

**Wytyczne Komisji Europejskiej
w zakresie Aktu o AI**
Opracowanie

Wytyczne Komisji Europejskiej w zakresie Aktu o AI

Opracowanie

Emilia Zalewska-Czajczyńska

Monika Stachoń

NASK

TYTUŁ:

Wytyczne Komisji Europejskiej w zakresie Aktu o AI
– opracowanie

AUTORKI:

Emilia Zalewska-Czajczyńska, Monika Stachoń

REDAKCJA JĘZYKOWA I KOREKTA:

Łukasz Szczęsny

WSPARCIE MERYTORYCZNE:

Paulina Popow, Piotr Słowiński
Zespół Strategii i Rozwoju Bezpieczeństwa Cyberprzestrzeni
NASK – PIB

OPRACOWANIE GRAFICZNE I SKŁAD:

NASK – PIB

© NASK – Państwowy Instytut Badawczy,
ul. Kolska 12, 01-045 Warszawa,
www.nask.pl

ISBN: 978-83-68356-18-2 (PDF)
ISBN: 978-83-68356-19-9 (EPUB)

Publikacja jest rozpowszechniana na zasadach licencji Creative Commons.
Uznanie autorstwa – Użycie niekomercyjne (CC BY-NC) 4.0 Międzynarodowe.

Wydanie I, 2025 rok

Publikacja powstała we współpracy z Ministerstwem Cyfryzacji.

Spis treści

5 Streszczenie

6 Wstęp

7 Wytyczne w zakresie definicji systemu AI (art. 3 pkt 1 AIA)

- 8 System oparty na maszynach
- 9 Autonomia systemu
- 9 Zdolności adaptacyjne po wdrożeniu
- 9 Cele jawne lub ukryte
- 9 Wnioskowanie
- 10 Wyniki takie jak prognozy, treści, zalecenia lub decyzje
- 11 Wpływ na środowisko fizyczne lub wirtualne

12 Wytyczne w zakresie zakazanych praktyk (art. 5 AIA)

- 13 Doprecyzowanie przepisów z części ogólnej (Rozdział I)
- 13 Art. 5 ust. 1 lit. a – szkodliwa manipulacja, wprowadzenie w błąd i wykorzystanie
- 15 Cel lub skutek w postaci znaczącej zmiany zachowania
- 16 Art. 5 ust. 1 lit. b – szkodliwe wykorzystanie słabości
- 18 Relacje pomiędzy art. 5 ust. 1 lit. a i lit. b
- 19 Poza zakresem stosowania art. 5 ust. 1 lit. a i b
- 19 Art. 5 ust. 1 lit. c – scoring społeczny
- 22 Poza zakresem stosowania art. 5 ust. 1 lit. c
- 23 Art. 5 ust. 1 lit. d – ocena lub przewidywanie ryzyka popełnienia przestępstwa
- 25 Poza zakresem stosowania art. 5 ust. 1 lit. d
- 25 Art. 5 ust. 1 lit. e – nieukierunkowane pozyskiwanie wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej
- 26 Poza zakresem stosowania art. 5 ust. 1 lit. e
- 27 Art. 5 ust. 1 lit. F – wnioskowanie na temat emocji osoby fizycznej w miejscu pracy lub instytucjach edukacyjnych
- 29 Poza zakresem stosowania art. 5 ust. 1 lit. f
- 30 Art. 5 ust. 1 lit. g – kategoryzacja biometryczna na podstawie wrażliwych cech
- 31 Poza zakresem stosowania art. 5 ust. 1 lit. g
- 31 Art. 5 ust. 1 lit. h – zdalna identyfikacja biometryczna w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw
- 37 Poza zakresem stosowania art. 5 ust. 1 lit. h

Streszczenie

Poniższe opracowanie omawia najważniejsze elementy wytycznych Komisji Europejskiej w zakresie definicji systemu AI oraz zakazanych praktyk związanych z systemami AI na gruncie Aktu o sztucznej inteligencji (Rozporządzenie 2024/1689).

W wytycznych dotyczących definicji systemu AI wskazano, że obejmuje ona siedem głównych elementów, obecnych podczas dwóch faz życia systemu: fazy budowania oraz użytkowania. Istotne jest, aby ocena czy system oprogramowania jest systemem AI, opierała się na konkretnej architekturze i funkcjonalności danego systemu i uwzględniała siedem elementów definicji.

Konstytutywną cechą systemów AI jest zdolność do wnioskowania, odróżniającą je od innych rodzajów systemów. Proste systemy oprogramowania, oparte na klasycznych heurystykach lub podstawowych mechanizmach predykcyjnych, nie są traktowane jako systemy AI. Niektóre systemy pomimo posiadania zdolności wnioskowania w wąskim zakresie, mogą nie wchodzić w zakres definicji systemu AI. W takim przypadku głównym powodem ich wyłączenia spod omawianych regulacji będzie ich ograniczona zdolność do analizowania wzorców i autonomicznego dostosowywania swoich wyników.

Druga część opracowania dotyczy wytycznych KE w sprawie systemów zakazanych. Akt o sztucznej inteligencji wprowadza katalog ośmiu kategorii systemów, których wykorzystanie może prowadzić m.in. do manipulacji, eksploatacji słabości ludzkich, kontroli społecznej lub inwigilacji.

Wytyczne Komisji mają charakter niewiążący. Ostateczna interpretacja przepisów Aktu o sztucznej inteligencji należy wyłącznie do kompetencji Trybunału Sprawiedliwości Unii Europejskiej (TSUE).

Komisja podkreśla, że stosowanie wymogów Aktu wymaga indywidualnej oceny każdego przypadku z uwzględnieniem jego specyfiki, dlatego też wytyczne i zawarte w nich przykłady mają charakter jedynie orientacyjny, nie zastępując konieczności indywidualnej interpretacji przepisów w odniesieniu do każdego systemu AI.

Wstęp

W dniu 4 lutego 2025 r. Komisja Europejska wydała wytyczne w zakresie stosowania art. 5 Rozporządzenia UE 2024/1689 (Aktu o sztucznej inteligencji, dalej AIA), zawierającego katalog zakazanych praktyk z użyciem systemów sztucznej inteligencji. Dwa dni później, 6 lutego, Komisja opublikowała wytyczne dotyczące definicji systemów AI, zawartej w art. 3 ust. 1 rozporządzenia. Obydwa te przepisy wchodzi w skład pierwszej grupy regulacji AIA, których okres stosowania rozpoczął się w dniu 2 lutego 2025 r.

Celem wytycznych jest zwiększenie przejrzystości prawnej i dostarczenie wskazówek na temat interpretacji przepisów rozporządzenia. Mają one z jednej strony pomóc podmiotom objętym zakresem AIA w dostosowaniu się i zapewnieniu zgodności z jego wymogami, a z drugiej – ułatwić unijnym i państwowym organom monitorowanie rynku i egzekwowanie prawa.

W niniejszym opracowaniu omówione zostały najważniejsze elementy obydwu dokumentów Komisji.

Wytyczne w zakresie definicji systemu AI (art. 3 pkt 1 AIA)

MONIKA STACHOŃ

System AI w rozumieniu art. 3 pkt 1 AIA to system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne.

Definicja ta obejmuje siedem głównych elementów:

1. system oparty na maszynach,
2. który jest zaprojektowany do działania z różnym poziomem autonomii,
3. który może wykazywać zdolności adaptacyjne po wdrożeniu
4. i który, dla celów jawnych lub ukrytych
5. wnioskuje, z danych wejściowych, które otrzymuje, w jaki sposób generować dane wyjściowe,
6. takie jak prognozy, treści, zalecenia lub decyzje,
7. które mogą wpływać na środowisko fizyczne lub wirtualne.

Jak wskazano w wytycznych, tak sformułowana definicja systemu sztucznej inteligencji przyjmuje perspektywę opartą na cyklu życia, obejmującą dwie główne fazy:

- fazę przedwdrożeniową („budowania”) systemu oraz
- fazę powdrożeniową („użytkowania”) systemu.

Wyróżnionych siedem elementów nie musi występować nieprzerwanie w obu fazach cyklu życia systemu AI. Określone czynniki mogą pojawić się w jednej z nich, ale nie wystąpić w drugiej.

System oparty na maszynach

Sformułowanie **oparty na maszynach** odnosi się do faktu, że systemy sztucznej inteligencji są opracowywane i uruchamiane na maszynach, rozumianych jako komponenty hardware (sprzętowe) i/lub software (oprogramowanie), które umożliwiają działanie systemu AI. Element „oparty na maszynach” w definicji systemu AI

podkreśla fakt, że systemy AI muszą być napędzane obliczeniowo i oparte na operacjach maszynowych.

Autonomia systemu

Drugi element definicji odnosi się do systemu **zaprojektowanego do działania na różnych poziomach autonomii**. Oznacza to, że systemy sztucznej inteligencji są zaprojektowane do działania z pewnym stopniem niezależności od zaangażowania człowieka i mogą działać bez jego interwencji. To poziom autonomii jest warunkiem koniecznym do określenia, czy system kwalifikuje się jako system sztucznej inteligencji.

Zdolności adaptacyjne po wdrożeniu

Ważnym elementem definicji systemu sztucznej inteligencji jest to, że **może on wykazywać zdolności adaptacyjne po wdrożeniu**. Adaptacyjność odnosi się do zdolności samouczenia się, pozwalając na zmianę zachowania systemu podczas jego użytkowania. Nowe zachowanie dostosowanego systemu może dawać inne wyniki niż poprzedni system dla tych samych danych wejściowych.

Cele jawne lub ukryte

Kolejnym aspektem definicji są cele systemu AI. **Mogą być one zdefiniowane jawnie (wyraźnie) lub niejawnie (w sposób dorożumiany)**. Wyrażne cele odnoszą się do celów bezpośrednio zakodowanych przez dewelopera w systemie. Z kolei niejawne cele odnoszą się do takich, które nie są wyraźnie określone, ale można je wywnioskować z zachowania lub podstawowych założeń systemu.

Cele samego systemu AI mogą różnić się od zamierzonego celu systemu AI w określonym kontekście. Pierwsza kategoria jest wewnętrzna dla samego systemu, odnosząc się do zadań, które mają zostać wykonane i ich wyników. Z kolei zamierzony cel jest zorientowany zewnętrznie i obejmuje kontekst, w którym system ma zostać wdrożony i jak musi być obsługiwany.

Wnioskowanie

Kluczowa dla zapewnienia autonomii jest **zdolność wnioskowania systemu AI**, tj. jego zdolność do generowania wyników.

Zdolność do wnioskowania jest cechą konstytutywną systemów sztucznej inteligencji i niezbędnym warunkiem, który odróżnia je od innych rodzajów systemów. W definicji nie będą mieściły się zatem prostsze tradycyjne systemy oprogramowania lub podejścia programistyczne, które opierają się wyłącznie na zasadach zdefiniowanych przez osoby fizyczne w celu automatycznego wykonywania operacji.

Należy także podkreślić, że określenie „**wnioskować, w jaki sposób**”, odnosi się do:

1. procesu uzyskiwania wyników, takich jak przewidywania, treści, zalecenia lub decyzje, które mogą wpływać na środowiska fizyczne i wirtualne, a także do
2. zdolności systemów AI do wyprowadzania modeli lub algorytmów, lub obu, z danych wejściowych.

Wyniki takie jak prognozy, treści, zalecenia lub decyzje

Kolejnym elementem definicji systemu sztucznej inteligencji jest wnioskowanie „**w jaki sposób generować dane wyjściowe, takie jak prognozy, treści, zalecenia lub decyzje**”. Ta zdolność do generowania wyników w oparciu o otrzymane dane wejściowe i przy użyciu uczenia maszynowego oraz podejść logicznych i opartych na wiedzy ma fundamentalne znaczenie dla działania systemów sztucznej inteligencji i odróżnia te systemy od innych form oprogramowania.

Wyniki generowane przez systemy AI należą do czterech szerokich kategorii, które różnią się poziomem zaangażowania człowieka. Są to:

- przewidywania,
- treści,
- zalecenia,
- decyzje.

Przewidywania to oszacowania nieznanego wartości (wyjście) na podstawie znanych wartości dostarczonych do systemu (wejście). **Treści** odnoszą się do generowania nowych materiałów przez system sztucznej inteligencji. Może to obejmować tekst, obrazy, filmy, muzykę i inne formy danych wyjściowych. **Rekomendacje** to sugestie konkretnych działań, produktów lub usług dla użytkowników w oparciu o ich preferencje, zachowania lub inne dane wejściowe. Z kolei **decyzje** odnoszą się do wniosków lub wyborów dokonywanych przez system. System sztucznej inteligencji, który

podejmuje decyzję, automatyzuje procesy, które są tradycyjnie obsługiwane przez ludzki osąd.

Wpływ na środowisko fizyczne lub wirtualne

Ostatnia część definicji odnosi się do **wpływu danych wyjściowych systemu AI na środowisko fizyczne lub wirtualne**. Element ten należy rozumieć jako podkreślenie faktu, że systemy sztucznej inteligencji nie są bierne, ale aktywnie wpływają na środowiska, w których są wdrażane.

Wytyczne w zakresie zakazanych praktyk (art. 5 AIA)

EMILIA ZALEWSKA-CZAJCZYŃSKA

Doprecyzowanie przepisów z części ogólnej (Rozdział I)

W pierwszej kolejności w wytycznych przedstawiono uwagi dotyczące aspektów istotnych dla zrozumienia art. 5 w kontekście całości AIA. Obejmują one wyjaśnienia w zakresie kluczowych pojęć, takich jak:

- **wprowadzanie do obrotu** – obejmuje wszelkie formy udostępniania systemów AI po raz pierwszy na rynku Unii, niezależnie od formy ich dostarczania (np. poprzez chmurę lub kopię fizyczną);
- **oddanie do użytku** – oznacza dostarczenie systemu AI do pierwszego użycia przez użytkownika lub do użytku własnego w Unii w określonym celu;
- **wykorzystanie** – należy je rozumieć jako użycie lub wdrożenie systemu w dowolnym momencie jego cyklu życia po wprowadzeniu go na rynek lub oddaniu do użytku; może obejmować również integrację systemu AI w usługach i procesach podmiotów, które z niego korzystają.

W wytycznych podano również dodatkowe wskazówki dotyczące pojęć „dostawcy” i „podmiotu stosującego” oraz katalogu włączeń ze stosowania Aktu (wymienionych w art. 2 ust. 3–6 i 8–12, takich jako obszar celów bezpieczeństwa narodowego, woj-skowych i obronnych, badania naukowe, pozaprofesjonalna działalność osoby fizycznej itd.).

Choć art. 5 AAI obowiązuje od 2 lutego 2025 r., rozdziały dotyczące zarządzania, egzekwowania przepisów i kar będą stosowane dopiero od 2 sierpnia 2025 roku. Z tego względu przepisy dotyczące kar za naruszenie zakazów z art. 5 również nie będą miały zastosowania przed 2 sierpnia 2025 roku.

Mimo to nawet w tym okresie przejściowym zakazy obejmują dostawców i podmioty stosujące systemy AI, w związku z czym powinni oni podjąć niezbędne środki, aby zapewnić, że nie wprowadzą na rynek, nie oddadzą do użytku ani nie będą stosować systemów AI w sposób, który może stanowić zabronioną praktykę. **Art. 5 umożliwia podmiotom dotkniętym jego naruszeniem egzekwowanie przestrzegania zakazów przed krajowymi sądami oraz wystąpienie o zasądzenie środków zabezpieczających przeciwko zabronionym praktykom.**

Art. 5 ust. 1 lit. a – szkodliwa manipulacja, wprowadzenie w błąd i wykorzystanie

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. a, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku lub wykorzystanie systemu AI.
2. System AI stosuje techniki klasyfikujące się jako podprogowe, celowo manipulacyjne lub wprowadzające w błąd.
3. Techniki stosowane przez system AI mają na celu lub skutkują znaczącą zmianą zachowania osoby lub grupy osób w sposób, który znacznie ogranicza ich zdolność do podjęcia świadomej decyzji, co sprawia, że podejmują decyzję, której nie podjęłyby w innych okolicznościach.
4. Znacząca zmiana zachowania powoduje lub istnieje uzasadnione prawdopodobieństwo, że spowoduje u tej osoby, grupy osób lub osoby trzeciej poważną szkodę.
5. Pomiędzy technikami zastosowanymi przez system AI, znaczącą zmianą zachowania osoby lub grupy osób oraz poważną szkodą istnieje wiarygodny związek przyczynowy.

Techniki podprogowe, celowo manipulacyjne i wprowadzające w błąd

Techniki podprogowe – oddziałują poza świadomą percepcją danej osoby. Oznacza to, że taka osoba pozostaje nieświadoma ich wpływu na podejmowane przez nią decyzje lub na kształtowanie jej wartości lub opinii.

Techniki celowo manipulacyjne – zostały zaprojektowane lub zastosowane w celu kształtowania, zmieniania lub kontrolowania zachowania danej osoby w sposób, który podważa jej indywidualną autonomię i swobodę wyboru. Wykorzystują one uprzedzenia poznawcze, podatności psychologiczne lub czynniki sytuacyjne, które sprawiają, że osoba jest bardziej podatna na ich wpływ. Manipulacja nie musi mieć na celu wyrządzenia szkody, wystarczy, że dąży do wywołania zmiany zachowania osoby poza jej kontrolą. Zakaz obejmuje także przypadki stosowania technik celowo manipulacyjnych przez system AI, w których robi on to bez intencji jego dostawcy lub podmiotu stosującego.

Techniki wprowadzające w błąd – mają na celu oszukać daną osobę i w ten sposób podważyć lub osłabić jej autonomię czy też wpłynąć na podejmowane przez nią decyzje, np. podając jej fałszywe informacje. Działają one w sposób, którego osoba nie jest świadoma lub, jeżeli ma świadomość stosowania techniki, nie jest zdolna oprzeć się jej wpływowi. Zakaz stosowania technik wprowadzających w błąd może dotyczyć także systemów AI, które stosują je bez intencji ich dostawcy lub podmiotu stosującego. Nie stanowią jednak naruszenia zakazu okazjonalne przypadki podania przez

system AI nieprawidłowych lub wprowadzających w błąd informacji lub halucynowania przez system.

W – zakaz z art. 5 ust. 1 lit. a obejmuje również przypadki, gdy system AI stosuje różne techniki łącznie (np. podprogowe z manipulacyjnymi).

Cel lub skutek w postaci znaczącej zmiany zachowania

Cel – rzeczywiste zamiary dostawcy lub podmiotu stosującego co do wykorzystania systemu AI, innymi słowy intencja stojąca za wykorzystaniem systemu AI. Nie jest więc to pojęcie tożsame z „**przeznaczeniem**” systemu AI, które zgodnie z definicją z art. 3 ust. 12 AIA, oznacza zastosowanie, do jakiego system AI został przeznaczony przez jego dostawcę, w tym konkretny kontekst i warunki wykorzystywania, określone w informacjach dostarczonych w instrukcji obsługi itd.

Skutek – może wystąpić niezależnie od tego, czy dostawca lub podmiot stosujący posiadają zamiar osiągnięcia efektu w postaci znaczącej zmiany zachowania.

Skutek działania systemu AI nie musi się w pełni zmaterializować, aby przyjąć, że przesłanka jego wystąpienia została spełniona. Wystarczy, że istnieje **dostatecznie duże prawdopodobieństwo**, że skutek wystąpi w przyszłości, oparte o istniejącą wiedzę i metody naukowe, a także na podstawie dostępnych danych o rzeczywistym wpływie systemu na zachowanie jednostek.

Konieczne jest, aby istniał **wiarygodny lub dostatecznie prawdopodobny związek przyczynowy** pomiędzy tym działaniem systemu AI a jego potencjalnym lub rzeczywistym wpływem na zachowanie.

Znacząca zmiana zachowania – występuje w sytuacji, w której zastosowane techniki wyraźnie osłabiają zdolność jednostek do podejmowania świadomych decyzji. Wyraźne osłabienie odnosi się do poważnego zniekształcenia procesu decydowania i kształtowania opinii, wykraczającego poza drobne lub nieistotne skutki.

Wystąpienie lub uzasadnione prawdopodobieństwo wystąpienia poważnej szkody

Poważna szkoda może mieć charakter:

- **fizyczny** – jakiegokolwiek zranienie lub uszczerbek dla zdrowia i życia osoby, a także szkoda na mieniu; ma natychmiastowe, poważne i nieodwracalne konsekwencje;

- **psychologiczny** – negatywny wpływ na zdrowie psychiczne i dobrostan emocjonalny osoby; jej skutki mogą kumulować się w czasie i nie być od razu widoczne; pomiar jej wielkości powinien uwzględniać indywidualne okoliczności danego przypadku;
- **ekonomiczny** – straty finansowe, wykluczenie lub niestabilność ekonomiczna.

System AI może spowodować **kilka szkód łącznie**, które dopiero jako całość mogą zostać uznane za wypełniające znamiona poważnej szkody.

Przykład: system AI, który spowoduje szkodę psychologiczną u osoby (np. traumę, depresję), co prowadzi do doznania przez nią szkody fizycznej (np. znaczne pogorszenie zdrowia fizycznego).

Za poważną szkodę mogą zostać uznane zarówno **natychmiastowe i bezpośrednie efekty** działania systemu, jak i te **systemowe, rozłożone w czasie i niebezpośrednie**.

Oceniając, czy szkoda jest poważna, należy wziąć pod uwagę **stopień**, w jakim szkoda jest skutkiem lub prawdopodobnym wynikiem działania systemu AI, **kontekst i łączne efekty** działania systemu AI, **skalę i intensywność** negatywnych skutków – np. liczbę osób, których dotyczy – **podatność osób poszkodowanych** – np. czy należą do grup szczególnie wrażliwych takich jak dzieci – oraz **trwałość i odwracalność** – np. czy szkoda jest długotrwała i możliwa do całkowitego naprawienia.

Istnienie wiarygodnego związku przyczynowego

Aby ocenić, czy występuje wiarygodny związek przyczynowy pomiędzy działaniem systemu AI oraz poważną szkodą, należy wziąć pod uwagę, czy dostawca lub podmiot stosujący mógł **racjonalnie przewidzieć wystąpienie poważnej szkody** jako rezultat działania systemu AI, wdrożyć **odpowiednie środki zabezpieczające lub zmniejszające ryzyko** wystąpienia poważnej szkody oraz w jakim stopniu zapewnił **przejrzystość** systemu AI i jego sposobu działania.

Art. 5 ust. 1 lit. b – szkodliwe wykorzystanie słabości

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. b, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku lub wykorzystanie systemu AI.

2. System AI wykorzystuje słabości osoby lub grupy osób wynikające z wieku, niepełnosprawności lub sytuacji społeczno-ekonomicznej.
3. Wykorzystywanie tych słabości przez system AI ma na celu lub skutkuje dokonaniem znaczącej zmiany zachowania osoby lub grupy osób.
4. Znacząca zmiana zachowania powoduje lub istnieje uzasadnione prawdopodobieństwo, że spowoduje u tej osoby, grupy osób lub osoby trzeciej poważną szkodę.
5. Pomiędzy wykorzystaniem słabości przez system AI, znaczącą zmianą zachowania osoby lub grupy osób oraz poważną szkodą istnieje wiarygodny związek przyczynowy.

Wykorzystanie słabości wynikające z wieku, niepełnosprawności lub sytuacji społeczno-ekonomicznej

Słabość – szeroki zakres podatności, obejmujących m. in. słabości poznawcze, emocjonalne i fizyczne, które mogą wpływać na zachowanie osoby lub grupy osób lub ich zdolność do podejmowania świadomych decyzji. Zakaz z tego przepisu odnosi się wyłącznie do osób posiadających słabości ze względu na przynależność do jednej wskazanych z nim kategorii (wiek, niepełnosprawność lub sytuacja społeczno-ekonomiczna).

Przesłanka słabości wynikających z **wieku** dotyczy dwóch grup: **dzieci** – jako osób poniżej 18. roku życia, o ograniczonym stopniu rozwoju poznawczego i społeczno-emocjonalnego oraz **osób starszych**, które mogą mieć ograniczone funkcje poznawcze lub nie rozumieć współczesnych technologii.

Przesłankę **niepełnosprawności** należy rozumieć szerokie spektrum długoterminowych problemów fizycznych, intelektualnych i sensorycznych, które w połączeniu z innymi barierami utrudniają dotkniętym nimi jednostkom uczestnictwo w społeczeństwie na równych zasadach z innymi.

Przesłanka **szczególnej sytuacji społecznej lub ekonomicznej** oznacza przynależność (formalną lub faktyczną) do określonej grupy społecznej lub ekonomicznej na podstawie stosunkowo stabilnych i długoterminowych cech – np. osoby żyjące w skrajnym ubóstwie, mniejszości etniczne i religijne, bezrobotni. Do tej kategorii nie zaliczają się sytuacje, które nie mają charakteru społeczno-ekonomicznego np. dotyczące osób pogrążonych w żałobie, samotnych.

Wystąpienie lub (uzasadnione prawdopodobieństwo wystąpienia) poważnej szkody

Poważna szkoda – w przypadku grup szczególnie wrażliwych szkody wywołane przez zakazane praktyki w zakresie AI mogą być szczególnie dotkliwe i wieloaspektowe. Z tego względu te same okoliczności, które można uznać za niosące akceptowalne ryzyko szkody dla przeciętnej osoby, w przypadku jednostki należącej do grupy wrażliwej mogą być już nieakceptowalne (np. inny będzie dopuszczalny próg ryzyka w przypadku dorosłych niż dzieci). Ponadto w przypadku szkód dotyczących dzieci należy zwrócić uwagę, że mogą być znacząco rozłożone w czasie, w szczególności jeśli wiążą się ze skutkami dla ich rozwoju i dobrostanu.

Przykład: system AI zaprojektowany w sposób antropomorficzny symuluje ludzkie reakcje emocjonalne w interakcjach z dziećmi, co może prowadzić do niezdrowego przywiązania i zaburzenia postrzegania prawdziwych relacji międzyludzkich.

Pomimo że w przeciwieństwie do art. 5 ust. 1 lit. a, art. 5 ust. 1 lit. b nie odnosi się do możliwości wyrządzenia szkody także grupie osób (a nie tylko jednostce), oba zakazy należy interpretować w sposób spójny, zgodny z logiką i celem AIA, jakim jest ochrona wszystkich osób należących do określonych grup szczególnie wrażliwych.

Relacje pomiędzy art. 5 ust. 1 lit. a i lit. b

Zakaz, o którym mowa w art. 5 ust. 1 lit. a, koncentruje się na **charakterze technik, które są wykorzystywane przez systemy AI**. Kluczowymi cechami zakazanej praktyki są **ukryty charakter wpływu systemu AI i skutki działania podważające autonomię osoby w zakresie podejmowania świadomych decyzji**.

Z kolei zakaz, o którym mowa w art. 5 ust. 1 lit. b, kładzie nacisk na **ochronę konkretnych rodzajów grup wrażliwych przed wykorzystaniem ich słabości przez systemy AI**. Kluczowymi cechami zakazanej praktyki są **charakterystyka osób**, na które systemy oddziałują i fakt, że **ich słabości są wykorzystywane**.

W przypadkach, gdy nie ma pewności, który zakaz powinien mieć zastosowanie, podstawowym kryterium różnicowania powinien być dominujący sposób wykorzystania przez system AI. Jeżeli nie jest on nakierowany na wykorzystywanie konkretnych słabości osoby lub grupy, art. 5 ust. 1 lit. a będzie miał zazwyczaj zastosowanie w pierwszej kolejności.

Poza zakresem stosowania art. 5 ust. 1 lit. a i b

Poza zakresem zakazów zawartych w art. 5 ust. 1 lit. a i b znajdują się systemy AI, które:

- **stosują zgodne z prawem przekonywanie** – w odróżnieniu od manipulacji, przekonywanie odbywa się w sposób transparentny i z poszanowaniem autonomii danej osoby; obejmuje ono prezentowanie argumentów lub informacji w sposób, który apeluje do rozsądku lub emocji, wyjaśnia cele systemu AI i jego sposób działania, tak, aby użytkownik mógł samodzielnie podjąć decyzję na podstawie otrzymanych informacji;
- **są manipulacyjne, wprowadzające w błąd i wykorzystujące, ale nie ma co do nich prawdopodobieństwa, że spowodują poważną szkodę** – np. chatbot terapeutyczny, który stosuje techniki podprogowe, by skłonić użytkowników do zdrowszego stylu życia, np. rzucenia palenia.

Art. 5 ust. 1 lit. c – scoring społeczny

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. c, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku lub wykorzystanie systemu AI.
2. System AI musi być przeznaczony lub wykorzystywany do oceny lub klasyfikacji osób fizycznych lub grup osób w określonym czasie na podstawie ich zachowań społecznych lub znanych, wywnioskowanych, lub przewidywanych cech osobistych lub osobowościowych.
3. Wynik uzyskany przy pomocy systemu AI musi prowadzić lub być w stanie prowadzić do krzywdzącego lub niekorzystnego traktowania osób lub grup w co najmniej jeden z poniższych sposobów:
 - a. w kontekstach społecznych niezwiązanych z tymi, w których dane zostały pierwotnie wygenerowane lub zebrane;
 - b. traktowania nieuzasadnionego lub nieproporcjonalnego w stosunku do zachowania społecznego lub jego wagi.

Wykorzystanie systemu AI do oceny lub klasyfikacji osób fizycznych lub grup w określonym czasie na podstawie ich zachowań społecznych lub cech

Ocena – dokonanie jakiejś formy osądu na temat osoby lub grupy osób.

Klasyfikacja – pojęcie szersze niż ocena; oznacza prostą kategoryzację osób lub grupy osób na podstawie ich cech, takich jak ich wiek, płeć i wzrost.

Profilowanie – wykorzystywanie informacji o osobie fizycznej lub grupie osób i ocenę ich cech lub wzorców zachowań w celu umieszczenia ich w określonej kategorii lub grupie, w szczególności w celu analizy lub prognozowania, np. ich zdolności do wykonania zadania, zainteresowań lub prawdopodobnego zachowania. Profilowanie dokonywane z wykorzystaniem systemów AI może być objęte zakresem zakazu z art. 5 ust. 1 lit. c AIA, choć nie jest wprost wskazane w przepisie.

Określony czas – w kontekście oceny lub klasyfikacji oznacza, że nie powinny one ograniczać się do jednorazowego oceniania danych lub zachowań w bardzo konkretnym kontekście.

Przykład: organ zajmujący się migracją i azylem wdraża częściowo zautomatyzowany system nadzoru w obozach dla uchodźców, który zbiera dane w pewnym odcinku czasu, podczas którego konkretne osoby są oceniane pod kątem ryzyka ucieczki.

Ocena i klasyfikacja muszą być dokonywane w odniesieniu do **zachowań społecznych lub znanych, wywnioskowanych lub przewidywanych cech osobistych i osobowościowych**. Mogą też dotyczyć obydwu elementów jednocześnie. Dane na ich temat mogą być dostarczane bezpośrednio przez osoby, których dotyczą lub gromadzone pośrednio, tj. poprzez nadzór, uzyskane od osób trzecich lub na podstawie wnioskowania z innych informacji.

Zachowania społeczne – szeroki termin, który może obejmować działania, zachowania, nawyki, interakcje społeczne itp. Pojęcie to może dotyczyć zachowań osób i grup osób w różnych kontekstach: prywatnym (np. podczas wydarzeń kulturalnych, wolontariatu), biznesowym (np. podczas korzystania z usług, spłaty długów), jak i w relacjach z podmiotami publicznymi, policją i prawem (np. przestrzeganie przepisów ruchu drogowego).

Cechy osobiste – mogą obejmować różne informacje dotyczące danej osoby, np. o płci, orientacji seksualnej, rasie, sytuacji rodzinnej, miejscu zamieszkania, dochodzie, poziomie zadłużenia itd.

Cechy osobowości – zasadniczo synonim cech osobistych, ale może odnosić się także tworzenia profili osobowościowych.

Znane cechy – to informacje i dane osobowe, które zostały dostarczone do systemu AI jako dane wejściowe i które w większości przypadków są informacjami weryfikowalnymi.

Cechy wywnioskowane – opierają się na informacjach, które zostały wywnioskowane na podstawie innych informacji, przy czym wnioskowanie jest zwykle dokonywane przez system AI.

Przewidywane cechy – są szacowane na podstawie wzorców z dokładnością mniejszą niż 100%.

Wynik prowadzący (lub mogący prowadzić) do szkodliwego lub niekorzystnego traktowania

Konieczne jest, aby istniał **związek przyczynowy** pomiędzy niekorzystnym lub krzywdzącym traktowaniem a scoringiem dokonany przez system AI lub z jego pomocą. Taki związek może istnieć również w przypadkach, w których szkodliwe konsekwencje jeszcze się nie zmaterializowały, ale system AI ma na celu lub jest w stanie wywołać taki niekorzystny wynik dla osoby lub grupy. Scoring dokonany przez system AI nie musi być jedyną przyczyną niekorzystnego lub krzywdzącego traktowania, aby narużyć zakaz – wystarczy, że odgrywa istotną rolę w procesie oceny.

Przykład: jeżeli organ publiczny wdraża system AI do oceny wiarygodności osób i łączy jego wynik z ludzką oceną na podstawie dodatkowych faktów, to taka praktyka będzie podlegała zakazowi tylko wtedy, gdy wynik generowany przez AI odgrywa wystarczająco ważną rolę w ostatecznej decyzji.

Krzywdzące lub niekorzystne traktowanie kontekstach społecznych niezwiązanych z tymi, w których dane zostały pierwotnie wygenerowane lub zebrane

Niekorzystne traktowanie – oznacza, że w wyniku oceny dana osoba lub grupa osób jest traktowana mniej korzystnie w porównaniu z innymi, co niekoniecznie musi wiązać się z konkretną szkodą lub uszczerbkiem (np. nałożenie dodatkowych kontroli).

Krzywdzące traktowanie – sytuacja, w której dana osoba lub grupa osób doznała określonej szkody lub uszczerbku w wyniku takiego traktowania.

Niezwiązane konteksty społeczne (z zebraniem / wygenerowaniem danych) – to takie, w których brak jest widocznego związku pomiędzy celem oceny a celem zebrania lub wygenerowania danych, w których system AI wykorzystuje te dane w sposób prowadzący do ogólnej inwigilacji danej osoby lub grupy.

Przykład: agencja opieki społecznej wykorzystuje system AI do szacowania prawdopodobieństwa nadużyć ze strony beneficjentów zasiłków dla gospodarstw domowych, który dokonuje oceny się na podstawie np. faktu posiadania małżonka określonej narodowości lub pochodzenia etnicznego.

Krzywdzącego lub niekorzystnego traktowania niektórych osób fizycznych lub grup osób, które jest nieuzasadnione lub nieproporcjonalne do ich zachowania społecznego lub jego wagi

Nieuzasadnione lub nieproporcjonalne traktowanie – ma miejsce, jeżeli z porównania skali wpływu i ingerencji danego działania w prawa podstawowe osoby lub grupy z wagą danego zachowania społecznego wynika, że działanie to jest nieproporcjonalne w stosunku do realizowanego celu, zgodnie z ogólną zasadą proporcjonalności i przy uwzględnieniu wartości etycznych i zasad sprawiedliwości społecznej. Za nieuzasadnione można uznać również traktowanie, które nie realizuje zgodnego z prawem celu.

Przykład: agencja rządowa wykorzystuje system AI do profilowania rodzin w celu wczesnego wykrywania dzieci „w stanie zagrożenia” w oparciu o kryteria takie jak zachowania społeczne rodziców (np. opuszczanie wizyt lekarskich lub otrzymywanie mandatów drogowych).

Obie przesłanki, o których mowa w art. 5 ust. 1 lit. c ppkt (i) i (ii) mogą być spełnione jednocześnie, kiedy krzywdzące lub niekorzystne traktowanie ma miejsce w kontekstach społecznych niezwiązanych z tymi, w których dane zostały pierwotnie wygenerowane lub zebrane, oraz jest nieuzasadnione lub nieproporcjonalne do zachowania społecznego lub jego wagi.

Poza zakresem stosowania art. 5 ust. 1 lit. c

Poza zakresem zakazu zawartego w art. 5 ust. 1 lit. c znajdują się systemy AI, które:

- **dokonyją oceny osób innych niż fizyczne (np. prawnych)** – nawet jeśli w niektórych przypadkach taki scoring może mieć pośredni wpływ na osoby fizyczne (np. na obywateli gminy w przypadku przydzielania budżetu); zakaz będzie jednak

obowiązywał, jeśli scoring osoby prawnej będzie bezpośrednio wpływał na ocenę lub klasyfikację osób fizycznych (np. wszyscy pracownicy firmy);

- **agregują indywidualne oceny użytkowników, którzy oceniają jakość usługi** (np. kierowcy na platformie car-sharingowej), chyba że dane są łączone z innymi informacjami i analizowane przez system AI w celu oceny lub klasyfikacji poszczególnych osób lub grup;
- **dokonyują oceny osób fizycznych zgodnie z prawem Unii lub krajowym** (np. scoring kredytowy).

Art. 5 ust. 1 lit. d – ocena lub przewidywanie ryzyka popełnienia przestępstwa

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. d, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku w tym konkretnym celu lub wykorzystanie systemu AI.
2. System AI musi dokonywać oceny ryzyka w celu oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną.
3. Ocena ryzyka lub przewidywanie muszą opierać się wyłącznie na poniższych przesłankach (łącznie lub jednej z):
 - a. profilowanie osoby fizycznej;
 - b. ocena cech osobowości i cech charakterystycznych osoby fizycznej.

Ocena ryzyka w celu oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną

Ocena ryzyka – odnosi się różnorodnych, zaawansowanych technologii AI i metod analitycznych, operujących na dużych zbiorach danych (np. dane socjo-ekonomiczne, zapisy policyjne), które w połączeniu z wiedzą kryminologiczną wykorzystywane są do prognozowania popełniania przestępstw na potrzeby podejmowania przez organy ścigania działań z obszaru zapobiegania, kontroli i walki z przestępczością. Takie oceny ryzyka są ukierunkowane na przyszłość i dotyczą przestępstw jeszcze niepopełnionych lub które mogą być na etapie przygotowań do popełnienia.

Profilowanie osoby fizycznej

Profilowanie – należy je rozumieć w odniesieniu do definicji z art. 4 ust. 4 RODO, której jednym z kluczowych elementów jest cel w postaci „oceny niektórych czynników osobowych osoby fizycznej”.

Profilowania grupowe – tworzenie i stosowanie opisowego profilu dla określonego rodzaju sprawców przestępstw (np. terrorystów) na podstawie danych historycznych dotyczących przestępstw popełnionych wcześniej przez inne osoby. Takie profilowanie może podlegać pod zakaz, jeżeli jego wyniki są wykorzystywane do oceny konkretnej jednostki.

Cechy osobowościowe i charakterystyczne – każdy indywidualny wskaźnik lub informacja dot. sytuacji osobistej, socjoekonomicznej lub zawodowej, np. narodowość, miejsce urodzenia, liczba dzieci, poziom zadłużenia czy posiadany samochód.

Aby stwierdzić, czy dana praktyka narusza zakaz, należy zbadać, czy w ocenie ryzyka lub przewidywaniu zostały uwzględnione **inne elementy** niż profilowanie osoby fizycznej lub ocena jej cech osobowościowych i charakterystycznych. Jeżeli system AI wykorzystywany jest do wspierania ludzkiej oceny, która opiera się na **obiektywnych i weryfikowalnych faktach związanych bezpośrednio z działalnością przestępczą**, przesłanki zakazu nie będą spełnione.

Dopuszczone do stosowania systemy AI do oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną, sklasyfikowane jako systemy wysokiego ryzyka, muszą zostać poddane nadzorowi ludzkiemu. Funkcja ta musi zostać powierzona osobom posiadającym odpowiednie kompetencje, wykształcenie oraz uprawnienia, które będą w stanie zrozumieć zdolności oraz ograniczenia systemów AI, a także prawidłowo interpretować wyniki ich pracy, w tym przeciwdziałać ryzyku błędów stroniczości.

Zakres obowiązywania podmiotów prywatnych

Zakaz z art. 5 ust. 1 lit. d może, poza organami ścigania, w pewnych przypadkach obejmować działania podmiotów prywatnych. Dotyczy to w szczególności przypadków, gdy są one upoważnione na mocy prawa do wykonywania publicznych uprawnień w zakresie wykrywania lub ścigania przestępstw lub wykonywania kar lub wykonują te zadania w imieniu organów ścigania. Zakaz może dotyczyć także wykonywania przez podmioty prywatne prognozy lub oceny ryzyka popełnienia przestępstwa, dokonywanej w celu wypełniania ciężących na nich obowiązków prawnych, np. w zakresie przeciwdziałania praniu brudnych pieniędzy i finansowania terroryzmu.

Poza zakresem stosowania art. 5 ust. 1 lit. d

Poza zakresem zakazów zawartych w art. 5 ust. 1 lit. d znajdują się systemy AI, które:

- **prognozują przestępczość na podstawie położenia, danych przestrzennych lub lokalizacji** – np. na podstawie danych dotyczących przestępczości w danej okolicy, w zakresie, w jakim wyniki tych prognoz nie są wykorzystywane jako element profilowania osoby fizycznej;
- **służą do analizy i oceny** dokonywanej przez podmioty prywatne niezorientowanej na ocenę ryzyka popełnienia przestępstwa i **dokonywanej w ramach normalnej biznesowej działalności podmiotu prywatnego** np. ochrony interesów finansowych;
- **są wykorzystywane do prognozowania przestępstw i ocen w odniesieniu do podmiotów prawnych** – wyjątek ten może nie mieć zastosowania, jeżeli za pośrednictwem podmiotu prawnego działa osoba fizyczna (np. jest to jednoosobowe przedsiębiorstwo);
- **są wykorzystywane do indywidualnych prognoz wykroczeń administracyjnych.**

Art. 5 ust. 1 lit. e – nieukierunkowane pozyskiwanie wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. e, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku w tym konkretnym celu lub wykorzystanie systemu AI.
2. System AI musi służyć do tworzenia lub rozszerzania baz danych służących do rozpoznawania twarzy.
3. Środki służące do zapełnienia bazy danych zapewnione są z wykorzystaniem narzędzi AI do nieukierunkowanego pozyskiwania (untargeted scraping).
4. Źródłem obrazów jest internet lub nagrania z telewizji przemysłowej (CCTV).

Bazy danych służące do rozpoznawania twarzy

Baza danych – każdy zbiór danych lub informacji, uporządkowany w celu szybkiego wyszukiwania i porównywania danych. Taka baza umożliwia porównanie ludzkiej twarzy z obrazu cyfrowego lub klatki wideo z obrazami, które zawiera, oraz określenie, czy istnieje prawdopodobne dopasowanie między nimi. Może ona mieć charakter tymczasowy, scentralizowany lub zdecentralizowany. Nawet jeśli baza nie została stworzona wyłącznie na potrzeby rozpoznawania twarzy, ale może być do tego wykorzystywana, może podlegać zakazowi.

Nieukierunkowane pozyskiwanie obrazów twarzy

Nieukierunkowane pozyskiwanie (scraping) – wykorzystywanie narzędzi automatycznych (np. botów) lub innych narzędzi do masowego pobierania danych z różnych źródeł. Nieukierunkowane oznacza technikę działającą na zasadzie „odkurzacza”, poprzez wchłanianie jak największej ilości danych bez ukierunkowania na konkretną osobę. Przestrzeganie klauzuli opt-out protokołów internetowych, takich jak robot.txt, nie ma wpływu na nieukierunkowany charakter pozyskiwania.

Selektywne pozyskiwanie (np. identyfikacja konkretnej osoby) nie jest objęte zakresem zakazu. Jeżeli jednak zbieranie danych polega na tworzeniu bazy danych krok po kroku, wybierając za każdym razem inne grupy osób lub kryteria, a końcowy rezultat jest funkcjonalnie taki sam, jak gdyby zbieranie danych miało charakter nieukierunkowany, takie działanie może naruszać zakaz.

Z internetu lub nagrań z telewizji przemysłowej

W wytycznych nie zawarto wskazówek dotyczących rozumienia pojęć internet i nagrania z telewizji przemysłowej. Wskazano jedynie, że materiały z CCTV to np. obrazy z kamer w miejscach takich jak lotniska, oraz zwrócono uwagę, że fakt, że osoba opublikowała swoje zdjęcia twarzy na platformie społecznościowej, nie oznacza, że wyraziła zgodę na włączenie tych zdjęć do bazy danych służącej do rozpoznawania twarzy.

Poza zakresem stosowania art. 5 ust. 1 lit. e

Poza zakresem zakazów zawartych w art. 5 ust. 1 lit. e znajdują się systemy AI, które:

- są stosowane do **nieukierunkowanego pozyskiwania innych danych biometrycznych (np. próbek głosu)**;

- zbierają obrazy twarzy w celu generowania obrazów fikcyjnych osób, jeżeli nie prowadzi to do rozpoznawania rzeczywistych osób.

Zakaz nie odnosi się również do baz danych służących do rozpoznawania twarzy, stworzonych przed wejściem w życie przepisów Aktu o AI, które nie są dalej rozszerzane za pomocą nieukierunkowanego zbierania danych.

Art. 5 ust. 1 lit. F – wnioskowanie na temat emocji osoby fizycznej w miejscu pracy lub instytucjach edukacyjnych

Główne elementy zakazu

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. e, jeśli łącznie zostaną spełnione następujące warunki:

1. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku w tym konkretnym celu lub wykorzystanie systemu AI.
2. System AI musi być wykorzystywany do wnioskowania o emocjach osoby fizycznej na podstawie danych biometrycznych.
3. Praktyka musi być stosowana w miejscu pracy lub w instytucjach edukacyjnych.
4. Praktyka nie obejmuje zastosowania systemu AI w celach medycznych lub związanych z bezpieczeństwem.

System AI wykorzystywany do wnioskowania o emocjach osoby fizycznej

System rozpoznawania emocji – zgodnie z definicją z art. 3 ust. 39 Aktu o AI, to system, który służy do identyfikacji lub wywnioskowania emocji lub zamiarów osób fizycznych na podstawie danych biometrycznych tych osób. Z kolei w treści zakazu z art. 5 ust. 1 lit. f mowa wyłącznie **o systemach AI wykorzystywanych do wnioskowania o emocjach osoby fizycznej**.

Wnioskowanie – polega na dedukcji informacji na podstawie innych danych, poprzez np. procesy analityczne systemu. W kontekście emocji obejmuje zazwyczaj także **identyfikację** (przetwarzanie danych biometrycznych, np. głosu, osoby fizycznej pozwalające bezpośrednio porównać i zidentyfikować emocję z inną, wcześniej zaprogramowaną w systemie rozpoznawania emocji) jako pierwszy etap procesu. Dlatego

zakaz należy rozumieć jako **obejmujący zarówno systemy AI identyfikujące, jak i wnioskowanie o emocjach lub intencjach**.

Emocje – np. radość, smutek, złość, zaskoczenie, wstręt, zakłopotanie, podekscytowanie, wstyd, pogarda, satysfakcja i rozbawienie. Pojęcie nie odnosi się do stanów fizycznych (takich jak np. ból, zmęczenie) lub gestów (np. unoszenie ramion).

Przykład: system rejestrujący, że dana osoba się uśmiecha, nie jest systemem rozpoznawania emocji.

Na podstawie danych biometrycznych

Dane biometryczne – można o nich wywnioskować z cech osobistych, takich jak atrybuty fizyczne (np. odciski palców, wzór tęczówki) lub behawioralne (np. gesty, ruchy, w tym ruchy poza świadomą kontrolą, np. rytm bicia serca). Zakaz nie obejmuje praktyk wnioskowania o emocjach na podstawie innych danych niż biometryczne.

Przykład: system AI, który wnioskuje o emocjach na podstawie tekstu pisanego (np. poprzez analizę treści/sentymentu), nie podlega zakazowi; jednak system AI wnioskujący o emocjach na podstawie sposobu pisania na klawiaturze lub postawy ciała podlega zakazowi.

W miejscu pracy lub w instytucjach edukacyjnych

Miejsce pracy – każda przestrzeń fizyczna lub wirtualna, w której osoby wykonują zadania i obowiązki powierzone przez pracodawcę lub organizację, z którą są związane, np. biura, fabryki, stadiony. Zakres obejmuje wszystkie osoby niezależnie od statusu (np. pracownik, kontraktor, wolontariusz), w tym także kandydatów w procesie rekrutacji.

Instytucje edukacyjne – wszelkie instytucje publiczne i prywatne, gdzie odbywa się nauczanie, niezależnie od trybu (stacjonarnego, online, hybrydowego), które mają uprawnienia do wydawania certyfikatów ukończenia. Zakaz należy rozumieć jako mający zastosowanie również do kandydatów w trakcie procesu rekrutacji do danej instytucji.

Wyłączenia ze względów medycznych lub bezpieczeństwa

Systemy AI służące do wnioskowania o emocjach w miejscu pracy lub edukacji mogą być stosowane w uzasadnionych przypadkach związanych z **względami medycznymi**

(np. w celach terapeutycznych) lub **względnami bezpieczeństwa**, rozumianymi jako ochrona życia i zdrowia (a nie innych interesów, takich jak mienie).

Wyjątki te należy rozumieć bardzo wąsko. Np. ogólne monitorowanie z wykorzystaniem systemów AI poziomu stresu w miejscu pracy nie może zostać uznane za uzasadnione względami ochrony zdrowia lub bezpieczeństwa.

Każde zastosowanie w celach medycznych i bezpieczeństwa powinno być zawsze ograniczone do tego, co jest ściśle **konieczne i proporcjonalne** oraz opatrzone **odpowiednimi zabezpieczeniami**. Mogą one obejmować np. uprzednią pisemną i uzasadnioną opinię eksperta dotyczącą konkretnego przypadku użycia. Konieczność stosowania powinna być oceniana obiektywnie w odniesieniu do celu medycznego lub bezpieczeństwa, a nie odnosić się do „potrzeb” pracodawcy lub instytucji edukacyjnej. Taka ocena powinna uwzględniać zbadanie, czy istnieją mniej inwazyjne alternatywne środki, które osiągnęłyby ten sam cel.

Przykład: system AI do rozpoznawania emocji jest stosowany, aby pomóc pracownikom lub studentom z autyzmem albo poprawić dostępność dla osób niewidomych lub niesłyszących.

Poza zakresem stosowania art. 5 ust. 1 lit. f

Poza zakresem zakazu zawartego w art. 5 ust. 1 lit. f znajdują się systemy AI, które:

- wnioskuje o emocjach na podstawie **danych innych niż biometryczne**;
- wnioskuje o **stanach fizycznych**, takich jak ból i zmęczenie;
- wykorzystywane są w **innych obszarach** niż miejsca pracy i instytucje edukacyjne;
- wykorzystywane są do **kontroli tłumy** (rozumianej jako monitorowanie i kontrolowanie zachowań grup w celu utrzymania porządku publicznego i bezpieczeństwa podczas wydarzeń).

Art. 5 ust. 1 lit. g – kategoryzacja biometryczna na podstawie wrażliwych cech

Główne elementy zakazu

1. Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. g, jeśli łącznie zostaną spełnione następujące warunki:
2. Praktyka obejmuje wprowadzenie do obrotu, oddanie do użytku w tym konkretnym celu lub wykorzystanie systemu AI.
3. System musi być biometrycznym systemem kategoryzacji.
4. Poszczególne osoby muszą zostać skategoryzowane na podstawie ich danych biometrycznych.
5. Na podstawie kategoryzacji następuje wnioskowanie o ich rasie, poglądach politycznych, przynależności do związków zawodowych, przekonaniach religijnych lub filozoficznych, życiu seksualnym lub orientacji seksualnej.

Biometryczny system kategoryzacji

Kategoryzacja osoby fizycznej przez system biometryczny – proces ustalania, czy dane biometryczne osoby fizycznej należą do grupy o pewnej z góry określonej charakterystyce, w celu przypisania osoby do określonej kategorii.

Dane biometryczne – to dane osobowe będące wynikiem specjalnego przetwarzania technicznego, które dotyczą cech fizycznych, fizjologicznych lub behawioralnych osoby fizycznej, takich jak wizerunek twarzy lub dane daktyloskopijne (zgodnie z art. 3 ust. 34 AIA).

Kategoryzacja biometryczna może opierać się na cechach fizycznych (np. rysy i kształt twarzy, kolor skóry), DNA lub aspektach behawioralnych, takich jak analiza naciśnięć klawiszy lub chód osoby.

Osoby fizyczne muszą być **kategoryzowane indywidualnie**. Jeśli nie jest to celem lub wynikiem kategoryzacji biometrycznej, np. jeśli kategoryzowana jest grupa, zakaz nie ma zastosowania.

Przykład: systemem do indywidualnej kategoryzacji będzie system zdolny do kategoryzowania osób i wyodrębniania ich na podstawie określonej cechy (np. blizna na prawym oku).

Art. 5 ust. 1 lit. g zakazuje jedynie biometrycznych systemów kategoryzacji, których celem jest wnioskowanie o ograniczonej liczbie wrażliwych cech: rasie, poglądach politycznych, przynależności do związków zawodowych, przekonaniach religijnych lub filozoficznych, życiu seksualnym lub orientacji seksualnej.

Poza zakresem stosowania art. 5 ust. 1 lit. g

Poza zakresem zakazu zawartego w art. 5 ust. 1 lit. g znajdują się systemy AI, które:

- służą do **oznaczania lub filtrowania legalnie pozyskanych zbiorów danych biometrycznych w oparciu o dane biometryczne**, w tym w obszarze egzekwowania prawa;
- kategoryzują **osoby na podstawie odzieży lub akcesoriów**, np. szaliki, krzyże, a także aktywności w mediach społecznościowych;
- mające **charakter pomocniczy w stosunku do innej usługi komercyjnej i będące bezwzględnie konieczne z obiektywnych względów technicznych** – oznacza to, że nie mogą być one używane samodzielnie, a jedynie w związku z główną usługą, jaką jest sprzedaż produktu – np. filtry kategoryzujące cechy twarzy lub ciała używane na platformach e-commerce, które pozwalają klientowi zobaczyć, jak produkt będzie wyglądał na nim.

Art. 5 ust. 1 lit. h – zdalna identyfikacja biometryczna w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw

Dana praktyka z wykorzystaniem AI narusza zakaz z art. 5 ust. 1 lit. h, jeśli łącznie zostaną spełnione następujące warunki:

1. System AI jest systemem zdalnej identyfikacji biometrycznej w czasie rzeczywistym (RBI).
2. Praktyka obejmuje wykorzystanie systemu AI.
3. System AI działa w czasie rzeczywistym.

4. System AI działa w przestrzeni publicznej.
5. System AI jest wykorzystywany do celów ścigania przestępstw.

System zdalnej identyfikacji biometrycznej w czasie rzeczywistym (RBI)

System zdalnej identyfikacji biometrycznej – system AI służący do identyfikacji osób fizycznych poprzez porównanie danych biometrycznych danej osoby z danymi biometrycznymi zawartymi w bazie referencyjnej, czyli bazie zawierającej dane biometryczne do celów porównawczych. Np. baza Systemu Informacyjnego Schengen (SIS II) może zostać wykorzystana jako baza referencyjna do rozpoznawania twarzy.

Identyfikacji biometryczna – automatyczne rozpoznawanie cech fizycznych, fizjologicznych, behawioralnych lub psychologicznych człowieka w celu ustalenia tożsamości osoby fizycznej poprzez porównanie jej danych biometrycznych z danymi biometrycznymi przechowywanymi w bazie danych (zgodnie z art. 3 pkt 35 AIA).

Weryfikacja biometryczna – polega na porównaniu danych prezentowanych przy czujniku z innym zbiorem wcześniej zapisanych danych przechowywanych na urządzeniu, np. w smartfonie lub paszporcie w celu potwierdzenia, że dana osoba jest tym, za kogo się podaje. Nie jest objęta zakazem z art. 5 ust. 1 lit. h.

Zdalność systemów – zdolność do identyfikacji osób bez ich aktywnego udziału, zazwyczaj na odległość. Aby uznać, że udział jest aktywny, nie wystarczy, żeby osoby były poinformowane o obecności kamer – muszą one brać świadomy i aktywny udział w identyfikacji.

Przykład: systemy stosowane do uzyskania dostępu do metra, takie jak biometryczne bilety, które zakładają, że osoby aktywnie podchodzą do czujnika biometrycznego, aby uzyskać dostęp, nie spełniają warunku zdalności.

Działanie systemu AI w czasie rzeczywistym

Rzeczywisty czas działania – oznacza, że system przechwytuje i przetwarza dane biometryczne natychmiastowo lub prawie natychmiastowo, bez istotnego opóźnienia. Wszystkie etapy przetwarzania danych biometrycznych, tj. wychwytywanie, porównywanie i identyfikacja, odbywają się jednocześnie lub z jedynie krótkim, ograniczonym opóźnieniem.

Działanie systemu AI w przestrzeni publicznej

Przestrzeń publiczna to każda fizyczna przestrzeń będącą własnością publiczną lub prywatną, dostępną dla nieokreślonej liczby osób fizycznych, niezależnie od tego, czy obowiązują określone warunki dostępu, i niezależnie od ewentualnych ograniczeń pojemności (zgodnie z art. 3 pkt 44 AIA).

Przestrzenią publiczną może być zarówno własność publiczna, jak i prywatna. Np. kawiarnia lub lotnisko może być własnością prywatną, ale stanowić przestrzeń publiczną. Decydujące nie jest w tym zakresie to, czy przestrzeń jest faktycznie otwarta (np. odblokowane drzwi), jeżeli z przepisów prawa lub oznaczeń wynika, że jest to przestrzeń z ograniczonym dostępem.

Przestrzenią publiczną nie są przestrzenie internetowe (np. media społecznościowe). Nie są nimi także więzienia i punkty kontroli granicznej.

Wykorzystywanie systemu AI do celów ścigania przestępstw

Ściganie przestępstw – działania prowadzone przez organy ścigania lub w ich imieniu w celu zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania lub ścigania czynów zabronionych lub wykonywania kar, w tym ochrony przed zagrożeniami dla bezpieczeństwa publicznego i zapobiegania takim zagrożeniom (zgodnie z art. 3 pkt 46 AIA).

Organy ścigania – organy publiczne lub inne organy lub podmioty, którym na podstawie prawa państwa członkowskiego powierzono sprawowanie władzy publicznej i wykonywanie uprawnień publicznych do celów ścigania przestępstw. Prowadzenie działań w imieniu organów ścigania ma miejsce w sytuacji, w której delegują one wykonanie czynności związanych ze ściganiem przestępstw do innego podmiotu lub osoby, w tym podmiotu prywatnego, lub poproszą je o wsparcie przy takich czynnościach. W obu przypadkach organy ścigania muszą przekazać instrukcje dotyczące wszystkich istotnych aspektów czynności i nadzorować jej wykonanie przez inne podmioty.

Jeżeli podmioty działają we własnym imieniu w celu wykrywania i przeciwdziałania przestępstwom (np. prania pieniędzy), nie podlega to pod zakres zakazu z art. 5 ust. 1 lit. h.

Wyjątki od zakazu

Zasadniczym warunkiem wykorzystania systemów zdalnej identyfikacji biometrycznej jest to, aby było to bezwzględnie konieczne do osiągnięcia jednego z trzech

wskazanych w art. 5 ust. 1 lit. h ppkt (i)–(iii) celów: **ukierunkowanego poszukiwania ofiar konkretnych przestępstw i osób zaginionych, zapobiegania bezpośrednim zagrożeniom życia lub bezpieczeństwa oraz lokalizacji i identyfikacji podejrzanych i sprawców niektórych poważnych przestępstw**. W tych przypadkach działania te są niezakazane na gruncie Aktu o AI pod warunkiem, że **istnieją w tym zakresie krajowe regulacje dopuszczające takie działania i są one zgodne z art. 5 ust. 2–7 AIA**.

Także inne kwestie dotyczące wyjątków od zakazu będą uzależnione od ustawodawstwa krajowego, np. zdefiniowanie i ocena tego, jakie zdarzenia można zakwalifikować jako bezpośrednie zagrożenie życia lub bezpieczeństwa fizycznego, a także ocena poziomu zagrożenia atakiem terrorystycznym.

Istnieją powiązanie pomiędzy poszczególnymi włączeniami, np.:

- zakaz z art. 5 ust. 1 lit. h ppkt (i) oraz z art. 5 ust. 1 lit. h ppkt (iii) – systemy zdalnej identyfikacji biometrycznej mogą zostać wykorzystane zarówno do lokalizacji ofiary, jak i sprawcy lub podejrzanego w sprawie handlu ludźmi (przestępstwo to znajduje się w załączniku II rozporządzenia);
- zakaz z art. art. 5 ust. 1 lit. h ppkt (ii) oraz z art. 5 ust. 1 lit. h ppkt (iii) – system zdalnej identyfikacji biometrycznej można wykorzystać do zapobiegnięcia zagrożeniu życia, a jeśli takie nastąpi, to może posłużyć do zidentyfikowania lub zlokalizowania uciekającego sprawcy;
- retrospektywne wykorzystanie systemów zdalnej identyfikacji biometrycznej nie jest objęte zakazem z art. 5, ale podlega przepisom dotyczącym systemów AI wysokiego ryzyka.

Warunki do wyjątków od zakazu z art. 5 ust. 1 lit. h i środki ochronne

Potwierdzenie tożsamości konkretnie wybranej osoby – w przeciwieństwie do identyfikacji, o której mowa w zakazie z art. 5 ust. 1 lit. h, ma na celu ochronę praw podstawowych poprzez nakierowanie wykorzystania systemu AI na daną osobę i ograniczanie w ten sposób ryzyka szerokiej inwigilacji. Pojęcie to powinno być rozumiane w ten sposób, że użycie systemu RBI w czasie rzeczywistym może być rozpoczęte tylko w celu poszukiwania konkretnych osób, co do których organy ścigania mają podstawy, by sądzić, że są ofiarami przestępstw wymienionych w artykule 5 ust. 1 lit. h ppkt (i) lub są zaangażowane w jedną z sytuacji, o których mowa w ppkt (ii) i (iii) tego przepisu.

W przypadku użycia systemu RBI do zapobiegania bezpośrednim zagrożeniom życia lub bezpieczeństwa, lub lokalizacji i identyfikacji podejrzanych i sprawców niektórych poważnych przestępstw organy ścigania nie muszą znać tożsamości osób, których szukają, zanim użyją systemu. Jeśli posiadają informacje o ataku terrorystycznym zaplanowanym przez daną grupę (nie wiedząc, kto konkretnie go wykona) w określonym czasie i miejscu, taki system może zostać użyty do zidentyfikowania sprawcy z tej grupy, pod warunkiem, że organy ścigania stworzyły bazę referencyjną zawierającą dane biometryczne osób wchodzących w skład tej grupy.

Przed wykorzystaniem systemu należy dokonać oceny charakteru sytuacji, która może prowadzić do jego użycia, w tym powagę, prawdopodobieństwo i skalę szkody, która zostałaby wyrządzona w przypadku niewykorzystania tego systemu, a następnie porównać ją z konsekwencjami dla praw i wolności osób, których dotyczy dana sytuacja, w tym powagę, prawdopodobieństwo i skalę tych konsekwencji. Dokonując porównania konsekwencji użycia systemu RBI z naruszeniem praw podstawowych, organy ścigania powinny sprawdzić, czy są dla nich dostępne **mniej inwazyjne, alternatywne rozwiązania**.

Użycie systemów RBI powinno być także ograniczone pod względem:

- **zasięgu geograficznego** – do wyraźnie określonego obszaru, co do którego istnieją podstawy do stwierdzenia, że dane zdarzenie ma mieć tam miejsce;
- **osób, których dotyczy** – do wyraźnie określonych osób lub kategorii osób;
- **czasu trwania** – do okresu, gdy ich zastosowanie jest absolutnie konieczne i który jest jasno oznaczony w oparciu o informacje stanowiące podstawę do użycia systemów.

Przed zastosowanie systemu RBI, organy ścigania przestępstw muszą przeprowadzić **ocenę wpływu na prawa podstawowe (Fundamental Rights Impact Assessment)**, zgodnie z wymaganiami określonymi w art. 27 AIA. W okresie pomiędzy momentem rozpoczęcia obowiązywania zakazów z art. 5 (2 lutego 2025 r.) początkiem stosowania przepisów o systemach AI wysokiego ryzyka (do 2 sierpnia 2026 r.), wymagania z art. 27 powinny zostać wdrożone przez podmioty wykorzystujące systemy RBI w czasie rzeczywistym, które spełniają warunki uprawniające je do skorzystania z jednej lub więcej wyjątków określonych w Artykule 5 ust. 1 lit. h.

Ocena wpływu na prawa podstawowe (FRIA) – oznacza nowy rodzaj oceny oddziaływania, który ma na celu zidentyfikowanie wpływu, jaki niektóre systemy sztucznej inteligencji wysokiego ryzyka, w tym systemy RBI, mogą wywierać na prawa podstawowe. Nie zastępuje ona już funkcjonujących ocen skutków dla ochrony danych (DPIA). W przypadku, gdy system AI przetwarza dane osobowe (co ma miejsce

w przypadku systemów RBI), FRIA powinna uzupełniać DPIA, bez powielania kwestii już w niej uwzględnionych.

Obowiązek przeprowadzenia FRIA spoczywa na podmiocie stosującym systemy RBI, a nie na podmiotach działających w jego imieniu. Muszą one jednak udostępnić odpowiednie informacje, aby podmiot stosujący mógł dokonać prawidłowej oceny.

Podmiot wykorzystujący system RBI w czasie rzeczywistym, stosowany w publicznie dostępnych przestrzeniach do celów egzekwowania prawa, zobowiązany jest do **zarejestrowania systemu** w bazie danych UE, zgodnie z art. 49 AIA. W uzasadnionych przypadkach (np. w obliczu bezpośredniego zagrożenia) zastosowanie systemu może rozpocząć się przed rejestracją, pod warunkiem, że następnie organ ścigania zarejestruje system bez zbędnej zwłoki, czyli w najkrótszym możliwym czasie, uwzględniając okoliczności sytuacji kryzysowej, które uniemożliwiły wcześniejszą rejestrację systemu.

Wymóg wcześniejszego zezwolenia dla każdego użycia systemu RBI w czasie rzeczywistym w publicznych przestrzeniach do celów ścigania przestępstw ma na celu zapewnienie, że zostanie ono:

- ocenione pod kątem konieczności i proporcjonalności do osiągnięcia jednego z celów wymienionych w art. 5 ust. 1 lit. h ppkt (i)–(iii);
- ograniczone do absolutnie niezbędnego minimum pod względem czasu trwania, zakresu geograficznego oraz liczby objętych nim osób.

Ocena konieczności i proporcjonalności przed użyciem systemu RBI powinna mieć podwójny charakter:

- oceny przez użytkownika podczas wykonywania FRIA, zgodnie z art. 5 ust. 2;
- ocena przez organ sądowy lub niezależny organ administracyjny, z uwzględnieniem prawa krajowego, Karty Praw Podstawowych oraz innych przepisów unijnych.

W **nagłej sytuacji** system RBI może zostać użyty bez wcześniejszego zezwolenia. Oznacza to przypadki, w których uzasadniona potrzeba skorzystania z systemu czyni obiektywnie niemożliwym uzyskanie wcześniejszego zezwolenia. Takie użycie powinno być ograniczone do absolutnego minimum oraz podlegać odpowiednim zabezpieczeniom określonym w prawie krajowym, a organ ścigania musi **niezwłocznie złożyć wniosek o zezwolenie, nie później niż w ciągu 24 godzin**. Za moment rozpoczęcia używania należy uznać chwilę, w której kamery zostaną włączone i pierwsze porównanie biometryczne dokonane w systemie.

Wymóg uzyskania zezwolenia dotyczy **każdorazowego przypadku uruchomienia systemu RBI**.

Przepisy krajowe uprawniające organy ścigania przestępstw do wykorzystywania systemów RBI nie mogą przekraczać ograniczeń określonych w Artykule 5 ust. 1 lit. h A AI oraz muszą być zgodne z wszystkimi innymi, powiązanymi wymogami zawartymi w Akcie o AI. Oznacza to, że państwa nie mogą rozszerzać celów, dla których systemy RBI mogą być wykorzystywane w publicznie dostępnych przestrzeniach. Mają jednak kompetencje do wprowadzenia bardziej rygorystycznych przepisów.

W prawie krajowym dotyczącym wykorzystania systemów RBI państwa mogą określić takie kwestie, jak np. to, które organy publiczne posiadają uprawnienia na gruncie art. 5 ust. 1 lit. h do użycia takich systemów, lub jakie są minimalne progi niezawodności i dokładności systemów. Państwa mogą wprowadzić również dodatkowe wymogi lub sankcje w przepisach dotyczących uzyskiwania zezwolenia na wykorzystanie systemów RBI lub w zakresie procedur ich rejestracji.

Poza zakresem stosowania art. 5 ust. 1 lit. h

Poza zakresem zakazu zawartego w art. 5 ust. 1 lit. h znajdują się systemy AI, które:

- są systemami **weryfikacji/uwierzytelnienia biometrycznego**;
- obejmują **identyfikację retrospektywną** (nie w czasie rzeczywistym);
- są systemami RBI wykorzystywanymi **w przestrzeni prywatnej lub internetowej** (np. w grze online);
- są wykorzystywane przez **podmioty inne niż organy ścigania oraz nie działające w ich imieniu**.

The background is a dark blue gradient. It features a subtle, glowing grid of dots that recedes into the distance, creating a sense of depth. Overlaid on this grid are streams of white binary code (0s and 1s) that appear to be floating or falling, reminiscent of the 'Matrix' effect. The overall aesthetic is futuristic and digital.

NASK