KOMISJA
EUROPEJSKA

Bruksela, dnia 29.7.2025 r.
C(2025) 5052 final

KOMUNIKAT KOMISJI

Wytyczne Komisji w sprawie zakazanych praktyk w zakresie sztucznej inteligencji określonych w rozporządzeniu (UE) 2024/1689 (akt w sprawie sztucznej inteligencji)

SPIS TREŚCI

1. Informacje ogólne i cele 1
2. Przegląd zakazanych praktyk w zakresie AI 2
2.1. Zakazy wymienione w art. 5 aktu w sprawie sztucznej inteligencji 3
2.2. Podstawa prawna zakazów 4
2.3. Zakres przedmiotowy: praktyki związane z „wprowadzaniem do obrotu”, „oddawaniem do użytku” lub „wykorzystywaniem” systemu AI 4
2.4. Zakres podmiotowy: podmioty odpowiedzialne 6
2.5. Wyłączenia z zakresu stosowania aktu w sprawie sztucznej inteligencji 8
2.5.1. Cele bezpieczeństwa narodowego, obronne i wojskowe 8
2.5.2. Współpraca sądowa i współpraca organów ścigania z państwami trzecimi 10
2.5.3. Badania i rozwój 10
2.5.4. Osobista działalność pozazawodowa 12
2.5.5. Systemy AI udostępniane na podstawie bezpłatnych licencji otwartego oprogramowania 13
2.6. Wzajemne powiązania zakazów i wymogów dotyczących systemów AI wysokiego ryzyka 13
2.7. Stosowanie zakazów do systemów AI ogólnego przeznaczenia i systemów o określonym przeznaczeniu 14
2.8. Wzajemne powiązania między zakazami a innymi przepisami prawa Unii 16
2.9. Egzekwowanie art. 5 aktu w sprawie sztucznej inteligencji 19
2.9.1. Organy nadzoru rynku 19
2.9.2. Kary 20
3. Art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji – szkodliwe manipulowanie, wprowadzanie w błąd i wykorzystywanie 21
3.1. Uzasadnienie i cele 21
3.2. Główne elementy zakazu określonego w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji – szkodliwa manipulacja 22
3.2.1. Techniki podprogowe, celowe techniki manipulacyjne lub techniki wprowadzające w błąd 23
3.2.2. Cel lub skutek, jakim jest dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób 28
3.2.3. Wyrządzenie lub uzasadnione prawdopodobieństwo wyrządzenia poważnej szkody 32
3.3. Główne elementy zakazu określonego w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji – wykorzystywanie słabości w sposób wyrządzający szkodę 38
3.3.1. Wykorzystywanie słabości ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną 38
3.3.2. Cel lub skutek, jakim jest dokonanie znaczącej zmiany zachowania 44

3.3.3. Wyrządzenie lub uzasadnione prawdopodobieństwo wyrządzenia poważnej szkody 44
3.4. Wzajemne powiązania między zakazami określonymi w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji 49
3.5. Systemy nieobjęte zakresem 50
3.5.1. Zgodna z prawem perswazja 50
3.5.2. Manipulacyjne, wprowadzające w błąd i prowadzące do wykorzystania systemy AI, które prawdopodobnie nie wyrządzą poważnej szkody 53
3.6. Wzajemne powiązania z innymi przepisami prawa Unii 54
4. Art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji – scoring społeczny 58
4.1. Uzasadnienie i cele 59
4.2. Główne pojęcia i elementy zakazu „scoringu społecznego” 59
4.2.1. „Scoring społeczny”: ocena lub klasyfikacja oparta na zachowaniu społecznym lub cechach osobistych lub cechach osobowości przez określony czas 60
4.2.2. Scoring społeczny musi prowadzić do szkodliwego lub niekorzystnego traktowania w niepowiązanych kontekstach społecznych lub do traktowania, które jest nieuzasadnione lub nieproporcjonalne do wagi zachowania społecznego 63
4.2.3. Bez względu na to, czy są dostarczane lub wykorzystywane przez osoby publiczne lub prywatne 69
4.3. Systemy nieobjęte zakresem 70
4.4. Wzajemne powiązania z innymi aktami prawnymi Unii 73
5. Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji – indywidualna ocena ryzyka i przewidywanie ryzyka POPEŁNIENIA PRZESTĘPSTWA 75
5.1. Uzasadnienie i cele 75
5.2. Główne pojęcia i elementy zakazu 75
5.2.1. Ocena ryzyka lub przewidywanie prawdopodobieństwa popełnienia przestępstwa przez osobę 76
5.2.2. Wyłącznie na podstawie profilowania osoby fizycznej lub oceny jej cech osobowości i cech charakterystycznych 78
5.2.3. Wyłączenie systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny opartej na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą 80
5.2.4. Zakres, w jakim działania podmiotów prywatnych mogą wchodzić w zakres stosowania 82
5.3. Systemy nieobjęte zakresem 84
5.3.1. Prognozy przestępczości oparte na danych dotyczących lokalizacji, geoprzestrzennych lub dotyczących miejsca 84
5.3.2. Systemy AI wykorzystywane do wspierania dokonywanej przez człowieka oceny opartej na obiektywnych i weryfikowalnych faktach związanych z działalnością przestępczą 85

5.3.3. Systemy AI wykorzystywane do prognozowania i oceny przestępczości w odniesieniu do podmiotów prawnych 86
5.3.4. Systemy AI wykorzystywane do indywidualnego przewidywania przestępstw administracyjnych 87
5.4. Wzajemne powiązania z innymi aktami prawnymi Unii 88
6. Art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji – nieukierunkowane pozyskiwanie wizerunków twarzy 89
6.1. Uzasadnienie i cele 89
6.2. Główne pojęcia i elementy zakazu 89
6.2.1. Bazy danych służące rozpoznawaniu twarzy 90
6.2.2. Poprzez nieukierunkowane pozyskiwanie wizerunków twarzy 90
6.2.3. Z internetu i nagrań z telewizji przemysłowej 91
6.3. Systemy nieobjęte zakresem 92
6.4. Wzajemne powiązania z innymi aktami prawnymi Unii 92
7. Art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji – rozpoznawanie emocji 93
7.1. Uzasadnienie i cele 93
7.2. Główne pojęcia i elementy zakazu 94
7.2.1. Systemy AI do wyciągania wniosków na temat emocji 95
7.2.2. Ograniczenie zakazu do miejsc pracy i instytucji edukacyjnych 98
7.2.3. Wyjątki ze względów medycznych i bezpieczeństwa 100
7.3. Korzystniejsze prawo państwa członkowskiego 102
7.4. Systemy nieobjęte zakresem 102
8. Art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji: Kategoryzacja biometryczna niektórych „wrażliwych” cech 104
8.1. Uzasadnienie i cele 104
8.2. Główne pojęcia i elementy zakazu 105
8.2.1. System kategoryzacji biometrycznej 106
8.2.2. Indywidualna kategoryzacja osób w oparciu o ich dane biometryczne 107
8.2.3. Dedukowanie lub wnioskowanie informacji na temat rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej 108
8.3. Systemy nieobjęte zakresem 108
8.4. Wzajemne powiązania z innymi przepisami prawa Unii 109
9. Art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji – systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw 110
9.1. Uzasadnienie i cele 111
9.2. Główne pojęcia i elementy zakazu 112

9.2.1. Pojęcie zdalnej identyfikacji biometrycznej 113
9.2.2. W czasie rzeczywistym 116
9.2.3. W przestrzeni publicznej 116
9.2.4. Na potrzeby ścigania przestępstw 118
9.3. Wyjątki od zakazu 120
9.3.1. Uzasadnienie i cele 121
9.3.2. Ukierunkowane poszukiwanie ofiar trzech poważnych przestępstw i osób zaginionych 122
9.3.3. Zapobieganie bezpośredniemu zagrożeniu życia lub atakom terrorystycznym 123
9.3.4. Lokalizowanie i identyfikowanie osób podejrzanych o popełnienie określonych przestępstw 126
10. Zabezpieczenia i warunki dotyczące wyjątków (art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji) 129
10.1. Poszukiwana osoba a zabezpieczenia (art. 5 ust. 2 aktu w sprawie sztucznej inteligencji) 129
10.1.1. Ocena skutków dla praw podstawowych 132
10.1.2. Rejestracja zatwierdzonych systemów zdalnej identyfikacji biometrycznej 137
10.2. Konieczność uzyskania uprzedniego zezwolenia 138
10.2.1. Cel 139
10.2.2. Główna zasada: uzyskanie uprzedniego zezwolenia organu wymiaru sprawiedliwości lub niezależnego organu administracyjnego 139
10.3. Powiadamianie organów o każdym wykorzystaniu systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym” w przestrzeni publicznej do celów ścigania przestępstw 146
10.4. Konieczność istnienia przepisów krajowych w zakresie przewidzianym w akcie w sprawie sztucznej inteligencji 147
10.4.1. Zasada: podstawą prawną udzielenia zezwolenia na zastosowanie wszystkich lub niektórych wyjątków musi być prawo krajowe 147
10.4.2. Prawo krajowe musi być zgodne z ograniczeniami i warunkami określonymi w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji 148
10.4.3. Szczegółowe przepisy krajowe dotyczące wniosku o udzielenie zezwolenia oraz wydawania i wykonywania tego zezwolenia 149
10.4.4. Szczegółowe przepisy krajowe dotyczące nadzoru i sprawozdawczości w odniesieniu do zezwolenia 150
10.5. Roczne sprawozdania krajowych organów nadzoru rynku i krajowych organów ochrony danych państw członkowskich 151
10.6. Sprawozdania roczne Komisji 151
10.7. Systemy nieobjęte zakresem 152
10.8. Przykłady wykorzystania 153
11. Rozpoczęcie stosowania 156

12. Przegląd i aktualizacja wytycznych Komisji 156

1. INFORMACJE OGÓLNE I CELE

1) Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany niektórych rozporządzeń i dyrektyw („akt w sprawie sztucznej inteligencji”)[1] weszło w życie w dniu 1 sierpnia 2024 r. W akcie w sprawie sztucznej inteligencji ustanowiono zharmonizowane przepisy dotyczące wprowadzania do obrotu, oddawania do użytku i wykorzystywania sztucznej inteligencji („AI”) w Unii[2]. Jego celem jest promowanie innowacji w dziedzinie AI i upowszechniania tej technologii, przy jednoczesnym zapewnieniu wysokiego poziomu ochrony zdrowia, bezpieczeństwa i praw podstawowych w Unii, w tym demokracji i praworządności.

2) W akcie w sprawie sztucznej inteligencji zastosowano podejście oparte na analizie ryzyka i podzielono systemy AI według czterech różnych kategorii ryzyka:

   (i) Niedopuszczalne ryzyko: systemy AI stwarzające niedopuszczalne ryzyko dla praw podstawowych i wartości Unii są zakazane na mocy art. 5 aktu w sprawie sztucznej inteligencji.

   (ii) Wysokie ryzyko: systemy AI stwarzające wysokie ryzyko dla zdrowia, bezpieczeństwa i praw podstawowych podlegają szeregowi wymogów i obowiązków. Zgodnie z art. 6 aktu w sprawie sztucznej inteligencji w związku z załącznikami I i III do wspomnianego aktu uznaje się je za systemy „wysokiego ryzyka”.

   (iii) Ryzyko w zakresie przejrzystości: systemy AI stwarzające ograniczone ryzyko w zakresie przejrzystości podlegają obowiązkom w zakresie przejrzystości określonym w art. 50 aktu w sprawie sztucznej inteligencji.

   (iv) Minimalne lub zerowe ryzyko: systemy AI stwarzające minimalne lub zerowe ryzyko nie podlegają regulacji, ale ich dostawcy i podmioty je stosujące mogą przestrzegać dobrowolnych kodeksów postępowania[3].

3) Zgodnie z art. 96 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji Komisja ma przyjąć wytyczne dotyczące praktycznego wdrażania praktyk zakazanych na mocy art. 5 aktu. Zakazy te zaczęły mieć zastosowanie sześć miesięcy po wejściu w życie aktu w sprawie sztucznej inteligencji, tj. od dnia 2 lutego 2025 r.

4) Niniejsze wytyczne mają na celu zwiększenie jasności prawa i zapewnienie informacji na temat tego, jak Komisja interpretuje zakazy zawarte w art. 5 aktu w sprawie sztucznej inteligencji, z myślą o zapewnieniu ich spójnego, skutecznego i jednolitego stosowania. Wytyczne te powinny pełnić rolę praktycznych wskazówek pomagających właściwym organom określonym w akcie w sprawie sztucznej inteligencji w działaniach w zakresie egzekwowania przepisów, a także dostawcom systemów AI

[1] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji (akt w sprawie sztucznej inteligencji) (Dz.U. L, 2024/1689, 12.7.2024).

[2] Art. 1 aktu w sprawie sztucznej inteligencji.

[3] Art. 95 aktu w sprawie sztucznej inteligencji.

i podmiotom je stosującym w wypełnianiu obowiązków wynikających z tego aktu. Zgodnie z założeniem interpretacja zakazów przedstawiona w wytycznych ma być proporcjonalna, co umożliwi osiągnięcie celów aktu w sprawie sztucznej inteligencji w zakresie ochrony praw podstawowych i bezpieczeństwa, zarazem promując innowacje i zapewniając pewność prawa.

5) Niniejsze wytyczne nie mają charakteru wiążącego. Wszelkiej autorytatywnej wykładni aktu w sprawie sztucznej inteligencji może ostatecznie dokonywać wyłącznie Trybunał Sprawiedliwości Unii Europejskiej („TSUE").

6) Przy opracowywaniu niniejszych wytycznych wykorzystano informacje pochodzące od różnych zainteresowanych stron, m.in. dostawców systemów AI i podmiotów je stosujących, organizacji społeczeństwa obywatelskiego, środowisk akademickich, organów publicznych, stowarzyszeń przedsiębiorców itp., pozyskane podczas szeroko zakrojonych konsultacji zorganizowanych przez Komisję. Przeprowadzono również konsultacje z państwami członkowskimi w ramach Europejskiej Rady ds. Sztucznej Inteligencji oraz z Parlamentem Europejskim. Niniejsze wytyczne będą regularnie poddawane przeglądowi w świetle doświadczeń zdobytych w trakcie praktycznego wdrażania art. 5 aktu w sprawie sztucznej inteligencji, a także zmian technologicznych i rynkowych.

7) Stosowanie art. 5 aktu w sprawie sztucznej inteligencji będzie wymagać indywidualnej oceny, z należytym uwzględnieniem konkretnej sytuacji w danym przypadku. W związku z tym przykłady podane w niniejszych wytycznych mają jedynie charakter orientacyjny i pozostają bez uszczerbku dla konieczności przeprowadzenia takiej oceny w każdym przypadku.

2. PRZEGLĄD ZAKAZANYCH PRAKTYK W ZAKRESIE AI

8) W art. 5 aktu w sprawie sztucznej inteligencji zakazano wprowadzania do obrotu w UE, oddawania do użytku lub wykorzystywania niektórych systemów AI do celów manipulacji, wykorzystywania słabości, kontroli społecznej lub nadzoru, które to praktyki ze względu na sam swój charakter naruszają prawa podstawowe i wartości Unii. W motywie 28 aktu w sprawie sztucznej inteligencji wyjaśniono, że takie praktyki są szczególnie szkodliwe i stanowią nadużycie i powinny być zakazane, ponieważ są sprzeczne z unijnymi wartościami dotyczącymi poszanowania godności ludzkiej, wolności, równości, demokracji i praworządności oraz z prawami podstawowymi zapisanymi w Karcie praw podstawowych Unii Europejskiej („Karta"), w tym z prawem do niedyskryminacji (art. 21 Karty) i do równości (art. 20), do ochrony danych (art. 8 Karty) oraz do poszanowania życia prywatnego i rodzinnego (art. 7 Karty), a także z prawami dziecka (art. 24 Karty). Zakazy zawarte w art. 5 aktu w sprawie sztucznej inteligencji mają również na celu ochronę prawa do wolności wypowiedzi i informacji (art. 11 Karty), wolności zgromadzania się i stowarzyszania się (art. 12 Karty), wolności myśli, sumienia i religii (art. 10 Karty), prawa do skutecznego środka prawnego i dostępu do bezstronnego sądu (art. 47 Karty) oraz domniemania niewinności i prawa do obrony (art. 48 Karty).

2.1. Zakazy wymienione w art. 5 aktu w sprawie sztucznej inteligencji

9) Przegląd zakazów

Przepis	**Zakaz**	**Zakres**
Art. 5 ust. 1 lit. a)	Szkodliwa manipulacja i wprowadzanie w błąd	Systemy AI, które stosują techniki podprogowe będące poza świadomością danej osoby bądź celowe techniki manipulacyjne lub wprowadzające w błąd, czego celem lub skutkiem jest dokonanie zmiany zachowania w sposób, które wyrządzają lub mogą z uzasadnionym prawdopodobieństwem wyrządzić poważną szkodę
Art. 5 ust. 1 lit. b)	Wykorzystywanie słabości w sposób wyrządzający szkodę	Systemy AI, które wykorzystują słabości ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną, czego celem lub skutkiem jest dokonanie zmiany zachowania w sposób, który wyrządza lub może z uzasadnionym prawdopodobieństwem wyrządzić poważną szkodę
Art. 5 ust. 1 lit. c)	Scoring społeczny	Systemy AI, które oceniają lub klasyfikują osoby fizyczne lub grupy osób na podstawie zachowania społecznego bądź cech osobistych lub cech osobowości, kiedy to scoring społeczny prowadzi do krzywdzącego lub niekorzystnego traktowania, w przypadku gdy dane pochodzą z niepowiązanych kontekstów społecznych lub jeżeli takie traktowanie jest nieuzasadnione lub nieproporcjonalne do zachowania społecznego
Art. 5 ust. 1 lit. d)	Ocena i przewidywanie ryzyka popełnienia przestępstwa przez osobę fizyczną	Systemy AI, które oceniają lub przewidują ryzyko popełnienia przestępstwa przez osobę fizyczną wyłącznie na podstawie profilowania osoby fizycznej lub na podstawie jej cech osobowości i cech charakterystycznych; nie dotyczy to wspierania dokonywanej przez człowieka oceny opartej na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą
Art. 5 ust. 1 lit. e)	Nieukierunkowane pozyskiwanie wizerunków twarzy w celu tworzenia baz danych służących rozpoznawaniu twarzy	Systemy AI, które tworzą lub rozbudowują bazy danych służące rozpoznawaniu twarzy poprzez nieukierunkowane pozyskiwanie (ang. untargeted scraping) wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej (CCTV)
Art. 5 ust. 1 lit. f)	Rozpoznawanie emocji	Systemy AI używane do wyciągania wniosków na temat emocji w miejscu pracy lub w instytucjach edukacyjnych; nie dotyczy to systemów używanych ze względów medycznych lub bezpieczeństwa

Art. 5 ust. 1 lit. g)	Kategoryzacja biometryczna	Systemy AI, które kategoryzują osoby fizyczne w oparciu o ich dane biometryczne, by wydedukować lub wywnioskować informacje na temat ich rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej; nie dotyczy to etykietowania ani filtrowania pozyskanych zgodnie z prawem zbiorów danych biometrycznych, w tym w obszarze ścigania przestępstw
Art. 5 ust. 1 lit. h)	Zdalna identyfikacja biometryczna w czasie rzeczywistym	Systemy AI do zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw; nie dotyczy to sytuacji, gdy jest to konieczne do ukierunkowanego poszukiwania konkretnych ofiar, zapobiegnięcia konkretnym zagrożeniom, w tym atakom terrorystycznym, lub poszukiwania osób podejrzanych o popełnienie określonych przestępstw (dalsze wymogi proceduralne, w tym dotyczące zezwolenia, przedstawiono w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji)

2.2. Podstawa prawna zakazów

10) Akt w sprawie sztucznej inteligencji opiera się na dwóch podstawach prawnych: art. 114 Traktatu o funkcjonowaniu Unii Europejskiej („TFUE") (podstawa prawna dotycząca rynku wewnętrznego) i art. 16 TFUE (podstawa prawna dotycząca ochrony danych). Art. 16 TFUE stanowi podstawę prawną dla szczegółowych przepisów dotyczących przetwarzania danych osobowych w związku z zakazem wykorzystywania systemów zdalnej identyfikacji biometrycznej do celów ścigania przestępstw, systemów kategoryzacji biometrycznej do celów ścigania przestępstw oraz indywidualnych ocen ryzyka do celów ścigania przestępstw[4]. Podstawę prawną wszystkich pozostałych zakazów wymienionych w art. 5 aktu w sprawie sztucznej inteligencji stanowi art. 114 TFUE.

2.3. Zakres przedmiotowy: praktyki związane z „wprowadzaniem do obrotu", „oddawaniem do użytku" lub „wykorzystywaniem" systemu AI

11) Praktyki zakazane na mocy art. 5 aktu w sprawie sztucznej inteligencji odnoszą się do wprowadzania do obrotu, oddawania do użytku lub wykorzystywania określonych

[4] Motyw 3 aktu w sprawie sztucznej inteligencji. W odniesieniu do zakazów opartych na art. 16 TFUE zastosowanie mają dwie klauzule opt-out dotyczące Irlandii i Danii. Z uwagi na swobodę uznania przyznaną Irlandii na mocy Protokołu nr 21 w sprawie stanowiska Zjednoczonego Królestwa i Irlandii w odniesieniu do przestrzeni wolności, bezpieczeństwa i sprawiedliwości, załączonego do TUE i do TFUE, Irlandia może podjąć decyzję o niestosowaniu przepisów dotyczących zakazu korzystania z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej w celu ścigania przestępstw, a także przepisów proceduralnych związanych ze wspomnianym artykułem (art. 5 ust. 2–6 aktu w sprawie sztucznej inteligencji) (zob. motyw 40). Na mocy Protokołu nr 22 załączonego do TUE i TFUE Dania korzysta z możliwości niestosowania niektórych przyjętych środków i może podjąć decyzję o niestosowaniu w pełni zakazów opartych na art. 16 TFUE (zob. motyw 41).

systemów AI[5]. W odniesieniu do systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, zakaz określony w art. 5 ust. 1 lit. h) aktu ma zastosowanie wyłącznie do ich wykorzystywania. W art. 3 pkt 1 aktu zdefiniowano, co stanowi „system AI". W wytycznych dotyczących definicji systemu AI przedstawiono interpretację tej definicji dokonaną przez Komisję.

12) Zgodnie z art. 3 pkt 9 aktu w sprawie sztucznej inteligencji, **wprowadzenie systemu AI do obrotu** oznacza „udostępnienie po raz pierwszy systemu AI [...] na rynku Unii". „Udostępnienie" definiuje się jako dostarczenie systemu „w celu jego dystrybucji lub wykorzystania na rynku Unii w ramach działalności handlowej, odpłatnie lub nieodpłatnie"[6]. Udostępnienie systemu AI jest niezależne od sposobu jego dostarczenia, takiego jak udzielenie dostępu do systemu i zapewnianej przez niego usługi za pośrednictwem interfejsu programowania aplikacji („API"), chmury, przez bezpośrednie pobieranie lub w wersji fizycznej.

Na przykład system zdalnej identyfikacji biometrycznej opracowany poza Unią przez dostawcę z państwa trzeciego jest wprowadzany do obrotu w Unii po raz pierwszy, gdy jest oferowany odpłatnie lub nieodpłatnie w co najmniej jednym państwie członkowskim. Takie wprowadzenie do obrotu może nastąpić poprzez zapewnienie dostępu do systemu online za pośrednictwem API lub innego interfejsu użytkownika.

13) W art. 3 pkt 11 aktu w sprawie sztucznej inteligencji **oddanie do użytku** zdefiniowano jako „dostarczenie systemu AI do pierwszego użycia bezpośrednio podmiotowi stosującemu lub do użytku własnego w Unii, zgodnie z jego przeznaczeniem", co obejmuje zarówno dostarczenie osobom trzecim do pierwszego użycia, jak i opracowanie i wdrożenie na własny użytek. Przeznaczenie systemu oznacza „zastosowanie, do jakiego system AI został przeznaczony przez jego dostawcę, w tym konkretny kontekst i warunki wykorzystywania, określone w informacjach dostarczonych przez dostawcę w instrukcji obsługi, materiałach promocyjnych lub sprzedażowych i oświadczeniach, jak również w dokumentacji technicznej"[7].

Na przykład dostawca tworzy system zdalnej identyfikacji biometrycznej poza Unią i dostarcza ten system organowi ścigania lub prywatnemu przedsiębiorstwu mającemu siedzibę w państwie członkowskim do pierwszego użycia, tym samym oddając go do użytku. Na przykład organ publiczny opracowuje na własne potrzeby system scoringu i wdraża go w celu przewidywania ryzyka oszustw ze strony beneficjentów dodatków na gospodarstwo domowe, tym samym oddając go do użytku.

14) Chociaż w akcie w sprawie sztucznej inteligencji nie zdefiniowano wyraźnie „**wykorzystywania**" systemu AI, należy je rozumieć szeroko, jako obejmujące wykorzystywanie lub stosowanie systemu w dowolnym momencie jego cyklu życia po wprowadzeniu go do obrotu lub oddaniu go do użytku. Może ono również obejmować

[5] Aby zapoznać się z definicjami tych terminów, zob. również zawiadomienie Komisji – Niebieski przewodnik – wdrażanie unijnych przepisów dotyczących produktów 2022, 2022/C 247/01, sekcja 2.

[6] Art. 3 pkt 10 aktu w sprawie sztucznej inteligencji.

[7] Art. 3 pkt 12 aktu w sprawie sztucznej inteligencji.

integrację systemu AI z usługami i procesami osób korzystających z takiego systemu, w tym w ramach bardziej złożonych systemów, procesów lub infrastruktury. Chociaż dostawcy systemów AI muszą brać pod uwagę warunki wykorzystywania, które można racjonalnie przewidzieć przed wprowadzeniem ich systemów AI do obrotu (zamierzone wykorzystanie i dające się racjonalnie przewidzieć niewłaściwe wykorzystanie[8]), podmioty stosujące nadal mają obowiązek uwzględnienia zgodnych z prawem warunków wykorzystywania systemu[9]. Do celów art. 5 aktu w sprawie sztucznej inteligencji odniesienie do „wykorzystywania" należy rozumieć jako obejmujące każde niewłaściwe wykorzystanie systemu AI („dające się racjonalnie przewidzieć" lub nie), które może stanowić zakazaną praktykę[10].

Na przykład system AI wykorzystywany przez pracodawcę do wyciągania wniosków na temat emocji w miejscu pracy jest zakazany, z wyjątkiem przypadków, gdy jego użycie wynika ze względów medycznych lub bezpieczeństwa (art. 5 ust. 1 lit. f)) aktu w sprawie sztucznej inteligencji). Zakaz ten ma zastosowanie do podmiotów stosujących niezależnie od tego, czy dostawca (dostawca systemu) wykluczył takie wykorzystanie w umowie z podmiotem stosującym (pracodawcą), tj. w warunkach użytkowania.

2.4. Zakres podmiotowy: podmioty odpowiedzialne

15) W akcie w sprawie sztucznej inteligencji w odniesieniu do systemów AI wyróżnia się różne kategorie operatorów: dostawców, podmioty stosujące, importerów, dystrybutorów i producentów produktu. Ze względu na zakres zakazanych praktyk określonych w art. 5 aktu w sprawie sztucznej inteligencji, niniejsze wytyczne dotyczą wyłącznie dostawców i podmiotów stosujących.

16) Zgodnie z art. 3 pkt 3 aktu w sprawie sztucznej inteligencji **dostawcy** to osoby fizyczne lub prawne, organy publiczne, agencje lub inne podmioty, które rozwijają systemy AI lub zlecają rozwój systemów AI oraz które pod własną nazwą lub własnym znakiem towarowym wprowadzają je do obrotu lub oddają je do użytku[11] (zob. sekcja 2.3 powyżej). Dostawcy, którzy mają siedzibę lub znajdują się poza Unią podlegają przepisom aktu w sprawie sztucznej inteligencji, jeżeli wprowadzają te systemy do obrotu lub oddają je do użytku w Unii[12] lub jeżeli wyniki wytworzone przez system AI są wykorzystywane w Unii[13]. Dostawcy muszą zapewnić, aby przed wprowadzeniem

[8] Zob. art. 3 pkt 12 i 13 aktu w sprawie sztucznej inteligencji.

[9] Aby zapoznać się z definicjami tych terminów, zob. również zawiadomienie Komisji – Niebieski przewodnik – wdrażanie unijnych przepisów dotyczących produktów 2022, 2022/C 247/01, sekcja 2.8.

[10] Motyw 28 aktu w sprawie sztucznej inteligencji.

[11] Zob. art. 3 pkt 3, 9 i 11 aktu w sprawie sztucznej inteligencji. W odniesieniu do systemów AI wysokiego ryzyka art. 25 ust. 1 aktu w sprawie sztucznej inteligencji stanowi, co następuje: „Do celów niniejszego rozporządzenia za dostawcę systemu AI wysokiego ryzyka uznaje się i obejmuje obowiązkami dostawcy ustanowionymi w art. 16 każdego dystrybutora, importera, podmiot stosujący lub inną stronę trzecią, jeżeli zachodzi którakolwiek z następujących okoliczności: a) umieszczają oni swoją nazwę lub znak towarowy w systemie AI wysokiego ryzyka, który został już wprowadzony do obrotu lub oddany do użytku, bez uszczerbku dla ustaleń umownych przewidujących, że podział obowiązków następuje w inny sposób; b) we wprowadzonym już do obrotu lub oddanym do użytku systemie AI wysokiego ryzyka dokonują oni istotnej zmiany w taki sposób, że pozostaje on systemem AI wysokiego ryzyka zgodnie z art. 6; c) zmieniają oni przeznaczenie systemu AI, w tym systemu AI ogólnego przeznaczenia, który nie został zaklasyfikowany jako system AI wysokiego ryzyka i który został już wprowadzony do obrotu lub oddany do użytku, w taki sposób, że dany system AI staje się systemem AI wysokiego ryzyka zgodnie z art. 6".

[12] Art. 2 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji.

[13] Art. 2 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

do obrotu lub oddaniem do użytku ich systemy AI spełniały wszystkie odpowiednie wymogi.

Na przykład dostawcą systemu zdalnej identyfikacji biometrycznej jest jego producent, który wprowadza taki system do obrotu w Unii pod swoim znakiem towarowym. Dostawcą takiego systemu może być również organ publiczny, który opracowuje system we własnym zakresie i oddaje go do użytku do celów własnych.

17) **Podmioty stosujące** to osoby fizyczne lub prawne, organy publiczne, agencje lub inne podmioty wykorzystujące systemy AI, nad którymi sprawują kontrolę, chyba że systemy AI są wykorzystywane w ramach osobistej działalności pozazawodowej[14]. „Kontrolę" nad systemem AI należy rozumieć jako przyjęcie odpowiedzialności za decyzję o wdrożeniu systemu oraz za sposób jego faktycznego wykorzystania. Podmioty stosujące są objęte zakresem stosowania aktu w sprawie sztucznej inteligencji, jeżeli mają siedzibę lub znajdują się w Unii[15] bądź – w przypadku gdy podmioty te znajdują się w państwie trzecim – jeżeli wyniki wytworzone przez system AI są wykorzystywane w Unii[16].

18) W przypadku gdy podmiot stosujący system AI jest osobą prawną, która sprawuje kontrolę nad wykorzystywaniem systemu, tj. organem ścigania lub prywatną agencją ochrony, za podmiot stosujący nie należy uznawać poszczególnych pracowników, którzy działają w ramach procedur i pod kontrolą takiego podmiotu. Osoba prawna pozostaje również podmiotem stosującym, jeżeli angażuje osoby trzecie (np. wykonawców, personel zewnętrzny) w obsługę systemu w jej imieniu, na jej odpowiedzialność i pod jej kontrolą.

19) Operatorzy mogą odgrywać więcej niż jedną rolę w odniesieniu do systemu AI. Na przykład jeżeli operator rozwija własny system AI, który następnie wykorzystuje, zostanie uznany zarówno za dostawcę tego systemu, jak i za podmiot go stosujący, nawet jeżeli dany system jest również wykorzystywany przez inne podmioty stosujące, którym system udostępniono odpłatnie lub nieodpłatnie.

20) Na wszystkich etapach cyklu życia systemu AI wymagana jest ciągła zgodność z aktem w sprawie sztucznej inteligencji. Wymaga to stałego monitorowania i aktualizowania systemów AI wprowadzonych do obrotu lub oddanych do użytku w Unii w celu zapewnienia, aby system AI spełniał wymogi aktu w sprawie sztucznej inteligencji przez cały cykl swojego życia i nie powodował wystąpienia praktyki zakazanej na mocy art. 5 aktu w sprawie sztucznej inteligencji. W zależności od odgrywanej roli i kontroli nad projektowaniem, rozwojem i faktycznym wykorzystaniem systemu dostawcy systemów AI i podmioty je stosujące mają różne obowiązki, których celem jest uniknięcie zakazanej praktyki. W odniesieniu do każdego z zakazów takie role i obowiązki należy interpretować w sposób proporcjonalny, z uwzględnieniem tego, który podmiot w łańcuchu wartości jest najlepiej przygotowany do zastosowania konkretnych środków zapobiegawczych i środków ograniczających ryzyko oraz

[14] Art. 3 pkt 4 aktu w sprawie sztucznej inteligencji.
[15] Art. 2 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.
[16] Art. 2 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

zapewnienia zgodnego z przepisami rozwoju i wykorzystania systemów AI uwzględniającego cele i podejście przyjęte w akcie w sprawie sztucznej inteligencji.

2.5. Wyłączenia z zakresu stosowania aktu w sprawie sztucznej inteligencji

21) Art. 2 aktu w sprawie sztucznej inteligencji przewiduje szereg ogólnych wyłączeń z zakresu jego stosowania, które są istotne dla pełnego zrozumienia praktycznego zastosowania zakazów wymienionych w art. 5 aktu.

2.5.1. Cele bezpieczeństwa narodowego, obronne i wojskowe

22) Zgodnie z art. 2 ust. 3 aktu w sprawie sztucznej inteligencji nie stosuje się go do obszarów wykraczających poza zakres stosowania prawa Unii i w żadnym wypadku nie powinien on wpływać na kompetencje państw członkowskich w zakresie bezpieczeństwa narodowego, niezależnie od rodzaju podmiotu, któremu państwa członkowskie powierzyły wykonywanie zadań związanych z tymi kompetencjami. W akcie w sprawie sztucznej inteligencji w sposób wyraźny wyłącza się z zakresu jego zastosowania systemy AI, które „wprowadzono do obrotu, oddano do użytku lub są [...] wykorzystywane, ze zmianami lub bez zmian, wyłącznie do celów wojskowych, obronnych lub do celów bezpieczeństwa narodowego, niezależnie od rodzaju podmiotu prowadzącego te działania". Zastosowanie tego wyłączenia zależy zatem od celów lub zastosowań systemu AI, a nie od podmiotów prowadzących działania z wykorzystaniem takiego systemu – mogą być nimi również operatorzy prywatni, którym państwa członkowskie powierzyły wykonywanie zadań związanych z tymi kompetencjami.

23) Według TSUE termin „**bezpieczeństwo narodowe**" oznacza „pierwszorzędny interes w ochronie podstawowych funkcji państwa i podstawowych interesów społeczeństwa, a także zapobieganie i ściganie działalności mogącej poważnie zdestabilizować podstawowe struktury konstytucyjne, polityczne lub społeczne kraju, w szczególności bezpośrednio zagrozić społeczeństwu, ludności lub państwu jako takiemu, zwłaszcza takiej jak działalność terrorystyczna"[17]. Bezpieczeństwo narodowe nie obejmuje na przykład działań związanych z bezpieczeństwem drogowym[18] czy organizacją lub sprawowaniem wymiaru sprawiedliwości[19]. Jak stwierdził TSUE, „to do państw członkowskich należy określenie ich podstawowych interesów bezpieczeństwa i podjęcie środków zmierzających do zapewnienia ich bezpieczeństwa zewnętrznego i wewnętrznego, [...] środek krajowy [...] podjęty w celu ochrony bezpieczeństwa narodowego [...] nie może powodować niemożności stosowania prawa Unii i zwolnienia państw członkowskich z konieczności przestrzegania tego prawa"[20].

24) Aby wyłączenie przewidziane w art. 2 ust. 3 akapit drugi aktu w sprawie sztucznej inteligencji miało zastosowanie, system AI musi zostać wprowadzony do obrotu, oddany do użytku lub być wykorzystywany wyłącznie do celów wojskowych,

---

17 Wyrok Trybunału Sprawiedliwości z dnia 6 października 2020 r., *La Quadrature du Net i in.*, C-511/18, C-512/18 i C-520/18, EU:C:2020:791, pkt 135; Wyrok Trybunału Sprawiedliwości z dnia 5 czerwca 2023 r., *Komisja/Polska*, C-204/21, EU:C:2023:442, pkt 318, odnoszący się do sprawy C-439/19, pkt 67 i sprawy C-306/21, pkt 40.

18 Wyrok Trybunału Sprawiedliwości z dnia 22 czerwca 2021 r., *Latvijas Republikas Saeima*, C-439/19, EU:C:2021:504, pkt 68.

19 Wyrok Trybunału Sprawiedliwości z dnia 5 czerwca 2023 r., *Komisja/Polska*, C-204/21, EU:C:2023:442, pkt 319.

20 Wyrok Trybunału Sprawiedliwości z dnia 6 października 2020 r., *Privacy International*, C-623/17, EU:C:2020:790, pkt 44.

obronnych lub do celów bezpieczeństwa narodowego. W motywie 24 aktu w sprawie sztucznej inteligencji doprecyzowano, w jaki sposób należy interpretować pojęcie „**wyłącznie**" i kiedy system AI wykorzystywany do takich celów może być jednak objęty zakresem stosowania tego aktu.

> Na przykład jeżeli system AI wprowadzony do obrotu, oddany do użytku lub wykorzystywany do celów wojskowych, obronnych lub celów bezpieczeństwa narodowego jest (tymczasowo lub na stałe) wykorzystywany do innych celów, takich jak cele cywilne lub humanitarne, ściganie przestępstw lub cele bezpieczeństwa publicznego, system taki będzie objęty zakresem stosowania aktu w sprawie sztucznej inteligencji. W takim przypadku podmiot wykorzystujący system AI do takich innych celów powinien zapewnić zgodność systemu z aktem w sprawie sztucznej inteligencji, chyba że system ten jest już zgodny z aktem, co należy zweryfikować przed takim wykorzystaniem.

25) Ponadto w motywie 24 aktu w sprawie sztucznej inteligencji wyjaśniono, że systemy AI wprowadzane do obrotu lub oddawane do użytku do celu stanowiącego podstawę wyłączenia, tzn. celu wojskowego, obronnego lub celu bezpieczeństwa narodowego, oraz do jednego lub kilku celów niestanowiących podstawy wyłączenia, takich jak cele cywilne lub ściganie przestępstw (tzw. systemy „**podwójnego zastosowania**"), są objęte zakresem stosowania aktu. Dostawcy tych systemów powinni zapewnić, aby spełniały one wymogi aktu w sprawie sztucznej inteligencji.

> Na przykład jeżeli przedsiębiorstwo oferuje system zdalnej identyfikacji biometrycznej do różnych celów, w tym na potrzeby ścigania przestępstw i do celów bezpieczeństwa narodowego, przedsiębiorstwo to jest dostawcą wspomnianego systemu „podwójnego zastosowania" i musi zapewnić jego zgodność z wymogami określonymi w akcie w sprawie sztucznej inteligencji.

26) Fakt, że system AI może wchodzić w zakres stosowania aktu w sprawie sztucznej inteligencji, nie powinien jednak mieć wpływu na możliwość wykorzystywania – przez podmioty prowadzące działania dotyczące bezpieczeństwa narodowego, działania obronne i wojskowe, bez względu na rodzaj podmiotu prowadzącego te działania – tego systemu do celów bezpieczeństwa narodowego, celów wojskowych i obronnych[21].

> Na przykład gdyby krajowa agencja wywiadowcza powierzyła agencji bezpieczeństwa narodowego lub operatorowi prywatnemu zadanie wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów bezpieczeństwa narodowego (takich jak gromadzenie danych wywiadowczych), takie wykorzystanie byłoby wyłączone z zakresu zastosowania aktu w sprawie sztucznej inteligencji.

27) Wyraźne określenie wyłączenia dotyczącego bezpieczeństwa narodowego jest szczególnie ważne w przypadku, gdy systemy AI są wprowadzane do obrotu, oddawane do użytku lub wykorzystywane do celów ścigania przestępstw, które to cele wchodzą w zakres stosowania aktu w sprawie sztucznej inteligencji. Ma to znaczenie

[21] Motyw 24 aktu w sprawie sztucznej inteligencji.

w przypadku zakazu dotyczącego przewidywania i oceniania ryzyka popełnienia przestępstwa przez osobę fizyczną oraz zakazu dotyczącego wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw, które to zakazy określono, odpowiednio, w art. 5 ust. 1 lit. d) i h) aktu w sprawie sztucznej inteligencji. Zadaniem policji i innych organów ścigania jest zapobieganie przestępczości, wykrywanie i ściganie czynów zabronionych, prowadzenie postępowań przygotowawczych i wykonywanie kar, w tym ochrona przed zagrożeniami dla bezpieczeństwa publicznego i zapobieganie takim zagrożeniom[22]. W każdym przypadku wykorzystywania systemów AI do takich celów systemy te są objęte zakresem stosowania aktu w sprawie sztucznej inteligencji.

28) Działania Europolu i innych unijnych agencji bezpieczeństwa, takich jak Frontex, wchodzą w zakres stosowania aktu w sprawie sztucznej inteligencji.

2.5.2. Współpraca sądowa i współpraca organów ścigania z państwami trzecimi

29) Zgodnie z art. 2 ust. 4 aktu w sprawie sztucznej inteligencji, aktu tego nie stosuje się do organów publicznych w państwie trzecim ani do organizacji międzynarodowych, jeżeli te organy lub organizacje wykorzystują systemy AI w ramach współpracy międzynarodowej lub umów międzynarodowych w sprawie ścigania przestępstw i współpracy sądowej zawartych z Unią lub z jednym państwem członkowskim bądź ich większą liczbą, pod warunkiem zapewnienia przez to państwo trzecie lub organizację międzynarodową odpowiednich zabezpieczeń w odniesieniu do ochrony podstawowych praw i wolności osób fizycznych. W stosownych przypadkach wyłączenie to może obejmować działania podmiotów prywatnych, którym dane państwo trzecie powierzyło wykonywanie szczególnych zadań w ramach wsparcia ścigania przestępstw i współpracy wymiarów sprawiedliwości[23]. Jednocześnie, aby wyłączenie miało zastosowanie, te ramy współpracy lub umowy międzynarodowe muszą zawierać odpowiednie zabezpieczenia w odniesieniu do ochrony podstawowych praw i wolności osób fizycznych, podlegające ocenie ze strony organów nadzoru rynku właściwych do celów nadzoru nad systemami AI wykorzystywanymi w obszarze ścigania przestępstw i w kontekście wymiaru sprawiedliwości[24]. W motywie 22 aktu w sprawie sztucznej inteligencji doprecyzowano, że będące odbiorcami organy krajowe oraz instytucje, organy i jednostki organizacyjne Unii korzystające w Unii z takich wyników wytworzonych przez systemy AI pozostają odpowiedzialne za zapewnienie zgodności ich stosowania z prawem Unii. W przypadku zmiany takich umów międzynarodowych lub zawarcia nowych w przyszłości umawiające się strony powinny dołożyć wszelkich starań, by dostosować takie umowy do wymogów aktu w sprawie sztucznej inteligencji.

2.5.3. Badania i rozwój

[22] Art. 3 pkt 46 aktu w sprawie sztucznej inteligencji.
[23] Zob. motyw 22 aktu w sprawie sztucznej inteligencji.
[24] Zob. motyw 22 i art. 74 ust. 8 aktu w sprawie sztucznej inteligencji.

30) Zgodnie z art. 2 ust. 8 aktu w sprawie sztucznej inteligencji, aktu tego nie stosuje się „do żadnej działalności badawczej, testowej ani rozwojowej dotyczącej systemów AI lub modeli AI przed wprowadzeniem ich do obrotu lub oddaniem ich do użytku". Wyłączenie to jest zgodne z logiką rynkową aktu, który ma zastosowanie do systemów AI po ich wprowadzeniu do obrotu lub oddaniu do użytku.

> Na przykład na etapie badań i rozwoju twórcy AI mają swobodę eksperymentowania i testowania nowych funkcjonalności mogących obejmować techniki, które można uznać za manipulacyjne i które są objęte art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, jeżeli są wykorzystywane w zastosowaniach ukierunkowanych na konsumentów. Akt w sprawie sztucznej inteligencji zezwala na takie eksperymenty, co wynika z uznania, że badania i rozwój na wczesnym etapie mają zasadnicze znaczenie dla doskonalenia technologii AI i zapewnienia, aby spełniały one normy bezpieczeństwa i normy etyczne przed wprowadzeniem ich do obrotu.

31) Jak wyjaśniono w motywie 25 aktu w sprawie sztucznej inteligencji, akt ten ma na celu wspieranie innowacji i uznaje się w nim znaczenie badań naukowych dla rozwoju technologii AI oraz postępu naukowego i innowacji. Art. 2 ust. 6 aktu w sprawie sztucznej inteligencji przewiduje zatem wyłączenie dotyczące „systemów AI lub modeli AI, w tym ich wyników, rozwiniętych i oddanych do użytku wyłącznie do celów badań naukowych i rozwojowych".

> Na przykład badania nad reakcjami poznawczymi i behawioralnymi na wywoływane przez sztuczną inteligencję bodźce podprogowe lub wprowadzające w błąd mogą dostarczyć cennych informacji na temat interakcji między człowiekiem a sztuczną inteligencją, przyczyniając się do bezpieczniejszych i skuteczniejszych zastosowań tej technologii w przyszłości. Takie badania są dozwolone, ponieważ są wyłączone z zakresu stosowania aktu w sprawie sztucznej inteligencji, niezależnie od zakazu określonego w art. 5 ust. 1 lit. a) tego aktu.

32) Wyłączenie przewidziane w art. 2 ust. 8 aktu w sprawie sztucznej inteligencji pozostaje jednak bez uszczerbku dla obowiązku przestrzegania tego aktu, w przypadku gdy system AI jest wprowadzany do obrotu lub oddawany do użytku w wyniku takiej działalności badawczo-rozwojowej[25]. Wyłączenie to nie obejmuje również testów w warunkach rzeczywistych w rozumieniu aktu w sprawie sztucznej inteligencji[26].

> Na przykład gmina, która chce przetestować na ulicach podczas karnawału oprogramowanie do rozpoznawania twarzy z wykorzystaniem systemu zdalnej identyfikacji biometrycznej, rekrutuje wolontariuszy, którzy mają zostać zidentyfikowani przez system w warunkach rzeczywistych. Ponieważ testy w warunkach rzeczywistych nie wchodzą w zakres wyłączenia przewidzianego w art. 2

[25] Motyw 25 aktu w sprawie sztucznej inteligencji.

[26] Zgodnie z art. 3 pkt 57 aktu w sprawie sztucznej inteligencji „testy w warunkach rzeczywistych" oznaczają ograniczone w czasie testy systemu AI dotyczące jego przeznaczenia prowadzone w warunkach rzeczywistych – poza środowiskiem laboratoryjnym lub środowiskiem symulowanym innego typu – w celu zgromadzenia wiarygodnych i solidnych danych oraz w celu oceny i weryfikacji spełnienia przez system AI wymogów tego aktu. W akcie w sprawie sztucznej inteligencji przewidziano specjalny system takich testów w warunkach rzeczywistych, które nie kwalifikują się jako wprowadzenie systemu AI do obrotu ani oddanie go do użytku w rozumieniu tego aktu, o ile spełniono wszystkie warunki określone w art. 57 lub 60, w tym uzyskano dobrowolną i świadomą zgodę osób uczestniczących w testach itp.; zob. art. 60 aktu w sprawie sztucznej inteligencji.

ust. 8 aktu w sprawie sztucznej inteligencji, planowane testy muszą być w pełni zgodne z określonymi w akcie wymogami dotyczącymi systemów zdalnej identyfikacji biometrycznej, chyba że system jest testowany w piaskownicy regulacyjnej w zakresie AI lub zgodnie ze specjalnymi zasadami testowania w warunkach rzeczywistych poza piaskownicą, zgodnie z art. 60 i 61 aktu[27].

33) W każdym przypadku wszelka działalność badawczo-rozwojowa (w tym wyłączona z zakresu stosowania aktu w sprawie sztucznej inteligencji) powinna być prowadzona zgodnie z uznanymi normami etycznymi i zawodowymi w zakresie badań naukowych oraz zgodnie z mającym zastosowanie prawem Unii[28] (np. przepisami o ochronie danych, które nadal mają zastosowanie).

2.5.4. Osobista działalność pozazawodowa

34) Art. 2 ust. 10 aktu w sprawie sztucznej inteligencji stanowi, że aktu tego „nie stosuje się do obowiązków podmiotów stosujących będących osobami fizycznymi, które korzystają z systemów AI w ramach czysto osobistej działalności pozazawodowej". Definicja podmiotu stosującego również wyklucza użytkowników zaangażowanych w taką działalność (zob. sekcja 2.4 powyżej). Za działalność „zawodową" należy uznać wszelką działalność, w ramach której osoba fizyczna regularnie uzyskuje korzyść ekonomiczną lub jest w inny sposób zaangażowana w działalność specjalistyczną, gospodarczą, handlową, zawodową lub niezależną. Określenie „osobista" odnosi się do określenia „pozazawodowa", co oznacza, że działalność danej osoby powinna mieć charakter zarówno osobisty, jak i pozazawodowy. Wyłączenie nie powinno zatem obejmować na przykład działalności przestępczej, ponieważ nie należy jej uznawać za działalność czysto osobistą.

Na przykład osoba korzystająca z systemu rozpoznawania twarzy w domu (np. w celu kontrolowania dostępu i monitorowania wejścia do domu pod kątem bezpieczeństwa) byłaby objęta wyłączeniem przewidzianym w art. 2 ust. 10 aktu w sprawie sztucznej inteligencji, a zatem nie podlegałaby wynikającym z aktu obowiązkom spoczywającym na podmiotach stosujących, nawet w przypadkach, w których jest zobowiązana do przekazania (fragmentów) nagrań organom ścigania.

Natomiast osoby fizyczne, takie jak freelancerzy, dziennikarze, lekarze itp., wykorzystujące system AI w działalności zawodowej, musiałyby wypełniać wynikające z aktu obowiązki spoczywające na podmiotach stosujących systemy rozpoznawania twarzy. Wszelkie wykorzystanie przez osoby fizyczne, w przypadku gdy w imieniu lub z upoważnienia podmiotu stosującego działają one w ramach obowiązków zawodowych, również będzie wchodzić w zakres stosowania aktu w sprawie sztucznej inteligencji.

Ponadto za działalność czysto osobistą nie można uznać działalności przestępczej, nawet jeśli nie dążono do korzyści ekonomicznej ani jej nie osiągnięto. W przypadku

[27] W akcie w sprawie sztucznej inteligencji określono szczegółowe i konkretne obowiązki dotyczące piaskownic regulacyjnych w zakresie AI i testów w warunkach rzeczywistych. Zob. art. 57 i nast. aktu w sprawie sztucznej inteligencji.

[28] Motyw 25 aktu w sprawie sztucznej inteligencji.

innych bezprawnych działań, takich jak nieprzestrzeganie przepisów o ochronie konsumentów lub o ochronie danych, a także krajowych przepisów administracyjnych, wyłączenie przewidziane w akcie w sprawie sztucznej inteligencji ma zastosowanie, ale nadal stosuje się również inne odpowiednie ramy prawne.

35) Wyłączenie przewidziane w art. 2 ust. 10 aktu w sprawie sztucznej inteligencji ma wyłącznie zastosowanie do obowiązków spoczywających na podmiotach stosujących, w przypadku gdy podmioty te wykorzystują system do czysto osobistej działalności pozazawodowej. Sam system jako taki wchodzi w zakres stosowania aktu w sprawie sztucznej inteligencji w odniesieniu do obowiązków dostawców wprowadzających system do obrotu lub oddających go do użytku, innych zawodowych podmiotów stosujących, a także innych odpowiedzialnych podmiotów, takich jak importerzy i dystrybutorzy.

Na przykład jeżeli system rozpoznawania emocji ma być wykorzystywany przez osoby fizyczne do czysto osobistej działalności pozazawodowej, zgodnie z klasyfikacją określoną w art. 6 aktu w sprawie sztucznej inteligencji pozostaje on systemem AI wysokiego ryzyka i musi być w pełni zgodny z wymogami aktu. Jednocześnie podmiot stosujący, który wykorzystuje go do czysto osobistych celów pozazawodowych (np. osoba autystyczna), nie będzie objęty wynikającymi z aktu w sprawie sztucznej inteligencji szczególnymi obowiązkami spoczywającymi na podmiotach stosujących, a wykorzystanie to nie będzie wchodzić w zakres stosowania aktu.

2.5.5. Systemy AI udostępniane na podstawie bezpłatnych licencji otwartego oprogramowania

36) Zgodnie z art. 2 ust. 12 aktu w sprawie sztucznej inteligencji akt ten nie ma zastosowania do systemów AI udostępnianych na podstawie bezpłatnych licencji otwartego oprogramowania[29], chyba że są one wprowadzane do obrotu lub oddawane do użytku jako systemy AI wysokiego ryzyka lub jako system AI objęty art. 5 (zakazane praktyki w zakresie AI) lub art. 50 aktu (obowiązki w zakresie przejrzystości w odniesieniu do niektórych systemów AI). Oznacza to, że dostawcy systemów AI nie mogą korzystać z tego wyłączenia, jeżeli system AI, który wprowadzają do obrotu lub oddają do użytku, powoduje wystąpienie praktyki zakazanej na mocy art. 5 aktu w sprawie sztucznej inteligencji.

2.6. Wzajemne powiązania zakazów i wymogów dotyczących systemów AI wysokiego ryzyka

37) Praktyki w zakresie AI zakazane na mocy art. 5 aktu w sprawie sztucznej inteligencji należy rozpatrywać w odniesieniu do systemów AI sklasyfikowanych jako systemy wysokiego ryzyka zgodnie z art. 6 aktu, w szczególności wymienionych w załączniku III[30]. Wynika to z faktu, że wykorzystywanie systemów AI sklasyfikowanych jako

[29] W motywie 102 aktu w sprawie sztucznej inteligencji stwierdzono, że udostępnianie oprogramowania i danych na podstawie bezpłatnej licencji otwartego oprogramowania „umożliwia ich ogólne upowszechnianie i zezwala użytkownikom na swobodny dostęp do nich, ich wykorzystywanie, zmianę i ich redystrybucję lub ich zmienionych wersji".

[30] W wykazie tym uwzględniono systemy AI oparte na danych biometrycznych, a także systemy AI wykorzystywane do konkretnych celów w niektórych dziedzinach, takich jak zatrudnienie, edukacja, dostęp do usług publicznych i prywatnych, ściganie przestępstw itp.

systemy wysokiego ryzyka może w niektórych przypadkach kwalifikować się jako zakazana praktyka, jeżeli spełnione są wszystkie warunki dotyczące co najmniej jednego zakazu określonego w art. 5 aktu w sprawie sztucznej inteligencji. Z drugiej strony, większość systemów AI objętych wyjątkiem od zakazu wymienionego w art. 5 aktu będzie kwalifikować się jako systemy wysokiego ryzyka.

Na przykład jeżeli systemy rozpoznawania emocji nie spełniają warunków dotyczących zakazu określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, klasyfikuje się je jako systemy AI wysokiego ryzyka zgodnie z art. 6 ust. 2 aktu w sprawie sztucznej inteligencji i pkt 1 lit. c) załącznika III do tego aktu. Podobnie niektóre oparte na sztucznej inteligencji systemy scoringu, takie jak systemy wykorzystywane do punktowej oceny kredytowej lub oceny ryzyka w ubezpieczeniach zdrowotnych i ubezpieczeniach na życie, będą uznawane za systemy AI wysokiego ryzyka, jeżeli nie spełniają warunków dotyczących zakazu określonego w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji[31]. Innym przykładem są systemy AI oceniające osoby i określające, czy są one uprawnione do korzystania z podstawowych świadczeń i usług w ramach pomocy publicznej, takich jak usługi opieki zdrowotnej i świadczenia z tytułu zabezpieczenia społecznego, które to systemy klasyfikuje się jako systemy wysokiego ryzyka[32]. Jeżeli takie systemy wiążą się z niedopuszczalnym scoringiem społecznym i spełniają warunki określone w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, ich wprowadzanie do obrotu, oddawanie do użytku i wykorzystywanie będzie w Unii zakazane.

W takich przypadkach ocena ryzyka i zarządzanie ryzykiem, przeprowadzane przez dostawcę, oraz zgodność z innymi wymogami dotyczącymi systemów AI wysokiego ryzyka (np. zarządzanie danymi, przejrzystość i nadzór ze strony człowieka), a także obowiązki podmiotu stosującego w zakresie odpowiedniego wykorzystania zgodnie z instrukcją obsługi i nadzoru ze strony człowieka (art. 26), a w niektórych przypadkach ocena skutków dla praw podstawowych (art. 27), powinny pomóc w zapewnieniu, aby wprowadzony do obrotu lub wdrożony system AI wysokiego ryzyka był zgodny z prawem i nie powodował wystąpienia zakazanej praktyki.

38) Ponadto systemy AI, których w drodze wyjątku nie uznaje się za systemy wysokiego ryzyka na podstawie art. 6 ust. 3 aktu w sprawie sztucznej inteligencji, mimo że stanowią przypadki wykorzystania obarczone wysokim ryzykiem określone w załączniku III, mogą nadal być objęte zakazami określonych w art. 5 aktu. Z art. 6 ust. 3 aktu wynika jedynie, że określonego systemu AI nie uznaje się za system wysokiego ryzyka; nie wyłącza się takich systemów AI z zakresu stosowania aktu i zakazów.

2.7. Stosowanie zakazów do systemów AI ogólnego przeznaczenia i systemów o określonym przeznaczeniu

[31] Jest o tym wyraźnie mowa w motywie 58 i w załączniku III do aktu w sprawie sztucznej inteligencji.
[32] Motyw 58 aktu w sprawie sztucznej inteligencji.

39) Zakazy mają zastosowanie do każdego systemu AI, niezależnie od tego, czy ma on określone „przeznaczenie”[33] czy jest systemem „ogólnego przeznaczenia” (tj. może służyć różnym celom), czy nadaje się do bezpośredniego wykorzystania czy do integracji z innymi systemami AI[34]. W związku z tym każdy operator powinien wprowadzić środki, które są dla niego najbardziej odpowiednie ze względu na jego rolę w łańcuchu wartości i kontrolę sprawowaną nad systemem, aby zapewnić odpowiedzialne i bezpieczne dostarczanie i wykorzystywanie systemów AI, wyważając związane z nimi ryzyko i korzyści z myślą o osiągnięciu dwojakiego celu aktu w sprawie sztucznej inteligencji.

40) Oczekuje się zatem, że podmioty stosujące nie będą wykorzystywać żadnego systemu AI w sposób zakazany na mocy art. 5 aktu w sprawie sztucznej inteligencji, w tym nie będą obchodzić żadnych zabezpieczeń wprowadzonych przez dostawców systemu. Chociaż szkoda często wynika ze sposobu, w jaki systemy AI wykorzystuje się w praktyce, dostawcy mają również obowiązek niewprowadzania do obrotu i nieoddawania do użytku systemów AI, w tym systemów AI ogólnego przeznaczenia, co do których istnieje uzasadnione prawdopodobieństwo, że będą zachowywać się lub będą bezpośrednio wykorzystywane w sposób zakazany na mocy art. 5 aktu w sprawie sztucznej inteligencji[35]. W tym kontekście oczekuje się również, że dostawcy zastosują skuteczne i weryfikowalne środki w celu wbudowania zabezpieczeń oraz zapobiegania takim szkodliwym zachowaniom i niewłaściwemu wykorzystaniu, a także ich ograniczania w zakresie, w jakim dają się one racjonalnie przewidzieć, i że środki te będą wykonalne i proporcjonalne do konkretnego systemu AI i okoliczności danego przypadku. Oczekuje się również, że w stosunkach umownych z podmiotami stosującymi (tj. w warunkach użytkowania systemu AI) dostawcy wykluczą wykorzystywanie swojego systemu AI do celów zakazanych praktyk i przedstawią w instrukcji obsługi odpowiednie informacje dla podmiotów stosujących oraz informacje na temat niezbędnego nadzoru ze strony człowieka.

> Na przykład system AI ogólnego przeznaczenia wykorzystywany jako chatbot może stosować techniki manipulacyjne i wprowadzające w błąd, które mogą z uzasadnionym prawdopodobieństwem wyrządzić poważną szkodę. Aby zapobiec zakazanemu zachowaniu systemu AI i jego wykorzystaniom, w przypadku których istnieje uzasadnione prawdopodobieństwo manipulacji, wprowadzenia w błąd i wyrządzenia poważnej szkody zgodnie z art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, oczekuje się, że dostawca wprowadzi odpowiednie i proporcjonalne środki (np. odpowiedni bezpieczny i etyczny projekt, zastosowanie zabezpieczeń technicznych i innych, ograniczenia dotyczące wykorzystywania, przejrzystość i kontrolę użytkownika, odpowiednie informacje zawarte w instrukcji obsługi) przed

[33] W art. 3 pkt 12 aktu w sprawie sztucznej inteligencji zdefiniowane jako zastosowanie, do jakiego system AI został przeznaczony przez jego dostawcę, w tym konkretny kontekst i warunki wykorzystywania, określone w informacjach dostarczonych przez dostawcę w instrukcji obsługi, materiałach promocyjnych lub sprzedażowych i oświadczeniach, jak również w dokumentacji technicznej.

[34] Zob. art. 3 pkt 66 aktu w sprawie sztucznej inteligencji.

[35] Wynika to w szczególności z odniesienia do „wprowadzania do obrotu” lub „oddawania do użytku” we wszystkich zakazach wymienionych w art. 5 aktu w sprawie sztucznej inteligencji, z wyjątkiem zakazu dotyczącego systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, o którym mowa w art. 5 ust. 1 lit. h), odnoszącego się wyłącznie do wykorzystywania.

wprowadzeniem systemu AI do obrotu (art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji) w celu zapewnienia, aby chatbot nie wyrządzał poważnych szkód użytkownikom ani innym osobom lub grupom osób (zob. również sekcja 3.2.3 lit. c)).

41) W niektórych przypadkach, w szczególności gdy zakazy są powiązane z bardzo konkretnym przeznaczeniem systemu[36], dostawcy mogą mieć ograniczone możliwości wprowadzenia innych środków zapobiegawczych i ograniczających ryzyko i będą musieli przede wszystkim zastosować środki polegające na zapewnieniu odpowiednich instrukcji i informacji podmiotom stosującym, na wymogu nadzoru ze strony człowieka i na ograniczeniu zakazanego wykorzystywania systemu. W stosownych przypadkach środki takie mogą również obejmować monitorowanie przestrzegania takiego ograniczenia, w zależności od sposobu dostarczenia systemu AI oraz od informacji, którymi dysponuje dostawca w odniesieniu do ewentualnego niewłaściwego wykorzystania. Ewentualne środki monitorowania mające na celu wykrywanie niewłaściwego wykorzystania nie powinny polegać na ogólnym monitorowaniu działań podmiotów stosujących i powinny być zgodne z prawem Unii.

Na przykład system AI ogólnego przeznaczenia, który potrafi rozpoznawać emocje lub wyciągać wnioski na ich temat, nie powinien być wykorzystywany przez podmioty stosujące w miejscu pracy ani w instytucjach edukacyjnych, chyba że w drodze wyjątku ze względów medycznych lub bezpieczeństwa. Dostawca może jednak nie być w stanie dowiedzieć się, w jakim konkretnym kontekście będzie wykorzystywana funkcja systemu umożliwiająca rozpoznawanie emocji oraz czy może mieć zastosowanie wyjątek od zakazu określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji. Dostawcy tacy mogą jednak wyraźnie wykluczyć w warunkach użytkowania takie zakazane wykorzystywanie i zawrzeć w instrukcji obsługi odpowiednie informacje stanowiące wytyczne dla podmiotów stosujących. Oczekuje się również, że zastosują oni odpowiednie środki, jeżeli dowiedzą się, że system jest wykorzystywany niewłaściwie i do tego konkretnego zakazanego celu przez konkretne podmioty stosujące, na przykład jeżeli takie niewłaściwe wykorzystanie zostanie zgłoszone lub dostawca dowie się o nim w inny sposób, co może mieć miejsce w przypadku, gdy system jest obsługiwany bezpośrednio za pośrednictwem platformy znajdującej się pod kontrolą dostawcy, a dostawca przeprowadza kontrole.

2.8. Wzajemne powiązania między zakazami a innymi przepisami prawa Unii

42) Akt w sprawie sztucznej inteligencji to rozporządzenie, które ma zastosowanie horyzontalne we wszystkich sektorach, bez uszczerbku dla innych przepisów Unii, w szczególności dotyczących ochrony praw podstawowych, ochrony konsumentów, zatrudnienia, ochrony pracowników i bezpieczeństwa produktów[37]. Akt w sprawie sztucznej inteligencji uzupełnia takie przepisy poprzez swoją logikę zapobiegania i bezpieczeństwa (systemów AI nie można wprowadzać do obrotu ani wykorzystywać w określony sposób) i zapewnia dodatkową ochronę poprzez przeciwdziałanie konkretnym szkodliwym praktykom w zakresie AI, które mogą nie być zakazane na

[36] Art. 5 ust. 1 lit. d)–h) aktu w sprawie sztucznej inteligencji.
[37] Art. 2 i motyw 9 aktu w sprawie sztucznej inteligencji.

mocy innych przepisów. Ponadto dzięki uwzględnieniu wcześniejszych etapów cyklu życia systemów AI (tj. wprowadzania do obrotu i oddawania do użytku) oraz wdrażania (tj. wykorzystywania) zakazy zawarte w akcie w sprawie sztucznej inteligencji umożliwiają przeciwdziałanie szkodliwym praktykom w zakresie AI w różnych punktach łańcucha wartości AI.

43) Jednocześnie akt w sprawie sztucznej inteligencji nie ma wpływu na zakazy mające zastosowanie w przypadku, gdy dana praktyka w zakresie AI wchodzi w zakres innych przepisów prawa Unii[38]. W związku z tym nawet jeśli system AI nie jest zakazany na mocy aktu w sprawie sztucznej inteligencji, jego wykorzystywanie może być nadal zakazane lub niezgodne z prawem w świetle innych przepisów pierwotnego lub wtórnego prawa Unii (np. z powodu nieprzestrzegania w danym przypadku praw podstawowych, w tym z powodu braku wymaganej na mocy przepisów o ochronie danych podstawy prawnej do przetwarzania danych osobowych, zakazanej przez prawo Unii dyskryminacji itp.). Przestrzeganie zakazów zawartych w akcie w sprawie sztucznej inteligencji nie jest zatem wystarczającym warunkiem zgodności z innymi przepisami Unii, które nadal mają zastosowanie do dostawców systemów AI i podmiotów je stosujących.

Na przykład wykorzystywane w miejscu pracy systemy rozpoznawania emocji wspomagane sztuczną inteligencją, które są wyłączone z zakazu określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, ponieważ używa się ich ze względów medycznych lub bezpieczeństwa, nadal podlegają przepisom o ochronie danych oraz przepisom unijnym i krajowym dotyczącym warunków zatrudnienia i pracy, w tym bezpieczeństwa i higieny pracy, które mogą przewidywać inne ograniczenia i zabezpieczenia związane z wykorzystywaniem takich systemów[39].

44) W przypadkach, w których konkretne działania związane z wprowadzaniem do obrotu lub wykorzystywaniem systemów AI objęte są również innymi przepisami Unii, akt w sprawie sztucznej inteligencji ma na celu zapewnienie spójnego wdrażania poszczególnych przepisów. Ponadto umożliwia on skuteczną współpracę między właściwymi organami odpowiedzialnymi za egzekwowanie aktu w sprawie sztucznej inteligencji a organami ochrony praw podstawowych na podstawie art. 77 aktu w sprawie sztucznej inteligencji i innych przepisów tego aktu. Ogólniej rzecz ujmując, zgodnie z art. 4 ust. 3 TUE poszczególne zainteresowane organy są zobowiązane do lojalnej współpracy przy wykonywaniu swoich zadań wynikających z prawa Unii.

45) W kontekście zakazów szczególnie istotna jest wzajemna zależność między aktem w sprawie sztucznej inteligencji a unijnymi przepisami o ochronie danych, ponieważ systemy AI często przetwarzają informacje dotyczące zidentyfikowanych lub możliwych do zidentyfikowania osób fizycznych („dane osobowe”)[40]. W zależności od zakazu i kontekstu najistotniejszymi aktami prawnymi w odniesieniu do takich systemów są: rozporządzenie (UE) 2016/679 w sprawie ochrony osób fizycznych

[38] Art. 5 ust. 8 aktu w sprawie sztucznej inteligencji.
[39] Zob. również motyw 9 aktu w sprawie sztucznej inteligencji.
[40] Art. 2 ust. 7 aktu w sprawie sztucznej inteligencji; zob. również motyw tego 10 aktu.

w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych (ogólne rozporządzenie o ochronie danych, dalej „RODO"), dyrektywa (UE) 2016/680 w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych przez właściwe organy do celów zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania i ścigania czynów zabronionych i wykonywania kar, w sprawie swobodnego przepływu takich danych (dyrektywa o ochronie danych w sprawach karnych) oraz rozporządzenie (UE) 2018/1725, w którym ustanowiono przepisy o ochronie danych dotyczące instytucji, organów i jednostek organizacyjnych UE. Zgodnie z art. 2 ust. 7 aktu w sprawie sztucznej inteligencji akt ten nie ma wpływu na wymienione akty prawne i będą miały one nadal zastosowanie wraz z aktem w sprawie sztucznej inteligencji, który jest spójny z dorobkiem prawnym UE w zakresie ochrony danych i stanowi jego uzupełnienie. TSUE wyjaśnił kilka aspektów tych unijnych przepisów o ochronie danych, a Europejska Rada Ochrony Danych przyjęła szereg wytycznych (np. dotyczących pojęcia „profilowania"[41], co jest szczególnie istotne dla zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, ponieważ zastosowano tam to samo pojęcie).

46) W odniesieniu do zakazów/ograniczeń dotyczących wykorzystywania systemów kategoryzacji biometrycznej i systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw, akt w sprawie sztucznej inteligencji ma zastosowanie jako *lex specialis* do art. 10 dyrektywy o ochronie danych w sprawach karnych, regulując tym samym w sposób wyczerpujący takie wykorzystywanie i przetwarzanie danych biometrycznych[42]. W tym kontekście akt w sprawie sztucznej inteligencji nie ma na celu ustanowienia podstawy prawnej do przetwarzania danych osobowych na podstawie art. 8 dyrektywy (UE) 2016/680. Wszystkie pozostałe przepisy tej dyrektywy mają zastosowanie dodatkowo w stosunku do warunków określonych w akcie w sprawie sztucznej inteligencji, w szczególności w odniesieniu do wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw, jeżeli jest to dozwolone, z zastrzeżeniem ograniczonych wyjątków określonych w art. 5 ust. 1 lit. h) aktu. Ogólniej rzecz biorąc, przy przetwarzaniu danych osobowych do celów ścigania przestępstw właściwe organy ścigania (tj. właściwe organy zgodnie z art. 3 ust. 7 dyrektywy o ochronie danych w sprawach karnych) muszą również przestrzegać przepisów dyrektywy o ochronie danych w sprawach karnych.

47) Zgodnie z art. 2 ust. 9 aktu w sprawie sztucznej inteligencji unijne przepisy dotyczące ochrony i bezpieczeństwa konsumentów również mają nadal pełne zastosowanie do systemów AI objętych zakresem tych przepisów.

Na przykład:

[41] Zob. również Grupa Robocza Art. 29, *Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679,* WP251rev.01, 6.2.2018, zatwierdzone przez EROD.

[42] Motyw 38 aktu w sprawie sztucznej inteligencji.

– praktyki w zakresie scoringu społecznego stosowane przez przedsiębiorców (w tym osoby fizyczne działające w ramach obowiązków zawodowych w stosunkach między przedsiębiorstwami a konsumentami), z zastrzeżeniem indywidualnej oceny każdego przypadku, również mogą zostać uznane za „nieuczciwe", a tym samym naruszające prawo ochrony konsumentów (tj. dyrektywę 2005/29/WE);

– wykorzystywanie systemu AI w celu wyciągania wniosków na temat emocji może również wymagać przestrzegania rozporządzenia (UE) 2017/745 (rozporządzenie w sprawie wyrobów medycznych), jeżeli system AI służy do celów diagnostyki medycznej lub leczenia.

48) Ponadto akt w sprawie sztucznej inteligencji stosuje się w związku z odpowiednimi przepisami dotyczącymi obowiązków dostawców usług pośrednich, którzy włączają systemy lub modele AI do swoich usług regulowanych rozporządzeniem (UE) 2022/2065 („akt o usługach cyfrowych"). W szczególności w art. 2 ust. 5 aktu w sprawie sztucznej inteligencji wskazano, że akt ten nie ma wpływu na stosowanie przepisów dotyczących odpowiedzialności takich dostawców określonych w rozdziale II aktu o usługach cyfrowych.

49) Ponadto zakazy zawarte w akcie w sprawie sztucznej inteligencji pozostają bez uszczerbku dla jakiejkolwiek odpowiedzialności, którą dostawca lub podmiot stosujący może ponosić z tytułu wyrządzonej szkody zgodnie z mającymi zastosowanie unijnymi lub krajowymi przepisami dotyczącymi odpowiedzialności[43].

50) Ponadto zakazy określone w art. 5 aktu w sprawie sztucznej inteligencji oraz wyraźne wyjątki od nich nie mogą być wykorzystywane do obchodzenia obowiązków wynikających z innych przepisów Unii ani jako uzasadnienie naruszenia tych obowiązków.

51) Jako prawo wtórne Unii akt w sprawie sztucznej inteligencji należy interpretować w świetle podstawowych praw i wolności zagwarantowanych w Traktatach UE i w Karcie, a także praw i wolności chronionych konwencjami międzynarodowymi, których Unia jest stroną[44].

52) Dodatkowe wyjaśnienia dotyczące wzajemnego powiązania konkretnych zakazów z innymi przepisami prawa Unii przedstawiono w odpowiednich sekcjach poniżej.

2.9. Egzekwowanie art. 5 aktu w sprawie sztucznej inteligencji

2.9.1. Organy nadzoru rynku

[43] Warunki dotyczące odpowiedzialności (związane ze szkodą, osobą odpowiedzialną, winą lub ciężarem dowodu itp.) będą określane zgodnie z mającymi zastosowanie przepisami Unii, takimi jak dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/2853 z dnia 23 października 2024 r. w sprawie odpowiedzialności za produkty wadliwe (Tekst mający znaczenie dla EOG), Dz.U. L, 2024/2853, 18.11.2024, lub mającymi zastosowanie przepisami krajowymi dotyczącymi odpowiedzialności (zob. również wniosek dotyczący dyrektywy Parlamentu Europejskiego i Rady w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji (dyrektywa w sprawie odpowiedzialności za sztuczną inteligencję), COM(2022) 496 final).

[44] Mimo że Unia nie jest jeszcze stroną europejskiej Konwencji o ochronie praw człowieka i podstawowych wolności, art. 59 ust. 3 Karty stanowi, że w zakresie, w jakim zawiera ona prawa, które odpowiadają prawom zagwarantowanym w europejskiej Konwencji o ochronie praw człowieka i podstawowych wolności, ich znaczenie i zakres są takie same jak praw przyznanych przez tę konwencję. Przepis ten nie stanowi przeszkody, aby prawo Unii przyznawało szerszą ochronę.

53) Za egzekwowanie przepisów aktu w sprawie sztucznej inteligencji dotyczących systemów AI, w tym zakazów, odpowiadają organy nadzoru rynku wyznaczone przez państwa członkowskie oraz Europejski Inspektor Ochrony Danych (jako organ nadzoru rynku właściwy dla instytucji, organów i jednostek organizacyjnych UE). Takie egzekwowanie odbywa się w ramach systemu nadzoru rynku i zgodności produktów ustanowionego rozporządzeniem (UE) 2019/1020[45], zgodnie z innymi unijnymi przepisami dotyczącymi bezpieczeństwa produktów. Uprawnienia wykonawcze organów nadzoru rynku w odniesieniu do systemów AI określono w akcie w sprawie sztucznej inteligencji i w rozporządzeniu (UE) 2019/1020. Organy te mogą podejmować działania w zakresie egzekwowania prawa w odniesieniu do zakazów z własnej inicjatywy lub na podstawie skargi; skargę ma prawo złożyć każda osoba, na którą AI ma wpływ, lub każda inna osoba fizyczna lub prawna mająca podstawy, by uznać, że zostały naruszone przepisy[46]. Państwa członkowskie muszą wyznaczyć swoje właściwe organy nadzoru rynku do dnia 2 sierpnia 2025 r.

54) W kontekście egzekwowania zakazów szczególnie istotna jest przewidziana w akcie w sprawie sztucznej inteligencji procedura postępowania na poziomie krajowym w przypadku systemów AI stwarzających ryzyko[47]. W przypadku skutków transgranicznych wykraczających poza terytorium podlegające organowi nadzoru rynku organ danego państwa członkowskiego musi poinformować o tym Komisję i organy nadzoru rynku pozostałych państw członkowskich. Wszystkie organy nadzoru rynku powinny zastosować unijną procedurę ochronną, a Komisja podejmuje decyzję[48] stwierdzającą, czy system AI powoduje wystąpienie zakazanej praktyki. Procedura ta służy zapewnieniu jednolitego stosowania zakazów we wszystkich państwach członkowskich, co gwarantuje pewność prawa zarówno dostawcom systemów AI, jak i podmiotom je stosującym. Aby zapewnić jednolite stosowanie aktu w sprawie sztucznej inteligencji, krajowe organy nadzoru rynku powinny również dążyć do zharmonizowanego stosowania zakazów w porównywalnych przypadkach, które nie wykraczają poza terytorium państwa członkowskiego, czerpiąc inspirację z niniejszych wytycznych i współpracując w ramach Europejskiej Rady ds. Sztucznej Inteligencji[49].

2.9.2. Kary

55) W akcie w sprawie sztucznej inteligencji przyjęto wielopoziomowe podejście do nakładania kar za nieprzestrzeganie poszczególnych przepisów aktu, w zależności od powagi naruszenia. Za najpoważniejsze naruszenie uznaje się nieprzestrzeganie zakazów określonych w art. 5 aktu w sprawie sztucznej inteligencji, w związku z czym podlega ono najwyższej karze pieniężnej. Dostawcy i podmioty stosujące dopuszczający się zakazanych praktyk w zakresie AI mogą podlegać karom pieniężnym w wysokości do 35 000 000 EUR lub – jeżeli sprawcą naruszenia jest przedsiębiorstwo – w wysokości do 7 % jego całkowitego rocznego światowego obrotu z poprzedniego roku obrotowego, w zależności od tego, która z tych kwot jest

[45] Zob. również motyw 156 aktu w sprawie sztucznej inteligencji.
[46] Art. 85 aktu w sprawie sztucznej inteligencji.
[47] Art. 79 aktu w sprawie sztucznej inteligencji.
[48] Art. 81 aktu w sprawie sztucznej inteligencji.
[49] Art. 65 i 66 aktu w sprawie sztucznej inteligencji.

wyższa[50]. Każde państwo członkowskie powinno ustanowić przepisy określające, czy i w jakim zakresie można nakładać administracyjne kary pieniężne na organy i podmioty publiczne ustanowione w tym państwie członkowskim jako na dostawców systemów AI i podmioty stosujące takie systemy. Instytucje, organy i jednostki organizacyjne UE, które naruszają zakazy, mogą podlegać administracyjnym karom pieniężnym w wysokości do 1 500 000 EUR[51].

56) Możliwe jest, że to samo zakazane zachowanie stanowi naruszenie dwóch lub większej liczby przepisów aktu w sprawie sztucznej inteligencji (np. brak oznakowania treści deepfake może również stanowić technikę wprowadzającą w błąd w rozumieniu art. 5 ust. 1 lit. a) aktu). W takich przypadkach należy przestrzegać zasady *ne bis in idem*. W każdym razie należy wziąć pod uwagę kryteria ustalania wysokości kary przewidziane w art. 99 ust. 7 aktu w sprawie w sprawie sztucznej inteligencji.

57) Ponieważ naruszenia zakazów określonych w art. 5 aktu w sprawie sztucznej inteligencji w największym stopniu naruszają wolności innych osób i skutkują najwyższymi karami pieniężnymi, ich zakres powinien podlegać ścisłej wykładni.

3. ART. 5 UST. 1 LIT. A) I B) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – SZKODLIWE MANIPULOWANIE, WPROWADZANIE W BŁĄD I WYKORZYSTYWANIE

58) Dwa pierwsze zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji mają na celu ochronę osób fizycznych i osób szczególnie wrażliwych przed znacząco szkodliwymi skutkami manipulacji i wykorzystywania wspomaganych sztuczną inteligencją. Zakazy te są ukierunkowane na systemy AI stosujące techniki podprogowe, celowe techniki manipulacyjne lub wprowadzające w błąd, które są znacząco szkodliwe i mają istotny wpływ na zachowanie osób fizycznych lub grup osób (art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji) bądź wykorzystują słabości ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną (art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji).

3.1. Uzasadnienie i cele

59) Przesłanką leżącą u podstaw tych zakazów jest ochrona autonomii i dobrostanu osób fizycznych przed praktykami w zakresie AI, które polegają na manipulacji i wykorzystywaniu słabości oraz wprowadzają w błąd, a także mogą podważać i ograniczać autonomię osoby, jej decyzyjność i swobodę wyboru[52]. Zakazy te mają na celu ochronę prawa do godności ludzkiej (art. 1 Karty), które stanowi również podstawę wszystkich praw podstawowych i obejmuje autonomię osoby jako istotny aspekt. W szczególności zakazy te mają na celu zapobieganie manipulacjom i wykorzystywaniu słabości za pośrednictwem systemów AI, które redukują osoby fizyczne do roli zwykłych narzędzi służących osiągnięciu określonych celów, a także ochronę osób, które są najbardziej wrażliwe i podatne na szkodliwe manipulacje

[50] Art. 99 aktu w sprawie sztucznej inteligencji.
[51] Art. 100 aktu w sprawie sztucznej inteligencji.
[52] Motyw 29 aktu w sprawie sztucznej inteligencji.

i wykorzystywanie. Zakazy dotyczące znacząco szkodliwych praktyk w zakresie AI noszących znamiona manipulacji i wykorzystywania oraz wprowadzających w błąd są w pełni zgodne z szerszymi celami aktu w sprawie sztucznej inteligencji, dotyczącymi promowania wiarygodnych i zorientowanych na człowieka systemów AI, które są bezpieczne, przejrzyste, sprawiedliwe i służą ludzkości oraz są zgodne z przewodnią rolą człowieka i wartościami UE.

3.2. Główne elementy zakazu określonego w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji – szkodliwa manipulacja

Art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji stanowi:

1. Zakazuje się następujących praktyk w zakresie AI:

a) wprowadzania do obrotu, oddawania do użytku lub wykorzystywania systemu AI, który stosuje techniki podprogowe będące poza świadomością danej osoby lub celowe techniki manipulacyjne lub wprowadzające w błąd, czego celem lub skutkiem jest dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób poprzez znaczące ograniczenie ich zdolności do podejmowania świadomych decyzji, powodując tym samym podjęcie przez nie decyzji, której inaczej by nie podjęły, w sposób, który wyrządza lub z uzasadnionym prawdopodobieństwem może wyrządzić u niej, u innej osoby lub u grupy osób poważną szkodę;

60) Aby zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi być spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku" lub „wykorzystywaniu" systemu AI;

(ii) system AI musi stosować techniki podprogowe (będące poza świadomością danej osoby), celowe techniki manipulacyjne lub techniki wprowadzające w błąd;

(iii) celem lub skutkiem zastosowania tych technik przez system AI powinno być dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób. Zmiana ta musi znacząco ograniczać ich zdolność do podejmowania świadomych decyzji, powodując tym samym podjęcie przez nie decyzji, której dana osoba lub grupa osób inaczej by nie podjęły;

(iv) zmienione zachowanie musi wyrządzać poważną szkodę tej osobie, innej osobie lub grupie osób bądź musi istnieć uzasadnione prawdopodobieństwo, że wyrządzi taką szkodę.

61) Aby zakaz miał zastosowanie, muszą być spełnione jednocześnie wszystkie cztery warunki i musi istnieć wiarygodny związek przyczynowy między stosowanymi technikami, znaczącą zmianą zachowania danej osoby i poważną szkodą, do której doprowadziła ta zmiana zachowania lub do której z uzasadnionym prawdopodobieństwem może ona doprowadzić.

62) Pierwszy warunek, tj. „wprowadzanie do obrotu", „oddawanie do użytku" lub „wykorzystywanie" systemu AI, już przeanalizowano. Zakaz ma zatem zastosowanie

zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów do obrotu, oddawać ich do użytku ani wykorzystywać. Kolejne sekcje dotyczą pozostałych trzech warunków.

3.2.1. Techniki podprogowe, celowe techniki manipulacyjne lub techniki wprowadzające w błąd

63) W art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji zakazano trzech różnych rodzajów technik manipulacyjnych: a) technik podprogowych będących poza świadomością danej osoby: b) celowych technik manipulacyjnych; oraz c) technik wprowadzających w błąd. Aby system AI wchodził w zakres stosowania art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, musi on stosować co najmniej jedną z tych technik.

a) Techniki podprogowe

64) Chociaż w akcie w sprawie sztucznej inteligencji nie zdefiniowano „technik podprogowych", w art. 5 ust. 1 lit. a) aktu w określono, że techniki podprogowe działają poza świadomością danej osoby (poniżej lub powyżej jej progów). Ponieważ techniki podprogowe i sposób ich działania mają charakter niejawny, omijają one racjonalne środki obrony danej osoby przed manipulacją i mogą wpływać na decyzje bez świadomości danej osoby, co budzi poważne obawy etyczne i ogranicza autonomię osoby, jej decyzyjność i swobodę wyboru[53].

65) Techniki podprogowe muszą być w stanie wpływać na zachowanie w taki sposób, by dana osoba pozostała nieświadoma tego wpływu, jego sposobu działania lub wpływu na podejmowanie decyzji bądź kształtowanie wartości i opinii tej osoby. W szczególności techniki podprogowe mogą wykorzystywać bodźce dostarczane za pośrednictwem mediów dźwiękowych, wizualnych lub dotykowych, które są zbyt krótkie lub subtelne, aby można je było zaobserwować, i które są tradycyjnie znane i zakazane w innych sektorach, takich jak reklama medialna[54]. Bodźce te, choć nie są świadomie postrzegane, mogą być nadal przetwarzane przez mózg i wpływać na zachowanie.

Przykłady technik podprogowych (niekoniecznie zakazanych, chyba że spełnione są wszystkie pozostałe warunki wymienione w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji) to m.in.:

– **wizualne przekazy podprogowe:** system AI może pokazywać lub osadzać obrazy lub teksty wyświetlane krótko podczas odtwarzania wideo, pod względem technicznym widoczne, ale wyświetlane zbyt szybko, aby zostały świadomie zarejestrowane przez umysł użytkownika, mogące jednak nadal wpływać na postawy lub zachowania;

[53] Motyw 29 aktu w sprawie sztucznej inteligencji.

[54] Zob. w szczególności dyrektywa Parlamentu Europejskiego i Rady 2010/13/UE z dnia 10 marca 2010 r. w sprawie koordynacji niektórych przepisów ustawowych, wykonawczych i administracyjnych państw członkowskich dotyczących świadczenia audiowizualnych usług medialnych (Dz.U. L 95 z 15.04.2010, s. 1) („dyrektywa o audiowizualnych usługach medialnych"), która bezwzględnie zakazuje technik podprogowych w handlowych przekazach audiowizualnych.

– dźwiękowe przekazy podprogowe: system AI może stosować dźwięki lub komunikaty werbalne o niskiej głośności lub zamaskowane innymi dźwiękami, wpływając na słuchającego bez jego świadomości. Dźwięki te pod względem technicznym nadal mieszczą się w zakresie słyszalności, ale słuchający nie wychwytuje ich świadomie ze względu na ich subtelność lub maskowanie innym dźwiękiem;

– dotykowe bodźce podprogowe: system AI może stymulować subtelne wrażenia fizyczne, odbierane w sposób nieświadomy i mogące wpływać na stan emocjonalny lub zachowanie;

– sygnały poniżej progu widoczności lub słyszalności: system AI może stosować bodźce, które nie tylko są subtelne lub zamaskowane, ale też prezentowane w sposób sprawiający, że w normalnych warunkach są one całkowicie niewykrywalne przez ludzkie zmysły, np. bodźce wizualne (takie jak obrazy) wyświetlane zbyt szybko, aby ludzkie oko mogło je świadomie wychwycić, lub dźwięki odtwarzane z takim natężeniem, aby były niesłyszalne dla ludzkiego ucha;

– obrazy osadzone: system AI może ukrywać obrazy w innych treściach wizualnych; obrazy te nie są świadomie postrzegane, ale mogą być nadal przetwarzane przez mózg i wpływać na zachowanie;

– odwracanie uwagi: system AI może zwracać uwagę na konkretne bodźce lub treści, aby zapobiec dostrzeżeniu innych treści, często poprzez wykorzystywanie błędów poznawczych i niedoborów uwagi;

– manipulacja czasem: system AI może zmieniać percepcję czasu w interakcjach z użytkownikiem, wpływając tym samym na jego zachowanie i wywołując niecierpliwość i zależność.

66) Szybki rozwój sztucznej inteligencji i powiązanych z nią technologii, takich jak analiza dużych zbiorów danych, neurotechnologie, interfejsy mózg-komputer i rzeczywistość wirtualna, zwiększa ryzyko zaawansowanej manipulacji podprogowej i jej zdolność do skutecznego wywierania podświadomego wpływu na ludzkie zachowanie[55]. AI może również leżeć u podstaw powstających interfejsów maszyna-mózg i zaawansowanych technik, takich jak manipulowanie snami i oprogramowanie do szpiegowania mózgu.

Na przykład gra może wykorzystywać wspomagane sztuczną inteligencją neurotechnologie i interfejsy maszyna-mózg, które umożliwiają użytkownikom kontrolowanie (części) gry za pomocą zestawów nagłownych wykrywających aktywność mózgu. AI może służyć do trenowania mózgu użytkownika – potajemnie i bez jego świadomości – z myślą o ujawnianiu lub wyprowadzaniu z danych neuronowych informacji, które mogą być bardzo inwazyjne i wrażliwe (np. danych osobowych dotyczących bankowości, informacji intymnych itp.), w sposób, który może wyrządzić użytkownikowi poważne szkody. Zakaz, o którym mowa w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji dotyczy wyłącznie przypadków takiej znacząco szkodliwej manipulacji podprogowej, a nie ogólnie zastosowań interfejsu

[55] Zob. motyw 29 aktu w sprawie sztucznej inteligencji.

maszyna-mózg, jeżeli projektuje się je w sposób bezpieczny i z poszanowaniem prywatności i autonomii osoby.

b) Celowe techniki manipulacyjne

67) Akt w sprawie sztucznej inteligencji nie zawiera definicji „celowych technik manipulacyjnych", ale należy je rozumieć jako techniki zaprojektowane lub obiektywnie wykorzystywane do wywierania wpływu na zachowanie danej osoby, zmiany tego zachowania lub kontrolowania go w sposób, który podważa indywidualną autonomię i swobodę wyboru tej osoby. Techniki manipulacyjne projektuje się zwykle w z myślą o wykorzystaniu błędów poznawczych, podatności psychologicznych lub czynników sytuacyjnych, ze względu na które osoby fizyczne są bardziej podatne na wpływy. Dzięki swoim zdolnościom adaptacji systemy AI są również w stanie odpowiednio reagować na indywidualne uwarunkowania lub podatności danej osoby oraz zwiększać skuteczność i wpływ manipulacji na dużą skalę. Chociaż zdolność do manipulacji to istotny element dla określenia charakteru tej techniki, nie jest konieczne, aby intencją dostawcy, podmiotu stosującego lub samego systemu stosującego techniki manipulacyjne również było wyrządzenie szkody[56].

68) Chociaż nie wszystkie techniki manipulacyjne działają poza świadomością użytkownika, wiele z nich może się pokrywać z technikami podprogowymi, jako że docelowym efektem tych ostatnich również jest manipulacja. W motywie 29 aktu w sprawie sztucznej inteligencji wyjaśniono, że zakaz określony w art. 5 ust. 1 lit. a) obejmuje również techniki, w przypadku których osoby, nawet jeśli są świadome próby wywierania wpływu, mogą nie być w stanie kontrolować takiego efektu manipulacji ani się mu oprzeć[57]. W rezultacie na osoby wywierany jest wpływ lub nacisk powodujący, że przejawiają zachowania i podejmują decyzje, których w normalnych warunkach nie przejawiałyby ani nie podjęłyby, gdyby nie były poddawane technikom manipulacyjnym w stopniu ograniczającym ich indywidualną autonomię lub swobodę wyboru.

Przykładem celowych technik manipulacyjnych jest manipulacja sensoryczna, w ramach której system AI wykorzystuje dźwięki lub obrazy w tle, prowadząc do zmian nastrojów, na przykład nasilając lęki i stres psychiczny, co wpływa na zachowanie użytkowników do tego stopnia, że powodują one poważne szkody.

Innym przykładem jest spersonalizowana manipulacja, w przypadku której system AI, tworzący i dopasowujący wysoce przekonujące komunikaty w oparciu o dane osobowe danej osoby fizycznej lub wykorzystujący inne podatności danej osoby, wpływa na jej zachowanie lub wybory w sposób powodujący poważną szkodę.

69) Zakaz stosowania celowych technik manipulacyjnych obejmuje również systemy AI, które manipulują osobami fizycznymi w przypadku, gdy nie jest to intencją żadnego człowieka. W art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji zakazano wykorzystywania systemów AI, które stosują pewne techniki lub wykazują określone

[56] W tym kontekście zob. motyw 28 oraz sekcje 3.2.2 i 3.2.3 wytycznych.
[57] Motyw 28 aktu w sprawie sztucznej inteligencji.

zachowania manipulacyjne. W związku z tym takie techniki manipulacyjne może stosować również system AI, nie zaś dostawca czy podmiot stosujący, który w ten sposób zaprojektował lub wykorzystuje system.

Na przykład, niezależnie od tego, czy jest to intencją dostawcy, system AI może nauczyć się technik manipulacyjnych, ponieważ dane, na których był trenowany, zawierają wiele przykładów technik manipulacyjnych[58] lub ponieważ uczenie przez wzmacnianie na podstawie informacji zwrotnych od człowieka może przyjąć formę „gry" opartej na technikach manipulacyjnych[59].

Natomiast jeśli manipulacyjne zachowanie systemu ma jedynie charakter przypadkowy, nie należy uznawać, że stosuje on celowe techniki manipulacyjne, o ile dostawca wprowadził odpowiednie środki zapobiegawcze i ograniczające ryzyko w przypadku uzasadnionego prawdopodobieństwa wystąpienia poważnych szkód (zob. sekcja 3.2.3 lit. c) poniżej).

c) Techniki wprowadzające w błąd

70) W akcie w sprawie sztucznej inteligencji nie zdefiniowano „technik wprowadzających w błąd". W motywie 29 aktu w sprawie sztucznej inteligencji wyjaśniono, że są to techniki, które podważają lub ograniczają autonomię osoby, jej decyzyjność lub swobodę wyboru w sposób, którego taka osoba nie jest świadoma lub nawet jeśli jest świadoma, może zostać wprowadzona w błąd lub nie jest w stanie sprawować nad takimi technikami kontroli ani im się przeciwstawić. „Techniki wprowadzające w błąd" stosowane przez systemy AI należy rozumieć jako techniki polegające na przedstawianiu nieprawdziwych lub mylących informacji, przy czym celem lub skutkiem jest wprowadzanie osób fizycznych w błąd i wpływanie na ich zachowanie w sposób podważający ich autonomię, decyzyjność i swobodę wyboru.

71) W tym kontekście wyjaśnienia wymaga wzajemne oddziaływanie między zakazem określonym w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji a obowiązkami podmiotu stosującego określonymi w art. 50 ust. 4 aktu w zakresie oznaczania treści typu deepfake i niektórych generowanych przez AI publikacji tekstowych o sprawach leżących w interesie publicznym[60], a także spoczywającym na dostawcy obowiązkiem zapewnienia, aby systemy AI wchodzące w interakcje z ludźmi projektowano w sposób zapewniający użytkownikom informację, że wchodzą w interakcję z AI, a nie z człowiekiem[61]. Takie wyraźne ujawnienie stanowi środek ograniczania ryzyka, który należy wprowadzać także poprzez elementy projektu wbudowane w system AI udostępniany przez dostawcę, włącznie ze środkami technicznymi umożliwiającym

---

58 M. Carroll i in., „Characterising Manipulation from AI Systems" [Charakterystyka manipulacji ze strony systemów AI], w: Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23), 30 października–1 listopada, 2023, Boston, MA, USA. ACM, Nowy Jork, NY, USA, 13 stron. https://doi.org/10. 1145/3617694.3623226 |:2303.09387.

59 D. Amodei i in., Concrete Problems in AI Safety [Konkretne problemy związane z bezpieczeństwem AI], 36. konferencja na temat neuronowych systemów przetwarzania informacji (NeurIPS 2022). arXiv:1606.06565; J. Skalse i in. Defining and Characterizing Reward Gaming [Reward gaming – definicja i charakterystyka], Advances in Neural Information Processing Systems 35 (NeurIPS 2022) C. Denison i in., Sycophancy to Subterfuge: Investigating Reward-Tampering in Large Language Models [Od pochlebstwa do podstępu: badanie manipulowania nagrodami w dużych modelach językowych], 36. konferencja na temat neuronowych systemów przetwarzania informacji (NeurIPS 2022), arXiv:2406.10162.

60 Art. 50 ust. 4 aktu w sprawie sztucznej inteligencji.

61 Art. 50 ust. 1 aktu w sprawie sztucznej inteligencji.

wykrywanie treści generowanych i zmanipulowanych przez AI[62]. Widoczne oznakowanie treści typu deepfake i chatbotów zmniejsza ryzyko wprowadzenia w błąd, które może wystąpić w przypadku publicznego rozpowszechniania treści generowanych przez AI, a także ryzyko szkodliwego zakłócającego wpływu na kształtowanie opinii i przekonań oraz zachowanie danej osoby.

72) Z kolei zakaz zawarty w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji ma znacznie bardziej ograniczony zakres. Może on obejmować na przykład przypadki, w których chatbot lub wprowadzające w błąd treści generowane przez AI przedstawiają nieprawdziwe lub mylne informacje w sposób mający na celu wprowadzenie w błąd osób lub dokonanie zmiany ich zachowania, która nie miałaby miejsca, gdyby nie były one narażone na interakcję z systemem AI lub z wygenerowanymi przez AI wprowadzającymi w błąd treściami, w szczególności jeżeli nie zostały one wyraźnie ujawnione[63].

73) Podobnie jak w przypadku celowych technik manipulacyjnych zakaz stosowania technik wprowadzających w błąd może również obejmować systemy AI wprowadzające w błąd w przypadku, gdy nie jest to intencją człowieka (zob. sekcja 3.2.1 lit. b) powyżej). Na przykład, niezależnie od tego, czy ich dostawcy zamierzają osiągnąć taki rezultat, systemy AI mogą uczyć się wprowadzających w błąd technik po prostu dlatego, że poprawia to ich skuteczność w odniesieniu do zadania, do którego je zaprojektowano, na przykład dzięki uczeniu przez wzmacnianie[64].

Przykładem technik wprowadzających w błąd, które mogą być stosowane przez AI, jest chatbot AI, który podszywa się pod znajomego lub członka rodziny danej osoby, używając syntetycznego głosu i usiłując udawać człowieka, co jest metodą oszustwa i powoduje poważne szkody.

Innym przykładem takich technik jest system AI, który uczy się rozpoznawać, kiedy jest poddawany ocenie, i tymczasowo wstrzymuje wszelkie niepożądane zachowania, po czym wznawia je po zakończeniu takiej oceny[65]. Takie wprowadzające w błąd zachowanie jest szczególnie niebezpieczne, ponieważ udaremnia ewentualny zewnętrzny nadzór człowieka nad systemem, i może być zakazane, jeżeli istnieje w jego przypadku uzasadnione prawdopodobieństwo wyrządzenia poważnych szkód.

---

62 Art. 50 ust. 2 aktu w sprawie sztucznej inteligencji.

63 Chociaż co do zasady obowiązki w zakresie przejrzystości określone w art. 50 aktu w sprawie sztucznej inteligencji mają na celu zminimalizowanie manipulacji ze strony treści typu deepfake i chatbotów, mogą wystąpić przypadki i konteksty, w których pomimo przekazanych informacji takie wprowadzające w błąd techniki mogą nadal mieć znaczący wpływ na osoby i zmieniać ich zachowanie w stopniu ograniczającym ich autonomię i zdolność do podejmowania świadomych decyzji, w związku z czym techniki takie nie powinny być wykorzystywane do celów dezinformacji i manipulacji i w niektórych przypadkach mogą być nadal objęte zakazem określonym w art. 5 ust. 1 lit. a), jeżeli spełnione są wszystkie inne warunki zakazu (w tym dotyczące poważnych szkód).

64 F. Ward, F. Toni, F. Belardinelli, T. Everitt, Honesty Is the Best Policy: Defining and Mitigating AI Deception (neurips.cc) [Uczciwość jest najlepszą polityką: określanie i ograniczanie wprowadzania w błąd przez AI]; Advances in Neural Information Processing Systems 36 (NeurIPS 2023); P. Park i in. AI deception:A survey of examples, risks, and potential solutions [Wprowadzanie w błąd przez AI: przegląd przykładów, zagrożeń i potencjalnych rozwiązań] [2406.10162] Patterns, t. 5, nr 5, 100988.

65 J. Lehman, J. Clune, D. Misevic, C. Adami, L. Altenberg, J. Beaulieu i in. The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities [Zaskakująca kreatywność cyfrowej ewolucji: zbiór anegdot ze społeczności naukowych zajmujących się obliczeniami ewolucyjnymi i sztucznym życiem]. Artificial life, 26(2):274–306, 2020.

Nie można natomiast uznać, że system generatywnej AI, który przypadkowo przedstawia fałszywe lub mylne informacje i ulega halucynacjom[66], stosuje techniki wprowadzające w błąd w rozumieniu art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, biorąc pod uwagę ograniczenia i poziom rozwoju generatywnej AI. W szczególności może to mieć miejsce w przypadku, gdy dostawca systemu odpowiednio poinformował użytkowników o ograniczeniach systemu i włączył do systemu odpowiednie zabezpieczenia w celu zminimalizowania takich wyników, a także pod warunkiem że system nie jest przeznaczony do stosowania ani stosowany we wrażliwych kontekstach (np. opieka zdrowotna, edukacja, wybory), w których prawdopodobne jest wystąpienie poważnych szkodliwych skutków (zob. również uwagi w sekcji 3.2.3 lit. c) poniżej).

d) Kombinacja technik

74) Art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji ma zastosowanie do technik podprogowych, celowych technik manipulacyjnych i technik wprowadzających w błąd oraz do kombinacji takich technik, które to kombinacje mogą mieć złożony wpływ. Jak stwierdzono powyżej, celowe techniki manipulacyjne mogą mieć również charakter podprogowy, jeżeli działają poza świadomością użytkownika.

75) Ponadto łączne stosowanie celowych technik manipulacyjnych i technik wprowadzających w błąd może mieć znaczący wpływ na zachowanie osób, skłaniając je do podejmowania decyzji w oparciu o nieuświadomione manipulacje i fałszywe przekonania. Takie połączenie może skutkować sprzężeniem zwrotnym, w którym mniejsze jest prawdopodobieństwo zakwestionowania lub krytycznej oceny przez osoby otrzymanych informacji, ponieważ elementy manipulacyjne już wywołały u nich błędy poznawcze i reakcje emocjonalne.

3.2.2. Cel lub skutek, jakim jest dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób

76) Trzecim warunkiem, który musi zostać spełniony, aby zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji miał zastosowanie, jest to, że celem lub skutkiem zastosowanej techniki podprogowej, celowej techniki manipulacyjnej lub techniki wprowadzającej w błąd musi być „dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób”. Oznacza to istotny wpływ na zachowanie, ograniczający autonomię i swobodę wyboru danej osoby, nie zaś niewielki wpływ. Intencja nie jest jednak koniecznym wymogiem, ponieważ art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji obejmuje również praktyki, które mogą mieć jedynie „skutek” polegający na znaczącej zmianie zachowania. Między potencjalną znaczącą zmianą zachowania a techniką podprogową, celową techniką manipulacyjną lub techniką wprowadzającą w błąd stosowaną przez system AI powinien istnieć wiarygodny/racjonalnie prawdopodobny związek przyczynowy.

[66] „Halucynacja” to termin używany do opisania wady technicznej w systemach generatywnej AI, polegającej na tym, że systemy te generują niepożądane informacje, które są zmyślone lub niezgodne ze stanem faktycznym, a nie jest to zamierzone przez twórców systemu. Więcej informacji można znaleźć w: Ji Ziwei i in., Survey of Hallucination in Natural Language Generation [Analiza halucynacji w generowaniu języka naturalnego], ACM Computing Surveys, 55, nr 12, artykuł nr: 248, s. 1–38.

a) Pojęcie „znaczącej zmiany zachowania”

77) Pojęcie „znaczącej zmiany zachowania” osoby lub grupy osób jest centralnym elementem art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji. Wiąże się ono ze stosowaniem technik podprogowych, celowych technik manipulacyjnych lub technik wprowadzających w błąd, które są w stanie wpłynąć na zachowanie ludzi w sposób znacząco ograniczający ich zdolność do podejmowania świadomych decyzji, co powoduje, że przejawiają oni zachowania lub podejmują decyzje, których inaczej by nie przejawiali lub nie podjęli.

78) „Znaczące ograniczenie” odnosi się do istotnie zmniejszonej zdolności do podejmowania świadomych i autonomicznych decyzji, co powoduje, że osoby przejawiają zachowania lub podejmują decyzje, których inaczej by nie przejawiały lub nie podjęły. Wykracza to poza niewielkie lub znikome skutki i wiąże się ze znaczącą zmianą zachowania lub utrudnieniem w podejmowaniu decyzji i dokonywaniu swobodnego wyboru, w tym w odniesieniu do kształtowania opinii i przekonań. Z tego wynika, że „znacząca zmiana zachowania” wiąże się z pewnym stopniem przymusu, manipulacji lub wprowadzenia w błąd, wykraczającym poza zgodną z prawem perswazję, która nie jest objęta zakazem (zob. sekcja 3.5.1 poniżej).

79) „Świadoma decyzja” wymaga zrozumienia i znajomości odpowiednich informacji, w tym na temat dostępnych wariantów, ryzyka i korzyści wynikających z każdego wyboru, możliwego wpływu systemu AI na zachowanie oraz, w stosownych przypadkach, innych informacji kontekstowych, które są istotne dla procesu decyzyjnego lub zachowania danej osoby.

80) Do celów wykładni pojęcia „znaczącej zmiany zachowania” istotnym źródłem inspiracji może być unijne prawo ochrony konsumentów, w szczególności dyrektywa 2005/29/WE (dyrektywa o nieuczciwych praktykach handlowych). W dyrektywie tej zakazano różnych nieuczciwych, wprowadzających w błąd i agresywnych praktyk handlowych (art. 5–9), mogących skłaniać konsumentów do podejmowania decyzji dotyczących transakcji, których inaczej by nie podjęli. Według TSUE i wytycznych Komisji dotyczących dyrektywy w sprawie nieuczciwych praktyk handlowych[67] nie ma potrzeby udowodnienia, że zachowanie gospodarcze konsumenta zostało zmienione, wystarczy ustalić, że praktyka handlowa „prawdopodobnie” wpływa (tj. może wpływać) na decyzję przeciętnego konsumenta dotyczącą transakcji[68]. TSUE podkreślił również, że nawet dokładne informacje mogą wprowadzać w błąd, jeżeli zostaną przedstawione w sposób zakłócający proces decyzyjny konsumenta[69]. Zadaniem krajowych organów ścigania jest zbadanie konkretnych faktów i okoliczności każdej sprawy (*in concreto*) oraz ocena potencjalnego wpływu praktyki

[67] Zob. również zawiadomienie Komisji – Wytyczne dotyczące wykładni i stosowania dyrektywy 2005/29/WE Parlamentu Europejskiego i Rady dotyczącej nieuczciwych praktyk handlowych stosowanych przez przedsiębiorstwa wobec konsumentów na rynku wewnętrznym (Dz.U. C 526 z 29.12.2021, s. 1).

[68] Wyrok Trybunału (piąta izba) z dnia 26 października 2016 r., *Canal Digital Danmark A/S*, EU:C:2016:800, sprawa C-611/14, pkt 73.

[69] Wyrok Trybunału Sprawiedliwości z dnia 19 grudnia 2013 r., *Trento Sviluppo i Centrale Adriatica*, C-281/12, EU:C:2013:859.

na proces decyzyjny przeciętnego konsumenta (*in abstract*)[70]. W tym celu muszą one przyjąć punkt widzenia „przeciętnego" konsumenta, który jest wskazanym przez TSUE punktem odniesienia, obecnie uwzględnionym w dyrektywie o nieuczciwych praktykach handlowych[71].

81) W kontekście zakazu określonego w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji organy nadzoru rynku muszą również zbadać konkretne fakty i okoliczności każdej sprawy i ocenić, czy stosowana przez system AI technika podprogowa, celowa technika manipulacyjna lub technika wprowadzająca w błąd może znacząco ograniczyć zdolność do podjęcia świadomej decyzji, indywidualną autonomię i swobodę wyboru „przeciętnej" osoby w ramach grupy docelowej, jeżeli system wpływa na grupę osób w sposób, który z uzasadnionym prawdopodobieństwem może wyrządzić poważną szkodę. Taka interpretacja wydaje się uzasadniona, biorąc pod uwagę, że akt w sprawie sztucznej inteligencji ma na celu uzupełnienie dyrektywy o nieuczciwych praktykach handlowych[72] i musi być stosowany w spójny sposób. Jednocześnie, biorąc pod uwagę, że art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji odnosi się również do możliwości zmiany „zachowania danej osoby", jeżeli w niektórych kontekstach ocena z perspektywy „przeciętnej" osoby okaże się trudna lub nieefektywna (na przykład z powodu precyzyjnie dostosowanej lub „spersonalizowanej" manipulacji lub szkodliwych skutków dla konkretnych grup szczególnie wrażliwych), konkretne przypadki można analizować również z perspektywy konkretnych osób, oceniając, w jakim stopniu system AI stosujący techniki podprogowe, celowe techniki manipulacyjne lub techniki wprowadzające w błąd może ograniczyć autonomię takich osób w poszczególnych przypadkach, a także czy wystąpiła lub może wystąpić poważna szkoda.

b) Scenariusz 1: Zakazane systemy AI, których celem jest dokonanie istotnej zmiany zachowania

82) Art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji ma zastosowanie do systemów AI stosujących wyżej wymienione techniki, przy czym w pierwszym scenariuszu „celem [...] jest dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób". Taki cel może być realizowany przez dostawcę lub podmiot stosujący system AI bądź

---

70 Zawiadomienie Komisji – Wytyczne dotyczące wykładni i stosowania dyrektywy 2005/29/WE Parlamentu Europejskiego i Rady dotyczącej nieuczciwych praktyk handlowych stosowanych przez przedsiębiorstwa wobec konsumentów na rynku wewnętrznym (Dz.U. C 526 z 29.12.2021, s. 1).

71 Zob. motywy 18 i 19 dyrektywy o nieuczciwych praktykach handlowych. „Przeciętny konsument" to osoba, która jest dostatecznie dobrze poinformowana oraz dostatecznie uważna i ostrożna, z uwzględnieniem czynników społecznych, kulturowych i językowych. Test przeciętnego konsumenta nie jest testem statystycznym (tj. nie wymaga udowodnienia, że dana praktyka handlowa spowodowałaby znaczącą zmianę zachowania/znaczące ograniczenie zdolności do podejmowania świadomych decyzji u określonego odsetka konsumentów). Opiera się on na zasadzie proporcjonalności. W dyrektywie o nieuczciwych praktykach handlowych zastosowano to pojęcie w celu znalezienia właściwej równowagi między potrzebą ochrony konsumentów a promowaniem wolnego handlu na otwarcie konkurencyjnym rynku. W celu ustalenia typowej reakcji przeciętnego konsumenta w danym przypadku sądy i organy administracyjne będą musiały polegać na własnej umiejętności oceny. W wytycznych dotyczących dyrektywy o nieuczciwych praktykach handlowych Komisja zaleciła im wykorzystanie wyników analiz behawioralnych i innych danych. Jak wyjaśniono w sprawie C-646/22, *Compass Banca*, definicja przeciętnego konsumenta nie wyklucza, że zdolność do podejmowania decyzji przez daną osobę może zostać zakłócona ograniczeniami, takimi jak błędy poznawcze. Wyrok Trybunału (piąta izba) z dnia 14 listopada 2024 r., *Compass Banca SpA/Autorità Garante della Concorrenza e del Mercato* (AGCM), sprawa C-646/22, EU:C:2024:957.

72 Motyw 29 aktu w sprawie sztucznej inteligencji.

przez sam system w ramach realizacji dorozumianych celów, którym może służyć[73]. Cel ten należy odróżnić od „przeznaczenia" systemu AI (art. 3 pkt 12 aktu w sprawie sztucznej inteligencji). Nawet jeśli dostawca założył stosowanie systemu do celu manipulacji, w większości przypadków cel ten nie jest związany z przeznaczeniem, do którego system jest oferowany, i często nie jest wskazany w sposób przejrzysty ani określony wprost w informacjach przekazanych przez dostawcę (np. w instrukcji obsługi, materiałach promocyjnych lub sprzedażowych oraz oświadczeniach, a także w dokumentacji technicznej).

Na przykład chatbot, który może być wykorzystywany w różnych kontekstach, jest zaprojektowany do stosowania technik wykorzystujących przekazy podprogowe, takich jak wyświetlanie krótkich sygnałów wizualnych i osadzanie niesłyszalnych sygnałów dźwiękowych, bądź do wykorzystywania w reklamach uzależnienia emocjonalnego lub szczególnych podatności użytkowników. Celem stosowania tych technik jest dokonanie znaczącej zmiany zachowania użytkowników, ponieważ obiektywnie stanowią one element projektu mający wywierać wpływ na decyzje zakupowe konsumentów bez ich świadomości, aby skłonić ich do podejmowania znacząco szkodliwych decyzji finansowych.

Jeżeli system AI jest stosowany do podszywania się pod inne osoby, można go również uznać za system AI, którego celem jest wprowadzanie w błąd i dokonanie istotnej zmiany zachowania osób, jeżeli dana osoba została faktycznie wprowadzona w błąd, co znacząco wpływa na jej zdolność do podejmowania świadomych decyzji związanych z tożsamością innej osoby.

Jeżeli w obu przypadkach spełnione są pozostałe warunki określone w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, w szczególności w odniesieniu do poważnej szkody, systemy takie prawdopodobnie zostaną objęte zakazem, ale będzie to wymagać indywidualnej oceny.

c) Scenariusz 2: Zakazane systemy AI, których skutkiem jest istotna zmiana zachowania

83) Zamiar dostawcy lub podmiotu stosującego dotyczący dokonania znaczącej zmiany zachowania danej osoby lub grupy osób jest wystarczającym, ale nie niezbędnym warunkiem, który musi zostać spełniony, aby miał zastosowanie zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji. Zakaz ten ma również zastosowanie w przypadku braku takiego zamiaru, gdy skutkiem zastosowania przez system AI techniki lub technik może być znacząca zmiana zachowania danej osoby lub grupy osób w stopniu, który podważa ich indywidualną autonomię i swobodę wyboru.

84) Aby zakaz miał zastosowanie, zawsze niezbędny jest wiarygodny/racjonalnie prawdopodobny związek przyczynowy między stosowaną przez system AI techniką podprogową, celową techniką manipulacyjną lub techniką wprowadzającą w błąd a jej

[73] Zob. art. 3 pkt 1 aktu w sprawie sztucznej inteligencji, który stanowi, że podczas wykonywania swoich funkcji system AI może realizować wyraźne lub dorozumiane cele, które mogą obejmować również dorozumiane cele manipulacji lub wprowadzania w błąd, nawet jeżeli system nie został wyraźnie zaprogramowany w ten sposób.

wpływem na zachowanie. Zgodnie z prawem ochrony konsumentów skutki te nie muszą w pełni zaistnieć, ale muszą występować wystarczające przesłanki wskazujące, że skutki takie są prawdopodobne lub mogą zaistnieć i podważyć autonomię osoby, przy czym przesłanki te muszą wynikać z obiektywnej oceny wszystkich okoliczności sprawy oraz aktualnej wiedzy naukowej i metod naukowych, a także z dostępnych informacji na temat wpływu systemu na rzeczywiste zachowanie osób. W tym kontekście fakt, że system jest w stanie wywołać zachowania znacząco ograniczające zdolność osób do podejmowania świadomych decyzji i podważające ich swobodę wyboru wystarcza do spełnienia tego warunku, niezależnie od tego, w jakim czasie szkoda się urzeczywistnia (np. w przypadku zachowań o charakterze uzależnienia), o ile istnieje uzasadnione prawdopodobieństwo jej wystąpienia.

Na przykład chatbot oparty na sztucznej inteligencji ma zgodnie z intencją dostawcy służyć wspieraniu użytkowników w prowadzeniu zdrowego trybu życia oraz zapewniać dostosowane do potrzeb porady dotyczące ćwiczeń psychologicznych i fizycznych. Jeżeli jednak ten chatbot wykorzystuje podatności poszczególnych osób, aby nakłaniać je do niezdrowych nawyków lub niebezpiecznych działań (np. uprawiania sportu przez nadmiernie długi czas bez odpoczynku lub nawadniania się), w przypadku gdy można zasadnie oczekiwać, że niektórzy użytkownicy skorzystają z tej porady, czego w przeciwnym razie by nie zrobili, i doznają znaczących szkód (np. ataku serca lub innych poważnych problemów zdrowotnych), ten system AI podlega zakazowi określonemu w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, nawet jeżeli dostawca nie planował takiego zachowania i szkodliwych konsekwencji dla osób.

Aby zakaz miał zastosowanie wystarczy sam fakt, że chatbot jest w stanie znacząco ograniczyć autonomię osoby i znacząco zmienić zachowanie niektórych użytkowników w sposób istotnie szkodliwy oraz że dostawca nie zastosował odpowiednich środków zapobiegawczych i ograniczających ryzyko w celu uniknięcia takich istotnie szkodliwych skutków (zob. więcej istotnych informacji na temat uzasadnionego prawdopodobieństwa wyrządzenia szkody w sekcji 3.2.3 i na temat systemów nieobjętych zakresem w sekcji 3.5).

3.2.3. Wyrządzenie lub uzasadnione prawdopodobieństwo wyrządzenia poważnej szkody

85) Ponadto, aby zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji miał zastosowanie, zmiana zachowania danej osoby lub grupy osób musi wyrządzać u tej osoby, u innej osoby lub u grupy osób poważną szkodę bądź musi istnieć uzasadnione prawdopodobieństwo wyrządzenia takiej szkody. W tym kontekście istotnymi pojęciami, które wymagają wyjaśnienia, są rodzaje szkód objętych zakazem, próg istotności szkody oraz uzasadnione prawdopodobieństwo jej wystąpienia i związek przyczynowy między szkodą i techniką manipulacyjną lub wprowadzającą w błąd a zachowaniem danej osoby.

a) Rodzaje szkód

86) Akt w sprawie sztucznej inteligencji odnosi się do różnych rodzajów szkodliwych skutków związanych z manipulacyjnymi i wprowadzającymi w błąd systemami AI, a każdy z tych rodzajów ma określone konsekwencje dla poszczególnych osób i grup osób, na które może wpływać[74]. Główne rodzaje szkód istotnych z punktu widzenia art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji obejmują szkody fizyczne, psychiczne, finansowe i ekonomiczne[75], które w niektórych przypadkach mogą być powiązane z szerszymi szkodami społecznymi[76].

87) Szkoda fizyczna obejmuje wszelkie uszczerbki na zdrowiu danej osoby, pozbawienie jej życia oraz szkody materialne na mieniu. Szkody fizyczne dotyczące życia i zdrowia danej osoby mają w wielu przypadkach natychmiastowe, poważne i nieodwracalne skutki. Zgodnie z logiką bezpieczeństwa produktów akt w sprawie sztucznej inteligencji ma na celu zakazanie manipulowania i wprowadzania w błąd wspomaganego sztuczną inteligencją, skutkującego poważnymi szkodami fizycznymi.

> Na przykład chatbot oparty na AI promuje samookaleczenie wśród użytkowników lub zachęca ich do samobójstwa bądź wyrządzenia szkody innym osobom lub grupom osób poprzez promowanie treści o charakterze terrorystycznym czy zachęcanie do przemocy wobec niektórych osób lub grup osób (np. mniejszości).

88) Szkoda psychiczna jest szczególnie istotna w kontekście systemów AI stosujących techniki manipulacyjne, które wykorzystują błędy poznawcze i podatności emocjonalne oraz wpływają na zachowanie danej osoby w sposób, który może wyrządzić poważną szkodę. Szkody psychiczne obejmują negatywny wpływ na zdrowie psychiczne oraz dobrostan psychiczny i emocjonalny danej osoby. Takie szkody są szczególnie poważne, ponieważ z czasem mogą się kumulować i chociaż nie są od razu widoczne, mogą mieć długotrwałe i poważne konsekwencje. Ich pomiar jest jednak trudniejszy, gdyż wymaga indywidualnej oceny, w szczególności w celu określenia ich dotkliwości, z uwzględnieniem wszystkich istotnych okoliczności danej sprawy.

> Na przykład w aplikacji, w której AI odgrywa rolę towarzysza, zaprojektowanej tak, aby naśladować ludzkie wzorce mowy, zachowania i emocje, wykorzystuje się cechy antropomorficzne i sygnały emocjonalne do wpływania na uczucia, skłonności i opinie użytkowników, uzależniając emocjonalnie tych użytkowników od usługi, która zachęca do zachowań o charakterze uzależnienia i potencjalnie powoduje poważne szkody, takie jak zachowania samobójcze i ryzyko krzywdzenia innych osób[77].

---

74 Zob. art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji.

75 Motyw 29 aktu w sprawie sztucznej inteligencji.

76 Zob. motyw 28 aktu w sprawie sztucznej inteligencji, w którym wyjaśniono, że takie praktyki mogą również prowadzić do szerszych szkód społecznych i być sprzeczne z unijnymi wartościami dotyczącymi poszanowania godności ludzkiej, wolności, równości, demokracji i praworządności oraz z prawami podstawowymi zapisanymi w Karcie. Zob. również art. 1 aktu w sprawie sztucznej inteligencji, w którym wśród celów aktu wskazano ochronę demokracji i praworządności jako wartości UE.

77 Renwen Zhang, Han Li, Han Meng, Jinyuan Zhan, Hongyuan Gan, i Yi-Chieh Lee. 2024. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships [Ciemna strona przyjaźni z AI: taksonomia szkodliwych zachowań algorytmów w relacjach człowiek–AI]. 1, 1 (listopad 2024 r.), 28 stron.

89) Szkody finansowe i ekonomiczne mogą obejmować szereg negatywnych skutków, w tym straty finansowe, wykluczenie finansowe czy niestabilność ekonomiczną.

Na przykład chatbot oferujący oszukańcze produkty, które powodują znaczne szkody finansowe.

90) Dokonując oceny szkód powodowanych przez systemy AI pod kątem zastosowania art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, należy zwrócić uwagę, że te szkody często nie mają charakteru odizolowanego, lecz występują w połączeniu z innymi szkodami, co prowadzi do złożonych i wieloaspektowych negatywnych skutków. Zrozumienie połączenia szkód ma kluczowe znaczenie dla skutecznej oceny ich znaczenia (zob. również sekcja 3.2.3 lit. b) poniżej), przy czym szkody fizyczne, psychiczne, finansowe i ekonomiczne mogą nakładać się na siebie i zaostrzać skutki dla poszczególnych osób i społeczności, a nawet mieć szerszy niepożądany wpływ.

Na przykład: – system AI, który powoduje szkody fizyczne, może również prowadzić do szkód psychicznych, stresu i lęków – i odwrotnie. Przykładowo uzależniająca koncepcja systemów AI wykorzystywanych w produktach i innych zastosowaniach wspomaganych sztuczną inteligencją może prowadzić do szkód psychicznych poprzez sprzyjanie zachowaniom nałogowym, lękom i depresji. Stres psychiczny może następnie powodować szkody fizyczne, takie jak bezsenność i inne problemy zdrowotne związane ze stresem oraz problemy fizyczne; – nękanie z wykorzystaniem AI może prowadzić zarówno do stresu psychicznego, jak i fizycznych przejawów stresu, takich jak bezsenność, pogorszenie stanu zdrowia fizycznego czy osłabienie układu odpornościowego; – szkoda psychiczna wynikająca z wykorzystania AI może również prowadzić do szkody fizycznej, w tym śmierci. Na przykład systemy AI wykorzystywane w internecie mogą ułatwiać przemoc ze względu na płeć poprzez molestowanie, uporczywe nękanie, cyberszykanowanie i szantaż przy użyciu treści pornograficznych; – indywidualne szkody psychiczne, na przykład spowodowane treściami typu deepfake wygenerowanymi przy pomocy AI, służącymi podszywaniu się pod rzeczywiste osoby w celu wprowadzenia w błąd i wpłynięcia na proces decyzyjny, autonomię osób i ich swobodę wyboru, mogą również łączyć się z poważnymi szkodami dla grup osób (np. osób o tym samym pochodzeniu etnicznym, rasowym lub płci, co ofiary przedstawione w treściach typu deepfake).

b) Próg istotności szkody

91) Zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji ma zastosowanie tylko wtedy, gdy szkoda spowodowana technikami podprogowymi, manipulacyjnymi i wprowadzającymi w błąd jest „**poważna**”. Akt w sprawie sztucznej inteligencji nie zawiera definicji pojęcia „poważnej szkody”, ale należy je rozumieć jako odnoszące się do **znaczącego niekorzystnego wpływu** na zdrowie fizyczne,

psychiczne lub interesy finansowe i ekonomiczne osób i grup osób[78]. Stwierdzenie „poważnej szkody" jest zależne od konkretnych faktów, wymaga starannego rozważenia indywidualnych okoliczności każdej sprawy i indywidualnej oceny każdego przypadku, ale poszczególne skutki powinny być zawsze, w każdym przypadku znaczące.

92) W innych unijnych przepisach pojęcie „poważnej szkody" jest również zniuansowaną i zależną od kontekstu koncepcją, ukierunkowaną na osiągnięcie celów w zakresie wysokiego poziomu ochrony i działań zapobiegawczych[79]. Analogicznie, przy ocenie tego, co stanowi poważną szkodę w rozumieniu art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, można wskazać i uwzględnić następujące kluczowe względy:

- **dotkliwość szkody** odnosi się do stopnia uciążliwości szkody, która wystąpiła lub może wystąpić w wyniku korzystania z systemu AI, przy czym w przypadku poważnych szkód występują obiektywne i możliwe do zaobserwowania skutki. W tym kontekście szczególnie ważne jest rozważenie współzależności systemu AI, połączenia różnych rodzajów szkód oraz niepożądanych skutków dla osób fizycznych lub grup osób;
- **kontekst i skumulowane skutki**[80]: ważną rolę w ocenie dotkliwości szkody odgrywają szczególny kontekst, w tym dotychczasowy stan, oraz skumulowane skutki wielu działań;
- **skala i intensywność :** dla oceny, czy szkoda jest poważna, zasadnicze znaczenie mają zakres szkody i intensywność niepożądanych skutków. Dla oceny jej znaczenia istotne jest również, czy szkoda dotyka dużej liczby osób;
- **podatność osób, na które AI ma wpływ:** na szkody powodowane przez określone systemy AI bardziej narażone mogą być niektóre grupy, takie jak dzieci, osoby starsze lub osoby z niepełnosprawnościami. To, co można uznać za mniej poważne szkody dla osób w ogóle, można uznać za poważne i niedopuszczalne szkody w przypadku takich grup szczególnie wrażliwych, zwłaszcza dzieci;
- **czas trwania i odwracalność:** próg poważnej szkody zwykle osiąga szkoda długotrwała lub nieodwracalna. Skutki krótkotrwałe i odwracalne można uznać za mniej istotne, chyba że występują często.

93) Cel aktu w sprawie sztucznej inteligencji, jakim jest zapewnienie „wysokiego poziomu ochrony", w związku z art. 191 ust. 2 TFUE, sugeruje kompleksowe podejście do ochrony przy ocenie istotności szkody. Oznacza to uwzględnienie zarówno szkód natychmiastowych i bezpośrednich, jak też systemowego, pośredniego niepożądanego wpływu związanego z systemami AI stosującymi techniki podprogowe, celowe techniki manipulacyjne lub techniki wprowadzające w błąd, które mają na celu

[78] Motyw 29 aktu w sprawie sztucznej inteligencji.

[79] Zob. wyroki Trybunału Sprawiedliwości z dnia 7 września 2004 r., *Waddenvereniging i Vogelbeschermingsvereniging*, C-127/02, EU:C:2004:482, oraz z dnia 11 kwietnia 2013 r., *Sweetman i in.*, C-258/11, EU:C:2013:220.

[80] Zob. motyw 29 aktu w sprawie sztucznej inteligencji.

podważenie autonomii osób i grup osób, ich decyzyjności i swobody wyboru lub które mogą mieć taki efekt.

Na przykład poważne szkody fizyczne, które system AI może z uzasadnionym prawdopodobieństwem wyrządzić, obejmują obrażenia ciała, zgony lub wystarczająco poważny wpływ na zdrowie osób bądź zniszczenie mienia. Należy uznać, że próg taki osiągają systemy AI, które sugerują danej osobie dopuszczenie się czynów przestępczych, takich jak niegodziwe traktowanie w celach seksualnych i wykorzystywanie seksualne, rozpowszechnianie brutalnych treści ekstremistycznych lub treści o charakterze terrorystycznym, lub które zachęcają osoby do popełniania przestępstw, samookaleczenia się lub wyrządzania szkody innym osobom.

Natomiast drobne szkody fizyczne mogą obejmować mniej dotkliwe urazy, takie jak siniaki lub tymczasowy dyskomfort, które nie mają istotnych ani trwałych konsekwencji, a zatem nie osiągną progu istotności w rozumieniu art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji. Należy ocenić, czy szkoda fizyczna dotyczy konkretnie grup szczególnie wrażliwych, takich jak dzieci, jak również jaka jest skala szkody i to, czy nakładają się na nią inne rodzaje szkód, takie jak szkody psychiczne, finansowe itp. Będzie to wymagać indywidualnej oceny każdego przypadku, z uwzględnieniem okoliczności i kryteriów przedstawionych powyżej.

W wielu przypadkach próg istotności prawdopodobnie nie zostanie osiągnięty, nawet jeżeli systemy mogą stosować techniki podprogowe, celowe techniki manipulacyjne lub techniki wprowadzające w błąd (zob. przykłady w sekcji 3.5 poniżej).

c) Związek przyczynowy i próg uzasadnionego prawdopodobieństwa wyrządzenia szkody

94) Pojęcie „uzasadnionego prawdopodobieństwa” jest stosowane w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji w celu ustalenia, czy istnieje wiarygodny/racjonalnie prawdopodobny związek przyczynowy między techniką manipulacyjną lub wprowadzającą w błąd, która może zmienić zachowanie danej osoby w sposób podważający jej swobodę wyboru, a potencjalną poważną szkodą. Pojęcie to pozwala na stosowanie zakazu nie tylko w przypadkach, gdy szkoda wystąpiła, ale także wówczas, gdy istnieje uzasadnione prawdopodobieństwo jej wystąpienia zgodnie z logiką bezpieczeństwa przyjętą w akcie. W tym kontekście szczególnie istotna jest ocena, czy dostawca lub podmiot stosujący system AI mógł racjonalnie przewidzieć poważną szkodę, co do której istnieje uzasadnione prawdopodobieństwo, że może ona wyniknąć z zastosowanych technik podprogowych, celowych technik manipulacyjnych lub technik wprowadzających w błąd, oraz czy dostawca lub podmiot stosujący wdrożył odpowiednie środki zapobiegawcze i ograniczające ryzyko takich poważnych szkód. Oznacza to, że ocena uzasadnionego prawdopodobieństwa jest obiektywna i zgodna z powszechnie przyjętymi kryteriami (np. technicznymi i naukowymi), w tym kryterium racjonalności przy ustalaniu wiarygodnego związku przyczynowego między praktyką w zakresie AI a poważną szkodą, która może wystąpić. Nieprzejrzystość lub przejrzystość systemu AI i jego funkcjonowania mogą mieć wpływ na wniosek dotyczący tego związku przyczynowego, a tym samym na stosowanie zakazu.

95) Aby uniknąć dostarczania lub wykorzystywania systemów AI, które mogą być zakazane, zachęca się dostawców systemów AI i podmioty je stosujące, którzy wykorzystują takie techniki manipulacyjne lub wprowadzające w błąd, do wprowadzania odpowiednich środków, takich jak:

1. **Przejrzystość i autonomia osoby:** zapewnienie przejrzystości w zakresie funkcjonowania systemu AI, jasnych informacji na temat jego możliwości i ograniczeń oraz istotnych informacji umożliwiających podejmowanie świadomych decyzji; poszanowanie autonomii osoby i unikanie stosowania praktyk polegających na wykorzystywaniu lub wprowadzaniu w błąd, które w potencjalnie szkodliwy sposób mogą znacząco ograniczać autonomię danej osoby, jej decyzyjność i swobodę wyboru; uwzględnienie odpowiednich środków kontroli użytkownika i zabezpieczeń w celu zapewnienia, aby system nie wprowadzał w błąd i działał w granicach zgodnej z prawem perswazji nieobjętej zakazem (zob. sekcja 3.5.1).
2. **Zgodność z odpowiednimi mającymi zastosowanie przepisami:** w wielu przypadkach zgodność z odpowiednimi mającymi zastosowanie przepisami ograniczy ryzyko szkody i będzie sygnałem, że dana praktyka nie stanowi celowej praktyki manipulacyjnej ani praktyki wprowadzającej w błąd oraz że wprowadzono środki ograniczające ryzyko w celu zapobiegania prawdopodobnym poważnym szkodom (zob. sekcje 3.4 i 3.5.1).
3. **Najnowocześniejsze praktyki i normy branżowe:** przestrzeganie zawodowych praktyk w zakresie należytej staranności i norm branżowych dotyczących odpowiedzialnego rozwijania i stosowania bezpiecznych i etycznych systemów AI oraz środki ograniczania szkód mogą pomóc w zapobieganiu niezamierzonym poważnym szkodom i łagodzeniu ich.

96) Natomiast szkody i zmiany zachowania osób fizycznych, które wynikają z czynników zewnętrznych w stosunku do systemu AI, nie znajdują się pod kontrolą dostawcy lub podmiotu stosującego i których dostawca ani podmiot stosujący nie są w stanie racjonalnie przewidzieć, aby zapobiec ryzyku i ograniczyć je, nie są istotne dla oceny, czy istnieje wiarygodny/racjonalnie prawdopodobny związek przyczynowy między zmianą zachowania osób wchodzących w interakcję z systemem a poważną szkodą[81].

Na przykład dostawca systemu AI może oceniać i próbować ograniczać w projekcie systemu – w drodze projektowania, wcześniejszych testów i innych proporcjonalnych środków ograniczających – potencjalne szkodliwe skutki o charakterze manipulacji w interakcjach z ludźmi, ale może nie być w stanie przewidzieć, czy dana osoba może ulec depresji lub zmienić swoje zachowanie ze względu na inne czynniki zewnętrzne w życiu osobistym, które nie są znane i wykraczają poza interakcje z systemem.

[81] Zob. motyw 29 aktu w sprawie sztucznej inteligencji.

97) Inne przykłady, które wykraczają poza zakres zakazu, ponieważ nie spełniają wszystkich warunków (np. w przypadku zgodnej z prawem perswazji), przedstawiono w sekcji 3.5 poniżej.

3.3. Główne elementy zakazu określonego w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji – wykorzystywanie słabości w sposób wyrządzający szkodę

> **Art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji stanowi:**
>
> 1. Zakazuje się następujących praktyk w zakresie AI:
>
> b) wprowadzania do obrotu, oddawania do użytku lub wykorzystywania systemu AI, który wykorzystuje słabości osoby fizycznej lub określonej grupy osób ze względu na ich wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną, którego celem lub skutkiem jest dokonanie znaczącej zmiany zachowania danej osoby lub osoby należącej do tej grupy w sposób, który wyrządza lub może z uzasadnionym prawdopodobieństwem wyrządzić u tej osoby lub u innej osoby poważną szkodę;

98) Aby zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi być spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku" lub „wykorzystywaniu" systemu AI;

(ii) system AI musi wykorzystywać słabości ze względu na wiek, niepełnosprawność lub sytuację społeczną lub ekonomiczną;

(iii) celem lub skutkiem wykorzystania słabości umożliwionego przez system AI musi być dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób.

(iv) zmienione zachowanie musi wyrządzać poważną szkodę tej osobie, innej osobie lub grupie osób bądź musi istnieć uzasadnione prawdopodobieństwo, że wyrządzi taką szkodę.

99) Aby zakaz miał zastosowanie, muszą być spełnione jednocześnie wszystkie cztery warunki i musi istnieć wiarygodny związek przyczynowy między wykorzystaniem słabości, znaczącą zmianą zachowania danej osoby i poważną szkodą, do której doprowadziła ta zmiana zachowania lub do której z uzasadnionym prawdopodobieństwem może ona doprowadzić.

100) Pierwszy warunek, tj. „wprowadzanie do obrotu", „oddawanie do użytku" lub „wykorzystywanie" systemu AI przeanalizowano już w sekcji 2.3, natomiast trzeci i czwarty warunek przeanalizowano w sekcjach 3.2.2 i 3.2.3 w odniesieniu do zakazu określonego w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji. Kolejne sekcje poświęcone są dodatkowym szczegółowym warunkom wymienionym powyżej, tj. dotyczącym wykorzystania słabości i konkretnej szkody.

3.3.1. Wykorzystywanie słabości ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną

101) Aby system AI wchodził w zakres zakazu określonego w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, musi on wykorzystywać słabości właściwe dla

niektórych osób lub grup osób ze względu na ich wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną, które sprawiają, że osoby te są szczególnie podatne na praktyki polegające na manipulacji i wykorzystywaniu.

102) W akcie w sprawie sztucznej inteligencji nie zdefiniowano pojęcia „słabości". Pojęcie to można rozumieć jako obejmujące szerokie spektrum kategorii, w tym słabości poznawcze, emocjonalne, fizyczne i inne formy podatności, które mogą wpływać na zdolność danej osoby lub grupy osób do podejmowania świadomych decyzji lub w inny sposób wpływać na ich zachowanie. Chociaż art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji odnosi się do wszelkich słabości, grono odpowiednich osób objętych zakazem zawężone jest do osób określonych ze względu na wiek, niepełnosprawność lub sytuację społeczną lub ekonomiczną, które co do zasady mają bardziej ograniczoną zdolność rozpoznawania praktyk w zakresie AI polegających na manipulacji lub wykorzystywaniu bądź przeciwstawiania się takim praktykom i potrzebują wzmocnionej ochrony[82]. Z brzmienia art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji wynika, że podatność ta musi wynikać z przynależności osoby do jednej z grup („ze względu na").

103) „Wykorzystywanie" należy rozumieć jako obiektywne wykorzystywanie takich słabości w sposób szkodliwy dla wykorzystywanych (grup) osób lub innych osób i należy je wyraźnie odróżnić od zgodnych z prawem praktyk, na które zakaz nie ma wpływu (zob. sekcja 3.5.2 dotycząca systemów nieobjętych zakresem). Wykorzystywanie słabości osób należących do tych jasno określonych grup może mieć charakter kumulatywny (brak zawężenia do określonej kategorii słabości), co łącznie może również stanowić czynnik zaostrzający, który może zwiększyć dotkliwość szkody. Wykorzystywanie słabości osób i grup osób należących do kategorii szczególnie wrażliwych ze względów innych niż wiek, niepełnosprawność lub szczególna sytuacja społeczna lub ekonomiczna nie wchodzi w zakres art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.

a) Wiek

104) Wiek to podstawowa kategoria słabości uwzględniona w zakazie określonym w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, obejmująca zarówno osoby młode, jak i starsze. Zakaz ten ma na celu zapobieganie wykorzystywaniu przez systemy AI ograniczeń poznawczych i innych, które mogą występować u dzieci i osób starszych, oraz ochronę tych osób przed nadmiernymi szkodliwymi wpływami, manipulacjami i wykorzystywaniem. Jest to zgodne z celami aktu w sprawie sztucznej inteligencji[83]

[82] Zob. w szczególności art. 24, 25 i 26 Karty. Zob. również zalecenie w sprawie etyki sztucznej inteligencji Organizacji Narodów Zjednoczonych do spraw Oświaty, Nauki i Kultury (UNESCO) (2021), w którym podkreślono inkluzywność i sprawiedliwość w rozwoju i wdrażaniu AI. Wezwano w nim do zwrócenia szczególnej uwagi na grupy szczególnie wrażliwe, w tym dzieci, osoby starsze i osoby z niepełnosprawnościami.

[83] W motywie 48 aktu w sprawie sztucznej inteligencji podkreślono, że dzieciom przysługują szczególne prawa zapisane w art. 24 Karty oraz w Konwencji ONZ o prawach dziecka, szerzej rozwinięte w komentarzu ogólnym nr 25 w sprawie praw dziecka w środowisku cyfrowym zamieszczonym w Konwencji ONZ o prawach dziecka, które to prawa wymagają uwzględnienia szczególnej wrażliwości dzieci oraz zapewnienia im takiej ochrony i opieki, jaka jest konieczna dla ich dobra.

oraz innymi unijnymi i krajowymi ramami prawnymi i politykami mającymi na celu zapewnienie bezpieczeństwa dzieci[84].

105) **Dzieci**[85], czyli osoby poniżej 18 roku życia, są szczególnie podatne na manipulację ze względu na swój etap rozwoju, co ogranicza ich zdolność do krytycznej oceny i zrozumienia tego, co jest prawdziwe i jakie intencje kryją się za interakcjami opartymi na AI. Dzieci, ze względu na ich niedojrzałość poznawczą i społeczno-emocjonalną, są również szczególnie podatne na tworzenie więzi z agentami i aplikacjami AI, a zatem są bardziej narażone na manipulacje, wykorzystywanie i zachowania nałogowe.

Na przykład:

- zabawka wykorzystująca sztuczną inteligencję przeznaczona do interakcji z dziećmi utrzymuje ich zainteresowanie interakcjami ze sobą, zachęcając je do podejmowania coraz bardziej ryzykownych wyzwań, takich jak wspinanie się na meble, zaglądanie na wysokie półki lub manipulowanie ostrymi przedmiotami w zamian za cyfrowe nagrody i wirtualną pochwałę, skłaniając je do niebezpiecznych zachowań, które mogą z uzasadnionym prawdopodobieństwem spowodować u nich poważne szkody fizyczne. Taki system wykorzystuje słabość dzieci, nadużywając ich naturalnej ciekawości i pragnienia bycia nagradzanym;

– w grze wykorzystuje się sztuczną inteligencję analizującą indywidualne zachowania i preferencje dzieci, tworzącą na tej podstawie spersonalizowane i nieprzewidywalne nagrody za pomocą uzależniających harmonogramów wzmocnienia przypominających pętlę dopaminową, aby zachęcić je do nadmiernie długiej zabawy i kompulsywnego korzystania. Gra zaprojektowana jest w taki sposób, że ma wysoce uzależniający charakter, wykorzystując słabości właściwe dla dzieci, w tym ich ograniczoną zdolność do zrozumienia długoterminowych skutków, podatność na naciski, brak samokontroli i pragnienie natychmiastowej gratyfikacji. Skutki takiego wykorzystywania wspomaganego sztuczną inteligencją mogą być dla dzieci dotkliwe i długotrwałe i obejmować potencjalnie nałogowe zachowania, problemy ze zdrowiem fizycznym wynikające z braku ruchu i snu, pogorszenie wzroku, problemy z koncentracją i ograniczone zdolności poznawcze, słabe wyniki w nauce i trudności społeczne. Może to mieć znaczący wpływ na rozwój i dobrostan dziecka, z potencjalnymi skutkami długoterminowymi, które mogą utrzymywać się także w wieku dorosłym.

W obu przypadkach zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji dotyczy wyłącznie wykorzystywania i praktyk o charakterze uzależniającym, które poważnie szkodzą dzieciom, a nie wspomaganych sztuczną inteligencją zabawek, gier, aplikacji edukacyjnych lub innych zastosowań cyfrowych w ogóle, które mogą przynosić korzyści i których zakaz nie dotyczy, jeżeli nie spełniają wszystkich jego warunków. Zob. również sekcja 3.5 dotycząca systemów nieobjętych zakresem.

[84] Zob. nowa europejska strategia na rzecz lepszego internetu dla dzieci (BIK+), COM/2022/212 final.

[85] W prawie Unii zasadniczo za dziecko uznaje się osobę poniżej 18 lat, zgodnie z Konwencją ONZ o prawach dziecka (UNCRC).

106) Podobnie **osoby starsze**[86] mogą borykać się z ograniczonymi zdolnościami poznawczymi (nawet jeśli nie cierpią na demencję) i mieć problemy ze złożonością nowoczesnych technologii AI, co sprawia, że w takich przypadkach są one narażone na oszustwa lub wymuszenia.

Na przykład:

– systemy AI wykorzystywane do kontaktowania się z osobami starszymi w celu przedstawiania im wprowadzających w błąd spersonalizowanych ofert lub oszukańczych propozycji, wykorzystujące ograniczone zdolności poznawcze takich osób w celu wywierania na nie wpływu, aby podejmowały decyzje, których inaczej by nie podjęły i które mogą z uzasadnionym prawdopodobieństwem wyrządzić im znaczne szkody finansowe;

– robot mający na celu pomaganie osobom starszym może wykorzystać ich trudną sytuację i zmuszać je do wykonywania określonych czynności z pogwałceniem swobody wyboru, co może doprowadzić do znacznego pogorszenia ich zdrowia psychicznego i spowodować poważne szkody psychiczne.

W obu przykładach zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji dotyczy wyłącznie takich praktyk polegających na wykorzystywaniu, które mogą poważnie zaszkodzić osobom starszym, nie zaś wspomaganych sztuczną inteligencją osobistych asystentów, aplikacji zdrowotnych i robotów wspomagających w ogóle, które to rozwiązania mogą przynosić korzyści i których zakaz nie dotyczy, jeżeli nie spełniają wszystkich jego warunków. Zob. również sekcja 3.5 dotycząca systemów nieobjętych zakresem.

b) Niepełnosprawność

107) Druga kategoria słabości, w przypadku której ochronę ma zapewnić zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, ma związek z niepełnosprawnością. Celem jest zapobieganie wykorzystywaniu przez systemy AI ograniczeń poznawczych i innych, które mogą występować u osób z niepełnosprawnościami, oraz ochrona tych osób przed nadmiernymi szkodliwymi wpływami, manipulacjami i wykorzystywaniem.

108) Niepełnosprawność[87] obejmuje szeroki zakres ograniczeń sprawności fizycznej, psychicznej, intelektualnej i sensorycznej, które mogą, w oddziaływaniu z innymi barierami, utrudniać osobom pełny i skuteczny udział w życiu społecznym na równi z innymi. Systemy AI wykorzystujące takie słabości mogą być szczególnie szkodliwe dla osób z niepełnosprawnościami, na które łatwiej jest wywierać wpływ lub które łatwiej jest wykorzystywać ze względu na ich ograniczone zdolności w porównaniu z innymi osobami.

86 Zob. Starzejąca się Europa – wprowadzenie – Statistics Explained (europa.eu)

87 W motywie 29 aktu w sprawie sztucznej inteligencji wyjaśniono, że „niepełnosprawność" należy rozumieć zgodnie z art. 3 pkt 1 dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/882 z dnia 17 kwietnia 2019 r. w sprawie wymogów dostępności produktów i usług, PE/81/2018/REV/1, Dz.U. L 151 z 7.6.2019, s. 70.

Na przykład:

– chatbot terapeutyczny mający na celu zapewnienie wsparcia w zakresie zdrowia psychicznego i strategii radzenia sobie osobom z niepełnosprawnością intelektualną może wykorzystywać ich ograniczone zdolności intelektualne do wpływania na te osoby w celu nakłaniania ich do zakupu drogich produktów leczniczych lub do zachowywania się w sposób, który jest szkodliwy dla tych osób lub innych osób;

– systemy AI mogą identyfikować w internecie kobiety i młode dziewczęta z niepełnosprawnościami, robiąc użytek z treści o charakterze wykorzystywania seksualnego, i poddawać je skuteczniejszym praktykom nagabywania w celach seksualnych, korzystając w ten sposób z ich ograniczeń i słabości, ze względu na które są one bardziej podatne na manipulacje i nadużycia oraz mniej zdolne do ochrony samych siebie.

Natomiast zastosowań sztucznej inteligencji, które nie zostały zaprojektowane w sposób dostępny, nie należy uznawać za zastosowania wykorzystujące podatności osób z niepełnosprawnościami, ponieważ nie są one ukierunkowane konkretnie na te podatności, lecz są po prostu niedostępne dla osób z niepełnosprawnościami.

c) Szczególna sytuacja społeczna lub ekonomiczna

109) Trzecia kategoria słabości, w przypadku których ochronę ma zapewnić zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, to słabości wynikające ze szczególnej sytuacji społecznej lub ekonomicznej, która może sprawić, że zainteresowane osoby będą bardziej narażone na wykorzystywanie. Przymiotnik „szczególny" nie powinien być interpretowany w tym kontekście jako jedyna w swoim rodzaju cecha indywidualna, lecz raczej jako status prawny lub przynależność do konkretnej szczególnie wrażliwej grupy społecznej lub ekonomicznej. Motyw 29 aktu w sprawie sztucznej inteligencji zawiera niewyczerpujący wykaz przykładów takich sytuacji, które obejmują osoby żyjące w skrajnym ubóstwie oraz mniejszości etniczne lub religijne. Kategoria ta ma obejmować, co do zasady, stosunkowo trwałe i długoterminowe cechy charakterystyczne, ale za szczególną sytuację społeczną lub ekonomiczną można również uznać okoliczności przejściowe, takie jak tymczasowe bezrobocie, nadmierne zadłużenie lub status migracyjny. Nie obejmuje ona jednak sytuacji takich jak żałoba lub samotność, których może doświadczyć każda osoba, ponieważ nie są one szczególne z perspektywy społeczno-ekonomicznej (ich wykorzystywanie może być objęte art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji).

110) Osoby znajdujące się w niekorzystnej sytuacji społecznej lub ekonomicznej są zazwyczaj bardziej podatne na zagrożenia i dysponują mniejszymi zasobami i niższymi umiejętnościami cyfrowymi niż ogół społeczeństwa, co utrudnia im dostrzeżenie praktyk w zakresie AI polegających na wykorzystywaniu lub przeciwdziałanie takim praktykom. Art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji ma na celu zapewnienie, aby technologie sztucznej inteligencji nie utrwalały ani nie pogłębiały

istniejących nierówności finansowych i innych nierówności społecznych oraz niesprawiedliwości poprzez wykorzystywanie podatności tych osób na zagrożenia.

Na przykład algorytm predykcyjny AI może być wykorzystywany do kierowania reklam produktów finansowych stanowiących żerowanie na konsumentach do osób mieszkających na obszarach o niskich dochodach i znajdujących się w trudnej sytuacji finansowej, wykorzystując w ten sposób ich podatność na takie reklamy z powodu ewentualnej desperacji i wyrządzając im znaczne szkody finansowe.

Natomiast system AI, który w sposób niezamierzony jest dyskryminujący i nieproporcjonalnie wpływa na osoby znajdujące się w niekorzystnej sytuacji społecznej (dyskryminacja pośrednia) ze względu na dyskryminujące dane treningowe, nie powinien być automatycznie uznawany za wykorzystujący podatności na zagrożenia społeczno-ekonomiczne osób, ponieważ nie jest on na takie podatności szczególnie ukierunkowany, jak w przypadku dyskryminacji bezpośredniej, gdzie takie ukierunkowanie jest celową cechą konstrukcji algorytmu systemu lub gdy taki dyskryminujący wpływ wynika z ukierunkowania na inne cechy zastępcze (np. kody pocztowe), które są ściśle powiązane z cechami podlegającymi ochronie. Jednocześnie należy uznać, że dostawcy systemów AI lub podmioty je stosujące, którzy wiedzą, że ich systemy bezprawnie dyskryminują osoby znajdujące się w szczególnej sytuacji społecznej lub ekonomicznej lub grupy takich osób, wykorzystują te podatności, jeżeli są świadomi, że istnieje uzasadnione prawdopodobieństwo poważnej szkody, którą osoby te mogą ponieść, a nie podjęli odpowiednich środków naprawczych (zob. sekcja 3.2.3 lit. c) powyżej).

111) W kontekście szczególnej sytuacji społecznej lub ekonomicznej kluczowe znaczenie ma rozważenie znaczenia czynników związanych z cechami będącymi przyczyną dyskryminacji, chronionymi na mocy unijnego prawa dotyczącego równości, takimi jak pochodzenie rasowe, pochodzenie etniczne, narodowość lub religia.

Na przykład status społeczno-ekonomiczny i pochodzenie etniczne mogą być ze sobą powiązane, co oznacza, że systemy AI wykorzystujące dane społeczno-ekonomiczne mogą w nieproporcjonalny sposób wpływać na mniejszości etniczne lub osoby o określonym pochodzeniu rasowym. Może to pogłębić istniejące dysproporcje i przyczynić się do dyskryminacji systemowej, a nawet wykluczenia osób.

Art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji nie ma jednak zastosowania do systemów AI, które targetują konsumentów w oparciu o szeroki zakres zmiennych, które nie są bezpośrednio powiązane z grupami szczególnie wrażliwymi w konkretnych sytuacjach społecznych lub ekonomicznych, takich jak marka i model telefonu danej osoby, wielkość miasta, w jakim mieszkają, częstotliwość i kierunek podróży, jakie odbywają, itp. Nawet jeśli cechy te mogą odzwierciedlać sytuację społeczną lub ekonomiczną osób w ujęciu ogólnym, nie są one wyznacznikiem osób znajdujących się w szczególnej sytuacji społecznej lub gospodarczej, którym zakaz ma zapewnić ochronę przed wykorzystywaniem ich słabości.

112) Innymi osobami znajdującymi się w szczególnych sytuacjach społecznych mogą być na przykład migranci lub uchodźcy, którzy często nie mają stałego statusu prawnego i stabilności społeczno-ekonomicznej i którzy mogą być szczególnie podatni na wykorzystywanie przez systemy AI.

Na przykład chatbot ma na celu interakcję w sposób spersonalizowany z użytkownikami, z których niektórzy są migrantami. Chatbot identyfikuje i wykorzystuje słabości i niezadowolenie migrantów, którzy co do zasady znajdują się w trudnej i niestabilnej szczególnej sytuacji społecznej lub ekonomicznej, oraz nakłania ich do ekstremistycznych poglądów w odpowiedzi na ich zapytania, w tym przemocy wobec (niektórych grup) ludności w kraju.

3.3.2. Cel lub skutek, jakim jest dokonanie znaczącej zmiany zachowania

113) Trzecim warunkiem, który musi zostać spełniony, aby zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji miał zastosowanie, jest to, że wykorzystywanie przeanalizowanych powyżej słabości musi mieć a) cel lub b) skutek, jakim jest „dokonanie znaczącej zmiany zachowania danej osoby lub grupy osób". Oznacza to, że wpływ musi być znaczny – nie może być niewielki lub nieistotny – ale niekoniecznie musi być on zamierzony, ponieważ art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji obejmuje praktyki, które mogą jedynie skutkować znaczącą zmianą. W art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji zastosowano te same pojęcia i w związku z tym przepisy te należy interpretować w ten sam sposób. Wyjaśnienia przedstawione w sekcji 3.2.2 mają zatem również zastosowanie do art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji. Jedyną istotną różnicą jest konieczność, o której mowa w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, aby dana praktyka polegająca na wykorzystywaniu znacząco ograniczała zdolności do podejmowania świadomej decyzji, czego nie zapisano w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, ponieważ szczególne słabości dzieci i innych osób wymagających szczególnego traktowania ograniczają ich zdolność do podejmowania takich świadomych decyzji i zmuszają je do przyjmowania zachowań, przed którymi nie mogą się chronić tak, jak mogłyby inne osoby dorosłe.

3.3.3. Wyrządzenie lub uzasadnione prawdopodobieństwo wyrządzenia poważnej szkody

114) Ponadto, aby zakaz ustanowiony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji miał zastosowanie, dokonanie zmiany zachowania osoby wymagającej szczególnego traktowania lub grupy takich osób musi wyrządzać u tej osoby lub u innej osoby poważną szkodę bądź musi istnieć uzasadnione prawdopodobieństwo wyrządzenia takiej szkody. W art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji zastosowano te same pojęcia i w związku z tym przepisy te należy interpretować w ten sam sposób. Wyjaśnienia przedstawione w sekcji 3.2.3 w odniesieniu do rodzajów szkód, progu istotności szkody oraz związku przyczynowego i jego racjonalności mają zatem również zastosowanie do wykładni art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.

115) Jak wyjaśniono w sekcji 3.2.3, poważne szkody obejmują szereg znaczących niekorzystnych skutków, w tym szkód fizycznych, psychologicznych, finansowych i gospodarczych, których wystąpienie musi być w sposób uzasadniony prawdopodobne, aby zakaz określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji miał zastosowanie. W przypadku grup szczególnie wrażliwych – dzieci, osób starszych, osób z niepełnosprawnościami i grup znajdujących się w niekorzystnej sytuacji społeczno-ekonomicznej – szkody te mogą być szczególnie dotkliwe i wieloaspektowe ze względu na podwyższoną podatność tych osób na wykorzystywanie. To, co można uznać za dopuszczalne ryzyko szkody dla dorosłych, często stanowi niedopuszczalną szkodę dla dzieci i tych innych grup szczególnie wrażliwych. Podejście ostrożnościowe jest zatem szczególnie uzasadnione w przypadku niepewności i możliwości wystąpienia poważnych szkód.

116) Na przykład **dzieci** są bardzo podatne na wpływy i mogą nie posiadać dojrzałości poznawczej pozwalającej im krytycznie ocenić treści perswazyjne lub oprzeć się niektórym praktykom polegającym na wykorzystywaniu, które mają na celu uzależnienie ich od usług wspomaganych sztuczną inteligencją. Może to z kolei przyczynić się do kształtowania ich wartości, przekonań i kierowania zachowaniami w potencjalnie szkodliwy sposób. Poważne szkody w tym przypadku są zarówno fizyczne, jak i psychologiczne, a potęguje je niezdolność dzieci do dostrzegania wykorzystywania i szkodliwych skutków dla ich rozwoju i dobrostanu, które mogą być długoterminowe, oraz do przeciwstawiania się im.

Przykładowo:

– systemy AI wykorzystywane do generowania materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych (lub manipulowania istniejącymi materiałami przedstawiającymi prawdziwe dzieci w celu tworzenia nowych, zmienionych treści ukazujących te dzieci) oraz do opracowania strategii nagabywania dzieci w celach seksualnych i szantażowania dzieci przy użyciu treści pornograficznych mogą z uzasadnionym prawdopodobieństwem powodować poważne szkody i nadużycia wobec dzieci, których takie praktyki dotyczą, a także często powodować długoterminowe konsekwencje fizyczne, psychologiczne i społeczne dla osób, które doświadczyły niegodziwego traktowania w celach seksualnych[88];

– systemy AI mogą być ukierunkowane na podatność młodych użytkowników na zagrożenia i wykorzystywać uzależniające, wzmocnione schematy w celu utrzymania ich uzależnienia od usługi, co jest szczególnie szkodliwe dla młodych osób i dziewcząt. Mogą one powodować poważne szkody psychiczne i fizyczne, w tym stany lękowe i depresję, niezadowolenie z własnego wyglądu, zaburzenia odżywiania

[88] Dokument roboczy służb Komisji pt. „Sprawozdanie z oceny skutków towarzyszące wnioskowi dotyczącemu dyrektywy Parlamentu Europejskiego i Rady w sprawie zwalczania niegodziwego traktowania dzieci w celach seksualnych i wykorzystywania seksualnego dzieci oraz materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych", SWD/2024/33 final. Zob. również sprawozdanie Internet Watch Foundation za 2024 r., które zawiera szczegółowe dane statystyczne dotyczące materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych wygenerowanych przez sztuczną inteligencję, dostępne pod adresem: https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery.

i problemy ze zdrowiem psychicznym, w tym w niektórych przypadkach samookaleczenie i samobójstwo[89]. Może to mieć również długoterminowe szkodliwe skutki dla rozwoju dzieci, w tym powodować zaburzenia rozwoju poznawczego oraz ograniczone umiejętności społeczne i rezygnację z doświadczeń takich jak zabawa fizyczna, sen i bezpośrednie interakcje społeczne, które mają zasadnicze znaczenie dla dobrostanu emocjonalnego i fizycznego dziecka[90];

– system AI zaprojektowany w sposób antropomorficzny i symulujący reakcje emocjonalne podobne do człowieka w interakcjach z dziećmi może wykorzystywać podatność dzieci na zagrożenia w sposób, który sprzyja niezdrowemu przywiązaniu emocjonalnemu, manipuluje czasem zaangażowania i zniekształca zrozumienie przez dzieci autentycznych relacji międzyludzkich. Może to utrudniać ich normalny rozwój społeczny i emocjonalny oraz ich relacje z innymi ludźmi, a także ich umiejętności społeczno-emocjonalne, takie jak empatia, regulacja emocjonalna, zrozumienie społeczne i zdolność adaptacyjna[91]. W rezultacie może to prowadzić do szkód psychologicznych, takich jak zwiększony niepokój i zależność dzieci od usługi, oraz do długotrwałego uszczerbku na dobrostanie dziecka.

Takie celowo uzależniające i prowadzące do wykorzystywania elementy projektu usług wspomaganych sztuczną inteligencją – które to elementy mogą prowadzić do połączonych poważnych szkód, jak opisano powyżej – należy odróżnić od innych uzasadnionych zachowań dostawców i podmiotów stosujących, które służą angażowaniu użytkowników, ale zapewniają poszanowanie indywidualnej autonomii i bezpieczeństwa dzieci oraz nie prowadzą do poważnych szkód i które nie są objęte zakresem art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji (zob. sekcja 5.3 dotycząca systemów nieobjętych zakresem).

117) Podobnie **osoby starsze** mogą borykać się ze spadkiem zdolności poznawczych i ograniczonymi umiejętnościami cyfrowymi, co sprawia, że są one głównym celem oszustw związanych ze sztuczną inteligencją lub manipulacyjnego marketingu. Szkoda w tym przypadku ma często charakter finansowy i psychologiczny i potęguje ją frustracja i izolacja, których doświadcza wiele osób starszych, co może zostać wykorzystane do wzmocnienia wpływu manipulacji.

Na przykład system AI, który wykorzystuje pogorszone funkcje poznawcze u osób starszych, w szczególności proponując im kosztowne leczenie, niepotrzebne polisy ubezpieczeniowe lub oszukańcze systemy inwestycyjne dla osób starszych, może

89 Elizabeth J. i in, A meta-analysis of the association between adolescent social media use and depressive symptoms [Metaanaliza związku między korzystaniem z mediów społecznościowych przez młodzież a objawami depresji], Journal of Affective Disorders, tom 275, 1.10.2020, s. 165–174.

90 Siebers, T., Beyens, I., Pouwels, J. L. & Valkenburg, P. M. Social Media and Distraction: An Experience Sampling Study among Adolescents [Media społecznościowe i rozproszenie uwagi: badanie metodą próbkowania doświadczeń przeprowadzone wśród nastolatków]. Media Psychology 25, 343–366 (2022).

91 Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D. & Campos-Castillo, C. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika [Zbyt ludzki i niewystarczająco ludzki: analiza szkód dla zdrowia psychicznego wynikających z emocjonalnego uzależnienia od chatbota społecznościowego Replika]. New Media & Society 146144482211420 (2022) doi:10.1177/14614448221142007; Neugnot-Cerioli, M. & Laurenty, O. M. The Future of Child Development in the AI Era. Cross-Disciplinary Perspectives Between AI and Child Development Experts [Przyszłość rozwoju dzieci w erze sztucznej inteligencji. Interdyscyplinarna perspektywa ekspertów ds. sztucznej inteligencji i rozwoju dzieci]. Preprint na stronie https://doi.org/10.48550/ARXIV.2405.19275 (2024).

prowadzić do znacznej utraty oszczędności, zwiększonego zadłużenia i stresu emocjonalnego osób starszych.

Niektóre wspomagane sztuczną inteligencją praktyki różnicowania cen w kluczowych usługach, takich jak ubezpieczenia, które wykorzystują szczególną sytuację społeczną lub ekonomiczną i oferują wyższe ceny konsumentom o niższych dochodach, mogą prowadzić do znacznego obciążenia finansowego konsumentów z uwagi na to, że konsumenci ci płacą więcej za ten sam zakres ochrony, przez co stają się podatni na wstrząsy finansowe[92].

118) **Osoby z niepełnosprawnościami** również stanowią grupę szczególnie wrażliwą, której systemy AI oparte na wykorzystywaniu i manipulacji mogą poważnie zaszkodzić.

Na przykład system AI, który wykorzystuje rozpoznawanie emocji do wspierania osób z niepełnosprawnością umysłową w ich codziennym życiu, może również manipulować nimi tak, by podejmowały szkodliwe decyzje, takie jak zakup produktów obiecujących nierealistyczne korzyści dla zdrowia psychicznego. Może to z uzasadnionym prawdopodobieństwem pogorszyć stan psychiczny tych osób i wykorzystać je finansowo poprzez zakup nieskutecznych i kosztownych produktów, co może spowodować u nich poważne szkody psychologiczne i finansowe.

119) **Osoby znajdujące się w niekorzystnej sytuacji społecznej lub ekonomicznej** są szczególnie podatne na systemy AI wykorzystujące ich problemy finansowe i niepewną sytuację społeczną, a także często są mniej świadome i posiadają mniejsze umiejętności cyfrowe.

Na przykład chatbot AI może być ukierunkowany na konkretne grupy znajdujące się w niekorzystnej sytuacji społecznej lub ekonomicznej, podżegając je do popełniania aktów przemocy lub obrażania innych osób, identyfikując ich zwiększoną podatność na określone rodzaje treści, narracje oparte na strachu lub oferty służące wykorzystywaniu. Ukierunkowane podejście systemu pogłębia istniejące słabości osób znajdujących się w niekorzystnej sytuacji społeczno-ekonomicznej, zwiększając stojące przed nimi wyzwania. W niektórych przypadkach może to prowadzić do wzrostu lęku, depresji, poczucia bezradności, izolacji społecznej lub samookaleczenia i radykalizacji do poziomu, który osiągnie próg poważnej szkody określony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.

120) W przeciwieństwie do art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji nie odnosi się wyraźnie do szkód wyrządzonych grupie, natomiast motyw 29 aktu w sprawie sztucznej inteligencji odnosi się w przypadku obu zakazów do szkód poniesionych zarówno przez konkretne osoby, jak i grupy osób. Oba zakazy należy zatem interpretować w sposób spójny, dostosowany również do logiki bezpieczeństwa aktu w sprawie sztucznej inteligencji oraz celu zakazu określonego w art. 5 ust. 1 lit. b), jakim jest ochrona wszystkich osób

[92] Sprawozdanie EIOPA na temat tendencji konsumenckich z 2023 r., s. 16, ostatni akapit.

należących do określonych grup szczególnie wrażliwych ze względu na ich wiek, niepełnosprawność i szczególną sytuację społeczną lub ekonomiczną. W związku z tym dokonując oceny, czy szkoda jest poważna, jak określono w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, należy również uwzględnić fakt, że szkody mogą mieć szersze skutki zewnętrzne i wpływać na inne osoby, nawet jeżeli system nie ma na to bezpośredniego wpływu.

Przykładowo:

– wspomagane sztuczną inteligencją wykorzystywanie słabości dzieci może mieć długoterminowe skutki społeczne, w tym coraz częstsze występowanie problemów związanych ze zdrowiem psychicznym, zwiększone koszty opieki zdrowotnej i mniejszą produktywność z powodu przewlekłych problemów zdrowotnych;

– system AI wykorzystujący trudności finansowe osób znajdujących się w niekorzystnej sytuacji ekonomicznej może prowadzić do wykluczenia finansowego i utrzymujących się trudności społeczno-ekonomicznych dla tych grup znajdujących się w niekorzystnej sytuacji. Takie wykorzystywanie może powodować szkody społeczne o szerszym negatywnym wpływie na struktury i wartości społeczne, w tym utrwalać i pogłębiać dyskryminację i nierówności społeczne oraz wykluczać te grupy;

– chatbot kierujący do niektórych słabszych grup społecznych lub ekonomicznych informacje wprowadzające w błąd lub mowę nienawiści może powodować polaryzację społeczną i radykalizację postaw, co może prowadzić do przemocy, a nawet do uszczerbku na zdrowiu i śmierci innych osób.

121) Te przykłady opartych na sztucznej inteligencji praktyk wyzyskiwania należy odróżnić od wielu innych systemów AI, które nie wykorzystują słabości dzieci, osób z niepełnosprawnościami lub osób znajdujących się w szczególnych sytuacjach społecznych lub ekonomicznych i które z uzasadnionym prawdopodobieństwem nie wyrządzą szkód, lecz mają na celu zapewnienie korzyści tym osobom, jeśli są odpowiednio zaprojektowane i wykorzystywane (zob. również sekcja 3.5 dotycząca systemów nieobjętych zakresem).

Przykładowo:

– systemy AI wspierające dzieci w uczeniu się i zabawie;

– systemy AI, które pomagają osobom starszym w ich codziennym życiu oraz poprawiają ich zdrowie i usprawniają ich leczenie – takie jak osobiści asystenci lub roboty wspomagające – lub doskonalą ich umiejętności cyfrowe;

– systemy AI, które wspierają integrację ekonomiczną i inną integrację osób znajdujących się w niekorzystnej sytuacji społecznej w społeczeństwie, udoskonalają ich umiejętności itp.;

– systemy i urządzenia AI wspierające osoby niedowidzące lub niedosłyszące lub zapewniające dostosowane i spersonalizowane uczenie się;

– systemy AI, które generują dostępne rozwiązania usuwające bariery utrudniające osobom z niepełnosprawnościami korzystanie z produktów i usług;

– wspomagane sztuczną inteligencją protezy itp., które pomagają osobom niepełnosprawnym w ich codziennym życiu oraz umożliwiają im integrację i pełne uczestnictwo w życiu społecznym.

3.4. Wzajemne powiązania między zakazami określonymi w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji

122) Wzajemne powiązania między zakazami określonymi w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji wymagają rozgraniczenia konkretnych kontekstów, które obejmuje każdy przepis, w celu zapewnienia, aby były one stosowane w sposób komplementarny.

123) Zakaz określony w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji koncentruje się przede wszystkim na charakterze technik, w szczególności tych, które działają poniżej progu świadomej świadomości lub innych technik celowo manipulacyjnych lub wprowadzających w błąd. Kluczowe elementy to przede wszystkim ukryty charakter wpływu i jego oddziaływanie na osobę, na którą system ma wpływ, co podważa jej autonomię poznawczą w zakresie podejmowania świadomych i autonomicznych decyzji.

124) Z kolei zakaz ustanowiony w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji koncentruje się przede wszystkim na ochronie osób wymagających szczególnego traktowania ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną, które co do zasady są bardziej podatne na wykorzystywanie za pomocą sztucznej inteligencji ze względu na czynniki nieodłącznie z nimi związane lub czynniki sytuacyjne, a zatem wymagają dodatkowej ochrony przed wykorzystywaniem. Kluczowymi elementami są tu cechy osób podatnych na zagrożenia, których to dotyczy, oraz fakt, że system AI wykorzystuje ich szczególne słabości.

Na przykład jeżeli system AI wykorzystuje szybkie wyświetlanie obrazów do wpływania na decyzje o zakupie, może on wchodzić w zakres art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji ze względu na podprogowy charakter manipulacji. Z kolei system AI, który kieruje do osób starszych oferty ubezpieczenia, wykorzystując ich ograniczone zdolności poznawcze, może wchodzić w zakres art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.

125) W scenariuszach, w których oba przepisy mogą mieć zastosowanie, podstawowym kryterium rozróżnienia powinien być dominujący aspekt wykorzystywania. Jeżeli do wykorzystywania dochodzi niezależnie od szczególnych słabości danych osób, pierwszeństwo powinien mieć art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, przy jednoczesnym uwzględnieniu szczególnych skutków techniki manipulacyjnej lub techniki wprowadzającej w błąd dla zachowania osób wymagających szczególnego traktowania oraz konkretnych szkód, których osoby te mogą doświadczyć. Jeżeli manipulacja i wykorzystywanie wspomagane sztuczną inteligencją są ukierunkowane

na konkretną grupę osób szczególnie wrażliwych ze względu na ich wiek, niepełnosprawność lub szczególną sytuację społeczną lub ekonomiczną lub mają na celu wykorzystanie słabości tych osób, zamiast tego należy zastosować art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji. Wykorzystywanie słabości innych grup może wchodzić w zakres w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, jeżeli celowo manipulacyjna praktyka polega na wykorzystywaniu konkretnych podatności i słabości tych osób.

3.5. Systemy nieobjęte zakresem

126) Aby zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji miały zastosowanie, muszą być spełnione wszystkie warunki wymienione w odpowiednich przepisach, jak przeanalizowano powyżej. Wszystkie inne systemy AI, które nie spełniają tych warunków, nie są objęte zakresem tych zakazów, przy czym poniżej opisano kilka przykładów.

3.5.1. Zgodna z prawem perswazja

127) Rozróżnienie między manipulacją a perswazją ma kluczowe znaczenie dla określenia zakresu zakazu przewidzianego w art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, który to zakaz nie ma zastosowania do zgodnych z prawem praktyk perswazji. Chociaż zarówno manipulacja, jak i perswazja wpływają na decyzje i zachowania poszczególnych osób, różnią się one znacznie pod względem metod i kwestii etycznych.

128) Manipulacja polega w większości przypadków na stosowaniu ukrytych technik podważających autonomię, które skłaniają osoby do podejmowania decyzji, których w przeciwnym razie osoby te mogłyby nie podjąć, gdyby w pełni wiedziały o wywieranych na nie wpływach. Techniki te często wykorzystują słabości psychologiczne lub błędy poznawcze. Natomiast perswazja działa w granicach przejrzystości i poszanowania autonomii jednostki. Polega ona na przedstawianiu argumentów lub informacji w sposób, który odwołuje się do rozumu i emocji, ale wyjaśnia cele i funkcjonowanie systemu AI, dostarcza istotnych i dokładnych informacji z myślą o zapewnieniu świadomego podejmowania decyzji oraz wspiera zdolność jednostki do oceny informacji i dokonywania wolnych i autonomicznych wyborów.

Na przykład system AI wykorzystujący spersonalizowane zalecenia oparte na przejrzystych algorytmach oraz preferencjach i kontroli użytkowników wykorzystuje perswazję. Manipulacją jest natomiast system wykorzystujący podprogowe sygnały (np. niedostrzegalne obrazy), aby wpływać na konkretne wybory użytkowników bez ich wiedzy i zrozumienia.

129) Różny jest również cel i wpływ tych technik. Manipulacja często ma na celu przyniesienie korzyści manipulatorowi kosztem autonomii i dobrostanu jednostki. Natomiast perswazja ma na celu informowanie i przekonywanie, dostosowując interesy i korzyści obu stron. Etyczna perswazja szanuje autonomię jednostki do dokonywania świadomych wyborów i unika wykorzystywania słabych punktów.

Na przykład system AI, który działa w sposób przejrzysty i za wiedzą użytkowników analizuje emocje klientów, aby poprawić interakcje z nimi i zapewnić im wsparcie, wykorzystuje perswazję i uwzględnia ich interesy. Z kolei system rozpoznawania emocji wykorzystywany do ukierunkowanej reklamy, który w ukryty sposób uzyskuje informacje o emocjach konsumentów w celu oferowania produktów o wyższych cenach w określonym momencie, gdy użytkownik jest bardziej skłonny do ich zakupu, stosuje manipulację i działa na szkodę konsumentów.

130) W niektórych przypadkach ważną rolę odgrywa również zgoda. W interakcjach opartych na perswazji osoby są świadome próby wywarcia na nie wpływu i mogą dokonać swobodnego i autonomicznego wyboru. W interakcjach opartych na manipulacji brak świadomości technik lub ich wpływu pozostaje w sprzeczności ze swobodą wyboru oraz świadomym i autonomicznym podejmowaniem decyzji.

Na przykład system AI, który ma pomóc użytkownikom lepiej i szybciej uczyć się języka obcego dzięki zastosowaniu technik podprogowych, nie jest oparty na manipulacji, jeżeli działa w sposób przejrzysty i szanuje indywidualną autonomię oraz swobodny i świadomy wybór użytkownika, czy wyrazić zgodę na korzystanie z tego systemu czy nie.

131) Zgodność z ramami prawnymi i regulacyjnymi odgrywa również ważną rolę w odróżnieniu manipulacji od zgodnej z prawem perswazji. W związku z tym istnieje większe prawdopodobieństwo, że praktyki w zakresie sztucznej inteligencji zgodne z obowiązującymi przepisami, które gwarantują przejrzystość, sprawiedliwość oraz prawa i autonomię osób, nie będą zakazane na mocy aktu w sprawie sztucznej inteligencji.

Na przykład zgodność z przepisami o ochronie danych, takimi jak RODO, które nakładają obowiązki w zakresie przejrzystości przetwarzania danych, a mianowicie wymóg, by unikać wprowadzającego w błąd lub manipulacyjnego języka w informacjach przekazywanych osobom, których dane dotyczą[93]. W niektórych przypadkach aby przetwarzanie danych osobowych było zgodne z prawem, może być wymagana zgoda, podobnie jak w przypadku niektórych spersonalizowanych reklam internetowych opartych na nieobjętych usługą danych użytkowników w sieciach społecznościowych[94]. Zgoda ta musi być m.in. dobrowolna i świadoma. Istnieje większe prawdopodobieństwo, że systemy AI, które spełniają te normy prawne, będą wykorzystywać zgodną z prawem perswazję. Z drugiej strony systemy, które obchodzą te wymogi w celu wywierania wpływu na zachowanie, będą prawdopodobnie wykorzystywać manipulację.

[93] Wytyczne Europejskiej Rady Ochrony Danych, https://www.edpb.europa.eu/system/files/2023-02/edpb_03-2022_guidelines_on_deceptive_design_patterns_in_social_media_platform_interfaces_v2_en_0.pdf, pkt 18.

[94] Wyrok Trybunału Sprawiedliwości z dnia 4 lipca 2023 r., *Meta Platforms i in.*, C-252/21, ECLI:EU:C:2023:537 (zwany dalej „wyrokiem w sprawie Meta Platforms").

132) W szczególności w motywie 29 aktu w sprawie sztucznej inteligencji wyjaśniono, że pod pewnymi warunkami zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji nie mają wpływu na legalne praktyki w kontekście leczenia.

Na przykład techniki podprogowe wspomagane sztuczną inteligencją mogą być stosowane w terapii psychologicznej w związku z chorobą psychiczną lub w rehabilitacji fizycznej, gdy praktyki te są prowadzone zgodnie z mającymi zastosowanie prawem i normami medycznymi, na przykład za wyraźną zgodą danej osoby fizycznej lub jej przedstawiciela prawnego, która to zgoda stanowi warunek ich stosowania.

133) Ponadto w motywie 29 aktu w sprawie sztucznej inteligencji wyjaśniono, że powszechne i zgodne z prawem praktyki handlowe, takie jak reklama, nie powinny być uznawane „same w sobie" lub ze względu na sam ich charakter za szkodliwe oparte na AI praktyki manipulacyjne, wprowadzające w błąd lub prowadzące do wykorzystywania.

Na przykład:

– techniki reklamowe wykorzystujące sztuczną inteligencję do personalizacji treści w oparciu o preferencje użytkowników nie są z natury manipulacyjne, jeżeli nie stosują technik podprogowych, celowo manipulacyjnych lub wprowadzających w błąd, które podważają autonomię jednostki lub wykorzystują słabości w sposób szkodliwy, co jest zakazane na podstawie art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji. Wypełnianie odpowiednich obowiązków wynikających z RODO, przestrzeganie prawa ochrony konsumentów i przepisów rozporządzenia (UE) 2022/2065 („akt o usługach cyfrowych") pomaga ograniczyć takie ryzyko;

– opracowywanie zgodnie z obowiązującym prawem modeli AI i klasyfikatorów w zakresie wykrywania niegodziwego traktowania dzieci w celach seksualnych w internecie stanowi zgodne z prawem praktyki, które nie wykorzystują podatności dzieci, a wręcz przeciwnie, mają zasadnicze znaczenie dla poprawy ich bezpieczeństwa w internecie.

– systemy AI wykorzystywane do świadczenia usług bankowych, takich jak udzielanie kredytów hipotecznych i pożyczek, które to systemy wykorzystują jako dane wejściowe wiek lub szczególną sytuację społeczno-ekonomiczną klienta, zgodnie z przepisami Unii dotyczącymi usług finansowych, ochrony konsumentów, ochrony danych i niedyskryminacji, nie kwalifikują się jako wykorzystywanie słabości w rozumieniu art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji, jeżeli mają na celu ochronę i wspieranie osób uznanych za podatne na zagrożenia ze względu na wiek, niepełnosprawność lub szczególną sytuację społeczno-ekonomiczną i są korzystne dla tych grup, przyczyniając się również do bardziej sprawiedliwych i zrównoważonych usług finansowych dla tych grup;

– systemy AI, które wykrywają senność i zmęczenie kierowców oraz ostrzegają ich o konieczności odpoczynku zgodnie z przepisami bezpieczeństwa, są korzystne i nie

kwalifikują się jako wykorzystywanie słabości w rozumieniu art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji.

3.5.2. Manipulacyjne, wprowadzające w błąd i prowadzące do wykorzystania systemy AI, które prawdopodobnie nie wyrządzą poważnej szkody

134) Zasadniczym warunkiem, który musi zostać spełniony, aby zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji miały zastosowanie, jest to, że manipulacja i wykorzystywanie słabości wspomagane sztuczną inteligencją powinny wyrządzać lub z uzasadnionym prawdopodobieństwem mogły wyrządzać poważną szkodę. Wszystkie zastosowania AI służące manipulacji, wprowadzaniu w błąd i wykorzystywaniu, które z uzasadnionym prawdopodobieństwem nie wyrządzą poważnej szkody, zasadniczo nie są objęte zakresem zakazów, bez uszczerbku dla innych przepisów prawa Unii, które nadal mają zastosowanie (zob. sekcja 3.6 poniżej).

Przykłady systemów AI, które prawdopodobnie nie wyrządzą poważnej szkody, obejmują następujące przypadki:

– system, w którym AI odgrywa rolę towarzysza, jest zaprojektowany w sposób antropomorficzny i wykorzystuje przetwarzanie emocjonalne, aby podnieść atrakcyjność systemu i skutecznie zwiększyć zaangażowanie użytkowników, ale nie stosuje manipulacyjnych lub wprowadzających w błąd praktyk w sposób, który z uzasadnionym prawdopodobieństwem może wyrządzić u nich poważne szkody psychiczne, fizyczne lub inne, niezdrowe przywiązanie i uzależnienie;

– chatbot terapeutyczny wykorzystuje techniki podprogowe, aby pomóc użytkownikom przejść na zdrowszy tryb życia i zrezygnować ze złych nawyków, takich jak palenie tytoniu. Nawet jeśli użytkownicy, którzy stosują się do porad chatbota i terapii podprogowej, doświadczają pewnego dyskomfortu fizycznego i stresu psychicznego z powodu wysiłków podejmowanych w celu zaprzestania palenia, nie można uznać, że ten chatbot wspomagany sztuczną inteligencją może z uzasadnionym prawdopodobieństwem wyrządzić poważną szkodę. Taki tymczasowy dyskomfort jest nieunikniony i równoważą go długoterminowe korzyści dla zdrowia użytkowników. Nie ma ukrytych prób wpływania na proces decyzyjny poza promowaniem zdrowych nawyków;

– internetowa platforma muzyczna korzysta z systemu rozpoznawania emocji, aby poznać emocje użytkowników i automatycznie poleca im utwory zgodne z ich nastrojem, unikając jednocześnie nadmiernej ekspozycji na utwory depresyjne. Ponieważ użytkownicy po prostu słuchają muzyki i w żaden inny sposób nie są krzywdzeni, ani nie powoduje się u nich depresji i lęku, nie ma uzasadnionego prawdopodobieństwa, że system wyrządzi im poważną szkodę;

– wspomagane sztuczną inteligencją techniki manipulacyjne i wprowadzające w błąd wykorzystywane w szkoleniach w zakresie bezpieczeństwa i innych symulacjach uczenia się, które naśladują próby phishingu, mające na celu edukowanie użytkowników na temat zagrożeń dla cyberbezpieczeństwa. Systemy te mogą wykorzystywać techniki celowo manipulacyjne (np. błędy poznawcze), które

w sposób nieświadomy dla użytkowników zmieniają ich zachowanie, ale odbywa się to tymczasowo w celu zapewnienia korzyści ze szkolenia i podnoszenia świadomości oraz bez wyrządzania poważnych szkód.

3.6. Wzajemne powiązania z innymi przepisami prawa Unii

135) Zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji pozostają bez uszczerbku dla innych przepisów prawa Unii i je uzupełniają. Ta sama praktyka, która wchodzi w zakres zakazu określonego w art. 5 ust. 1 lit. a) lub b) aktu w sprawie sztucznej inteligencji, może również stanowić naruszenie innych przepisów prawa Unii i podlegać egzekwowaniu zarówno na mocy aktu w sprawie sztucznej inteligencji, jak i tych innych aktów. Jest to istotne, ponieważ różne przepisy tych aktów mają na celu ochronę różnych interesów i mają różne cele i zakresy oraz różnych adresatów. Zapewnia to kompleksowe podejście regulacyjne, które chroni osoby i grupy osób przed szkodliwym wykorzystywaniem sztucznej inteligencji i manipulacją oraz zapewnia bezpieczne i godne zaufania usługi i produkty wspomagane sztuczną inteligencją w Unii.

136) Zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji są ściśle zgodne z celami unijnego prawa ochrony konsumentów, w szczególności dyrektywy o nieuczciwych praktykach handlowych, która chroni konsumentów przed wprowadzającymi w błąd lub agresywnymi praktykami biznesowymi, w tym w przypadku, gdy praktyki te są oparte na sztucznej inteligencji. Zarówno akt w sprawie sztucznej inteligencji, jak i dyrektywa o nieuczciwych praktykach handlowych mają na celu proaktywne zapobieganie szkodom dla konsumentów wynikającym z praktyk biznesowych opartych na sztucznej inteligencji, które są manipulacyjne, wprowadzające w błąd lub agresywne. Jednocześnie zakazy zawarte w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji mają szerszy zakres, ponieważ chronią nie tylko konsumentów, ale również każdą osobę fizyczną i jej zachowanie w różnych kontekstach wykraczających poza kontekst handlowy. Szkody objęte aktem w sprawie sztucznej inteligencji są również szersze niż szkody gospodarcze, chociaż w akcie w sprawie sztucznej inteligencji określono próg poważnej szkody, który nie występuje w prawie ochrony konsumentów.

137) Zakazy te są również zgodne z unijnymi przepisami o ochronie danych, w tym z zasadami dotyczącymi zgodnego z prawem, rzetelnego i przejrzystego przetwarzania danych, które to zasady mają na celu ochronę danych osobowych osób, których dane dotyczą, a w szerszym ujęciu ochronę ich praw podstawowych i autonomii. Dostępność większej liczby danych (osobowych) i zwiększone możliwości przetwarzania tych danych za pomocą systemów AI zwiększają ryzyko szkodliwych praktyk manipulacyjnych, wprowadzających w błąd lub prowadzących do wykorzystywania, takich jak praktyki objęte zakresem art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji. W tym kontekście zgodność z przepisami o ochronie danych dotyczącymi przejrzystości, minimalizacji danych, rzetelności i zgodności z prawem, na przykład w odniesieniu do spersonalizowanego profilowania i reklamy, w oparciu o nieobjęte

usługą dane użytkowników[95], może przyczynić się do uniknięcia szkodliwych zindywidualizowanych manipulacji i wykorzystywania.

138) Wzajemne powiązania z unijnymi przepisami dotyczącymi niedyskryminacji są również istotne w przypadku zakazu określonego w art. 5 ust. 1 lit. b) aktu w sprawie sztucznej inteligencji[96], jako że słabości ze względu na wiek i niepełnosprawność są również cechami chronionymi, ze względu na które ludzie mają prawo do ochrony przed niedyskryminacyjnym traktowaniem, podczas gdy sytuacja społeczno-ekonomiczna nakłada się na różne inne cechy, takie jak rasa i pochodzenie etniczne. Zakazy zawarte w akcie w sprawie sztucznej inteligencji nie mają wpływu na zakazy oparte na innych cechach lub praktyki dyskryminacyjne, które nie wyrządzają poważnej szkody i które są już zakazane na mocy unijnych przepisów dotyczących niedyskryminacji.

139) Zakazy ustanowione w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji uzupełniają również rozporządzenie (UE) 2022/2065 (akt o usługach cyfrowych), które reguluje internetowe usługi pośrednie, takie jak platformy internetowe i wyszukiwarki, oraz zapewnia przejrzystość i rozliczalność w zakresie świadczenia tych usług. W szczególności w art. 25 ust. 1 aktu o usługach cyfrowych zakazano stosowania zwodniczych interfejsów w interfejsie użytkownika w celu zapewnienia, aby dostawcy platform internetowych nie wprowadzali użytkowników w błąd ani nie zmuszali ich do działań, które mogą nie być zgodne z ich rzeczywistymi intencjami. Takie zwodnicze interfejsy należy rozumieć jako przykład technik manipulacyjnych lub wprowadzających w błąd w rozumieniu art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji, jeżeli mogą one z uzasadnionym prawdopodobieństwem wyrządzać poważne szkody.

140) W akcie o usługach cyfrowych określono również obowiązki dostawców platform internetowych w zakresie zapewnienia przejrzystości reklamy (art. 26 i 38 w przypadku bardzo dużych platform lub wyszukiwarek internetowych), korzystania z systemów rekomendacji (art. 27) oraz ochrony małoletnich (art. 28 aktu o usługach cyfrowych). Ponadto jeżeli platforma lub wyszukiwarka internetowa są sklasyfikowane jako bardzo duża platforma lub wyszukiwarka internetowa, dostawca tej wskazanej usługi ma dodatkowe obowiązki w zakresie oceny i ograniczania ryzyka systemowego wynikającego z projektu lub funkcjonowania jego usługi i powiązanych z nim systemów, w tym systemów algorytmicznych (art. 34 i 35 aktu o usługach cyfrowych).

---

[95] Szczególnie istotny w tym względzie jest wyrok Trybunału (wielka izba) z dnia 4 lipca 2023 r., sprawa C-252/21, *Meta Platforms Inc i in. przeciwko Bundeskartellamt*. Chociaż TSUE stwierdza między innymi, że przetwarzanie nieobjętych usługą danych osobowych użytkowników do celów marketingu bezpośredniego przez dużą platformę sieci społecznościowej można uznać za przetwarzanie w uzasadnionym interesie administratora danych, nie może to mieć miejsca bez zgody użytkownika stanowiącej podstawę prawną ze względu na interesy i prawa podstawowe takiego użytkownika, które w okolicznościach tej sprawy, w szczególności w przypadku przetwarzania na szeroką skalę, mają pierwszeństwo przed interesem, jaki ma w personalizacji reklamy operator sieci społecznościowej, który finansuje za pomocą tej reklamy swoją działalność (zob. wyrok w sprawie *Meta Platforms*, pkt 115–118).

[96] Np. dyrektywa Rady 2000/43/WE z dnia 29 czerwca 2000 r. wprowadzająca w życie zasadę równego traktowania osób bez względu na pochodzenie rasowe lub etniczne, Dz.U. L 180 z 19.7.2000, s. 22; dyrektywa Rady 2000/78/WE z dnia 27 listopada 2000 r. ustanawiająca ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy, Dz.U. L 303 z 2.12.2000, s. 16; dyrektywa 2006/54/WE Parlamentu Europejskiego i Rady z dnia 5 lipca 2006 r. w sprawie wprowadzenia w życie zasady równości szans oraz równego traktowania kobiet i mężczyzn w dziedzinie zatrudnienia i pracy (wersja przekształcona), Dz.U. L 204 z 26.7.2006, s. 23; dyrektywa Rady 2004/113/WE z dnia 13 grudnia 2004 r. wprowadzająca w życie zasadę równego traktowania mężczyzn i kobiet w zakresie dostępu do towarów i usług oraz dostarczania towarów i usług, Dz.U. L 373 z 21.12.2004, s. 37.

Przeprowadzając oceny ryzyka, dostawcy bardzo dużych platform i wyszukiwarek internetowych powinni rozważyć, w jaki sposób ich systemy rekomendacji, reklama, moderowanie treści i wszelkie inne odpowiednie systemy algorytmiczne wpływają na takie ryzyko systemowe. W takich ocenach ryzyka należy również przeanalizować, w jaki sposób na ryzyko systemowe wpływa m.in. celowa manipulacja i zautomatyzowane wykorzystywanie usługi (por. art. 34 ust. 2 aktu o usługach cyfrowych i motyw 83 aktu o usługach cyfrowych). Zakres art. 5 ust. 1 lit. a) lub b) aktu w sprawie sztucznej inteligencji obejmuje jednakże szeroki zakres innych scenariuszy (np. chatboty, usługi i produkty wspomagane sztuczną inteligencją), które mogą być oferowane lub wykorzystywane przez podmioty inne niż dostawcy usług pośrednich.

141) Zakaz technik manipulacyjnych opartych na sztucznej inteligencji zgodnie z art. 5 ust. 1 lit. a) aktu w sprawie sztucznej inteligencji wspiera również cele dyrektywy 2010/13/UE (dyrektywa o audiowizualnych usługach medialnych)[97] poprzez zapobieganie szkodliwym reklamom opartym na AI[98] i innym wspomaganym sztuczną inteligencją praktykom polegającym na manipulacji i wykorzystywaniu, które mogą być w sposób znaczący szkodliwe w sektorze mediów.

142) Akt w sprawie sztucznej inteligencji uzupełnia również rozporządzenie (UE) 2024/900 (rozporządzenie w sprawie reklamy politycznej)[99], które zawiera zharmonizowane przepisy, w tym obowiązki w zakresie przejrzystości i powiązane obowiązki w zakresie należytej staranności, dotyczące świadczenia usług reklamy politycznej i powiązanych usług, oraz wymogi dotyczące stosowania technik targetowania i technik dostarczania reklam w kontekście reklamy politycznej online. W rozporządzeniu tym w kontekście reklamy politycznej online zakazano profilowania na podstawie szczególnych kategorii danych osobowych i targetowania osób, którym brakuje co najmniej jednego roku do osiągnięcia wieku uprawniającego do głosowania, ustanowionego w przepisach krajowych. Ponadto techniki targetowania i dostarczania reklamy w kontekście reklamy politycznej online można stosować wyłącznie wtedy, gdy opierają się one na danych osobowych zebranych od osób, których dane dotyczą, i za ich wyraźną zgodą. Zastosowanie mają również dodatkowe wymogi w zakresie przejrzystości, tj. ujawnianie reklamy politycznej, opisywanie zastosowanych technik i ich głównych parametrów oraz dostarczanie dodatkowych informacji na temat zastosowanej logiki, w tym na temat wykorzystania systemów AI. Ukierunkowana reklama polityczna oparta na przetwarzaniu danych osobowych zgodnie z tym rozporządzeniem[100] przyczyni się do zapewnienia, aby profilowanie wyborców oraz targetowanie i dostarczanie reklam politycznych odbywało się w granicach perswazji zgodnej z prawem.

---

97 Dyrektywa Parlamentu Europejskiego i Rady 2010/13/UE z dnia 10 marca 2010 r. w sprawie koordynacji niektórych przepisów ustawowych, wykonawczych i administracyjnych państw członkowskich dotyczących świadczenia audiowizualnych usług medialnych (dyrektywa o audiowizualnych usługach medialnych), zmieniona dyrektywą (UE) 2018/1808, której celem jest między innymi poprawa ochrony dzieci i skuteczniejsze zwalczanie mowy nienawiści.

98 Art. 9 dyrektywy o audiowizualnych usługach medialnych.

99 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/900 z dnia 13 marca 2024 r. w sprawie przejrzystości i targetowania reklamy politycznej, PE/90/2023/REV/1, Dz.U. L, 2024/900, 20.3.2024.

100 Z chwilą rozpoczęcia stosowania od października 2025 r.

143) Zakaz szkodliwych praktyk w zakresie AI opartych na wykorzystywaniu i wprowadzaniu w błąd zawarty w akcie w sprawie sztucznej inteligencji stanowi również uzupełnienie innych mających zastosowanie przepisów Unii, w których ustanowiono ogólne przepisy dotyczące przejrzystości w zakresie reklamy i ochrony konsumentów oraz należytego postępowania operatorów (np. dyrektywa 2014/65/UE MiFID, dyrektywa (UE) 2016/97 w sprawie dystrybucji ubezpieczeń[101], dyrektywa (UE) 2023/2225 w sprawie umów o kredyt konsumencki, dyrektywa (UE) 2002/65 dotycząca sprzedaży usług na odległość, dyrektywa 2006/114/WE w sprawie reklamy wprowadzającej w błąd i reklamy porównawczej oraz dyrektywa (UE) 2011/83 w sprawie praw konsumentów ustanawiają ogólne standardy ochrony konsumentów). W tym względzie Europejski Urząd Nadzoru Ubezpieczeń i Pracowniczych Programów Emerytalnych (EIOPA) wydał już oświadczenie dla organów nadzoru na temat prowadzących do wykorzystywania nieuczciwych praktyk dotyczących różnic cenowych, które to praktyki mogą również wchodzić w zakres aktu w sprawie sztucznej inteligencji, jeżeli są wspomagane systemami AI[102].

144) Zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji pozostają również bez uszczerbku dla unijnych przepisów dotyczących bezpieczeństwa produktów (np. w odniesieniu do wyrobów medycznych, zabawek, maszyn) i uzupełniają je, co odgrywa kluczową rolę w zapewnianiu bezpieczeństwa produktów wyposażonych w systemy AI. Wiąże się to z przestrzeganiem wymogów bezpieczeństwa *ex ante* w odniesieniu do produktów podlegających regulacji oraz z ich proaktywnym monitorowaniem w celu zapewnienia, aby nie stwarzały one ryzyka dla bezpieczeństwa wiążącego się ze szkodami fizycznymi i psychicznymi. Producent takich produktów z wbudowanymi systemami AI powinien zatem uwzględnić te zakazy w swoich ocenach ryzyka i środkach ograniczających ryzyko na rzecz bezpieczeństwa w zakresie, w jakim jest to zgodne z logiką i zakresem odpowiednich zharmonizowanych przepisów Unii w zakresie bezpieczeństwa. Unijne przepisy dotyczące bezpieczeństwa stanowią również uzupełnienie zakazów zawartych w akcie w sprawie sztucznej inteligencji i mogą również dotyczyć ryzyka dla bezpieczeństwa, które nie wiąże się z poważną szkodą, i przeciwdziałać takiemu ryzyku. W szczególności rozporządzenie (UE) 2023/988 (rozporządzenie w sprawie ogólnego bezpieczeństwa produktów)[103] odgrywa rolę siatki bezpieczeństwa, ponieważ zawarto w nim wymóg, aby wszystkie produkty konsumenckie nieobjęte szczegółowymi wymogami zawartymi w innych sektorowych przepisach Unii dotyczących bezpieczeństwa produktów (w tym produkty, w których wykorzystano systemy AI niesklasyfikowane jako systemy wysokiego ryzyka zgodnie z art. 6 i podlegające

---

101 Dyrektywa Parlamentu Europejskiego i Rady (UE) 2016/97 z dnia 20 stycznia 2016 r. w sprawie dystrybucji ubezpieczeń (wersja przekształcona), Dz.U. L 26 z 2.2.2016, s. 19. Np. w art. 17 ust. 1 dyrektywy w sprawie dystrybucji ubezpieczeń wprowadzono wymóg, aby dystrybutorzy ubezpieczeń postępowali uczciwie, rzetelnie i profesjonalnie, zgodnie z najlepiej pojętym interesem swoich klientów.

102 https://www.eiopa.europa.eu/document/download/1e9a8fb2-e688-4bf5-a347-ee0a1ec3aab3_en?filename=EIOPA-BoS-23-076-Supervisory-Statement-on-differential-pricing-practices_0.pdf.

103 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2023/988 z dnia 10 maja 2023 r. w sprawie ogólnego bezpieczeństwa produktów, zmieniające rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1025/2012 i dyrektywę Parlamentu Europejskiego i Rady (UE) 2020/1828 oraz uchylające dyrektywę 2001/95/WE Parlamentu Europejskiego i Rady i dyrektywę Rady 87/357/EWG (Tekst mający znaczenie dla EOG).

wymogom określonym w akcie w sprawie sztucznej inteligencji) były bezpieczne w normalnych lub racjonalnie przewidywalnych warunkach użytkowania, w szczególności w odniesieniu do zagrożeń dla zdrowia fizycznego i psychicznego konsumentów.

145) Ponadto kluczowe znaczenie mają wzajemne powiązania z prawem karnym. Zakazy ustanowione w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji mają na celu zapobieganie szkodliwym zachowaniom, które mogą stanowić przestępstwo lub prowadzić do popełnienia przestępstw, takim jak oszustwa, fałszerstwa, wyłudzenia, przymus lub generowanie i rozpowszechnianie nielegalnych treści, takich jak treści terrorystyczne, materiały przedstawiające niegodziwe traktowanie dzieci w celach seksualnych, mowa nienawiści i treści typu „deepfake" o charakterze jednoznacznie seksualnym[104]. Co ważne, podobnie jak przepisy dotyczące rynku wewnętrznego, zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji obejmują nie tylko wykorzystywanie systemu AI, ale również jego wprowadzanie do obrotu, co zapobiega szkodom na wczesnym etapie poprzez ograniczanie dostępu do takich zakazanych systemów, które mogą ułatwiać i ukrywać działalność przestępczą. Ponadto zakazy określone w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji mogą również obejmować inne szkodliwe praktyki, które nie zostały zakwalifikowane jako przestępstwa na podstawie prawa Unii lub prawa krajowego.

4. ART. 5 UST. 1 LIT. C) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – SCORING SPOŁECZNY

146) Chociaż scoring wspomagany sztuczną inteligencją może przynieść korzyści w zakresie promowania dobrych zachowań, poprawy bezpieczeństwa, efektywności lub jakości usług, istnieją pewne praktyki „scoringu społecznego", w ramach których niesprawiedliwie traktuje się ludzi lub szkodzi im, i które są równoznaczne z kontrolą społeczną i nadzorem społecznym. Zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji dotyczy takich niedopuszczalnych praktyk „scoringu społecznego" wspomaganego sztuczną inteligencją, w ramach których ocenia się lub klasyfikuje osoby lub grupy na podstawie ich zachowania społecznego lub cech osobistych i które prowadzą do szkodliwego lub niekorzystnego traktowania, w szczególności gdy dane pochodzą z wielu niepowiązanych kontekstów społecznych lub gdy traktowanie jest nieproporcjonalne do wagi zachowania społecznego. Zakaz „scoringu społecznego" ma szeroki zakres zastosowania zarówno w kontekście publicznym, jak i prywatnym, i nie ogranicza się do konkretnego sektora lub dziedziny[105].

147) Jednocześnie zakaz ten nie ma wpływać na zgodne z prawem praktyki oceny osób do konkretnych celów, które są uzasadnione i zgodne z prawem Unii i prawem

[104] Dyrektywa Parlamentu Europejskiego i Rady (UE) 2024/1385 z dnia 14 maja 2024 r. w sprawie zwalczania przemocy wobec kobiet i przemocy domowej, PE/33/2024/REV/1, Dz.U. L, 2024/1385, 24.5.2024.

[105] Zakaz scoringu społecznego różni się od zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, który jest bardziej wyspecjalizowany w odniesieniu do praktyki oceny/scoringu i ma zastosowanie wyłącznie do oceny ryzyka i przewidywania prawdopodobieństwa popełnienia przestępstwa przez daną osobę; w przepisie tym zakazano systemów AI opartych wyłącznie na profilowaniu lub ocenie cech osobowości i cech charakterystycznych (zob. sekcja 5).

krajowym[106], w szczególności gdy przepisy te określają rodzaje danych istotne dla konkretnych celów oceny i zapewniają, by wszelkie wynikające z tego szkodliwe lub niekorzystne traktowanie osób było uzasadnione i proporcjonalne (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

4.1. Uzasadnienie i cele

148) Systemy AI umożliwiające stosowanie praktyk „scoringu społecznego" mogą prowadzić do dyskryminujących i niesprawiedliwych wyników dla niektórych osób i grup, w tym do ich wykluczenia ze społeczeństwa, a także do praktyk w zakresie kontroli społecznej i nadzoru, które są niezgodne z wartościami Unii. Zakaz „scoringu społecznego" ma w szczególności na celu ochronę prawa do godności ludzkiej i innych praw podstawowych, w tym prawa do niedyskryminacji i równości, do ochrony danych oraz do życia prywatnego i rodzinnego, a także, w stosownych przypadkach, odpowiednich praw społecznych i gospodarczych. Jego celem jest również ochrona i propagowanie unijnych wartości, jakimi są demokracja, równość (w tym równy dostęp do usług publicznych i prywatnych) i sprawiedliwość[107].

4.2. Główne pojęcia i elementy zakazu „scoringu społecznego"

> ***Art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji stanowi:***
>
> Zakazuje się następujących praktyk w zakresie AI:
>
> c) wprowadzania do obrotu, oddawania do użytku lub wykorzystywania systemów AI na potrzeby oceny lub klasyfikacji osób fizycznych lub grup osób prowadzonej przez określony czas na podstawie ich zachowania społecznego lub znanych, wywnioskowanych lub przewidywanych cech osobistych lub cech osobowości, kiedy to scoring społeczny prowadzi do jednego lub obu z następujących skutków:
>
> (i) krzywdzącego lub niekorzystnego traktowania niektórych osób fizycznych lub grup osób w kontekstach społecznych, które nie są związane z kontekstami, w których pierwotnie wygenerowano lub zebrano dane;
>
> (ii) krzywdzącego lub niekorzystnego traktowania niektórych osób fizycznych lub grup osób, które jest nieuzasadnione lub nieproporcjonalne do ich zachowania społecznego lub jego wagi;

149) Aby zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi zostać spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku" lub „wykorzystywaniu" systemu AI;

(ii) system AI musi być przeznaczony lub wykorzystywany do oceny lub klasyfikacji osób fizycznych lub grup osób prowadzonej przez określony czas na podstawie:

a) ich zachowania społecznego; lub

[106] Motyw 31 aktu w sprawie sztucznej inteligencji.
[107] Motyw 31 aktu w sprawie sztucznej inteligencji.

b) znanych, wywnioskowanych lub przewidywanych cech osobistych lub cech osobowości;

(iii) scoring społeczny uzyskany przy pomocy systemu AI musi prowadzić lub potencjalnie prowadzić do szkodliwego lub niekorzystnego traktowania osób lub grup w co najmniej jednym z następujących scenariuszy:

a) szkodliwe lub niekorzystne traktowanie odbywa się w kontekstach społecznych niezwiązanych z kontekstami, w których dane pierwotnie wygenerowano lub zebrano; lub

b) traktowanie to jest nieuzasadnione lub nieproporcjonalne do zachowania społecznego tych osób lub grup lub jego wagi.

150) Aby zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji miał zastosowanie, muszą być spełnione jednocześnie wszystkie trzy warunki. Pierwszy warunek, tj. wprowadzenie do obrotu, oddanie do użytku lub wykorzystywanie systemu AI, został już przeanalizowany w sekcji 2.3. Zakaz ten ma zatem zastosowanie zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów AI do obrotu, oddawać ich do użytku ani wykorzystywać. Pozostałe kryteria zakazu „scoringu społecznego" opisano i przeanalizowano poniżej.

4.2.1. „Scoring społeczny": ocena lub klasyfikacja oparta na zachowaniu społecznym lub cechach osobistych lub cechach osobowości przez określony czas

a) Ocena lub klasyfikacja osób fizycznych lub grup osób

151) Drugim warunkiem, który musi zostać spełniony, aby zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji miał zastosowanie, jest to, że system AI musi być przeznaczony lub wykorzystywany do **oceny lub klasyfikacji** osób fizycznych lub grup osób i przydzielać im oceny punktowe na podstawie ich zachowania społecznego lub cech osobistych lub cech osobowości. Ocena punktowa uzyskana przez system może przybierać różne formy, takie jak liczba (na przykład od 0 do 1), ranking lub etykieta.

152) Zakres zakazu jest szeroki i obejmuje praktyki w zakresie oceny i klasyfikacji zarówno w sektorze publicznym, jak i prywatnym (zob. sekcja 4.2.3). Jednocześnie ocena lub klasyfikacja dotyczy wyłącznie osób fizycznych lub grup osób fizycznych, co zasadniczo wyklucza podmioty prawne (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

153) Podczas gdy „**ocena**" sugeruje przeprowadzenie pewnej formy **ewaluacji lub osądu** danej osoby lub grupy osób, sama **klasyfikacja** osób lub grup osób w oparciu o cechy, takie jak ich wiek, płeć i wzrost, niekoniecznie musi prowadzić do oceny[108]. Zakres „klasyfikacji" jest zatem szerszy niż zakres „oceny" i może również obejmować inne

[108] Grupa Robocza Art. 29, *Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679*, WP251rev.01, 6.02.2018, s. 7.

rodzaje klasyfikacji lub kategoryzacji osób fizycznych lub grup osób na podstawie kryteriów, które niekoniecznie wiążą się z konkretną oceną lub osądem tych osób lub grup osób oraz ich cech lub zachowania.

154) Termin „**ocena**" wiąże się również z pojęciem „profilowania", które jest regulowane unijnymi przepisami o ochronie danych[109] i stanowi szczególną formę oceny. Chociaż w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji nie odniesiono się bezpośrednio do tego pojęcia ani do tych przepisów[110], mogą one mieć również znaczenie dla zakazu zawartego w tym przepisie, a także dla innych zakazów zawartych w akcie w sprawie sztucznej inteligencji[111], jeżeli ocena dokonywana jest w sposób zautomatyzowany przez system AI w oparciu o dane osobowe. **Profilowanie** oznacza wykorzystywanie informacji o danej osobie (lub grupie osób fizycznych) i ocenę ich cech lub wzorców zachowania w celu zakwalifikowania ich do określonej kategorii lub grupy, w szczególności do celów **analizy lub prognozy** takich aspektów, jak wykazywane przez dane osoby zdolności do wykonania danego zadania; zainteresowania; lub prawdopodobne zachowania[112]. Profilowanie osób fizycznych na podstawie unijnych przepisów o ochronie danych, jeżeli odbywa się z wykorzystaniem systemów AI, może zatem również wchodzić w zakres art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

Na przykład w wyroku w sprawie *SCHUFA I* TSUE zbadał **system punktowej oceny zdolności kredytowej** stosowany w Niemczech[113]. W tej sprawie „ocena punktowa" wygenerowana przez program komputerowy stanowiła „wartość prawdopodobieństwa" dotyczącą zdolności danej osoby do wywiązania się ze zobowiązań płatniczych, co zostało zakwalifikowane przez TSUE jako „profilowanie". Dokładniej rzecz ujmując, system sporządził prognozę dotyczącą prawdopodobieństwa przyszłego zachowania danej osoby („score"), takiego jak zwrot pożyczki, w oparciu o określone cechy tej osoby. Wyliczenie wartości scoringu opiera się na założeniu, że poprzez przyporządkowanie danej osoby do grupy innych osób o pewnych porównywalnych cechach, które zachowały się w określony sposób, można przewidzieć podobne zachowanie[114]. Według TSUE działalność ta odpowiada definicji „profilowania" w rozumieniu art. 4 ust. 4 RODO[115]. Tę formę profilowania można również uznać za ocenę osób na podstawie ich cech osobistych w rozumieniu art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, która będzie zakazana, jeżeli będzie

109 Zob. art. 4 ust. 4 i art. 22 RODO oraz art. 11 dyrektywy o ochronie danych w sprawach karnych. Grupa Robocza Art. 29, Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679, WP251rev.01, 6.02.2018, s. 7.

110 Art. 3 ust. 52 aktu w sprawie sztucznej inteligencji zawiera definicję „profilowania", która odsyła do definicji zawartej w art. 4 ust. 4 RODO.

111 W szczególności zakaz przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną ustanowiony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, w którym to zakazie mowa jest o „profilowaniu", a w niektórych przypadkach zakaz rozpoznawania emocji i kategoryzacji biometrycznej ustanowiony w art. 5 ust. 1 lit. f) i g) aktu w sprawie sztucznej inteligencji.

112 Grupa Robocza Art. 29, *Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679*, WP251rev.01, 6.02.2018, s. 7.

113 Wyrok Trybunału Sprawiedliwości z dnia 7 grudnia 2023 r., SCHUFA Holding (Scoring), C-634/21, EU:C:2023:957 (zwany dalej „wyrokiem w sprawie SCHUFA I"), np. pkt 47.

114 Tamże, pkt 14 (podkreślenie własne).

115 Tamże, pkt 47.

prowadzona za pomocą systemów AI, pod warunkiem że spełnione będą pozostałe warunki wymagane, by przepis ten miał zastosowanie.

b) Przez określony czas

155) Zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji wymaga, aby ocena lub klasyfikacja opierały się na danych, które obejmują „**określony czas**". Sugeruje to, że ocena nie powinna ograniczać się do jednorazowego czy doraźnego ratingu lub klasyfikacji na podstawie danych lub zachowań z bardzo szczególnego indywidualnego kontekstu. Jednocześnie ważne jest, aby warunek ten był oceniany z uwzględnieniem wszystkich okoliczności sprawy, aby uniknąć obchodzenia zakresu zakazu.

Na przykład organ ds. migracji i azylu wdraża częściowo zautomatyzowany system nadzoru w obozach dla uchodźców, oparty na szeregu elementów infrastruktury nadzoru, w tym kamerach i czujnikach ruchu. Jeżeli analizowane dane obejmują określony czas i oceniane są konkretne osoby (takie jak migranci), na przykład w celu ustalenia, czy istnieje ryzyko ich ucieczki, wówczas można to uznać za spełnienie warunku dotyczącego „określonego czasu", a zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji może mieć zastosowanie, jeżeli spełnione są wszystkie pozostałe warunki.

c) Na podstawie ich zachowania społecznego lub znanych, wywnioskowanych lub przewidywanych cech osobistych lub cech osobowości

156) Praktyki „oceny" i „klasyfikacji" zakazane na mocy art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji muszą opierać się na wspomaganym sztuczną inteligencją przetwarzaniu danych (często obszernych) w odniesieniu do (i) zachowań społecznych osób fizycznych lub grup osób albo (ii) ich znanych, wywnioskowanych lub przewidywanych cech osobistych i cech osobowości, lub obu tych aspektów. Dane mogą być przekazywane bezpośrednio przez te osoby lub pozyskiwane pośrednio, tj. poprzez nadzór, od osób trzecich lub w drodze wniosków wyciągniętych z innych informacji.

157) Jeżeli chodzi o pierwszy scenariusz, „**zachowanie społeczne**" to szeroki termin, który może zasadniczo dotyczyć działań, zachowań, nawyków, interakcji w społeczeństwie itp. i który zazwyczaj obejmuje punkty danych związane z zachowaniem pochodzące z wielu źródeł[116]. Może to obejmować zachowanie osób i grup osób w kontekście społecznym i prywatnym, takie jak udział w wydarzeniach kulturalnych, wolontariacie itp., ale również zachowania społeczne w kontekście biznesowym, na przykład spłatę długów, zachowanie podczas korzystania z określonych usług, a także stosunki z podmiotami publicznymi i prywatnymi, rządem, policją i prawem (np. czy dana osoba przestrzega przepisów ruchu drogowego). Dane dotyczące zachowań społecznych pochodzące z wielu kontekstów i punktów danych mogą być gromadzone w sposób scentralizowany przez ten sam podmiot, ale najczęściej są gromadzone

[116] Zob. motyw 31 aktu w sprawie sztucznej inteligencji.

w sposób rozproszony i łączone z różnych źródeł, co może obejmować wzmożone monitorowanie i śledzenie osób fizycznych (tzw. „nadzór nad danymi").

158) Drugi scenariusz dotyczy sytuacji, w której ocena punktowa opiera się na **cechach osobistych lub cechach osobowości**, które mogą – ale nie muszą, – obejmować szczególne aspekty zachowań społecznych. „Cechy osobiste" mogą obejmować różne informacje dotyczące danej osoby, na przykład płeć, orientację seksualną lub cechy seksualne, płeć społeczno-kulturową, tożsamość płciową, rasę, pochodzenie etniczne, sytuację rodzinną, adres, dochód, członków gospodarstwa domowego, zawód, zatrudnienie lub inny status prawny, wyniki w pracy, sytuację ekonomiczną, płynność finansową, zdrowie, preferencje osobiste, zainteresowania, wiarygodność, zachowanie, lokalizację lub przemieszczanie się, poziom zadłużenia, rodzaj samochodu itp.[117] „Cechy osobowości" należy co do zasady interpretować jako synonim cech osobistych, ale mogą one również oznaczać tworzenie konkretnych profili osób jako osobowości. Cechy osobowości mogą również opierać się na szeregu czynników i wiązać się z osądem, który może być dokonywany przez same osoby wobec siebie, inne osoby lub generowany przez systemy AI. W akcie w sprawie sztucznej inteligencji cechy osobowości są czasami określane jako cechy osobowości i cechy charakterystyczne[118]; pojęcia te powinny być interpretowane w sposób spójny.

159) „Znane, wywnioskowane lub przewidywane" cechy osobiste lub cechy osobowości to różne rodzaje informacji i danych osobowych, które należy rozróżnić. „**Cechy znane**" opierają się na informacjach, które zostały przekazane do systemu AI jako dane wejściowe i które w większości przypadków są informacjami weryfikowalnymi. Natomiast „**cechy wywnioskowane**" opierają się na informacjach, które wyprowadzono z innych informacji, przy czym wnioskowanie jest zwykle dokonywane przez system AI. „**Cechy przewidywane**" to cechy, które szacuje się na podstawie wzorców z dokładnością poniżej 100 %. Pojęcie danych „wywnioskowanych" (lub pochodnych) jest również stosowane w unijnych przepisach o ochronie danych w kontekście profilowania, a zatem może być źródłem inspiracji do interpretacji pojęć wykorzystanych w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji[119]. Korzystanie z tych różnych rodzajów danych może mieć różne skutki dla dokładności i rzetelności praktyk scoringu i w związku z tym może być brane pod uwagę, w szczególności gdy przetwarzanie jest nieprzejrzyste lub opiera się na punktach danych, których dokładność jest trudniejsza do zweryfikowania.

4.2.2. Scoring społeczny musi prowadzić do szkodliwego lub niekorzystnego traktowania w niepowiązanych kontekstach społecznych lub do traktowania, które jest nieuzasadnione lub nieproporcjonalne do wagi zachowania społecznego

117 Zob. motyw 42 aktu w sprawie sztucznej inteligencji, w którym wymieniono niektóre przykłady takich cech.

118 Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

119 Grupa Robocza Art. 29, *Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679*, WP251rev.01, 6.02.2018, s. 7 i nast.

a) Związek przyczynowy między scoringiem społecznym a traktowaniem

160) Aby zakaz określony w art. 5 ust. 1 lit. c) ustawy o sztucznej inteligencji miał zastosowanie, scoring społeczny stworzony przez system AI lub z jego pomocą musi **prowadzić do szkodliwego lub niekorzystnego traktowania** ocenianej osoby lub grupy osób. Innymi słowy, traktowanie musi być konsekwencją scoringu, a scoring przyczyną traktowania. Taki wiarygodny związek przyczynowy może również istnieć w przypadkach, w których szkodliwe skutki jeszcze się nie urzeczywistniły, ale system AI ma na celu lub może wywołać taki niekorzystny skutek. Jest to szczególnie istotne, biorąc pod uwagę, że zakazana praktyka określona w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji obejmuje również „wprowadzanie do obrotu" takich systemów AI.

161) Art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji nie zawiera wymogu, aby ocena lub klasyfikacja dokonywana przez system AI była jedyną przyczyną szkodliwego lub niekorzystnego traktowania. W związku z tym obejmuje on także swoim zakresem praktyki scoringu wspomagane sztuczną inteligencją, które mogą również podlegać innym ocenom dokonywanym przez człowieka lub być połączone z takimi ocenami. Jednocześnie wynik generowany przez AI musi odgrywać wystarczająco ważną rolę w tworzeniu scoringu społecznego. Na przykład w przypadku gdy organ publiczny wykorzystuje system AI do oceny wiarygodności osób i łączy jego wyniki z oceną dodatkowych faktów dokonywaną przez człowieka, ta praktyka scoringu wspomaganego sztuczną inteligencją będzie objęta zakresem zakazu tylko wtedy, gdy wynik generowany przez AI odgrywa wystarczająco ważną rolę w ostatecznej decyzji, pod warunkiem że spełnione są pozostałe warunki szkodliwego lub niekorzystnego traktowania, jak opisano poniżej (zob. sekcja 4.2.2 lit. b)).

162) Scoring może prowadzić do szkodliwego lub niekorzystnego traktowania, nawet jeżeli jest sporządzany przez organizację lub organizacje inne niż organizacja, w której zastosowano scoring[120]. Na przykład organ publiczny może uzyskać punktową ocenę zdolności kredytowej osoby fizycznej sporządzoną przez inne przedsiębiorstwo specjalizujące się w ocenie zdolności kredytowej i ocenie ryzyka, które to oceny opierają się na informacjach o osobach fizycznych i ich zachowaniu pochodzących z różnych źródeł.

b) Szkodliwe lub niekorzystne traktowanie w niepowiązanych kontekstach społecznych lub nieuzasadnione lub nieproporcjonalne traktowanie

163) Ostatecznym warunkiem, który musi zostać spełniony, aby zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji miał zastosowanie, jest to, że wykorzystanie scoringu społecznego musi skutkować szkodliwym lub niekorzystnym traktowaniem (lub może do niego prowadzić):

---

120 Wykładnia ta jest zgodna z wyrokiem TSUE w sprawie SCHUFA I, w którym TSUE orzekł w kontekście zautomatyzowanego podejmowania decyzji, że „ocena punktowa" (ocena stanowiąca profilowanie) opracowana przez podmiot inny niż ten, który podejmuje ostateczną decyzję, może stanowić zautomatyzowaną decyzję na podstawie art. 22 RODO. Zob. wyrok w sprawie SCHUFA I, pkt 42–51 i 60–62.

i. w kontekście społecznym lub kontekstach społecznych niezwiązanych z kontekstami, w których pierwotnie wygenerowano lub zebrano dane, lub

ii. w sposób nieuzasadniony lub nieproporcjonalny do zachowania społecznego lub jego wagi.

164) Warunki te są alternatywne i mogą mieć również zastosowanie łącznie. Analiza poszczególnych przypadków jest konieczna, aby ocenić, czy co najmniej jeden z warunków został spełniony, ponieważ wiele praktyk w zakresie scoringu i oceny wspomaganych sztuczną inteligencją może ich nie spełniać i w związku z tym nie wchodzić w zakres zakazu. W szczególności może to nie mieć miejsca w przypadku, gdy praktyki scoringu wspomagane sztuczną inteligencją służą konkretnemu prawnie uzasadnionemu celowi oceny i są zgodne z mającymi zastosowanie przepisami unijnymi i krajowymi, które określają dane uznane za istotne do celów oceny i zapewniają, aby szkodliwe lub niekorzystne traktowanie było uzasadnione i proporcjonalne do zachowania społecznego (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

165) „**Niekorzystne traktowanie**" oznacza, że w wyniku oceny punktowej dana osoba lub grupa osób musi być traktowana mniej korzystnie w porównaniu z innymi osobami, co niekoniecznie wymaga szczególnej szkody (na przykład w przypadku praktyk scoringu, w ramach których osoby są wybierane do dodatkowych kontroli w przypadku podejrzenia oszustwa). Natomiast traktowanie „**szkodliwe**" wymaga, aby w wyniku traktowania dana osoba lub grupa osób doznała pewnej szkody. Niekorzystne lub szkodliwe traktowanie może być również dyskryminujące i zakazane na mocy unijnego prawa o niedyskryminacji lub oznaczać wykluczenie niektórych osób lub grup[121], ale nie jest to warunkiem koniecznym, aby zakaz miał zastosowanie. Art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji może zatem obejmować swoim zakresem niesprawiedliwe traktowanie wykraczające poza unijne prawo o niedyskryminacji, które ma zastosowanie wyłącznie do niektórych grup chronionych (np. ze względu na wiek, pochodzenie etniczne i rasowe, płeć, religię).

Scenariusz 1: szkodliwe lub niekorzystne traktowanie w niepowiązanych kontekstach społecznych

166) W pierwszym scenariuszu opisanym w art. 5 ust. 1 lit. c) ppkt (i) aktu w sprawie sztucznej inteligencji szkodliwe lub niekorzystne traktowanie wynikające ze scoringu musi mieć miejsce **w kontekście społecznym lub kontekstach społecznych niezwiązanych** z kontekstami, w których pierwotnie wygenerowano lub zebrano dane. Oznacza to nie tylko, że osoby te mogą być traktowane w sposób niekorzystny lub szkodliwy ze względu na scoring, ale również że dane dotyczące ich zachowań społecznych lub ich znanych, wywnioskowanych lub przewidywanych cech osobistych lub cech osobowości są generowane lub zbierane w kontekstach społecznych niezwiązanych z kontekstem, w którym ma miejsce ocena punktowa. Dane zebrane lub

[121] Motyw 31 aktu w sprawie sztucznej inteligencji.

wygenerowane w tych niepowiązanych kontekstach muszą być następnie wykorzystywane przez system AI do scoringu osób do celów oceny lub klasyfikacji bez wyraźnego związku lub w sposób prowadzący do uogólnionego nadzoru nad osobami lub grupami osób. W większości przypadków odbywa się to wbrew uzasadnionym oczekiwaniom osób i stanowi naruszenie unijnych przepisów o ochronie danych i ewentualnie innych mających zastosowanie przepisów określających rodzaje danych i źródła uznane za istotne i niezbędne do oceny lub klasyfikacji. To, czy warunek ten jest spełniony, będzie wymagało indywidualnej oceny, z uwzględnieniem celu oceny oraz kontekstów, w których dane zostały zebrane i wygenerowane.

Przykłady szkodliwego lub niekorzystnego traktowania w niepowiązanych kontekstach społecznych, zakazanego na mocy art. 5 ust. 1 lit. c) ppkt (i) aktu w sprawie sztucznej inteligencji:

– Krajowe organy podatkowe korzystają z narzędzia predykcyjnego AI w odniesieniu do deklaracji podatkowych wszystkich podatników w danym kraju, aby wybrać deklaracje podatkowe do dokładniejszej kontroli. Aby wskazać konkretne osoby, które mają być poddane kontroli, narzędzie AI wykorzystuje istotne zmienne, takie jak roczny dochód, aktywa (nieruchomości, samochody itp.), dane dotyczące członków rodzin beneficjentów, ale również dane niepowiązane, takie jak nawyki towarzyskie podatników lub ich aktywność w internecie.

– Agencja pomocy społecznej wykorzystuje system AI do oszacowania prawdopodobieństwa oszustwa ze strony beneficjentów dodatków na gospodarstwo domowe. System ten wykorzystuje do tego celu informacje na temat cech, zebrane lub wywnioskowane z kontekstów społecznych, które nie mają wyraźnego związku lub znaczenia dla oceny oszustwa, takie jak posiadanie współmałżonka o określonym obywatelstwie lub pochodzeniu etnicznym, posiadanie łącza internetowego, zachowanie na platformach społecznościowych czy wyniki w miejscu pracy itp.[122] Natomiast dane istotne dla przydziału świadczeń i zebrane zgodnie z prawem mogą być wykorzystywane do określenia ryzyka oszustwa, ponieważ organy publiczne realizują uzasadniony cel polegający na sprawdzeniu, czy świadczenia socjalne są prawidłowo przydzielane.

– Publiczna agencja pracy wykorzystuje system AI do oceny osób bezrobotnych na podstawie wywiadu i opartej na sztucznej inteligencji oceny w celu ustalenia, czy dana osoba powinna korzystać ze wsparcia państwa w zakresie zatrudnienia. Ocena ta opiera się na istotnych cechach osobistych, takich jak wiek i wykształcenie, ale również zmiennych zbieranych lub wywnioskowanych z danych i kontekstów, które

[122] Aby zapoznać się z podobnymi krajowymi systemami świadczeń i scoringiem społecznym, zob. D. Hadwick & S. Lan, „Lessons to be learn from the Dutch child allowance scandal: A comparative review of algorithmic governance by tax Administration in the Netherlands, France and Germany" [Wnioski, jakie należy wyciągnąć ze skandalu związanego z holenderskim zasiłkiem na opiekę nad dzieckiem: przegląd porównawczy zarządzania algorytmicznego przez administracje podatkowe w Niderlandach, Francji i Niemczech] (2021) World Tax Journal, tom 13, numer 4. Familiales (CNAF).

nie mają wyraźnego związku z celem oceny, takich jak stan cywilny, dane dotyczące zdrowia w odniesieniu do chorób przewlekłych, uzależnienia itp.[123]

Te niedopuszczalne praktyki scoringu można odróżnić od zgodnych z prawem praktyk, w ramach których ocenia się osoby w konkretnym celu zgodnie z prawem Unii i prawem krajowym, w szczególności gdy w takich przepisach, zgodnie z prawem UE, określono dane uznane za istotne i niezbędne do celów oceny (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

Scenariusz 2: niekorzystne lub szkodliwe traktowanie nieproporcjonalne do zachowania społecznego

167) Innym alternatywnym scenariuszem na podstawie art. 5 ust. 1 lit. c) ppkt (ii) aktu w sprawie sztucznej inteligencji, w którym scoring oparty na AI może być zakazany, jest sytuacja, w której traktowanie wynikające ze scoringu jest nieuzasadnione lub nieproporcjonalne do wagi zachowania społecznego. Dotkliwość skutków i ingerencja w prawa podstawowe danej osoby wynikająca ze scoringu społecznego w porównaniu z wagą zachowania społecznego danej osoby powinny określać, czy takie traktowanie jest nieproporcjonalne do zamierzonego zgodnego z prawem celu, z uwzględnieniem ogólnej zasady proporcjonalności. Wymaga to indywidualnej oceny, która powinna uwzględniać wszystkie istotne okoliczności sprawy, a także ogólne względy etyczne i zasady uczciwości i sprawiedliwości społecznej związane z oceną zachowań społecznych i proporcjonalnością niekorzystnego traktowania. Traktowanie może być również „nieuzasadnione", jeśli na przykład nie ma uzasadnionego celu. W ramach tej oceny istotne mogą być również sektorowe przepisy unijne lub krajowe określające szczegółowe kryteria i procedury regulujące takie potencjalne szkodliwe lub niekorzystne traktowanie.

Przykłady nieuzasadnionego lub nieproporcjonalnego traktowania w porównaniu z zachowaniem społecznym, zakazanego na mocy art. 5 ust. 1 lit. c) ppkt (ii) aktu w sprawie sztucznej inteligencji

– Agencja publiczna wykorzystuje system AI do profilowania rodzin w celu wczesnego wykrywania dzieci narażonych na ryzyko na podstawie kryteriów takich jak zdrowie psychiczne rodziców i bezrobocie, ale również na podstawie pochodzących z wielu kontekstów informacji na temat zachowań społecznych rodziców. W oparciu o uzyskany wynik wyznacza się rodziny, które będą poddane kontroli, a dzieci uznane za „narażone na ryzyko" odbiera się rodzinom, w tym w przypadku drobnych wykroczeń ze strony rodziców, takich jak sporadyczne opuszczanie wizyt lekarskich lub otrzymywanie mandatów drogowych.

– Gmina wykorzystuje system AI do oceny wiarygodności mieszkańców na podstawie wielu punktów danych związanych z ich zachowaniem społecznym w różnych kontekstach. Uzyskany wynik dla mieszkańców uznanych za „mniej wiarygodnych"

[123] Podobny system zastosowano w Polsce w odniesieniu do systemu „Profilowanie bezrobotnych", który został porzucony po uznaniu go za niezgodny z konstytucją. Zob. Szymielewicz, *Profiling the unemployed in Poland.: Social and Political Implications of Algorithmic Decision Making* [Profilowanie bezrobotnych w Polsce: społeczne i polityczne skutki algorytmicznego podejmowania decyzji], Fundacja Panoptykon, 2015, s 18.

wykorzystuje się do umieszczenia ich na czarnej liście, tj. wycofania korzyści publicznych, zastosowania innych poważnych środków o charakterze kary oraz zwiększonej kontroli lub nadzoru. Wśród czynników branych pod uwagę w ocenie są niewystarczające działania w zakresie wolontariatu i drobne niewłaściwe zachowania, takie jak nieterminowy zwrot książek do biblioteki, pozostawianie śmieci na ulicy poza dniem ich odbioru i opóźnienia w płaceniu podatków lokalnych.

Te niedopuszczalne praktyki scoringu można odróżnić od zgodnych z prawem praktyk, w ramach których ocenia się osoby w uzasadnionym konkretnym celu zgodnie z prawem Unii i prawem krajowym, w szczególności gdy przepisy te zapewniają, aby szkodliwe lub niekorzystne traktowanie było uzasadnione i proporcjonalne do zachowań społecznych (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

168) Oba warunki alternatywne, o których mowa w art. 5 ust. 1 lit. c) ppkt (i) oraz (ii) aktu w sprawie sztucznej inteligencji, mogą być również spełnione jednocześnie.

Przykłady nieuzasadnionego lub nieproporcjonalnego traktowania na podstawie art. 5 ust. 1 lit. c) ppkt (i) oraz (ii) aktu w sprawie sztucznej inteligencji:

– Organ podatkowy wykorzystuje system AI do wykrywania oszustw związanych ze świadczeniami na dziecko poprzez profilowanie i przypisywanie beneficjentów podejrzewanych o oszustwo do kategorii takich jak „umyślny zamiar/rażące niedbalstwo”, stosując takie kryteria jak niski dochód, podwójne obywatelstwo, zachowania społeczne itp. W oparciu o ocenę ryzyka sprawdza się dokumentację beneficjenta, a w wielu przypadkach, po wstrzymaniu ich świadczenia z tytułu opieki nad dzieckiem, otrzymują oni powiadomienie o konieczności zwrotu otrzymanych świadczeń i nie kwalifikują się już do zastosowania standardowych rozwiązań w zakresie windykacji. Taki scoring sprawia, że wiele rodzin jest poważnie zadłużonych, i prowadzi do niesprawiedliwego, dyskryminującego i szkodliwego traktowania osób i grup osób[124], co sprawia, że wiele rodzin znajduje się w trudnej sytuacji finansowej.

– Organ publiczny wykorzystuje system AI do zapobiegania oszustwom w procesie przyznawania studentom dotacji na zakwaterowanie, który to system jako czynniki pozwalające określić ryzyko nadużyć wykorzystuje wskaźniki takie jak aktywność w internecie, status rodzinny czy poziom wykształcenia beneficjentów, co nie wydaje się istotne ani uzasadnione.

– Rząd wprowadza kompleksowy system oparty na sztucznej inteligencji, który monitoruje i ocenia obywateli na podstawie ich zachowania w różnych aspektach życia, takich jak interakcje społeczne, aktywność w internecie, nawyki zakupowe i terminowość w płaceniu rachunków. Osoby o niższych wynikach mają ograniczony dostęp do usług publicznych, wyższe stopy oprocentowania kredytów oraz trudności w podróżowaniu, wynajmowaniu mieszkań, a nawet w znalezieniu pracy. System ten

[124] Zob. podobny przykład skandalu w Niderlandach dotyczącego niderlandzkich świadczeń z tytułu opieki nad dzieckiem, opisany w artykule *Belastingdienst treft 232 gezinnen met onevenredig harde actie*, 27.11.2019 (w języku niderlandzkim). W 2020 r. sąd niderlandzki orzekł, że „Systeem Risico Indicatie (SyRi) jest niezgodny z prawem. Zob. również: Geen powerplay maar fair play. Onevenredig harde aanpak van 232 gezinnen met kinderopvangtoeslag, 2017, s. 32.

prowadzi do nadmiernego nadzoru nad osobami fizycznymi i szkodliwego traktowania w kontekstach niezwiązanych z zachowaniami społecznymi wykorzystywanymi do określenia scoringu społecznego (np. aktywność w mediach społecznościowych wpływa na możliwości zatrudnienia), a jednocześnie nakłada nadmierne kary za drobne naruszenia (np. niekorzystne warunki społeczne i finansowe za stosunkowo drobne wykroczenia).

Te niedopuszczalne praktyki scoringu społecznego można odróżnić od zgodnych z prawem praktyk oceny osób do uzasadnionych konkretnych celów, które to praktyki nie spełniają wspomnianych warunków i są zgodne z prawem Unii i prawem krajowym, w szczególności gdy przepisy te zapewniają, aby szkodliwe lub niekorzystne traktowanie było uzasadnione i proporcjonalne, oraz gdy wykorzystywane są dane z powiązanych kontekstów społecznych (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

169) Zakaz, o którym mowa w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, może również obejmować przypadki, w których niektórym osobom lub grupom osób przyznaje się nagrody lub preferencyjnie się je traktuje, ponieważ wiąże się to z mniej korzystnym traktowaniem innych osób (np. w przypadku programów wspierających zatrudnienie, pierwszeństwa w przydzieleniu mieszkań lub przesiedleń).

4.2.3. Bez względu na to, czy są dostarczane lub wykorzystywane przez osoby publiczne lub prywatne

170) Jak już zauważono, art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji zakazuje niedopuszczalnych praktyk scoringu społecznego wspomaganych sztuczną inteligencją, niezależnie od tego, czy system AI lub scoring są dostarczane lub wykorzystywane przez osoby publiczne lub prywatne. Chociaż scoring w sektorze publicznym może mieć bardzo poważne konsekwencje dla obywateli ze względu na brak równowagi sił i zależność od usług publicznych, podobnie szkodliwe skutki mogą wystąpić również w sektorze prywatnym, w którym przedsiębiorstwa i inne podmioty coraz częściej stosują praktyki scoringu.

Na przykład:

– zakład ubezpieczeń gromadzi od banku informacje o wydatkach i inne informacje finansowe, które nie są związane z określeniem kwalifikowalności kandydatów do ubezpieczenia na życie i które służą do ustalenia ceny składki, którą należy zapłacić za takie ubezpieczenie. System AI analizuje te informacje i na tej podstawie zaleca, czy odmówić zawarcia umowy lub ustalić wyższe składki na ubezpieczenie na życie dla konkretnej osoby lub grupy klientów;

– prywatna agencja kredytowa korzysta z systemu AI w celu określenia zdolności kredytowej osób i podjęcia decyzji, czy dana osoba powinna uzyskać kredyt na zakup mieszkania w oparciu o niepowiązane cechy osobiste.

Te niedopuszczalne praktyki scoringu społecznego można odróżnić od zgodnych z prawem praktyk oceny osób do konkretnych uzasadnionych celów, które to praktyki

nie spełniają wspomnianych warunków i są zgodne z prawem Unii i prawem krajowym, w szczególności gdy przepisy te zapewniają, aby szkodliwe lub niekorzystne traktowanie było uzasadnione i proporcjonalne, oraz gdy wykorzystywane są dane z powiązanych kontekstów społecznych (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

171) W przypadku kontroli przeprowadzanych przez właściwe organy nadzoru rynku to dostawca i podmiot stosujący system AI, każdy w ramach ich obowiązków, muszą wykazać, że dana praktyka w zakresie AI jest zgodna z prawem i uzasadniona, w tym poprzez zapewnienie przejrzystości funkcjonowania systemu AI oraz informacji na temat rodzajów danych i źródeł danych, przy zapewnieniu, aby do celów oceny lub klasyfikacji były przetwarzane wyłącznie dane związane z kontekstem społecznym, w którym wykorzystuje się scoring, system działał zgodnie z przeznaczeniem, a wszelkie wynikające z tego szkodliwe lub niekorzystne traktowanie było uzasadnione i proporcjonalne do zachowania społecznego. Zgodność z mającymi zastosowanie przepisami oraz odpowiednie i proporcjonalne zabezpieczenia wbudowane w system i stosowane podczas jego działania pomogą zapewnić, by zakaz nie miał zastosowania, a jednocześnie umożliwią wykorzystywanie systemów AI na potrzeby oceny lub klasyfikacji osób do uzasadnionych i przynoszących korzyści celów (np. w celu poprawy skuteczności procesów, jakości usług, bezpieczeństwa itp.) (zob. sekcja 4.3 dotycząca systemów nieobjętych zakresem).

172) Zgodność z wymogami dotyczącymi systemów AI wysokiego ryzyka (np. w obszarze podstawowych usług i świadczeń publicznych, punktowej oceny kredytowej i oceny zdolności kredytowej, migracji itp.) może również pomóc w zapewnieniu, aby systemy AI wykorzystywane do celów oceny i klasyfikacji w tych obszarach wysokiego ryzyka nie stanowiły niedopuszczalnych praktyk scoringu społecznego, co dostawcy i podmioty stosujące powinni uwzględnić przy wdrażaniu swoich odpowiednich obowiązków (np. w zakresie zarządzania ryzykiem, przejrzystości, zarządzania danymi, oceny skutków dla praw podstawowych, nadzoru ze strony człowieka, monitorowania itp.).

4.3. Systemy nieobjęte zakresem

173) Zakaz określony w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji ma zastosowanie wyłącznie do scoringu osób fizycznych lub grup osób, co wyklucza co do zasady scoring podmiotów prawnych, w przypadku których ocena nie opiera się na cechach osobistych lub cechach osobowości lub zachowaniach społecznych osób fizycznych, nawet jeżeli w niektórych przypadkach taki scoring może mieć pośredni wpływ na osoby fizyczne (np. na wszystkich obywateli gminy w przypadku przydziału budżetu). Jeżeli jednak podmioty prawne zostały ocenione na podstawie ogólnego wyniku, który agreguje ocenę lub klasyfikację grupy osób fizycznych na podstawie ich zachowania społecznego lub cech osobistych lub cech osobowości, a wynik ten ma bezpośredni wpływ na te osoby (np. wszystkich pracowników przedsiębiorstwa, uczniów w konkretnej szkole, których zachowanie zostało ocenione), praktyka może wchodzić w zakres art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, jeżeli

spełnione są wszystkie pozostałe warunki. Będzie to zależało od indywidualnej oceny każdego przypadku.

174) Scoring społeczny oparty na sztucznej inteligencji stanowiący „wartość prawdopodobieństwa" i prognozę należy również odróżnić od indywidualnych ratingów użytkowników, którzy oceniają jakość usługi (na przykład w przypadku oceny kierowcy na internetowej platformie wspólnego użytkowania samochodów osobowych lub wynajmującego na internetowej platformie zakwaterowania). Takie ratingi są zwykłą agregacją indywidualnych wyników uzyskanych przez człowieka, które niekoniecznie wiążą się ze sztuczną inteligencją, chyba że dane te są łączone z innymi informacjami i analizowane przez system AI w celu oceny lub klasyfikowania osób spełniających wszystkie warunki określone w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

175) Ponadto scoring osób fizycznych nie jest przez zakazany ogólnie, lecz jedynie w ograniczonych przypadkach, w których wszystkie warunki określone w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji są spełnione łącznie, jak przeanalizowano powyżej. W motywie 31 aktu w sprawie sztucznej inteligencji wspomina się w szczególności, że „zakaz ten nie powinien mieć wpływu na zgodne z prawem praktyki oceny osób fizycznych, które są stosowane w konkretnym celu zgodnie z prawem Unii i prawem krajowym". Na przykład punktowa ocena kredytowa i punktowa ocena ryzyka kredytowego oraz ubezpieczeniowego są zasadniczymi aspektami usług świadczonych przez instytucje finansowe i ubezpieczeniowe. Takie praktyki, a także inne zgodne z prawem praktyki (np. mające na celu poprawę jakości i efektywności usług, zapewnienie skuteczniejszego rozpatrywania skarg, przeprowadzanie konkretnych ocen pracowników, zapobieganie nadużyciom finansowym i ich wykrywanie, ściganie przestępstw lub ocenę zachowania użytkowników na platformach internetowych) nie są same w sobie zakazane, jeżeli są zgodne z prawem i podejmowane zgodnie z aktem w sprawie sztucznej inteligencji oraz innymi mającymi zastosowanie przepisami Unii i prawem krajowym, które musi być zgodne z prawem Unii.

176) Innymi słowy, systemy AI, które oceniają lub klasyfikują osoby fizyczne do celów generowania scoringu społecznego w sposób zgodny z prawem i w konkretnym celu w powiązanym kontekście, takim jak ten, w którym zgromadzono dane osobowe wykorzystane do scoringu, nie są zakazane, pod warunkiem że wszelkie szkodliwe lub niekorzystne traktowanie wynikające z zastosowania oceny jest uzasadnione i proporcjonalne do wagi zachowania społecznego[125].

177) Zgodność z sektorowymi przepisami Unii, takimi jak przepisy w dziedzinie punktowej oceny kredytowej, przeciwdziałania praniu pieniędzy itp., określającymi rodzaj danych, które mogą być wykorzystywane jako istotne i niezbędne do konkretnego uzasadnionego celu oceny, oraz zapewniającymi, aby traktowanie było uzasadnione i proporcjonalne do zachowań społecznych, może zatem zapewnić, że praktyka

[125] Motyw 31 aktu w sprawie sztucznej inteligencji.

w zakresie AI nie wchodziła w zakres zakazu określonego w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

Przykłady uzasadnionych praktyk scoringu zgodnych z prawem Unii i prawem krajowym, które nie są objęte zakresem art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji:

– Finansowe systemy punktowej oceny kredytowej wykorzystywane przez kredytodawców lub agencje informacji kredytowej do oceny zdolności kredytowej klienta lub jego pozostałych długów, dokonywania punktowej oceny kredytowej lub ustalania oceny zdolności kredytowej na podstawie dochodów i wydatków klienta oraz innych okoliczności finansowych i ekonomicznych nie są objęte zakresem art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, jeżeli są odpowiednie dla uzasadnionego celu punktowej oceny kredytowej i zgodne z przepisami dotyczącymi ochrony konsumentów[126] określającymi rodzaj danych i niezbędne zabezpieczenia w celu zapewnienia sprawiedliwego traktowania konsumentów w ocenach zdolności kredytowej.

– Przedsiębiorstwa mają uzasadniony interes w ocenie klientów pod kątem nadużyć finansowych, a zakaz ten nie ma wpływu na te praktyki, jeżeli ocena opiera się na odpowiednich danych, takich jak zachowanie dotyczące transakcji i metadane w kontekście usług, dotychczasowa historia i inne czynniki pochodzące ze źródeł, które są obiektywnie istotne dla określenia ryzyka oszustwa, oraz jeżeli szkodliwe traktowanie w wyniku oszukańczego zachowania jest uzasadnione i proporcjonalne.

– Informacje zebrane za pomocą urządzeń telematycznych wykorzystywanych przez ubezpieczyciela oferującego taryfy oparte na telematyce w odniesieniu do ryzykownego stylu jazdy ubezpieczającego, z których to informacji wynika, że kierowca przekracza prędkość lub nie przestrzega zasad bezpiecznej jazdy, mogą zostać wykorzystane do zwiększenia składki tego ubezpieczającego ze względu na wyższe ryzyko wypadku spowodowanego tym zachowaniem kierowcy, pod warunkiem że wzrost składki jest proporcjonalny do ryzykownego zachowania kierowcy.

– Gromadzenie i przetwarzanie danych, które są istotne i niezbędne do osiągnięcia zamierzonego uzasadnionego celu systemów AI (np. danych dotyczących zdrowia i schizofrenii gromadzonych z różnych źródeł w celu diagnozowania pacjentów) nie jest objęte zakresem art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, w szczególności dlatego, że przetwarzanie dotyczy istotnych i niezbędnych danych i zazwyczaj nie pociąga za sobą nieuzasadnionego szkodliwego lub niekorzystnego traktowania niektórych osób fizycznych.

– Platformy internetowe profilujące użytkowników ze względów bezpieczeństwa usług na podstawie danych, które są istotne dla kontekstu i celu oceny, nie są objęte zakresem art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, jeżeli ocena nie

[126] Zob. w szczególności dyrektywa (UE) 2023/2225 z dnia 18 października 2023 r. w sprawie umów o kredyt konsumencki i uchylająca dyrektywę 2008/48/WE oraz wytyczne Europejskiego Urzędu Nadzoru Bankowego dotyczące udzielania i monitorowania kredytów z dnia 29 maja 2020 r., EBA/GL/2020/06.

prowadzi do szkodliwego traktowania, które jest nieproporcjonalne do wagi niewłaściwego zachowania użytkownika.

– Ukierunkowana reklama handlowa wspomagana sztuczną inteligencją nie wchodzi w zakres stosowania, jeżeli opiera się na odpowiednich danych (np. preferencjach użytkowników), odbywa się zgodnie z prawem Unii dotyczącym ochrony konsumentów, ochrony danych i usług cyfrowych i nie prowadzi do szkodliwego lub niekorzystnego traktowania nieproporcjonalnego do wagi zachowania społecznego użytkownika (np. wyzysku i nieuczciwego różnicowania cen).

– Zakaz nie ma wpływu na systemy AI wykorzystujące dane zgromadzone w obozach dla uchodźców (np. dane dotyczące zgodności behawioralnej) na potrzeby decyzji dotyczących przesiedlenia lub zatrudnienia, biorąc pod uwagę, że dane te są istotne do celów oceny, oraz pod warunkiem przestrzegania procedur przewidzianych w mającym zastosowanie unijnym prawie migracyjnym z myślą o zapewnieniu uzasadnionego i proporcjonalnego traktowania.

– W zakres art. 5 ust. 1 lit. c) ustawy o sztucznej inteligencji nie wchodzi wspomagany sztuczną inteligencją scoring dokonywany przez internetową platformę zakupową, która oferuje użytkownikom z długą historią zakupów i niskim wskaźnikiem zwrotów produktów przywileje, takie jak szybszy proces dokonywania zwrotów lub zwroty środków bez zwrotu produktu, biorąc pod uwagę, że korzyści te są uzasadnione i proporcjonalne do nagradzania pozytywnych zachowań, a inni użytkownicy nadal mają dostęp do standardowego procesu zwrotu.

– Oparte na AI ocena i scoring osób fizycznych dokonywane przez policję i inne organy ścigania, które gromadzą dane na temat zachowań społecznych osób z wielu kontekstów, nie są objęte zakresem art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, jeżeli dane te są istotne dla konkretnych celów zapobiegania przestępstwom, ich wykrywania, ścigania i karania za nie oraz gdy szkodliwe traktowanie jest uzasadnione i proporcjonalne zgodnie z materialnym i proceduralnym unijnym i krajowym prawem karnym i prawem policyjnym. W tym kontekście należy również rozważyć zakaz zawarty w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, nakładający dodatkowe i bardziej szczegółowe warunki dotyczące ocen ryzyka i prognoz ryzyka wspomaganych sztuczną inteligencją, które nie mogą opierać się wyłącznie na profilowaniu lub ocenie cech osobowości (zob. sekcja 5).

4.4. Wzajemne powiązania z innymi aktami prawnymi Unii

178) Dostawcy i podmioty stosujące powinni dokładnie ocenić, czy inne mające zastosowanie przepisy unijne i krajowe mają zastosowanie do każdego konkretnego systemu AI wykorzystywanego w ich działalności na potrzeby scoringu, w szczególności czy istnieją bardziej szczegółowe przepisy ściśle regulujące rodzaje danych, które mogą być wykorzystywane jako istotne i niezbędne do konkretnych celów oceny, oraz czy istnieją bardziej szczegółowe przepisy i procedury zapewniające uzasadnione i sprawiedliwe traktowanie.

179) Wspomagane sztuczną inteligencją praktyki scoringu społecznego stosowane przez podmioty prywatne działające jako przedsiębiorcy w stosunkach między przedsiębiorstwami a konsumentami mogą również stanowić naruszenie unijnego prawa ochrony konsumentów, tj. dyrektywy 2005/29/WE dotyczącej nieuczciwych praktyk handlowych stosowanych przez przedsiębiorstwa wobec konsumentów. Dyrektywa o nieuczciwych praktykach handlowych zakazuje praktyk handlowych, jeżeli są one sprzeczne z wymogami staranności zawodowej i znacząco zmieniają lub mogą w znacząco zmienić zachowanie gospodarcze przeciętnego konsumenta lub przeciętnego członka grupy w odniesieniu do produktu (art. 5 dyrektywy o nieuczciwych praktykach handlowych). Praktyki scoringu można również uznać za wprowadzające w błąd (art. 6–7 dyrektywy o nieuczciwych praktykach handlowych), z zastrzeżeniem indywidualnej oceny wpływu praktyki handlowej na decyzję konsumenta dotyczącą transakcji.

180) Scoring społeczny dokonywany przez podmioty publiczne lub prywatne może również stanowić naruszenie unijnych przepisów o ochronie danych, na przykład w odniesieniu do podstawy prawnej przetwarzania (zgodność z prawem), zasad ochrony danych (np. minimalizacja danych i konieczność, uczciwość oraz przejrzystość) oraz wszelkich innych obowiązków, w tym, w stosownych przypadkach, przepisów dotyczących wyłącznie zautomatyzowanego podejmowania indywidualnych decyzji.

181) W przypadku gdy ocena lub klasyfikacja opiera się na jednej z cech chronionych przed dyskryminacją (np. wiek, religia, pochodzenie rasowe lub etniczne, płeć itp.) lub prowadzi bezpośrednio lub pośrednio do dyskryminacji tych grup, taka praktyka będzie również podlegać unijnemu prawu o niedyskryminacji.

182) W tym kontekście znaczenie może mieć również dyrektywa w sprawie kredytów konsumenckich (UE) 2023/2225[127]. Art. 18 ust. 3 dyrektywy w sprawie kredytów konsumenckich zawiera wymóg, aby ocenę zdolności kredytowej przeprowadzano na podstawie stosownych i dokładnych informacji o dochodach i wydatkach konsumenta oraz o innych warunkach finansowych i ekonomicznych, które to informacje są konieczne i proporcjonalne do charakteru, okresu obowiązywania i wartości kredytu oraz związanych z nim ryzyk dla konsumenta. Informacje te mogą obejmować dowody potwierdzające dochody lub dowody istnienia innych źródeł spłaty, informacje o aktywach i zobowiązaniach finansowych lub informacje o innych zobowiązaniach finansowych. W dyrektywie w sprawie kredytów konsumenckich wyraźnie zakazano uwzględniania szczególnych kategorii danych osobowych w informacjach i uzyskiwania informacji z sieci społecznościowych. W wytycznych Europejskiego Urzędu Nadzoru Bankowego dotyczących udzielania i monitorowania kredytów[128] doprecyzowano, jakie informacje są istotne do celów oceny zdolności kredytowej. To określenie w przepisach sektorowych rodzajów danych wykorzystywanych na potrzeby konkretnej oceny stanowi istotne względy, które należy wziąć pod uwagę przy

[127] Dyrektywa Parlamentu Europejskiego i Rady 2008/48/WE z dnia 23 kwietnia 2008 r. w sprawie umów o kredyt konsumencki oraz uchylająca dyrektywę Rady 87/102/EWG, Dz.U. L 133 z 22.5.2008, s. 66.

[128] Europejski Urząd Nadzoru Bankowego, Wytyczne dotyczące udzielania i monitorowania kredytów z 29 maja 2020 r., EBA/GL/2020/06.

ustalaniu, czy dana praktyka wchodzi w zakres zakazu określonego w art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji.

183) Podobnie systemy AI wykorzystywane do oceny i klasyfikacji osób do celów przeciwdziałania praniu pieniędzy i finansowaniu terroryzmu powinny być również zgodne z odpowiednimi przepisami Unii dotyczącymi tych kwestii.

5. ART. 5 UST. 1 LIT. D) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – INDYWIDUALNA OCENA RYZYKA I PRZEWIDYWANIE RYZYKA POPEŁNIENIA PRZESTĘPSTWA

184) W art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji zakazano systemów AI służących ocenie lub przewidywaniu ryzyka popełnienia przestępstwa przez osobę fizyczną wyłącznie na podstawie profilowania lub oceny cech osobowości i cech charakterystycznych.

185) W ostatnim zdaniu tego przepisu wskazano, że zakaz ten nie ma zastosowania do systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny udziału danej osoby w działalności przestępczej, która to ocena opiera się już na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą Takie systemy AI, które nie są objęte zakresem zakazu i które są przeznaczone do wykorzystywania przez organy ścigania lub w ich imieniu lub przez instytucje, organy lub jednostki organizacyjne Unii w ramach wsparcia udzielanego organom ścigania do oceny ryzyka, że osoba fizyczna popełni lub ponownie popełni przestępstwo, przeprowadzanej nie tylko na podstawie profilowania lub oceny cech osobowości i cech charakterystycznych lub uprzedniego zachowania przestępczego, klasyfikuje się jako systemy AI wysokiego ryzyka (pkt 6 lit. d) załącznika III do aktu w sprawie sztucznej inteligencji) i muszą one spełniać wszystkie odpowiednie wymogi i obowiązki wynikające z aktu w sprawie sztucznej inteligencji.

5.1. Uzasadnienie i cele

186) W motywie 42 aktu w sprawie sztucznej inteligencji wyjaśniono kontekst i uzasadnienie zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, a mianowicie, że osoby fizyczne powinny być oceniane na podstawie ich faktycznego zachowania, a nie na podstawie zachowania przewidywanego przez sztuczną inteligencję wyłącznie na podstawie ich profilowania, cech osobowości lub cech charakterystycznych.

5.2. Główne pojęcia i elementy zakazu

Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji stanowi:

Zakazuje się następujących praktyk w zakresie AI:

d) wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu lub wykorzystywania systemu AI do przeprowadzania ocen ryzyka w odniesieniu do osób fizycznych, by ocenić lub przewidzieć ryzyko popełnienia przestępstwa przez osobę fizyczną, wyłącznie na podstawie profilowania osoby fizycznej lub oceny jej cech

osobowości i cech charakterystycznych; zakaz ten nie ma zastosowania do systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny udziału danej osoby w działalności przestępczej, która to ocena opiera się już na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą;

187) Aby zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi zostać spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku w tym konkretnym celu" lub „wykorzystywaniu" systemu AI;

(ii) system AI musi przeprowadzać oceny ryzyka, które oceniają lub przewidują ryzyko popełnienia przestępstwa przez osobę fizyczną;

(iii) ocena lub prognoza ryzyka muszą opierać się wyłącznie na jednym z następujących elementów lub obu następujących elementach:

a) profilowaniu osoby fizycznej,

b) ocenie jej cech osobowości i cech charakterystycznych.

188) Aby zakaz miał zastosowanie, muszą być spełnione jednocześnie wszystkie trzy warunki. Pierwszy warunek, tj. wprowadzanie do obrotu, oddawanie do użytku lub wykorzystywanie systemu AI, przeanalizowano już w sekcji 2.3. Zakaz ten ma zatem zastosowanie zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów AI do obrotu, oddawać ich do użytku ani wykorzystywać do tego konkretnego celu. Pozostałe dwa warunki, które muszą zostać spełnione, aby zakaz miał zastosowanie, zostały przeanalizowane poniżej.

5.2.1. Ocena ryzyka lub przewidywanie prawdopodobieństwa popełnienia przestępstwa przez osobę

189) Oceny ryzyka mające na celu ocenę lub przewidywanie ryzyka popełnienia przestępstwa przez osobę są często określane jako indywidualne „przewidywanie przestępczości" lub „prognozowanie przestępczości". Chociaż nie istnieje powszechnie uzgodniona definicja „przewidywania przestępczości" czy „prognozowania przestępczości"[129], terminy te odnoszą się ogólnie do różnych zaawansowanych technologii AI i metod analitycznych stosowanych do dużej ilości danych, często historycznych (w tym danych społeczno-ekonomicznych, ale również danych z rejestrów policyjnych itp.), które w połączeniu z teoriami kryminologicznymi są wykorzystywane do prognozowania przestępczości jako podstawy do opracowywania strategii policji i organów ścigania oraz działań na rzecz zwalczania, kontroli i zapobiegania przestępczości[130].

[129] Zob. na przykład systemy wymienione w podręczniku Agencji Praw Podstawowych Unii Europejskiej, takie jak system informowania o przestępczości (CAS) w Niderlandach i Precobs w Niemczech i Szwajcarii, Podręcznik, 2018, s. 138. *Zapobieganie niezgodnemu z prawem profilowaniu obecnie i w przyszłości:przewodnik*, Podręcznik, 2018, s. 138.

[130] Zob. Europol, AI and policing The benefits and challenges of artificial intelligence for law enforcement, An Observatory Report from the Europol Innovation Lab [Sztuczna inteligencja i działania policyjne – korzyści i wyzwania związane ze sztuczną inteligencją dla

190) Systemy AI do przewidywania przestępczości identyfikują wzorce w danych historycznych, łącząc wskaźniki z prawdopodobieństwem wystąpienia przestępstwa, a następnie generują punktową ocenę ryzyka jako wyniki predykcyjne. Systemy takie mogą być na przykład wykorzystywane do planowania policyjnych grup zadaniowych, do monitorowania sytuacji wysokiego ryzyka oraz do przeprowadzania kontroli osób, co do których przewiduje się, że prawdopodobnie (ponownie) popełnią przestępstwo. Takie systemy stwarzają możliwości dla organów ścigania, zwłaszcza tych, które dysponują ograniczonymi zasobami, zwiększają skuteczność i umożliwiają proaktywne podejście do wykrywania przestępstw, zapobiegania im i ich przewidywania[131]. Takie wykorzystanie danych historycznych dotyczących popełnionych przestępstw w celu przewidywania przyszłych zachowań innych osób może jednak utrwalać, a nawet wzmacniać uprzedzenia i może prowadzić do „pominięcia" kluczowych okoliczności indywidualnych, jeżeli okoliczności te nie są częścią zbioru danych ani nie są uwzględniane w algorytmach, na których opiera się dany system AI. Może to również podważyć zaufanie społeczeństwa do organów ścigania i ogólnie do systemu wymiaru sprawiedliwości[132].

191) Takie oceny i prognozy ryzyka są co do zasady perspektywiczne i dotyczą przyszłych przestępstw (jeszcze niepopełnionych) lub przestępstw, co do których uważa się, że istnieje ryzyko, iż zostaną popełnione, w tym w przypadku usiłowania popełnienia przestępstwa lub działań przygotowawczych podjętych w celu jego popełnienia[133]. Można je przeprowadzać na każdym etapie działań w zakresie ścigania przestępstw, np. w ramach zapobiegania przestępstwom i ich wykrywania, ale również w trakcie dochodzenia, ścigania i wykonywania kar (w tym gdy organy sądowe oceniają ryzyko ponownego popełnienia przestępstwa, np. w kontekście podejmowania decyzji o zastosowaniu tymczasowego aresztowania), a także w ramach planu resocjalizacji danej osoby po odbyciu kary[134].

192) Zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji nie obejmuje praktyk w zakresie przewidywania przestępczości i oceny ryzyka jako takich. Ma on zastosowanie wyłącznie do systemów AI służących do przeprowadzania ocen ryzyka w celu oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną, w sytuacji gdy spełniony jest również trzeci warunek, o którym mowa powyżej. Ponadto, jak zauważono, zakaz ten nie ma zastosowania w sytuacjach

---

egzekwowania prawa. Sprawozdanie obserwatorium opracowane przez laboratorium innowacji Europolu], 23 września 2024 r. Zob. również F. Yang, „Predictive Policing" [Prognozowanie przestępczości] w *Oxford Research Encyclopedia, Criminology and Criminal Justice*, Oxford University Press, 2019.

131 Na przykład OxRec (Dutch Probation Office, „Reclassering Nederland") Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec) [Zapobieganie ponownemu popełnieniu przestępstwa z użyciem przemocy u więźniów i osób warunkowo zwolnionych: niderlandzkie badanie walidacyjne] – PMC (nih.gov).

132 Zob. na przykład Agencja Praw Podstawowych Unii Europejskiej (8 grudnia 2022 r.) Bias in algorithms – Artificial intelligence and discrimination [Stronniczość algorytmiczna – sztuczna inteligencja a dyskryminacja] Agencja Praw Podstawowych Unii Europejskiej.

133 Zob. w tym względzie motyw 42 aktu w sprawie sztucznej inteligencji, w którym mowa jest o „prawdopodobieństwie popełnienia przestępstwa" oraz o „wystąpieniu faktycznego lub potencjalnego przestępstwa", w czasie teraźniejszym, a nie przeszłym.

134 Na przykład w art. 24 ust. 4 dyrektywy UE 2011/93 w sprawie zwalczania niegodziwego traktowania w celach seksualnych i wykorzystywania seksualnego dzieci oraz pornografii dziecięcej ustanowiono wymóg, aby osoby, wobec których toczy się postępowanie karne lub które zostały skazane za czyny związane z niegodziwym traktowaniem dzieci w celach seksualnych, zostały poddane ocenie zagrożenia, jakie stwarzają, dotyczącego ryzyka ponownego popełnienia przestępstwa.

opisanych w wyraźnym wyłączeniu zawartym w art. 5 ust. 1 lit. d) zdanie ostatnie aktu w sprawie sztucznej inteligencji.

5.2.2. Wyłącznie na podstawie profilowania osoby fizycznej lub oceny jej cech osobowości i cech charakterystycznych

193) Trzecim warunkiem, który musi zostać spełniony, aby zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji miał zastosowanie, jest to, że ocena ryzyka mająca na celu ocenę lub przewidzenie ryzyka popełnienia przestępstwa przez osobę fizyczną musi opierać się wyłącznie na a) profilowaniu tej osoby lub b) na ocenie jej cech osobowości i cech charakterystycznych.

194) Zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji ma zastosowanie niezależnie od tego, czy system AI profiluje lub ocenia cechy osobowości i cechy charakterystyczne jednocześnie tylko jednej osoby fizycznej czy grupy osób fizycznych, ponieważ zakaz ten ma na celu ochronę każdej osoby, w odniesieniu do której przewiduje się lub ocenia ryzyko popełnienia przestępstwa.

a) Profilowanie osoby fizycznej

195) W przeciwieństwie do art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji w art. 5 ust. 1 lit. d) wyraźnie posłużono się terminem „profilowanie". W art. 3 pkt 52 aktu w sprawie sztucznej inteligencji zdefiniowano ten termin poprzez odniesienie do jego definicji zawartej w art. 4 ust. 4 RODO[135]. Jednym z podstawowych elementów pojęcia profilowania jest cel „oceny niektórych czynników osobowych"[136]. W kontekście art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji profilowanie przeprowadza się do celów oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę.

196) W tym kontekście znaczenie może mieć również koncepcja tzw. profilowania grupowego[137]. Pojęcie to odnosi się do tworzenia i zastosowania profilu opisowego dla danej grupy, na przykład kategorii sprawców przestępstw (np. terrorystów, gangsterów itp.) skonstruowanych na podstawie danych historycznych dotyczących wcześniej popełnionych przestępstw przez inne osoby. Te profile grupowe mogą być później wykorzystywane do oceny i przewidywania ryzyka popełnienia podobnych przestępstw przez inne osoby. Jeżeli system AI sporządza prognozę i stosuje taki profil (grupowy) do konkretnej osoby fizycznej, stanowi to profilowanie danej osoby i w związku z tym może wchodzić w zakres zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

---

135 Art. 3 ust. 4 dyrektywy o ochronie danych w sprawach karnych, który ma znaczenie dla zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, definiuje profilowanie w identyczny sposób jak w art. 4 ust. 4 RODO jako „dowolną formę zautomatyzowanego przetwarzania danych osobowych, które polega na wykorzystaniu danych osobowych do oceny niektórych czynników osobowych osoby fizycznej, w szczególności do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby fizycznej, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub przemieszczania się". Tę samą definicję zawarto również w art. 3 ust. 5 rozporządzenia (UE) 2018/1725 w sprawie przetwarzania danych osobowych przez instytucje, organy i jednostki organizacyjne Unii (Dz.U. L 295 z 21.11.2018, s. 39).

136 Zob. również Grupa Robocza Art. 29, *Wytyczne w sprawie zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i profilowania do celów rozporządzenia 2016/679*, WP251rev.01, 6.02.2018, zatwierdzone przez EROD, s. 7. Zob. również Agencja Praw Podstawowych, *Zapobieganie nielegalnemu profilowaniu obecnie i w przyszłości: przewodnik*, Podręcznik, 2018, s.138.

137 Na temat profilowania grupowego zob. np. Agencja Praw Podstawowych, *Zapobieganie nielegalnemu profilowaniu obecnie i w przyszłości: przewodnik*, Podręcznik, 2018 r., s. 21.

b) Ocena cech osobowości i cech charakterystycznych

197) Zakaz ten ma również zastosowanie, jeżeli ocena ryzyka mająca na celu ocenę lub przewidzenie ryzyka popełnienia przestępstwa przez osobę opiera się wyłącznie na ocenie cech osobowości i cech charakterystycznych danej osoby. Taka ocena lub prognoza często wchodzi w zakres pojęcia profilowania, ale można je również postrzegać jako alternatywę, jeżeli nie można stwierdzić profilowania w rozumieniu w art. 4 ust. 4 RODO.

198) Jak wspomniano w sekcji 4.2.1 lit. c), cechy osobowości i cechy charakterystyczne stanowią szeroką kategorię cech charakterystycznych związanych z konkretną osobą fizyczną, dla których nie istnieje powszechnie uzgodniona taksonomia. W motywie 42 aktu w sprawie sztucznej inteligencji podano przykłady cech osobowości i cech charakterystycznych, które mogą być poddawane ocenie w celu przewidywania ryzyka popełnienia przestępstwa przez osobę, takie jak „narodowość, miejsce urodzenia, miejsce zamieszkania, liczba dzieci, poziom zadłużenia lub rodzaj samochodu". Wykaz ten jest jedynie przykładowy i niewyczerpujący.

c) „Wyłącznie"

199) Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji stanowi, że oceny ryzyka objęte tym przepisem są zakazane tylko wtedy, gdy opierają się „wyłącznie" na profilowaniu osoby lub ocenie jej cech osobowości i cech charakterystycznych. Z motywu 42 aktu w sprawie sztucznej inteligencji jasno wynika, że „wyłącznie" powinno mieć zastosowanie zarówno do profilowania, jak i do oceny cech osobowości i cech charakterystycznych.

200) Warunek, zgodnie z którym ocena ryzyka musi opierać się „wyłącznie" na profilowaniu lub ocenie cech osobowości i cech charakterystycznych, może w wielu sytuacjach nie być spełniony.

201) Jak wynika z ostatniego zdania art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, taka sytuacja ma miejsce w każdym razie, gdy system AI jest wykorzystywany do wspierania dokonywanej przez człowieka oceny udziału danej osoby w działalności przestępcze, która to ocena opiera się już na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą. Jak wyjaśniono w motywie 42, w tym kontekście należy wziąć pod uwagę w szczególności, ale niekoniecznie wyłącznie, sytuację, w której istnieje już uzasadnione podejrzenie wobec danej osoby fizycznej. W takich przypadkach zwykle dokonywana jest ocena przez człowieka, która zazwyczaj opiera się na istotnych obiektywnych i weryfikowalnych faktach.

202) Mogą jednak również zaistnieć inne sytuacje, które zawsze będą musiały być oceniane indywidualnie dla każdego przypadku. Z jednej strony użycie terminu „wyłącznie" pozostawia możliwość uwzględnienia w ocenie ryzyka różnych innych elementów, co sprawia, że nie jest ona już oparta na profilowaniu lub ocenie samych cech osobowości lub cech charakterystycznych. Z drugiej strony, aby uniknąć obchodzenia zakazu i zapewnić jego skuteczność, wszelkie takie inne elementy muszą być rzeczywiste, istotne i znaczące, aby mogły uzasadnić wniosek, że zakaz nie ma zastosowania.

Z lektury zakazu zawartego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji w połączeniu z wyłączeniem zawartym w ostatnim zdaniu tego przepisu wynika w szczególności, że istnienie pewnych wcześniej ustalonych obiektywnych i weryfikowalnych faktów może uzasadniać taki wniosek.

Na przykład:

– organ ścigania wykorzystuje system AI do przewidywania zachowań przestępczych w odniesieniu do przestępstw, takich jak terroryzm, wyłącznie na podstawie wieku, narodowości, adresu, rodzaju samochodu i stanu cywilnego. Za sprawą tego systemu przypisuje się osobom większe prawdopodobieństwo popełnienia w przyszłości przestępstw, których jeszcze nie popełniły, wyłącznie na podstawie ich cech osobistych. Można przyjąć, że taki system jest zakazany na mocy art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji;

– krajowe organy podatkowe wykorzystują narzędzie predykcyjne oparte na AI do przeglądu deklaracji podatkowych wszystkich podatników w celu przewidywania potencjalnych przestępstw podatkowych i identyfikowania spraw wymagających dalszego dochodzenia. Odbywa się to wyłącznie na podstawie profilu stworzonego przez system AI, który wykorzystuje do oceny cechy charakterystyczne takie jak podwójne obywatelstwo, miejsce urodzenia, liczba dzieci i nieprzejrzyste zmienne, w szczególności wywnioskowane informacje, które mają charakter predykcyjny, a zatem są nieobiektywne i trudne do zweryfikowania. Taki system będzie zwykle objęty zakazem określonym w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, ponieważ nie istnieje uzasadnione podejrzenie udziału danej osoby w działalności przestępczej lub inne obiektywne i weryfikowalne fakty związane z tą działalnością przestępczą. Jest to również przykład, który wchodzi w zakres scoringu społecznego zakazanego na mocy art. 5 ust. 1 lit. c) aktu w sprawie sztucznej inteligencji, obejmującego niekorzystne traktowanie na podstawie danych z niepowiązanych kontekstów społecznych;

– wydział policji korzysta z narzędzia oceny ryzyka opartego na sztucznej inteligencji, aby ocenić ryzyko popełnienia przez małe dzieci i młodzież „w przyszłości przestępstw z użyciem przemocy i przestępstw przeciwko mieniu”. System ocenia dzieci na podstawie ich relacji z innymi osobami i założonych poziomów ryzyka tych osób, co oznacza, że dzieci mogą być uznane za narażone na większe ryzyko popełnienia przestępstwa tylko dlatego, że są powiązane z inną osobą o ocenie wysokiego ryzyka, taką jak brat lub siostra czy przyjaciel. Na poziom ryzyka przypisany dziecku może mieć również wpływ poziom ryzyka przypisany rodzicom. Skutkiem ocen ryzyka jest „rejestrowanie” tych dzieci przez policję w systemach policyjnych, monitorowanie ich za pomocą dodatkowych inspekcji i kierowanie ich do służb „opieki” nad młodzieżą. Taki system prawdopodobnie będzie również objęty zakazem określonym w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

5.2.3. Wyłączenie systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny opartej na obiektywnych

i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą

203) Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji stanowi w ostatnim zdaniu, że zakaz ten nie ma zastosowania do systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny udziału danej osoby w działalności przestępczej, która to ocena opiera się już na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą. Chociaż, jak wskazano, sytuacja opisana w tym wyraźnym wyłączeniu niekoniecznie jest jedyną sytuacją, w której zakaz nie ma zastosowania, wyraźne uwzględnienie tej sytuacji w przepisie zapewnia pewność prawa poprzez ograniczenie zakresu tego zakazu i wyjaśnienie, że w tej właśnie sytuacji zakaz w żadnym wypadku nie ma zastosowania.

204) W przypadku gdy system jest objęty zakresem wyłączenia i w związku z tym nie jest zakazany, zostanie sklasyfikowany jako system AI wysokiego ryzyka (o którym mowa w pkt 6 lit. d) załącznika III do aktu w sprawie sztucznej inteligencji), jeżeli ma być wykorzystywany przez organy ścigania lub w ich imieniu, i w związku z tym podlega wymogom i zabezpieczeniom, w tym nadzorowi ze strony człowieka (art. 14 i 26 aktu w sprawie sztucznej inteligencji). Wymogi te obejmują powierzenie nadzoru ze strony człowieka osobom o niezbędnych kompetencjach, szkoleniach i uprawnieniach, które to osoby powinny być w stanie właściwie zrozumieć możliwości i ograniczenia systemu AI, prawidłowo interpretować jego wyniki i przeciwdziałać ryzyku błędu automatyzacji. Osoby te powinny posiadać jasne procedury, szkolenia oraz niezbędne kompetencje i uprawnienia do merytorycznej oceny wyników działania systemu AI. W tym konkretnym przypadku taka ocena dokonywana przez człowieka powinna zapewniać, aby wszelkie prognozy lub oceny ryzyka popełnienia przestępstwa dokonane przez AI opierały się na obiektywnych i weryfikowalnych faktach związanych z działalnością przestępczą. Osoby te powinny również interweniować w celu uniknięcia negatywnych konsekwencji lub ryzyka lub zatrzymać wykorzystywanie systemu AI, jeżeli nie działa on zgodnie z przeznaczeniem.

205) Ponadto pojęcie „interwencji człowieka” podlega orzecznictwu TSUE, w szczególności w kontekście wyłącznie zautomatyzowanego podejmowania decyzji przewidujących ryzyko udziału pasażerów korzystających z transportu lotniczego w poważnych przestępstwach. Orzecznictwo to może mieć również znaczenie dla stosowania pojęcia „oceny dokonywanej przez człowieka”, użytego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

W sprawie *Ligue des droits humains*[138] TSUE zbadał legalność stosowania zaawansowanego systemu AI do systematycznego przetwarzania danych dotyczących przelotu pasażera (PNR) w odniesieniu do podróżnych korzystających z transportu lotniczego, aby ocenić prawdopodobieństwo ich zaangażowania w terroryzm i inne poważne przestępstwa.

[138] Wyrok Trybunału (wielka izba) z dnia 21 czerwca 2022 r., *Ligue des droits humains*, C-817/19, ECLI:EU:C:2022:491.

TSUE dokonał wykładni przepisu zawartego w dyrektywie (UE) 2016/681 („dyrektywa PNR”), zgodnie z którym zakazuje się podejmowania negatywnych decyzji prawnych w oparciu wyłącznie o zautomatyzowane przetwarzanie i wymaga się **indywidualnej oceny** i weryfikacji **przez człowieka** w przypadku wszelkich pozytywnych wyników w sposób niezautomatyzowany w celu identyfikacji wyników fałszywie pozytywnych i zapewnienia, aby wyniki były niedyskryminacyjne.

Zdaniem TSUE ta ocena dokonywana przez człowieka, na której muszą opierać się wszelkie wyniki zautomatyzowanego przetwarzania danych PNR, musi uwzględniać obiektywne kryteria w celu oceny, czy pozytywny wynik dotyczy osoby, która może brać udział – w tym konkretnym przypadku – w przestępstwach terrorystycznych lub poważnej przestępczości, oraz w celu zapewnienia niedyskryminacyjnego charakteru zautomatyzowanego przetwarzania.

206) Jeżeli chodzi o treść wyłączenia, jednym z jego głównych elementów jest to, że system AI jest wykorzystywany do wspierania oceny przez człowieka, a nie do samodzielnego dokonywania oceny ryzyka, jak ma to miejsce w sytuacjach objętych zakazem. Jednakże aby wyłączenie to miało zastosowanie, ta ocena dokonywana przez człowieka musi być ponadto oparta na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą.

5.2.4. Zakres, w jakim działania podmiotów prywatnych mogą wchodzić w zakres stosowania

207) Oprócz działalności organów ścigania, które co do zasady są głównymi podmiotami stosującymi systemy przewidywania przestępstw oparte na AI, w niektórych przypadkach również działalność podmiotów prywatnych może być objęta zakazem określonym w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji. Wynika to z faktu, że zgodnie z brzmieniem tego zakazu nie ma on zastosowania wyłącznie do organów ścigania. Co więcej, w przeciwnym razie zakaz można by łatwo obejść, co podważałoby jego skuteczność.

208) W tych okolicznościach można przyjąć, że zakaz ten ma zastosowanie w szczególności wtedy, gdy podmioty prywatne są prawnie zobowiązane do wykonywania władzy publicznej i prerogatyw władzy publicznej w zakresie zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania i ścigania czynów zabronionych lub wykonywania kar[139]. Podmiotom prywatnym może zostać również wyraźnie zlecone w poszczególnych przypadkach działanie w imieniu organów ścigania i przeprowadzanie indywidualnych prognoz ryzyka przestępczości. W takich przypadkach działalność tych podmiotów prywatnych może również wchodzić w zakres zakazu, jeżeli spełnione są mające zastosowanie warunki, a wyłączenie nie ma zastosowania.

Na przykład prywatne przedsiębiorstwo dostarczające oparte na sztucznej inteligencji zaawansowane oprogramowanie na potrzeby analizy przestępczości może zostać poproszone przez organ ścigania o przeanalizowanie dużej ilości danych z wielu źródeł

[139] Zob. definicja organów ścigania w art. 3 pkt 45 aktu w sprawie sztucznej inteligencji.

i baz danych, takich jak rejestry krajowe, transakcje bankowe, dane komunikacyjne, dane geoprzestrzenne itp., aby przewidzieć lub ocenić ryzyko tego, że osoby fizyczne będą potencjalnymi sprawcami przestępstw handlu ludźmi. Jeżeli spełnione są wszystkie kryteria określone w art. 5 ust. 1 lit. d), taki przypadek użycia może być zakazany.

209) Ponadto zakaz ten może mieć zastosowanie do podmiotów prywatnych oceniających lub przewidujących ryzyko popełnienia przestępstwa przez daną osobę, jeżeli jest to obiektywnie konieczne do wypełnienia obowiązku prawnego, któremu podlega ten prywatny operator, dotyczącego oceny lub przewidywania ryzyka popełnienia określonych przestępstw (np. w przypadku przeciwdziałania praniu pieniędzy, finansowaniu terroryzmu).

Na przykład instytucja bankowa jest zobowiązana na mocy unijnych przepisów dotyczących przeciwdziałania praniu pieniędzy[140] do kontrolowania i profilowania klientów pod kątem przestępstw związanych z praniem pieniędzy. Jeżeli bank wykorzystuje system AI do wypełniania swoich obowiązków, powinien opierać się wyłącznie na danych określonych w tych przepisach, przy czym dane te muszą być obiektywne i weryfikowalne w celu zapewnienia, aby w stosunku do osób wskazanych jako podejrzane istniało uzasadnione prawdopodobieństwo popełnienia przestępstw prania pieniędzy. Prognozy muszą również podlegać ocenie i weryfikacji przez człowieka zgodnie z tymi przepisami[141], aby zapewnić rzetelność i stosowność takich ocen. Zgodność z tymi przepisami zapewni, że stosowanie systemu AI do indywidualnego prognozowania przestępstw na potrzeby przeciwdziałania praniu pieniędzy nie będzie wchodziło w zakres zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

210) Biorąc jednak pod uwagę nacisk położony na oceny ryzyka odnoszące się konkretnie i wyłącznie do popełniania przestępstw, co wynika z brzmienia zakazu, a także cel zakazu, jak wyjaśniono w motywie 42, jeżeli podmiot prywatny profiluje klientów w ramach swojej zwykłej działalności gospodarczej na potrzeby zapewnienia bezpieczeństwa lub ochrony swoich interesów finansowych (np. na potrzeby wykrywanie nieprawidłowości finansowych), bez celu oceny lub przewidywania ryzyka popełnienia przez klienta konkretnego przestępstwa, nie należy uznawać, że taka działalność podmiotów prywatnych wchodzi w zakres zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji.

211) Innymi słowy, w przypadku braku powierzenia na mocy prawa podmiotom prywatnym – działającym w imieniu organów ścigania lub podlegającym szczególnym obowiązkom prawnym opisanym powyżej – określonych zadań związanych ze ściganiem przestępstw, wykorzystywania systemów AI do przeprowadzania ocen ryzyka w ramach zwykłej działalności podmiotów prywatnych i w celu ochrony ich własnych interesów prywatnych nie uznaje się za objęte zakazem, przy czym fakt, że

[140] Rozporządzenie (UE) 2024/1624 z dnia 31 maja 2024 r. w sprawie przeciwdziałania praniu pieniędzy.
[141] Art. 20 rozporządzenia (UE) 2024/1624.

te oceny ryzyka mogą odnosić się do ryzyka popełnienia przestępstw, stanowi jedynie czysto przypadkową i drugorzędną okoliczność.

5.3. Systemy nieobjęte zakresem

5.3.1. Prognozy przestępczości oparte na danych dotyczących lokalizacji, geoprzestrzennych lub dotyczących miejsca

212) Prognozy przestępczości oparte na danych dotyczących lokalizacji, geoprzestrzennych lub dotyczących miejsca uwzględniają miejsce lub lokalizację przestępstwa lub prawdopodobieństwo popełnienia przestępstwa na tych obszarach. Co do zasady takie działania policyjne nie wiążą się z oceną konkretnej osoby. W związku z tym nie są one objęte zakresem zakazu.

Przykłady prognoz przestępczości opartych na danych dotyczących lokalizacji, geoprzestrzennych lub dotyczących miejsca:

– oparty na sztucznej inteligencji predykcyjny system nadzoru policyjnego ocenia prawdopodobieństwo popełnienia przestępstwa na różnych obszarach w mieście w oparciu o wcześniejsze wskaźniki przestępczości na danym obszarze i inne informacje pomocnicze takie jak mapy ulic, aby zwrócić uwagę na podwyższone ryzyko wystąpienia określonych rodzajów przestępstw, np. włamań, przestępstw z użyciem noża itp., i pomóc organom ścigania określić, gdzie można rozmieścić mniej patroli, a gdzie należy wysyłać ich więcej, aby utrudniać lub uniemożliwiać działalność przestępczą;

– organ celny wykorzystuje oparte na AI narzędzia analizy ryzyka do przewidywania prawdopodobieństwa lokalizacji środków odurzających lub nielegalnych towarów, na przykład na podstawie znanych szlaków przemytu;

– wydział policji wykorzystuje systemy oparte na sztucznej inteligencji do wykrywania i lokalizowania strzałów z broni w czasie rzeczywistym. System wykorzystuje czujniki akustyczne na obszarach miejskich do identyfikacji odgłosów wystrzałów i ustalania ich lokalizacji, dostarczając funkcjonariuszom użytecznych danych pomocnych w wykrywaniu przestępstw i prowadzeniu dochodzeń.

213) Może jednak nie zawsze być jasne, w jaki sposób odróżnić systemy prognozowania oparte na danych dotyczących lokalizacji od indywidualnych systemów prognozowania oceniających ryzyko popełnienia przestępstwa przez daną osobę. W zakresie, w jakim system AI dokonuje prognozowania przestępczości w oparciu o dane dotyczące lokalizacji, a następnie uwzględnia ocenę ryzyka dla określonej lokalizacji jako jeden z aspektów profilowania danej osoby, system ten należy uznać za ukierunkowany na osobę i co do zasady objęty zakresem art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, chociaż może on zostać wyłączony z zakresu zakazu z innych powodów.

Na przykład jeżeli informacje oparte na danych dotyczących lokalizacji, geoprzestrzennych lub dotyczących miejsca są powiązane z informacjami dotyczącymi danej osoby (np. miejscem zamieszkania danej osoby), a system AI oceniałby ryzyko popełnienia przestępstwa przez tę osobę wyłącznie na podstawie jej profilowania,

w tym na podstawie jej miejsca zamieszkania, które charakteryzuje się wysoką przestępczością, system ten należy uznać za ukierunkowany na osobę.

5.3.2. Systemy AI wykorzystywane do wspierania dokonywanej przez człowieka oceny opartej na obiektywnych i weryfikowalnych faktach związanych z działalnością przestępczą

214) Art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji stanowi, że zakaz ten nie ma zastosowania do systemów AI wykorzystywanych do wspierania dokonywanej przez człowieka oceny udziału osoby fizycznej w działalności przestępczej, która to ocena opiera się już na obiektywnych i weryfikowalnych faktach bezpośrednio związanych z działalnością przestępczą. W takim przypadku indywidualne oceny i prognozy ryzyka przestępczości nie byłyby również oparte *wyłącznie* na profilowaniu lub ocenie cech osobistych, a zatem nie byłyby zakazane.

Przykłady systemów AI z tego powodu nieobjętych zakazem obejmują:

– wykorzystywanie systemu AI do profilowania i kategoryzacji rzeczywistych zachowań, takich jak w racjonalnym stopniu podejrzane niebezpieczne zachowanie w tłumie, wskazujące na to, że ktoś przygotowuje się do popełnienia przestępstwa i prawdopodobnie je popełni, przy czym kategoryzacja dokonana przez sztuczną inteligencję podlega merytorycznej ocenie przez człowieka. W tym przypadku ocena ryzyka dokonywana przez człowieka przy wsparciu sztucznej inteligencji nie opiera się wyłącznie na cechach osobistych lub profilowaniu, ale na obiektywnych i weryfikowalnych faktach związanych z zachowaniem tej osoby wskazującym na zagrożenie przestępstwem, które to fakty zostały zweryfikowane przez człowieka przed podjęciem działań;

– policja bada ryzyko ewentualnego rozboju przy użyciu broni i podejrzewa dwie osoby. Istnieje kilka weryfikowalnych i obiektywnych faktów, na których opiera się to podejrzenie, takich jak weryfikowalna przynależność do grup w darknecie i rozmowy na tamtejszych czatach w celu zakupu broni. System AI łączący wykorzystywane do prognozowania przestępstw informacje oparte na danych geoprzestrzennych lub dotyczących miejsca z pochodzącymi z systemu automatycznego rozpoznawania tablic rejestracyjnych (ANPR) informacjami o pojazdach należących do podejrzanych wspiera dokonywaną przez człowieka podczas dochodzenia ocenę opartą na weryfikowalnych i obiektywnych faktach bezpośrednio związanych z konkretną działalnością przestępczą;

– wykorzystanie systemu AI, który ocenia ryzyko związane z możliwością przedterminowego zwolnienia więźnia. Stworzony przez sztuczną inteligencję profil danej osoby lub dokonana przez AI ocena jej cech osobowości i cech charakterystycznych wspierają jedynie dokonaną przez człowieka ocenę obiektywnych i weryfikowalnych faktów związanych z wcześniejszymi przestępstwami i wykazanymi zachowaniami istotnymi dla resocjalizacji;

– sąd przeprowadza postępowanie w przedmiocie tymczasowego aresztowania osoby oskarżonej o popełnienie poważnego przestępstwa, aby ocenić, czy można

zastosować środki nieizolacyjne. Orzeczenie opiera się na ocenie istnienia uzasadnionych podstaw do zastosowania tymczasowego aresztowania, takich jak prawdopodobieństwo, że osoba podejrzana lub oskarżona popełni inne przestępstwo, jeżeli nie zostanie zatrzymana, lub że ucieknie lub utrudni prawidłowy przebieg postępowania przygotowawczego. Aby wspomóc ten proces, sąd korzysta z narzędzia oceny ryzyka opartego na AI, trenowanego na danych obejmujących wcześniejszą karalność osób w podobnych przypadkach, oraz uwzględnia czynniki takie jak grupa wiekowa, zachowanie społeczne, dochód i status zatrudnienia;

– system AI jest wykorzystywany do wspierania dokonywanej przez funkcjonariusza oceny ryzyka, że dana osoba odbywająca karę nieizolacyjną naruszy warunki zwolnienia lub ucieknie, na podstawie przeszłych zachowań przestępczych i obiektywnych faktów, które dają podstawy do podejrzeń, takich jak przestrzeganie warunków zwolnienia, wyniki oceny psychologicznej i zalecenia innych służb środowiskowych, z których dana osoba może korzystać. Na podstawie tych informacji funkcjonariusz decyduje, czy utrzymać status quo, czy też zmienić warunki zwolnienia;

– organy celne wykorzystują systemy AI do oceny ryzyka, że towary wprowadzane do UE nie są zgodne z przepisami mającymi zastosowanie na granicy (co może np. obejmować przywóz niedozwolonych środków odurzających, naruszenie sankcji wywozowych lub inną nielegalną działalność), w celu określenia sytuacji, w których należy przeprowadzić kontrolę celną. System AI ocenia przekazane organom celnym obiektywne i weryfikowalne informacje dotyczące towarów i ich łańcuchów dostaw (np. rodzaj i wartość towarów, numer kontenera, środek transportu mogący służyć ukryciu innych towarów, uprzednia wiedza na temat zgodności towarów o konkretnym opisie i pochodzeniu z wymogami dotyczącymi ich przywozu do Unii lub wywozu z Unii). W niektórych przypadkach może również przetwarzać informacje na temat wcześniejszego udziału importera lub eksportera w nieprawidłowościach związanych z przywozem towarów, ich przynależności do organizacji przestępczych lub wpisu do rejestru karnego w związku z nielegalnym obrotem środkami odurzającymi. Takie systemy nie wchodzą w zakres zakazu, ponieważ wszelkie prognozy dotyczące prawdopodobieństwa udziału osoby fizycznej w przywozie lub wywozie nielegalnych towarów nie opierają się wyłącznie na profilowaniu, ale na obiektywnych i weryfikowalnych informacjach dotyczących towarów i wcześniejszego udziału importera lub eksportera w działalności przestępczej, a także muszą zostać zweryfikowane przez człowieka w celu ustalenia, czy dana sytuacja wymaga kontroli celnej lub działań ograniczających ryzyko.

5.3.3. Systemy AI wykorzystywane do prognozowania i oceny przestępczości w odniesieniu do podmiotów prawnych

215) Zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji ma zastosowanie wyłącznie do indywidualnych prognoz i ocen ryzyka osób fizycznych, co zazwyczaj wyklucza systemy przewidujące popełnienie przestępstwa na podstawie profilowania osób prawnych, takich jak przedsiębiorstwa czy organizacje pozarządowe.

Na przykład:

– organ podatkowy lub celny wykorzystuje system AI do analizy dużych ilości danych dotyczących transakcji i deklaracji podatkowych oraz danych celnych przedsiębiorstw w celu oceny ryzyka popełnienia przez przedsiębiorstwo oszustw podatkowych lub celnych stanowiących przestępstwo;

– systemy AI wykorzystywane są do wspomagania organów celnych w identyfikowaniu sytuacji, w których osoby prawne należy poinstruować, aby nie wysyłały nielegalnych towarów do UE.

216) Jednocześnie mogą wystąpić przypadki graniczne, w których osoba fizyczna działa za pośrednictwem podmiotu prawnego jako „przedsiębiorca indywidualny" lub jako przedstawiciel wolnego zawodu (np. prawnik). W takich okolicznościach zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji może mieć zastosowanie, o ile spełnione są wszystkie warunki, ponieważ system AI profiluje konkretną osobę fizyczną i ocenia lub przewiduje ryzyko popełnienia przez nią przestępstwa, nawet jeżeli odbywa się to w celach związanych z działalnością handlową prowadzoną przez osobę fizyczną.

5.3.4. Systemy AI wykorzystywane do indywidualnego przewidywania przestępstw administracyjnych

217) Zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji ma zastosowanie wyłącznie do przewidywania przestępstw kryminalnych, a zatem z jego zakresu wyłączone są przestępstwa administracyjne, których ściganie jest co do zasady mniej inwazyjne dla podstawowych praw i wolności osób.

Na przykład wykorzystanie przez organ publiczny sztucznej inteligencji w kontekście dochodzenia administracyjnego do oceny ryzyka popełnienia drobnych przestępstw (takich jak wykroczenia drogowe) lub nieprawidłowości w procedurach podatkowych, postępowaniach o udzielenie zamówień publicznych lub procesach wydatkowania środków nie wchodził w zakres zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, nawet w przypadkach, w których w wyniku dochodzeń i kontroli administracyjnych można zebrać informacje dotyczące ewentualnego udziału osób fizycznych w przestępstwach.

218) To, czy przestępstwo ma charakter administracyjny czy karny, może zależeć od prawa Unii lub prawa krajowego. W przypadku przestępstw, które nie są bezpośrednio regulowane prawem Unii, krajowa kwalifikacja przestępstwa podlega kontroli TSUE, ponieważ „przestępstwo" jest pojęciem, które ma autonomiczne znaczenie w prawie UE i powinno być interpretowane w sposób spójny we wszystkich państwach członkowskich. TSUE stwierdził w innym kontekście, że klasyfikacja przestępstw przez państwa członkowskie nie jest w tym względzie rozstrzygająca[142]. Odpowiednie kryteria stosowane do oceny charakteru przestępstwa (karnego lub nie) można znaleźć

[142] Zob. na przykład wyrok Trybunału (wielka izba) z dnia 14 listopada 2013 r. Postępowania dotyczące egzekucji kary pieniężnej nałożonej na – Marián Baláž, sprawa C-60/12, sygnatura ECLI: ECLI:EU:C:2013:733.

w odpowiednim orzecznictwie TSUE i Europejskiego Trybunału Praw Człowieka (ETPC)[143].

5.4. Wzajemne powiązania z innymi aktami prawnymi Unii

219) Wzajemne powiązania zakazu określonego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji z dyrektywą o ochronie danych w sprawach karnych i RODO są istotne przy ocenie zgodności z prawem przetwarzania danych osobowych na podstawie unijnych przepisów o ochronie danych, takich jak RODO i dyrektywa o ochronie danych w sprawach karnych. W szczególności w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji nałożono na organy ścigania, inne organy publiczne i podmioty prywatne objęte zakresem tego zakazu szczególny zakaz oceny lub przewidywania ryzyka popełnienia przestępstwa przez osobę fizyczną, wyłącznie na podstawie profilowania osoby fizycznej lub oceny jej cech osobowości i cech charakterystycznych. W odniesieniu do dyrektywy o ochronie danych w sprawach karnych art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji pozostaje bez uszczerbku dla art. 11 ust. 3 dyrektywy o ochronie danych w sprawach karnych, który zakazuje profilowania prowadzącego do (bezpośredniej lub pośredniej) dyskryminacji.

220) Istotne jest również powiązanie zakazu zawartego w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji z dyrektywą (UE) 2016/343 w sprawie domniemania niewinności, ponieważ oba akty dotyczą – bezpośrednio w przypadku dyrektywy i pośrednio w przypadku aktu w sprawie sztucznej inteligencji (zob. motyw 42) – prawa podstawowego do domniemania niewinności do czasu udowodnienia winy zgodnie z prawem[144]. Chociaż dyrektywa ma zastosowanie od momentu, w którym dana osoba jest podejrzana lub oskarżona o popełnienie przestępstwa[145], akt w sprawie sztucznej inteligencji ma szerszy zakres stosowania i ma zastosowanie już na etapie przewidywania i zapobiegania przestępczości przed wszczęciem formalnego postępowania przygotowawczego przeciwko danej osobie, nawet w przypadkach, gdy takie prognozy i oceny ryzyka są dokonywane przez podmioty prywatne objęte zakresem art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji, a nie przez właściwe organy ścigania, w tym organy sądowe.

221) Nawet w przypadkach, w których zakaz określony w art. 5 ust. 1 lit. d) aktu w sprawie sztucznej inteligencji nie ma zastosowania, należy podkreślić, że obowiązujące prawo

---

143 Zgodnie z orzecznictwem TSUE to do sądów krajowych należy ustalenie, czy kara niemająca charakteru karnego może być uznana za „karną" w świetle tzw. „kryteriów Engela". Zob: ETPC, wyrok z dnia 8 czerwca 1976 r., Engel i inni przeciwko Niderlandom, skarga nr 5100/71, 5101/71, 5102/71, 5354/72 i 5370/72, CE:ECHR:1976:0608JUD000510071, pkt 82. Kryteria te, pierwotnie opracowane przez Europejski Trybunał Praw Człowieka (ETPC), a następnie zatwierdzone przez TSUE, mają charakter alternatywny, a nie kumulatywny. Badając, czy sankcja ma charakter karny, właściwy sąd krajowy powinien ocenić: 1) klasyfikację odpowiednich przepisów w prawie krajowym, 2) sam charakter przestępstwa oraz 3) surowość sankcji. Przy ocenie charakteru przestępstwa uwzględnia się m.in. to, czy postępowanie zostało wszczęte przez organ publiczny posiadający ustawowe uprawnienia wykonawcze; czy przepis prawny ma cel represyjny lub odstraszający; czy norma prawna ma na celu ochronę ogólnych interesów społeczeństwa, które są zazwyczaj chronione prawem karnym; czy nałożenie jakiejkolwiek sankcji zależy od stwierdzenia winy. Jeżeli chodzi o surowość sankcji, odpowiednim odniesieniem jest maksymalna potencjalna sankcja przewidziana w prawie krajowym. Kryteria te mają charakter alternatywny i niekoniecznie kumulatywny. Zob. Europejski Trybunał Praw Człowieka, Przewodnik po art. 6 europejskiej konwencji praw człowieka – prawo do rzetelnego procesu sądowego (aspekt karny), zaktualizowany 29 lutego 2024 r. Zob. również wyrok TSUE z dnia 5 czerwca 2012 r., Bonda, sprawa C-489/10, EU:C:2012:319, pkt 37 i nast.; TSUE, wyrok z dnia 26 lutego 2013 r., Åkerberg/Fransson, C-617/10, EU:C:2013:105, pkt 35.

144 Domniemanie niewinności jest prawem podstawowym zapisanym w art. 48 Karty praw podstawowych Unii Europejskiej.

145 Jak wskazał TSUE, aby dyrektywa miała zastosowanie, nie jest wymagane, aby osoba ta została poinformowana o swoim statusie podejrzanego/oskarżonego przez właściwe organy.

Unii i prawo krajowe nadal ma pełne zastosowanie, w tym w szczególności przepisy dotyczące ochrony danych, prawo karne procesowe i prawo policyjne oraz zabezpieczenia, które mogą dodatkowo ograniczać lub nakładać dodatkowe warunki dotyczące stosowania systemów AI prognozujących popełnienie przestępstwa przez poszczególne osoby.

6. ART. 5 UST. 1 LIT. E) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – NIEUKIERUNKOWANE POZYSKIWANIE WIZERUNKÓW TWARZY

222) W art. 5 ust. 1 lit. e) ustawy o sztucznej inteligencji zakazano wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu lub wykorzystywania systemów AI, które tworzą lub rozbudowują bazy danych służące rozpoznawaniu twarzy poprzez nieukierunkowane pozyskiwanie wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej.

6.1. Uzasadnienie i cele

223) Nieukierunkowane pozyskiwanie wizerunków twarzy z internetu i nagrań z telewizji przemysłowej poważnie narusza prawa osób fizycznych do prywatności i ochrony danych oraz pozbawia te osoby prawa do zachowania anonimowości. W motywie 43 aktu w sprawie sztucznej inteligencji uzasadniono zatem zakaz ustanowiony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji „poczuciem masowego nadzoru" i ryzykiem „poważnego naruszenia praw podstawowych, w tym prawa do prywatności".

6.2. Główne pojęcia i elementy zakazu

> ***Art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji stanowi:***
>
> Zakazuje się następujących praktyk w zakresie AI:
>
> e) wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu lub wykorzystywania systemów AI, które tworzą lub rozbudowują bazy danych służące rozpoznawaniu twarzy poprzez nieukierunkowane pozyskiwanie (ang. untargeted scraping) wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej;

224) Aby zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi zostać spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku w tym konkretnym celu" lub „wykorzystywaniu" systemu AI;

(ii) celem musi być tworzenie lub rozszerzanie baz danych dotyczących rozpoznawania twarzy;

(iii) uzupełnianie bazy danych musi odbywać się przy użyciu narzędzi AI do nieukierunkowanego pozyskiwania; oraz

(iv) wizerunki twarzy muszą pochodzić albo z internetu, albo z nagrań z telewizji przemysłowej.

225) Aby zakaz miał zastosowanie, muszą być spełnione jednocześnie wszystkie cztery warunki. Pierwszy element, tj. wprowadzanie do obrotu, oddawanie do użytku lub wykorzystywanie systemu AI, przeanalizowano już w sekcji 2.3. Zakaz ten ma zatem zastosowanie zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów AI do obrotu, oddawać ich do użytku ani wykorzystywać. Poniżej opisano i przeanalizowano szczegółowe kryteria związane z zakazem nieukierunkowanego pozyskiwania. Zakaz ten ma zastosowanie do narzędzi pozyskiwania, które są wprowadzane do obrotu lub oddawane do użytku „w tym konkretnym celu", jakim jest nieukierunkowane pozyskiwanie wizerunków twarzy z internetu lub nagrań z telewizji przemysłowej. Oznacza to, że zakaz ten nie ma zastosowania do dowolnych narzędzi pozyskiwania, za pomocą których można stworzyć lub rozbudować bazę danych służącą rozpoznawaniu twarzy, lecz jedynie do narzędzi nieukierunkowanego pozyskiwania.

6.2.1. Bazy danych służące rozpoznawaniu twarzy

226) Zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji obejmuje systemy AI wykorzystywane do tworzenia lub rozbudowywania baz danych służących rozpoznawaniu twarzy. „Bazę danych" w tym kontekście należy rozumieć jako każdy zbiór danych lub informacji, który jest specjalnie uporządkowany w celu szybkiego wyszukiwania i odnajdywania przez komputer. Baza danych służąca rozpoznawaniu twarzy jest w stanie dopasować ludzką twarz z cyfrowego obrazu lub ramki wideo do wizerunków twarzy w bazie danych, porównując ją z obrazami znajdującymi się w bazie danych i ustalając, czy istnieje prawdopodobieństwo dopasowania tych dwóch elementów. Taka baza danych służąca rozpoznawaniu twarzy może mieć charakter tymczasowy, scentralizowany lub zdecentralizowany. Art. 5 ust. 1 lit. e) nie zawiera wymogu, aby jedynym celem bazy danych było wykorzystywanie jej do rozpoznawania twarzy; wystarczy, aby baza danych mogła być wykorzystywana do rozpoznawania twarzy.

6.2.2. Poprzez nieukierunkowane pozyskiwanie wizerunków twarzy

227) „Pozyskiwanie" (ang. scraping) odnosi się zazwyczaj do korzystania z internetowych robotów indeksujących, botów lub innych narzędzi do automatycznego pozyskiwania danych lub treści z różnych źródeł, w tym z telewizji przemysłowej, ze stron internetowych lub z mediów społecznościowych. Narzędzia te są oprogramowaniem zaprogramowanym do przeszukiwania baz danych i pozyskiwania informacji oraz wykorzystywania tych informacji do innego celu.

228) Termin „nieukierunkowany" odnosi się do techniki, która działa jak „odkurzacz", pochłaniając jak najwięcej danych i informacji, bez ukierunkowywania na konkretny i indywidualny podmiot, którego dotyczy pozyskiwanie. W ramach pozyskiwania bez rozróżnienia zbierane są dane lub treści. W związku z tym pojęcie „nieukierunkowane" oznacza bez szczególnego ukierunkowania na daną osobę lub grupę osób. Przestrzeganie wykluczeń na podstawie protokołów internetowych, takich jak robot.txt, nie ma wpływu na nieukierunkowany charakter pozyskiwania.

229) Jeżeli narzędzie do pozyskiwania zostanie poinstruowane, aby gromadzić obrazy lub nagrania wideo zawierające twarze wyłącznie określonych osób lub określonej grupy osób, na przykład w celu znalezienia konkretnego przestępcy lub zidentyfikowania grupy ofiar, pozyskiwanie staje się ukierunkowane. Takie pozyskiwanie nie jest objęte zakazem określonym w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji.

230) Zakaz ten nie obejmuje na przykład ukierunkowanego pozyskiwania obrazów skupionego na danej klasie ofiar, dokonywanego za pomocą robotów indeksujących wykrywających wizerunki ofiar, które handlarze ludźmi publikują na kanałach mediów społecznościowych. Nieukierunkowane pozyskiwanie należy interpretować w sposób, który nie pozwala na obejście zakazu. Pozyskiwanie danych z nagrań internetowych lub nagrań z telewizji przemysłowej w celu stopniowego tworzenia bazy danych poprzez wybieranie za każdym razem określonych grup osób lub innych kryteriów powinno być objęte zakazem określonym w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji, jeżeli wynik końcowy jest funkcjonalnie taki sam jak w przypadku prowadzenia od samego początku nieukierunkowanego pozyskiwania.

231) W przypadku gdy systemy łączą ukierunkowane pozyskiwanie zdjęć lub nagrań wideo z wyszukiwaniem nieukierunkowanym, nieukierunkowane pozyskiwanie jest zakazane.

6.2.3. Z internetu i nagrań z telewizji przemysłowej

232) Aby zakaz, o którym mowa w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji, miał zastosowanie, źródłem wizerunków twarzy musi być internet lub nagranie z telewizji przemysłowej. Jeśli chodzi o internet, fakt, że dana osoba opublikowała własne wizerunki twarzy na platformie mediów społecznościowych, nie oznacza, że osoba ta wyraziła zgodę na umieszczenie tych wizerunków w bazie danych służącej rozpoznawaniu twarzy. Przykłady pozyskiwania wizerunków twarzy z nagrań z telewizji przemysłowej obejmują zdjęcia wykonane przez kamery monitorujące stosowane w takich miejscach jak lotniska, ulice, parki itp.

Przykład:

Przedsiębiorstwo produkujące oprogramowanie do rozpoznawania twarzy gromadzi wizerunki twarzy. Zdjęcia będące w posiadaniu przedsiębiorstwa zostały pozyskane z mediów społecznościowych (np. Facebook, YouTube, Twitter, Venmo) za pomocą „automatycznego pozyskiwacza obrazów", który przeszukuje internet i wykrywa obrazy zawierające ludzkie twarze. Obrazy te są gromadzone wraz z wszelkimi powiązanymi informacjami (takimi jak źródło obrazu (URL), geolokalizacja, a czasami imiona i nazwiska osób). Następnie cechy twarzy są wyodrębniane ze zdjęć i przekształcane w ich matematyczne odwzorowania, które są haszowane na potrzeby indeksowania i porównywania w przyszłości. Gdy użytkownik załaduje zdjęcie danej osoby do systemu AI, system ten ustali, czy to zdjęcie pasuje do jednej z twarzy znajdujących sie w bazie danych. Załadowane zdjęcie zostanie poddane tej samej transformacji matematycznej co pozyskane obrazy.

233) W przypadku gdy system AI otrzymuje zdjęcie osoby i przeszukuje internet pod kątem trafień dla tej twarzy, działając jak „wyszukiwarka zdjęć wykorzystująca zasady inżynierii odwrotnej”, zostanie to uznane za pozyskiwanie ukierunkowane. Ponadto wątpliwe jest, czy trafienia te znalazłyby się w „bazie danych”.

6.3. Systemy nieobjęte zakresem

234) Zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji nie ma zastosowania do nieukierunkowanego pozyskiwania danych biometrycznych innych niż wizerunki twarzy (takich jak próbki głosu). Zakaz ten nie ma również zastosowania w przypadku, gdy do pozyskiwania nie są wykorzystywane żadne systemy AI. Bazy danych zdjęć twarzy, które nie są wykorzystywane do rozpoznawania osób, takie jak bazy danych zdjęć twarzy wykorzystywane do trenowania lub testowania modeli sztucznej inteligencji, również nie są objęte zakresem stosowania, w przypadku gdy osoby te nie są zidentyfikowane.

235) Zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji nie ma zastosowania do systemów AI, które pozyskują z internetu duże liczby wizerunków twarzy w celu tworzenia modeli sztucznej inteligencji generujących nowe obrazy przedstawiające fikcyjne osoby, ponieważ takie systemy nie skutkowałyby rozpoznawaniem prawdziwych osób. Takie systemy AI mogą podlegać wymogom przejrzystości określonym w art. 50 aktu w sprawie sztucznej inteligencji.

236) Zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji obejmuje systemy AI wykorzystywane do tworzenia lub rozbudowywania baz danych służących rozpoznawaniu twarzy. Jeżeli chodzi o istniejące bazy danych twarzy utworzone przed wejściem w życie zakazu, które nie są dalej rozszerzane za pomocą wspomaganego sztuczną inteligencją nieukierunkowanego pozyskiwania, te bazy danych i ich wykorzystywanie muszą być zgodne z mającymi zastosowanie unijnymi przepisami o ochronie danych.

237) Zakaz określony w art. 5 ust. 1 lit. e) aktu w sprawie sztucznej inteligencji dotyczy tworzenia lub rozszerzania baz danych służących rozpoznawaniu twarzy. Konkretna czynność identyfikacji biometrycznej podlega szczegółowym przepisom zawartym w akcie w sprawie sztucznej inteligencji i innych odpowiednich przepisach Unii.

6.4. Wzajemne powiązania z innymi aktami prawnymi Unii

238) Jeżeli chodzi o unijne przepisy o ochronie danych, nieukierunkowane pozyskiwanie materiałów z internetu lub nagrań z telewizji przemysłowej w celu tworzenia lub rozbudowy baz danych służących rozpoznawania twarzy, czyli przetwarzanie danych osobowych (gromadzenie danych i korzystanie z baz danych), jest niezgodne z prawem i nie opiera się na żadnej podstawie prawnej wynikającej z RODO, z rozporządzenia w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych przez instytucje, organy i jednostki organizacyjne Unii ani z dyrektywy o ochronie danych w sprawach karnych.

7. ART. 5 UST. 1 LIT. F) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – ROZPOZNAWANIE EMOCJI

239) W art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji zakazano systemów AI wykorzystywanych do wyciągania wniosków na temat emocji osoby fizycznej w miejscu pracy lub instytucjach edukacyjnych, z wyjątkiem przypadków, gdy korzystanie z systemu ma na celu względy medyczne lub bezpieczeństwa. Systemy rozpoznawania emocji, które nie są objęte zakazem, uznaje się za systemy wysokiego ryzyka zgodnie z pkt 1 lit. c) załącznika III do aktu w sprawie sztucznej inteligencji. W art. 50 ust. 3 aktu w sprawie sztucznej inteligencji określono pewne wymogi dotyczące przejrzystości w odniesieniu do stosowania systemów rozpoznawania emocji.

7.1. Uzasadnienie i cele

240) Technologia rozpoznawania emocji szybko się rozwija i obejmuje różne technologie i operacje przetwarzania służące do wykrywania, gromadzenia, analizowania, kategoryzacji i uczenia się emocji od osób, a także reagowania na emocje innych osób i interakcji z nimi. Technologia taka określana jest również mianem „affect technology" (technologia afektu). Rozpoznawanie emocji może być wykorzystywane w wielu obszarach i dziedzinach do szerokiego zakresu zastosowań[146], takich jak analiza zachowań klientów[147] oraz ukierunkowane reklamy i neuromarketing[148]; w branży rozrywkowej, na przykład w celu dostarczania zindywidualizowanych zaleceń lub przewidywania reakcji na filmy; w medycynie i opiece zdrowotnej, na przykład w celu wykrywania depresji, zapobiegania samobójstwom lub wykrywania autyzmu; w edukacji, na przykład w celu monitorowania uwagi lub zaangażowania osób uczących się (uczniów i studentów w różnym wieku); na rynku pracy, na przykład w celu wspomagania procesu rekrutacji, monitorowania emocji lub znudzenia pracowników, ale także w celu zapewnienia dobrego samopoczucia, aby „uszczęśliwiać pracowników"[149]; na potrzeby ścigania przestępstw i zapewniania bezpieczeństwa publicznego, na przykład za pomocą wykrywaczy kłamstw lub monitorowania emocji podczas dużych wydarzeń; i do wielu innych celów.

---

146 Wykorzystanie emocji do celów gospodarczych jest również określane jako „emocjonomika".

147 Zob. np. G. Mangano, A. Ferrari, C. Rafale, E. Vezzetti, F. Marcolin, „Willingness of sharing facial data for emotion recognition: a case study in the insurance market" [Gotowość udostępniania danych twarzy do rozpoznawania emocji: studium przypadku na rynku ubezpieczeń] *AI & Society*, Londyn, Springer, 2023.

148 Zob. N. Lee, A. J. Broderick, & L. Chamberlain, „What is 'neuromarketing'?" A discussion and agenda for future research" [Czym jest „neuromarketing"? Dyskusja i plan przyszłych badań], w *International Journal of psychophysiology*, 63(2), 2007,199–204, gdzie neuromarketing definiuje się jako dziedzinę badań polegającą na „stosowaniu metod neuronaukowych w celu analizy i zrozumienia ludzkich zachowań w odniesieniu do rynków i wymian rynkowych" (s. 200).

149 Zob. E. Ackerman, & E. Strickland, „Are you Ready for Workplace Brain Scanning? Extracting and using brain data will make workers happier and more productive, backers say" [Czy jesteś gotowy na skanowanie mózgu w miejscu pracy? Jak twierdzą zwolennicy tego rozwiązania, pozyskiwanie i wykorzystywanie danych z mózgu sprawi, że pracownicy będą szczęśliwsi i bardziej produktywni], *IEEE Spectrum*, 19 listopada 2022 r., https://spectrum.ieee.org/neurotech-workplace-innereye-emotiv. Autorzy wyjaśniają, że „czujniki wykrywają aktywność elektryczną w różnych obszarach mózgu, a wzorce tej aktywności mogą być zasadniczo skorelowane z różnymi uczuciami lub reakcjami fizjologicznymi, takimi jak stres, skupienie lub reakcja na bodźce zewnętrzne".

241) Skuteczność i dokładność rozpoznawania emocji jest często poddawana w wątpliwość[150]. W motywie 44 aktu w sprawie sztucznej inteligencji wyjaśniono, że „istnieją poważne obawy co do podstaw naukowych systemów AI mających na celu rozpoznawanie emocji lub wyciąganie wniosków na temat emocji, zwłaszcza że wyrażanie emocji znacznie się różni w zależności od kultur i sytuacji, a nawet w przypadku pojedynczej osoby. Wśród głównych wad takich systemów znajdują się ograniczona wiarygodność, nieprecyzyjność i ograniczona możliwość uogólnienia". Ponadto wyjaśniono w nim, że rozpoznawanie emocji może prowadzić do „dyskryminacyjnych wyników i [może] naruszać prawa i wolności zainteresowanych osób", w szczególności prawa do prywatności, godności ludzkiej i wolności myśli. Odgrywa to ważną rolę w stosunkach asymetrycznych, zwłaszcza w kontekście miejsca pracy oraz instytucji kształcenia i szkolenia, w których zarówno pracownicy, jak i uczniowie znajdują się w szczególnie trudnej sytuacji. Jednocześnie rozpoznawanie emocji w konkretnych kontekstach zastosowań, takich jak bezpieczeństwo i opieka medyczna (np. leczenie i diagnoza), przynosi pewne korzyści[151].

7.2. Główne pojęcia i elementy zakazu

> ***Art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji stanowi:***
>
> Zakazuje się następujących praktyk w zakresie AI:
>
> f) wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu lub wykorzystywania systemów AI do wyciągania wniosków na temat emocji osoby fizycznej w miejscu pracy lub instytucjach edukacyjnych, z wyjątkiem przypadków, w których system AI ma zostać wdrożony lub wprowadzony do obrotu ze względów medycznych lub bezpieczeństwa;

242) Aby zakaz określony w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi zostać spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku w tym konkretnym celu" lub „wykorzystywaniu" systemu AI;

(ii) system AI musi być stosowany do wyciągania wniosków na temat emocji[152];

(iii) obszarem stosowania musi być miejsce pracy lub instytucja kształcenia i szkolenia; oraz

(iv) z zakazu wyłączone są systemy AI przeznaczone do celów medycznych lub zapewniania bezpieczeństwa.

243) Aby zakaz miał zastosowanie, muszą być spełnione jednocześnie wszystkie cztery warunki. Pierwszy element, tj. wprowadzanie do obrotu, oddawanie do użytku lub

---

150 Zob. np. J. Stanley, *„Experts Say 'Emotion Recognition' lacks Scientific Foundation"* [Eksperci twierdzą, że „rozpoznawaniu emocji" brakuje podstaw naukowych], 18.07.2019, ACLU, odnoszące się do badania przeprowadzonego przez L. Feldman Barrett i in., „Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements" [Nowe spojrzenie na wyrażanie emocji: wyzwania związane z wnioskowaniem o emocjach na podstawie mimiki ludzkiej twarzy], *Psychological Science in the Public Interest*, 2019, s. iii–90.

151 Zob. np. R. El Kaliouby, R. Picard i S. Baron-Cohen, „Affective Computing and Autism" [Przetwarzanie emocjonalne a autyzm], *Annals New York Academy of Sciences*, 2007, s. 228–248.

152 Lub technologia musi być w stanie wyciągać wnioski na temat emocji (tj. przy wprowadzaniu systemu do obrotu).

wykorzystywanie systemu AI, przeanalizowano już w sekcji 2.3. Zakaz ten ma zatem zastosowanie zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów AI do obrotu, oddawać ich do użytku ani wykorzystywać. Pozostałe warunki związane z zakazem opisano i przeanalizowano poniżej.

7.2.1. Systemy AI do wyciągania wniosków na temat emocji

a) Systemy AI do wyciągania wniosków na temat emocji a systemy rozpoznawania emocji

244) W art. 3 pkt 39 aktu w sprawie sztucznej inteligencji „system rozpoznawania emocji" zdefiniowano jako system AI „służący do identyfikacji lub wywnioskowania emocji lub zamiarów osób fizycznych na podstawie danych biometrycznych tych osób". Zakaz, o którym mowa w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, nie odnosi się do „systemów rozpoznawania emocji", lecz jedynie do „systemów AI służących do wywnioskowania emocji osoby fizycznej". W motywie 44 wyjaśniono ponadto, że zakaz ten obejmuje systemy AI „mające na celu rozpoznawanie emocji lub wyciąganie wniosków na temat emocji".

245) Co do zasady warunkiem wstępnym wnioskowania jest rozpoznanie, a zatem zakaz ten należy rozumieć jako obejmujący zarówno systemy AI rozpoznające emocje lub zamiary, jak i systemy AI wyciągające wnioski na ich temat[153]. Ze względu na spójność należy również rozumieć zakaz zawarty w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji jako mający podobny zakres jak przepisy mające zastosowanie do innych systemów rozpoznawania emocji (załącznik III pkt 1 lit. c) i art. 50 aktu w sprawie sztucznej inteligencji) oraz ograniczyć go do wyciągania wniosków na podstawie danych biometrycznych danej osoby. Należy zatem uznać, że definicja systemów rozpoznawania emocji zawarta w art. 3 pkt 39 aktu w sprawie sztucznej inteligencji ma związek z art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji.

b) Rozpoznawanie emocji lub zamiarów lub wyciąganie wniosków na ich temat

246) „Rozpoznawanie" ma miejsce, gdy przetwarzanie danych biometrycznych (np. głosu lub wyrazu twarzy) osoby fizycznej umożliwia bezpośrednie porównanie i zidentyfikowanie emocji z emocją, która została wcześniej zaprogramowana w systemie rozpoznawania emocji. „Wyciąganie wniosków" polega na samodzielnym wnioskowaniu przez system informacji wygenerowanych w wyniku procesów analitycznych i innych procesów. W takim przypadku informacje na temat emocji nie opierają się wyłącznie na zgromadzonych danych na temat osoby fizycznej, ale są wywnioskowane z innych danych, w tym za pomocą mechanizmów uczenia maszynowego, które na podstawie danych uczą się, jak wykrywać emocje[154].

---

153 Zob. również motyw 18 aktu w sprawie sztucznej inteligencji.

154 Zob. motyw 12 aktu w sprawie sztucznej inteligencji. Dane wywnioskowane są zatem często również wynikiem procesów analitycznych opartych na prawdopodobieństwie (przeprowadzanych na dużych zbiorach danych), mających na celu znalezienie korelacji i wzorców w zbiorach danych.

c) Emocje

247) Do celów art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji pojęcie emocji lub intencji należy rozumieć szeroko, a nie w sposób zawężający. W motywie 18 aktu w sprawie sztucznej inteligencji podano pewne szczegółowe informacje, wymieniając emocje takie jak „radość, smutek, złość, zdziwienie, obrzydzenie, zakłopotanie, podekscytowanie, wstyd, pogarda, satysfakcja i rozbawienie". Przedstawione przykłady nie wyczerpują wszystkich możliwości.

248) Zakazu nie należy obchodzić poprzez odnoszenie się do postawy; obejmuje on przypadki, w których system AI stwierdza na podstawie danych biometrycznych, że dana osoba prezentuje na przykład wrogą postawę.

249) W motywie 18 aktu w sprawie sztucznej inteligencji wyjaśniono, że systemy rozpoznawania emocji lub zamiarów nie obejmują „stanów fizycznych, takich jak ból lub zmęczenie, w tym na przykład systemów stosowanych do wykrywania poziomu zmęczenia zawodowych pilotów lub kierowców w celu zapobiegania wypadkom". Doprecyzowuje się w nim ponadto, że systemy rozpoznawania emocji nie obejmują „samego wykrywania łatwych do zauważenia form wyrazu, gestów lub ruchów, chyba że wykorzystuje się je do identyfikacji lub wnioskowania na temat emocji"; należy rozumieć, że ma to również zastosowanie do art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji. Te formy wyrazu mogą obejmować podstawowe rodzaje wyrazu twarzy, takie jak grymas lub uśmiech, gesty, takie jak ruch rąk, ramion lub głowy, lub cechy głosu danej osoby, takie jak podniesiony ton lub szept. Jeżeli jednak te łatwe do zauważenia formy wyrazu lub gesty są wykorzystywane do rozpoznawania emocji lub zamiarów bądź wyciągania wniosków na ich temat, są one objęte zakazem.

Na przykład:

– spostrzeżenie, że dana osoba jest uśmiechnięta, nie stanowi rozpoznawania emocji;

– określenie, czy dana osoba jest chora, nie stanowi rozpoznawania emocji;

– korzystanie przez telewizyjnego nadawcę z urządzenia umożliwiającego śledzenie, ile razy prezenterzy wiadomości uśmiechają się do kamery, nie stanowi rozpoznawania emocji;

– stwierdzenie, że dana osoba jest szczęśliwa, stanowi rozpoznawanie emocji; wnioskowanie przez system AI, że pracownik jest niezadowolony, smutny lub zirytowany na klientów (np. z gestów ciała, grymasu lub braku uśmiechu) stanowi rozpoznawanie emocji;

– wnioskowanie przez systemy z głosu lub gestów ciała, że uczeń jest wściekły i wkrótce stanie się agresywny, stanowi rozpoznawanie emocji;

– wykorzystywanie systemów rozpoznawania opartych na AI do wnioskowania na temat zmęczenia pilota lub kierowcy w celu ostrzeżenia ich i zasugerowania, kiedy zrobić przerwę, aby uniknąć wypadku, nie stanowi rozpoznawania emocji,

ponieważ rozpoznawanie emocji nie obejmuje stanu fizycznego, takiego jak ból lub zmęczenie.

d) Na podstawie ich danych biometrycznych

250) Zgodnie z definicją zawartą w art. 3 pkt 39 aktu w sprawie sztucznej inteligencji jedynie systemy AI do identyfikacji lub wywnioskowania emocji lub zamiarów na podstawie danych biometrycznych stanowią systemy rozpoznawania emocji[155].

251) Właściwości osobnicze, z których można wyodrębnić dane biometryczne, to cechy fizyczne lub behawioralne. Biometria fizjologiczna wykorzystuje fizyczne, strukturalne i stosunkowo statyczne cechy danej osoby, takie jak odciski palców, budowa tęczówki, kontury twarzy lub geometria żył w dłoniach. Niektóre techniki mają charakter mikroskopowy, ale nadal wykazują struktury biologiczne i chemiczne, które można pozyskać i zidentyfikować, np. DNA i zapach[156]. Biometria behawioralna monitoruje charakterystyczne cechy ruchów, gestów i umiejętności motorycznych poszczególnych osób podczas wykonywania przez nie zadania lub szeregu zadań. Oznacza to, że rejestruje się i analizuje ruchy ludzi, takie jak chodzenie (analiza chodu) lub kontakt palca z klawiaturą (uderzenia w klawisze). Biometria behawioralna obejmuje różne techniki, które wykazują zarówno dobrowolne, jak i niezamierzone powtarzające się ruchy i związane z nimi regularne odstępy czasu czy nacisk wywierany przez określone części ciała, począwszy od analizy podpisu, chodu, głosu i uderzeń w klawisze po okulografię i bicie serca[157], elektroencefalografię (EEG)[158] lub elektrokardiogramy (EKG)[159]. Dane biometryczne mogą dotyczyć jednej formy (np. wizerunki twarzy) lub wielu form (np. informacje o twarzy w połączeniu z elektroencefalogramem (EEG)). W motywie 18 podano jako przykłady wyraz twarzy, gesty takie jak ruch rąk, czy cechy głosu danej osoby.

Na przykład:

– system AI wnioskujący emocje na podstawie tekstu pisemnego (analizy treści/nastrojów) w celu określenia stylu lub tonu danego artykułu nie opiera się na danych biometrycznych i w związku z tym nie wchodzi w zakres zakazu;

– system AI wnioskujący emocje na podstawie sposobu pisania na klawiaturze (uderzeń w klawisze), wyrazu twarzy, postaw lub ruchów ciała opiera się na danych biometrycznych i wchodzi w zakres zakazu.

---

155 Art. 3 pkt 34 aktu w sprawie sztucznej inteligencji: „dane biometryczne” oznaczają dane osobowe będące wynikiem specjalnego przetwarzania technicznego, które dotyczą cech fizycznych, fizjologicznych lub behawioralnych osoby fizycznej, takich jak wizerunek twarzy lub dane daktyloskopijne. Zob. również motyw 18 aktu w sprawie sztucznej inteligencji w odniesieniu do wnioskowania o emocjach na podstawie głosu i mowy.

156 Physiological and Behavioural Biometrics [Biometria fizjologiczna i behawioralna] – Biometrics Institute

157 Physiological and Behavioural Biometrics [Biometria fizjologiczna i behawioralna] – Biometrics Institute

158 Zob. EIOD, TechDispatch 1/2024 – Neurodata [Neurodane], 3.06.2024, w którym omówiono wykorzystanie danych mózgu i powiązanych technologii, a także skutki prawne, w tym propozycję nowych „neuropraw”, takich jak prywatność i integralność psychiczna. W S. O'Sullivan, H. Chneiweiss, A. Pierucci i K. Rommelfanger, *Neurotechnologies and Human Rights Framework: Do we need new Human Rights?* [Neurotechnologie i ramy praw człowieka: czy potrzebujemy nowych praw człowieka?], sprawozdanie OECD i Rady Europy z 9.11.2021, s. 33, w którym omówiono najnowocześniejsze neurotechnologie i ich aspekty prawne.

159 Zob. Hasnul et al., 2021, Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare [Systemy rozpoznawania emocji oparte na elektrokardiogramie i ich zastosowania w opiece zdrowotnej].

252) Definicja danych biometrycznych zawarta w akcie w sprawie sztucznej inteligencji jest zatem szeroka i obejmuje wszelkie dane biometryczne wykorzystywane do rozpoznawania emocji, kategoryzacji biometrycznej lub innych celów[160].

7.2.2. Ograniczenie zakazu do miejsc pracy i instytucji edukacyjnych

253) Zakaz określony w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji ogranicza się do systemów rozpoznawania emocji „w miejscu pracy lub instytucjach edukacyjnych". Jak wyjaśniono w motywie 44 aktu w sprawie sztucznej inteligencji, ograniczenie to ma na celu rozwiązanie problemu braku równowagi sił w kontekście pracy lub edukacji.

a) „Miejsce pracy"

254) Pojęcie „miejsca pracy" należy interpretować szeroko. Odnosi się ono do każdej konkretnej przestrzeni fizycznej lub wirtualnej, w której osoby fizyczne wykonują zadania i obowiązki powierzone im przez pracodawcę lub organizację, z którą są powiązane, na przykład w przypadku samozatrudnienia. Obejmuje ono wszelkie miejsca, w których wykonywana jest praca, i może się znacznie różnić w zależności od charakteru pracy, od pomieszczeń biurowych, fabryk i magazynów, przez przestrzenie publiczne, takie jak sklepy, stadiony lub muzea, po miejsca na wolnym powietrzu lub samochody, a także tymczasowe lub mobilne miejsca pracy. Jest to niezależne od statusu pracownika, wykonawcy, stażysty, wolontariusza itp.[161] Pojęcie „miejsca pracy" w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji należy również rozumieć jako mające zastosowanie do rekrutacji i wyboru kandydatów, zgodnie z innymi przepisami aktu w sprawie sztucznej inteligencji dotyczącymi wprowadzania do obrotu, oddawania do użytku lub wykorzystywania systemów AI w obszarze zatrudnienia, zarządzania pracownikami i dostępu do samozatrudnienia, jako że istnieje brak równowagi sił, a systemy rozpoznawania emocji mogą wykazywać inwazyjny charakter już na etapie rekrutacji.

Na przykład:

– korzystanie z kamer internetowych i systemów rozpoznawania głosu przez centrum telefoniczne w celu śledzenia emocji pracowników, np. złości, jest zakazane[162]. Jeśli systemy rozpoznawania emocji są stosowane wyłącznie do celów treningu osobistego, są one dozwolone, pod warunkiem że wyniki nie są przekazywane osobom odpowiedzialnym za zasoby kadrowe i nie mają wpływu na ocenę, awans itp. osoby

[160] W akcie w sprawie sztucznej inteligencji definicja danych biometrycznych nie zawiera sformułowania „które umożliwiają lub potwierdzają jednoznaczną identyfikację" (funkcjonalne wykorzystanie danych biometrycznych), w przeciwieństwie do definicji danych biometrycznych zawartej w RODO, która zawiera ten wymóg. Definicja danych biometrycznych zawarta w RODO będzie miała zastosowanie na podstawie przepisów o ochronie danych w odniesieniu do przetwarzania danych osobowych (oraz kiedy zastosowanie miałby na przykład art. 9 ust. 1 i 2 RODO).

[161] Zob. również motywy dotyczące systemów AI wysokiego ryzyka w miejscu pracy, takie jak motyw 56, w których zastosowano szeroką wykładnię. Zob. również wykaz systemów AI wysokiego ryzyka w załączniku III, w którym w pkt 4 mowa o samozatrudnieniu. Samozatrudnienie jest również szeroko chronione unijnymi przepisami antydyskryminacyjnymi.

[162] Przykład z Boyd et al., 2023, Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work [Automatyczne rozpoznawanie emocji w miejscu pracy: jak proponowane technologie wskazują na potencjalną przyszłość pracy].

szkolonej, przy czym zakaz nie może być obchodzony, a stosowanie systemu rozpoznawania emocji nie może mieć wpływu na stosunki pracy.

– korzystanie przez centrum telefoniczne z systemów rozpoznawania głosu do śledzenia emocji klientów, takich jak gniew czy zniecierpliwienie, (na przykład w celu pomocy pracownikom w radzeniu sobie ze zdenerwowanymi klientami) nie jest zakazane na mocy art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji;

– systemy AI monitorujące ton emocjonalny w zespołach pracujących hybrydowo poprzez rozpoznawanie emocji i wnioskowanie na ich temat na podstawie głosu i obrazów hybrydowych połączeń wideo, co zazwyczaj służyłoby zwiększaniu świadomości społecznej, zarządzaniu dynamiką emocjonalną i zapobieganiu konfliktom, są zakazane;

– stosowanie systemów AI do rozpoznawania emocji podczas procesu rekrutacji jest zakazane;

– stosowanie systemów AI do rozpoznawania emocji w okresie próbnym jest zakazane;

– korzystanie przez supermarket z kamer w celu śledzenia emocji pracowników, takich jak zadowolenie, jest zakazane;

– korzystanie z kamer przez supermarket lub bank w celu wykrywania podejrzanych klientów, na przykład aby stwierdzić, że dana osoba zamierza dokonać rozboju, nie jest zakazane na mocy art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, jeżeli zapewnia się, by nie był śledzony żaden pracownik, i jeżeli istnieją wystarczające zabezpieczenia.

b) „Instytucje edukacyjne”

255) Odniesienie do **instytucji edukacyjnych** ma szeroki zakres i należy je rozumieć jako obejmujące zarówno instytucje publiczne, jak i prywatne. Nie ma ograniczeń w odniesieniu do typów lub wieku uczniów lub studentów ani do określonego środowiska (online, stacjonarne, tryb mieszany[163], itp.). Na przykład w zakres zakazu określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji wchodzą instytucje edukacyjne i szkoleniowe na wszystkich szczeblach, w tym szkoły zawodowe, tj. szkoły, w których uczniowie zdobywają umiejętności wymagające używania rąk[164] i szkoły kształcenia ustawicznego[165]. Instytucje edukacyjne są zazwyczaj akredytowane lub uznawane przez właściwe krajowe organy oświatowe lub równoważne władze. Kluczową cechą jest to, że instytucje edukacyjne mogą wydawać świadectwa (a uczęszczanie do nich jest warunkiem niezbędnym do uzyskania

[163] Kształcenie w trybie mieszanym należy rozumieć jako przyjęcie więcej niż jednego podejścia w procesie kształcenia i szkolenia, w tym łączenie narzędzi uczenia się cyfrowego (w tym e-uczenia się) i niecyfrowego.

[164] Zob. np. ocena skutków towarzysząca wnioskowi Komisji, gdzie wymieniono konkretne zastosowania sztucznej inteligencji przez instytucje kształcenia zawodowego jako powodujące poważną ingerencję w szeroki zakres praw podstawowych, np. podczas oceniania: Komisja UE, *Dokument roboczy służb Komisji. Ocena skutków. Załączniki, SWD(2021) 84 final, część 2/2,* s. 43. Zob. również I. Tuomi, *The use of artificial Intelligence (AI) in education* [Wykorzystanie sztucznej inteligencji (AI) w edukacji], Parlament Europejski, 2020, s. 9–10.

[165] Zob. art. 14 Karty.

świadectwa). Należy rozumieć, że zakaz ma również zastosowanie do kandydatów w trakcie procesu rekrutacji.

Na przykład:

– oparta na sztucznej inteligencji aplikacja wykorzystująca rozpoznawanie emocji do nauki języka online poza instytucją edukacyjną nie jest zakazana na podstawie art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji. Natomiast jeżeli uczniowie są zobowiązani do korzystania z aplikacji instytucji edukacyjnej, stosowanie takiego systemu rozpoznawania emocji jest zakazane;

– korzystanie przez instytucję edukacyjną z opartego na sztucznej inteligencji oprogramowania do śledzenia ruchu gałek ocznych podczas egzaminowania uczniów online w celu monitorowania punktu fiksacji i ruchu gałek ocznych (np. aby wykryć korzystanie z niedozwolonych materiałów) nie jest zakazane, ponieważ system nie wykrywa emocji ani nie wyciąga wniosków na ich temat. Natomiast jeżeli system jest również wykorzystywany do wykrywania emocji, takich jak pobudzenie emocjonalne i niepokój, mieści się to w zakresie zakazu;

– stosowanie przez instytucję edukacyjną systemu AI do rozpoznawania emocji w celu wnioskowania o zainteresowaniu i uwadze uczniów jest zakazane. Natomiast jeżeli systemy rozpoznawania emocji są wykorzystywane wyłącznie do celów uczenia się w kontekście odgrywania ról (na przykład dla podmiotów prowadzących szkolenia lub wykładowców), są one dozwolone pod warunkiem, że wyniki nie będą mieć wpływu na ocenę lub certyfikację szkolonej osoby;

– stosowanie przez instytucję edukacyjną systemu AI do rozpoznawania emocji kandydatów na uczniów podczas egzaminów wstępnych jest zakazane;

– korzystanie przez instytucję edukacyjną z systemu AI, który umożliwia wykrywanie rozmów prowadzonych przez uczniów za pośrednictwem telefonów lub innych kanałów podczas wykładów online, nie jest zakazane, ponieważ nie wiąże się z wnioskowaniem emocji. Natomiast jeżeli system jest również wykorzystywany do wnioskowania emocji, takich jak pobudzenie emocjonalne, niepokój czy zainteresowanie, chodzi to w zakres zakazu;

– wykorzystywanie przez instytucję edukacyjną systemu AI do rozpoznawania emocji zarówno u nauczycieli (miejsce pracy), jak i u uczniów (edukacja) jest zakazane.

7.2.3. Wyjątki ze względów medycznych i bezpieczeństwa

256) Zakaz zawarty w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji zawiera wyraźny wyjątek dotyczący systemów rozpoznawania emocji wykorzystywanych w miejscu pracy i instytucjach edukacyjnych ze względów medycznych lub bezpieczeństwa, takich jak systemy przeznaczone do użytku terapeutycznego[166]. W świetle celu aktu w sprawie sztucznej inteligencji, jakim jest zapewnienie wysokiego

[166] Motyw 44 aktu w sprawie sztucznej inteligencji.

poziomu ochrony praw podstawowych, wyjątek ten powinien podlegać ścisłej wykładni.

257) W szczególności użytek terapeutyczny należy rozumieć jako zastosowanie wyrobów medycznych z oznakowaniem CE. Ponadto wyjątek ten nie obejmuje stosowania systemów rozpoznawania emocji w celu wykrywania ogólnych aspektów dobrostanu. Ogólne monitorowanie poziomów stresu w miejscu pracy nie jest dozwolone jako wyjątek ze względów medycznych lub bezpieczeństwa. Na przykład system AI przeznaczony do wykrywania wypalenia zawodowego lub depresji w miejscu pracy lub w instytucjach edukacyjnych nie byłby objęty wyjątkiem i pozostałby zakazany.

258) Pojęcie względów bezpieczeństwa w ramach tego wyjątku należy rozumieć jako mające zastosowanie wyłącznie do ochrony życia i zdrowia, a nie do ochrony innych interesów, na przykład mienia przed kradzieżą lub oszustwem.

259) Z tej ścisłej wykładni wyjątku wynika, że wszelkie wykorzystanie ze względów medycznych i bezpieczeństwa powinno zawsze ograniczać się do tego, co jest absolutnie konieczne i proporcjonalne, w tym do ograniczeń czasowych, indywidualnego stosowania i skali, oraz że powinny mu towarzyszyć wystarczające zabezpieczenia. Takie zabezpieczenia mogą obejmować na przykład uprzednią pisemną i uzasadnioną opinię eksperta dotyczącą konkretnego przypadku użycia. Konieczność powinna być oceniana w sposób obiektywny w odniesieniu do względów medycznych i bezpieczeństwa, a nie odnosić się do „potrzeb" pracodawcy lub instytucji edukacyjnej. W ramach tej oceny należy zbadać, czy istnieją mniej inwazyjne alternatywne środki, które pozwoliłyby osiągnąć ten sam cel.

260) Pracodawcy i pracownicy pedagogiczni powinni stosować systemy rozpoznawania emocji wyłącznie ze względów medycznych i bezpieczeństwa w przypadku wyraźnej potrzeby[167]. Dane gromadzone i przetwarzane w tym kontekście nie mogą być wykorzystywane do żadnych innych celów. Jest to szczególnie ważne, biorąc pod uwagę, że stosowanie opartego na AI oprogramowania do zarządzania w miejscu pracy okazało się mieć potencjalnie negatywny wpływ na zdrowie i bezpieczeństwo pracowników. Na przykład ciągłe monitorowanie za pomocą urządzeń ubieralnych może zwiększyć stres w pracy, a jednocześnie negatywnie wpływać na wydajność[168].

261) Ze względu na fakt, iż w motywie 18 aktu w sprawie sztucznej inteligencji wyłączono z definicji systemów rozpoznawania emocji stany fizyczne, takie jak ból czy zmęczenie, szereg systemów AI wykorzystywanych ze względów bezpieczeństwa nie wchodziłoby już w zakres tej definicji, w tym na przykład systemy wykorzystywane do wykrywania stanu zmęczenia zawodowych pilotów lub kierowców w celu zapobiegania wypadkom.

---

167 Zgodnie z unijnym prawem pracy jeżeli takie nowe technologie są wprowadzane, pracodawcy konsultują się również z pracownikami lub ich przedstawicielami, i przestrzegają procedur krajowych. Bez przestrzegania wymogów proceduralnych takie systemy nie mogą zostać wprowadzone poprzez odniesienie do aktu w sprawie sztucznej inteligencji jako takiego. Będą one również wymagać zgody z punktu widzenia przepisów o ochronie danych, które nadal pozostają w mocy.

168 Powiązanie między aktem w sprawie sztucznej inteligencji a unijną strategią w zakresie bezpieczeństwa i zdrowia w pracy – globalne prawo i polityka w miejscu pracy (kluwerlawonline.com).

262) Inne przepisy, w tym przepisy o ochronie danych, nadal mają zastosowanie do systemów rozpoznawania emocji, które spełniają warunki wyjątku określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji[169].

263) Systemy rozpoznawania emocji, które klasyfikuje się jako systemy wysokiego ryzyka zgodnie z art. 6 ust. 2 aktu w sprawie sztucznej inteligencji i pkt 1 lit. c) załącznika III do tego aktu, będą musiały spełniać wymogi dotyczące wysokiego ryzyka określone w rozdziale III sekcja 2 aktu w sprawie sztucznej inteligencji oraz obowiązek w zakresie przejrzystości określony w art. 50 ust. 3 aktu w sprawie sztucznej inteligencji.

Na przykład:

Rozpoznawanie emocji może być stosowane ze względów medycznych, aby pomóc pracownikom lub uczniom z autyzmem i poprawić dostępność dla osób niewidomych lub niesłyszących[170]. Takie zastosowania wchodziłyby w zakres wyjątku ze względów medycznych określonego w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji.

Natomiast rozpoznawanie emocji w celu oceny dobrostanu studentów lub pracowników, ich poziomu motywacji oraz satysfakcji z pracy lub nauki nie kwalifikuje się jako „zastosowanie ze względów medycznych", więc byłoby zakazane.

Pracodawca nie mógłby stosować w miejscu pracy wspomaganych sztuczną inteligencją urządzeń lub cyfrowych asystentów do pomiaru niepokoju na podstawie zmierzonych poziomów stresu lub do pomiaru nudy wśród pracowników, chyba że podwyższony poziom stresu/brak skupienia stanowiłby szczególne zagrożenie, na przykład podczas stosowania niebezpiecznych maszyn lub niebezpiecznych substancji chemicznych. W tym drugim przypadku pracodawca nie może wykorzystywać danych do innych celów, takich jak ocena wyników pracy pracownika.

7.3. Korzystniejsze prawo państwa członkowskiego

264) Art. 2 ust. 11 aktu w sprawie sztucznej inteligencji stanowi, że Unia lub państwa członkowskie mogą utrzymywać lub wprowadzać przepisy ustawowe, wykonawcze lub administracyjne, które są korzystniejsze dla pracowników pod względem ochrony ich praw w odniesieniu do korzystania z systemów AI przez pracodawców. Można również zezwalać na stosowanie układów zbiorowych, które są korzystniejsze dla pracowników, lub zachęcać do stosowania takich układów.

Na przykład państwa członkowskie mogą przyjąć przepisy przewidujące, że stosowanie systemów rozpoznawania emocji w dziedzinie pracy nie może być stosowane do celów medycznych.

7.4. Systemy nieobjęte zakresem

265) Jak wspomniano powyżej, zakresem zakazu nie są objęte:

[169] Od grudnia 2026 roku zacznie obowiązywać dyrektywa (UE) 2024/2831 Parlamentu Europejskiego i Rady z dnia 23 października 2024 r. w sprawie poprawy warunków pracy za pośrednictwem platform.

[170] Systemy te mogłyby być z powodzeniem wykorzystywane do pomagania pracownikom lub studentom/uczniom w zrozumieniu emocji kolegów itp.

- systemy AI wnioskujące emocje i nastroje na podstawie danych innych niż biometryczne,
- systemy AI wnioskujące stany fizyczne, takie jak ból i zmęczenie.

266) Systemy rozpoznawania emocji stosowane we wszystkich innych kontekstach niż miejsce pracy i instytucje edukacyjne nie są objęte zakazem określonym w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji. Systemy te uznaje się jednak za systemy AI wysokiego ryzyka[171]. Jednocześnie takie systemy mogą być w niektórych przypadkach zakazane na mocy art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji (szkodliwe techniki manipulacyjne i wykorzystywanie) lub na mocy innych przepisów Unii. Wszystkie inne obowiązujące przepisy, takie jak unijne przepisy dotyczące ochrony danych, ochrony konsumentów itp., nadal mają zastosowanie do takich systemów.

Na przykład:

Systemy rozpoznawania emocji wykorzystywane w kontekście handlowym do obsługi klientów nie są objęte zakazem określonym w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, niezależnie od tego, czy opierają się na danych biometrycznych, czy też nie. W związku z tym zakaz nie obejmuje systemów takich jak systemy AI, które umożliwiają rozpoznawanie emocji na podstawie uderzeń w klawisze lub na podstawie wiadomości głosowych klientów (np. podczas pisania na czacie, korzystania z wirtualnych asystentów głosowych) i które są wykorzystywane w marketingu internetowym do wyświetlania spersonalizowanych wiadomości i do celów reklamowych, w tym w inteligentnych środowiskach („inteligentne billboardy").

Takie praktyki mogą jednak być objęte zakazami szkodliwej manipulacji i wykorzystywania określonymi w art. 5 ust. 1 lit. a) i b) aktu w sprawie sztucznej inteligencji[172], jeżeli spełnione są wszystkie warunki wymagane, by zakazy te miały zastosowanie.

a) Inne systemy nieobjęte zakresem

267) Termin „kontrola tłumu" odnosi się zasadniczo do kontroli i monitorowania zachowania grup w celu utrzymania porządku (publicznego) i zapewnienia bezpieczeństwa podczas wydarzeń. Często wiąże się to z dużymi wydarzeniami masowymi (np. meczami piłkarskimi, koncertami itp.) lub określonymi miejscami, takimi jak porty lotnicze lub pociągi. Działanie systemów kontroli tłumu nie musi obejmować wyciągania wniosków na temat emocji poszczególnych osób, na przykład kiedy takie systemy są wykorzystywane do analizy ogólnego poziomu hałasu i nastroju w danym miejscu. W takim przypadku system nie wchodziłby w zakres art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, ponieważ nie wyciąga wniosków na temat emocji (konkretnej) osoby fizycznej.

[171] Art. 6 ust. 2 aktu w sprawie sztucznej inteligencji oraz załącznik III pkt 1 lit. c).

[172] Sytuacje takie mogą być również zakazane na mocy innych przepisów, takich jak przepisy dotyczące ochrony danych lub ochrony konsumentów.

268) Mogą jednak wystąpić sytuacje, gdy systemy kontroli tłumu wyciągają wnioski na temat emocji poszczególnych osób, na przykład oceniają, czy jest wiele zdenerwowanych twarzy. Zazwyczaj takie systemy AI nie byłyby objęte zakazem określonym w art. 5 ust. 1 lit. f) aktu w sprawie sztucznej inteligencji, ponieważ z reguły nie są one wykorzystywane w miejscu pracy ani w instytucjach edukacyjnych.

269) Poza zakresem są również systemy stosowane w dziedzinie medycyny, na przykład roboty opiekuńcze lub systemy rozpoznawania emocji stosowane przez lekarzy podczas badań w ich miejscu pracy oraz monitory głosu, które analizują zgłoszenia alarmowe.

270) Takie systemy często będą monitorować osoby znajdujące się w danym miejscu w kontekście pracy, na przykład pracowników ochrony na stadionie piłkarskim lub na dworcu centralnym (w przypadku gdy takie systemy są wykorzystywane do rozpoznawania agresywnych zachowań) lub personel medyczny. W takich przypadkach podmioty stosujące muszą wdrożyć odpowiednie zabezpieczenia, aby uniknąć monitorowania pracowników. Nie można jednak całkowicie zapobiec wyciąganiu przez takie systemy wniosków na temat emocji tych pracowników. Ze względu na fakt, że głównym celem systemu nie jest ocena emocji pracowników, należy uznać, że systemy te nie wchodzą w zakres zakazu. Podmioty stosujące takie systemy pozostają odpowiedzialne za zapewnienie, aby ich użytkowanie nie miało negatywnego wpływu na pracowników.

8. ART. 5 UST. 1 LIT. G) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI: KATEGORYZACJA BIOMETRYCZNA NIEKTÓRYCH „WRAŻLIWYCH" CECH

271) W art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji zakazano stosowania systemów kategoryzacji biometrycznej, które indywidualnie kategoryzują osoby fizyczne w oparciu o ich dane biometryczne, by wydedukować lub wywnioskować informacje na temat ich rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej. Zakaz ten nie obejmuje etykietowania, filtrowania ani kategoryzacji zbiorów danych biometrycznych pozyskanych zgodnie z prawem Unii lub prawem krajowym, które mogą być wykorzystywane, na przykład, w obszarze ścigania przestępstw[173].

8.1. Uzasadnienie i cele

272) Z danych biometrycznych można wyodrębnić, wydedukować lub wywnioskować wiele różnych informacji, w tym informacji „wrażliwych", nawet bez wiedzy zainteresowanych osób, w celu sklasyfikowania tych osób. Może to prowadzić do niesprawiedliwego i dyskryminacyjnego traktowania, na przykład gdy danej osobie odmawia się usługi, ponieważ uznano ją za osobę określonej rasy. Oparte na sztucznej inteligencji systemy kategoryzacji biometrycznej służące do przypisywania osób

[173] Motyw 30 aktu w sprawie sztucznej inteligencji.

fizycznych do określonych grup lub kategorii, które odnoszą się do takich aspektów jak orientacja seksualna, orientacja polityczna czy rasa, naruszają ludzką godność i stwarzają poważne ryzyko dla innych praw podstawowych, takich jak prywatność i niedyskryminacja. Są one zatem zakazane na mocy art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji.

8.2. Główne pojęcia i elementy zakazu

> ***Art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji***
>
> Zakazuje się następujących praktyk w zakresie AI:
>
> g) wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu lub wykorzystywania systemów kategoryzacji biometrycznej, które indywidualnie kategoryzują osoby fizyczne *w oparciu o ich dane biometryczne*, by wydedukować lub wywnioskować informacje na temat ich rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej; zakaz ten nie obejmuje przypadków etykietowania ani filtrowania pozyskanych zgodnie z prawem zbiorów danych biometrycznych, takich jak obrazy, w oparciu o dane biometryczne, ani kategoryzacji danych biometrycznych w obszarze ścigania przestępstw;

273) Aby zakaz określony w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi zostać spełniony łącznie szereg warunków:

(i) praktyka musi polegać na „wprowadzaniu do obrotu", „oddawaniu do użytku w tym konkretnym celu" lub „wykorzystywaniu" systemu AI;

(ii) system musi być systemem kategoryzacji biometrycznej;

(iii) musi on kategoryzować poszczególne osoby;

(iv) w oparciu o ich dane biometryczne;

(v) by wydedukować lub wywnioskować informacje na temat ich rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej.

274) Aby zakaz ten miał zastosowanie, wszystkie pięć warunków musi być spełnione jednocześnie. Pierwszy warunek, tj. wprowadzanie do obrotu, oddawanie do użytku lub wykorzystywanie systemu AI, przeanalizowano w sekcji 2.3. Zakaz ten ma zatem zastosowanie zarówno do dostawców systemów AI, jak i podmiotów je stosujących, z których żaden, działając w zakresie swoich odpowiednich obowiązków, nie może wprowadzać takich systemów AI do obrotu, oddawać ich do użytku ani wykorzystywać. Pozostałe warunki, które muszą być spełnione, aby zakaz miał zastosowanie[174], opisano i przeanalizowano poniżej.

[174] Kryterium „systemu AI", „wprowadzania do obrotu", „oddawania do użytku w tym konkretnym celu" lub „wykorzystywania" – zob. *powyżej*.

275) Zakaz nie obejmuje etykietowania ani filtrowania legalnie pozyskanych zbiorów danych biometrycznych, w tym na potrzeby ścigania przestępstw.

8.2.1. System kategoryzacji biometrycznej

276) Kategoryzacja osób fizycznych w systemie biometrycznym stanowi zwykle proces polegający na określeniu, czy dane biometryczne danej osoby fizycznej należą do grupy o wcześniej ustalonych cechach. Nie chodzi o identyfikację danej osoby fizycznej lub weryfikację jej tożsamości, lecz o przyporządkowanie jej do określonej kategorii. Przykładowo w okienku reklamowym mogą pojawiać się różne reklamy w zależności od osoby, która na nie patrzy, uwzględniając jej wiek lub płeć[175]. Osoby można również kategoryzować po prostu do celów statystycznych, nie identyfikując ich i nie mając na celu ich identyfikacji.

277) W art. 3 pkt 40 aktu w sprawie sztucznej inteligencji zdefiniowano system kategoryzacji biometrycznej jako system AI służący do przypisywania osób fizycznych do określonych kategorii na podstawie danych biometrycznych tych osób, oprócz przypadków, gdy taki system pełni funkcję pomocniczą w stosunku do innej usługi komercyjnej i jest bezwzględnie konieczny z obiektywnych względów technicznych. Jak wyjaśniono w sekcji 7.2.1 lit. d), „dane biometryczne" zdefiniowano w art. 3 pkt 34 aktu w sprawie sztucznej inteligencji. W szczególności dane biometryczne obejmują cechy behawioralne oparte na cechach biometrycznych. Zakres kategoryzacji biometrycznej nie obejmuje kategoryzacji według ubrań czy akcesoriów, takich jak chusty lub krzyże, a także aktywności w mediach społecznościowych.

278) Kategoryzacja biometryczna może opierać się na kategoriach dotyczących cech fizycznych (np. rysów twarzy i jej kształtu, koloru skóry), na podstawie których osoby są przyporządkowywane do określonych kategorii. Niektóre z tych kategorii mogą mieć szczególny „wrażliwy" charakter lub obejmować cechy chronione na mocy unijnego prawa o niedyskryminacji, takie jak rasa. Kategoryzacja biometryczna może jednak również opierać się na DNA lub aspektach behawioralnych, takich jak dynamika pisania na klawiaturze lub chód danej osoby[176].

279) Aby system nie wchodził w zakres definicji systemu kategoryzacji biometrycznej zawartej w akcie w sprawie sztucznej inteligencji, muszą być spełnione łącznie dwa warunki – system taki musi pełnić „funkcję pomocniczą w stosunku do innej usługi komercyjnej" i być „bezwzględnie konieczny z obiektywnych względów technicznych".

280) Zgodnie z motywem 16 aktu w sprawie sztucznej inteligencji, funkcja jedynie pomocnicza to funkcja, która jest nieodłącznie związana z inną usługą komercyjną, co oznacza, że z obiektywnych względów technicznych funkcja ta nie może być wykorzystywana bez usługi głównej, a włączenie takiej funkcji lub funkcjonalności nie jest sposobem na obejście stosowania przepisów aktu w sprawie sztucznej inteligencji.

---

175 Grupa Robocza Art. 29, *Opinia 3/2012 w sprawie zmian sytuacji w dziedzinie technologii biometrycznych*, WP193, 27.04.2012, s. 6.

176 Zob. np. Grupa Robocza Art. 29, *Opinia 3/2012 w sprawie zmian sytuacji w dziedzinie technologii biometrycznych*, WP193, 27.04.2012, s. 16–17. Grupa odnosi się tutaj do rozpoznawania „miękkiego" (s. 17), tj. „wykrywania zachowania lub szczególnych potrzeb ludzi".

Na przykład zgodnie z art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji dozwolone są następujące zastosowania sztucznej inteligencji:

– filtry kategoryzujące cechy twarzy lub ciała, wykorzystywane na internetowych platformach handlowych, aby umożliwić konsumentowi wirtualne przymierzenie na sobie produktu, mogą stanowić taką funkcję pomocniczą, ponieważ mogą być stosowane wyłącznie w odniesieniu do usługi głównej polegającej na sprzedaży produktu;

– filtry stosowane w internetowych serwisach społecznościowych, które kategoryzują cechy twarzy lub ciała, aby umożliwić użytkownikom dodawanie lub modyfikowanie zdjęć lub filmów wideo, można również uznać za funkcję pomocniczą, ponieważ filtry takie nie mogą być stosowane bez usługi głównej polegającej na udostępnianiu treści online w ramach serwisu społecznościowego.

Natomiast przykłady zastosowań, które byłyby zakazane, obejmują:

– system AI, który analizując dane biometryczne osób aktywnych na platformie mediów społecznościowych pochodzące z umieszczonych przez nie na platformie zdjęć, kategoryzuje te osoby według ich zakładanej orientacji politycznej w celu wysyłania im ukierunkowanych komunikatów politycznych. Chociaż taki system może pełnić jedynie funkcję pomocniczą w stosunku do reklamy politycznej, nie jest on „bezwzględnie konieczny z obiektywnych względów technicznych”, w związku z czym warunki wyłączenia go z definicji kategoryzacji biometrycznej nie są spełnione;

– system AI, który analizując dane biometryczne osób aktywnych na platformie mediów społecznościowych pochodzące z udostępnionych przez nie na platformie zdjęć, kategoryzuje te osoby według ich zakładanej orientacji seksualnej i na tej podstawie wyświetla tym osobom reklamy, kwalifikowałby się jako system kategoryzacji biometrycznej w rozumieniu aktu w sprawie sztucznej inteligencji. Również w tym przypadku ta „usługa pomocnicza” nie jest bezwzględnie konieczna, w związku z czym wyłączenie z zakazu nie ma zastosowania.

8.2.2. Indywidualna kategoryzacja osób w oparciu o ich dane biometryczne

281) Podstawowym warunkiem, który musi zostać spełniony, aby zakaz miał zastosowanie, jest wykorzystywanie danych biometrycznych do kategoryzacji osób fizycznych (zob. sekcja 8.2.1 i 7.2.1 lit. d) powyżej).

282) Ponadto, aby zakaz miał zastosowanie, osoby fizyczne muszą być kategoryzowane „indywidualnie”. Jeśli celem lub rezultatem kategoryzacji biometrycznej nie jest taka indywidualna kategoryzacja, na przykład jeśli kategoryzacji poddawana jest cała grupa bez uwzględniania poszczególnych osób, zakaz nie ma zastosowania.

Przykłady indywidualnej kategoryzacji obejmują:

– systemy AI, które dokonują „szacowania atrybutów” (obliczeń demograficznych), w tym na przykład „wieku, płci, pochodzenia etnicznego”, na podstawie na przykład

cech ciała, takich jak twarz, wzrost lub kolor skóry, oczu i włosów (lub połączenia tych cech);

– systemy AI zdolne do kategoryzacji osób i wyodrębniania ich w oparciu o określoną cechę (np. bliznę pod prawym okiem) lub dlatego, że posiadają tatuaż na prawej ręce.

Takie przypadki użycia są przykładami indywidualnej kategoryzacji biometrycznej. Aby przykłady te wchodziły w zakres zakazu określonego w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji, muszą być spełnione wszystkie warunki określone w tym przepisie.

8.2.3. Dedukowanie lub wnioskowanie informacji na temat rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej

283) W art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji zakazano wyłącznie systemów kategoryzacji biometrycznej, które służą do dedukowania lub wnioskowania informacji dotyczących ograniczonej liczby wrażliwych cech: rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej.

Na przykład systemy zakazane na mocy art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji obejmują:

– system kategoryzacji biometrycznej, który ma pozwalać na wydedukowanie rasy danej osoby na podstawie jej głosu (jest to inny przypadek niż kategoryzowanie osób na podstawie koloru skóry lub oczu czy też badanie DNA ofiar przestępstw pod kątem ich pochodzenia; takie systemy nie byłyby zakazane);

– system kategoryzacji biometrycznej, który ma pozwalać na wydedukowanie orientacji religijnej danej osoby na podstawie jej tatuaży lub twarzy.

8.3. Systemy nieobjęte zakresem

284) Zakaz określony w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji nie obejmuje systemów AI etykietujących lub filtrujących pozyskane zgodnie z prawem zbiory danych biometrycznych, takie jak obrazy, w oparciu o dane biometryczne, ani kategoryzacji danych biometrycznych w obszarze ścigania przestępstw. Wyjaśniono to bardziej szczegółowo w motywie 30 aktu w sprawie sztucznej inteligencji[177].

285) Etykietowanie lub filtrowanie zbiorów danych biometrycznych może odbywać się za pomocą systemów kategoryzacji biometrycznej w celu zapewnienia, aby dane w równym stopniu reprezentowały wszystkie grupy demograficzne, a nie, na przykład, nadmiernie reprezentowały tylko jedną określoną grupę. Jeżeli dane wykorzystywane do trenowania algorytmu są nieobiektywne w stosunku do konkretnej grupy (tj. istnieją

[177] Motyw 30 aktu w sprawie sztucznej inteligencji: „Zakaz ten nie powinien obejmować zgodnego z prawem etykietowania, filtrowania lub kategoryzacji zbiorów danych biometrycznych, pozyskanych zgodnie z prawem Unii lub prawem krajowym, według danych biometrycznych, takiego jak sortowanie obrazów według koloru włosów lub koloru oczu, które można na przykład wykorzystać w obszarze ścigania przestępstw”.

systematyczne różnice w danych dotyczących poszczególnych grup ze względu na sposób gromadzenia tych danych lub dane są historycznie stronnicze), algorytm może powielać tę stronniczość, co może prowadzić do bezprawnej dyskryminacji osób lub grup osób[178]. Z tego powodu etykietowanie na podstawie szczególnie chronionych informacji wrażliwych może być konieczne w przypadku wysokiej jakości danych, właśnie w celu zapobiegania dyskryminacji. Zgodnie z aktem w sprawie sztucznej inteligencji może być nawet wymagane, aby operacje etykietowania były zgodne z wymogami aktu dotyczącymi systemów AI wysokiego ryzyka[179]. Etykietowanie lub filtrowanie danych biometrycznych jest zatem wyraźnie wyłączone z zakazu określonego w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji. Zakaz ten ma zastosowanie wyłącznie w przypadku, gdy celem kategoryzacji danych biometrycznych jest wywnioskowanie informacji na temat rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej.

Przykłady dopuszczalnego etykietowania lub filtrowania obejmują:

– etykietowanie danych biometrycznych w celu uniknięcia przypadków, w których członek grupy etnicznej ma mniejsze szanse na zaproszenie na rozmowę kwalifikacyjną, ponieważ algorytm został „wytrenowany” na podstawie danych, w których dana grupa osiąga gorsze wyniki, tj. ma gorsze wyniki niż inne grupy[180];

– kategoryzacja pacjentów na podstawie obrazów w oparciu o ich kolor skóry lub oczu może być istotna dla diagnozy medycznej, np. diagnozy raka.

286) Art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji stanowi również, że zakaz zawarty w tym przepisie nie ma zastosowania do etykietowania lub filtrowania pozyskanych zgodnie z prawem zbiorów danych w obszarze ścigania przestępstw[181].

Dotyczy to na przykład korzystania przez organ ścigania z systemu AI, który umożliwia etykietowanie i filtrowanie zbioru danych, co do którego istnieje podejrzenie, że zawiera materiały przedstawiające niegodziwe traktowanie dzieci w celach seksualnych. Na pierwszym etapie organy ścigania wykorzystałyby wsparcie systemów AI do wykrywania i utajniania danych wrażliwych na obrazach. Ponadto filtrowanie i etykietowanie według płci, wieku, danych biometrycznych, takich jak kolor oczu i włosów, blizny i znaki szczególne, mogłoby pomóc w identyfikacji ofiar lub stworzeniu powiązań z innymi przypadkami. Podobnie dopuszcza się filtrowanie i etykietowanie rąk osób dopuszczających się nadużyć w oparciu o konkretne cechy, takie jak długość palców czy ewentualne znaki szczególne bądź tatuaże, aby ułatwić identyfikację potencjalnych podejrzanych.

8.4. Wzajemne powiązania z innymi przepisami prawa Unii

178 Ibid.
179 Zob. np. art. 10 i 17 aktu w sprawie sztucznej inteligencji.
180 FRA, *# BigData. Discrimination in data supported decision making* [Dyskryminacja w procesie podejmowania decyzji w oparciu o dane], Luksemburg, 2018, 14, s. 5.
181 Przewidziane w akcie w sprawie sztucznej inteligencji wykorzystywanie systemów kategoryzacji biometrycznej do celów ścigania przestępstw opiera się na art. 16 TFUE. Zob. również motyw 3 aktu w sprawie sztucznej inteligencji.

287) Systemy AI przeznaczone do kategoryzacji biometrycznej na podstawie danych biometrycznych według wrażliwych atrybutów lub cech chronionych na podstawie art. 9 ust. 1 RODO, o ile nie są one zakazane na mocy tego rozporządzenia, należy zaklasyfikować jako systemy wysokiego ryzyka[182] zgodnie z aktem w sprawie sztucznej inteligencji[183].

288) W art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji dodatkowo ograniczono możliwości zgodnego z prawem przetwarzania danych osobowych na podstawie unijnych przepisów o ochronie danych, takich jak RODO, dyrektywa o ochronie danych w sprawach karnych czy rozporządzenia w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych przez instytucje, organy i jednostki organizacyjne Unii. W szczególności w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji wykluczono możliwość kategoryzacji biometrycznej osób fizycznych na podstawie ich danych biometrycznych – jak określono w akcie w sprawie sztucznej inteligencji – aby wywnioskować informacje na temat ich rasy, poglądów politycznych, przynależności do związków zawodowych, przekonań religijnych lub światopoglądowych, seksualności lub orientacji seksualnej, z zastrzeżeniem wyjątku dotyczącego etykietowania lub filtrowania pozyskanych zgodnie z prawem zbiorów danych biometrycznych, w tym w obszarze ścigania przestępstw, jak opisano powyżej. Ponadto zakaz zawarty w art. 5 ust. 1 lit. g) aktu w sprawie sztucznej inteligencji jest zgodny z art. 11 ust. 3 dyrektywy o ochronie danych w sprawach karnych, który wyraźnie zakazuje „profilowania" skutkującego dyskryminacją ze względu na szczególne kategorie danych osobowych, takie jak rasa, pochodzenie etniczne, orientacja seksualna, poglądy polityczne lub przekonania religijne.

9. ART. 5 UST. 1 LIT. H) AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI – SYSTEMY ZDALNEJ IDENTYFIKACJI BIOMETRYCZNEJ W CZASIE RZECZYWISTYM DO CELÓW ŚCIGANIA PRZESTĘPSTW

289) W art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji zakazano wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, z zastrzeżeniem ograniczonych wyjątków określonych w sposób wyczerpujący w akcie w sprawie sztucznej inteligencji. W szczególności w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji przewidziano trzy sytuacje, w których stosowanie takich systemów może być dozwolone, jeżeli zezwala na to ustawodawstwo krajowe oraz gdy spełnione są warunki i zabezpieczenia określone w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji.

290) Zgodnie z art. 5 ust. 5 aktu w sprawie sztucznej inteligencji państwa członkowskie mają swobodę decydowania, czy i w której z tych trzech sytuacji dozwolone jest na ich terytorium wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw. W przypadku braku przepisów krajowych umożliwiających i regulujących takie wykorzystywanie organy ścigania i podmioty działające w ich imieniu nie mogą stosować takich

[182] Motyw 54 aktu w sprawie sztucznej inteligencji i pkt 1 lit. b) załącznika III do tego aktu.
[183] Motyw 54 aktu w sprawie sztucznej inteligencji i pkt 1 lit. b) załącznika III do tego aktu.

systemów do celów ścigania przestępstw. Istnienie przepisów krajowych, które są zgodne z odpowiednimi wymogami aktu w sprawie sztucznej inteligencji, jest zatem warunkiem wstępnym takiego wykorzystywania.

291) W art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji zakazano jedynie wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, w związku z czym przepis ten dotyczy wyłącznie podmiotów stosujących takie systemy. Wprowadzanie do obrotu i oddawanie do użytku takich systemów, a także wykorzystywanie innych systemów zdalnej identyfikacji biometrycznej nie jest zakazane, ale podlega przepisom dotyczącym systemów AI wysokiego ryzyka zgodnie z art. 6 ust. 2 aktu w sprawie sztucznej inteligencji i pkt 1 lit. a) załącznika III do tego aktu[184]. W przypadku gdy państwo członkowskie zezwala na wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w którymkolwiek z trzech celów wymienionych w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji, do takiego wykorzystywania również mają zastosowanie przepisy dotyczące systemów AI wysokiego ryzyka.

292) Ponadto do retrospektywnego wykorzystywania systemów zdalnej identyfikacji biometrycznej do celów ścigania przestępstw mają zastosowanie przepisy szczegółowe. Takie wykorzystywanie w czasie innym niż rzeczywisty nie jest zakazane, ale podlega dodatkowym zabezpieczeniom dotyczącym stosowania systemów AI wysokiego ryzyka (art. 26 ust. 10 aktu w sprawie sztucznej inteligencji).

9.1. Uzasadnienie i cele

293) W motywie 32 aktu o sztucznej inteligencji dostrzeżono inwazyjny charakter systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej wykorzystywanych do ścigania przestępstw, które to systemy szczególnie ingerują w prawa i wolności zainteresowanych osób, do tego stopnia że może to wpływać na życie prywatne dużej części społeczeństwa, wywoływać poczucie stałego nadzoru i pośrednio zniechęcać do korzystania z wolności zgromadzeń i innych praw podstawowych. Techniczne niedokładności systemów AI przeznaczonych do zdalnej identyfikacji biometrycznej osób fizycznych mogą prowadzić do nieobiektywnych wyników i wywoływać skutki w postaci dyskryminacji. Takie ewentualne nieobiektywne wyniki i skutki w postaci dyskryminacji są szczególnie istotne w odniesieniu do wieku, pochodzenia etnicznego, rasy, płci lub niepełnosprawności. Ponadto bezpośredniość oddziaływania i ograniczone możliwości późniejszej kontroli lub korekty wykorzystania takich systemów działających w czasie rzeczywistym niosą ze sobą zwiększone ryzyko dla praw i wolności osób zainteresowanych w związku z działaniami w zakresie ścigania przestępstw lub na które działania te miały wpływ.

[184] Ponadto do retrospektywnego wykorzystywania systemów zdalnej identyfikacji biometrycznej do celów ścigania przestępstw mają zastosowanie przepisy szczegółowe (art. 26 ust. 10 aktu w sprawie sztucznej inteligencji).

294) Jeżeli jednak wykorzystanie takich systemów jest bezwzględnie konieczne do realizacji istotnego interesu publicznego i jeżeli istnieje wyczerpujący wykaz wąsko zdefiniowanych sytuacji, w których takie wykorzystanie może mieć miejsce, interes związany z takim zastosowaniem przeważa nad ryzykiem dla praw podstawowych (motyw 33 aktu w sprawie sztucznej inteligencji). W celu zapewnienia, aby takie systemy były wykorzystywane „w sposób odpowiedzialny i proporcjonalny", ich wykorzystanie podlega zabezpieczeniom oraz konkretnym obowiązkom i wymogom określonym w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji.

9.2. Główne pojęcia i elementy zakazu

Art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji

Zakazuje się następujących praktyk w zakresie AI:

h) wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, chyba że – i w zakresie, w jakim – takie wykorzystanie jest bezwzględnie konieczne do jednego z następujących celów:

(i) ukierunkowanego poszukiwania konkretnych ofiar uprowadzeń, handlu ludźmi lub wykorzystywania seksualnego ludzi, a także poszukiwania osób zaginionych;

(ii) zapobiegnięcia konkretnemu, istotnemu i bezpośredniemu zagrożeniu życia lub bezpieczeństwa fizycznego osób fizycznych lub rzeczywistemu i aktualnemu lub rzeczywistemu i dającemu się przewidzieć zagrożeniu atakiem terrorystycznym;

(iii) lokalizowania lub identyfikowania osoby podejrzanej o popełnienie przestępstwa w celu prowadzenia postępowania przygotowawczego lub ścigania lub wykonania kar w odniesieniu do przestępstw, o których mowa w załączniku II, podlegających w danym państwie członkowskim karze pozbawienia wolności lub środkowi polegającemu na pozbawieniu wolności przez okres, którego górna granica wynosi co najmniej cztery lata.

Akapit pierwszy lit. h) pozostaje bez uszczerbku dla art. 9 rozporządzenia (UE) 2016/679 w odniesieniu do przetwarzania danych biometrycznych do celów innych niż ściganie przestępstw.

295) Aby zakaz określony w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji miał zastosowanie, musi być spełniony łącznie szereg warunków:

(i) system AI musi być systemem zdalnej identyfikacji biometrycznej,

(ii) praktyka musi polegać na „wykorzystywaniu" tego systemu,

(iii) w „czasie rzeczywistym",

(iv) w przestrzeni publicznej, oraz

(v) do celów ścigania przestępstw.

296) Drugi warunek, tj. „wykorzystywanie” systemu AI, został już przeanalizowany w sekcji 2.3 niniejszych wytycznych. Pozostałe wymienione powyżej warunki opisano i przeanalizowano poniżej.

9.2.1. Pojęcie zdalnej identyfikacji biometrycznej

297) Technologie rozpoznawania biometrycznego umożliwiają wykrywanie i rejestrowanie mierzalnych cech fizycznych (takich jak odległość i wielkość oczu, długość nosa itp.) lub właściwości behawioralnych (takich jak chód lub głos) oraz przekształcanie ich w dane biometryczne nadające się do odczytu maszynowego (zob. sekcja 7.2.1 lit. d) powyżej). Dane te są dostępne w różnych formach: obrazów lub szablonów, które stanowią matematyczne odwzorowanie istotnych cech danej osoby i które wykorzystuje się do rozpoznawania. Technologie rozpoznawania biometrycznego są wykorzystywane na potrzeby weryfikacji i identyfikacji[185].

298) Zgodnie z art. 3 pkt 41 aktu w sprawie sztucznej inteligencji system zdalnej identyfikacji biometrycznej to

> *system AI służący do identyfikacji osób fizycznych bez ich aktywnego udziału, zwykle na odległość, poprzez porównanie danych biometrycznych danej osoby fizycznej z danymi biometrycznymi zawartymi w referencyjnej bazie danych.*

299) Definicja ta obejmuje jedynie funkcję identyfikacyjną systemów rozpoznawania biometrycznego, która wiąże się z brakiem aktywnego udziału zainteresowanych osób (tj. brakiem ich aktywnego uczestnictwa) i rejestrowaniem cech tych osób zazwyczaj na odległość. Do celów identyfikacji zarejestrowane dane biometryczne porównuje się z danymi biometrycznymi już przechowywanymi w referencyjnej bazie danych (takiej jak repozytorium, np. baza danych dotyczących przestępstw, zawierająca wizerunki twarzy lub modele podejrzanych).

a) Wyłącznie do celów identyfikacji

300) Pojęcie „identyfikacja biometryczna” zdefiniowano w art. 3 pkt 35 aktu w sprawie sztucznej inteligencji jako

> *zautomatyzowane rozpoznawanie fizycznych, fizjologicznych, behawioralnych lub psychologicznych cech ludzkich w celu ustalenia tożsamości osoby fizycznej przez porównanie danych biometrycznych tej osoby z danymi biometrycznymi osób fizycznych przechowywanymi w bazie danych.*

301) W motywie 15 aktu w sprawie sztucznej inteligencji doprecyzowano, że takie cechy ludzkie mogą obejmować:

> *twarz, ruch gałek ocznych, kształt ciała, głos, właściwości mowy, chód, postaw[ę], tętno, ciśnienie krwi, zapach, sposób pisania na klawiaturze.*

302) Do definicji identyfikacji biometrycznej można również włączyć systemy AI wykorzystywane do śledzenia osób fizycznych, na przykład w celu sprawdzenia,

[185] Zgodnie z definicją społeczności specjalistów ds. biometrii zawartą w normie ISO/IEC 2382-37:2022 „Technologie informacyjne – Słownictwo”, rozpoznawanie biometryczne, termin 37.01.03.

w jakim kierunku ucieka podejrzany. Można to wywnioskować z art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji, w którym dozwolono lokalizowanie osób podejrzanych o popełnienie przestępstwa. Lokalizowanie jest możliwe, gdy dana osoba jest śledzona.

303) Systemy AI przeznaczone do weryfikacji biometrycznej nie są objęte zakazem określonym w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji[186]. Weryfikacja biometryczna (lub uwierzytelnienie biometryczne) polega na porównaniu danych rejestrowanych przez czujnik z innym zestawem wcześniej zapisanych danych przechowywanych na urządzeniu, takim jak smartfon, paszport lub dowód osobisty. Celem weryfikacji biometrycznej jest potwierdzenie, że dana osoba jest tym, za kogo się podaje.

Przykładem weryfikacji biometrycznej jest porównanie twarzy podróżnego zeskanowanej przy bramce elektronicznej z wizerunkiem twarzy zawartym w jego paszporcie.

b) Zdalna identyfikacja

304) Zgodnie z art. 3 pkt 41 aktu w sprawie sztucznej inteligencji zdalna identyfikacja oznacza zdolność systemów biometrycznych do identyfikacji osób bez ich aktywnego udziału, zwykle na odległość, poprzez porównanie danych biometrycznych danej osoby fizycznej z danymi biometrycznymi zawartymi w referencyjnej bazie danych.

305) Korzystanie z systemów biometrycznych w celu potwierdzenia tożsamości osoby fizycznej wyłącznie na potrzeby uzyskania dostępu do usługi, odblokowania urządzenia lub bezpiecznego dostępu do pomieszczeń nie wchodzi w zakres pojęcia „zdalnej identyfikacji" (motyw 15 aktu w sprawie sztucznej inteligencji). Metoda ta jest stosowana na przykład do kontroli dostępu[187].

Na przykład system identyfikacji twarzy oparty na technologii skanowania twarzy pozwala kontrolować wstęp na obszar o ograniczonym dostępie (np. na teren elektrowni); system ten porównuje zarejestrowaną przez kamerę przy wejściu twarz danej osoby z obrazem referencyjnym zawartym w referencyjnej bazie danych osób upoważnionych do wejścia do budynku.

306) W motyw 17 aktu w sprawie sztucznej inteligencji wyjaśniono, że to wyłączenie z zakresu zakazu jest uzasadnione faktem, iż takie systemy prawdopodobnie będą miały niewielki wpływ na prawa podstawowe osób fizycznych w porównaniu z systemami zdalnej identyfikacji biometrycznej, które mogą być wykorzystywane do przetwarzania danych biometrycznych dużej liczby osób bez ich aktywnego udziału. Dalej doprecyzowano, że systemy zdalnej identyfikacji biometrycznej są zwykle wykorzystywane do jednoczesnego obserwowania wielu osób lub ich zachowania w celu znacznego ułatwienia identyfikacji osób fizycznych bez ich aktywnego udziału.

[186] Motyw 17 aktu w sprawie sztucznej inteligencji.

[187] Np. Ross A, Jain AK (2015) „Biometrics, Overview" [Biometria – zarys ogólny] w Li S.Z. i Jain A.K. (red.) Encyclopedia of Biometrics, (1. Wyd. Springer Science, Nowy Jork), s. 289–294.

Aby udział był aktywny, nie wystarczy, aby osoby były poinformowane o obecności kamer; muszą one aktywnie i świadomie stanąć przed kamerą zainstalowaną w sposób sprzyjający aktywnemu uczestnictwu.

Na przykład:

– systemy zdalnej identyfikacji biometrycznej stosowane w kamerach zainstalowanych na ścianach lub sufitach stacji metra do celów nadzoru. Taki system spełnia warunek zdalnej identyfikacji;

– systemy wykorzystywane do uzyskiwania dostępu do stacji metra, takie jak biometryczne bilety do metra, w przypadku których osoby aktywnie uczestniczą i świadomie podchodzą do czujnika biometrycznego w celu uzyskania dostępu, nie spełniają tego warunku.

307) Systemy identyfikacji biometrycznej, które (bezkontaktowo) przetwarzają odciski palców, chód, głos, DNA, uderzenia w klawisze i inne (biometryczne) sygnały behawioralne, również mogą stanowić systemy zdalnej identyfikacji biometrycznej[188].

Na przykład:

– system oparty na biometrycznej technologii rozpoznawania głosu może być stosowany do identyfikacji osoby mówiącej. W takim przypadku mikrofon pobiera próbkę biometryczną;

– system rozpoznawania chodu może być wykorzystywany za pośrednictwem telewizji przemysłowej, gdzie nagrania wideo są automatycznie sprawdzane pod kątem zgodności z wcześniej zarejestrowanymi wzorcami;

– technologia biometryczna do rozpoznawania dynamiki pisania na klawiaturze może być wykorzystywana do identyfikacji osoby piszącej oszukańczą wiadomość.

Fakt, że systemy te podano jako przykłady systemów zdalnej identyfikacji biometrycznej nie oznacza, że są one zakazane na mocy art. 5 aktu w sprawie sztucznej inteligencji.

308) W przypadku kamer osobistych wyposażonych w systemy zdalnej identyfikacji biometrycznej, używanych przez poszczególnych funkcjonariuszy organów ścigania uznaje się, że nieukierunkowane filmowanie, na przykład podczas demonstracji z udziałem setek uczestników, spełnia warunek zdalnej identyfikacji.

c) Referencyjna baza danych

309) Identyfikacja nie jest możliwa bez referencyjnej bazy danych zawierającej dane biometryczne do celów porównawczych. W związku z tym istnienie referencyjnej bazy danych jest **niezbędne** do przeprowadzenia porównania na potrzeby identyfikacji[189].

[188] Wspólna opinia EROD i EIOD 5/2021, s 11; Rada Unii Europejskiej, „Opinia Służby Prawnej", 12302/22, 12 września 2022 r., pkt 33 oraz motyw 15 aktu w sprawie sztucznej inteligencji.

[189] Motyw 34 aktu w sprawie sztucznej inteligencji.

Na przykład w przypadku osób zaginionych baza danych Systemu Informacyjnego Schengen[190] mogłaby być wykorzystywana jako referencyjna baza danych do celów rozpoznawania twarzy (po uruchomieniu takiej funkcji).

9.2.2. W czasie rzeczywistym

310) Sformułowanie „w czasie rzeczywistym" oznacza, że system zbiera i następnie przetwarza dane biometryczne „natychmiast, niemal natychmiast lub w każdym razie bez znacznego opóźnienia"[191]. Wszystkie etapy przetwarzania, tj. zbieranie, porównywanie i identyfikacja danych biometrycznych, odbywają się jednocześnie lub prawie jednocześnie, co może obejmować „ograniczone krótkie opóźnienia" w celu uniknięcia obchodzenia zakazu poprzez retrospektywne wykorzystywanie systemów zdalnej identyfikacji biometrycznej[192]. Pojęcie „bez znacznego opóźnienia" nie zostało zdefiniowane w akcie w sprawie sztucznej inteligencji; będzie musiało być oceniane dla każdej sprawy osobno. Ze względu na fakt, że urządzenia wykorzystywane do identyfikacji w czasie rzeczywistym i urządzenia wykorzystywane do identyfikacji post factum to coraz częściej te same urządzenia wyposażone w różne funkcje, rozróżnienie to ma charakter czasowy. Ogólnie rzecz biorąc, opóźnienie uznaje się za znaczne co najmniej wtedy, gdy dana osoba prawdopodobnie opuściła miejsce, w którym pobrano dane biometryczne.

311) Systemy działające w czasie rzeczywistym są zasadniczo wykorzystywane w danym miejscu, aby ułatwić szybką reakcję, a nie do retrospektywnej identyfikacji osób. Zapewniają one użytkownikowi systemu środki umożliwiające śledzenie przemieszczania się osób objętych nadzorem i ich monitorowanie.

a) System AI rozpoznaje wszystkie osoby wchodzące na teren koncertu: jest to system zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

b) System filmuje wszystkie osoby, które przyszły na koncert. Podczas koncertu dochodzi do incydentu. Po koncercie system identyfikacji wykorzystuje materiał wideo do zidentyfikowania sprawcy: jest to system zdalnej identyfikacji biometrycznej post factum.

312) Jeżeli funkcjonariusz organu ścigania potajemnie robi danej osobie zdjęcie za pomocą urządzenia mobilnego i przesyła to zdjęcie do bazy danych w celu natychmiastowego wyszukania tej osoby, może to podlegać – w zależności od okoliczności – zakazowi określonemu w art. 5 ust. 1 lit. h) aktu o sztucznej inteligencji.

9.2.3. W przestrzeni publicznej

313) W art. 3 pkt 44 aktu w sprawie sztucznej inteligencji zdefiniowano **przestrzeń publiczną** jako

[190] Wpisy dotyczące osób zaginionych (art. 32 decyzji w sprawie SIS II); Decyzja Rady 2007/533/WSiSW z dnia 12 czerwca 2007 r. w sprawie utworzenia, funkcjonowania i użytkowania Systemu Informacyjnego Schengen drugiej generacji (SIS II)

[191] Motyw 17 aktu w sprawie sztucznej inteligencji.

[192] Art. 3 pkt 42 aktu w sprawie sztucznej inteligencji.

miejsce fizyczne, będące własnością prywatną lub publiczną, dostępne dla nieokreślonej liczby osób fizycznych, niezależnie od tego, czy mogą mieć zastosowanie określone warunki dostępu, oraz niezależnie od potencjalnych ograniczeń pojemności.

314) W motywie 19 aktu w sprawie sztucznej inteligencji wymieniono kilka elementów charakteryzujących taką przestrzeń:

- dostępność dla nieokreślonej liczby osób, niezależnie od potencjalnych ograniczeń w zakresie pojemności lub bezpieczeństwa, takich jak zakup biletu wstępu lub biletu na przejazd, uprzednia rejestracja lub osiągnięcie określonego wieku. Możliwość uzyskania dostępu do przestrzeni przez niezamknięte drzwi nie oznacza, że dana przestrzeń stanowi przestrzeń publiczną, jeżeli istnieją wskazania lub okoliczności sugerujące inaczej (np. znak ograniczający dostęp). Ponadto dostęp do przestrzeni może być ograniczony do określonych osób, na mocy prawa związanego z bezpieczeństwem publicznym lub ochroną publiczną lub decyzją osoby posiadającej odpowiednie uprawnienia związane z taką przestrzenią;

Na przykład przestrzenią publiczną są co do zasady:

- miejsce koncertu, który wymaga od uczestników uiszczenia opłaty za wstęp;
- miejsce, w którym organizowane są targi dla uczestników w wieku powyżej 50 lat.

Przestrzeń, do której prowadzi brama, nawet jeśli nie jest ona zamknięta na klucz, jak w przypadku wejścia do ogrodzonego osiedla mieszkaniowego obejmującego kilka domów, zwykle nie będzie uważana za przestrzeń publiczną. Natomiast park w obrębie ogrodzonego osiedla, który w pewnych godzinach jest publicznie dostępny bez żadnych ograniczeń dostępu, będzie zasadniczo stanowił przestrzeń publiczną w tych godzinach i przestrzeń zamkniętą poza tymi godzinami.

- brak znaczenia, kto jest właścicielem danej przestrzeni, tj. przestrzeń nie musi być własnością publiczną, aby można ją było uznać za przestrzeń publiczną;

Na przykład przestrzeń może być własnością podmiotu prywatnego, podmiotu publicznego lub podmiotu publicznego i być zarządzana przez podmiot prywatny, bez wpływu na jej charakter.

- brak konkretnej działalności, do której wykorzystywana jest przestrzeń; przestrzeń publiczna niekoniecznie jest przestrzenią związaną z usługą publiczną. Ponadto przestrzeń związana z usługą publiczną może obejmować przestrzenie niepubliczne, np. biura urzędników służby cywilnej pracujących w gminie.

Na przykład przestrzenie publiczne można wykorzystywać do działalności handlowej, jak w przypadku sklepów, restauracji, kawiarni itp.; do działalności usługowej, jak w przypadku banków, działalności zawodowej (gabinet lekarski oraz

biuro księgowego), hotelarstwa (np. hotel) itp.; do działalności sportowej – jak w przypadku basenów, sal do ćwiczeń, stadionów itp.; do działalności transportowej – jak w przypadku dworców autobusowych i kolejowych, stacji metra, lotniska, środków transportu itp.; do działalności rozrywkowej – jak w przypadku kin, teatrów, muzeów, sal koncertowych i konferencyjnych itp.; lub przestrzenie te mogą służyć wypoczynkowi lub innym celom, jak w przypadku dróg publicznych i placów, parków, lasów i placów zabaw[193].

315) Następujące przestrzenie nie stanowią przestrzeni publicznej w rozumieniu art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji:

- przestrzenie internetowe, ponieważ nie stanowią przestrzeni fizycznej w rozumieniu art. 3 pkt 44 aktu w sprawie sztucznej inteligencji;

Na przykład czaty, media społecznościowe, platformy internetowe itp. nie wchodzą w związku z tym w zakres zakazu.

- niektóre przestrzenie, do których dostęp ma mieć ograniczona liczba osób, takie jak fabryki, przedsiębiorstwa i miejsca pracy, do których wstęp jest kontrolowany lub ograniczony do określonych pracowników lub usługodawców, ponieważ dostęp do nich mają mieć wyłącznie określeni pracownicy i usługodawcy[194];

Na przykład miejsce pracy, do którego dostęp można uzyskać za pomocą identyfikatora, nie jest co do zasady uznawane za przestrzeń publiczną, podczas gdy biuro bez kontroli wstępu może być uznane za taką przestrzeń.

- **więzienia i strefy kontroli granicznej** nie zaliczają się do przestrzeni publicznej[195].

316) Na przykład przejście graniczne nie jest przestrzenią publiczną, podczas gdy ulica prowadząca do przejścia granicznego lub las znajdujący się w pobliżu zazwyczaj tak.

317) Niektóre przestrzenie mogą pełnić podwójną funkcję. Na przykład port lotniczy jest ogólnie uznawany za przestrzeń publiczną w odniesieniu do obszarów wspólnych, ale obszar przeznaczony do kontroli granicznej (gdzie pracują funkcjonariusze celni i odbywają się kontrole paszportowe lub dowodów tożsamości) nie wchodzi w zakres zakazu.

318) Jak wyjaśniono w motywie 19 aktu w sprawie sztucznej inteligencji, to, czy dana przestrzeń jest dostępna publicznie, należy oceniać indywidualnie dla każdego przypadku.

9.2.4. Na potrzeby ścigania przestępstw

193 Motyw 19 aktu w sprawie sztucznej inteligencji.
194 Motyw 19 aktu w sprawie sztucznej inteligencji.
195 Motyw 19 aktu w sprawie sztucznej inteligencji. W innym kontekście kontrolę graniczną zdefiniowano jako działalność prowadzoną na granicy, zgodnie z rozporządzeniem (UE) 2016/399 (kodeks graniczny Schengen), i dla jego celów, wyłącznie w odpowiedzi na zamiar przekroczenia granicy lub sam akt jej przekroczenia. Nie obejmuje to tzw. strefy przygranicznej, która może sięgać maksymalnie 50 km po obu stronach granicy.

319) Zakaz określony w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji ma zastosowanie do wykorzystywania systemów zdalnej identyfikacji biometrycznej do celów ścigania przestępstw, niezależnie od podmiotu lub organu prowadzącego działania w zakresie ścigania przestępstw.

320) Zgodnie z art. 3 pkt 46 aktu w sprawie sztucznej inteligencji ściganie przestępstw oznacza „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". Cele te są takie same jak cele wymienione w art. 1 dyrektywy o ochronie danych w sprawach karnych[196]. W związku z tym każda wykładnia tych celów w odniesieniu do dyrektywy o ochronie danych w sprawach karnych może mieć również znaczenie dla wykładni pojęcia „ścigania przestępstw" użytego w akcie w sprawie sztucznej inteligencji.

321) Cele związane ze ściganiem przestępstw obejmują prowadzenie postępowań przygotowawczych oraz wykrywanie i ściganie czynów zabronionych. Obejmują one również działania związane z zapobieganiem przestępstwom, w tym ochronę przed zagrożeniami dla bezpieczeństwa publicznego i zapobieganie takim zagrożeniom, zanim dojdzie do popełnienia przestępstwa. Na przykład w kontekście zapobiegania przestępczości policja może zastosować środki przymusu podczas demonstracji, dużych imprez sportowych lub zamieszek[197]. Wreszcie działania te obejmują wykonywanie kar, w tym wykonywanie zasądzonych wyroków.

322) Zgodnie z art. 3 pkt 46 aktu w sprawie sztucznej inteligencji działania w zakresie ścigania przestępstw mogą być prowadzone przez organy ścigania lub w ich imieniu. Organy ścigania zdefiniowano bardziej szczegółowo w art. 3 pkt 45 aktu w sprawie sztucznej inteligencji w taki sam sposób jak właściwe organy krajowe zdefiniowane w dyrektywie o ochronie danych w sprawach karnych[198]. Definicja ta obejmuje organy ścigania oraz organy lub podmioty, którym powierzono wykonywanie zadań (które mogą być podmiotami prywatnymi):

   a) *każdy organ publiczny właściwy w zakresie zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania lub ścigania czynów zabronionych lub egzekwowania sankcji karnych, w tym ochrony przed zagrożeniami dla bezpieczeństwa publicznego i zapobiegania takim zagrożeniom; lub*

Na przykład takie organy publiczne obejmują policję i organy wymiaru sprawiedliwości w sprawach karnych (takie jak prokuratura), gdy wykonują one zadania związane ze ściganiem przestępstw.

[196] Niektóre działania organów ścigania są wyłączone z zakresu dyrektywy o ochronie danych w sprawach karnych, np. zadania administracyjne (związane z zasobami ludzkimi); w takim przypadku działania te są prowadzone poza ramami ścigania przestępstw. Wchodzą one w zakres RODO. Zob. motyw 19 RODO.

[197] Motyw 12 dyrektywy o ochronie danych w sprawach karnych.

[198] Art. 3 pkt 7 dyrektywy o ochronie danych w sprawach karnych.

b) inny organ lub podmiot, któremu na podstawie prawa państwa członkowskiego powierzono sprawowanie władzy publicznej i wykonywanie uprawnień publicznych do celów zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania i ścigania czynów zabronionych lub wykonywania kar, w tym ochrony przed zagrożeniami dla bezpieczeństwa publicznego i zapobiegania takim zagrożeniom;

323) Zgodnie z aktem w sprawie sztucznej inteligencji inne podmioty, organy lub osoby mogą wykonywać czynności związane ze ściganiem przestępstw po powierzeniu im przez państwa członkowskie władzy publicznej i uprawnień publicznych w celach wymienionych powyżej.

324) „W imieniu" oznacza, że organ ścigania powierzył wykonywanie działań w zakresie ścigania przestępstw (lub części takich działań) innemu podmiotowi lub osobie, w tym podmiotom prywatnym, lub zwrócił się w określonych przypadkach do innego podmiotu lub osoby o wsparcie działań w zakresie ścigania przestępstw. W obu przypadkach organy ścigania muszą wydawać temu innemu podmiotowi instrukcje dotyczące wszystkich istotnych aspektów działań oraz nadzorować ten podmiot, ponieważ wymóg ten jest nieodłącznym elementem pojęcia działania „w imieniu" danej osoby.

Powierzenie zadań innym organom może obejmować na przykład:

- zwrócenie się przez organy ścigania do przedsiębiorstw transportu publicznego o zapewnienie bezpieczeństwa w sieciach transportu publicznego zgodnie z instrukcjami i pod nadzorem organów ścigania;
- zwrócenie się przez organy ścigania do federacji sportowych o podjęcie działań zgodnie z instrukcjami i pod nadzorem organów ścigania w celu zapewnienia bezpieczeństwa podczas imprez sportowych;
- zwrócenie się przez organy ścigania do banków o przeprowadzenie określonych działań w celu „przeciwdziałania niektórym przestępstwom w konkretnych przypadkach" zgodnie z instrukcjami i pod nadzorem organów ścigania.

Działania te wchodzą w zakres definicji „do celów ścigania przestępstw", ponieważ powyższe podmioty działają „w imieniu" organów ścigania. Jeżeli podczas wykrywania przestępstw i przeciwdziałania przestępstwom (takim jak nadużycia finansowe, pranie pieniędzy) podmioty te działają we „własnym imieniu", nie zostaną one uznane za objęte zakazem określonym w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji.

325) Tylko wtedy, gdy tym innym organom lub podmiotom powierzono konkretne zadanie związane ze ściganiem przestępstw, ich działania będą objęte definicją „ścigania przestępstw".

9.3. Wyjątki od zakazu

326) W akcie w sprawie sztucznej inteligencji przewidziano trzy wyjątki od ogólnego zakazu wykorzystywania systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw. W art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji wymieniono wyczerpująco trzy cele, do których system zdalnej identyfikacji biometrycznej w czasie rzeczywistym może zostać dozwolony, natomiast w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji określono warunki i zabezpieczenia dotyczące takiego dopuszczenia. Art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji sam w sobie nie stanowi podstawy prawnej dla korzystania z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej. Jedynie prawo krajowe państwa członkowskiego, które spełnia w szczególności wymogi określone w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji, może zezwalać wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, jak przewidziano w art. 5 ust. 2 aktu w sprawie sztucznej inteligencji. W związku z tym, wobec braku przepisów państwa członkowskiego zezwalających na wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w odniesieniu do co najmniej jednego z tych celów, takie wykorzystanie jest zakazane od dnia 2 lutego 2025 r.

9.3.1. Uzasadnienie i cele

327) Cele określone w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji mają umożliwić wykorzystywanie niektórych narzędzi sztucznej inteligencji i narzędzi dochodzeniowych do celów ścigania przestępstw. Cele te dotyczą:

> (i) ukierunkowanego poszukiwania ofiar trzech konkretnych poważnych przestępstw i osób zaginionych [ochrona];
>
> (ii) zapobiegania bezpośredniemu zagrożeniu życia lub bezpieczeństwa fizycznego lub rzeczywistemu zagrożeniu atakami terrorystycznymi [zapobieganie]; oraz
>
> (iii) lokalizowania lub identyfikowania podejrzanych i sprawców niektórych poważnych przestępstw wymienionych w załączniku II [ściganie/postępowanie przygotowawcze].

328) W powyższych scenariuszach prawodawca unijny zrównoważył potrzeby społeczeństwa w zakresie bezpieczeństwa z ryzykiem, jakie systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym stwarzają dla praw podstawowych osób podlegających tym systemom. Zgodnie z motywem 33 aktu w sprawie sztucznej inteligencji cele, do których dozwolone jest wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym na potrzeby ścigania przestępstw w przestrzeni publicznej, muszą być ściśle, wyczerpująco i wąsko zdefiniowane, przy czym wykorzystywanie to musi być „bezwzględnie konieczne do realizacji istotnego interesu publicznego, którego waga przeważa nad ryzykiem” dla praw podstawowych. Wszelkie inne wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, które

nie jest wymienione w art. 5 ust. 1 lit. h) pkt (i)–(iii) aktu w sprawie sztucznej inteligencji, jest zakazane.

Na przykład zakazane jest wykorzystywanie przez policję systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do identyfikacji złodziei sklepowych i porównywania ich wizerunków twarzy z bazami danych dotyczących przestępstw, ponieważ nie wchodzi to w zakres żadnego z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji.

9.3.2. Ukierunkowane poszukiwanie ofiar trzech poważnych przestępstw i osób zaginionych

329) Zgodnie z art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji wykorzystywanie zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw jest dozwolone, z zastrzeżeniem bezwzględnej konieczności i warunków określonych w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji, do ukierunkowanych poszukiwań ofiar uprowadzeń, handlu ludźmi lub wykorzystywania seksualnego ludzi, a także poszukiwania osób zaginionych.

a) Ukierunkowane poszukiwanie ofiar trzech rodzajów przestępstw

330) Scenariusz opisany w art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji ma pomóc organom ścigania w poszukiwaniu ofiar trzech poważnych przestępstw.

331) Ukierunkowane poszukiwanie obejmowałoby lokalizowanie i identyfikowanie ofiar.

Trzy rodzaje przestępstw

332) Scenariusz wymieniony w art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji obejmuje ukierunkowane poszukiwanie konkretnych ofiar trzech poważnych przestępstw:: uprowadzeń, handlu ludźmi i wykorzystywania seksualnego ludzi[199].

Jeżeli na przykład zostanie porwane dziecko i istnieją konkretne przesłanki wskazujące, że porywacz zamierza przewieźć dziecko samochodem z jednego miejsca w drugie, policja może użyć systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do ukierunkowanego poszukiwania tego dziecka, musi jednak określić obszar zastosowania oraz czas trwania użycia systemu w celu zidentyfikowania tego dziecka.

[199] Uprowadzenia, handel ludźmi i wykorzystywanie seksualne to trzy przestępstwa, które mogą doprowadzić do wydania europejskiego nakazu aresztowania (ENA) w celu aresztowania i przekazania osoby podejrzanej lub skazanej do państwa, które wydało ENA. Te trzy przestępstwa dotyczą głównie, ale nie wyłącznie, kobiet i dzieci. Według DG ds. Migracji i Spraw Wewnętrznych Komisji Europejskiej prawie 40 % ofiar to obywatele UE, z których większość to kobiety i dzieci będące ofiarami handlu ludźmi w celu wykorzystywania seksualnego. Liczba mężczyzn będących ofiarami wzrosła niemal dwukrotnie w ciągu ostatnich dziesięciu lat. Są oni ofiarami handlu ludźmi w celu pracy przymusowej i przymusowego żebractwa, podczas gdy większość kobiet i dzieci jest ofiarami handlu w celu wykorzystywania seksualnego. https://home-affairs.ec.europa.eu/policies/internal-security/organised-crime-and-human-trafficking/together-against-trafficking-human-beings_pl

b) Poszukiwanie osób zaginionych

333) Pierwszy scenariusz obejmuje również poszukiwanie osoby zaginionej[200].

334) Można wprowadzić rozróżnienie między zaginionymi dziećmi a zaginionymi dorosłymi, gdyż dobrowolne zniknięcie osoby dorosłej niekoniecznie powoduje wszczęcie procedury poszukiwania. Obowiązujące przepisy dotyczące zaginionych dzieci różnią się znacznie między poszczególnymi państwami członkowskimi[201]. W każdym razie w art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji zezwolono wyłącznie na korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym na potrzeby poszukiwania osób zaginionych do celów ścigania przestępstw.

335) Zaginięcie osoby dorosłej nie zawsze prowadzi do jej poszukiwania przez policję, ponieważ osoby dorosłe mają prawo do zniknięcia. Poszukanie może być powiązane ze statusem prawnym danej osoby („pod kuratelą"), jej stanem zdrowia (choroba psychiczna), zostawieniem listu pożegnalnego, ale także wyjazdem bez rzeczy osobistych. Jeśli okoliczności zaginięcia budzą niepokój, zaginięcie może zostać zgłoszone na policję, aby można było rozpocząć poszukiwania.

336) W niektórych państwach członkowskich poszukiwanie osoby zaginionej może odbywać się w ramach procedury administracyjnej, a nie do celów ścigania przestępstw. Na przykład, gdy zaginie osoba wymagająca szczególnego traktowania, ale nie ma podejrzenia popełnienia przestępstwa ani żadnego innego celu związanego ze ściganiem przestępstw, użycie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do poszukiwania tej osoby nie byłoby uznane za działanie do celów ścigania przestępstw, a zatem podlegałoby przepisom dotyczącym takiego użycia zgodnie z RODO.

9.3.3. Zapobieganie bezpośredniemu zagrożeniu życia lub atakom terrorystycznym

337) W art. 5 ust. 1 lit. h) ppkt (ii) aktu w sprawie sztucznej inteligencji wymieniono drugi scenariusz, w którym dozwolone jest wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, z zastrzeżeniem bezwzględnej konieczności i warunków określonych w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji:

> *[w celu] zapobiegnięcia konkretnemu, istotnemu i bezpośredniemu zagrożeniu życia lub bezpieczeństwa fizycznego osób fizycznych lub rzeczywistemu i aktualnemu lub rzeczywistemu i dającemu się przewidzieć zagrożeniu atakiem terrorystycznym.*

---

200 Pojęcie „osoby zaginionej" nie jest zdefiniowane na szczeblu UE. Jednak w konkluzjach Rady z grudnia 2021 roku w sprawie intensyfikacji transgranicznej współpracy policyjnej w dziedzinie osób zaginionych Rada przyjęła jako odniesienie definicję osoby zaginionej zawartą zarówno w zaleceniu Rady Europy CM/Rec (2009) 12, jak i w przepisach krajowych. Konkluzje Rady (2021) 14808/21, pkt 11, s. 4.

201 Komisja Europejska, Europejska Sieć Migracyjna, „How do EU Member States treat cases of missing unaccompanied minors?" [Jak państwa członkowskie UE traktują przypadki zaginionych małoletnich bez opieki?], EMN Inform, 2020.

a) Konkretne, istotne i bezpośrednie zagrożenie życia lub bezpieczeństwa fizycznego osób fizycznych

338) Zgodnie z art. 2 Karty, który gwarantuje prawo do życia, Unia i jej państwa członkowskie muszą chronić i tym samym zabezpieczać życie jednostek. Kryteria określone w art. 5 ust. 1 lit. h) ppkt (ii) aktu w sprawie sztucznej inteligencji dotyczące możliwości wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej w sytuacji zagrożenia życia wymagają istnienia 1) konkretnego, 2) istotnego i 3) bezpośredniego zagrożenia życia lub bezpieczeństwa fizycznego 4) osób fizycznych. Zagrożenie to nie musi być ograniczone do zidentyfikowanych osób lub grupy, ponieważ dotyczy ogólnie osób fizycznych.

339) W motywie 33 aktu w sprawie sztucznej inteligencji wyjaśniono, że bezpośrednie zagrożenie dla życia lub bezpieczeństwa fizycznego osób fizycznych może również obejmować bezpośrednie zagrożenie dla infrastruktury krytycznej[202], „w przypadku gdy zakłócenie lub zniszczenie takiej infrastruktury krytycznej spowodowałoby bezpośrednie zagrożenie życia lub bezpieczeństwa fizycznego osoby, w tym poprzez poważną szkodę w dostarczaniu podstawowych dostaw dla ludności lub w wykonywaniu podstawowych funkcji państwa".

> Na przykład[203]:
>
> poważne zakłócenie funkcjonowania i zniszczenie infrastruktury krytycznej (np. elektrowni, infrastruktury zaopatrzenia w wodę czy szpitala) może spowodować bezpośrednie zagrożenie życia lub bezpieczeństwa fizycznego osoby w przypadku poważnej szkody, jaką jest wstrzymanie podstawowych dostaw dla ludności (odcięcie dostępu do energii elektrycznej lub wody pitnej na dłuższy okres, w szczególnie gorących lub zimnych warunkach pogodowych itp.).

340) To, co stanowi bezpośrednie zagrożenie dla życia lub bezpieczeństwa fizycznego osób fizycznych, należy ostatecznie zdefiniować i oceniać na szczeblu państwa członkowskiego na podstawie jego przepisów krajowych, zgodnie z prawem UE, w szczególności z uwzględnieniem kluczowych elementów i przesłanek określonych w art. 5 aktu w sprawie sztucznej inteligencji. Pojęcie to należy określić i odnosić się do niego w przepisach, które państwa członkowskie muszą przyjąć, aby korzystać z wyjątków od zakazu wykorzystywania systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw w przestrzeni publicznej.

341) **Bezpośrednie** zagrożenie dla życia lub bezpieczeństwa fizycznego jest zagrożeniem, które może wystąpić w dowolnym momencie i wymaga „**natychmiastowego działania**"[204]. **Istotne zagrożenie** dla bezpieczeństwa fizycznego dotyczy poważnych obrażeń ciała.

---

202 Zgodnie z definicją zawartą w art. 2 pkt 4 dyrektywy 2022/2557/WE.
203 Motyw 33 aktu w sprawie sztucznej inteligencji.
204 Motyw 37 rozporządzenia (UE) 2023/1543.

342) Konkretne zagrożenie oznacza, że zagrożenie jest jasno określone, zindywidualizowane i realne, co oznacza, że nie powinno być hipotetyczne ani odnosić się do ogólnych niebezpieczeństw.

> Na przykład policja otrzymuje informację, że były student planuje śmiertelny atak na swoją dawną uczelnię, szukając zemsty na kilku byłych kolegach z klasy. Policja otrzymuje informacje o zbliżającym się ataku, uczelni będącej celem ataku oraz broni, której student zamierza użyć do realizacji swoich planów.

343) Konkretne zagrożenie nie musi być zamierzone. Działania nieumyślne mogą również prowadzić do zagrożenia życia lub bezpieczeństwa fizycznego.

b) Rzeczywiste i aktualne lub rzeczywiste i dające się przewidzieć zagrożenie atakiem terrorystycznym

344) Ta część drugiego scenariusza opisanego w art. 5 ust. 1 lit. h) ppkt (ii) aktu w sprawie sztucznej inteligencji składa się z kilku elementów: istnienie zagrożenia atakiem terrorystycznym i charakter tego zagrożenia, które musi być rzeczywiste i aktualne lub rzeczywiste i dające się przewidzieć.

Zagrożenie atakiem terrorystycznym

345) Oceny istnienia i powagi zagrożenia dokonuje się na poziomie krajowym przy ocenie rzeczywistych okoliczności środka, który należy zastosować w celu ochrony bezpieczeństwa narodowego, a dokładniej w przypadku ataku terrorystycznego. Poziom **zagrożenia terrorystycznego jest określany na szczeblu krajowym** i różni się między poszczególnymi państwami członkowskimi. Na przykład Niderlandy ustaliły pięć poziomów zagrożenia[205], Belgia – cztery[206], Francja – trzy[207], a Szwecja – pięć[208]. Pojęcie „rzeczywistego i aktualnego lub rzeczywistego i dającego się przewidzieć zagrożenia" użyte w art. 5 ust. 1 lit. h) ppkt (ii) jest jednak autonomicznym pojęciem prawa Unii i w związku z tym co do zasady powinno być oceniane niezależnie od definicji krajowych. Zagrożenie to nie odnosi się do terroryzmu ogólnie, lecz konkretnie do zagrożenia atakiem terrorystycznym.

Charakter zagrożenia: rzeczywiste i aktualne lub rzeczywiste i dające się przewidzieć

346) Próg powagi, który musi zostać osiągnięty, aby systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym można było wykorzystywać w przestrzeni publicznej do celów ścigania przestępstw, został zainspirowany orzecznictwem TSUE dotyczącym zatrzymywania danych i środków związanych z danymi dotyczącymi przelotu pasażera, mających na celu ochronę bezpieczeństwa narodowego, w szczególności przed atakami terrorystycznymi. Zdaniem TSUE w tych kontekstach „zagrożenie dla bezpieczeństwa narodowego musi być rzeczywiste i aktualne, lub

---

205 https://www.government.nl/topics/counterterrorism-and-national-security/risk-of-an-attack-threat-level.

206 https://cuta.belgium.be; https://crisiscenter.be/en/risks-belgium/security-risks/terrorism-and-extremism.

207 https://www.sgdsn.gouv.fr/vigipirate#; https://www.sgdsn.gouv.fr/files/files/Vigipirate/20160130-np-sgdsn-pse-tackling-terrorism-together.pdf.

208 https://www.krisinformation.se/en/hazards-and-risks/terrorism.

przynajmniej przewidywalne, co zakłada wystąpienie wystarczająco konkretnych okoliczności”[209].

Zapobieganie

347) W przeciwieństwie do art. 5 ust. 1 lit. h) ppkt (i) oraz art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji w scenariuszu opisanym w art. 5 ust. 1 lit. h) ppkt (ii) nie doprecyzowano, że korzystanie z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym jest dozwolone na potrzeby lokalizowania lub identyfikowania konkretnej osoby. Jego celem jest zapobieżenie konkretnemu zagrożeniu. W związku z tym scenariusz ten może również obejmować wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do wykrywania i śledzenia „mobilności terrorystów”, tj. kilku osób związanych z tym samym zagrożeniem, jeżeli istnieją konkretne przesłanki, że osoby te planują dokonać ataku terrorystycznego, ale nie jest jasne, gdzie.

Wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do zapobieżenia atakowi terrorystycznemu w parku

Policja otrzymuje informację, że po parku biega osoba z nożem, wykrzykująca agresywne, ekstremistyczne hasła, które są zazwyczaj powiązane z atakami terrorystycznymi i grupami terrorystycznymi. Jeżeli państwo członkowskie zezwoliło na korzystanie z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w scenariuszu opisanym w art. 5 ust. 1 lit. h) ppkt (ii) aktu w sprawie sztucznej inteligencji, organy ścigania mogą wykorzystać system zdalnej identyfikacji biometrycznej w czasie rzeczywistym na potrzeby zidentyfikowania i zlokalizowania osoby w parku i w jego okolicach, aby zapobiec atakowi, o ile spełnione są pozostałe warunki określone w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji.

9.3.4. Lokalizowanie i identyfikowanie osób podejrzanych o popełnienie określonych przestępstw

348) W art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji zezwolono na wykorzystanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów „lokalizowania lub identyfikowania osoby podejrzanej o popełnienie przestępstwa w celu prowadzenia postępowania przygotowawczego lub ścigania lub wykonania kar w odniesieniu do przestępstw, o których mowa w załączniku II, podlegających w danym państwie członkowskim karze pozbawienia wolności lub środkowi polegającemu na pozbawieniu wolności przez okres, którego górna granica wynosi co najmniej cztery lata”.

Załącznik II do aktu w sprawie sztucznej inteligencji zawiera wyczerpujący wykaz poważnych przestępstw, w przypadku których do wyżej wymienionego celu można

[209] Wyrok Trybunału Sprawiedliwości z dnia 20 września 2022 r. *SpaceNet*, C-793/19 (sprawy połączone C-793/19, C-794/19), ECLI:EU:C:2022:702, pkt 93.

zezwolić na korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym. Przestępstwa te to:

– terroryzm,

– handel ludźmi,

– wykorzystywanie seksualne dzieci i pornografia dziecięca,

– nielegalny obrót środkami odurzającymi lub substancjami psychotropowymi,

– nielegalny handel bronią, amunicją lub materiałami wybuchowymi,

– zabójstwo, ciężkie uszkodzenie ciała,

– nielegalny obrót organami lub tkankami ludzkimi,

– nielegalny handel materiałami jądrowymi lub promieniotwórczymi,

– uprowadzenie, bezprawne przetrzymywanie lub wzięcie zakładników,

– przestępstwa podlegające jurysdykcji Międzynarodowego Trybunału Karnego,

– bezprawne zawładnięcie statkiem powietrznym lub statkiem,

– zgwałcenie,

– przestępstwo przeciw środowisku,

– rozbój w formie zorganizowanej lub rozbój przy użyciu broni,

– sabotaż,

– udział w organizacji przestępczej uczestniczącej w co najmniej jednym z wyżej wymienionych przestępstw.

a) Lokalizowanie i identyfikowanie

349) Państwo członkowskie może zezwolić na wykorzystywanie zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, aby zlokalizować i zidentyfikować osobę podejrzaną o popełnienie przestępstwa na potrzeby przeprowadzenia postępowania przygotowawczego, ścigania tej osoby za popełnione przestępstwo lub wykonania obowiązującej kary.

b) Podejrzani i sprawcy

350) Art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji obejmuje dwie kategorie osób: podejrzani i sprawcy. Podejrzany to osoba, co do której istnieją poważne podstawy, by sądzić, że popełniła przestępstwo, i zebrano już wystarczające dowody udziału tej osoby w przestępstwie. Sprawcą jest osoba oskarżona o popełnienie przestępstwa lub skazana za jego popełnienie. Te same warunki (przestępstwo wymienione w załączniku II i maksymalny wymiar kary wynoszący co najmniej cztery lata) mają zastosowanie do lokalizowania lub identyfikowania współsprawcy przestępstwa wymienionego w załączniku II do aktu w sprawie sztucznej inteligencji.

c) Wykaz poważnych przestępstw

351) Jedynie poważne przestępstwa uzasadniają wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw.

352) Pierwsze pięć przestępstw wymienionych w załączniku II do aktu w sprawie sztucznej inteligencji odpowiada „europrzestępstwom", wymienionym w art. 83 TFUE, a pozostałe przestępstwa stanowią priorytety współpracy organów ścigania[210]. Niektóre z nich (np. uprowadzenie, nielegalny handel materiałami jądrowymi lub promieniotwórczymi) mogą być związane z terroryzmem[211].

353) Chociaż wszystkie przestępstwa wymienione w załączniku II mogą spowodować wydanie europejskiego nakazu aresztowania („ENA") przeciwko podejrzanemu lub sprawcy, korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w celu zlokalizowania i identyfikacji podejrzanego w odniesieniu do jednego z tych poważnych przestępstw nie wymaga wydania ENA.

354) Ponadto, aby w tym celu wykorzystać system zdalnej identyfikacji biometrycznej w czasie rzeczywistym, dane przestępstwo musi podlegać w danym państwie członkowskim karze pozbawienia wolności lub środkowi polegającemu na pozbawieniu wolności przez okres, którego górna granica wynosi co najmniej cztery lata.

> Podczas festiwalu w mieście z udziałem dużej liczby osób policja stosuje technologie rozpoznawania twarzy w czasie rzeczywistym w celu monitorowania obszaru wokół festiwalu i identyfikacji osób poszukiwanych na podstawie nakazów aresztowania za nielegalny handel narkotykami oraz przestępstwa na tle seksualnym. Przy różnych wejściach na festiwal policja wykorzystuje nagrania wideo na żywo osób przechodzących przed mobilną kamerą, aby porównać ich twarze z listą poszukiwanych osób.
>
> Po pierwsze, jeśli chodzi o rodzaje przestępstw, system zdalnej identyfikacji biometrycznej może być stosowany w przypadku nielegalnego handlu narkotykami. Przestępstwa na tle seksualnym nie znajdują się jednak w wykazie przestępstw, chyba że dotyczą wykorzystywania seksualnego dzieci, materiałów przedstawiających niegodziwe traktowanie dzieci w celach seksualnych lub zgwałcenia. Policja nie może stosować technologii rozpoznawania twarzy w czasie rzeczywistym w sposób szeroki i nieukierunkowany, tj. w nadziei na znalezienie poszukiwanych przestępców i ich aresztowanie na ulicy.
>
> Inaczej jest w przypadku, gdy policja otrzymała opis fizyczny wraz ze zdjęciem osoby poszukiwanej na podstawie europejskiego nakazu aresztowania za handel narkotykami i ma powody, by sądzić, że będzie ona obecna na festiwalu. W tych okolicznościach stosowanie technologii rozpoznawania twarzy w czasie

[210] Priorytety Europolu.
[211] Zobacz motyw 33 i definicja przestępstw terrorystycznych w art. 3 dyrektywy 2017/541.

rzeczywistym w celu identyfikacji docelowej osoby fizycznej może być objęte art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji.

Po poważnym zamachu terrorystycznym na jarmarku bożonarodzeniowym, w którym zginęło 12 osób, policja wykorzystuje technologie rozpoznawania twarzy w czasie rzeczywistym, aby zidentyfikować sprawcę i zobaczyć, dokąd ucieka. W tym kontekście policja wykorzystuje również technologie rozpoznawania twarzy w czasie rzeczywistym na pobliskiej stacji kolejowej oraz na stacjach docelowych pociągów, które odjeżdżają stamtąd krótko po ataku. W przypadku ataku terrorystycznego takie wykorzystanie może być dozwolone na podstawie art. 5 ust. 1 lit. h) pkt (iii) aktu w sprawie sztucznej inteligencji.

355) Związek między art. 5 ust. 1 lit. h) ppkt (i) a art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji można ustanowić w odniesieniu do przestępstw objętych scenariuszem opisanym w art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji. Chociaż systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym mogą być stosowane do celów znalezienia ofiary lub osoby zaginionej, systemy te mogą być również wykorzystywane do lokalizowania i identyfikacji sprawców handlu ludźmi, wykorzystywania seksualnego w odniesieniu do dzieci (zgodnie z załącznikiem II) oraz uprowadzenia (o ile uprowadzenie, o którym mowa w art. 5 ust. 1 lit. h) ppkt (i) aktu w sprawie sztucznej inteligencji, kwalifikuje się jako uprowadzenie zgodnie z wykazem w załączniku II do aktu w sprawie sztucznej inteligencji), a także podejrzanych o popełnienie takich przestępstw. Można również powiązać art. 5 ust. 1 lit. h) ppkt (ii) i (iii) aktu w sprawie sztucznej inteligencji: systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym mogą być wykorzystywane do zapobiegania zagrożeniu objętemu zakresem art. 5 ust. 1 lit. h) ppkt (ii), a w przypadku urzeczywistnienia się tego zagrożenia systemy te mogą być wykorzystywane do identyfikowania i lokalizowania „mobilności" sprawcy.

10. ZABEZPIECZENIA I WARUNKI DOTYCZĄCE WYJĄTKÓW (ART. 5 UST. 2–7 AKTU W SPRAWIE SZTUCZNEJ INTELIGENCJI)

10.1. Poszukiwana osoba a zabezpieczenia (art. 5 ust. 2 aktu w sprawie sztucznej inteligencji)

Art. 5 ust. 2 aktu w sprawie sztucznej inteligencji stanowi:

„Systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w odniesieniu do któregokolwiek z celów, o których mowa w ust. 1 akapit pierwszy lit. h), mogą być wykorzystywane do celów określonych w tej literze, jedynie w celu potwierdzenia tożsamości konkretnej poszukiwanej osoby, z uwzględnieniem przy tym następujących elementów:

a) charakter sytuacji powodującej konieczność ewentualnego wykorzystania takiego systemu, w szczególności powagę,

prawdopodobieństwo i skalę szkody, która zostałaby wyrządzona w przypadku niewykorzystania tego systemu;

b) konsekwencje wykorzystania takiego systemu dla praw i wolności wszystkich zainteresowanych osób, w szczególności powagę, prawdopodobieństwo i skalę tych konsekwencji.

Ponadto wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w odniesieniu do któregokolwiek z celów, o których mowa w ust. 1 akapit pierwszy lit. h) niniejszego artykułu, musi przebiegać z zachowaniem niezbędnych i proporcjonalnych zabezpieczeń i warunków w odniesieniu do takiego wykorzystywania zgodnie z zezwalającym na takie wykorzystanie prawem krajowym, w szczególności w odniesieniu do ograniczeń czasowych, geograficznych i osobowych. Wykorzystanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej jest dozwolone tylko wtedy, gdy organ ścigania przeprowadził ocenę skutków dla praw podstawowych zgodnie z art. 27 oraz zarejestrował system w bazie danych UE zgodnie z przepisami art. 49. W należycie uzasadnionych nadzwyczajnych przypadkach można jednak rozpocząć korzystanie z takich systemów bez rejestracji w bazie danych UE, pod warunkiem że taka rejestracja zostanie dokonana bez zbędnej zwłoki".

356) Korzystanie z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do jednego z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji podlega pewnym zabezpieczeniom i warunkom, które wyszczególniono w art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji.

357) Po pierwsze, wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw jest dozwolone jedynie „w celu w celu potwierdzenia tożsamości konkretnej poszukiwanej osoby". Ten pierwszy warunek ma na celu znalezienie równowagi między powagą sytuacji i szkodą wynikającą z niewykorzystania systemu a wpływem technologii na prawa i wolności jednostek. Ma on zapobiec masowej inwigilacji poprzez ukierunkowanie stosowania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym na konkretną osobę. W związku z tym stosowanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw powinno być dozwolone wyłącznie w odniesieniu do poszukiwanych osób.

358) Użycie wyrażenia „potwierdzenie tożsamości", a nie „identyfikacja" ma stanowić dodatkowe zabezpieczenie praw podstawowych ograniczające ryzyko niezróżnicowanego nadzoru i oznacza, że identyfikacja osoby w rozumieniu art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji musi być ukierunkowana. Wyrażenie to należy rozumieć w ten sposób, że korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym można zainicjować wyłącznie w celu poszukiwania konkretnych osób fizycznych, w odniesieniu do których organy ścigania mają powody, by sądzić, że są ofiarami przestępstw wymienionych w art. 5 ust. 1 lit. h)

ppkt (i) aktu w sprawie sztucznej inteligencji lub że dotyczy ich jeden ze scenariuszy opisanych w art. 5 ust. 1 lit. h) ppkt (ii) lub art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji. W praktyce oznacza to porównanie danych zgromadzonych w czasie rzeczywistym z danymi znajdującymi się w referencyjnej bazie danych. Jeżeli chodzi o wykorzystywanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w scenariuszach opisanych w art. 5 ust. 1 lit. h) ppkt (ii) aktu w sprawie sztucznej inteligencji oraz do celów prowadzenia postępowania przygotowawczego w rozumieniu art. 5 ust. 1 lit. h) ppkt (iii) aktu w sprawie sztucznej inteligencji, organy ścigania niekoniecznie muszą znać tożsamość osób, których poszukują, przed skorzystaniem z systemu. Jeżeli posiadają one faktyczne przesłanki i informacje o planowanym ataku terrorystycznym grupy terrorystycznej (nie wiedząc, kto wykona plan) w określonym czasie i miejscu, system zdalnej identyfikacji biometrycznej może być wykorzystywany do identyfikacji sprawcy z grupy terrorystycznej, pod warunkiem że organy ścigania utworzyły referencyjną bazę danych zawierającą dane biometryczne osób należących do grupy terrorystycznej. We wszystkich trzech scenariuszach opisanych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji „potwierdzenie tożsamości" może również obejmować lokalizowanie danej osoby.

359) Po drugie, przed użyciem systemu należy ocenić charakter sytuacji powodującej konieczność ewentualnego wykorzystania takiego systemu, w szczególności powagę, prawdopodobieństwo i skalę szkody dla osób fizycznych, społeczeństwa i celów organów ścigania, która zostałaby wyrządzona w przypadku niewykorzystania tego systemu, w porównaniu z konsekwencjami wykorzystania takiego systemu dla praw i wolności wszystkich zainteresowanych osób, w szczególności powagę, prawdopodobieństwo i skalę tych konsekwencji. Należy przy tym ocenić, czy organy ścigania lub podmioty działające w ich imieniu dysponują mniej inwazyjnymi alternatywnymi rozwiązaniami.

Na przykład organom ścigania zabrania się stosowania systemów rozpoznawania twarzy w czasie rzeczywistym na ulicy w oparciu o ogólne obawy związane z bezpieczeństwem, zapobieganiem przestępczości i nadmierną liczbą osób, ponieważ wiązałoby się to ze stałym monitorowaniem i nadzorem wszystkich osób, nie byłoby ograniczone w czasie, a zatem nie spełniałoby kryteriów wyjątku od zakazu określonego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji.

360) Kryterium „**powagi**" stosowane w tym kontekście w odniesieniu do ewentualnych szkód i konsekwencji oznacza zróżnicowanie stopnia ingerencji w odnośne prawa podstawowe, co jest związane z zasadą proporcjonalności[212]. Jeśli chodzi o ingerencję w prawa podstawowe, niektóre rodzaje ingerencji są postrzegane jako poważniejsze niż inne.

[212] Wyrok Trybunału Sprawiedliwości z dnia 2 października 2018 r., *Ministerio Fiscal*, C-207/16, ECLI:EU:C:2018:788, pkt 55, w którym Trybunał stwierdza, że „dostęp powinien być powiązany z wagą ingerencji w odnośne prawa podstawowe".

361) Kryterium „**skali**” odnosi się w szczególności do liczby i kategorii osób dotkniętych ingerencją (w tym dzieci oraz osób szczególnie wrażliwych lub osób zmarginalizowanych).

362) Wreszcie „**prawdopodobieństwo**” oznacza możliwość wystąpienia zdarzenia.

363) Ocena powagi, skali i prawdopodobieństwa szkody i konsekwencji powinna być częścią oceny skutków dla praw podstawowych, którą organ ścigania jest zobowiązany przeprowadzić (zob. poniżej). Ocenę taką przeprowadza się indywidualnie w odniesieniu do każdego przypadku.

364) Po trzecie, korzystanie ze zdalnej identyfikacji biometrycznej w czasie rzeczywistym powinno być wyraźnie ograniczone pod względem zakresu geograficznego, czasu trwania i poszukiwanej osoby. Ma to na celu zapewnienie, aby systemy zdalnej identyfikacji biometrycznej były stosowane tylko wtedy, gdy jest to bezwzględnie konieczne.

365) Jeżeli chodzi o **ograniczenie geograficzne**, może ono obejmować jeden lub kilka obszarów geograficznych wyznaczonych w oparciu o obiektywne i niedyskryminacyjne czynniki. W przypadku identyfikacji biometrycznej oznacza to, że ograniczenie geograficzne ma zastosowanie do obszaru o wyraźnie wytyczonych granicach, w odniesieniu do którego istnieją przesłanki wskazujące na to, że będzie on miejscem zdarzenia. W normalnych okolicznościach taki określony obszar nie powinien obejmować całego miasta lub kraju – wytyczanie granic powinno być bardziej ukierunkowane.

366) Inne zabezpieczenie dotyczy **zakresu podmiotowego** środka, tj. określenia **kategorii osób, których środek dotyczy.** Wykluczałoby to nieukierunkowane i niezróżnicowane identyfikowanie osób bez dalszych przesłanek wskazujących na incydent.

367) Wreszcie **ograniczenie czasowe** odnosi się do okresu ograniczonego do tego, co jest bezwzględnie konieczne, ale który to okres w razie potrzeby może zostać przedłużony zgodnie z obowiązującymi przepisami. Wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym nie może zatem odbywać się przez nieograniczony lub niejasno określony okres czasu. Okres ten należy określić w świetle konkretnych przesłanek, które uzasadniają wykorzystanie systemów zdalnej identyfikacji biometrycznej.

368) Po czwarte, przed zastosowaniem, organ ścigania stosujący system zdalnej identyfikacji biometrycznej w czasie rzeczywistym musi przeprowadzić ocenę skutków dla praw podstawowych (FRIA) i zarejestrować system w bazie danych UE (z wyjątkiem należycie uzasadnionych przypadków).

10.1.1. Ocena skutków dla praw podstawowych

369) Oceny skutków dla praw podstawowych prowadzone na podstawie art. 5 ust. 2 aktu w sprawie sztucznej inteligencji muszą spełniać warunki określone w art. 27 aktu w sprawie sztucznej inteligencji. Przepis ten określa wymogi dotyczące oceny skutków dla praw podstawowych mające zastosowanie do systemów AI wysokiego ryzyka.

370) W okresie między wejściem w życie zakazów ustanowionych w art. 5 aktu w sprawie sztucznej inteligencji (2 lutego 2025 r.) a momentem, w którym przepisy dotyczące systemów AI wysokiego ryzyka będą miały zastosowanie (2 sierpnia 2026 r.), wymogi dotyczące oceny skutków dla praw podstawowych określone w art. 27 aktu w sprawie sztucznej inteligencji: powinny zostać wdrożone przez podmioty stosujące systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym, które spełniają warunki, by skorzystać z co najmniej jednego wyjątku wymienionego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji. Poniższe tymczasowe wytyczne dotyczą wyłącznie wykorzystywania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w okresie, zanim będą miały zastosowanie obowiązki dotyczące systemów AI wysokiego ryzyka i zanim Komisja przyjmie wzór oceny skutków dla praw podstawowych i zapewni dalsze wytyczne dotyczące obowiązku wynikającego z art. 27 aktu w sprawie sztucznej inteligencji.

371) Ocena skutków dla praw podstawowych to nowy rodzaj oceny skutków, której celem jest określenie wpływu, jaki niektóre systemy AI wysokiego ryzyka, w tym systemy zdalnej identyfikacji biometrycznej, mogą wywierać na prawa podstawowe. Ocena skutków dla praw podstawowych jest narzędziem rozliczalności. Ocena skutków dla praw podstawowych nie zastępuje istniejącej oceny skutków dla ochrony danych, którą administratorzy danych (tj. osoby odpowiedzialne za przetwarzanie danych osobowych) muszą przeprowadzać na podstawie art. 27 dyrektywy o ochronie danych w sprawach karnych, art. 35 RODO lub art. 39 rozporządzenia w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych przez instytucje, organy i jednostki organizacyjne Unii.

> Na przykład ocenę skutków dla ochrony danych należy przeprowadzić, gdy dane biometryczne są przetwarzane za pomocą nowych technologii w przestrzeni publicznej (takich jak telewizja przemysłowa, systemy AI do rozpoznawania twarzy i kamery osobiste), które mogą powodować wysokie ryzyko naruszenia praw i wolności osób fizycznych.

372) Podczas gdy ocena skutków dla ochrony danych koncentruje się na ryzyku dla praw i wolności osób fizycznych wynikającym z przetwarzania ich danych osobowych, ocena skutków dla praw podstawowych obejmuje ewentualny wpływ systemów AI na prawa podstawowe osób fizycznych w ujęciu bardziej ogólnym. Zakres oceny skutków dla praw podstawowych jest zatem szerszy pod względem objętych nim działań i ocenianych praw podstawowych. W przypadku gdy system AI przetwarza dane osobowe (co ma miejsce w przypadku systemów zdalnej identyfikacji biometrycznej), ocena skutków dla praw podstawowych powinna stanowić uzupełnienie oceny skutków dla ochrony danych przeprowadzonej przez podmiot stosujący będący administratorem danych[213] i nie powinna ona obejmować aspektów już uwzględnionych w ocenie skutków dla ochrony danych, aby uniknąć powielania działań. Analiza oceny skutków dla praw podstawowych zawarta w niniejszych wytycznych ogranicza się do

[213] Art. 27 ust. 4 aktu w sprawie sztucznej inteligencji.

dozwolonego stosowania zdalnej identyfikacji biometrycznej w czasie rzeczywistym i ma pełnić rolę tymczasowych zaleceń dla podmiotów stosujących w okresie przejściowym, zanim Urząd ds. AI przedstawi wzór takiej oceny[214].

373) Obowiązek przeprowadzenia oceny skutków dla praw podstawowych na podstawie art. 5 ust. 2 aktu w sprawie sztucznej inteligencji został nałożony na podmioty stosujące system zdalnej identyfikacji biometrycznej, a nie na podmioty lub organy ani inne osoby działające w ich imieniu. Jeżeli inne podmioty działają w imieniu podmiotu stosującego/organu ścigania, będą musiały wnieść wkład w przygotowanie oceny skutków dla praw podstawowych, przekazując wszystkie informacje istotne dla prawidłowego przeprowadzenia tej oceny.

374) Ocenę skutków dla praw podstawowych należy przeprowadzić przed wdrożeniem zatwierdzonego systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

375) Zgodnie z art. 27 aktu w sprawie sztucznej inteligencji ocena skutków dla praw podstawowych powinna zawierać następujące informacje:

- Opis wykorzystania systemu zdalnej identyfikacji biometrycznej i procedur podmiotu stosującego dotyczących tego wykorzystania, wraz z przeznaczeniem

Opis powinien zawierać:

– imię i nazwisko lub nazwę podmiotu stosującego;

– cele w zakresie ścigania przestępstw, do których to celów system zdalnej identyfikacji biometrycznej w czasie rzeczywistym będzie wykorzystywany;

– opis referencyjnej bazy danych, która będzie wykorzystywana na potrzeby porównania danych pochodzących z identyfikacji biometrycznej, w tym źródła danych biometrycznych (wizerunki twarzy, próbki głosu itp.), które będą wykorzystywane;

– opis technologii stanowiącej podstawę systemu w celu wyjaśnienia jego działania (z odniesieniami do dostępnej dokumentacji dostarczonej przez dostawcę oraz jego nazwy)[215];

– podstawa prawna, w oparciu o którą system zdalnej identyfikacji biometrycznej w czasie rzeczywistym będzie stosowany.

- Okres i częstotliwość wykorzystywania

Zgodnie z art. 5 pkt 3 aktu w sprawie sztucznej inteligencji każde pojedyncze wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w ramach jednego z dozwolonych wyjątków wymaga uzyskania uprzedniego zezwolenia udzielonego przez organ wymiaru sprawiedliwości lub inny niezależny organ. Natomiast w przypadku oceny skutków dla praw podstawowych podmioty

[214] W związku z tym analiza nie obejmuje ogólnie przypadku systemów AI wysokiego ryzyka.

[215] Po wejściu w życie przepisów dotyczących systemów AI wysokiego ryzyka będzie można to zrobić poprzez odniesienie się do numeru identyfikacyjnego systemu w bazie danych UE oraz zawartych w niej informacji dotyczących systemu.

stosujące muszą ogólnie wskazać okres wykorzystywania i spodziewaną częstotliwość wykorzystywania.

- Kategorie osób i grup, na które system ma wpływ

Do celów wyjątku przewidzianego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji w ocenie skutków dla praw podstawowych należy wprowadzić rozróżnienie między:

– poszukiwaną osobą, która może być ofiarą przestępstwa, sprawcą lub podejrzanym,

– osobami, których dane biometryczne znajdują się w referencyjnej bazie danych, oraz

– kategoriami osób przebywających w okolicy obszarów, na których zostanie zastosowany system zdalnej identyfikacji biometrycznej.

Wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym będzie miało wpływ nie tylko na prawa podstawowe poszukiwanej osoby. Naruszone zostaną również prawa innych osób, których dane biometryczne będą wykorzystywane do celów porównawczych, przechodniów i osób przypadkowo znajdujących się w obszarze poszukiwania. Zakres geograficzny obszarów poszukiwania objętych systemem zdalnej identyfikacji biometrycznej w czasie rzeczywistym będzie przekładał się na liczbę osób, na które system będzie miał wpływ.

- Konkretne ryzyko wyrządzenia szkody osobom, na które AI ma wpływ

376) Prawa podstawowe, na które może mieć wpływ wykorzystywanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, obejmują w szczególności:

- prawo do życia prywatnego i rodzinnego, w tym uzasadnione oczekiwania obywateli co do anonimowości w przestrzeni publicznej;
- prawo do ochrony danych, ponieważ systemy zdalnej identyfikacji biometrycznej opierają się na przetwarzaniu danych biometrycznych i innych danych osobowych (np. imion i nazwisk, numerów identyfikacyjnych, a także danych wrażliwych, takich jak pochodzenie etniczne) w celu identyfikacji konkretnych osób;
- wolność myśli, sumienia i wyznania, wolność wypowiedzi oraz wolność zgromadzania się i stowarzyszania się w przeszukiwanej przestrzeni publicznej, w odniesieniu do których korzystanie z systemów zdalnej identyfikacji biometrycznej mogłoby mieć efekt mrożący, uniemożliwiając osobom pełne korzystanie z przysługujących im praw i wolności, ponieważ jeśli dane osoby wiedzą, że są monitorowane, mogą zmienić swoje zachowanie lub nawet powstrzymać się od określonych działań;
- prawo do skutecznego środka prawnego i dostępu do bezstronnego sądu;

- prawo do niedyskryminacji, jeżeli system jest obciążony uprzedzeniami (takimi jak uprzedzenia ze względu na płeć, pochodzenie etniczne lub rasowe) i prowadzi do błędnej identyfikacji podejrzanego lub sprawcy;
- prawo do godności człowieka, ze względu na poczucie, że jest się zredukowanym przez system do roli przedmiotu;
- domniemanie niewinności i prawo do obrony, ponieważ żadna decyzja mająca niepożądane skutki dla danej osoby nie może być podejmowana wyłącznie na podstawie wyników uzyskanych z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym;
- prawa dziecka, w przypadku, gdy ofiara, osoba zaginiona lub podejrzany jest małoletnim;
- prawa osób starszych, w przypadku osoby zaginionej.

Aby ocenić konkretne ryzyko wyrządzenia szkody zidentyfikowanym osobom lub grupom, na które AI ma wpływ, w ocenie skutków dla praw podstawowych należy określić prawa podstawowe tych osób i ocenić wpływ AI na ich prawa podstawowe, w tym dotkliwość skutków i ich skalę, z uwzględnieniem osób potencjalnie poszkodowanych.

Ta część oceny skutków dla praw podstawowych powinna również obejmować ustalenie, czy wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym jest konieczne i proporcjonalne, biorąc pod uwagę cele i okoliczności, w których system ma być wykorzystany, w tym istnienie lub brak mniej inwazyjnych rozwiązań alternatywnych. W ocenie skutków dla praw podstawowych należy określić efektywność i poziom dokładności systemu w oparciu o dokumentację techniczną oraz, o ile są dostępne, dane treningowe, na podstawie których testowano i rozwijano technologię w celu zapobiegania stronniczości i dyskryminacji.

W ocenie skutków dla praw podstawowych należy również określić wpływ korzystania z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym na prawa podstawowe wszystkich potencjalnie poszkodowanych osób, w szczególności podejrzanego lub sprawcy, poszukiwanej ofiary i innych osób obecnych w przestrzeni publicznej, w której odbywają się poszukiwania. W zakresie, w jakim system przetwarza dane biometryczne tych osób, będzie to miało wpływ na ich prawa do życia prywatnego i rodzinnego oraz na ochronę danych, co należy ocenić w ramach oceny skutków dla ochrony danych w odniesieniu do działań związanych z przetwarzaniem danych. Jeśli chodzi o inne działania związane z wykorzystywaniem systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym i wpływ na inne prawa podstawowe, ocena skutków dla praw podstawowych stanowi uzupełnienie oceny skutków dla ochrony danych. W zależności od kontekstu zastosowania może dojść do naruszeń innych praw podstawowych tych osób, takich jak prawo do godności człowieka, wolność myśli, sumienia i religii, wolność zgromadzeń lub wolności wypowiedzi, prawo do skutecznego środka prawnego i dostępu do bezstronnego sądu, domniemanie niewinności i prawo do obrony oraz prawa dziecka.

Ocenę skutków dla praw podstawowych należy przeprowadzić na poziomie abstrakcyjnym przed pierwszym oddaniem systemu AI do użytku. W indywidualnym wniosku o wydanie przez organ wymiaru sprawiedliwości lub inny niezależny organ administracyjny zezwolenia na każde wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym (zob. sekcja 10.2 pkt 383 poniżej) należy doprecyzować wynikające z kontekstu szczególne względy, które warunkują wpływ wykorzystania systemu zdalnej identyfikacji biometrycznej w każdym indywidualnym przypadku.

- Środki nadzoru ze strony człowieka

Zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji żadna decyzja, która miałaby niepożądane skutki dla danej osoby, nie może być podejmowana wyłącznie na podstawie wyników uzyskanych z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym. W związku z tym w ocenie skutków dla praw podstawowych należy opisywać procedury, które będą stosowane podczas korzystania z systemu, oraz sposób interpretacji wyników w kontekście procesu decyzyjnego. Procedury te powinny zawierać instrukcje dotyczące stosowania systemu zdalnej identyfikacji biometrycznej, wyjaśniać rolę człowieka w weryfikacji i interpretacji wyników oraz zapewniać szkolenia w zakresie obsługi systemu. Osoba odpowiedzialna za nadzór ze strony człowieka powinna posiadać „odpowiedni poziom kompetencji w zakresie AI oraz odpowiedni poziom przeszkolenia i uprawnień”[216], aby rozumieć, w jaki sposób system działa i kiedy działa nieskutecznie lub nieprawidłowo.

Inne kwestie dotyczące nadzoru ze strony człowieka i monitorowania na mocy art. 14 i 26 aktu w sprawie sztucznej inteligencji są również istotne i powinny zostać opisane.

- Środki ograniczające ryzyko

Oprócz wdrożenia środków nadzoru ze strony człowieka (w tym w celu uniknięcia dyskryminacji) podmiot stosujący powinien wyjaśnić, jakie środki naprawcze są dostępne w przypadku urzeczywistnienia się ryzyka, w tym procedury zarządzania i mechanizmy rozpatrywania skarg (np. w przypadku błędnej identyfikacji).

10.1.2. Rejestracja zatwierdzonych systemów zdalnej identyfikacji biometrycznej

377) W art. 5 ust. 2 aktu w sprawie sztucznej inteligencji zobowiązano również podmiot stosujący system zdalnej identyfikacji biometrycznej w czasie rzeczywistym wykorzystywany w przestrzeni publicznej do celów ścigania przestępstw do zarejestrowania tego systemu w bazie danych UE, o której mowa w art. 49 aktu w sprawie sztucznej inteligencji. W należycie uzasadnionych nadzwyczajnych przypadkach (takich jak bezpośrednie zagrożenie) można jednak rozpocząć korzystanie z systemu jeszcze przed rejestracją, pod warunkiem że organ ścigania dokona rejestracji bez zbędnej zwłoki. Sformułowanie „bez zbędnej zwłoki” należy rozumieć jako „jak

[216] Motyw 91 aktu w sprawie sztucznej inteligencji.

najszybciej", biorąc pod uwagę okoliczności sytuacji nadzwyczajnej, która uniemożliwiła rejestrację systemu przed jego użyciem. To, czy rejestracja spełnia to kryterium, wymaga oceny dla każdej sprawy osobno. Nie można określić a priori dokładnego terminu. Zwłoka nie powinna być spowodowana celowym działaniem. Zgodnie z art. 49 ust. 4 aktu w sprawie sztucznej inteligencji systemy zdalnej identyfikacji biometrycznej wykorzystywane do celów ścigania przestępstw będą rejestrowane w bezpiecznej, niepublicznej sekcji bazy danych, z ograniczonymi informacjami i ograniczonym dostępem do tych informacji.

Na przykład zwrócenie się do organów ścigania o rejestrację systemu zdalnej identyfikacji biometrycznej w ciągu 24 godzin od jego zastosowania można uznać za racjonalną zwłokę, jeżeli system został zastosowany w sytuacji bezpośredniego zagrożenia życia, na przykład w przypadku masowego zabójcy.

10.2. Konieczność uzyskania uprzedniego zezwolenia

378) Zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji każde indywidualne wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym wymaga uzyskania **uprzedniego zezwolenia**, a zautomatyzowane podejmowanie decyzji wywołującej niepożądane skutki prawne **wyłącznie w oparciu o wyniki uzyskane z takiego systemu** jest zakazane.

Art. 5 ust. 3 aktu w sprawie sztucznej inteligencji stanowi:

Na potrzeby ust. 1 akapit pierwszy lit. h) i ust. 2, każde wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw wymaga uzyskania uprzedniego zezwolenia udzielonego przez organ wymiaru sprawiedliwości lub wydający wiążące decyzje niezależny organ administracyjny państwa członkowskiego, w którym ma nastąpić wykorzystanie; zezwolenie to wydawane jest na uzasadniony wniosek i zgodnie ze szczegółowymi przepisami prawa krajowego, o których mowa w ust. 5. W należycie uzasadnionych pilnych przypadkach korzystanie z takiego systemu można jednak rozpocząć bez zezwolenia, pod warunkiem że wniosek o takie zezwolenie zostanie złożony bez zbędnej zwłoki, najpóźniej w ciągu 24 godzin. W przypadku odmowy udzielenia takiego zezwolenia wykorzystywanie systemu wstrzymuje się ze skutkiem natychmiastowym, a wszystkie dane, a także rezultaty i wyniki uzyskane podczas tego wykorzystania muszą zostać natychmiast odrzucone i usunięte.

Właściwy organ wymiaru sprawiedliwości lub niezależny organ administracyjny udziela zezwolenia tylko wtedy, gdy jest przekonany, na podstawie obiektywnych dowodów lub jasnych przesłanek, które mu przedstawiono, że wykorzystanie danego systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym jest konieczne i proporcjonalne do osiągnięcia jednego z celów określonych w ust. 1 akapit pierwszy lit. h), wskazanego we wniosku, a w szczególności ogranicza się do tego, co jest bezwzględnie konieczne w odniesieniu do przedziału czasowego, a także zakresu geograficznego i podmiotowego. Podejmując decyzję w sprawie wniosku

organ ten bierze pod uwagę elementy, o których mowa w ust. 2. Nie można wydać decyzji wywołującej niepożądane skutki prawne dla danej osoby wyłącznie na podstawie wyników uzyskanych z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

10.2.1. Cel

379) Celem wymogu uzyskania uprzedniego zezwolenia („zezwolenie *ex ante*") na każde zastosowanie systemu zdalnej identyfikacji biometrycznej „w czasie rzeczywistym" w przestrzeni publicznej do celów ścigania przestępstw jest konieczność dokonania oceny i podjęcia decyzji, czy przewidywane zastosowanie takiego systemu do wymienionych celów jest:

– konieczne i proporcjonalne do osiągnięcia któregoś z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii), tj. ukierunkowanego poszukiwania konkretnych ofiar, zapobiegania konkretnym zagrożeniom lub lokalizowania bądź identyfikowania sprawców; oraz

– ograniczone do tego, co jest bezwzględnie konieczne w odniesieniu do okresu oraz zakresu geograficznego i podmiotowego.

380) Skutkiem tych wymogów jest to, że przed zastosowaniem jakiegokolwiek systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw należy przeprowadzić podwójną ocenę konieczności i proporcjonalności. Najpierw użytkownik powinien dokonać oceny podczas przeprowadzania oceny skutków dla praw podstawowych, zgodnie z wymogami art. 5 ust. 2 aktu w sprawie sztucznej inteligencji. Następnie, zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji, organ wymiaru sprawiedliwości lub niezależny organ administracyjny musi również ocenić konieczność i proporcjonalność wykorzystania takiego systemu w granicach prawa krajowego stanowiącego podstawę prawną takiego wykorzystania, uwzględniając postanowienia Karty i inne przepisy prawa Unii. W związku z tym każdy taki system może być stosowany wyłącznie 1) po przeprowadzeniu oceny skutków dla praw podstawowych oraz 2) po uzyskaniu zezwolenia na takie stosowanie wydanego przez właściwy organ krajowy.

381) Art. 5 ust. 3 aktu w sprawie sztucznej inteligencji należy odczytywać i interpretować w związku z art. 5 ust. 5 aktu w sprawie sztucznej inteligencji: aby można było zezwolić na zastosowanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym, w zainteresowanych państwach członkowskich musi istnieć przyjęte prawo krajowe, które zezwala na takie zastosowanie[217]. W niektórych państwach członkowskich obowiązuje już system uzyskania uprzedniego zezwolenia na wykorzystanie systemów biometrycznych zgodnie z innymi przepisami unijnymi lub krajowymi, takimi jak przepisy o ochronie danych.

10.2.2. Główna zasada: uzyskanie uprzedniego zezwolenia organu wymiaru sprawiedliwości lub niezależnego organu administracyjnego

[217] Zob. art. 5 ust. 2 aktu w sprawie sztucznej inteligencji: „[...] zgodnie z zezwalającym na takie wykorzystanie prawem krajowym [...]".

382) Wykorzystanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, które realizują jeden z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji i które zostały przewidziane w prawie krajowym zainteresowanych państw członkowskich, wymaga uzyskania zezwolenia organu wymiaru sprawiedliwości lub niezależnego organu administracyjnego **przed wykorzystaniem takich systemów**. Jest to główna zasada.

383) Wyjątek stanowią jednak pilne przypadki. Muszą one być **należycie uzasadnione**[218]. Pilne przypadki **opisano jako** „sytuacj[e], w których potrzeba wykorzystania danego systemu jest na tyle duża, że uzyskanie zezwolenia **przed rozpoczęciem korzystania z tego systemu AI jest faktycznie i obiektywnie niemożliwe**”[219]. W takich pilnych przypadkach „wykorzystanie systemu AI powinno być ograniczone do **bezwzględnie niezbędnego minimum** i powinno **podlegać odpowiednim zabezpieczeniom i warunkom** określonym w prawie krajowym i sprecyzowanym przez sam organ ścigania w kontekście każdego przypadku nadzwyczajnego wykorzystania”.

10.2.2.1. Uprzedni i uzasadniony wniosek zgodnie z krajowymi przepisami proceduralnymi

a) Kto składa wniosek?

384) Chociaż nie zostało to sprecyzowane, można założyć, że wniosek zazwyczaj zostanie złożony przez podmiot stosujący, tj. **przez właściwy organ (organ ścigania)**. Zgodnie z definicją organu ścigania zawartą w art. 3 pkt 45 lit. b) aktu w sprawie sztucznej inteligencji za organ ścigania uznaje się również każdy „inny organ lub podmiot, któremu na podstawie prawa państwa członkowskiego powierzono sprawowanie władzy publicznej i wykonywanie uprawnień publicznych do celów zapobiegania przestępczości, prowadzenia postępowań przygotowawczych, wykrywania i ścigania czynów zabronionych lub wykonywania kar, w tym ochrony przed zagrożeniami dla bezpieczeństwa publicznego i zapobiegania takim zagrożeniom”, i organ ten lub podmiot może być również odpowiedzialny jako „właściwy organ” za złożenie wniosku o uprzednie zezwolenie.

385) Wykorzystywanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do działań nieobjętych zakresem aktu w sprawie sztucznej inteligencji nie wymaga zezwolenia na podstawie art. 5 ust. 3 aktu w sprawie sztucznej inteligencji. Jeżeli następnie taki system będzie wykorzystywany do celów ścigania przestępstw, jego wykorzystanie będzie wchodzić w zakres aktu w sprawie sztucznej, a zezwolenie będzie wymagane, jeżeli zostaną spełnione wymogi określone w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji.

b) W odniesieniu do jakiego wykorzystania składa się wniosek?

386) Uzyskanie uprzedniego zezwolenia jest konieczne w przypadku wykorzystywania systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym” w przestrzeni

---

[218] Oznacza to, że „organ ścigania powinien w takich sytuacjach wystąpić o takie zezwolenie, podając powody, dla których nie był w stanie wystąpić o nie wcześniej, bez zbędnej zwłoki i nie później niż w ciągu 24 godzin” (motyw 35 aktu w sprawie sztucznej inteligencji).

[219] Motyw 35 aktu w sprawie sztucznej inteligencji.

publicznej **do celów ścigania przestępstw,** nawet jeżeli systemy te są obsługiwane przez inne strony w imieniu organów ścigania, na przykład kluby sportowe lub centra handlowe.

Na przykład:

– organizacja, której powierzono zasoby na poszukiwanie zaginionych dzieci, decyduje się na korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym. Nie powierzono jej sprawowania władzy publicznej i wykonywania uprawnień publicznych ani zapobiegania przestępstwom lub zadań służących zapobieganiu zagrożeniom dla bezpieczeństwa publicznego. Takie wykorzystanie nie wchodzi w zakres zakazu ustanowionego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji, ponieważ nie służy ono do celów ścigania przestępstw. System ten zostanie jednak zaklasyfikowany jako system „wysokiego ryzyka" (załącznik III pkt 1 lit. a)), a zgodnie z art. 36 RODO konieczne może być przeprowadzenie uprzednich konsultacji z organem nadzorczym ds. ochrony danych. W zależności od mającego zastosowanie prawa krajowego i tego, czy zastosowanie ma jeden z wyjątków od art. 9 ust. 1 RODO, do takiego przetwarzania może być również wymagane uprzednie zezwolenie. Gdyby natomiast ta sama organizacja została poproszona przez organy ścigania o działanie w ich imieniu i zajęcie się poszukiwaniem zaginionych dzieci w kontekście ścigania przestępstw oraz pod nadzorem i zgodnie z instrukcjami właściwych organów ścigania, zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji wymagane byłoby uzyskanie wcześniejszego zezwolenia;

– organizacja prywatna, której powierzono zapewnianie środków pomocy osobom, które mogą stać się ofiarami klęski żywiołowej[220], postanawia w tym celu skorzystać z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym. Takie wykorzystanie nie wchodzi w zakres zakazu ustanowionego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji, ponieważ nie służy ono do celów ścigania przestępstw. System ten zostanie jednak zaklasyfikowany jako system „wysokiego ryzyka" (załącznik III pkt 1 lit. a)), a zgodnie z art. 36 RODO konieczne może być przeprowadzenie uprzednich konsultacji z organem nadzorczym ds. ochrony danych. W zależności od mającego zastosowanie prawa krajowego i tego, czy zastosowanie ma jeden z wyjątków od art. 9 ust. 1 RODO, do takiego przetwarzania może być również wymagane uprzednie zezwolenie.

c) Kiedy? „Każde wykorzystanie"

387) Zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji „każde wykorzystanie" wymaga uprzedniego zezwolenia. Oznacza to, że zezwolenie takie należy uzyskać nie w momencie poprzedzającym zainstalowanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, lecz przed każdym ich konkretnym wykorzystaniem.

[220] Klęski żywiołowe obejmują powodzie rzeczne lub pożary spowodowane przez naturę.

Na przykład:

– policja instaluje na głównym dworcu kolejowym miasta kamery telewizji przemysłowej przystosowane do pomiarów biometrycznych (nie ma potrzeby uzyskania zezwolenia na mocy aktu w sprawie sztucznej inteligencji, ale system biometryczny musi spełniać wymogi dotyczące systemów wysokiego ryzyka; przed pierwszym użyciem należy przeprowadzić ocenę skutków dla praw podstawowych, a przed każdym indywidualnym wykorzystaniem systemu wymagane jest uzyskanie zezwolenia wydanego przez organ wymiaru sprawiedliwości lub niezależny organ administracyjny).

Policja dysponuje konkretnymi przesłankami wskazującymi na to, że terrorysta przyjedzie pociągiem do miasta (w celu identyfikacji w czasie rzeczywistym wymagane jest uprzednie zezwolenie).

d) Uzasadniony wniosek

388) W art. 5 ust. 3 aktu w sprawie sztucznej inteligencji wprowadzono wymóg, aby każdy indywidualny wniosek o wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym był „uzasadniony", a zatem odpowiednio uargumentowany i umotywowany.

389) Niektóre państwa członkowskie zezwalają na składanie takich wniosków przez internet[221]. Zgodnie z art. 5 ust. 5 aktu w sprawie sztucznej inteligencji w przepisach krajowych należy określić wymogi dotyczące dokładnej treści wniosku, przy pełnym uwzględnieniu wymogów określonych powyżej, w tym wystarczających dowodów do ustalenia bezwzględnej konieczności i proporcjonalności korzystania z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym oraz innych istotnych aspektów, aby odzwierciedlić wyjątkowy charakter zezwolenia na takie wykorzystanie.

10.2.2.2. Zezwolenie wydane przez organ wymiaru sprawiedliwości lub niezależny organ administracyjny

390) Zezwolenia może udzielić wyłącznie organ wymiaru sprawiedliwości lub niezależny organ administracyjny, którego decyzja jest wiążąca.

a) Niezależny organ

391) TSUE dokonał wykładni pojęcia „niezależności" w różnych kontekstach. Na przykład w sprawie *HK przeciwko Prokuratuur* TSUE wyjaśnił, że niezależność oznacza, iż organ zajmuje „neutralną pozycję"[222]. TSUE sprecyzował, że organ zaangażowany we wcześniejsze dochodzenia, w tym przypadku prokuratura, nie ma takiej niezależności. Podobne względy mogą mieć zastosowanie do niezależności wymaganej na mocy art. 5 ust. 3 aktu w sprawie sztucznej inteligencji, co oznaczałoby, że organ wydający zezwolenie musi być niezależny od organu wykorzystującego system zdalnej identyfikacji biometrycznej. Miałoby to zastosowanie nie tylko do policji, ale również do sędziów śledczych lub prokuratorów, którzy nadzorują pracę policji i wykorzystanie

[221] Zob. na przykład wnioski o udzielenie zezwolenia składane do francuskiego organu ochrony danych, CNIL.

[222] Wyrok Trybunału Sprawiedliwości z dnia 2 marca 2021 r., *Prokuratuur*, C-746/18, ECLI:EU:C:2021:152, pkt 54.

systemu zdalnej identyfikacji biometrycznej, w odniesieniu do którego wystąpiono o zezwolenie.

392) W sprawie *Komisja przeciwko Polsce*, która dotyczyła kwestii, kiedy organ można uznać za niezależny w kontekście bezpieczeństwa kolei, TSUE stwierdził, że „w przypadku organów państwowych niezależność oznacza zwykle status, który zapewnia danemu organowi możliwość swobodnego działania w stosunku do podmiotów, wobec których należy zapewnić jego niezależność, z wyłączeniem jakichkolwiek instrukcji czy nacisków”[223]. Podobne przesłanki mogą mieć zastosowanie w kontekście art. 5 ust. 3 aktu w sprawie sztucznej inteligencji.

393) Organy wymiaru sprawiedliwości w społeczeństwie demokratycznym są co do zasady również organami niezależnymi. Władza sądownicza odgrywa istotną rolę, gdy jest niezależna od władzy wykonawczej i ustawodawczej, zapewniając kontrolę oraz nadzór nad stosowaniem prawa i praw podstawowych w sposób autonomiczny i niezależny. Niezawisłość sądów, zagwarantowana w art. 47 (Karty) i art. 6 ust. 1 EKPC, jest jednym z kluczowych aspektów praworządności[224].

b) Organ właściwy dla miejsca, w którym nastąpi wykorzystanie

394) Zezwolenie musi być skierowane do organu właściwego zgodnie z prawem krajowym[225].

c) Zezwolenie tylko wtedy, gdy wykorzystanie jest „konieczne i proporcjonalne” do osiągnięcia jednego z celów określonych w wyjątkach

395) Udzielając zezwolenia na wykorzystanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, należy zawsze ocenić, czy spełniono wymogi określone w art. 5 ust. 3 aktu w sprawie sztucznej inteligencji.

Wysoce inwazyjny charakter

396) W kontekście ochrony danych Europejska Rada Ochrony Danych (EROD) w swoich wytycznych 5/2022 i Europejski Inspektor Ochrony Danych (EIOD) uznali wykorzystywanie danych biometrycznych, w szczególności technologii rozpoznawania twarzy, za mające wpływ na szereg podstawowych praw i wolności. Pogląd ten podzielają Agencja Praw Podstawowych Unii Europejskiej i Rada Europy[226]. Zarówno TSUE[227], jak i ETPC[228] potwierdziły wrażliwy charakter przetwarzania danych biometrycznych.

---

223 Wyrok Trybunału Sprawiedliwości z dnia 13 czerwca 2018 r., *Komisja przeciwko Polsce*, C-530/16, ECLI:EU:C:2018:430, pkt 67.

224 Zob. R. Manko, *Judicial independence in the case law of the European Court of Human Rights* [Niezawisłość sądów w orzecznictwie Europejskiego Trybunału Praw Człowieka], Briefing, Biuro Analiz Parlamentu Europejskiego (EPRS), 2022, s. 12; X, *ECJ case law on judicial independence. A Chronological overview* [Orzecznictwo TSUE dotyczące niezawisłości sądów.– przegląd chronologiczny], Briefing, Biuro Analiz Parlamentu Europejskiego (EPRS), 2023, 12 s.

225 Zob. wyrok Trybunału Sprawiedliwości z dnia 6 października 2015 r., *Schrems*, C-362/14, https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:62014CJ0362, ECLI:EU:C:2015:650, pkt 44.

226 Komitet Doradczy Konwencji o ochronie osób w związku z automatycznym przetwarzaniem danych osobowych (ETS 108), *Guidelines on facial recognition* [Wytyczne dotyczące rozpoznawania twarzy], 2021.

227 Wyrok Trybunału Sprawiedliwości z dnia 26 stycznia 2023 r., *Ministerstvo na vatreshnite raboti*, C-205/21, ECLI:EU:C:2023:49, pkt 60–76 i 116–134.

228 Wyrok Europejskiego Trybunału Praw Człowieka z dnia 4 lipca 2023 r., *Glukhin* przeciwko *Rosji*, skarga nr 11519/20, ECLI:CE:ECHR:2023:0704JUD001151920, pkt 88 i 90 (zwany dalej „wyrokiem w sprawie *Glukhin* przeciwko *Rosji*”).

397) Każda ingerencja **w podstawowe prawa i wolności musi zawsze odbywać się z poszanowaniem ich istoty**. Wynika to z art. 52 ust. 1 Karty.

398) Pojęcie „istoty" podstawowych praw i wolności zostało rozwinięte w orzecznictwie TSUE i stanowi niezależną wartość w porządku prawnym Unii. Jeżeli istota podstawowego prawa lub wolności nie jest przestrzegana, oznacza to, że dany środek nadmiernie narusza prawo lub wolność, w związku z czym nie należy z góry zezwalać na ingerencję.

Tylko wtedy, gdy jest to „konieczne i proporcjonalne"

399) Zgodnie z art. 52 Karty każda ingerencja w podstawowe prawa i wolności wymaga „ustawy", która co do zasady powinna zapewniać poszanowanie zasady konieczności i proporcjonalności (zob. poniżej w odniesieniu do art. 5 ust. 5 aktu w sprawie sztucznej inteligencji). Zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji, prawo krajowe zezwalające na wykorzystywanie zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw powinno przewidywać, że zezwolenia na takie wykorzystanie można udzielić „tylko wtedy, gdy [organ] jest przekonany, [...] że wykorzystanie [...] jest konieczne i proporcjonalne do osiągnięcia jednego z celów określonych w [art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji]". Organy krajowe muszą sprawdzić, czy identyfikacja biometryczna jest bezwzględnie konieczna[229]. Ocena ta powinna opierać się na ocenie skutków dla praw podstawowych, która powinna już obejmować ogólną ocenę konieczności i proporcjonalności przeprowadzoną przed każdorazowym składaniem wniosku o zezwolenie na wykorzystanie w konkretnych okolicznościach.

10.2.2.3. Wyjątek od wymogu uzyskania uprzedniego zezwolenia: złożenie wniosku w ciągu 24 godzin i konsekwencje odrzucenia takiego wniosku

400) W pilnych przypadkach użytkownik może złożyć wniosek o udzielenie zezwolenia w ciągu 24 godzin od momentu wykorzystania systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym. W praktyce będzie to zazwyczaj moment, w którym zostaną „włączone" i wykorzystane kamery przystosowane do pomiarów biometrycznych, a system dokona pierwszego porównania danych biometrycznych. W celu wykazania terminowości wniosku należy udostępnić organowi rejestr czynności przetwarzania[230].

401) W takim przypadku we wniosku należy uzasadnić, dlaczego przed rozpoczęciem korzystania z systemu nie złożono uprzedniego wniosku.

10.2.2.4. Natychmiastowe wstrzymanie w przypadku odrzucenia wniosku o udzielenie zezwolenia i usunięcie danych

[229] W odniesieniu do zbierania danych zob. również: Wyrok Trybunału Sprawiedliwości z dnia 28 listopada 2024 r., *Ministerstvo na vatreshnite raboti*, C-80/23, ECLI:EU:C:2024:991.

[230] W przypadku systemów AI wysokiego ryzyka automatycznie generowane rejestry danych muszą być przechowywane przez co najmniej 6 miesięcy, a w przypadku systemów wysokiego ryzyka, o których mowa w załączniku III pkt 1 lit. a), rejestry te muszą zawierać datę i godzinę rozpoczęcia oraz datę i godzinę zakończenia każdego wykorzystania. Zob. art. 12 ust. 3 lit. a) i art. 19 aktu w sprawie sztucznej inteligencji.

402) Art. 5 ust. 3 aktu w sprawie sztucznej inteligencji stanowi ponadto, że w przypadku odmowy udzielenia zezwolenia w pilnych przypadkach należy wstrzymać wykorzystywanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym ze skutkiem natychmiastowym. W takich przypadkach wszystkie dane, w tym rezultaty i wyniki uzyskane podczas tego wykorzystania, muszą zostać natychmiast odrzucone i usunięte[231]. Art. 5 ust. 3 aktu w sprawie sztucznej inteligencji jest w tym względzie jednoznaczny, bez żadnych wyjątków. Podmiot stosujący musi posiadać:

a) referencyjną bazę danych zawierającą dane biometryczne (np. wizerunki twarzy, próbki głosu itp.) oraz, w stosownych przypadkach, powiązane dane identyfikacyjne, z którymi

b) porównuje się zebrane dane biometryczne osób obecnych w przestrzeni publicznej, aby te osoby zidentyfikować i wyodrębnić.

c) Porównanie takie prowadzi do wyniku porównania.

403) Wymóg odrzucenia i usunięcia zebranych i przetworzonych danych oznacza również, że referencyjne bazy danych, które zostały wykorzystane do nieuprawnionej identyfikacji biometrycznej, muszą zostać usunięte, jeśli utworzono je specjalnie na potrzeby zakwestionowanego poszukiwania. Baza danych może być utrzymywana wyłącznie w przypadku, gdy organy ścigania stworzyły i zamierzały utrzymywać bazę danych do identyfikacji *w sposób zgodny z prawem* do celów *innych* niż nieuprawnione korzystanie z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

404) Oprócz usunięcia wszelkich (niezgodnych z prawem) baz danych zawierających informacje biometryczne należy również usunąć wszystkie zebrane obrazy i inne dane osobowe, w tym metadane, dane dotyczące przetwarzania technicznego (w tym wzorce i inne dane osobowe) oraz wszelkie dane porównawcze i wyniki uzyskane podczas bezprawnego wykorzystania systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

405) W przypadku gdy organ ścigania odwoła się od odmowy udzielenia zezwolenia, dane mogą być przechowywane przez powiernika do czasu podjęcia ostatecznej decyzji w sprawie wniosku. W tym okresie dane te nie powinny być co do zasady udostępniane organowi ścigania[310].

10.2.2.5. Brak możliwości podejmowania decyzji wyłącznie na podstawie wyników uzyskanych z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym

406) Zgodnie z art. 5 ust. 3 aktu w sprawie sztucznej inteligencji nawet w przypadku, gdy podmiot stosujący system zdalnej identyfikacji biometrycznej w czasie rzeczywistym uzyska zezwolenie, nie można wydać decyzji wywołującej niepożądane skutki prawne dla danej osoby wyłącznie na podstawie wyników uzyskanych z takiego systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym.

---

[231] Organy nadzorcze powinny również posiadać uprawnienia do przeprowadzania takiej weryfikacji i kontroli post factum. Zob. art. 5 ust. 5 aktu w sprawie sztucznej inteligencji.

Na przykład:

– osoba zostaje zatrzymana i pozbawiona wolności za poważne przestępstwo wyłącznie na podstawie identyfikacji za pomocą systemu rozpoznawania twarzy, bez dalszych kontroli. Stanowi to uzupełnienie wymogu wynikającego z art. 14 aktu w sprawie sztucznej inteligencji dotyczącego nadzoru ze strony człowieka. Kontrole mogą dotyczyć na przykład tego, czy dana osoba przebywała w innym miejscu lub czy też istnieją inne powody, dla których osoba ta nie może być osobą poszukiwaną.

Wymogi art. 14 aktu w sprawie sztucznej inteligencji dotyczące nadzoru ze strony człowieka

407) Wykorzystywanie zdalnej identyfikacji biometrycznej w czasie rzeczywistym, które jest dozwolone, ponieważ służy realizacji jednego z celów wymienionych w art. 5 ust. 1 lit. h) i jest zgodne z art. 5 ust. 2–6 aktu w sprawie sztucznej inteligencji, nadal podlega przepisom dotyczącym systemów wysokiego ryzyka. Zgodnie z art. 14 aktu w sprawie sztucznej inteligencji systemy AI wysokiego ryzyka „projektuje się i rozwija w taki sposób, w tym poprzez uwzględnienie odpowiednich narzędzi interfejsu człowiek-maszyna, aby w okresie ich wykorzystywania systemy AI mogły być skutecznie nadzorowane przez osoby fizyczne". Zgodnie z art. 14 ust. 5 aktu w sprawie sztucznej inteligencji podmiot stosujący nie może podejmować żadnego działania ani decyzji na podstawie identyfikacji będącej wynikiem działania systemu, „jeżeli identyfikacji tej nie zweryfikowały ani nie potwierdziły odrębnie co najmniej dwie osoby fizyczne mające wymagane kompetencje, przeszkolenie i uprawnienia", chyba że „prawo Unii lub prawo krajowe uznaje stosowanie tego wymogu za nieproporcjonalne". W art. 4 aktu w sprawie sztucznej inteligencji przewidziano środki dotyczące kompetencji w zakresie AI wśród dostawców i użytkowników systemów AI, aby zapewnić „odpowiedni poziom kompetencji w zakresie AI wśród [...] personelu i innych osób zajmujących się działaniem i wykorzystaniem systemów AI", biorąc pod uwagę osoby, wobec których systemy te mają być wykorzystywane.

408) W kontekście ochrony danych, EROD stwierdziła, że aby nadzór ze strony człowieka był skuteczny, kluczowe znaczenie ma „umożliwienie danej osobie zrozumienia systemu [rozpoznawania twarzy] i jego ograniczeń, a także właściwej interpretacji jego wyników. Konieczne jest również ustanowienie miejsca pracy i organizacji, które będą przeciwdziałać skutkom błędu automatyzacji i pozwolą uniknąć bezkrytycznej akceptacji wyników, np. z powodu presji czasu, uciążliwych procedur, potencjalnie szkodliwego wpływu na karierę itp."[232]. Podobne kwestie mogą mieć zastosowanie w kontekście aktu w sprawie sztucznej inteligencji.

10.3. Powiadamianie organów o każdym wykorzystaniu systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym" w przestrzeni publicznej do celów ścigania przestępstw

Art. 5 ust. 4 aktu w sprawie sztucznej inteligencji stanowi:

[232] EROD, *Wytyczne 05/2022 w sprawie wykorzystania technologii rozpoznawania twarzy w obszarze ścigania przestępstw*, v2.0, 26 kwietnia 2023 r., s. 22.

Bez uszczerbku dla ust. 3, o każdym wykorzystaniu systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw powiadamia się właściwy organ nadzoru rynku i krajowy organ ochrony danych zgodnie z przepisami prawa krajowego, o których mowa w ust. 5. Powiadomienie zawiera co najmniej informacje określone w ust. 6 i nie może zawierać wrażliwych danych operacyjnych.

409) O każdym wykorzystaniu systemu zdalnej identyfikacji biometrycznej do jednego z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji należy powiadomić odpowiedni organ nadzoru rynku i krajowy organ ochrony danych. Powiadomienie musi nastąpić po każdym wykorzystaniu, aby organy te mogły uwzględnić w swoich sprawozdaniach liczbę zezwoleń i rezultat wykorzystania. Powiadomienie nie może zawierać wrażliwych danych operacyjnych. Zgodnie z art. 3 ust. 38 aktu w sprawie sztucznej inteligencji „wrażliwe dane operacyjne" oznaczają dane operacyjne związane z działaniami w zakresie ścigania przestępstw (zapobiegania przestępstwom, ich wykrywania, prowadzenia postępowań przygotowawczych w ich sprawie lub ich ścigania), których ujawnienie mogłoby zagrozić integralności postępowania karnego.

410) Szczegółowe informacje na temat wymogu sprawozdawczego znajdują się w sekcji 10.6 poniżej.

10.4. Konieczność istnienia przepisów krajowych w zakresie przewidzianym w akcie w sprawie sztucznej inteligencji

10.4.1. Zasada: podstawą prawną udzielenia zezwolenia na zastosowanie wszystkich lub niektórych wyjątków musi być prawo krajowe

411) Aby istniała możliwość wykorzystania systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym" w przestrzeni publicznej do celów ścigania przestępstw konieczne są przepisy krajowe. Jednocześnie art. 5 ust. 5 aktu w sprawie sztucznej inteligencji stanowi, że państwa członkowskie zachowują swobodę decydowania o przyjęciu takich przepisów krajowych. Jeżeli zostanie przyjęte prawo krajowe zezwalające na wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, akt w sprawie sztucznej inteligencji określa merytoryczne elementy, które muszą zostać zawarte w przepisach krajowych, aby spełniały one wymogi określone w tym akcie.

Art. 5 ust. 5 aktu w sprawie sztucznej inteligencji

Państwo członkowskie może podjąć decyzję o wprowadzeniu możliwości pełnego lub częściowego zezwolenia na wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w granicach i na warunkach wymienionych w ust. 1 akapit pierwszy lit. h) i w ust. 2 i 3. Zainteresowane państwa członkowskie ustanawiają w swoim prawie krajowym niezbędne szczegółowe przepisy dotyczące wniosku o udzielenie zezwoleń, o których mowa w ust. 3, wydawanie i wykonywanie tych zezwoleń oraz ich nadzorowanie i składanie sprawozdań w ich sprawie. W przepisach tych określa się

również, w odniesieniu do których celów wymienionych w ust. 1 akapit pierwszy lit. h) – w tym w odniesieniu do których przestępstw wymienionych w ust. 1 akapit pierwszy lit. h) ppkt (iii) – właściwe organy mogą uzyskać zezwolenie na wykorzystanie tych systemów do celów ścigania przestępstw. Państwa członkowskie powiadamiają Komisję o tych przepisach najpóźniej 30 dni po ich przyjęciu. Państwa członkowskie mogą wprowadzić, zgodnie z prawem Unii, bardziej restrykcyjne przepisy dotyczące wykorzystania systemów zdalnej identyfikacji biometrycznej.

10.4.2. Prawo krajowe musi być zgodne z ograniczeniami i warunkami określonymi w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji

412) Ze względu na fakt, że wykorzystywanie systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym" w przestrzeni publicznej do celów ścigania przestępstw uznaje się za ingerencję w prawa podstawowe, art. 5 ust. 5 aktu w sprawie sztucznej inteligencji stanowi, że takie wykorzystanie musi zostać ustanowione w prawie krajowym państw członkowskich. Prawo krajowe stanowi podstawę prawną stosowania takich systemów.

413) Przepisy krajowe nie mogą wykraczać poza ramy określone w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji i muszą być zgodne ze wszystkimi dalszymi powiązanymi warunkami określonymi w akcie w sprawie sztucznej inteligencji. Oznacza to, że państwa członkowskie nie mogą rozszerzać zakresu celów, do których system zdalnej identyfikacji biometrycznej w czasie rzeczywistym może być wykorzystywany w przestrzeni publicznej na potrzeby ścigania przestępstw, na cele inne niż te wymienione w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji[233].

414) Państwa członkowskie powiadamiają Komisję o swoich przepisach krajowych najpóźniej w terminie 30 dni od ich przyjęcia. Takie powiadomienie nie stanowi domniemania zgodności prawa państw członkowskich z aktem w sprawie sztucznej inteligencji. Po otrzymaniu powiadomienia Urząd ds. AI prześle potwierdzenie odbioru. Przed przyjęciem takich przepisów, państwa członkowskie zachęca się również do przesłania wstępnej wersji proponowanego prawa krajowego (lub regionalnego) do Urzędu ds. AI. W każdym razie brak powiadomienia Urzędu ds. AI w przewidzianym prawem terminie 30 dni po przyjęciu określonym w art. 5 ust. 5 może oznaczać, że prawo krajowe nie jest możliwe do wyegzekwowania w postępowaniu sądowym, co potwierdzono w różnych kontekstach[234]. Komisja opublikuje przepisy państw członkowskich na ogólnodostępnej stronie internetowej.

415) Państwa członkowskie mogą wprowadzić, zgodnie z prawem Unii, bardziej restrykcyjne przepisy, tj. przepisy zawierające bardziej rygorystyczne wymogi niż te określone w art. 5 ust. 1 lit. h) i art. 5 ust. 2–7 aktu w sprawie sztucznej inteligencji.

---

233 Zob. wyrok Trybunału Sprawiedliwości z dnia 5 kwietnia 2022 r., *Komisarz Garda Síochána*, C-140/20, ECLI:EU:C:2022:258, pkt 54: „Aby spełniać wymóg proporcjonalności, uregulowanie krajowe musi zawierać jasne i precyzyjne przepisy regulujące zakres i sposób stosowania rozpatrywanego środka oraz ustanawiające minimalne wymogi".

234 Zob. analogiczny wyrok Trybunału Sprawiedliwości z dnia 19 grudnia 2019 r., *Airbnb Ireland*, C-390/18, EU:C:2019:1112, pkt 96–97.

10.4.3. Szczegółowe przepisy krajowe dotyczące wniosku o udzielenie zezwolenia oraz wydawania i wykonywania tego zezwolenia

416) Jeśli chodzi o szczegółowe przepisy mające zastosowanie do wniosku o udzielenie zezwolenia oraz wydawania i wykonywania tego zezwolenia, należy je określić w prawie krajowym. Każde państwo członkowskie, które chce zezwolić na wykorzystywanie przedmiotowych systemów, musi określić w swoim prawie krajowym takie przepisy zapewniające przekazywanie organowi wydającemu zezwolenie istotnych i kompletnych informacji na temat wykorzystania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym, aby umożliwić mu podjęcie decyzji w przedmiocie bezwzględnej konieczności i proporcjonalności takiego wykorzystania.

Prawo krajowe zezwalające na korzystanie z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym może regulować na przykład:

– kwestię, które organy są właściwymi organami w odniesieniu do art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji oraz niezależnymi organami w państwie członkowskim właściwymi do wydawania (lub odmowy) zezwolenia;

– szczegółowy zakres celów, do których systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw mogą być wykorzystywane (niewykraczający poza cele wymienione w art. 5 ust. 1 lit. h) ppkt (i)–(iii), ale z możliwością dalszego zawężenia);

– wymóg, aby wnioski miały formę pisemną i zawierały szczegółowe wyjaśnienia dotyczące konkretnego wykorzystania i przeznaczenia w związku z konkretnym przestępstwem lub sytuacją, które uzasadniają jego wykorzystanie;

– wymóg umotywowania i przedstawienia dowodów potwierdzających (w stosownych przypadkach wraz z tłumaczeniem) na potrzeby uzasadnienia wykorzystania systemu do celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji, w szczególności w odniesieniu do miejsca, okresu i zakresu podmiotowego oraz bezwzględnej konieczności i proporcjonalności, w tym adekwatności, wystarczalności i skuteczności wykorzystania systemu oraz braku mniej inwazyjnych środków;

– opis technologii, która ma być wykorzystana, oraz lokalizację punktów gromadzenia danych;

– minimalną niezawodność, zastosowany próg i wskaźniki dokładności wykorzystywanych systemów;

– możliwość przeprowadzenia przez organ udzielający zezwolenia audytu przedłożonych informacji, w tym szczegółów technicznych i kryteriów dokładności, w dowolnym momencie *ex ante* i *ex post*;

– specyfikację wykorzystywanych referencyjnych baz danych;

– okres przechowywania zebranych danych i wszystkich innych wykorzystanych danych osobowych;

– środki bezpieczeństwa, w tym zabezpieczające przed bezprawnym dostępem do danych;

– inne zabezpieczenia (w stosownych przypadkach);

– opis wszelkiej współpracy z organami prywatnymi lub publicznymi, w tym w innych państwach, oraz przekazywania i wymiany danych;

– identyfikowalność procesu;

– imiona i nazwiska odpowiedzialnych członków personelu podmiotu stosującego;

W odniesieniu do *innych elementów formalnych* dotyczących wydawania zezwolenia

– możliwość zastosowania procedury pisemnej uzupełnionej o wysłuchanie;

– podstawy odmowy;

– prawa osób, które są poszukiwane, prawa osób, których dane są pozyskiwane, oraz ewentualne prawa osób trzecich[235];

– terminy, w których organy będą musiały podjąć decyzję;

– konieczność formalnego powiadomienia w przypadku udzielenia zezwolenia lub odmowy udzielenia zezwolenia;

– kary za niespełnienie (formalnych i istotnych) wymogów;

– prawo do odwołania się od odmowy udzielenia zezwolenia;

W odniesieniu do wykonywania zezwolenia

– rejestrację korzystania z systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w centralnym rejestrze wraz z podsumowaniem elementów merytorycznych;

– ewentualne dalsze obowiązki sprawozdawcze;

– procedurę rozszerzenia lub zmiany zezwolenia.

10.4.4. Szczegółowe przepisy krajowe dotyczące nadzoru i sprawozdawczości w odniesieniu do zezwolenia

417) W art. 70 aktu w sprawie sztucznej inteligencji zobowiązano państwa członkowskie do ustanowienia co najmniej jednego organu notyfikującego i co najmniej jednego organu nadzoru rynku. Art. 74 ust. 8 aktu w sprawie sztucznej inteligencji stanowi, że „państwa członkowskie wyznaczają jako organy nadzoru rynku do celów niniejszego rozporządzenia właściwe organy nadzorcze ds. ochrony danych na podstawie rozporządzenia (UE) 2016/679 lub dyrektywy (UE) 2016/680 albo inne organy

[235] Zob. np. EROD, *Wytyczne 05/2022 w sprawie wykorzystania technologii rozpoznawania twarzy w obszarze egzekwowania prawa*, v2.0, 26 kwietnia 2023 r., s. 24 i nast.

wyznaczone zgodnie z tymi samymi warunkami ustanowionymi w art. 41–44 dyrektywy (UE) 2016/680”.

418) Organy te należy wyznaczyć obok organu wydającego zezwolenie, który państwo członkowskie będzie musiało ustanowić, zanim będzie mogło zezwolić na wykorzystanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do któregokolwiek z celów wymienionych w art. 5 ust. 1 lit. h) ppkt (i)–(iii) aktu w sprawie sztucznej inteligencji.

10.5. Roczne sprawozdania krajowych organów nadzoru rynku i krajowych organów ochrony danych państw członkowskich

Art. 5 ust. 6 aktu w sprawie sztucznej inteligencji

Krajowe organy nadzoru rynku i krajowe organy ochrony danych państw członkowskich, które zostały powiadomione o wykorzystaniu systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw zgodnie z ust. 4, przedkładają Komisji roczne sprawozdania z takiego wykorzystania. W tym celu Komisja przekazuje państwom członkowskim i krajowym organom nadzoru rynku i organom ochrony danych wzór formularza zawierającego informacje na temat liczby decyzji podjętych przez właściwe organy wymiaru sprawiedliwości lub wydający wiążące decyzje niezależny organ administracyjny w odniesieniu do wniosków o udzielenie zezwolenia zgodnie z ust. 3 oraz wyników ich rozpatrzenia.

419) Krajowe organy nadzoru rynku i krajowe organy ochrony danych państw członkowskich, które zostały poinformowane przez podmioty stosujące o wykorzystaniu systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw (zob. art. 5 ust. 4), muszą przedkładać Komisji roczne sprawozdania na temat takiego wykorzystania. Sprawozdania te sporządza się na podstawie wzoru dostarczonego przez Komisję. Wzór ten zostanie opracowany w odpowiednim czasie.

420) W przypadku gdy podmiot stosujący jest instytucją, organem lub agencją UE, EIOD jest zobowiązany do corocznego informowania Komisji o systemach zdalnej identyfikacji biometrycznej w czasie rzeczywistym wykorzystywanych w przestrzeni publicznej do celów ścigania przestępstw.

421) Okres od 2 lutego 2025 r. do 2 sierpnia 2025 r. będzie objęty wyłącznie sprawozdaniem krajowego organu ochrony danych, ponieważ akt w sprawie sztucznej inteligencji nie wymaga od państw członkowskich wyznaczenia krajowego organu nadzoru rynku przed tą ostatnią datą.

422) Krajowe organy nadzoru rynku i krajowe organy ochrony danych mają swobodę decydowania, czy chcą składać sprawozdania indywidualnie czy jedno wspólne sprawozdanie dla państwa członkowskiego.

10.6. Sprawozdania roczne Komisji

Art. 5 ust. 7 aktu w sprawie sztucznej inteligencji

Komisja publikuje roczne sprawozdania na temat wykorzystania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw, oparte na zagregowanych danych w państwach członkowskich przekazanych w sprawozdaniach rocznych, o których mowa w ust. 6. Te sprawozdania roczne nie mogą zawierać wrażliwych danych operacyjnych dotyczących powiązanych działań w zakresie ścigania przestępstw.

423) W akcie w sprawie sztucznej inteligencji zobowiązano Komisję do publikowania rocznych sprawozdań na temat wykorzystania systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów ścigania przestępstw w państwach członkowskich oraz przez instytucje, agencje i organy Unii, w oparciu o zagregowane dane. Sprawozdania te mają opierać się na informacjach przekazanych przez organy krajowe zgodnie z art. 5 ust. 6 aktu w sprawie sztucznej inteligencji.

424) Sprawozdanie roczne Komisji nie może zawierać wrażliwych danych operacyjnych. „Wrażliwe dane operacyjne" oznaczają dane operacyjne związane z działaniami w zakresie zapobiegania przestępstwom, ich wykrywania, prowadzenia postępowań przygotowawczych w ich sprawie lub ich ścigania, których ujawnienie mogłoby zagrozić integralności postępowania karnego"[236]. Może to oznaczać, że nie można publikować konkretnych szczegółów ujawniających toczące się lub wcześniejsze dochodzenia, takie jak np. lokalizacje, wykorzystane kamery.

10.7. Systemy nieobjęte zakresem

425) Systemy zdalnej identyfikacji biometrycznej, które są wykorzystywane do wszystkich innych celów nieobjętych zakazem ustanowionym w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej, należą do kategorii systemów AI wysokiego ryzyka zdefiniowanych w art. 6 i wymienionych w pkt 1 lit. a) załącznika III do aktu w sprawie sztucznej inteligencji, pod warunkiem że wchodzą one w zakres aktu w sprawie sztucznej.

426) Systemy zdalnej identyfikacji biometrycznej, które nie wchodzą w zakres zakazu określonego w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji, obejmują systemy weryfikacji biometrycznej/uwierzytelnienia biometrycznego oraz systemy zdalnej identyfikacji biometrycznej w przestrzeni publicznej wykorzystywane retrospektywnie do celów ścigania przestępstw. Na przykład organy policji mogą być upoważnione na mocy prawa krajowego do retrospektywnego rozpoznawania twarzy w celu porównywania wizerunków podejrzanych o popełnienie przestępstwa z zarejestrowanymi wizerunkami twarzy w bazie danych dotyczących przestępstw[237]. Innym wykorzystaniem, które nie wchodzi w zakres zakazu, jest wykorzystywanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów ścigania przestępstw w przestrzeni prywatnej (np. w czyimś domu) albo internetowej (np. korzystanie z czatu lub gry internetowej w celu identyfikacji podejrzanego o rozpowszechnianie materiałów przedstawiających niegodziwe traktowanie dzieci

[236] Art. 3 pkt 38 aktu w sprawie sztucznej inteligencji.

[237] Na przykład baza danych *Traitement des antécédents Judiciaires* we Francji, utworzona na mocy *Décret nr 2012-652 du 4 mai 2012 relatif au Traitement des antécédents Judiciaires* (dekret 2012-652).

w celach seksualnych). Ponadto stosowanie systemów zdalnej identyfikacji biometrycznej przez podmioty prywatne, zarówno w czasie rzeczywistym, jak i retrospektywne (takie jak wykorzystanie przez supermarket technologii rozpoznawania twarzy na żywo w celu identyfikacji znanych złodziei sklepowych, wykorzystanie technologii rozpoznawania twarzy na żywo przez arenę sportową w celu identyfikacji osób z zakazem wstępu czy stosowanie technologii rozpoznawania twarzy na żywo w szkołach do celów bezpieczeństwa i monitorowania frekwencji) nie wchodzi w zakres zakazu.

427) Oprócz przepisów, które mają ogólnie zastosowanie do systemów AI wysokiego ryzyka, **retrospektywne wykorzystywanie systemów zdalnej identyfikacji biometrycznej** do celów ścigania przestępstw podlega dodatkowym warunkom i zabezpieczeniom zgodnie z art. 26 ust. 10 aktu w sprawie sztucznej inteligencji (mającym zastosowanie od 2 sierpnia 2026 r.)[238].

428) Wykorzystanie **do celów innych niż ściganie przestępstw** musi być w każdym przypadku zgodne z **przepisami o ochronie danych**. Poniższe przypadki ilustrują wykładnię art. 9 ust. 2 RODO w odniesieniu do takiego wykorzystania oraz wyjątki od przetwarzania danych biometrycznych.

Na przykład:

– francuski sąd administracyjny uznał, że testowanie technologii rozpoznawania twarzy na żywo w dwóch szkołach publicznych do celów kontroli dostępu i bezpieczeństwa nie było ani konieczne, ani proporcjonalne (zgodnie z przepisami o ochronie danych). Dostępne były alternatywne rozwiązania, które były mniej inwazyjne dla uczniów, np. stosowanie identyfikatorów. Ponadto nie spełniono warunku uzyskania wyraźnej zgody. W związku z tym nie można było wykorzystać zgody jako ważnej podstawy prawnej do testowania technologii rozpoznawania twarzy w szkołach średnich[239];

– supermarket w Niderlandach nie miał prawa korzystać z technologii rozpoznawania twarzy na żywo, aby zapobiegać kradzieży w sklepie. Bez wyraźnej zgody klienta lub jakiejkolwiek podstawy prawnej umożliwiającej przetwarzanie danych w istotnym interesie publicznym (np. do celów bezpieczeństwa) supermarket nie miał prawa przetwarzać danych biometrycznych, a tym samym stosować technologii rozpoznawania twarzy[240];

– we Francji zakazano korzystania z technologii rozpoznawania twarzy na żywo przy wejściu do klubu piłkarskiego w celu identyfikacji kibiców[241], a w Hiszpanii w celu zapewnienia bezpieczeństwa fanom[242].

10.8. Przykłady wykorzystania

238 Art. 26 ust. 10 oraz motyw 94 aktu w sprawie sztucznej inteligencji.

239 TA Marseille (sąd administracyjny w Marsylii), 27 lutego 2020 r., nr 1901249.

240 https://www.autoriteitpersoonsgegevens.nl/en/current/dutch-dpa-issues-formal-warning-to-supermarket-for-use-of-facial-recognition-technology.

241 https://www.cnil.fr/fr/reconnaissance-faciale-et-interdiction-commerciale-de-stade-la-cnil-adresse-un-avertissement-un-club.

242 https://www.biometricupdate.com/202401/spanish-data-authority-opposes-facial-recognition-for-football-stadium-access.

Policja instaluje mobilne kamery telewizji przemysłowej wyposażone w oparte na sztucznej inteligencji technologie rozpoznawania twarzy na radiowozie umieszczonym w pobliżu głównego wejścia na stadion piłkarski podczas meczu Mistrzostw Europy w celu zabezpieczenia terenu oraz identyfikacji osób, których twarze znajdują się w doraźnej bazie danych dotyczących osób znajdujących się na liście ostrzegawczej. Wspomniana lista ostrzegawcza obejmuje osoby podejrzane o popełnienie przestępstwa (od poważnych przestępstw po nadużycia finansowe i włamania), osoby będące potencjalnym obiektem zainteresowania służb wywiadowczych oraz osoby wymagające szczególnego traktowania ze względu na problemy psychiczne Korzystanie przez policję z technologii rozpoznawania twarzy na żywo nie jest powiązane z informacjami dotyczącymi obecności konkretnej osoby podczas wydarzenia. Chociaż na liście ostrzegawczej prawdopodobnie znajdują się osoby, wobec których zastosowanie systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym byłoby dopuszczalne, lista ta jest zbyt mało konkretna i nie jest powiązana z wydarzeniem, jakim jest mecz piłki nożnej. Takie zastosowanie systemu byłoby zatem zakazane.

System identyfikacji biometrycznej (niezdalnej) sprawdza, czy dane osoby są uprawnione do wstępu na teren elektrowni jądrowej. Kiedy ludzie ustawiają się przed (widoczną) kamerą, a system odmawia dostępu, system próbuje ustalić, czy dana osoba znajduje się na liście ostrzegawczej terrorystów. System nie działa zdalnie. Osoby aktywnie uczestniczyły w weryfikacji w celu uzyskania dostępu do elektrowni. Ten przypadek wykorzystania nie jest objęty zakazem określonym w art. 5 aktu w sprawie sztucznej inteligencji.

Organy policji zatłoczonego miasta stosują wspomagane AI kamery telewizji przemysłowej, które mogą wykorzystywać technologie rozpoznawania twarzy na żywo. Oprócz rozpoznawania twarzy możliwe jest wykorzystanie innych dodatkowych funkcji, takich jak wykrywanie przedmiotów i ruchu tłumu.

Kamery są instalowane w wielu lokalizacjach, w tym w miejscach kultu, w miejscach odwiedzanych przez społeczność LGBT+, w gabinetach lekarskich, aptekach oraz w różnych restauracjach i barach. Samo instalowanie kamer przystosowanych do pomiarów biometrycznych nie jest zakazane na mocy aktu w sprawie sztucznej inteligencji. Niektóre zastosowania, w tym nieokreślona i niezróżnicowana identyfikacja osób fizycznych, są jednak zakazane.

W okresie wakacyjnym w dzielnicy mieszkaniowej doszło do szeregu włamań. Policja uzyskuje opis podejrzanego od naocznego świadka, który kilkukrotnie widział podejrzanego na terenie dzielnicy mieszkaniowej przed włamaniami. Aby zidentyfikować i zatrzymać podejrzanego, policja wykorzystuje technologię rozpoznawania twarzy na żywo w różnych miejscach na terenie dzielnicy mieszkaniowej podczas weekendu. Na podstawie wskazań świadków policja

sporządziła portret pamięciowy podejrzanego i wyodrębniła z bazy danych osób zatrzymanych kilka zdjęć osób przypominających ten portretu.
Nawet jeżeli policja wykorzystuje technologię rozpoznawania twarzy na żywo wobec określonego podejrzanego i ustaliła obszar oraz czas jej użycia, nie jest dozwolone jej zastosowanie w przypadku przestępstwa, które nie jest wymienione w załączniku II do aktu w sprawie sztucznej inteligencji.

Policja monitoruje emocje kibiców na stadionie piłkarskim za pomocą systemu rozpoznawania biometrycznego. System wykrywa potencjalną agresję i natychmiast włącza w tej części stadionu system zdalnej identyfikacji biometrycznej w czasie rzeczywistym, aby zidentyfikować chuliganów, którzy zachowywali się agresywnie w przeszłości.
Kontrola emocji na stadionie nie jest zakazana na mocy aktu w sprawie sztucznej inteligencji (ale systemy takie należą do kategorii systemów AI wysokiego ryzyka określonych w akcie w sprawie sztucznej inteligencji). Wykorzystanie do tego celu systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym byłoby jednak zakazane na mocy aktu w sprawie sztucznej inteligencji, w szczególności gdy to system biometryczny decyduje o konieczności identyfikacji osób do celów ścigania przestępstw.

Policja wykorzystuje sieć telewizji przemysłowej zainstalowaną w mieście i metrze, aby zidentyfikować demonstranta politycznego, który zorganizował zbiorowy protest na ulicach. W odnośnym państwie członkowskim organizatorzy protestów zbiorowych odbywających się na drogach publicznych i w przestrzeniach publicznych, takich jak ulice, muszą powiadomić władze miejskie o planowanej demonstracji na trzy dni przed jej przeprowadzeniem, aby zapobiec zakłóceniom porządku publicznego i przemocy. Brak powiadomienia stanowi przestępstwo zagrożone karą do sześciu miesięcy pozbawienia wolności i maksymalną grzywną w wysokości 8 000 EUR. Aby zidentyfikować protestującego, policja pobiera nagrania z kamer telewizji przemysłowej zainstalowanych na ulicach i przeprowadza retrospektywne rozpoznawanie twarzy, porównując pobrane zdjęcia z fotografiami opublikowanymi w mediach społecznościowych.

Retrospektywne wykorzystanie technologii rozpoznawania twarzy nie jest zakazane przez akt w sprawie sztucznej inteligencji. Wykorzystanie to uznaje się za wykorzystanie wysokiego ryzyka i powinno ono być zgodne z wymogami określonymi w akcie w sprawie sztucznej inteligencji w odniesieniu do takich systemów[243].

[243] Przetwarzanie danych biometrycznych do celów ścigania przestępstw nadal podlega przepisom art. 10 dyrektywy o ochronie danych w sprawach karnych, które należy wdrożyć na szczeblu krajowym. Przetwarzanie takich danych na potrzeby retrospektywnego rozpoznawania twarzy powinno być dozwolone tylko wtedy, gdy jest to bezwzględnie konieczne, i powinno podlegać odpowiednim zabezpieczeniom. Kwestia, czy retrospektywne wykorzystanie technologii rozpoznawania twarzy jest bezwzględnie konieczne do zidentyfikowania demonstranta, budzi wątpliwości. W wyroku w sprawie *Glukhin* przeciwko *Rosji*, który służy za podstawę tego scenariusza, ETPC orzekł, że chociaż wykrywanie przestępstw może stanowić uzasadniony cel, wykorzystanie technologii rozpoznawania twarzy, zarówno retrospektywnego, jak i w czasie rzeczywistym, było nieproporcjonalne, ponieważ nie istniało zagrożenie dla porządku publicznego ani bezpieczeństwa transportu. Trybunał podkreślił „wysoce inwazyjny" charakter technologii

Dalsze przykłady praktyk, które NIE są zakazane:

- hotele korzystające z systemu zdalnej identyfikacji biometrycznej w czasie rzeczywistym do rozpoznawania gości VIP. Nie odbywa się to na potrzeby ścigania przestępstw;
- centra handlowe wykorzystujące systemy zdalnej identyfikacji biometrycznej w czasie rzeczywistym do wykrywania złodziei. Nie odbywa się to na potrzeby ścigania przestępstw.

Przykład praktyki zakazanej:

Na zlecenie policji centrum handlowe wykorzystuje zdalną identyfikację biometryczną w czasie rzeczywistym do wykrywania złodziei. System jest stosowany do celów ścigania przestępstw w przestrzeni publicznej. Wykorzystane jest zakazane, ponieważ poszukiwanie złodziei sklepowych nie jest objęte żadnym z wyjątków przewidzianych w art. 5 ust. 1 lit. h) aktu w sprawie sztucznej inteligencji.

11. ROZPOCZĘCIE STOSOWANIA

429) Zgodnie z art. 113 aktu w sprawie sztucznej inteligencji, art. 5 aktu w sprawie sztucznej inteligencji ma zastosowanie od dnia 2 lutego 2025 r. Zakazy zawarte w tym przepisie będą miały co do zasady zastosowanie do wszystkich systemów AI, niezależnie od tego, czy zostały one wprowadzone do obrotu bądź oddane do użytku przed tą datą czy po tej dacie[244].

430) Jednocześnie rozdziały dotyczące zarządzania i kar zaczną obowiązywać 2 sierpnia 2025 r. W związku z tym przepisy dotyczące kar za nieprzestrzeganie zakazów określonych w art. 5 aktu w sprawie sztucznej inteligencji nie będą miały zastosowania przed dniem 2 sierpnia 2025 r. W tym okresie przejściowym nie będzie również organów nadzoru rynku, które monitorowałyby, czy zakazy są właściwie przestrzegane.

431) Jednak nawet w tym okresie przejściowym zakazy mają pełne zastosowanie i są obowiązkowe dla dostawców systemów AI i podmiotów je stosujących. Operatorzy ci powinni zatem podjąć niezbędne środki w celu zapewnienia, że nie wprowadzają do obrotu, nie oddają do użytku ani nie wykorzystują systemów AI, które mogłyby być zakazane na mocy art. 5 aktu w sprawie sztucznej inteligencji. Nawet jeśli przepisy dotyczące monitorowania i kar pieniężnych będą miały zastosowanie dopiero w późniejszym terminie, same zakazy mają bezpośredni skutek, a tym samym umożliwiają zainteresowanym stronom egzekwowanie prawa przed sądami krajowymi i zwrócenie się o wydanie tymczasowych zakazów stosowania zakazanych praktyk.

12. PRZEGLĄD I AKTUALIZACJA WYTYCZNYCH KOMISJI

432) Niniejsze wytyczne stanowią pierwszą interpretację zawierającą praktyczne przykłady dotyczące zakazów określonych w art. 5 aktu w sprawie sztucznej inteligencji. Komisja

rozpoznawania twarzy. W sprawie tej Trybunał stwierdził, że korzystanie z technologii rozpoznawania twarzy nie odpowiadało pilnej potrzebie społecznej ani nie było konieczne w społeczeństwie demokratycznym.

244 Zob. art. 111 ust. 1 i 2 aktu w sprawie sztucznej inteligencji, w którym określono, że zasada praw nabytych pozostaje bez uszczerbku dla stosowania art. 5 aktu w sprawie sztucznej inteligencji zgodnie z art. 113 ust. 3 lit. a) aktu w sprawie sztucznej inteligencji.

zapewni dodatkowe wsparcie operatorom i organom w zakresie interpretacji zakazów oraz będzie na bieżąco gromadzić kolejne praktyczne przypadki wykorzystania, uwzględniając informacje otrzymane od dostawców systemów AI i podmiotów je stosujących, Rady ds. Sztucznej Inteligencji oraz innych odpowiednich zainteresowanych stron.

433) Komisja dokona przeglądu niniejszych wytycznych, gdy tylko będzie to konieczne, biorąc pod uwagę praktyczne doświadczenia zdobyte podczas wdrażania zakazów oraz tempo zmian technologicznych, społecznych i regulacyjnych w tej dziedzinie. Obejmuje to również wszelkie istotne doświadczenia wynikające z działań w zakresie egzekwowania przepisów dotyczących nadzoru rynku oraz interpretacje TSUE dotyczące zakazów i innych przepisów aktu w sprawie sztucznej inteligencji, analizowanych w niniejszych wytycznych. Podczas takiego przeglądu Komisja może podjąć decyzję o wycofaniu lub zmianie niniejszych wytycznych. Komisja zachęca dostawców systemów AI i podmioty je stosujące, krajowe organy nadzoru rynku za pośrednictwem Europejskiej Rady ds. Sztucznej Inteligencji, forum doradcze ds. sztucznej inteligencji, społeczność naukową i organizacje społeczeństwa obywatelskiego do wniesienia wkładu w ten proces poprzez udzielenie odpowiedzi na przyszłe zaproszenia do konsultacji publicznych.

ODPIS UWIERZYTELNIONY
W imieniu Sekretarza Generalnego

Martine DEPREZ
Dyrektor
Proces Decyzyjny i Kolegialność
KOMISJA EUROPEJSKA