

Systemy rekomendacji a AI Act

– omówienie zakazanych praktyk z art. 5.

Analiza opiera się na AI Act, czyli na unijnym rozporządzeniu UE 2024/1689, oraz na wytycznych interpretacyjnych wydanych przez Komisję Europejską, które dotyczą zakazanych praktyk opisanych w art. 5.

Kluczowe jest zrozumienie, od kiedy konkretne przepisy obowiązują. AI Act jako całość zaczyna być stosowany od 2 sierpnia 2026 roku, ale niektóre jego części weszły w życie wcześniej. Dotyczy to przede wszystkim przepisów ogólnych oraz zakazów, w tym zakazanych praktyk z art. 5, które obowiązują już od 2 lutego 2025 roku. Wytyczne Komisji Europejskiej potwierdzają, że zakazy z art. 5 rzeczywiście zaczęły obowiązywać od 2 lutego 2025 roku. Jednocześnie podkreślają, że same wytyczne nie są prawnie wiążące. Oznacza to, że mają one charakter pomocniczy i służą jako wskazówka, jak rozumieć przepisy. Ostateczna, wiążąca interpretacja prawa należy do Trybunału Sprawiedliwości Unii Europejskiej. Innymi słowy, to Trybunał ma ostatnie słowo w sprawie tego, jak należy rozumieć i stosować przepisy AI Act.

O czym mówi AI Act?

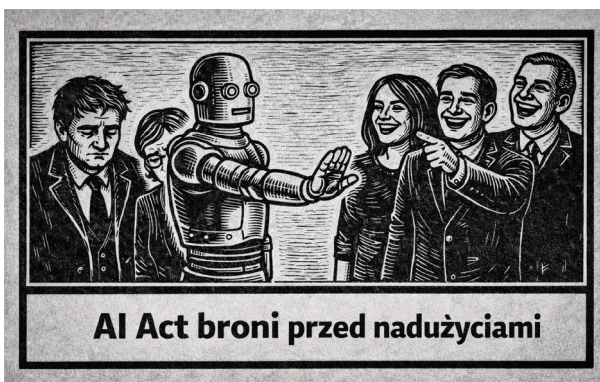
Po pierwsze – są rzeczy, których nie wolno robić, czyli zakazy z art. 5. Po drugie – są systemy wysokiego ryzyka, które można stosować, ale tylko pod warunkiem spełnienia określonych wymagań, takich jak zarządzanie ryzykiem lub kontrola stronniczości danych. Po trzecie – AI Act nie zastępuje innych przepisów Unii Europejskiej, więc nadal obowiązują m.in. przepisy dotyczące ochrony konsumenta oraz prawa antydyskryminacyjnego.

AI Act określa również, gdzie pojawia się problem dyskryminacji. Najlepszym przykładem jest zakaz tzw. social scoring, czyli punktowania ludzi.

O co chodzi z tą dyskryminacją?

Zgodnie z art. 5(1)(c), AI Act zabrania systemów, które oceniają lub klasyfikują

osoby albo grupy osób w czasie, czyli na podstawie historii ich zachowań. Taka ocena może opierać się zarówno na zachowaniach społecznych, jak i na cechach, które są znane, wywnioskowane lub przewidywane, np. dotyczących osobowości czy stylu życia. Taki system jest



zakazany wtedy, gdy prowadzi do gorszego traktowania tych osób. Dzieje się tak w dwóch sytuacjach. Pierwsza ma miejsce wtedy, gdy dane są wykorzystywane w innym kontekście niż ten, w którym zostały zebrane, np. gdy dane z zakupów są używane do oceny dostępu do innej usługi. Druga – występuje wtedy, gdy reakcja systemu jest nieuzasadniona albo nieproporcjonalna, czyli jest zbyt surowa w stosunku do zachowania lub nie ma racjonalnego uzasadnienia.

Wytyczne Komisji Europejskiej doprecyzowują, że zakaz social scoring jest szeroki i obejmuje zarówno sektor publiczny, jak i prywatny, bez ograniczenia do konkretnej branży. Jednocześnie podkreślają, że nie każda forma punktacji klientów jest automatycznie zakazana. Oznacza to, że można stosować social scoring klientów, o ile nie spełnia on opisanych wcześniej warunków prowadzących do niedozwolonego traktowania.

Kluczowa rzecz jest taka, że wytyczne wskazują, iż „cechy osobiste” mogą obejmować m.in. płeć, rasę, adres, dochód,

sytuację ekonomiczną, zachowanie oraz lokalizację lub sposób poruszania się. To oznacza, że systemy scoringowe oparte na takich cechach mogą podlegać analizie w kontekście art. 5(1)(c). Jednocześnie sam fakt wykorzystania tych cech nie jest jeszcze wystarczający, aby uznać system za zakazany. Nadal konieczne jest wykazanie, że prowadzi to do niekorzystnego traktowania w jednej z dwóch sytuacji, czyli albo poza kontekstem, w którym dane zostały zebrane, albo w sposób nieuzasadniony lub nieproporcjonalny.

To dość enigmatyczne i trudno z tego wyciągnąć jakieś praktyczne wnioski. Np. mamy do czynienia z przedsiębiorstwem, które analizuje swoich klientów nie tylko na podstawie historii zakupów i preferencji, lecz także na podstawie transkrypcji rozmów, z których wnioskuje ich cechy psychologiczne, takie jak skłonność do ryzyka czy sposób podejmowania decyzji. Na tej podstawie firma tworzy długoterminowy profil klientów i klasyfikuje ich pod kątem wartości biznesowej, a następnie różnicuje ofertę, przy czym część klientów zostaje wykluczona z bardziej atrakcyjnych warunków sprzedaży. Istotnym elementem jest również fakt, że taka analiza osobowości odbywa się bez wiedzy tych osób.

Z perspektywy AI Act, a konkretnie art. 5 ust. 1 lit. c, taka praktyka bardzo wyraźnie zbliża się do zakazanego social scoringu. System dokonuje bowiem oceny osób w czasie, opierając się nie tylko na ich zachowaniach, ale również na cechach wywnioskowanych, w tym cechach psychologicznych. Następnie ta ocena prowadzi do rzeczywistego zróżnicowania traktowania klientów, ponieważ część z nich zostaje pozbawiona dostępu do korzystniejszych

ofert. Taki skutek należy uznać za gorsze traktowanie w rozumieniu regulacji.

Dodatkowo występuje tu wykorzystanie danych w innym kontekście niż pierwotny, ponieważ rozmowy prowadzone w celu obsługi klienta zostają użyte do budowy profilu psychologicznego i podejmowania decyzji biznesowych o charakterze selekcyjnym. Jednocześnie reakcja systemu wydaje się nieproporcjonalna, gdyż cechy takie jak ostrożność finansowa czy styl podejmowania decyzji prowadzą do wykluczenia z oferty, co trudno uzasadnić obiektywnie i racjonalnie w kontekście samej sprzedaży produktu.

W konsekwencji należy uznać, że taki system bardzo prawdopodobnie narusza zakaz określony w art. 5 ust. 1 lit. c AI Act. Nie jest to już zwykła personalizacja oferty, lecz profilowanie o charakterze psychologicznym, które prowadzi do nieuzasadnionego i nieproporcjonalnego różnicowania klientów.

Dyskryminacja rasowa?

Drugi ważny zakaz, który ma znaczenie przy takich cechach jak kolor skóry czy kolor oczu, dotyczy biometrycznej kategoryzacji i wynika z art. 5(1)(g). Przepis ten zabrania używania systemów, które na podstawie danych biometrycznych przypisują ludzi do określonych kategorii w celu wnioskowania np. o ich rasie lub innych wrażliwych cechach.

AI Act wyjaśnia, czym są dane biometryczne. To dane osobowe powstające w wyniku specjalnego, technicznego przetwarzania cech fizycznych, fizjologicznych lub zachowań człowieka, takich jak np. wizerunek twarzy. W przepisach doprecyzowano również, czym jest biometryczna kategoryzacja. Oznacza ona przypisywanie osoby do określonej kategorii właśnie na podstawie danych biometrycznych. Jako przykłady takich kategorii podaje się m.in. płeć, wiek, kolor włosów czy kolor oczu.

W praktyce oznacza to, że jeżeli system rekomendacji potrafi „wyciągnąć” rasę, pochodzenie lub inne wrażliwe cechy na podstawie obrazu twarzy albo innego sygnału biometrycznego i wykorzystuje to do profilowania, to ryzyko



wejścia w zakres art. 5(1)(g) jest zupełnie innego rodzaju niż w przypadku zwykłej segmentacji marketingowej opartej na danych podanych przez użytkownika. Jeżeli natomiast system jest uznany za system wysokiego ryzyka, to AI Act nie ogranicza się do prostego zakazu dyskryminacji jednym zdaniem, lecz wprowadza konkretne obowiązki dotyczące jego projektowania i działania. Obejmują one m.in. obowiązek zarządzania ryzykiem zgodnie z art. 9 oraz wymagania dotyczące jakości i kontroli danych zgodnie z art. 10.

Szukanie dyskryminacji

W art. 10 AI Act wprost wskazuje się, że trzeba badać czy w danych i modelu występują stronniczości, które mogą prowadzić do dyskryminacji zakazanej w prawie Unii Europejskiej. Jednocześnie przepis ten nakłada obowiązek wdrożenia konkretnych działań, które pozwalają takie stronniczości wykrywać, zapobiegać im oraz ograniczać ich skutki, jeśli już się pojawiają.

Jeżeli chodzi o sytuację, w której system rekomendacji faworyzuje klientów na podstawie cech stałych, takich jak płeć, rasa lub kolor skóry, wiek czy miejsce zamieszkania, a innych eliminuje albo utrudnia im dostęp do benefitów, to wymaga to doprecyzowania.

Istnieją konkretne przepisy, pod które takie zachowanie może w pewnych sytuacjach podpadać, ale zawsze trzeba je ocenić w kontekście tych przepisów, a nie jednego ogólnego punktu.

Jak identyfikować złamanie zasad AI Act?

Poniżej przedstawiony jest „werdykt warunkowy” w czterech wariantach.

W pierwszym wariantcie, w którym hipoteza najczęściej jest fałszywa, chodzi

o typowy marketing i e-commerce. Jeżeli mamy do czynienia z klasycznym sklepem internetowym lub programem lojalnościowym, gdzie system przewiduje atrakcyjność klienta i na tej podstawie przyznaje zniżki lub graty, to nie jest to automatycznie zakazana praktyka w rozumieniu art. 5(1)(c). Wytyczne Komisji Europejskiej podają wprost przykłady takich sytuacji, które

nie podpadają pod ten przepis. Przykładowo platforma zakupowa może oferować dodatkowe przywileje użytkownikom, którzy mają silną historię zakupów i rzadko dokonują zwrotów, podczas gdy inni użytkownicy nadal mają dostęp do standardowej procedury zwrotu. W takim przypadku uznaje się, że korzyści są uzasadnione i proporcjonalne jako forma nagrody za określone zachowanie.

Podobnie wskazuje się, że reklama komercyjna wspierana przez AI, czyli tak zwana *AI-enabled targeted commercial advertising*, również nie wchodzi w zakres zakazu, o ile opiera się na istotnych danych, takich jak preferencje użytkownika, jest zgodna z innymi przepisami prawa Unii Europejskiej, np. dotyczącymi ochrony konsumenta, danych osobowych czy usług cyfrowych, oraz nie prowadzi do nieproporcjonalnie niekorzystnego traktowania użytkowników.

W takim ujęciu sama logika polegająca na tym, że rekomendujemy promocje tym klientom, którzy mają wyższą skłonność do zakupu, nie jest jeszcze zakazana przez AI Act.

W drugim wariantcie, w którym hipoteza może być prawdziwa, sytuacja zmienia się wtedy, gdy mamy do czynienia z tzw. social scoring. Twoja hipoteza może być trafna, jeżeli system scoringowy spełnia warunki określone w art. 5(1)(c), czyli w uproszczeniu jest to system oceny ludzi, który prowadzi do jednej z dwóch zakazanych sytuacji opisanych wcześniej.

W praktyce najłatwiej „wpaść” w art. 5(1)(c)(i) w sytuacji, gdy bierzesz dane z jednego kontekstu, np. z aktywności internetowej, nawyków społecznych, informacji o pochodzeniu czy relacjach społecznych, a następnie wykorzystujesz je do podejmowania decyzji

w zupełnie innym, niezwiązanym kontekście, takim jak przyznawanie świadczeń socjalnych, decyzje o zatrudnieniu czy dostęp do mieszkania, i prowadzi to do gorszego traktowania danej osoby. Wytyczne Komisji podają przykłady, w których do oceny ryzyka nadużyć dołącza się cechy, które nie mają związku z daną sytuacją, np. narodowość lub pochodzenie etniczne współmałżonka, co w efekcie tworzy mechanizm selekcji, który jest niezwiązany z celem i może być arbitralny.

Wariant opisany w art. 5(1)(c)(ii)¹, czyli dotyczący nieuzasadnionego lub nieproporcjonalnego traktowania, jest bardziej ocenny i trudniejszy do jednoznacznego stwierdzenia. Wytyczne Komisji wskazują wprost, że w takich przypadkach każdą sytuację należy analizować indywidualnie, czyli tak zwanym podejściem case-by-case.

Istnieje również wariant, w którym hipoteza jest prawdziwa dlatego, że w grę wchodzi dane biometryczne i wrażliwe kategorie. Jeżeli takie cechy jak kolor skóry, kolor oczu czy płeć są pozyskiwane poprzez analizę biometryczną, np. na podstawie obrazu twarzy z kamery lub zdjęcia, a następnie system wykorzystuje je do wnioskowania o rasie lub innych wrażliwych kategoriach wskazanych w art. 5(1)(g)², to mamy do czynienia z sytuacją, która może stanowić zakazaną praktykę, czyli biometryczną kategoryzację prowadzącą do wnioskowania o takich cechach.

To jest najczystsza, najbardziej bezpośrednia podstawa w AI Act, która może odnosić się do kwestii takich jak kolor skóry czy rasa w Twoim przykładzie, ale tylko wtedy, gdy mówimy o wykorzystaniu danych biometrycznych.

Istnieje też wariant, w którym problem nie polega na naruszeniu art. 5, ale wchodzi w grę reżim systemów wysokiego ryzyka oraz związane z nim wymogi dotyczące przeciwdziałania dyskryminacji. Jeżeli system rekomendacji

nie służy wyłącznie do marketingu, lecz jest wykorzystywany do podejmowania decyzji w obszarach wskazanych w Załączniku III, takich jak rekrutacja, decyzje w relacji pracowniczej, ocena zdolności kredytowej czy ustalanie ceny lub ryzyka w ubezpieczeniach zdrowotnych lub na życie, to może zostać zakwalifikowany jako system wysokiego ryzyka.

Jakie działania należy podjąć?

W takich przypadkach AI Act nie wprowadza prostego zakazu, lecz nakłada konkretne obowiązki. Obejmują one m.in. konieczność wdrożenia procesu zarządzania ryzykiem w kontekście praw podstawowych zgodnie z art. 9 oraz obowiązek badania i ograniczania stronniczości danych, które mogą prowadzić do dyskryminacji zakazanej prawem Unii Europejskiej, zgodnie z art. 10.

Wniosek główny w pierwszej części jest taki, że w typowym scenariuszu dotyczącym promocji i zniżek w e-commerce sama segmentacja lub klasyfikacja klientów, nawet jeśli może budzić wątpliwości etyczne, nie oznacza automatycznie naruszenia AI Act. Wynika to z tego, że AI Act nie zawiera prostego przepisu w stylu „zakaz dyskryminacji w rekomendacjach” na poziomie art. 5.

Jednocześnie takie działanie może jednak wejść w zakres naruszenia AI Act w określonych sytuacjach. Może się tak stać po pierwsze wtedy, gdy system spełnia warunki social scoring opisane w art. 5(1)(c)³. Po drugie wtedy, gdy wykorzystuje dane biometryczne do wnioskowania o cechach wrażliwych w sposób objęty art. 5(1)(g). Po trzecie wtedy, gdy mamy do czynienia z systemem wysokiego ryzyka i nie są spełnione wymagania dotyczące przeciwdziałania stronniczości danych z art. 10 oraz zarządzania ryzykiem z art. 9, co może prowadzić do sankcji.

Jeżeli chodzi o relację systemów rekomendacyjnych i dyskryminacji do innych przepisów Unii Europejskiej, AI Act wprost wskazuje, że jego stosowanie odbywa się bez uszczerbku dla innych aktów prawnych, w szczególności tych

dotyczących ochrony konsumentów oraz bezpieczeństwa produktów. Wytyczne Komisji Europejskiej dodatkowo podkreślają, że przepisy AI Act, w tym zakazy i wyjątki, nie mogą być wykorzystywane do obchodzenia obowiązków wynikających z innych regulacji unijnych.

To ma duże znaczenie dla przykładu zastosowania systemów rekomendacji, ponieważ nawet jeśli dana praktyka nie zostanie uznana za zakazaną na podstawie art. 5(1)(c), to wciąż może być problematyczna z punktu widzenia innych przepisów. Jeżeli decyzje są oparte na cechach chronionych, takich jak rasa, pochodzenie etniczne czy płeć, to taka praktyka może podlegać unijnemu prawu antydyskryminacyjnemu. Wytyczne Komisji mówią wprost, że jeżeli ocena lub klasyfikacja opiera się na takich chronionych podstawach albo prowadzi do dyskryminacji, to taka praktyka będzie również podlegać przepisom Unii Europejskiej dotyczącym niedyskryminacji.

W praktyce oznacza to, że AI Act warto rozumieć jako zestaw określonych granic i obowiązków, a nie jako jedyne źródło regulacji w tym obszarze. Z jednej strony wyznacza on wyraźne „czerwone linie”, czyli sytuacje całkowicie zakazane, takie jak określone formy scoringu czy wykorzystania biometrii opisane w art. 5. Z drugiej strony, w przypadku systemów wysokiego ryzyka, nakłada obowiązek technicznego i organizacyjnego podejścia do ograniczania stronniczości, czyli tak zwanego podejścia „antibias”, wynikającego m.in. z art. 9 i 10. Jednocześnie nie zastępuje on innych przepisów i nie jest jedyną podstawą do oceny dyskryminacji w relacji między firmą a klientem.

Wojciech Moszczyński



Wojciech Moszczyński – absolwent Katedry Ekonometrii i Statystyki Uniwersytetu Mikołaja Kopernika w Toruniu, specjalista z zakresu ekonometrii, finansów, data science oraz rachunkowości zarządczej. Specjalizuje się w optymalizacji procesów produkcyjnych i logistycznych. Prowadzi badania w obszarze rozwoju i zastosowania sztucznej inteligencji. Od lat zaangażowany w popularyzację machine learning oraz data science w środowiskach biznesowych.

¹ Art. 5(1)(c)(ii) nie wolno tworzyć systemów, które „punktują” ludzi i na tej podstawie traktują ich gorzej w sposób przesadzony albo bez sensownego powodu.

² Art. 5(1)(g) nie wolno używać AI, które analizuje na przykład twarz lub inne cechy biometryczne człowieka, żeby „odgadnąć” jego wrażliwe cechy i na tej podstawie go klasyfikować.

³ Art. 5(1)(c) nie wolno tworzyć systemów, które „punktują” ludzi i na tej podstawie traktują ich gorzej, jeśli robią to w oderwaniu od kontekstu albo w sposób przesadzony lub bez sensownego uzasadnienia.