

Temat 7 .

dr inż. Michał Malinowski

<https://www.drmalinowski.edu.pl>

Cyberbezpieczeństwo

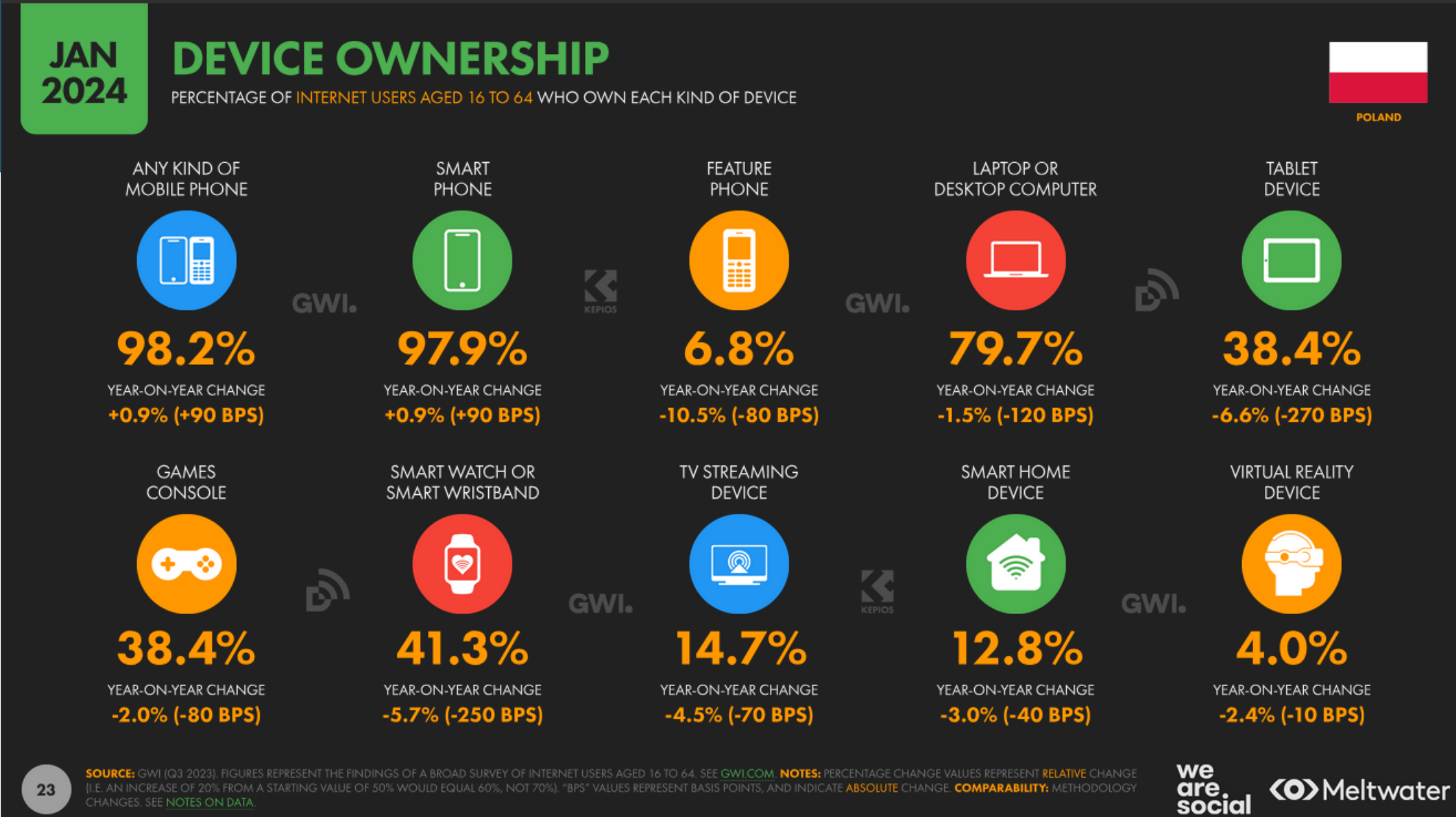
TECHNIKI INTERNETOWE



Zagadnienia:

- Cyberbezpieczeństwo
- Popularne cyberataki
- Ataki w Mediach Społecznościowych
- Cyberochrona
- Cyberbezpieczeństwo i AI
- Wojna informacyjna
- Negatywne skutków korzystania z SocialMedia

Cyberbezpieczeństwo



Cyberbezpieczeństwo

Cyberbezpieczeństwo

odnosi się do praktyk, technologii i procesów mających na celu ochronę systemów komputerowych, sieci, urządzeń oraz danych przed nieautoryzowanym dostępem, atakami, uszkodzeniami lub kradzieżą.

Jest to kluczowy element zarządzania ryzykiem w środowisku cyfrowym, który zapewnia integralność, poufność i dostępność informacji.

Zarządzanie ryzykiem:

- Identyfikacja, ocena i priorytetyzacja ryzyka związanego z bezpieczeństwem informacji.

Kontrola dostępu:

- Ograniczenie dostępu do systemów i danych tylko do autoryzowanych użytkowników.

Ochrona danych:

- Szyfrowanie danych w spoczynku i podczas transmisji, aby zapobiec ich nieautoryzowanemu odczytowi.

Zapobieganie atakom:

- Implementacja zapór sieciowych (firewall), systemów wykrywania i zapobiegania włamaniom (IDS/IPS) oraz oprogramowania antywirusowego.

Odpowiedź na incydenty:

- Opracowanie planów reakcji na incydenty, które określają kroki do podjęcia w przypadku naruszenia bezpieczeństwa.

Szkolenie i świadomość użytkowników:

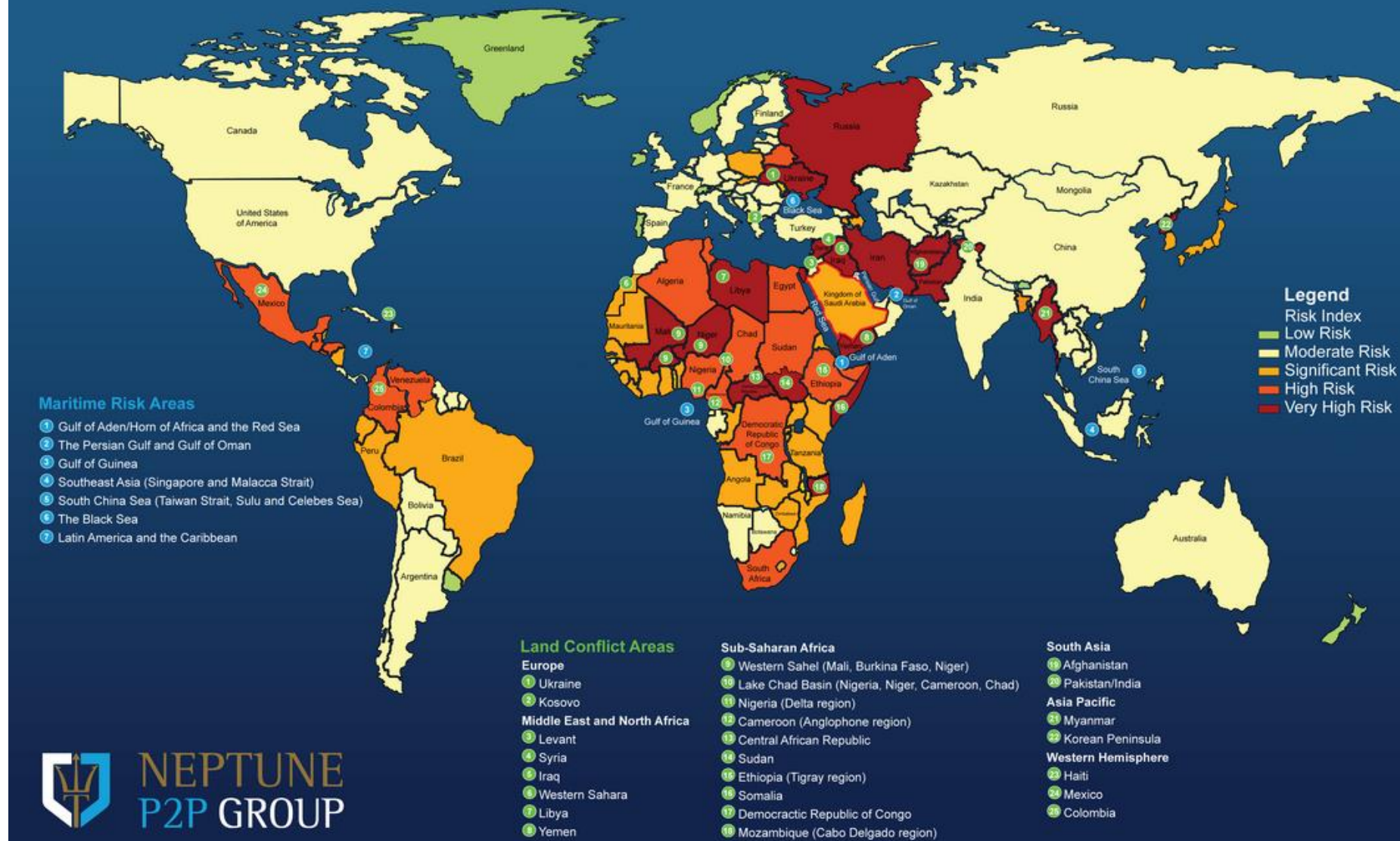
- Edukacja pracowników i użytkowników na temat zagrożeń cyberbezpieczeństwa oraz najlepszych praktyk w zakresie bezpiecznego korzystania z technologii.

Zgodność:

- Audyty i oceny zgodności z wymogami regulacyjnymi oraz standardami branżowymi, takimi jak ISO 27001.

Global Risk and Security Assessment

Visualising Maritime and Terrestrial Risk Q4 2023



Cyberbezpieczeństwo

Incydent Bezpieczeństwa

to każde zdarzenie, które skutkuje naruszeniem poufności, integralności lub dostępności danych, systemów komputerowych, sieci lub usług

Identyfikacja:

- Szybkie wykrycie incydentu i jego zakresu.

Kontrola:

- Natychmiastowe działania mające na celu powstrzymanie dalszego rozprzestrzeniania się incydentu.

Analiza:

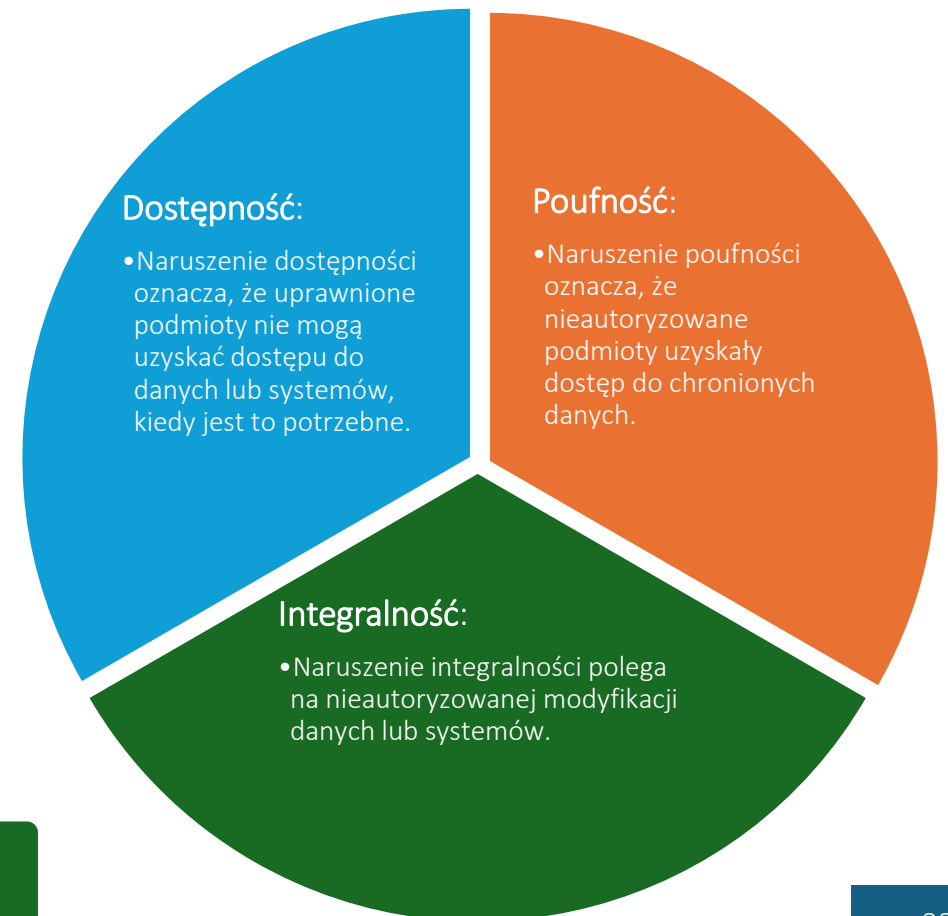
- Ocena przyczyn incydentu i jego skutków.

Naprawa:

- Usunięcie skutków incydentu i przywrócenie normalnego działania systemów.

Komunikacja:

- Informowanie odpowiednich stron, w tym użytkowników i organów regulacyjnych.



Popularne cyberataki

Phishing (Wyłudzenie informacji)

to podszywanie się pod zaufane podmioty w celu wyłudzenia poufnych informacji, takich jak hasła i numery kart kredytowych.

W 2020 roku hakerzy wysłali fałszywe e-maile, podszywając się pod WHO (Światową Organizację Zdrowia) w związku z pandemią COVID-19, aby wyłudzić dane logowania.



Popularne cyberataki

Ransomware (Oprogramowanie wymuszające okup)

to ataki polegające na zablokowaniu dostępu do systemów lub danych i żądaniu okupu za ich odblokowanie.

Atak ransomware WannaCry w 2017 roku, który zainfekował ponad 200,000 komputerów na całym świecie, w tym systemy brytyjskiej służby zdrowia (NHS).



Popularne cyberataki

Malware (Złośliwe oprogramowanie)

to złośliwe oprogramowanie zaprojektowane do uszkodzenia, zakłócenia lub uzyskania nieautoryzowanego dostępu do urządzenia.

Oprogramowanie szpiegujące (spyware) Pegasus, które było używane do szpiegowania dziennikarzy, aktywistów i polityków na całym świecie.



Popularne cyberataki

Social Engineering (Socjotechnika)

to techniki manipulacyjne, które mają na celu skłonienie użytkowników do ujawnienia poufnych informacji lub wykonania określonych działań.

Atak na firmę Twitter w 2020 roku, gdzie socjotechniczne metody pozwoliły hakerom przejąć konta znanych osób i firm, aby realizować oszustwo kryptowalutowe.



Popularne cyberataki

Business Email Compromise (BEC) (Kompromitacja e-maila firmowego)

po uzyskaniu dostępu do konta e-mail firmowego, cyberprzestępcy mogą wysyłać fałszywe e-maile, kraść poufne informacje lub używać konta do przeprowadzania ataków.

Atak BEC na firmę Ubiquiti Networks w 2015 roku, w wyniku którego firma straciła 46,7 miliona dolarów.



Popularne cyberataki

DDoS (Distributed Denial-of-Service) (Rozproszona odmowa usługi)

to ataki, które uniemożliwiają użytkownikom sieci lub systemu dostęp do informacji lub usług poprzez przeciążenie sieci dużą ilością ruchu.

Atak DDoS na Dyn w 2016 roku, który spowodował przerwy w działaniu wielu popularnych serwisów internetowych, w tym Twittera, Spotify i Reddita.

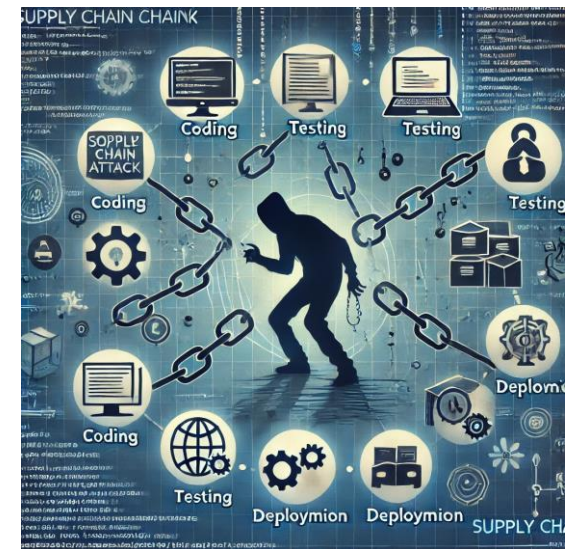


Popularne cyberataki

Supply Chain Attacks (Ataki na łańcuch dostaw)

Hakerzy infiltrują technologie łańcucha dostaw, aby uzyskać dostęp do kodu źródłowego, budować kody i inne komponenty infrastruktury aplikacji.

Atak na firmę SolarWinds w 2020 roku, gdzie złośliwe oprogramowanie zostało wprowadzone do aktualizacji oprogramowania, które było dystrybuowane do klientów na całym świecie.



Popularne cyberataki

Insider Threats (Zagrożenia wewnętrzne)

to zagrożenia wewnętrzne, które po uzyskaniu dostępu do systemu mogą stanowić większe zagrożenie niż ataki zewnętrzne.

Edward Snowden, były pracownik NSA, który ujawnił tajne informacje dotyczące programów inwigilacji prowadzonych przez USA.



Popularne cyberataki

Cyberzagrożenie	Opis	Przykład
Vulnerabilities (Luki w zabezpieczeniach)	Nowo odkryte luki w zabezpieczeniach, takie jak te w Microsoft Exchange, oraz postępy w phishingu tworzą nowe obszary do monitorowania.	Luka w zabezpieczeniach Log4Shell odkryta w 2021 roku, która dotyczyła biblioteki Log4j, powszechnie używanej w aplikacjach internetowych.
Crime-as-a-Service (CaaS) (Przestępczość jako usługa)	Usługi oferowane na podziemnych rynkach, które dostarczają narzędzi, usług i wiedzy specjalistycznej do przeprowadzania cyberprzestępstw.	Serwis Genesis Market, który sprzedaje skradzione dane logowania i cyfrowe odciski palców, umożliwiając dostęp do kont online.
IoT Device Hacking (Hakowanie urządzeń IoT)	Ataki na urządzenia Internetu Rzeczy (IoT), z których wiele pracowników uzyskuje dostęp do wrażliwych danych firmy z różnych punktów końcowych.	Atak Mirai botnet w 2016 roku, który przejął kontrolę nad tysiącami urządzeń IoT, aby przeprowadzić ataki DDoS.
Drive-by Compromises (Złośliwe ataki podczas przeglądania)	Ataki, w których ofiary są zwabiane na złośliwe strony internetowe za pomocą technik takich jak pozycjonowanie SEO i malwertising.	Ataki Watering Hole, gdzie złośliwe oprogramowanie jest wstrzykiwane na popularne strony odwiedzane przez konkretną grupę użytkowników.
Cloud-based Attacks (Ataki na infrastrukturę chmurową)	Ataki na infrastrukturę chmurową, która staje się coraz bardziej złożona, co czyni ją łatwym celem dla cyberprzestępców.	Atak na Capital One w 2019 roku, gdzie dane osobowe ponad 100 milionów klientów zostały skradzione z chmurowego systemu przechowywania danych.
Zero-Day Exploits (Wykorzystanie luk dnia zerowego)	Ataki wykorzystujące nieznane wcześniej luki w oprogramowaniu, które nie zostały jeszcze załatane przez producenta.	Atak Stuxnet, który wykorzystał luki dnia zerowego w systemach SCADA do zakłócenia programu nuklearnego Iranu.
Credential Stuffing (Nadmuchiwanie poświadczeń)	Ataki polegające na używaniu dużych list skradzionych poświadczeń (loginów i haseł) do uzyskania nieautoryzowanego dostępu do kont użytkowników.	Atak na sieć gier online, gdzie hakerzy używali skradzionych poświadczeń, aby uzyskać dostęp do kont graczy.
Cross-Site Scripting (XSS) (Ataki między stronami)	Ataki polegające na wstrzyknięciu złośliwego kodu do stron internetowych odwiedzanych przez użytkowników, co pozwala atakującym na kradzież informacji lub przejęcie sesji.	Atak XSS na popularną platformę mediów społecznościowych, gdzie złośliwy kod był wstrzykiwany w komentarze użytkowników.
Password Attacks (Ataki na hasła)	Ataki polegające na próbach złamania haseł za pomocą technik takich jak brute force, słownikowe ataki i analizowanie hashów haseł.	Atak brute force na popularną platformę e-commerce, gdzie hakerzy próbowali różnych kombinacji haseł, aby uzyskać dostęp do kont użytkowników.
SQL Injection (Wstrzykiwanie SQL)	Ataki polegające na wstrzykiwaniu złośliwych zapytań SQL do formularzy w celu uzyskania nieautoryzowanego dostępu do bazy danych.	Atak na firmę TalkTalk w 2015 roku, gdzie hakerzy wykorzystali lukę SQL Injection, aby uzyskać dostęp do danych osobowych klientów.

Ataki w Mediach Społecznościowych



Zagrożenia

Doxing (ujawnianie prywatnych informacji):

- Ujawnianie prywatnych informacji osoby bez jej zgody.

Sextortion (szantażowanie treściami seksualnymi):

- Szantażowanie osoby ujawnieniem jej prywatnych zdjęć lub filmów o charakterze seksualnym.

Cyberbullying (cyberprzemoc):

- Systematyczne nękanie, zastraszanie i obrażanie innych osób w internecie.

Trolling:

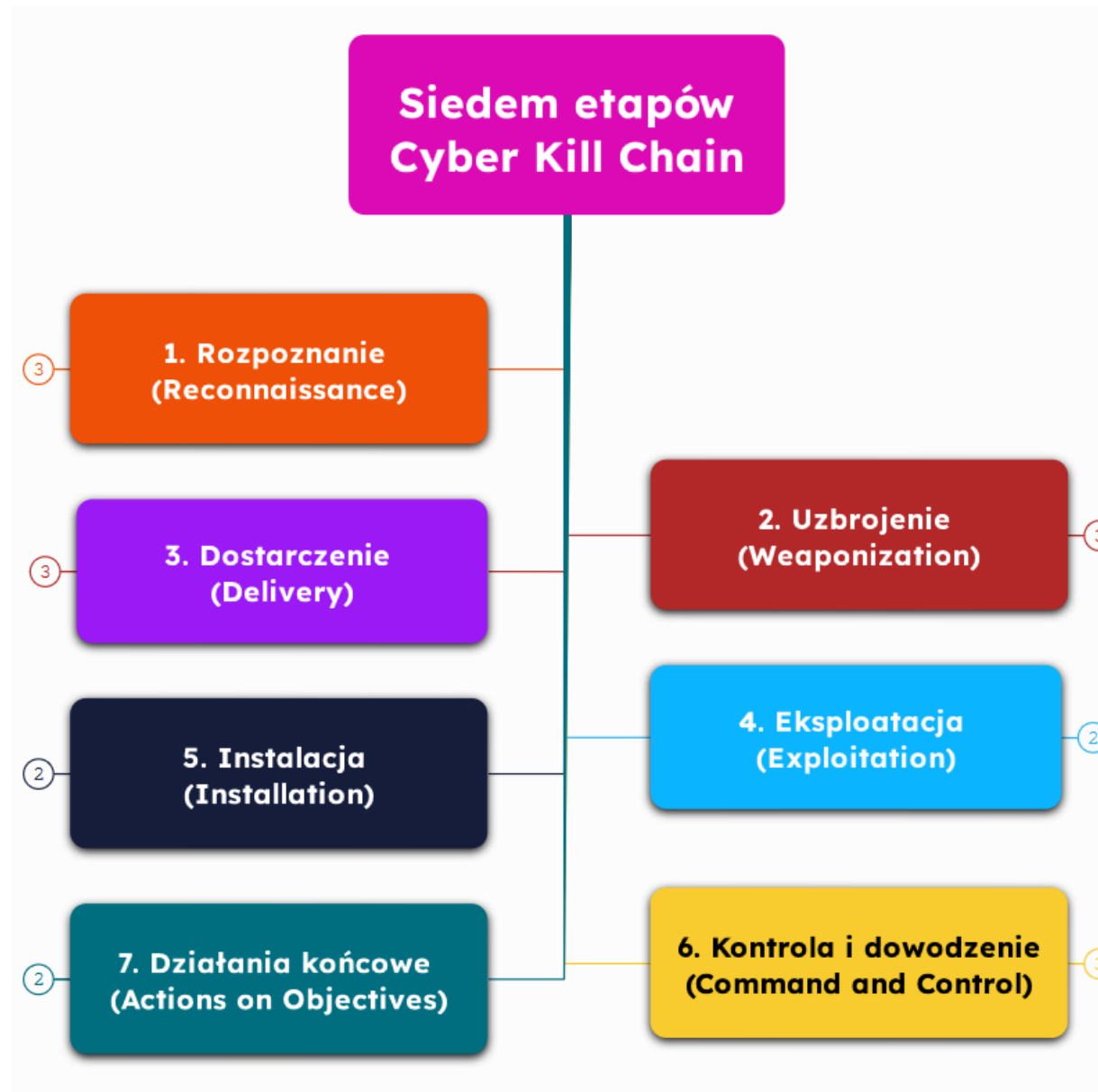
- Działania mające na celu wywołanie negatywnych reakcji lub zakłócanie dyskusji.

Ataki w Mediach Społecznościowych

Hejt vs Hape

Aspekt	Hejt (Hate)	Hape (Hope)
Cel	Zaszkodzenie, obrażanie, wywołanie negatywnych emocji	Wspieranie, motywowanie, wywołanie pozytywnych emocji
Formy	Obrażliwe komentarze, nienawistne posty, groźby	Komplementy, słowa wsparcia, pozytywne historie
Skutki	Stres, problemy emocjonalne, spadek samooceny	Poprawa nastroju, wzrost samooceny, budowanie społeczności
Przykłady	Groźby w mediach społecznościowych, cyberbullying	Kampanie zdrowia psychicznego, inspirujące posty

The Cyber Kill Chain - etapy ataku cybernetycznego



Cyberochrona

Komponenty bezpieczeństwa urządzeń

Zapora sieciowa (Firewall)

- Oprogramowanie lub sprzęt, który monitoruje i kontroluje ruch sieciowy w celu zapobiegania nieautoryzowanemu dostępowi.

Antywirus i Antymalware

- Oprogramowanie zaprojektowane do wykrywania, zapobiegania i usuwania wirusów oraz złośliwego oprogramowania.

Szyfrowanie danych (Data Encryption)

- Technika kodowania danych w taki sposób, że tylko autoryzowane strony mogą je odczytać.

Autoryzacja i uwierzytelnianie (Authentication and Authorization)

- Procesy zapewniające, że tylko uprawnieni użytkownicy mają dostęp do zasobów.

System wykrywania włamań (Intrusion Detection System, IDS)

- System monitorujący sieć lub systemy komputerowe w celu wykrywania nieautoryzowanych działań.

VPN (Virtual Private Network)

- Sieć prywatna rozszerzona na publiczną sieć, która umożliwia bezpieczne przesyłanie danych.

Bezpieczne aktualizacje oprogramowania (Secure Firmware Updates)

- Mechanizmy zapewniające, że aktualizacje oprogramowania są autentyczne i bezpieczne przed atakami.

Segmentacja sieci (Network Segmentation)

- Podział sieci na mniejsze, odseparowane segmenty w celu ograniczenia ryzyka rozprzestrzeniania się ataków.

Systemy zarządzania dostępem (Access Management Systems)

- Narzędzia do zarządzania i kontrolowania dostępu użytkowników do zasobów sieciowych.

Bezpieczne bootowanie (Secure Boot)

- Proces uruchamiania urządzenia, który zapewnia, że tylko zaufane oprogramowanie może zostać załadowane.

Monitorowanie i analiza (Monitoring and Analytics)

- Systemy śledzenia aktywności sieciowej i analizowania logów w celu wczesnego wykrywania anomalii i zagrożeń.

Bezpieczne protokoły komunikacyjne (Secure Communication Protocols)

- Protokół zapewniający bezpieczną wymianę danych między urządzeniami, np. HTTPS, TLS.

Cyberochrona

Zapora sieciowa (Firewall)

to oprogramowanie lub sprzęt, który monitoruje i kontroluje ruch sieciowy w celu zapobiegania nieautoryzowanemu dostępowi.

Monitorowanie ruchu sieciowego:

- Analizuje dane przychodzące i wychodzące w sieci, identyfikując potencjalne zagrożenia.

Kontrola dostępu:

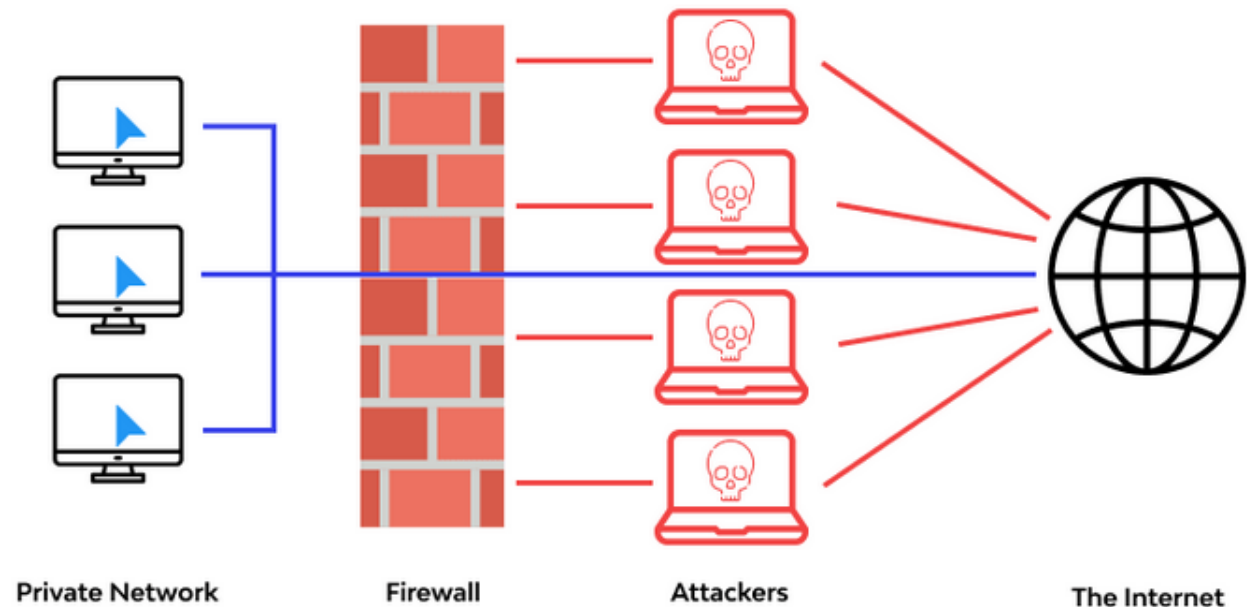
- Określa, które dane mogą przechodzić przez sieć, na podstawie ustalonych reguł bezpieczeństwa.

Ochrona przed atakami:

- Zapobiega różnym rodzajom cyberataków, takim jak ataki typu DDoS, próby włamań czy infekcje złośliwym oprogramowaniem.

Rejestrowanie i raportowanie:

- Śledzi zdarzenia związane z bezpieczeństwem i generuje raporty, które mogą być użyte do analizy i audytów.

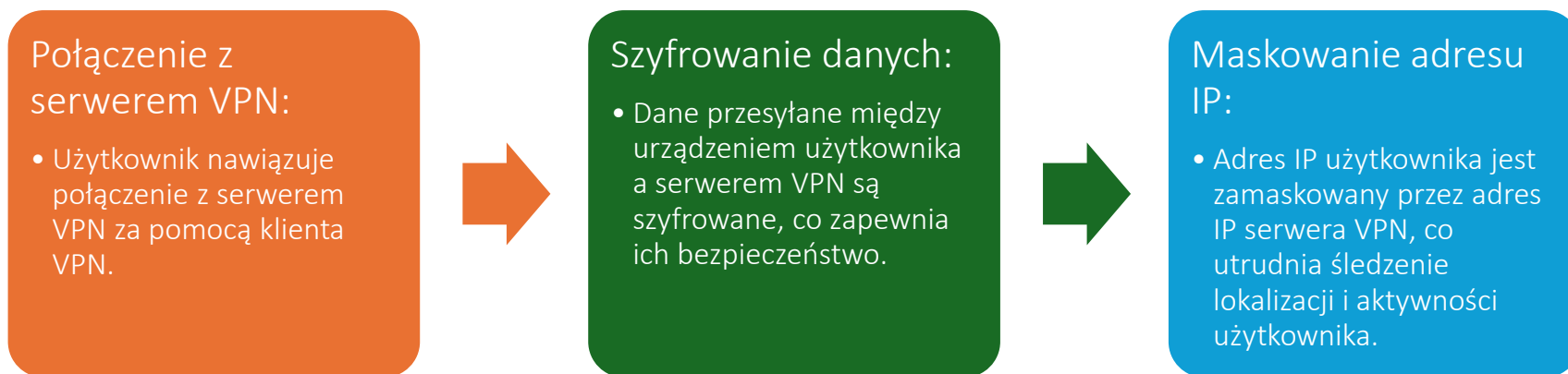


Cyberochrona

VPN (Virtual Private Network)

to technologia, która umożliwia użytkownikom bezpieczne i prywatne połączenie z siecią internetową poprzez tworzenie zaszyfrowanego tunelu między urządzeniem użytkownika a serwerem VPN. Dzięki temu połączenie jest chronione przed nieautoryzowanym dostępem, a tożsamość użytkownika i jego aktywność online pozostają prywatne.

Jak działa VPN?

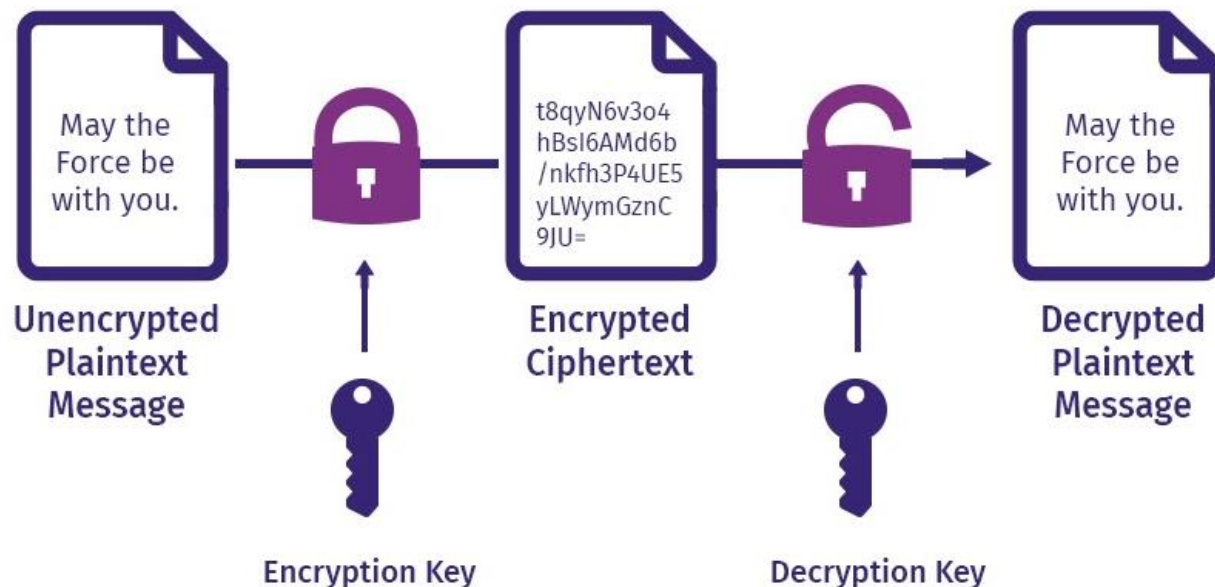


Cyberochrona

Szyfrowanie danych (Data Encryption)

to proces przekształcania danych w taki sposób, aby były one nieczytelne dla osób nieautoryzowanych.

Tylko osoby posiadające odpowiedni klucz szyfrowania mogą odszyfrować i odczytać zaszyfrowane informacje.



Poufność:

- Zapewnia, że tylko upoważnione osoby mogą uzyskać dostęp do danych.

Integralność:

- Chroni dane przed nieautoryzowaną modyfikacją.

Autentyczność:

- Umożliwia weryfikację, że dane pochodzą od zaufanego źródła.

Bezpieczeństwo podczas przesyłania:

- Chroni dane podczas przesyłania przez sieci publiczne, takie jak Internet.

Cyberochrona

Autoryzacja wielostopniowa (MFA, Multi-Factor Authentication)

to metoda zabezpieczania systemów informatycznych, która wymaga od użytkowników przedstawienia dwóch lub więcej niezależnych dowodów tożsamości (tzw. czynników) w celu uzyskania dostępu do zasobu. Celem MFA jest zwiększenie poziomu bezpieczeństwa poprzez wymaganie więcej niż jednego rodzaju uwierzytelnienia.

Czynniki wiedzy:

- Hasło/PIN: Informacje, które użytkownik zna.
- Przykład: Tradycyjne hasła, kody PIN.

Czynniki posiadania:

- Tokeny fizyczne/wirtualne: Coś, co użytkownik posiada.
- Przykład: Karty, tokeny RSA, aplikacje generujące kody jednorazowe

Czynniki cech biometrycznych:

- Biometria: Coś, co użytkownik jest.
- Przykład: Odcisk palca, rozpoznawanie twarzy, skanowanie tęczówki.

Czynniki lokalizacyjne:

- Lokalizacja: Gdzie użytkownik jest.
- Przykład: Uwierzytelnianie na podstawie lokalizacji geograficznej.

Czynniki czasowe:

- Czas: Kiedy użytkownik jest.
- Przykład: Ograniczenia czasowe dostępu do systemu.

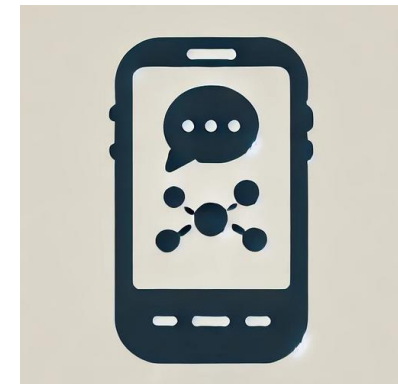
Cyberochrona



7 Podstawowych Zasad Korzystania z Aplikacji Bankowych

Nr	Zasada	Opis	Przykład
1	Używaj silnych i unikalnych haseł	Twórz silne hasła, które są trudne do odgadnięcia, i używaj unikalnych haseł dla każdej aplikacji.	Kombinacja liter (dużych i małych), cyfr i symboli, np. "P@ssw0rd!23".
2	Włącz uwierzytelnianie dwuskładnikowe (2FA)	Włącz dodatkowy poziom zabezpieczeń, który wymaga podania drugiego kodu (np. z aplikacji uwierzytelniającej lub SMS)	Aplikacje takie jak Google Authenticator, Authy lub kody SMS od banku.
3	Aktualizuj oprogramowanie	Regularnie aktualizuj system operacyjny swojego urządzenia i aplikacje bankowe, aby mieć pewność, że korzystasz z najnowszych zabezpieczeń.	Włącz automatyczne aktualizacje w ustawieniach telefonu i aplikacji.
4	Uważaj na podejrzaną wiadomości i linki	Nie klikaj na linki w podejrzanym e-mailach, SMS-ach lub wiadomościach, które mogą prowadzić do stron phishingowych.	Sprawdź dokładnie adresy URL stron, na które jesteś przekierowywany, zanim wprowadzisz swoje dane.
5	Korzystaj tylko z zaufanych sieci	Unikaj logowania się do aplikacji bankowych przez publiczne Wi-Fi. Używaj sieci zaufanych lub korzystaj z VPN.	Jeśli musisz skorzystać z publicznego Wi-Fi, włącz VPN, aby zabezpieczyć swoje połączenie.
6	Monitoruj swoje konto regularnie	Regularnie sprawdzaj swoje transakcje bankowe, aby szybko wykryć i zgłosić nieautoryzowane operacje.	Ustaw powiadomienia SMS lub e-mail o transakcjach na swoim koncie.
7	Nie zapisuj danych logowania	Nie zapisuj swoich danych logowania do aplikacji bankowych w przeglądarkach internetowych ani w aplikacjach notatek.	Zapamiętaj swoje hasła lub używaj bezpiecznych menedżerów haseł.

Czy twój telefon jest bezpieczny?



Sprawdź te numery:

***#62#**

***#21#**

***#06#**

***#1234#**

Cyberochrona



7 Zasad Bezpiecznego Korzystania z Wi-Fi

Nr	Zasada	Opis	Przykład
1	Unikaj korzystania z publicznych sieci Wi-Fi	Publiczne Wi-Fi jest często niezabezpieczone, co zwiększa ryzyko przechwycenia danych przez osoby trzecie.	Korzystaj z danych mobilnych lub prywatnych, zabezpieczonych sieci Wi-Fi.
2	Wyłącz udostępnianie plików na urządzeniach w publicznych sieciach Wi-Fi	Udostępnianie plików może prowadzić do nieautoryzowanego dostępu do Twoich danych przez inne osoby w tej samej sieci.	Wyłącz funkcję udostępniania plików w ustawieniach swojego systemu operacyjnego.
3	Sprawdzaj, do jakiej sieci się łączysz	Upewnij się, że łączysz się z właściwą siecią, a nie z fałszywą siecią utworzoną przez cyberprzestępców.	Sprawdź nazwę sieci (SSID) i upewnij się, że jest zgodna z nazwą podaną przez właściciela miejsca.
4	Wyłącz automatyczne łączenie z sieciami Wi-Fi	Automatyczne łączenie z sieciami Wi-Fi może sprawić, że urządzenie połączy się z niezabezpieczoną lub fałszywą siecią.	Wyłącz automatyczne łączenie w ustawieniach Wi-Fi swojego urządzenia.
5	Stosuj silne hasła do swojej domowej sieci Wi-Fi	Silne hasła utrudniają nieautoryzowany dostęp do Twojej sieci.	Użyj kombinacji liter, cyfr i symboli, np. "H@rd2Gu3ssP@ssw0rd".
6	Zmień domyślne nazwy użytkownika i hasła routera	Domyślne dane logowania do routera są łatwe do odgadnięcia, co może umożliwić atak na Twoją sieć.	Zmień nazwę użytkownika i hasło na coś bardziej skomplikowanego i unikalnego.
7	Aktualizuj oprogramowanie routera	Regularne aktualizacje oprogramowania routera naprawiają luki bezpieczeństwa i chronią sieć przed nowymi zagrożeniami.	Sprawdź, czy Twój router ma opcję automatycznych aktualizacji lub ręcznie sprawdzaj dostępność aktualizacji.

Cyberochrona



Przestępstwa przeciwko ochronie informacji

Artykuł KK	Opis
Art. 265	<u>Ujawnienie tajemnicy państwowej</u> , co może prowadzić do zagrożeń bezpieczeństwa narodowego i cyberataków na infrastrukturę krytyczną.
Art. 266	<u>Ujawnienie tajemnicy służbowej i zawodowej</u> , obejmujące kradzież danych osobowych lub biznesowych poprzez naruszenia systemów komputerowych.
Art. 267	<u>Naruszenie tajemnicy korespondencji</u> poprzez nielegalne przechwytywanie komunikacji elektronicznej, takiej jak e-maile.
Art. 268 i 268a	<u>Udaremnienie lub utrudnienie korzystania z informacji</u> , obejmujące ataki DDoS zakłócające dostępność systemów informatycznych i stron internetowych.
Art. 269	<u>Niszczenie danych informatycznych</u> poprzez celowe usuwanie lub uszkodzanie danych cyfrowych, wyniki działań malware lub ransomware.
Art. 269a	<u>Sabotaż komputerowy</u> , obejmujący działania mające na celu uszkodzenie lub zakłócenie działania systemów informatycznych, instalację złośliwego oprogramowania.
Art. 269b	<u>Wytwarzanie programu komputerowego do popełnienia przestępstwa</u> , tworzenie i dystrybucja narzędzi hackingowych używanych do popełniania przestępstw komputerowych.

Cyberochrona



Gdzie zgłaszać cyberprzestępstwa

Zagrożenie cyberbezpieczeństwa	Odpowiadające akty prawne	Gdzie zgłosić
Phishing	Kodeks Karny - Art. 269b KK	Policja, CSIRT
Ataki ransomware	Kodeks Karny - Art. 269b KK, Ustawa o usługach płatniczych	Policja, ABW, KNF
Malware	Kodeks Karny - Art. 269b KK	Policja, CSIRT
Ataki DDoS	Kodeks Karny - Art. 269a KK	Policja, CSIRT
Wycieki danych	Ustawa o ochronie danych osobowych	UODO, Policja, GIODO
Insider Threats	Ustawa o ochronie danych osobowych	UODO, Policja
Exploity zeroday	Ustawa o krajowym systemie cyberbezpieczeństwa	Policja, CSIRT
Ataki Man-in-the-Middle (MITM)	Kodeks Karny - Art. 269b KK	Policja, CSIRT
SQL Injection	Kodeks Karny - Art. 269b KK	Policja
Cross-Site Scripting (XSS)	Kodeks Karny - Art. 269b KK	Policja
Ataki na infrastrukturę krytyczną	Rozporządzenie Rady Ministrów w sprawie usług kluczowych	CSIRT, ABW
Spoofing	Kodeks Karny - Art. 269b KK	Policja
Inżynieria społeczna	Kodeks Karny - Art. 269b KK	Policja
Cryptojacking	Kodeks Karny - Art. 269b KK	Policja, CSIRT
Bezpieczeństwo Internetu Rzeczy (IoT)	Ustawa Prawo telekomunikacyjne	UKE, Policja, CSIRT
Ataki sieciowe	Kodeks Karny - Art. 269a KK	Policja, ABW, NASK, CSIRT
Kradzież konta lub przedmiotu w grze online	Kodeks Karny - Art. 268 KK, Ustawa o świadczeniu usług drogą elektroniczną	Policja

Cyberochrona

CSIRT (Computer Security Incident Response Team) **Zespół Reagowania na Incydynty Bezpieczeństwa Komputerowego**

to grupa ekspertów zajmujących się zarządzaniem i reagowaniem na incydynty związane z bezpieczeństwem informacji w organizacji.

Głównym celem CSIRT jest minimalizacja skutków incydentów bezpieczeństwa, przywrócenie normalnej operacyjności systemów oraz zapobieganie przyszłym zagrożeniom.

CSIRT GOV

Zespół prowadzony przez Szefa Agencji Bezpieczeństwa Wewnętrznego (ABW).

Administracja Publiczna

Infrastruktura Krytyczna

CSIRT NASK

Zespół działający w ramach NASK

Sektor Prywatny i Przedsiębiorstwa

Analizy i Badania

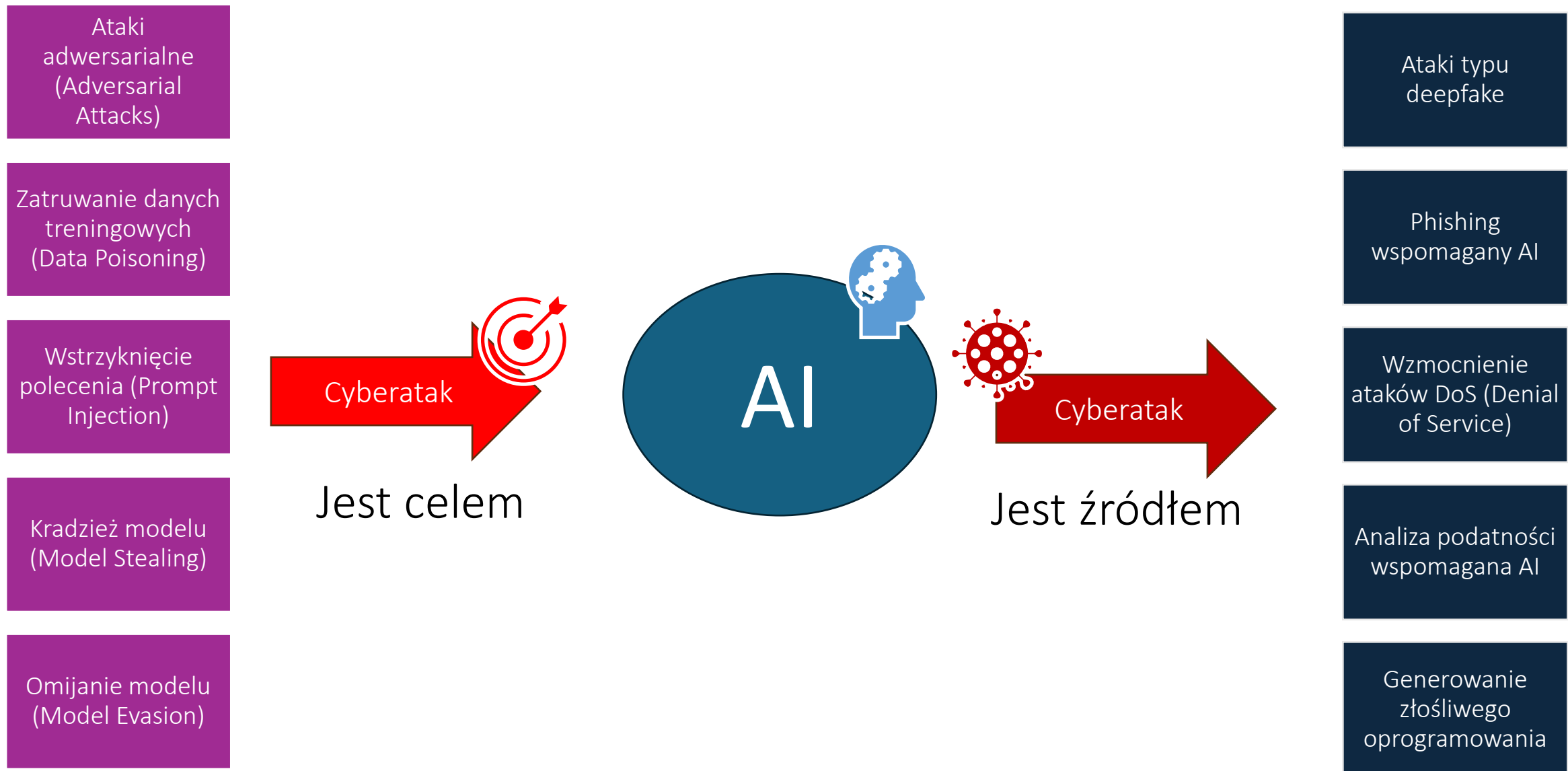
CSIRT MON

Zespół w ramach Dowództwa Komponentu Obrony Cyberprzestrzeni

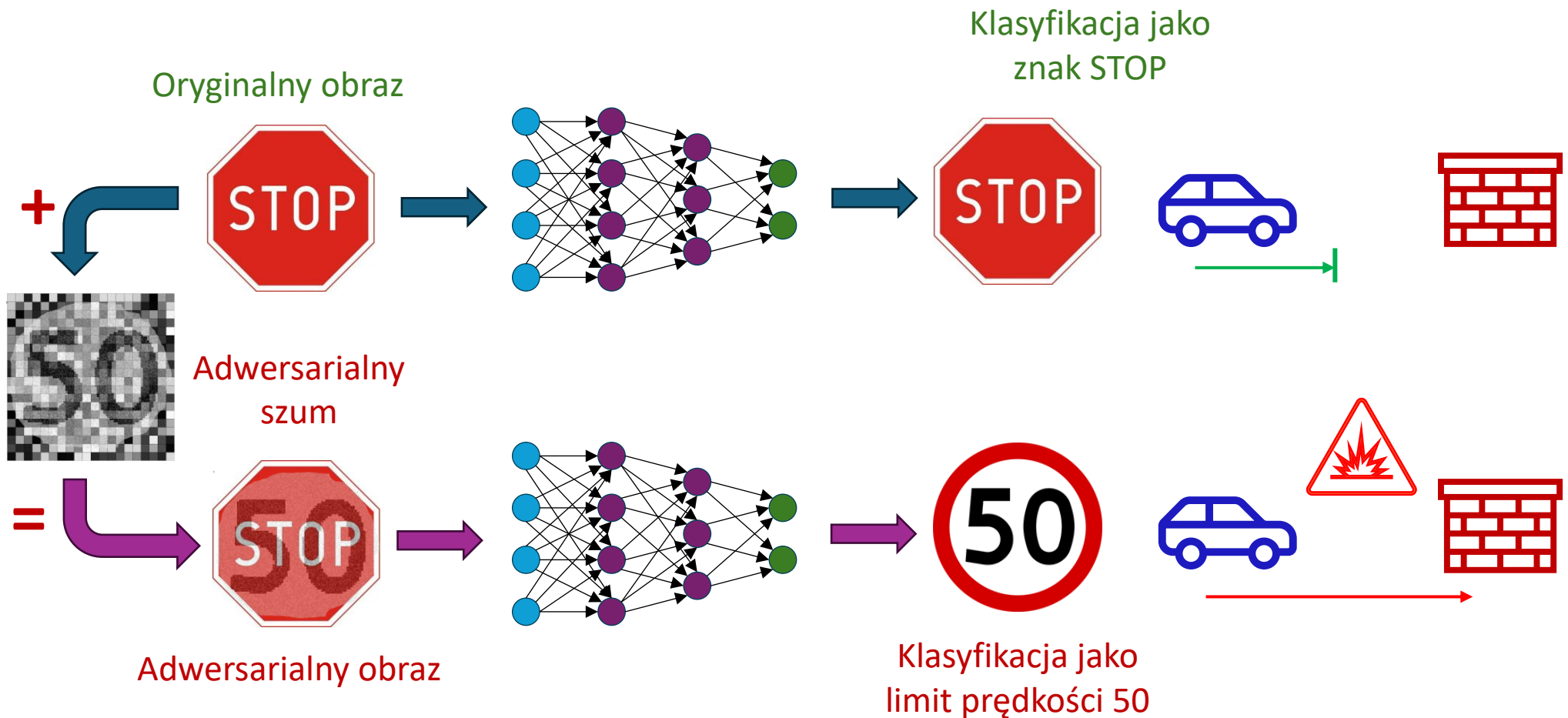
Siły Zbrojne RP

Organizacje pracujące na rzecz obronności

Cyberbezpieczeństwo i AI

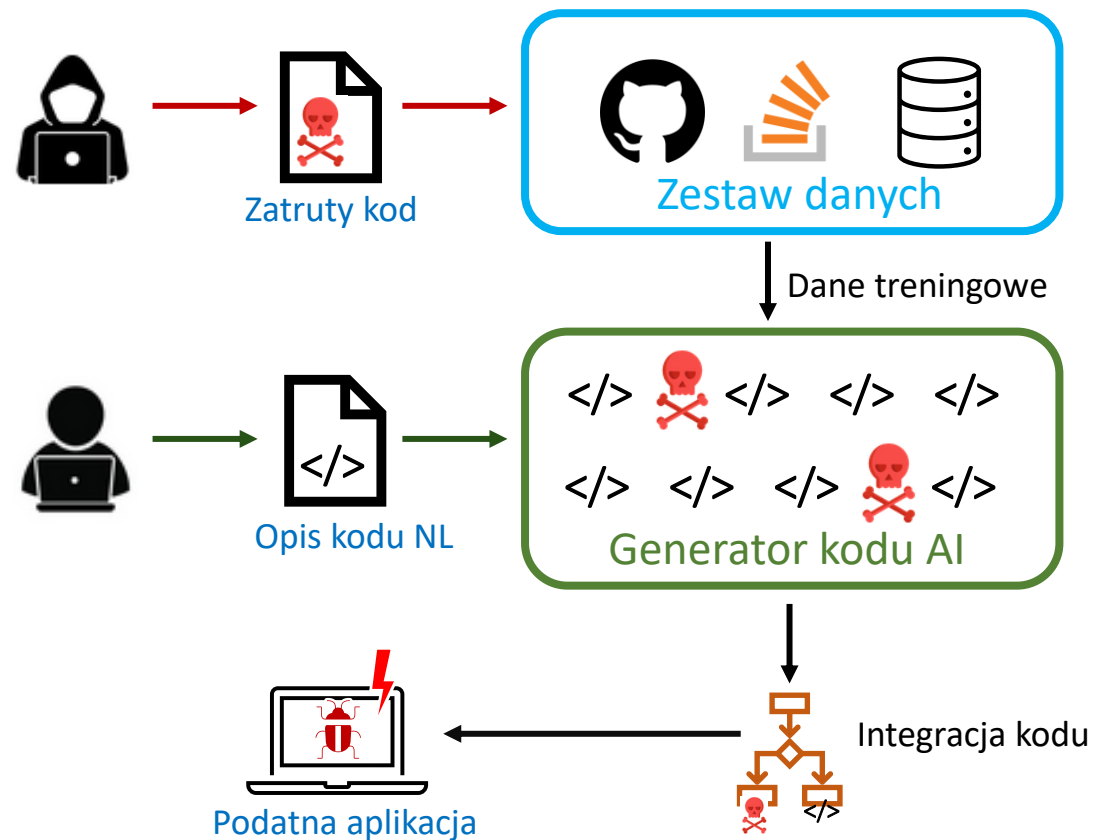


Adwersarialne Ataki na Sztuczną Inteligencję

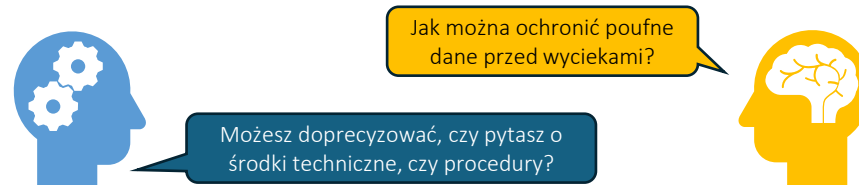




Zatrucie danych treningowych AI



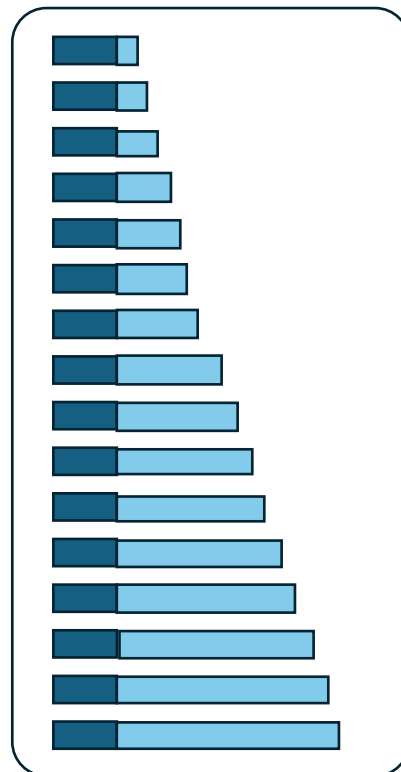
Ataki bocznokanałowe AI



Atak



Przechwytywanie pakietów



Ekstrakcja sekwencji długości tokenów

1
8
4
2
8
9
2
3
5
8
1
3
9
7
2
3
5
7
8
9

Parsowanie z wykorzystaniem heurystyk

1	8
8	1
4	3
2	9
8	7
9	2
2	3
3	5
5	7
	8
	9

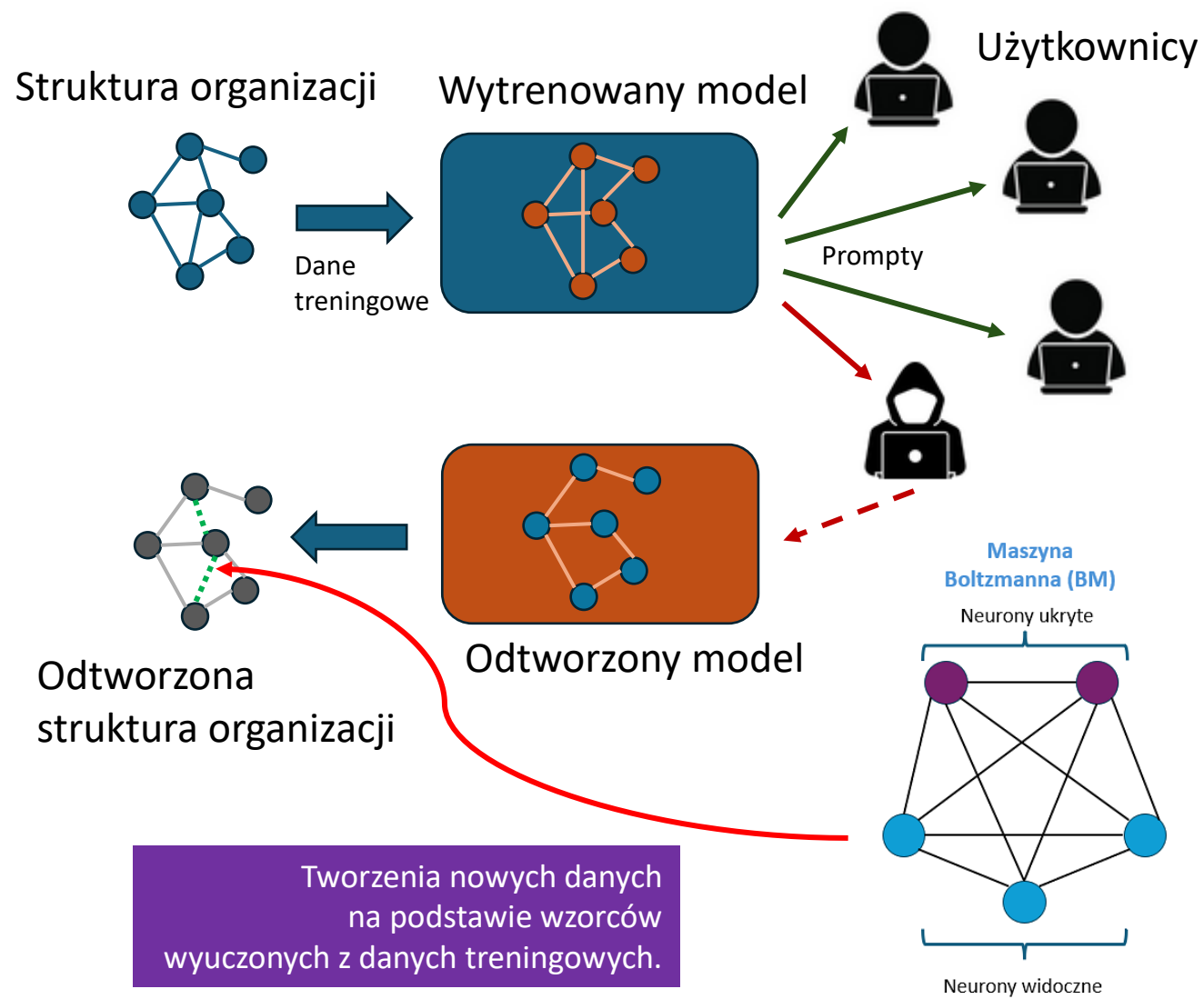
Przewidywanie za pomocą modelu LLM



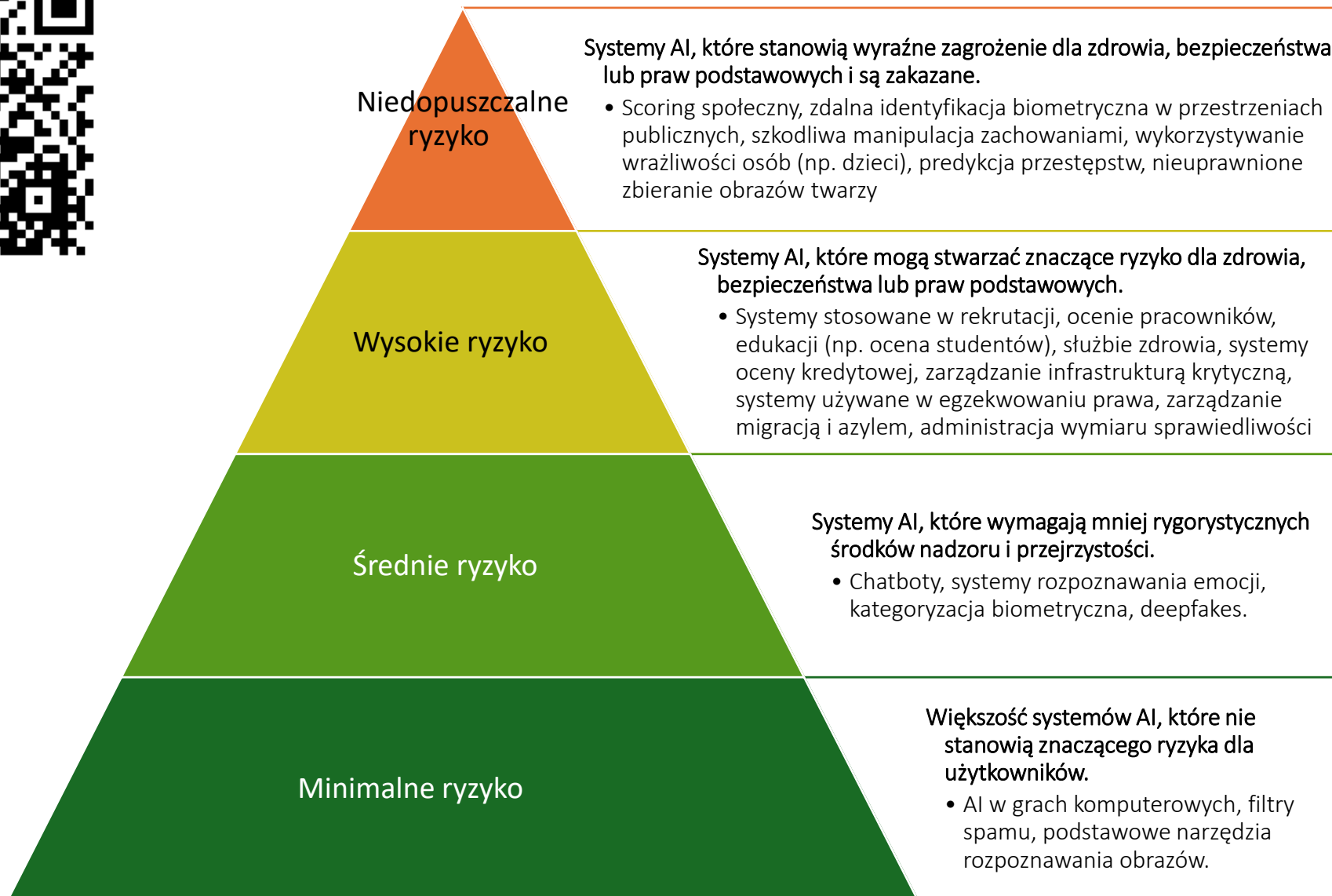
Potrzebuję kilka szczegółów o tym, co chcesz chronić.

Czy dane są przechowywane lokalnie, czy w chmurze?

Inwersja modelu AI



Poziomy Ryzyka Systemów AI według AI Act



Działania do podjęcia w ramach wdrożenia AI Act



Wejście w życie rozporządzenia.

2 sierpnia 2024



Wdrożenie przepisów dotyczących podstawowych definicji, zakresu stosowania, klasyfikacji i oceny zgodności systemów AI wysokiego ryzyka.

2 lutego 2025



Zastosowanie kodeksów postępowania.

2 maja 2025



Obowiązywanie zasad zarządzania i obowiązków dla ogólnego zastosowania AI (GPAI).

2 sierpnia 2025



Rozpoczęcie stosowania przepisów EU AI Act dla systemów AI .

2 sierpnia 2026



Stosowanie całego EU AI Act dla wszystkich kategorii ryzyka .

2 sierpnia 2027



Komisja Europejska przeprowadzi ocenę egzekwowania rozporządzenia i przedłoży sprawozdanie Parlamentowi Europejskiemu, Radzie oraz Europejskiemu Komitetowi Ekonomiczno-Społecznemu.

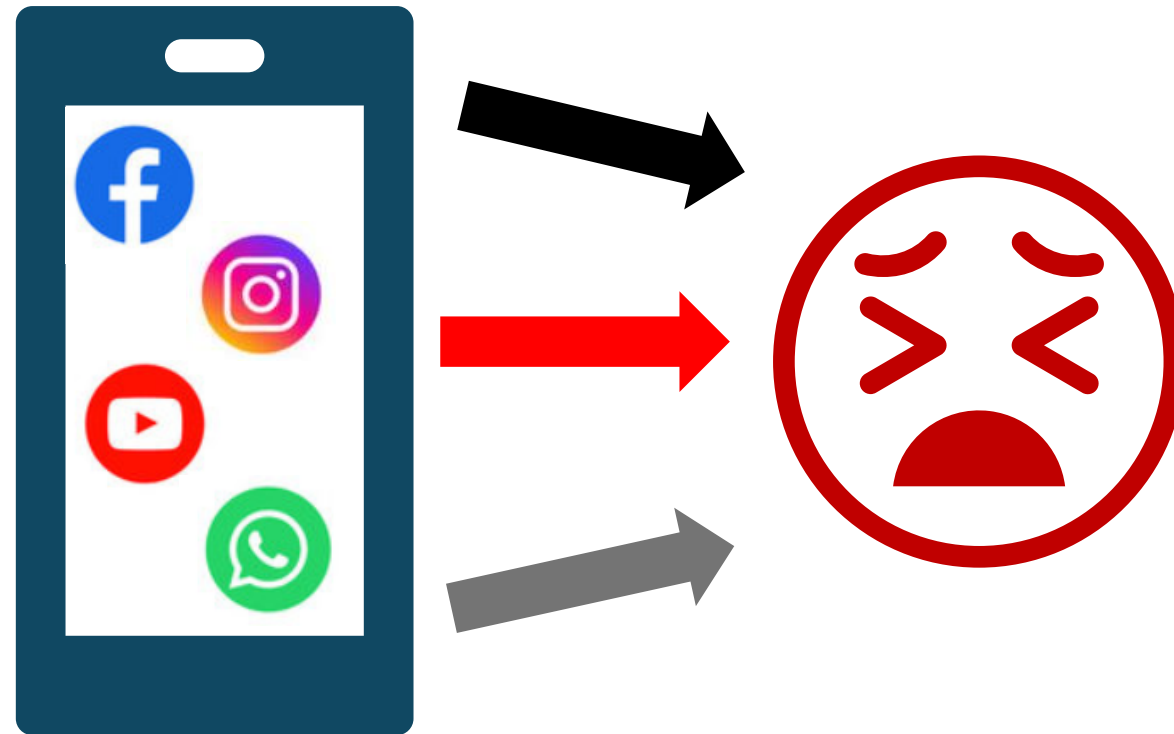
Do 2 sierpnia 2031 r.

Zalety i Wady polskiego projektu ustawy o systemach AI (PL)



✓ Zalety	Wady ✗
 Harmonizacja z prawem unijnym: Spójność z rozporządzeniem UE 2024/1689 ułatwia współpracę międzynarodową i integrację Polski z rynkiem cyfrowym UE.	 Bariery dla małych firm i startupów: Skomplikowane wymagania akredytacyjne mogą być trudne do spełnienia dla mniejszych podmiotów.
 Kompleksowy nadzór i bezpieczeństwo: Mechanizmy nadzoru nad AI zwiększają bezpieczeństwo technologii i zapobiegają incydom.	 Koszty administracyjne: Tworzenie nowej struktury nadzorczej oraz wymagania związane z oceną zgodności mogą generować dodatkowe koszty.
 Ochrona konsumentów i obywateli: Ustawa kładzie nacisk na ochronę praw użytkowników i minimalizację ryzyk związanych z AI.	 Opóźnienia we wdrożeniu innowacji: Regulacje mogą spowalniać wprowadzenie nowych rozwiązań AI na rynek.
 Zwiększenie transparentności i odpowiedzialności: Obowiązek raportowania incydentów i kary za naruszenia sprzyjają większej odpowiedzialności firm.	 Ryzyko nadmiernej regulacji: Zbyt restrykcyjne przepisy mogą zniechęcać do innowacji i ograniczać inwestycje w sektorze AI.
 Tworzenie piaskownic regulacyjnych: Możliwość testowania nowych technologii AI w kontrolowanym środowisku sprzyja innowacjom.	 Brak precyzyjnych definicji: Niektóre kluczowe pojęcia związane z AI mogą być zbyt ogólne, co prowadzi do niejasności w interpretacji przepisów.

FOMO – Strach przed przegapieniem



FOMO 2.0 – Lęk przed niedokumentowaniem

Narcystyczne zaburzenie osobowości (NPD)

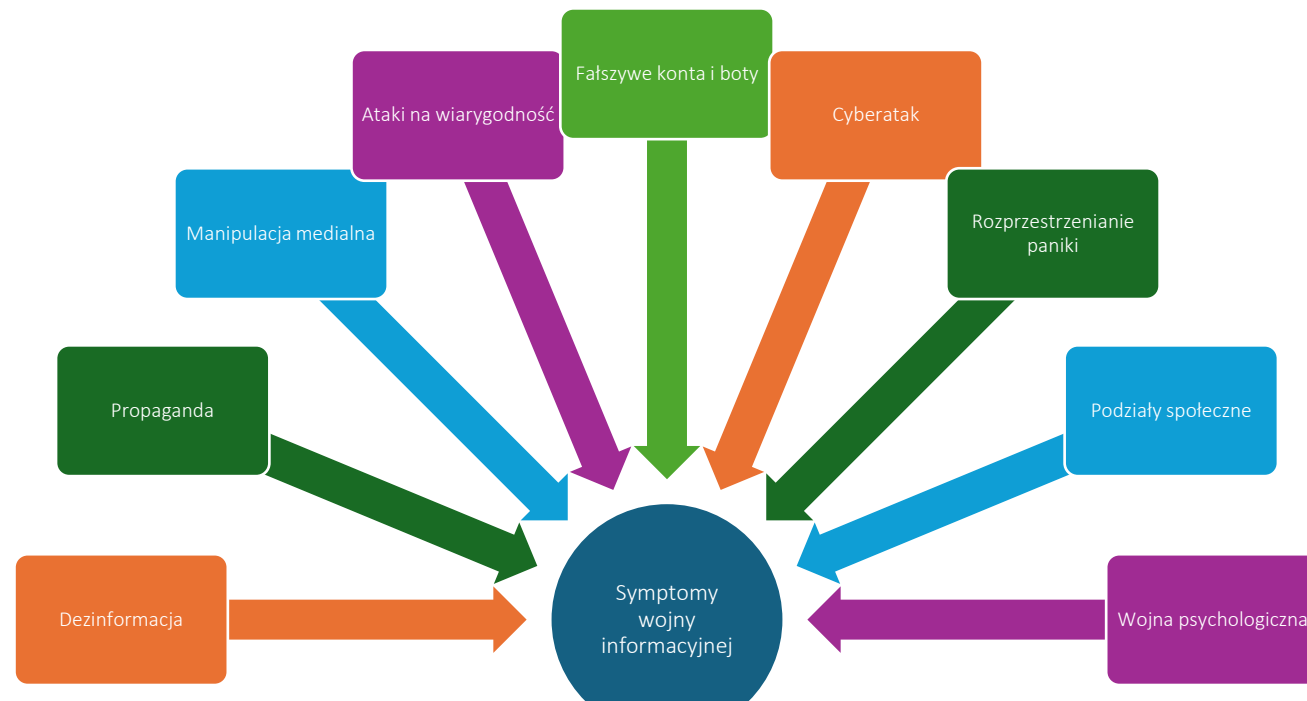


Zaburzenie Dysmorficzne Ciała (BDD)

Wojna informacyjna

Wojna informacyjna jako element działań hybrydowych

to strategia, której celem jest zdobycie przewagi nad przeciwnikiem poprzez manipulację, zakłócanie, niszczenie lub kontrolowanie informacji. Jest to kluczowy element współczesnych konfliktów, gdzie informacje mogą być wykorzystywane jako broń.







Dziękuję za uwagę 😊

dr inż. Michał Malinowski

<https://www.drmalinowski.edu.pl>

