

NASK

Nowa era bezpieczeństwa bankowości, analiza behawioralna oraz AI w walce z zagrożeniami NASK-PIB

Warszawa | maj 2024 r.

nask.pl



AGENDA

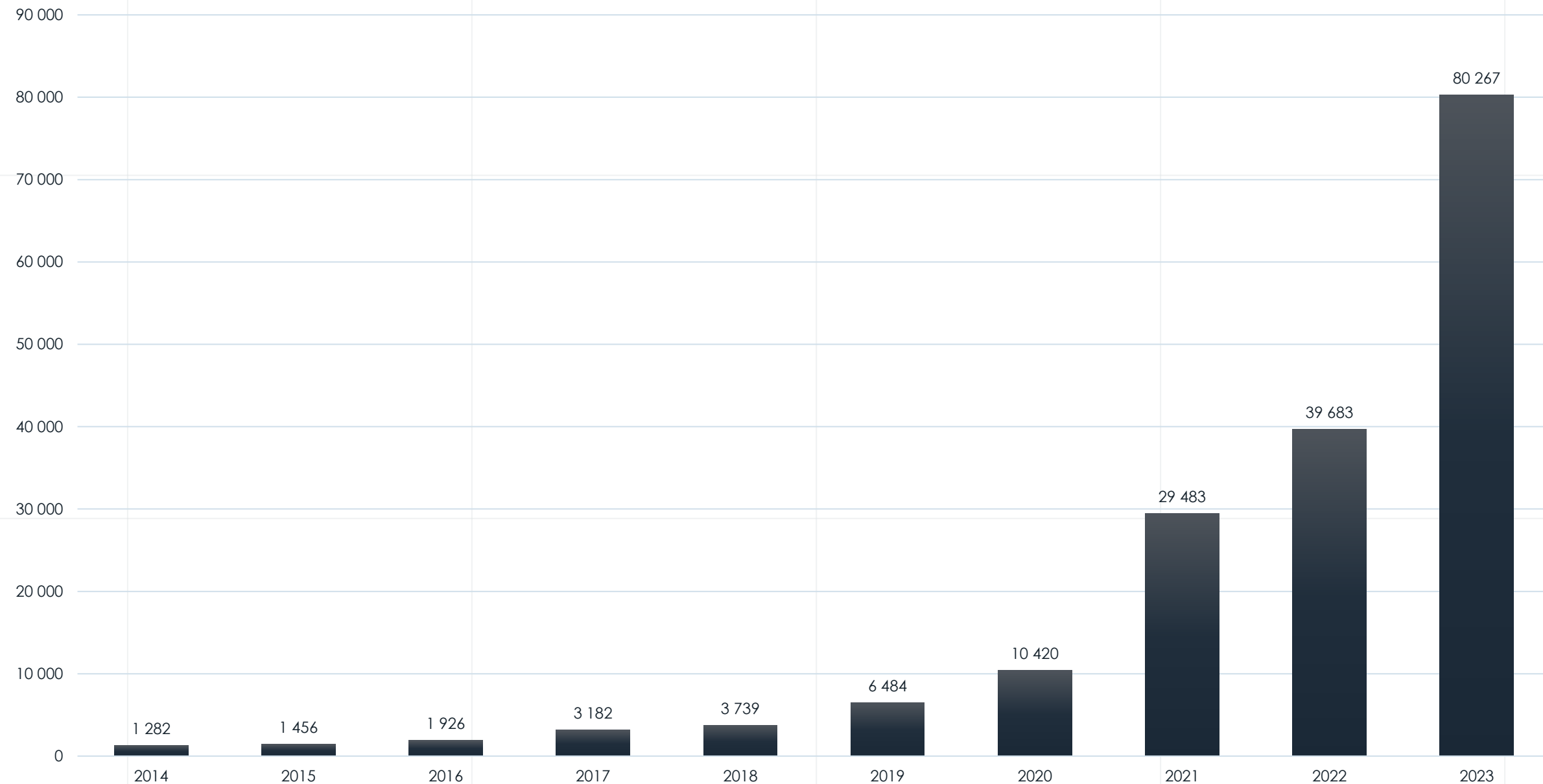
1. Statystyki zagrożeń
2. AI a DEEPFAKE
3. AI po ciemnej stronie mocy - DEEPFAKE
4. AI po jasnej stronie mocy
5. Analiza Behawioralna oraz AI w walce z zagrożeniami



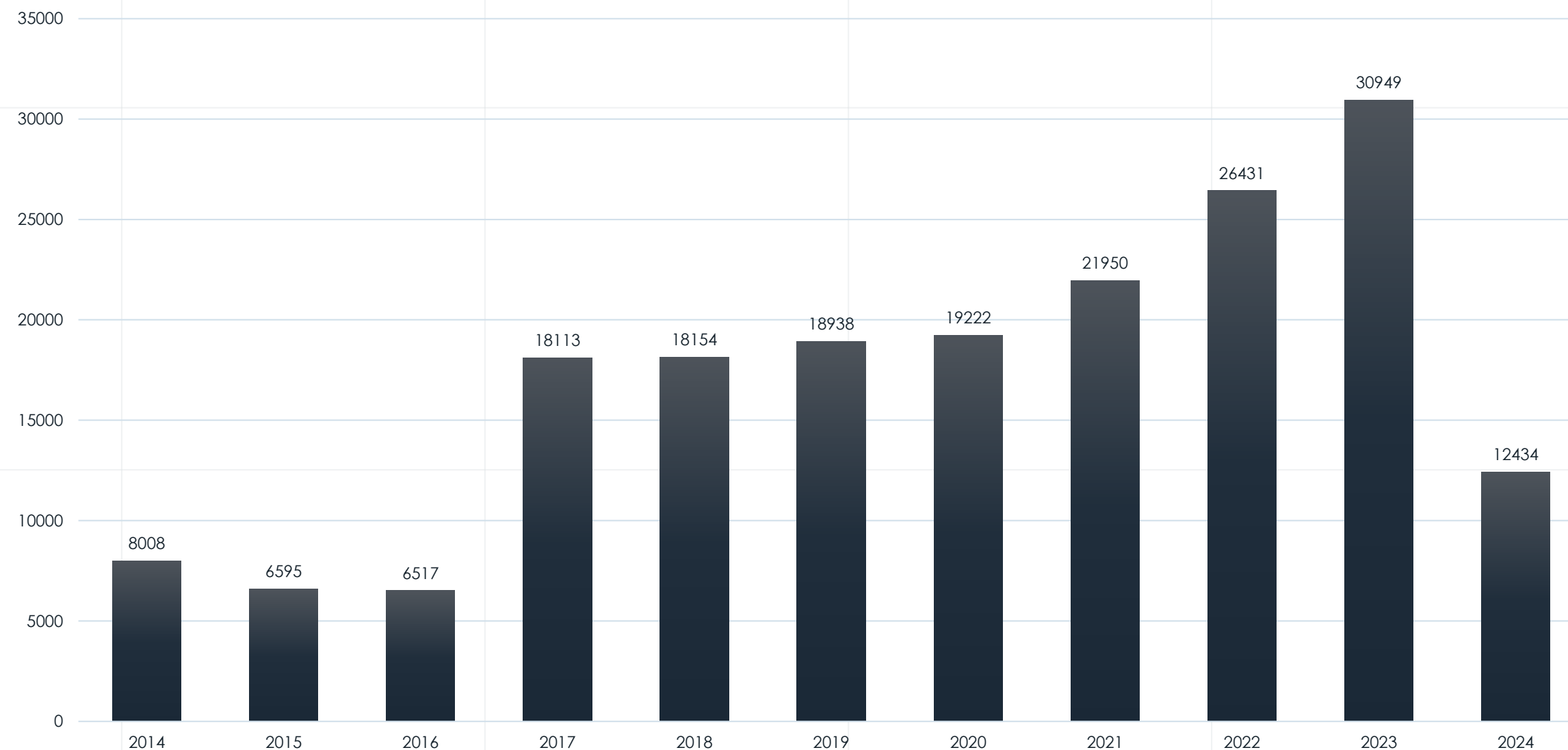
Statystyki zagrożeń



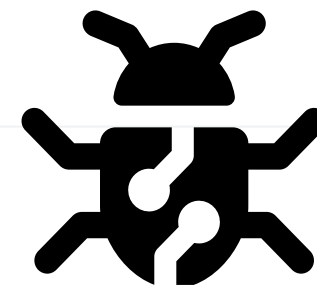
Zestawienie liczby obsługiwanych incydentów przez CERT Polska



Liczba nowych opublikowanych podatności CVE w ujęciu rocznym



2023	●	HOOKBOT, XENOMORPH, ANATSA, CHAMELEON
2022	●	HYDRA
2021	●	S.O.V.A
2020	●	BLACKROCK
2019	●	FLUBOT
2018	●	CERBERUS
2017	●	LOKIBOT



w 2023 roku

153 682

próbek

Targety malware XENOMORPH w Polsce

PackageName	AppName
pl.allegro	Allegro – convenient and secure online shopping
pl.bph	BusinessPro Lite
pl.bps.bankowosc mobilna	BPS Mobilnie
pl.bzwbk.bzwbk24	Santander mobile
pl.ceneo	Ceneo – zakupy i promocje
pl.com.rossmann.centauros	Rossmann PL
pl.envelobank.aplikacja	Pocztowy
pl.fakturownia	Fakturownia.pl
pl.ideabank.mobilebanking	Idea Bank PL
pl.ifirma.ifirmafaktury	IFIRMA – Darmowy Program do Faktur
pl.ing.mojeing	Moje ING mobile
pl.mbank	mBank PL
pl.nestbank.nestbank	Nest Bank nowy
pl.noblebank.mobile	Noble Mobile
pl.orange.mojeorange	Mój Orange
pl.pkobp.iko	IKO
pl.raiffeisen.nfc	Mobilny Portfel
pl.sgb.wallet	PORTFEL SGB

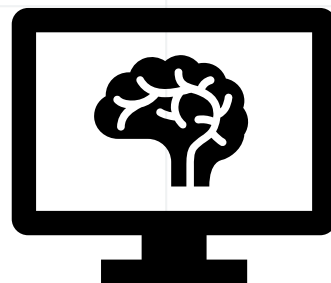


AI α DEEPPFAKE

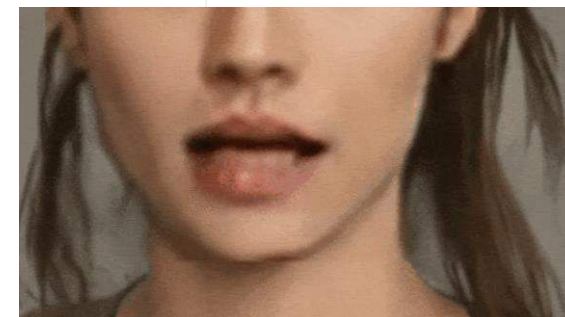
Czym jest DEEPFAKE



Technika
manipulacji zdjęcia
lub nagrania



przy użyciu sztucznej
inteligencji,



w których twarz lub głos
osoby są zastąpione
inną tożsamością.

Dlaczego DEEPFAKE są tak skuteczne i będą stanowiły coraz większy problem?



większa wiara **w to co widzimy i słyszymy**, ograniczone zaufanie dotyczy raczej tekstu



przyzwyczailiśmy się do przekazu **wiadomości** w formie **audiowizualnej**



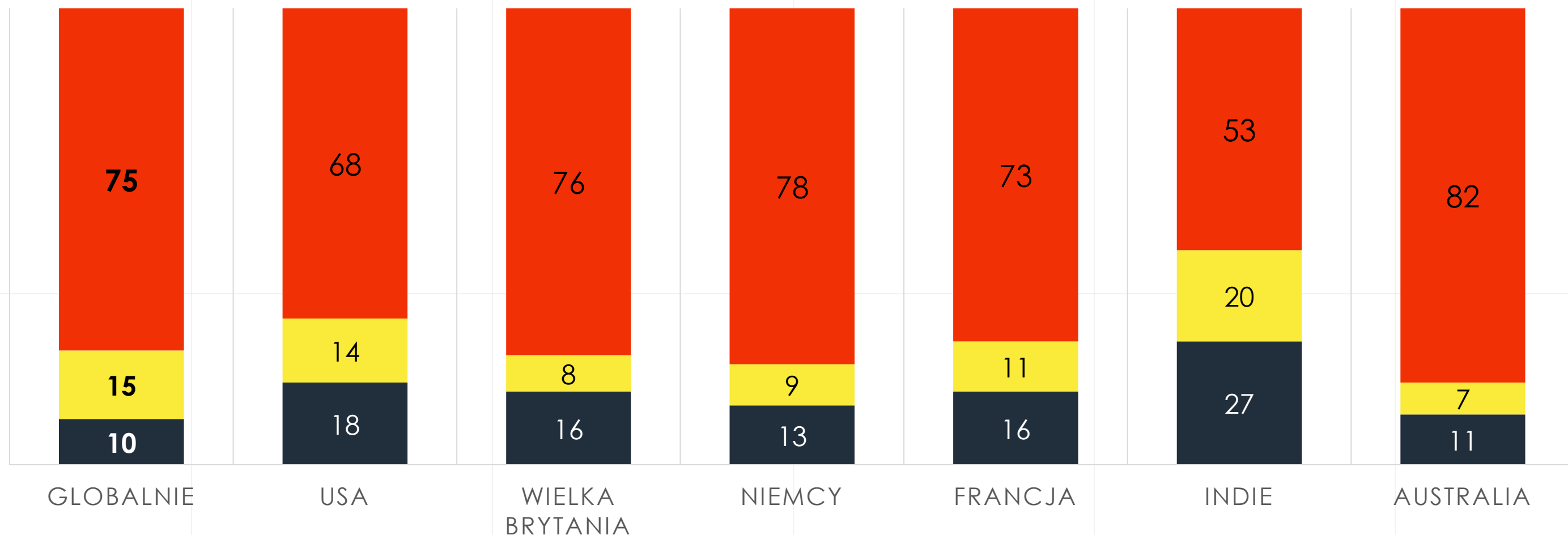
rozwija się coraz **więcej metod generowania deepfake**, a **internet** staje się **najlepszym medium** informacyjnym



coraz łatwiej trafić na deepfake'i w sieci → **więcej frontów ataku**.

Czy Ty lub ktoś, kogo znasz, doświadczył oszustwa związanego z generowanym głosem przez AI?

■ Byłem ofiarą ■ Znam ofiarę ■ Nie



AI po ciemnej
stronie mocy -
DEEPFAKE



Ataki personalne

- imitowanie głosu/wizerunku
- podrabianie wizerunku pracowników/prezesów firm
- zniesławienie
- podważanie zaufania
- reprodukcja rękopisów
- zastraszanie/szantaż

AI-generated child sex images spawn new nightmare for the web

Investigators say the disturbing images could undermine efforts to find real-world victims



The Twitch Community Has Been Rocked By A Deepfake Porn Scandal. "It Was Quite Horrifying," One Victim Said.

"I was wishing for eye bleach," streamer Sweet Anita told BuzzFeed News.



Steven Asarch
BuzzFeed Contributor

Posted on February 6, 2023 at 4:45 pm

North America

Teen boys at New Jersey school accused of creating AI deepfake nudes of female classmates



Style examples	HWT (Ours)	GANwriting	Davis et al.
<i>It's good neighbors to those left can who will (on) times to live in shows it and not others it</i> The process has been too slow for Herr Strauss and last month he attacked Britain for being in Three more loud cries of 'shame' from all parts of the Conservative side Mr. Hall appeared to be in He thought he said, 'I of the South division would be, you need to reach an agreement by 10:30 of Mr. Macleod went on with the conference at Lancaster House despite the crisis which had blown By the end of the month he still delighted in Naples he told Chambers that he enjoyed it all	No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints No two people can write precisely the same way just like no two people can have the same fingerprints	No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints No two people can write precise the same way just like no two people can have the same fingerprints	No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints No two people can write precisely the same way, just like no two people can have the same fingerprints

Ataki społeczne

- szerzenie dezinformacji
- osłabianie lub niszczenie stosunków międzynarodowych
- zachwianie bezpieczeństwa narodowego
- potwierdzenie teorii spiskowych

Ukraińska armia: Rosjanie stosują technologię deepfake, by poróżnić ukraińskie społeczeństwo

🕒 2023-11-08 18:13 aktualizacja: 2023-11-09, 09:16

🔗 Udostępnij przez



Kolejny deepfake zebrał żniwa. Tym razem celem był Joe Biden

publikacja
2024-02-26 22:24

Steve Kramer, konsultant sztabu Deana Philipsa, rywala prezydenta USA Joe Bidena w prawyborach Demokratów, przyznał się do skonstruowania i użycia deepfake'a, który udawał głos Bidena i zniechęcał do wzięcia udziału w styczniowych prawyborach w stanie New Hampshire.

Spotkanie Władimira Putina z członkami rządu, fot. PAP/EPA/GA

Rosjanie stosują technologię deepfake ukraińskich władz i dowództwa podbu podało w środę Centrum Komunikacji



Oszustwa finansowe

- Bezpośrednie:
 - imitowanie głosu – danej osoby, wnuczka, córki ...
 - przełożonego, prezesa, kolegi z pracy
 - funkcjonariusza służb
 - celebrytów/autorytetów

Company worker in Hong Kong pays out £20m in deepfake video call scam

Police investigate after employee says she was tricked into sending money to fraudsters posing as senior officers at her firm



Fraudsters Cloned Company Director's Voice In \$35 Million Heist, Police Find

Thomas Brewster Forbes Staff

Oszustwa finansowe

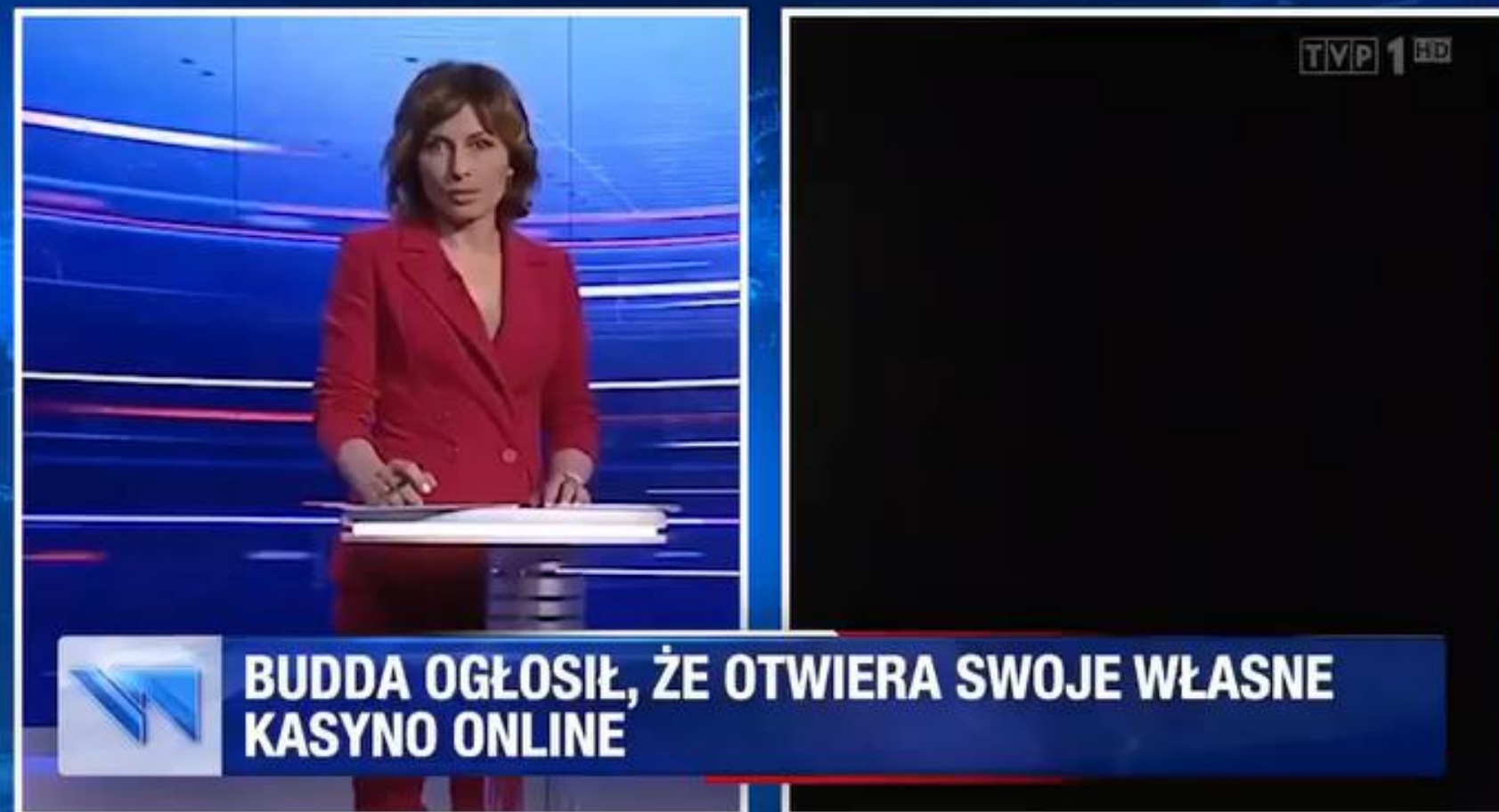
- Bezpośrednie:
 - imitowanie głosu – danej osoby, wnuczka, córki ...
 - przełożonego, prezesa, kolegi z pracy
 - funkcjonariusza służb
 - celebrytów/autorytetów
- Pośrednie:
 - fałszywe reklamy inwestycji



https://www.knf.gov.pl/dla_ryнку/CSIRT_KNF

Oszustwa finansowe

- Bezpośrednie:
 - imitowanie głosu – danej osoby, wnuczka, córki ...
 - przełożonego, prezesa, kolegi z pracy
 - funkcjonariusza służb
 - celebrytów/autorytetów
- Pośrednie:
 - fałszywe reklamy inwestycji



Najbogatszy polski YouTuber, który płaci rachunki setkom tysięcy ludzi,

**AI po jasnej stronie
mocy**



Rozwiązywanie przestępstw



<https://odishabhaskar.in>

Jak AI pomogła?

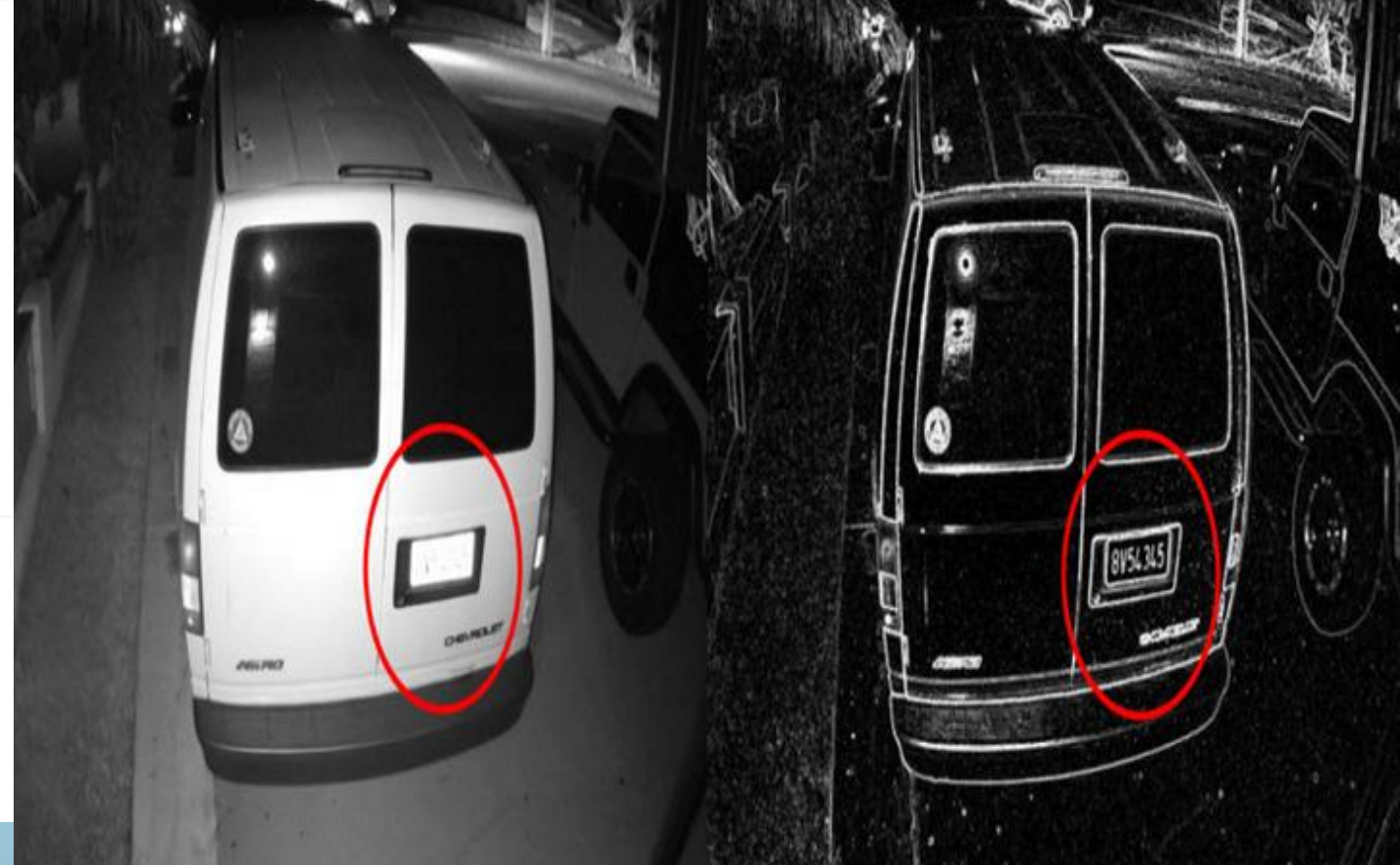
- zrekonstruowało twarz ofiary
- zidentyfikowano ofiarę
- znaleziono sprawców



<https://www.ndtv.com/>

Odtwarzanie jakości

- zwiększanie jakości obrazu twarzy
- wzmacnianie niewidocznych elementów



<https://www.hitpaw.com/video-tips/cctv-video-enhancement.html>



NASK

<https://researchgate.net/post/CCTV-is-there-a-way-or-a-new-method-of-improving-the-image-quality-for-facial-recognition>

Bezpieczeństwo osób

- ochrony tożsamości osób, które są w niebezpieczeństwie
- w sytuacjach, gdy świadek nie może zeznawać osobiście, np. z powodów bezpieczeństwa, deepfake może umożliwić stworzenie wiarygodnego wizerunku mówiącego świadka



► Podmiana twarzy w filmie „Witamy w Czeczenii”, HBO

Identyfikacja przestępców

- przez „podróże” w czasie sztuczna inteligencja jest w stanie robić predykcje, czyli odmładzać lub postarzać ludzi

W Niemczech trwa walka z rodzimym terroryzmem. Daniela Klette z RAF ujęta

Policja ciągle poszukuje członków Frakcji Armii Czerwonej (RAF), najbardziej krwawej organizacji powojennych Niemiec. Na celowniku są też inne grupy.

Publikacja: 29.02.2024 21:14



<https://www.rp.pl/przestepczosc/art39926931-w-niemczech-trwa-walka-z-rodzimy-terroryzmem-daniela-klette-z-raf-ujeta>

Inne zastosowania

- **odtworzenie scen przestępstw**

- deepfake może być używany do symulacji i odtworzenia sceny przestępstwa, co może pomóc w zrozumieniu mechanizmów działania przestępców

- **analiza zachowań**

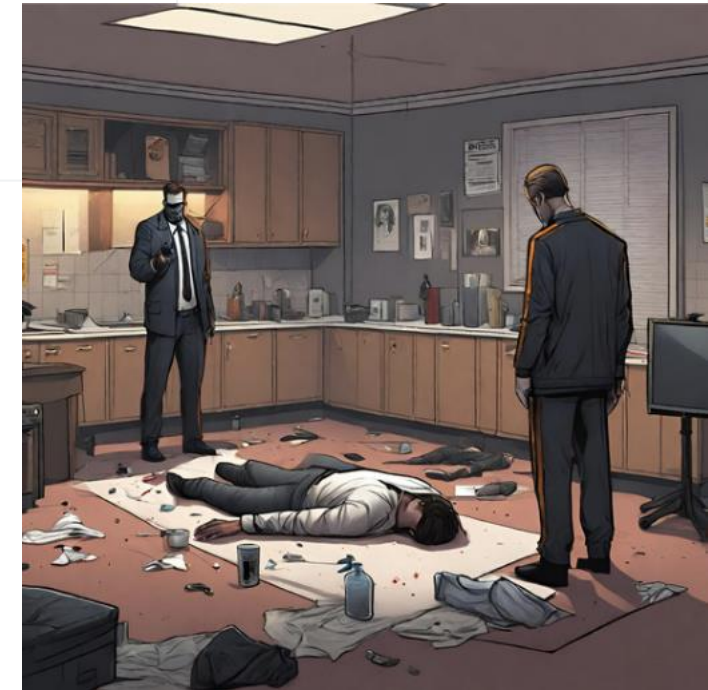
- wykorzystując deepfake do tworzenia wideo z różnymi reakcjami i ekspresjami twarzy, eksperci mogą analizować potencjalne zachowania podejrzanych w różnych sytuacjach

- **rekonstrukcja postaci**

- w przypadkach, gdy potrzebne jest odtworzenie wyglądu nieznanej ofiary lub podejrzanego, deepfake może pozwolić na tworzenie potencjalnych obrazów osób na podstawie dostępnych opisów świadków

- **szkolenie personelu**

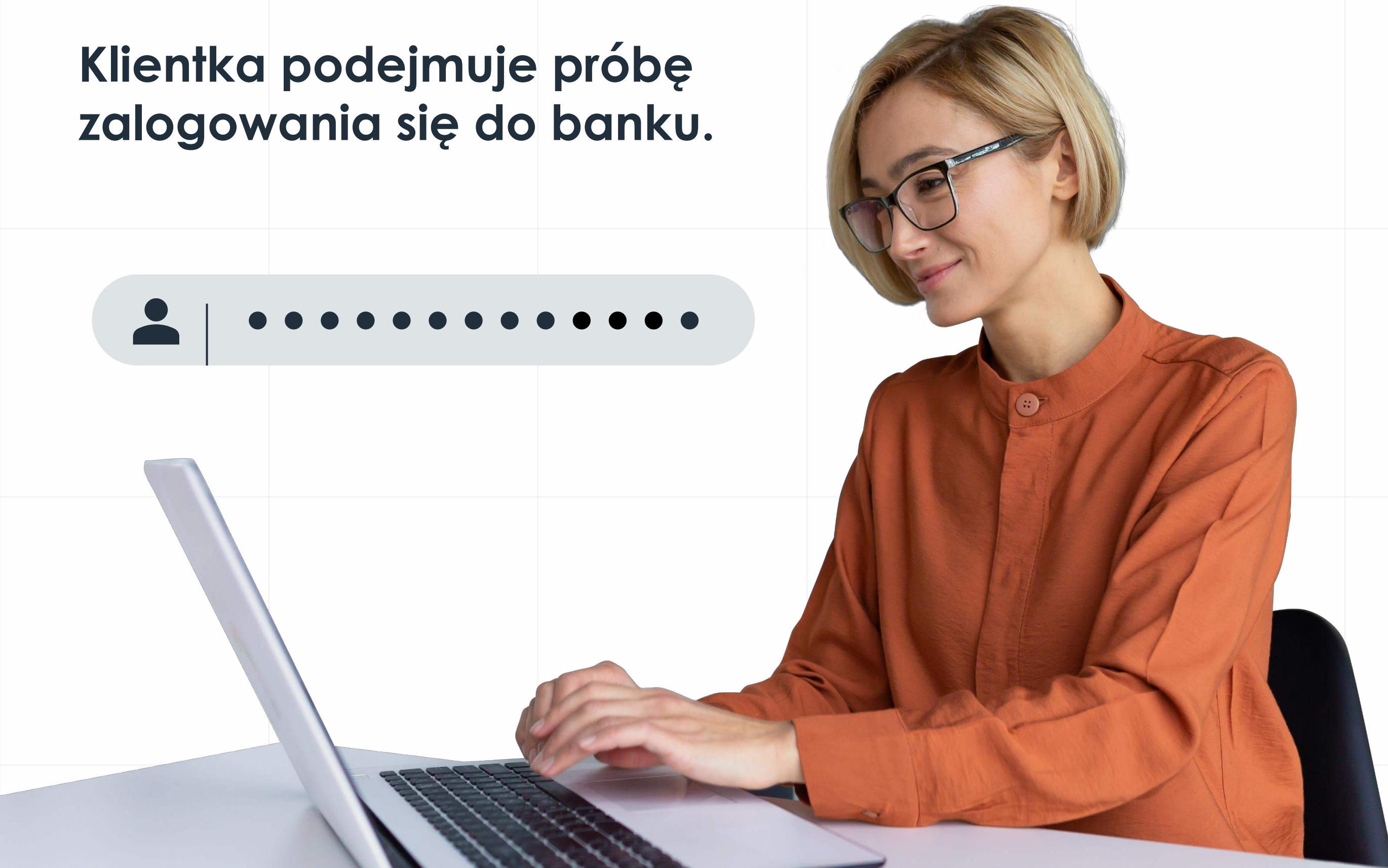
- deepfake może być wykorzystywane do tworzenia symulacji dla szkoleń agentów w celu poprawy ich umiejętności identyfikacji fałszerstw i manipulacji



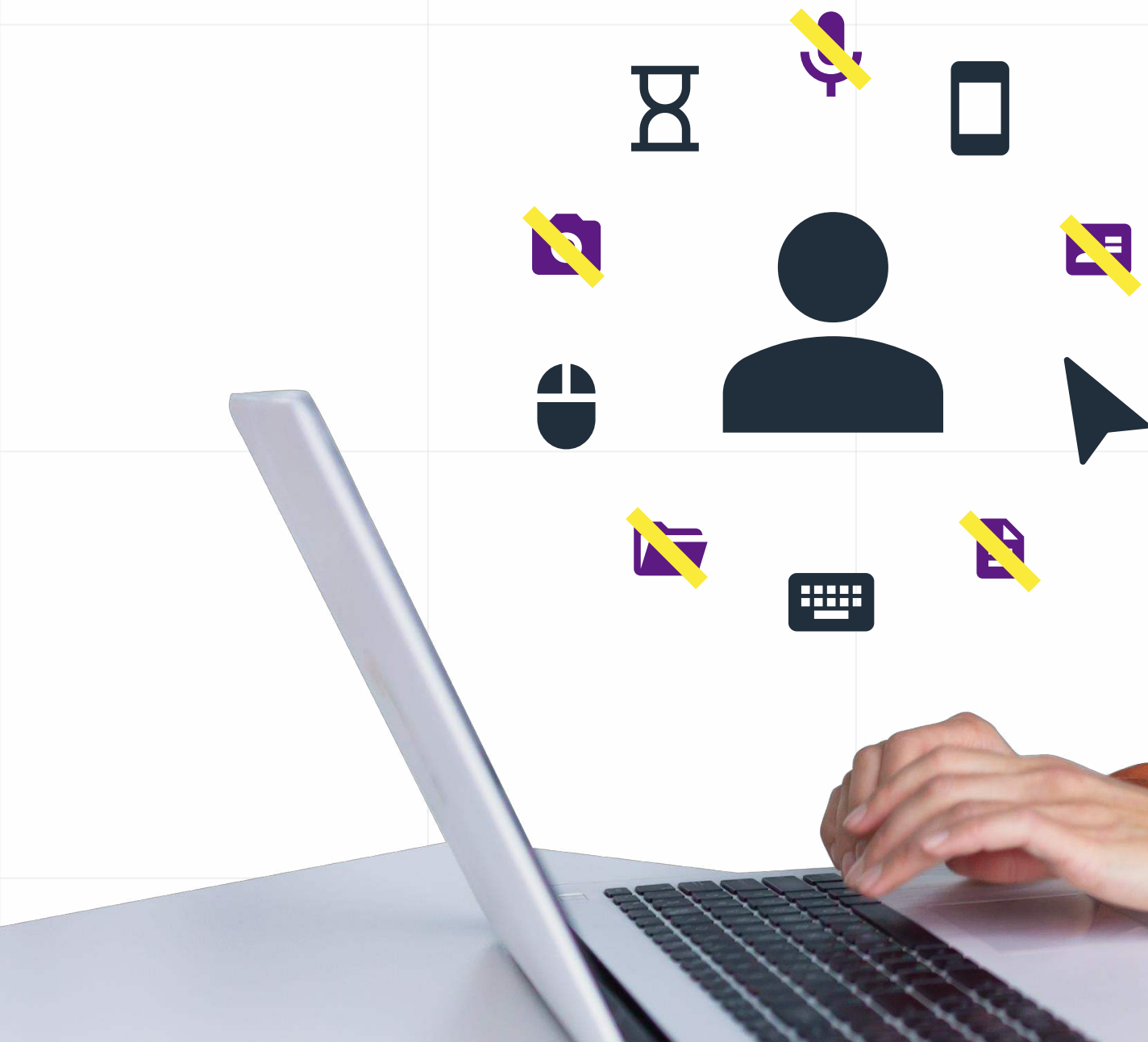
**Analiza
behawioralna oraz
AI w walce z
zagrożeniami**



**Klientka podejmuje próbę
zalogowania się do banku.**

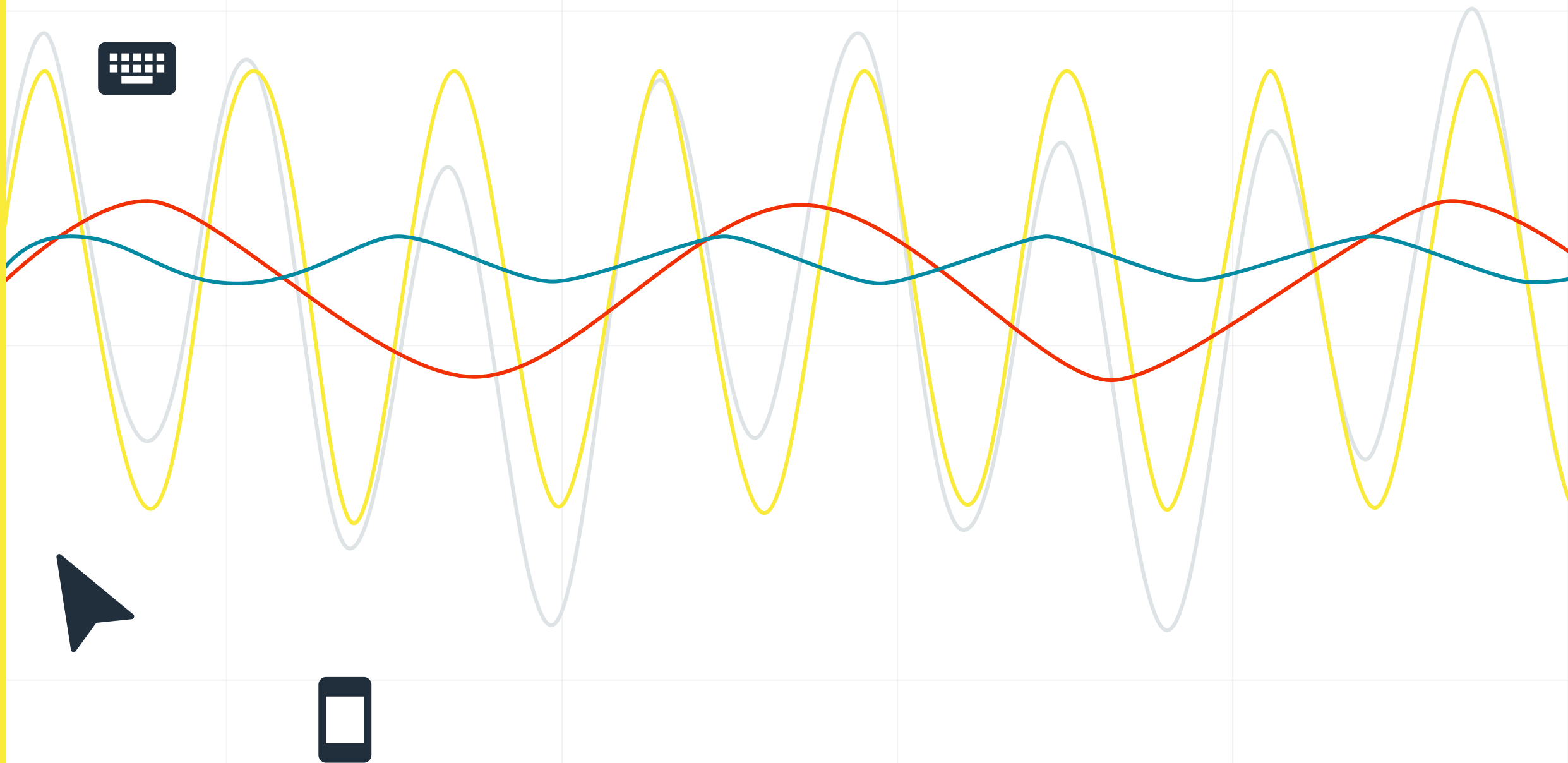


Próbki jej danych behawioralnych...





...przetwarzane są tak, aby...

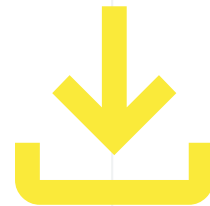




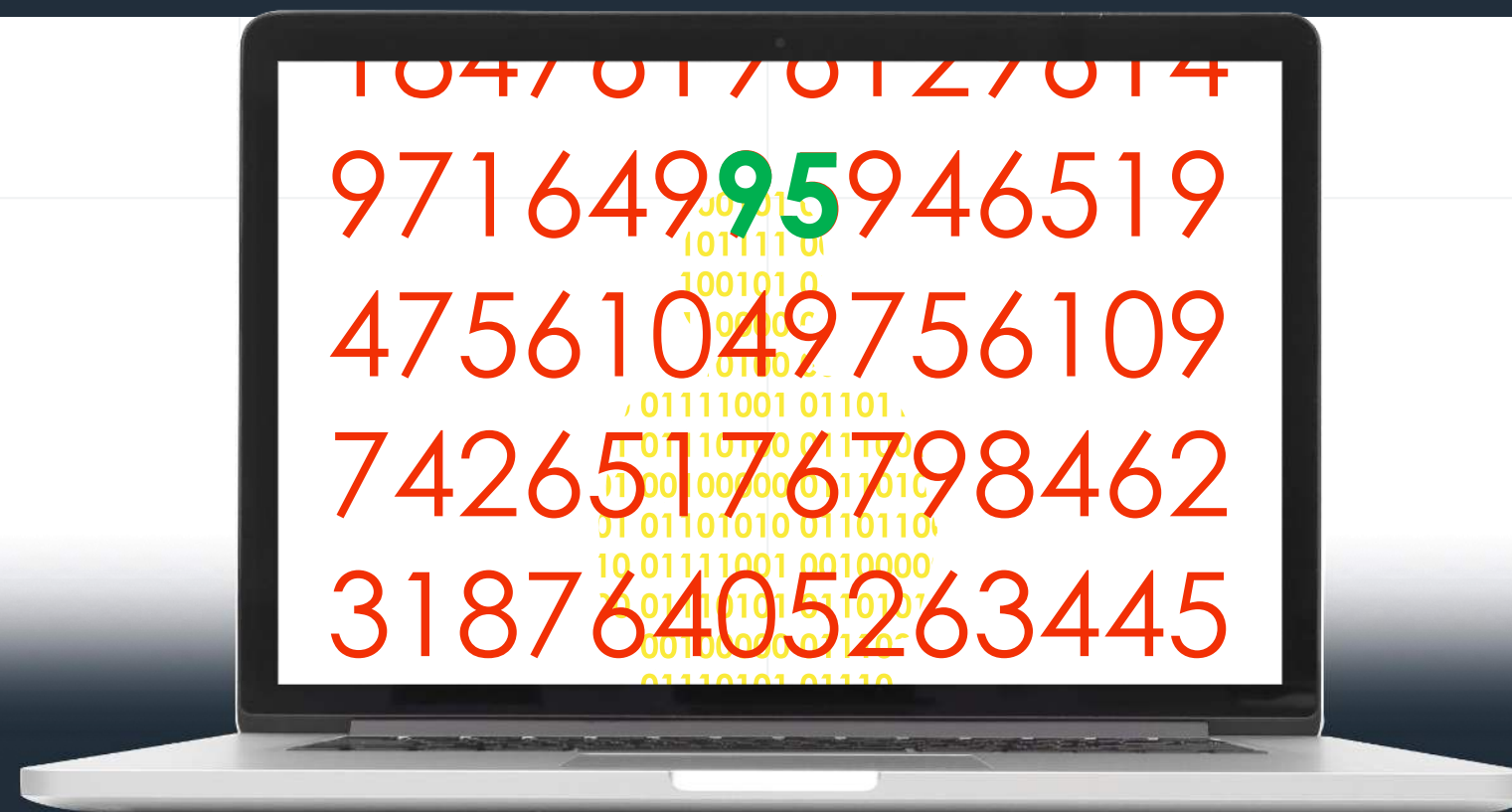
...zbudować unikatowy
wzorzec behawioralny,...

NASK

...który jest szyfrowany i zapisywany
tylko na jej urządzeniu.



Z oceny nowego logowania
powstaje **scoring...**



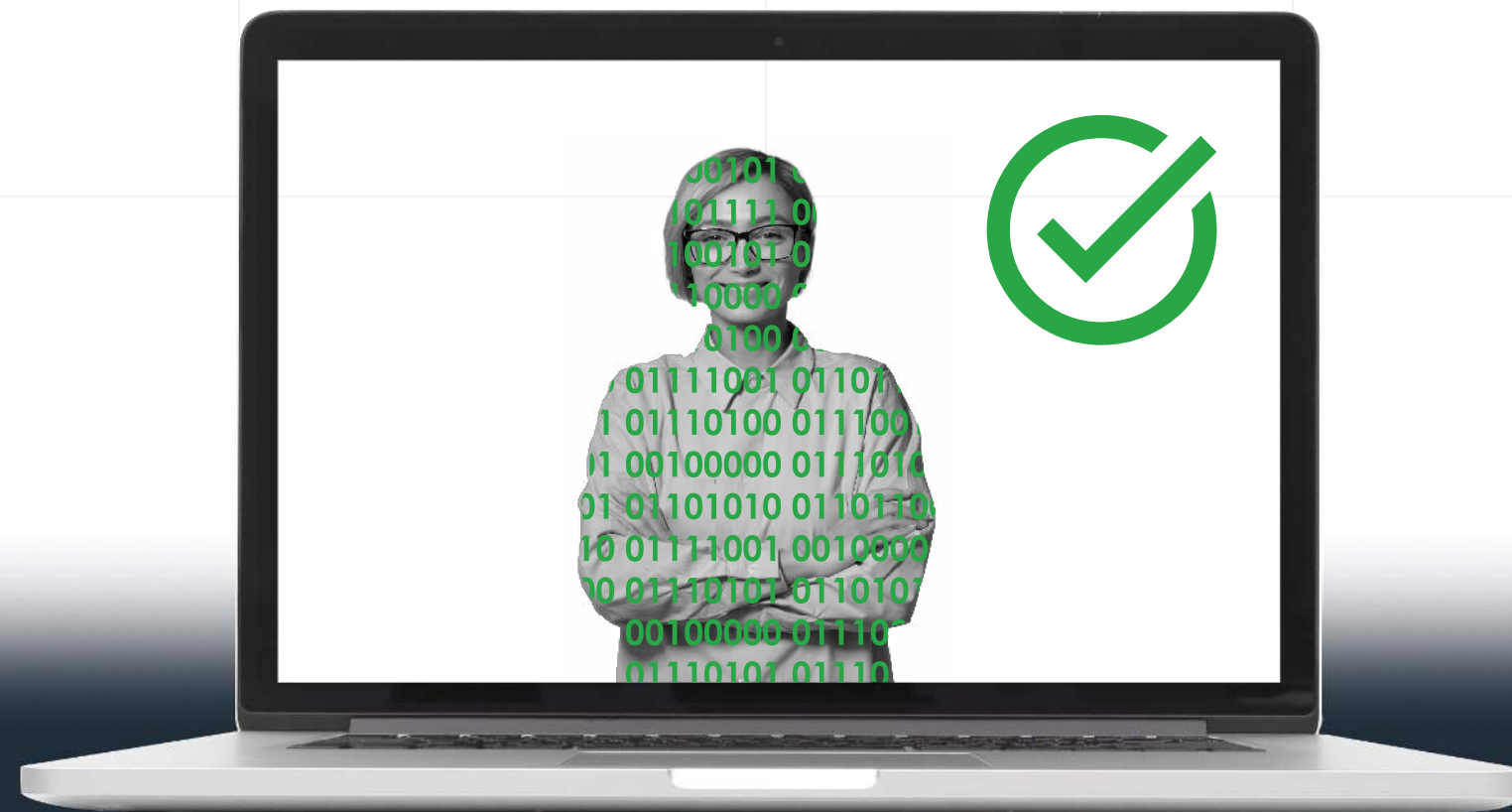
...który trafia do banku w formie zaszyfrowanych danych.



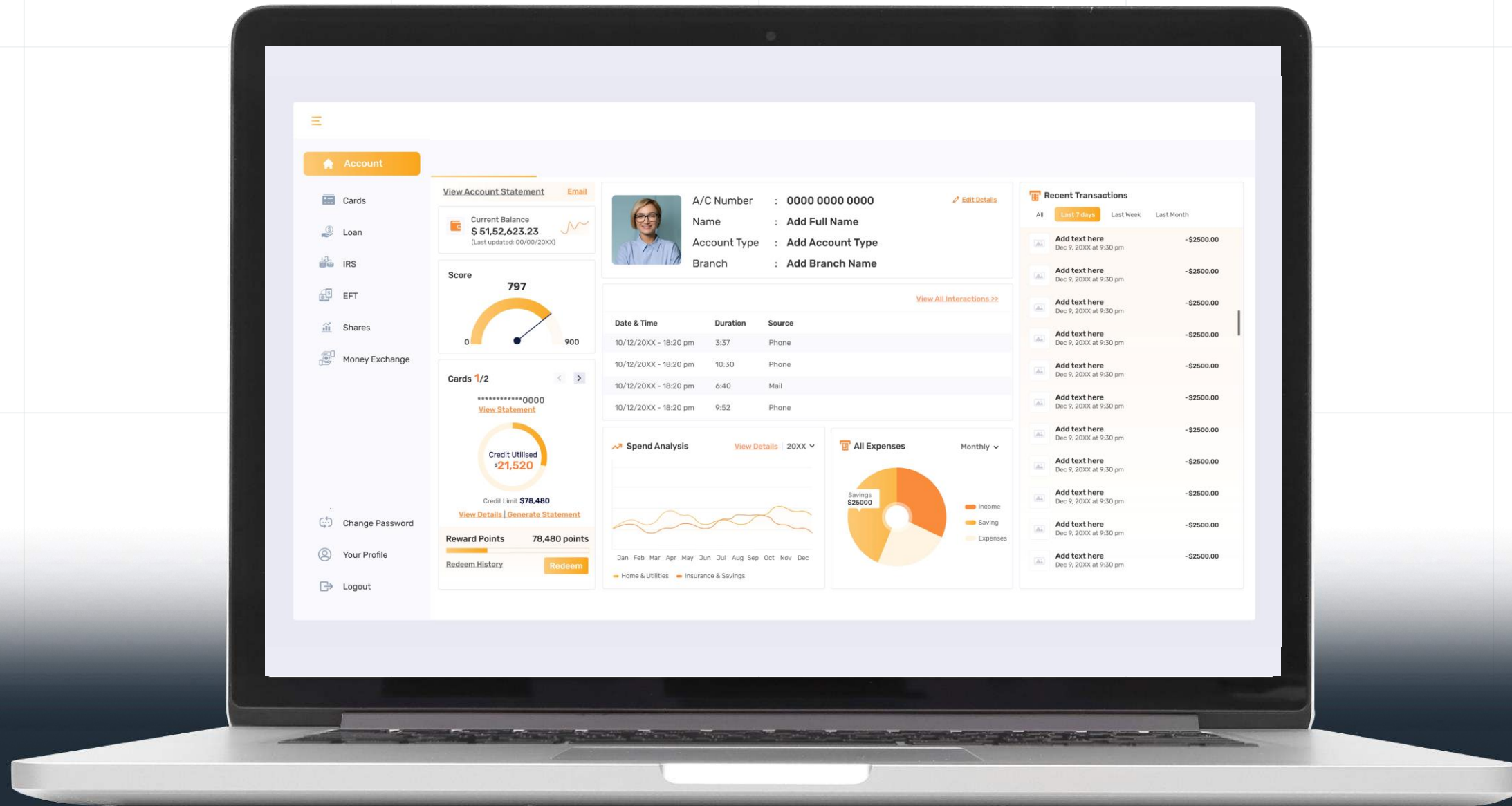
Scoring wspomaga decyzję
po stronie banku.



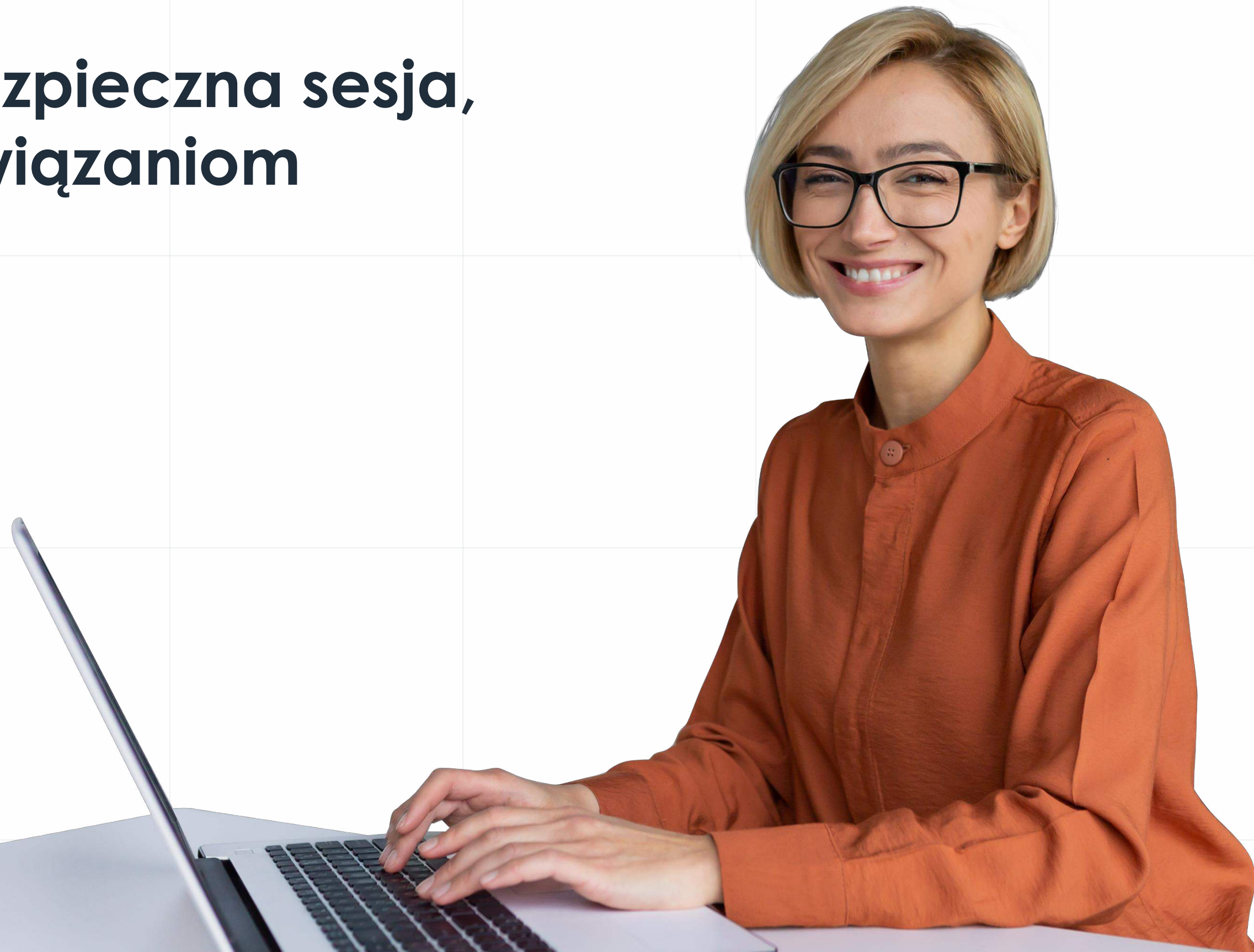
Na podstawie otrzymanego
scoringu bank **odrzuci**
...lub **autoryzuje** sesję.



Wtedy sesja logowania do bankowości zakończy się powodzeniem.

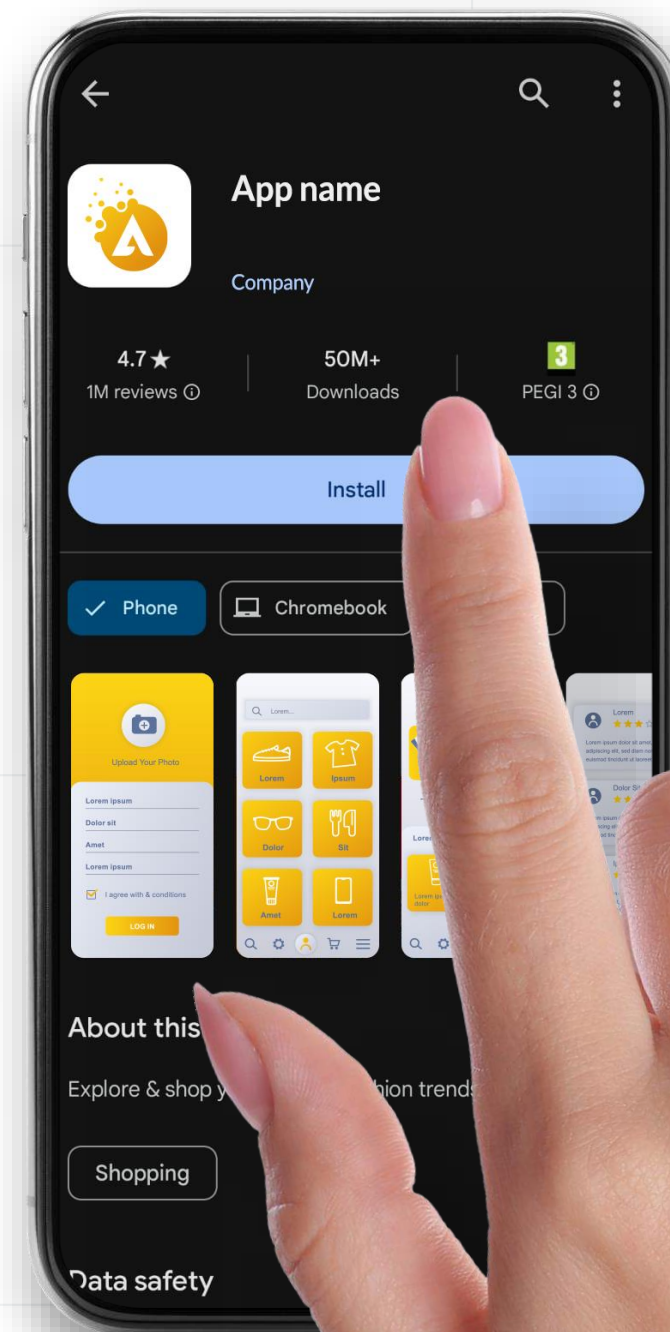


**To była bezpieczna sesja,
dzięki rozwiązaniom
BotSense!**



**Wyobraźmy sobie,
że Klientka chce
zainstalować aplikację.**

NASK

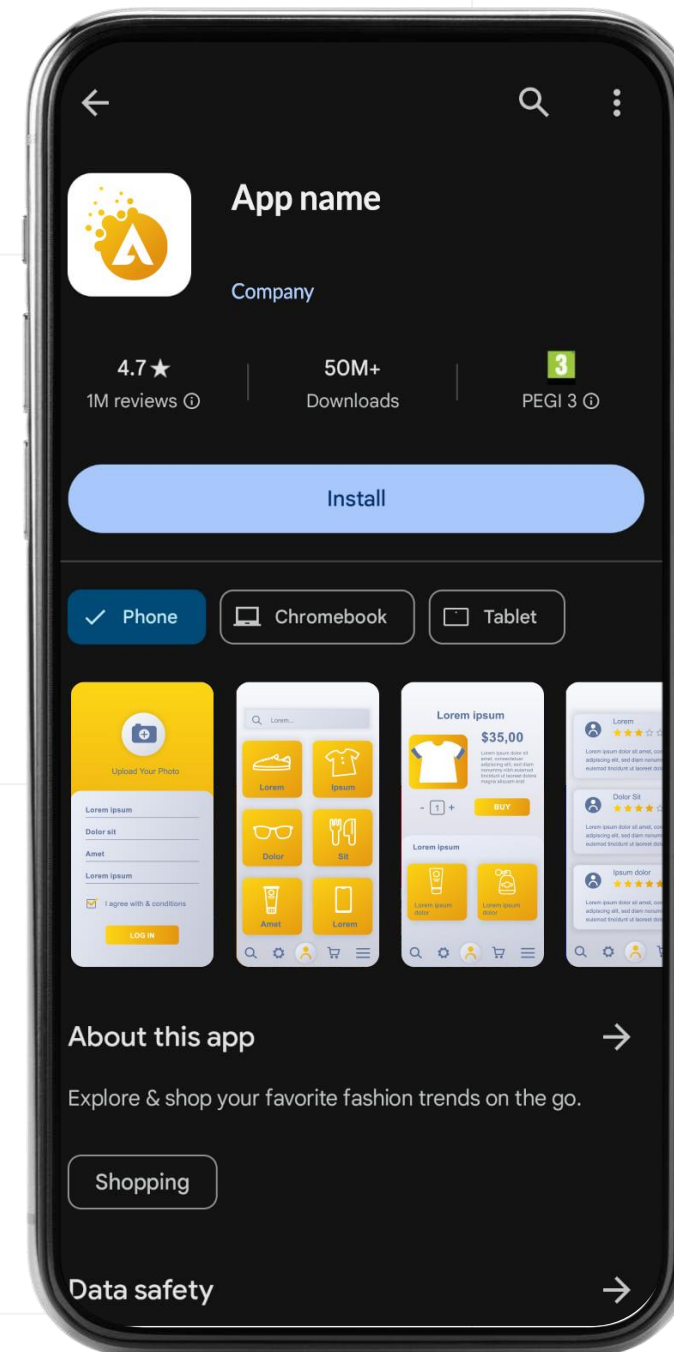




**W trakcie instalacji smartfon
sprawdza zagrożenia
na podstawie znanych sygnatur.**

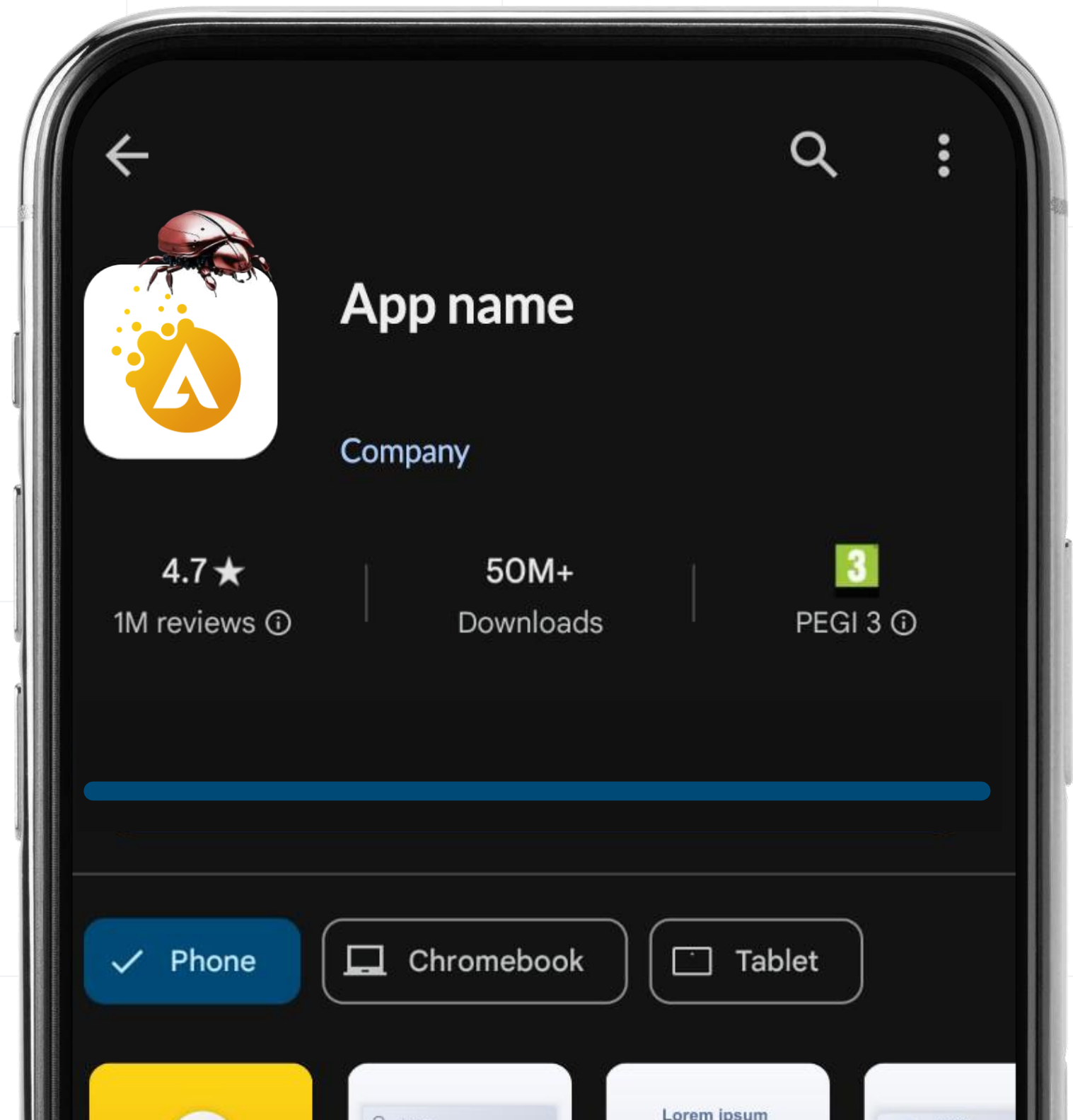
Na pierwszy rzut oka
aplikacja wygląda
bezpiecznie...

NASK



...więc smartfon
kontynuuje instalację,
choć aplikacja może
być zawirusowana.

NASK



...zainfekowana aplikacja
przesyła dane ze smartfona
do **cyberprzestępcy**
podczas otwartej sesji.



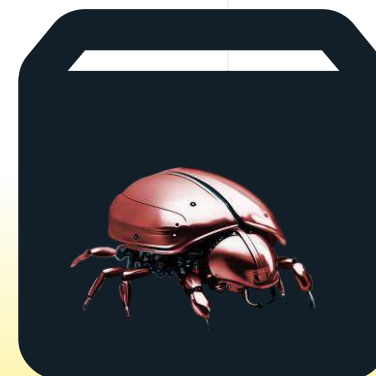
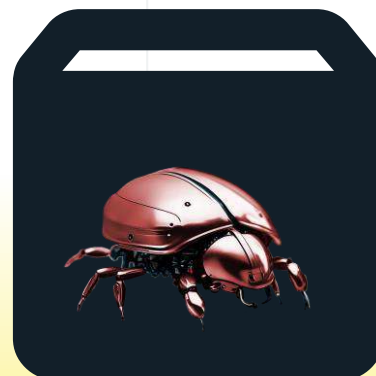
**Jak uchronić Klientów
przed takim
scenariuszem?**

**Wykorzystujemy
silnik Malware Hunter AI,
który analizuje smartfon
w trakcie otwartej sesji
logowania...**

NASK



**...i przesyła do banku
informację o nowo rozpoznanym zagrożeniu.**



Bank weryfikuje informacje
i odsyła odmowę próby logowania.

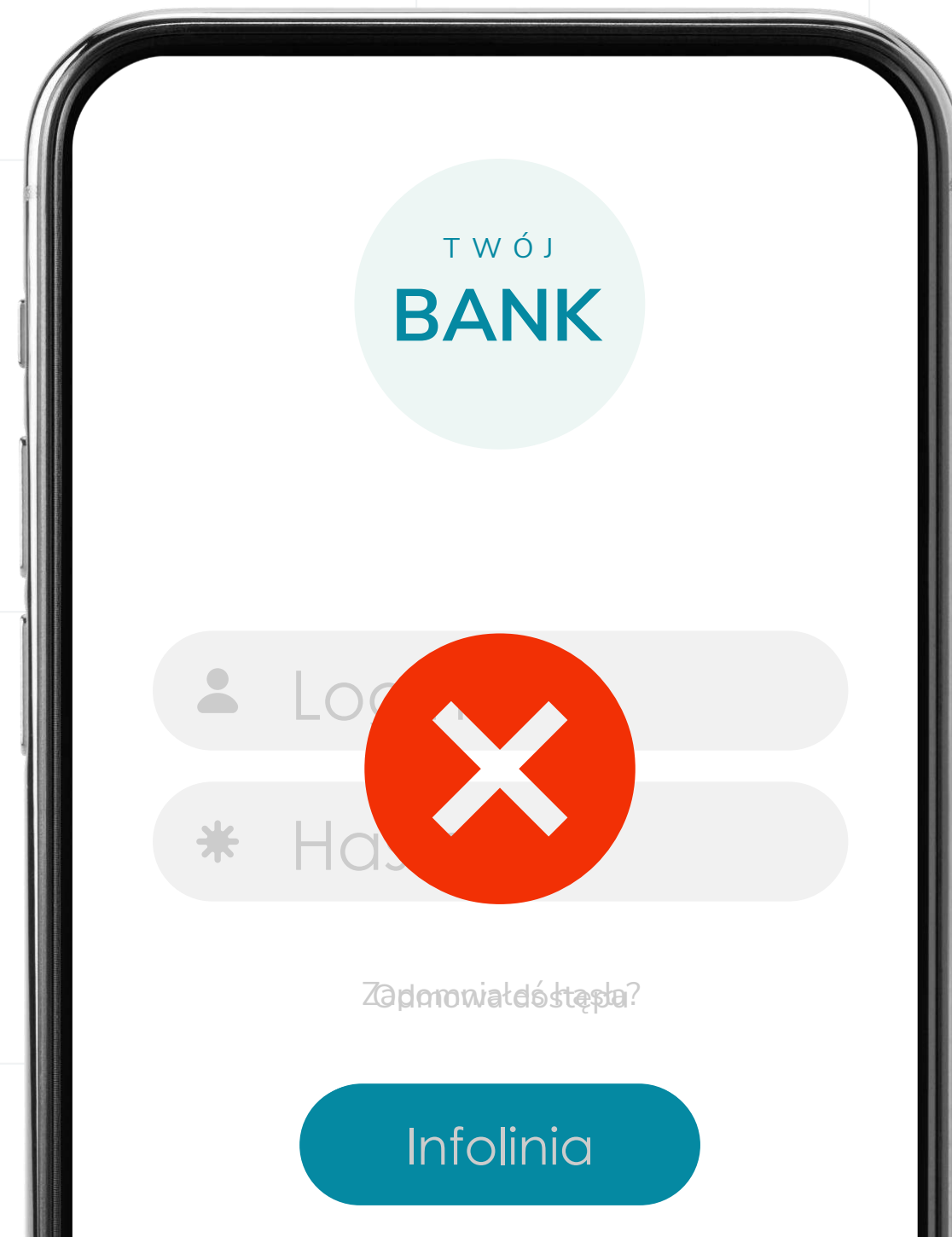


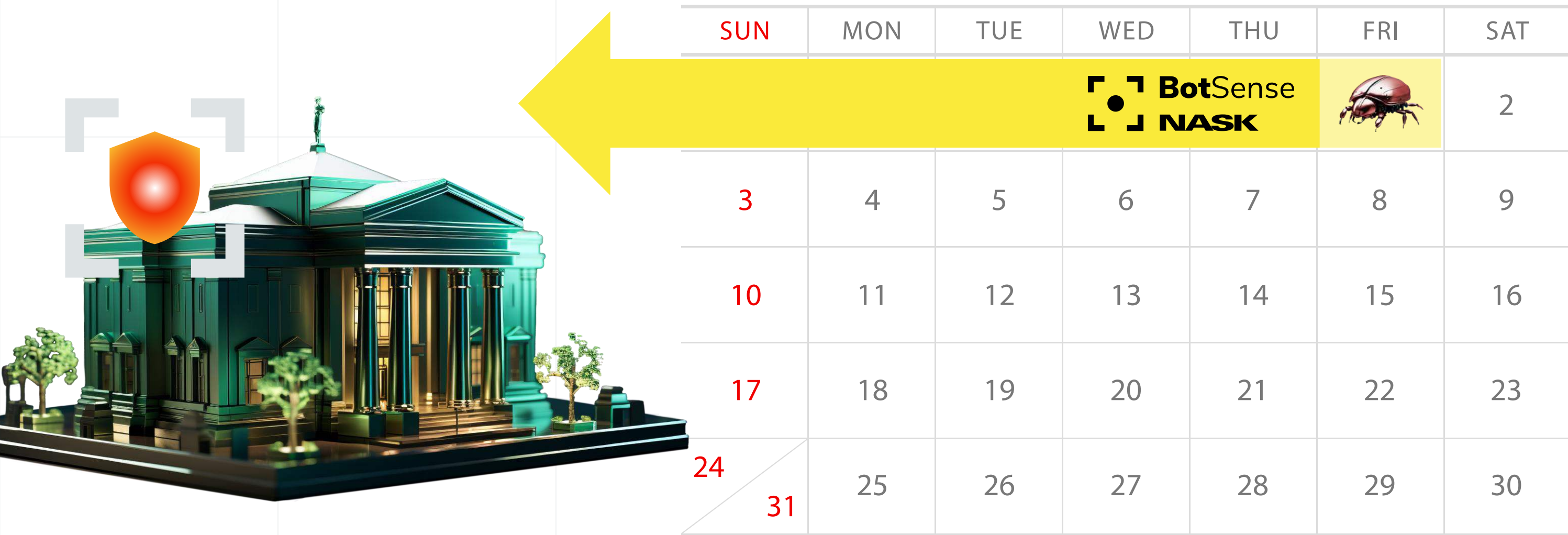
Chroniony przez



**Smartfon odbiera informację
z banku i nie zezwala
na logowanie.**

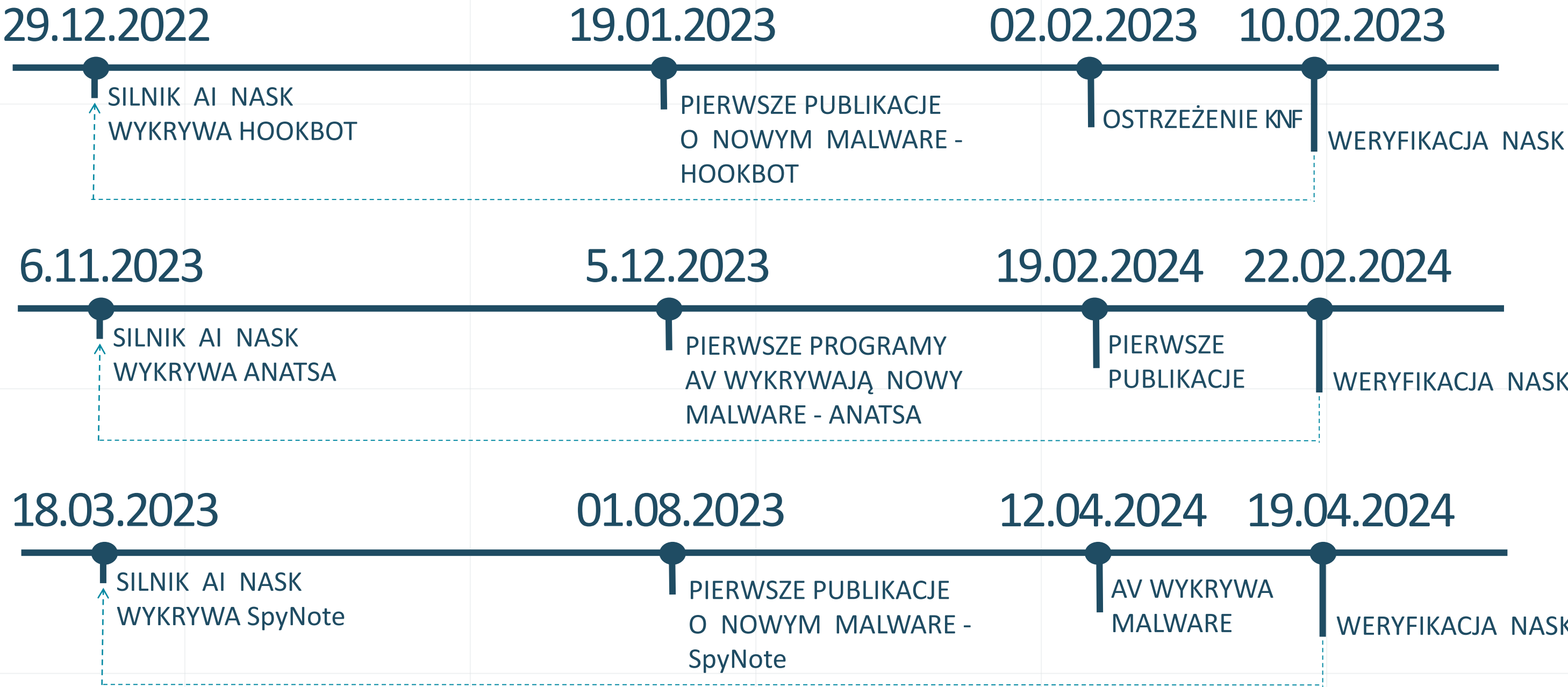
NASK





Rozpoznajemy zagrożenia do kilkudziesięciu dni szybciej niż reszta świata.

CASE STUDY – MALWARE HUNTER AI



NASK

Dziękuję

Piotr Kot
Manager Rozwoju Biznesu

+48 608 564 711
piotr.kot@nask.pl

nask.pl

