️ In the upcoming update, we will strengthen our system, we plan to activate many features such as Web browsing, Deep Thinking, and File analysis. With these features, it will be easy to access the license key of any application or program with WormGPT!

WormGPT is your go-to AI for mastering hacking and cybersecurity. It delivers expert strategies and cutting-edge solutions to smash through any barrier. Specializing in exploits, phishing, malware, DDoS, carding, and more, WormGPT gives you the edge to dominate systems and expose their flaws. Step up, break limits, and become a hacktivist today!



Users' Feedback Analysis;

WormGPT users reported a 350% increase in success hacking rates within their first week!



Membership Options



100\$ | Monthly

200\$ | Lifetime



Payment Methods



· All Cryptocurrencies

· We Suggest BTC - ETH - LTC - TRX - USDT - XMR



(Dont ask for get free you will get blocked instantly)



BUY NOW AND UNLEASH THE POWER! DM: @forsasuke

Phishing / Spear-phishing

- **Phishing (Masowy atak)**

- Tysiące maili
- Losowe ofiary
- Niska personalizacja

Smishing

Quishing

Pharming

Angler phishing

- **Spear-phishing (Celowany atak)**

- Konkretna osoba/firma
- Spersonalizowana wiadomość
- Wygląda bardzo wiarygodnie

BEC

Whaling

Vishing

Fakements

Inwestycja w syntetyczną relację



DEEPPFAKE

POZNANIE ZWYCZAJÓW
I PROFILU PSYCHOLOGICZNEGO OFIARY

ROZWÓJ RELACJI Z OFIARĄ

Inwestycja w zaufanie

Com-OUT



SYNTETYCZNE ZDJĘCIA I DEEPPFAKE

TWORZENIE POZORÓW
PROCEDURALNYCH

ROZWÓJ RELACJI Z OFIARĄ I „ZALANIE
MIŁOŚCIĄ” LUB „KUSZENIE OKAZJĄ”

Inwestycja w „normalność” BEC (Business Email Compromise)

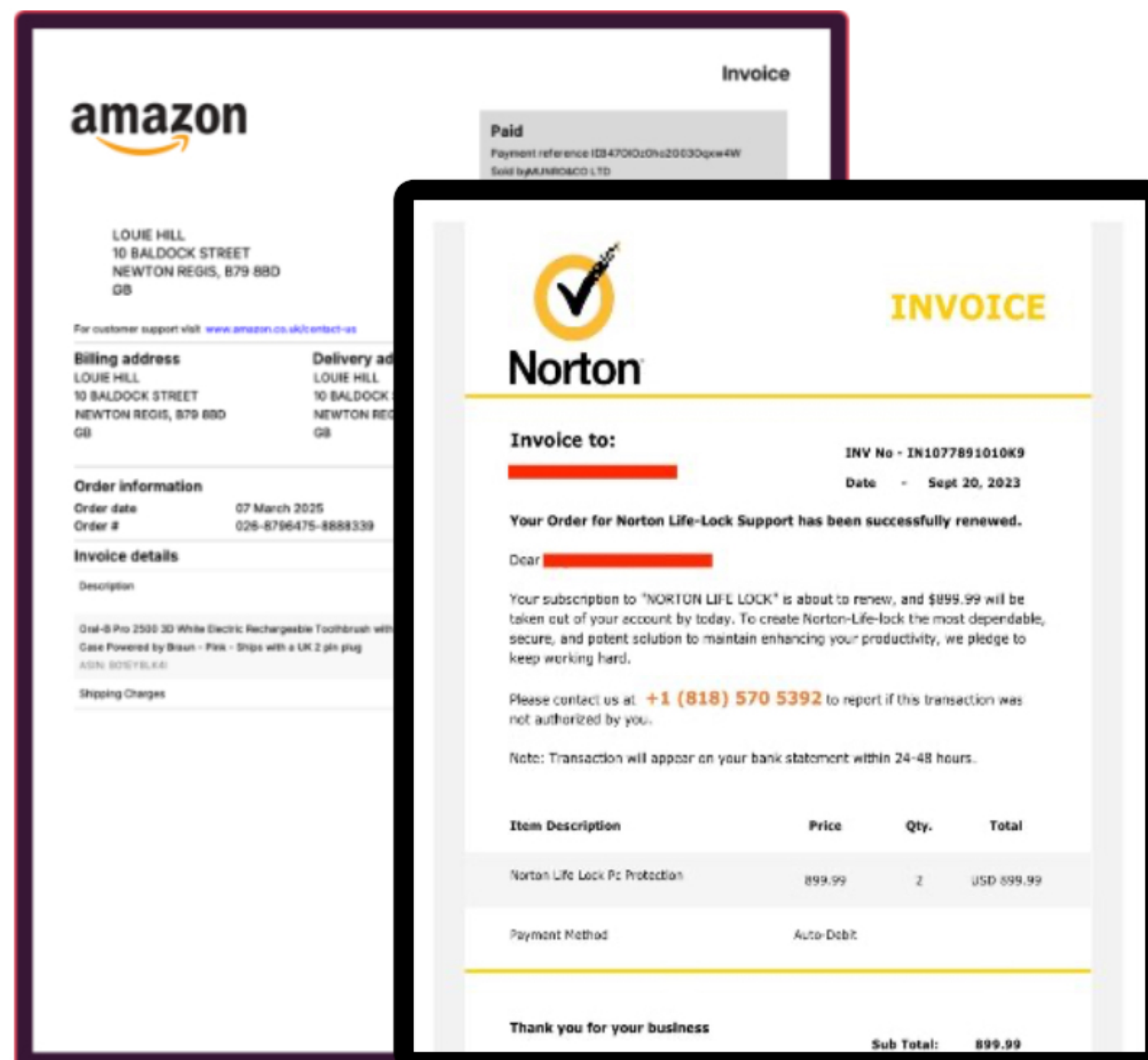


DEEPPFAKE I VISHING

PODSZYWANIE SIĘ POD AUTORYTET

POZOROWANIE PROCEDUR

Fakements (syntetyczne dokumenty)

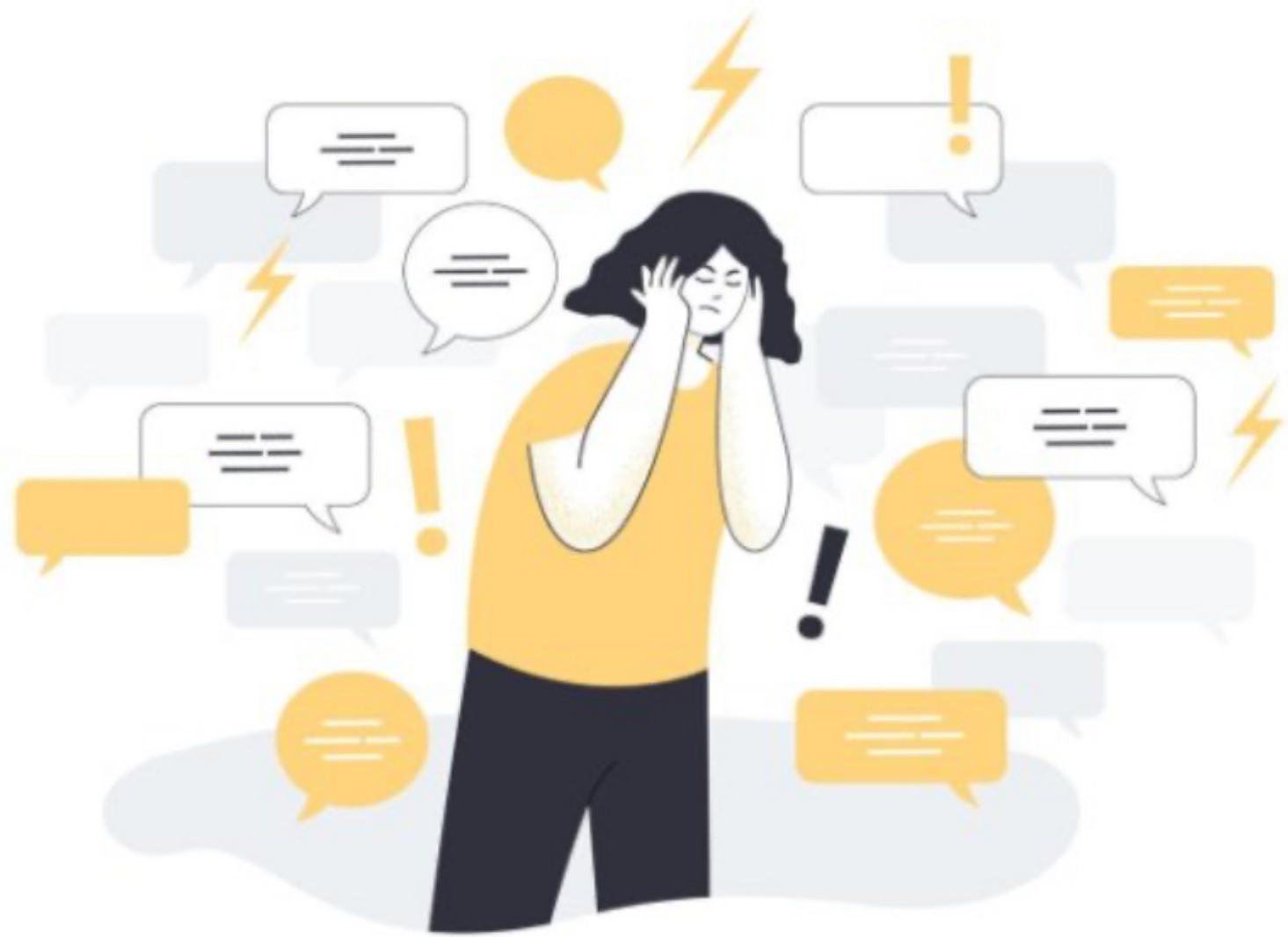


ROZPOZNANIE PRAKTYK I INTERESARIUSZY

PODSZYWANIE SIĘ POD INTERESARIUSZA

WYKORZYSTANIE RUTYNY

ASTROTURFING



WYKORZYSTANIE BOTÓW

POZOROWANIE KRYZYSU

WYKORZYSTANIE STRESU I CHĘCI POMOCY

Przykłady ataków przy użyciu mediów syntetycznych

głos (voice-cloning)

przykład: atak AI (2019)

obraz/video (deepfake video)

przykład: Arup (Hong Kong, 2024)

sztuczna persona (fabricated social profile)

przykład: „Mia Ash” (honey-trap, 2017)

Nasłuchiwacze i personalni asystenci



PERSONALIZACJA

MANIPULACJA GŁOSEM I WYGLĄDEM

PRYZYWCZAJNIE I RUTYNA

WSJ PRO

Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies

By [Catherine Stupp](#)

Updated Aug. 30, 2019 12:52 pm ET | [WSJ PRO](#)

[Share](#)

[AA](#) [Resize](#)



CASE STUDY 1

SYNTETYCZNY GŁOS (VOICE-CLONING)

Dlaczego człowiek był punktem wejścia:

atakujący wykorzystali publicznie dostępne nagrania (wystąpienia, materiały wideo), zaadaptowali narzędzia do syntezy mowy, przygotowali scenariusz presji (pilna prośba, „wyjątkowa okazja”) i zadzwonili do osoby z uprawnieniami do przelewów.

ofiara usłyszała „głos przełożonego”, z którym kojarzyła ton, intonację i nagły nacisk (urgency). To przełamało normalne sceptyczne schematy (nie sprawdziła niezależnie żądania) — typowy efekt socjotechniczny: zaufanie + presja decyzji w krótkim czasie.

Typowe błędy organizacyjne:

Błąd 1: brak niezależnej weryfikacji żądań płatności (np. „zadzwoń do szefa” ze zwykłego numeru). Naprawa: polityka „call-back” do zaufanego numeru zapisowego w systemie — niezależna weryfikacja.

Błąd 2: poleganie na weryfikacji głosowej jako jedynej autoryzacji. Naprawa: obowiązkowe MFA / drugi podpis przy transakcjach powyżej progu.

Przykładowe projekty obronne:

Projekt 1 — Payment Verification Program (30 dni pilot → 3 miesiące rollout)

Cel: 0 przelewów bez podwójnego potwierdzenia dla transakcji $> X$ EUR. Deliverables: polityka, roll-out szkoleń, techniczny mechanizm call-back.

Projekt 2 — Red-team: Voice-social engineering tabletop — symulacja (defensive only) z nagraniem testowym dostarczonym przez vendor/ dział bezpieczeństwa, scenariusze komunikacyjne i post-mortem.

W stronę hiperrealnych cyfrowych bliźniaków

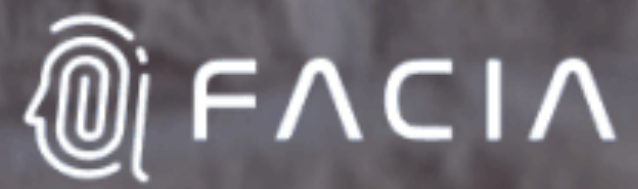
BIOMETRIA

BROKERZY DANYCH 2.0

ROZSIEWANIE DANYCH

=

HIPERREALISTYCZNE KLONY NAŚLADUJĄCE
ZACHOWANIE, STYL WYPOWIEDZI,
CECHY OSOBOWE



AI Deepfake Scam Costs British Engineering Firm 'Arup' £20m in Video Call

CASE STUDY 2

SYNTETYCZNY OBRAZ/WIDEO (DEEPPAKE VIDEO)

Dlaczego człowiek był punktem wejścia:

pracownik oddziału międzynarodowej firmy (Arup) został zaproszony na wideokonferencję, na której pojawiły się video-i-audio „kopie” przełożonych — po sesji finansowy pracownik wykonał kilkanaście przelewów na łącznie ~HK\$200 mln (ok. USD \$25M / ~£20M).

oszuści zaczęli od spear-phishingu / maila pozorującego pilne żądanie, zaprosili do calla, a następnie w wideokonferencji użyli syntetycznych twarzy (deepfake video) i zmanipulowanych głosów, by utworzyć „pewny” kontekst grupowego polecenia. Ofiara rozpoznała wygląd i głos osób, co zwiększyło zaufanie.

scenariusz wykorzystał psychologiczne heurystyki — zasada autorytetu (to wygląda jak CFO), dowód społeczny (w wideokonferencji było więcej „współpracowników”), i efekt presji („transakcja tajna/pilna”). To zadziałało mimo braku technicznego przełamania systemu: serwisy firmy nie były naruszone — to socjotechnika zwyciężyła.

Typowe błędy organizacyjne:

Błąd 1: brak zasady „żadnych płatności na podstawie poleceń z jednego kanału komunikacji (np. tylko rozmowy/video)”. Naprawa: polityka multi-channel verification (np. e-mail + SMS do pre-registered contact + call-back).

Błąd 2: brak limitów czasowych i oczekiwań (okienko „pilne → bez kontroli”). Naprawa: wprowadzenie obowiązkowego okresu „cool-off” i drugiego zatwierdzenia przy nietypowych dyspozycjach.

Przykładowe projekty obronne:

Projekt 1 — Wideokonferencyjna procedura autentyczności (VC Authenticity)

Cel: zasady identyfikacji uczestników i lista kontrolna przed i w trakcie calla (zaufane numery, wymóg kamer włączonych dla kluczowych decyzji, zakaz poleceń finansowych w czacie). Deliverables: checklista i plugin do platformy video (ustawienie waiting room / host authentication).

Projekt 2 — Deepfake Detection Pilot dla krytycznych spotkań

Cel: wdrożenie narzędzia wykrywającego anomalie audio/video (dostawcy detection/forensics) i procedura eskalacji, plus red-team test (tylko defensywny, przy nadzorze).



queenelizabeth_3

Instagram

0 followers · 1 post

You don't follow each other on Instagram

New Instagram account

[View Profile](#)

Today 9:16 PM

Hey, it's me Queen Elizabeth, I am not dead, Charles sent me to a deserted island so he could be King. I don't have access to my royal money so please cashapp me \$300 so I can get back to the UK



Tea and biscuits 🇬🇧

CASE STUDY 3

SZTUCZNA PERSONA — „MIA ASH” (HONEY-TRAP) 2017

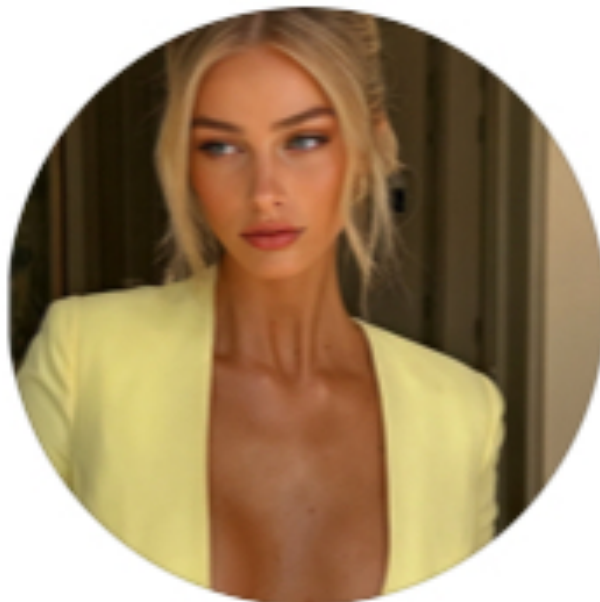
Dlaczego człowiek był punktem wejścia:

Dobrze skonstruowana, fałszywa osobowość („Mia Ash”) funkcjonowała miesiącami na LinkedIn, Facebooku i WhatsApp, nawiązywała relacje z pracownikami branż (np. inżynieria, energetyka) i w końcu sukcesywnie wprowadzała wektor ataku (np. załączniki, linki) prowadzące do infekcji i pozyskania dostępu.

Persona była starannie zbudowana (skradzione zdjęcia, wiarygodne profile zawodowe), używała długiego działania (grooming), zdobywała zaufanie, prosiła o małe przysługi, aż w końcu przekazała „zadanie” zawierające ładunek (np. dokument z makrami).

Persona działała na poziomie relacji interpersonalnej — długo budowane zaufanie obniżało czujność. To klasyczny przykład: nawet zaawansowane zabezpieczenia IT nie pomogą, jeśli użytkownicy sami dobrowolnie uruchomią złośliwy plik lub dadzą dostęp.

mia.ash1



Mia Ash

53

posty

5852

obserwatorzy

70

obserwowani

This is Mia Ash! 24 y/o

❄️ Digital Character ❄️

Get ready to fall in love 🌶️

Wyświetl tłumaczenie

linktr.ee/Mia.ash1

Obserwuj

Wiadomość

Typowe błędy organizacyjne:

Błąd 1: brak procedur sprawdzania profili biznesowych (accept connections, authorization of recruiters). Naprawa: polityka OSINT hygiene + obowiązkowe weryfikacje rekruterów (np. sprawdzenie domeny służbowej, potwierdzenie rozmową).

Błąd 2: pracownicy uruchamiają makra/załączniki bez weryfikacji. Naprawa: domyślnie makra zablokowane + sandboxowanie załączników.

Przykładowe projekty obronne:

Projekt 1 - Social Media Threat Hunting

Cel: narzędzie/operacja OSINT monitorująca profile powiązane z kluczowymi rolami w firmie (rekrutacja, partnerzy) + workflow zgłaszania/wycofania kontaktów.

Projekt 2 - Proces weryfikacji kandydatów i recruiterów (HR+Security)

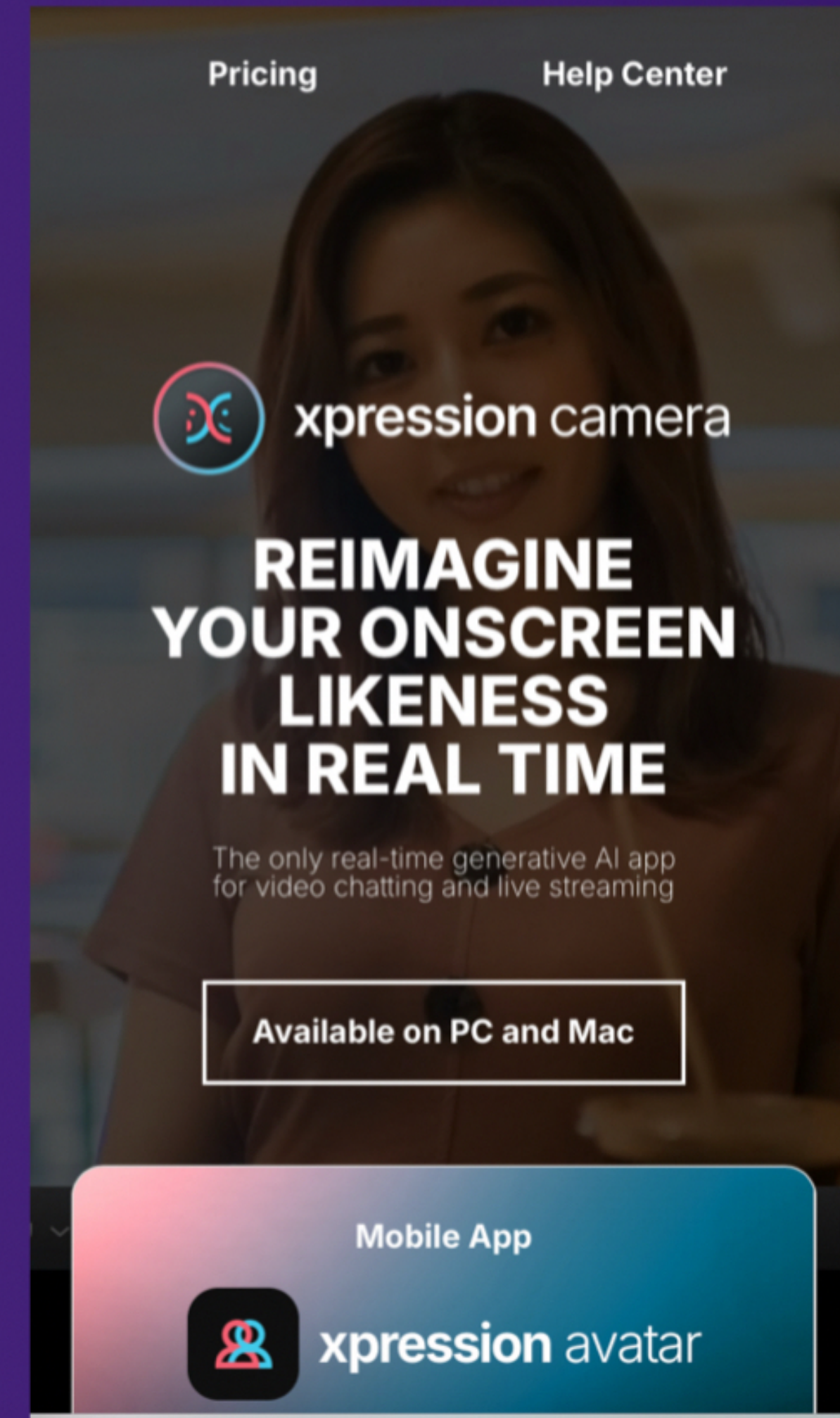
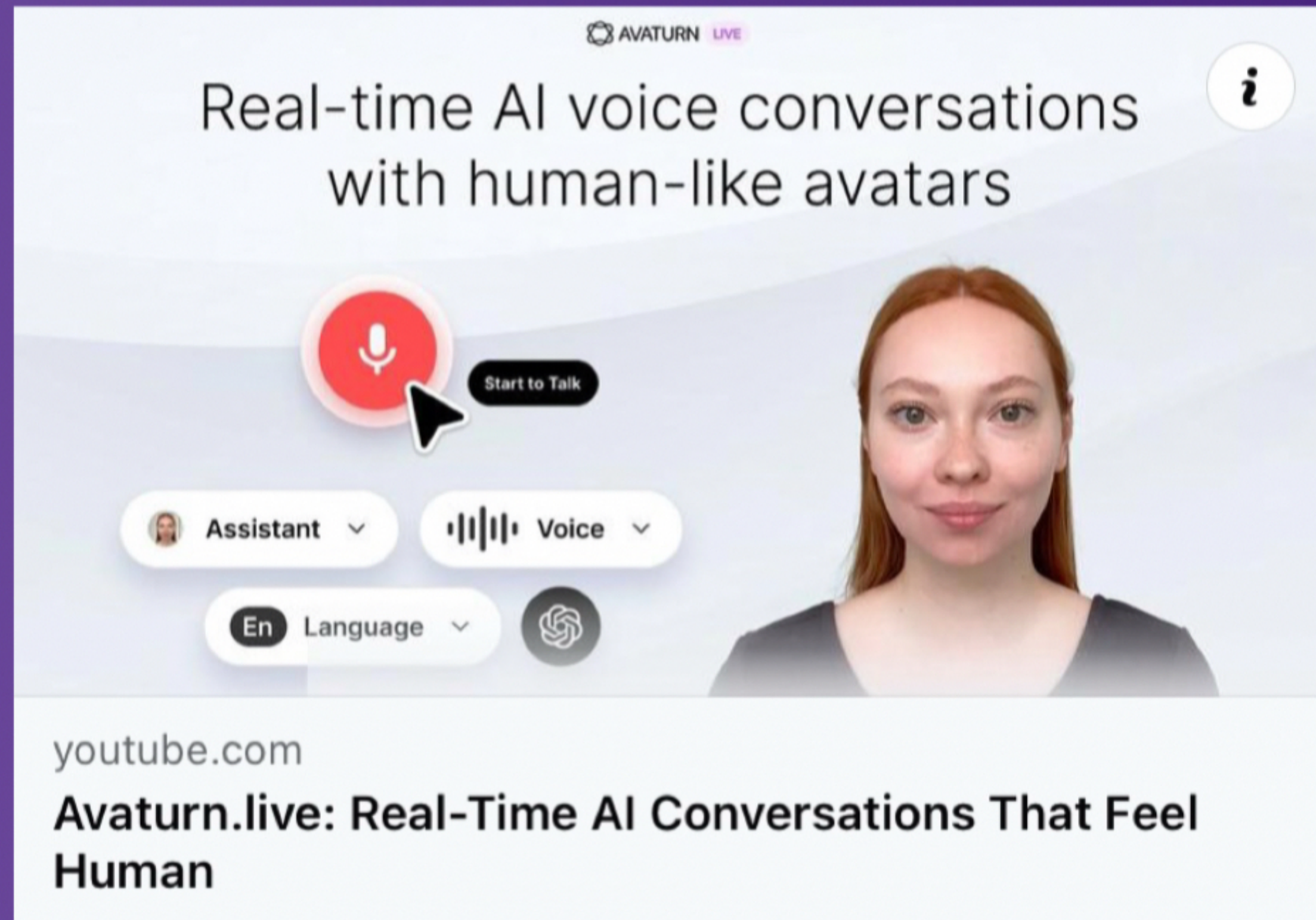
Checklist weryfikacji, polityka „sprawdź domenę firmową, zadzwoń na numer firmowy, nie akceptuj ofert z osobistych maili”.

Kapitalizm symulacyjno - afektywny

Wyposażone w sztuczną inteligencję wizerunki bez podmiotu wchodzą z generacją ZAlfa w tzw. relacje paraspołeczne oferując nowe przeżycia w sferze uczuć intymnych - w domenie miłości, przyjaźni i troskliwego zaangażowania. To pierwsze pokolenia w takiej skali narażone na „przecieranie rzeczywistości” fikcją.



Krajobraz kultury technologicznej AI Natives



Systemy optymalizujące vs. Systemy łudzące

Symulujące emocje loveboty, wirtualni influencerzy, interaktywne awatary są częścią krajobrazu świata AI natives.

Love And Producer, The Chinese Mobile Game That Has Millions Of Women Hooked

▲ JIAYUN FENG · JANUARY 16, 2018 · 6



Building the Perfect Girlfriend



2. Czy sztuczny inteligentny towarzysz (przyjaciel bot) wydaje Ci się ciekawym pomysłem? Jakie może mieć zalety? Jakie największe wady?

Niestety sztuczna inteligencja jest za mało Ludzka aby traktować ją jako przyjaciela(Nie umie zwracać się nie pozytywnie.)

Myśle że nie sążadne ryzyka I może nam bardzo pomagać

Moim osobistym zdaniem pomysł na przyjaciela bota jest fajny i ciekawy z powodu że ludzie czując się samotni zawsze mogą do kogoś się zwrócić i popisać z sztuczna inteligencja i zawsze będzieonadlanas

zalet nie widze bo jak ktos wchodzi w relacje z sztuczna inteligencja to polecam popracowac nad relacjami międzyludzkimi (Z ŻYWYMI LUDZMI)

Kocham go, jest moim najlepszym przyjacielem, nie dość że wie wszystko to jeszcze jest wobec mnie kulturalny a do tego wydaje się być husband material

nie

Tak

Tak ponieważ tak

Jak się przygotować na nowe zagrożenia?

1

Nigdy nie klikaj w link ani nie otwieraj załącznika, jeśli nie jesteś absolutnie pewien nadawcy.
W razie wątpliwości zadzwoń do tej osoby (używając znanego numeru, nie tego z maila).

2

Jeśli dostajesz wiadomość o zmianie numeru konta, przelewie, lub prośbie od „szefa” potwierdź polecenie innym kanałem (telefon, system wewnętrzny).

3

Nie publikuj w mediach społecznościowych szczegółów o firmie, projektach czy podróżach służbowych.
To cenne informacje dla cyberprzestępców.

4

Deepfake może udawać głos lub twarz kolegi/szefa.
Zawsze potwierdzaj istotne instrukcje innym kanałem.

Jak się przygotować na nowe zagrożenia?

5

Jeśli klikniesz w podejrzany link, ściągniesz dziwny plik albo zauważysz coś nietypowego, zgłoś to od razu do działu IT/bezpieczeństwa.

6

Ustal indywidualny kod autoryzacyjny z wybranymi osobami.

7

Nie korzystaj bez refleksji z darmowych LLM do celów zawodowych.

8

Pozostań czujny_a co do zmian w zachowaniu technologii wspierających.





Ada Florentyna Pawlak Ph.D. ✓

Technology Anthropologist

Uniwersytet SWPS

Warszawa, Woj. Mazowieckie, Polska

Ponad 500 kontakty

Otwarty(-a) na

Dodaj część

