Jakub Kozłowski

Autor jest doktorantem w Katedrze Prawa Europejskiego Uniwersytetu Jagiellońskiego (ORCID: https://orcid.org/0000-0003-2889-1426).

Akt w sprawie sztucznej inteligencji w praktyce – charakterystyka unijnej regulacji opartej na ryzyku

Słowa kluczowe: *sztuczna inteligencja, system sztucznej inteligencji, systemy wysokiego ryzyka, prawa podstawowe*

Celem artykułu jest analiza kluczowych aspektów unijnej regulacji sztucznej inteligencji. Autor, wychodząc od omówienia definicji systemu sztucznej inteligencji zawartej w akcie w sprawie sztucznej inteligencji[1] oraz wpływu tej definicji na praktykę stosowania rozporządzenia, porównuje aktualną definicję z zawartymi we wcześniejszych wersjach aktu, podkreślając korekty, które mają na celu zwiększenie jej precyzji i odporności na zmiany technologiczne. W artykule przeprowadzono charakterystykę aktu jako regulacji, której podstawą jest kryterium ryzyka, wyjaśniając, jak unijny prawodawca klasyfikuje systemy AI według poziomu zagrożenia, które mogą stwarzać. Rozważania obejmują także interpretację pojęcia zakazanych praktyk na gruncie rozporządzenia oraz szczegółową charakterystykę systemów wysokiego ryzyka, wskazując na wyzwania związane ze stosowaniem tych przepisów i potencjalne konsekwencje prawne.

1. Wprowadzenie

Dyskurs publiczny oraz naukowy skupiony wokół sztucznej inteligencji znacząco zintensyfikował się w 2023 r. na skutek jej masowej popularyzacji. Sztuczna inteligencja trafiła „pod strzechy" wraz z rozwojem oraz upowszechnieniem dużych modeli językowych takich jak ChatGPT[2]. Unijne rozporządzenie – akt w sprawie sztucznej inteligencji – nie stanowi bezpośredniej odpowiedzi na to zjawisko. Prace nad unijną regulacją rozpoczęły się bowiem znacznie wcześniej[3]. Wśród kamieni milowych tego procesu możemy wyróżnić m.in. wydanie rezolucji Parlamentu Europejskiego zawierającej zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki[4], komunikat Komisji: *Sztuczna inteligencja dla Europy*[5], publikację Białej księgi: *Europejskie podejście do doskonałości i zaufania*[6] oraz zaprezentowanie przez Komisję projektu rozporządzenia ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji i zmieniającego niektóre akty ustawodawcze Unii, który określa się jako akt w sprawie sztucznej inteligencji *(AI Act)*[7]. Proces legislacyjny zakończył się uchwaleniem ostatecznej wersji aktu przez Parlament Europejski w marcu 2024 r. Regulacja oficjalnie weszła w życie 1.08.2024 r., chociaż jej kluczowe przepisy będą obowiązywały zgodnie z wyznaczonym przez prawodawcę kalendarium, stopniowo wchodząc w życie w najbliższych latach.

1 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz.Urz. UE L 2024/1689) – dalej akt w sprawie sztucznej inteligencji lub rozporządzenie.

2 Duży model językowy (LLM) to model oparty na głębokiej sieci neuronowej, wytrenowany na dużych zbiorach danych tekstowych, który potrafi generować tekst oraz odpowiadać na pytania w sposób zbliżony do ludzkiego, wykorzystując olbrzymią liczbę parametrów do modelowania złożonych zależności językowych.

3 Zob. J. Mazur, *Unia Europejska wobec rozwoju sztucznej inteligencji: proponowane strategie regulacyjne a budowanie jednolitego rynku cyfrowego*, „Europejski Przegląd Sądowy" 2020/9, s. 13–18.

4 Rezolucja Parlamentu Europejskiego z 16.02.2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)) (Dz.Urz. UE C 252 z 2018 r., s. 239).

5 Komunikat Komisji z 25.04.2018 r.: *Sztuczna inteligencja dla Europy*, COM(2018) 237 *final*.

6 Komisja Europejska, *Biała księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania*, COM (2020) 65 *final*.

7 Wniosek: rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniające niektóre akty ustawodawcze Unii, COM(2021) 206 *final*.

Struktura prowadzonych rozważań zakłada wyodrębnienie kilku węzłowych dla regulacji zagadnień, przedstawienie ich ogólnego kontekstu celem ułatwienia czytelnikowi zrozumienia niuansów determinujących zachowanie unijnego prawodawcy, prowadząc do charakterystyki wybranych problemów. Selekcja poruszonych kwestii ma na celu naświetlenie istotnych zagadnień, które w przyszłości związanej ze stosowaniem przepisów rozporządzenia będą przedmiotem interpretacji. Wybór poruszonych zagadnień jest także uzasadniony bieżącym dyskursem doktrynalnym oraz debatą branżową, które koncentrują się na problematyce wynikającej z wdrażania aktu w sprawie sztucznej inteligencji. Jednym z głównych obszarów zainteresowania interesariuszy są konsekwencje prawne związane z implementacją systemów AI w różnych sektorach oraz wpływ przepisów regulacji na rozwój innowacji technologicznych. Kluczowa jest także kwestia praktyki w obsłudze prawnej podmiotów wykorzystujących lub planujących wykorzystywać systemy sztucznej inteligencji. Z kolei debata branżowa skupia się na praktycznych aspektach regulacji, takich jak wdrażanie nowych obowiązków przez podmioty tworzące i stosujące systemy AI, wyzwania związane z klasyfikacją systemów pod kątem ryzyka oraz wpływ przepisów na funkcjonowanie rynku technologii sztucznej inteligencji. Analiza przedstawiona w artykule stanowi głos w tej dyskusji, koncentrując się na szczególnie wrażliwych, właśnie z perspektywy praktycznej, kwestiach, oferując jednocześnie refleksję nad praktycznymi konsekwencjami aktu oraz możliwymi kierunkami jego interpretacji w przyszłości.

Analiza ewolucji definicji systemu AI, od pierwszych wersji aktu po wersję ostateczną, pozwala na zrozumienie, w jaki sposób unijny prawodawca dążył do precyzyjnego określenia zakresu przedmiotowego regulacji, uwzględniając dynamiczny rozwój technologii. Dla zrozumienia natury unijnej regulacji konieczne jest przybliżenie tego, w jaki sposób unijny prawodawca projektuje typologię systemów opartych na ryzyku. Charakterystyka praktyk zakazanych oraz systemów wysokiego ryzyka stanowi podniesienie kluczowej z punktu widzenia praktyki kwestii. Jej istotność wynika ze szczególnie wrażliwej i szeroko dyskutowanej kwestii, czyli limitowania technologii w celu realizacji celów stawianych przez prawodawcę.

Poniższe rozważania nie aspirują do kompleksowego omówienia unijnej regulacji sztucznej inteligencji. Celem analizy jest ocena efektywności wybranych rozwiązań przyjętych przez unijnego prawodawcę, z uwzględnieniem ich potencjalnych konsekwencji dla przyszłych interpretacji prawnych oraz praktyki stosowania, a także zaproponowanie w tym zakresie postulatów *de lege ferenda*.

2. Regulacja oparta na kryterium ryzyka – założenia i zakres aktu w sprawie sztucznej inteligencji

Podejście przyjęte w akcie w sprawie sztucznej inteligencji zawiera elementy charakterystyczne dla regulacji „produktowych", kompilując je z zakreślaniem sfer szczególnie wrażliwych w kontekście ochrony praw podstawowych, ale niemożliwych do opisania przez swoją specyfikę za pomocą obiektywnych wskaźników[8]. Regulacje tego typu bazują na mierzalnych kryteriach pozwalających na ocenę zgodności z przepisami. W kontekście AI oznaczałoby to możliwość sformułowania wymagań technicznych i standardów, według których można ocenić systemy sztucznej inteligencji. Niemniej ze względu na specyfikę systemów sztucznej inteligencji określenie takich wskaźników okazuje się problematyczne. Wynika to z cech specyficznych dla systemów AI, takich jak zdolność do adaptacji, uczenia się czy funkcjonowania na podstawie złożonych algorytmów, które mogą być trudne do jednoznacznej parametryzacji. Dodatkowo unikalna charakterystyka systemów sztucznej inteligencji tworzy szczególne zagrożenia dla ochrony praw podstawowych. Sztuczna inteligencja immanentnie wiąże się z występowaniem tzw. czarnych skrzynek (*black boxes*)[9] czy mnogością przyczyn mogących wpływać na generowanie dyskryminujących wyników przez systemy sztucznej inteligencji. Uniemożliwia to dokonanie prostej parametryzacji poszczególnych systemów, analogicznie na przykład do substancji chemicznych. Unijny prawodawca przyjmuje zatem rozwiązanie polegające na podziale systemów sztucznej inteligencji ze względu na ryzyko, jakie generują dla wyróżnionych dóbr, adresując jednocześnie obszary szczególnie narażone na negatywne skutki wykorzystania sztucznej inteligencji. Następnie przyporządkowuje określone ograniczenia oraz obowiązki względem systemu o określonej kategorii. Ich adresatami najczęściej są dostawcy[10] oraz podmioty stosujące systemy sztucznej inteligencji[11]. Rozwiązania przyjęte w akcie w sprawie sztucznej inteligencji sformułowano w taki sposób, aby dostosować przepisy do różnych etapów cyklu życia systemu AI. Przykładem takiego podejścia jest różnicowanie obowiązków wynikających z aktu z uwagi na podmiot, który miałby je realizować, a także moment wdrażania obowiązków i wymagań mający odpowiadać na konkretne potrzeby związane z danym momentem funkcjonowania systemu.

Samo wyróżnienie kategorii ryzyka zawartych w akcie, pomimo pozornej przejrzystości w tym zakresie, nie jest intuicyjne oraz jednoznaczne. Należy zauważyć niespójność w identyfikowaniu oraz nazywaniu poszczególnych poziomów ryzyka, a tym samym kategorii ryzyka, wskazując, że nie jest klarowne to, czy rozporządzenie identyfikuje trzy czy cztery grupy. Kiedy sięgniemy do materiałów przygotowanych przez instytucje

8 T. Mahler, *Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal*, „Nordic Yearbook of Law and Informatics" 2021, s. 259.

9 W kontekście sztucznej inteligencji termin „czarne skrzynki" (ang. *black boxes*) odnosi się do modeli lub systemów AI, których wewnętrzne mechanizmy działania są nieprzejrzyste i trudne do zrozumienia, zarówno dla użytkowników, jak i dla samych twórców tych systemów. Charakterystyka ta wynika z wysokiego stopnia złożoności modeli, takich jak głębokie sieci neuronowe, które przetwarzają dane wejściowe i generują wyniki w sposób, który nie jest łatwo interpretowalny. Mimo że takie systemy mogą dostarczać dokładnych i skutecznych wyników, ich decyzje nie są bezpośrednio wyjaśnialne, co rodzi wyzwania związane z interpretowalnością, przejrzystością, zaufaniem oraz odpowiedzialnością za decyzje podejmowane przez te modele, zwłaszcza w kontekście zastosowań wrażliwych społecznie, takich jak diagnostyka medyczna, decyzje kredytowe czy egzekwowanie prawa.

10 Art. 3 pkt 3 aktu w sprawie sztucznej inteligencji: „dostawca" oznacza osobę fizyczną lub prawną, organ publiczny, agencję lub inny podmiot, które rozwijają system AI lub model AI ogólnego przeznaczenia lub zlecają rozwój systemu AI lub modelu AI ogólnego przeznaczenia oraz które – odpłatnie lub nieodpłatnie – pod własną nazwą lub własnym znakiem towarowym wprowadzają do obrotu lub oddają do użytku system AI.

11 Art. 3 pkt 4 aktu w sprawie sztucznej inteligencji: „podmiot stosujący" oznacza osobę fizyczną lub prawną, organ publiczny, agencję lub inny podmiot, które wykorzystują system AI, nad którym sprawują kontrolę, z wyjątkiem sytuacji, gdy system AI jest wykorzystywany w ramach osobistej działalności pozazawodowej.

unijne, których celem jest przybliżenie regulacji zawartych w akcie, często powielaną ilustracją służącą zobrazowaniu tej kwestii jest piramida. Patrząc od jej wierzchołka, tworzą ją systemy zaliczane kolejno do kategorii ryzyka nieakceptowalnego, wysokiego, ograniczonego (systemy objęte specjalnymi obowiązkami informacyjnymi) i minimalnego[12]. Nie ulega wątpliwości, że możemy wyróżnić dwie klarowne kategorie ryzyka. Obiema bowiem posługuje się unijny prawodawca w treści normatywnej aktu. Akt w sprawie sztucznej inteligencji w art. 5 wyróżnia kategorię zakazanych praktyk związanych ze sztuczną inteligencją, co determinuje istnienie niedopuszczalnego poziomu ryzyka. W kolejnym i następnych przepisach prawodawca posługuje się natomiast kategorią określaną jako wysokie ryzyko. Niejasny pozostaje jednak zakres systemów cechujących się „ograniczonym" oraz „minimalnym" ryzykiem.

Dodatkowo problematyczne pozostają także specjalne rodzaje systemów, dodane do regulacji na etapie prac legislacyjnych. Przykładem w tym zakresie są *deep fakes*. Pojęcie to zostało zdefiniowane w art. 3 pkt 60 aktu w sprawie sztucznej inteligencji[13] i wykorzystane do sformułowania obowiązków dotyczących przejrzystości zawartych w art. 50 ust. 4 aktu w sprawie sztucznej inteligencji. Patrząc z perspektywy oceny ryzyka, prawodawca dostrzega, że materiały typu *deep fake* generują specyficzne zagrożenia, w związku z czym ustanawia dedykowane normy, aby im zapobiec. Nie kwalifikuje jednocześnie w sposób jednoznaczny systemów pozwalających generować fałszywe nagrania do systemów zakazanych lub charakteryzujących się wysokim ryzykiem. Jednocześnie podwyższa standard ochrony odpowiadający na specyficzne zagrożenia, które wynikają ze zdiagnozowanego ryzyka. W konsekwencji pozycjonuje to pewne systemy niekwalifikujące się do kategorii objętych wysokim lub niedopuszczalnym ryzykiem w niesymetrycznej względem siebie pozycji, jeśli chodzi o zakres regulacji. Co więcej, istnieje możliwość, że niektóre typy systemów spełnią przesłanki ogólne z art. 6 ust. 1 aktu, co spowoduje zakwalifikowanie ich do systemów wysokiego ryzyka. Opisywany problem zaburza wizję prostej, piramidalnej struktury ryzyka, na której opiera się narracja na temat rozporządzenia prezentowana przez instytucje UE, a także większość piśmiennictwa. Ma to istotne znaczenie praktyczne, ogranicza bowiem poczucie pewności prawa w sytuacji, w której warunkiem efektywności regulacji jest prawidłowe funkcjonowanie systemu opartego na ryzyku.

Jaki jest zakres regulacji? Rozporządzenie operuje pojęciem systemu sztucznej inteligencji zamiast sztucznej inteligencji, co ma istotne znaczenie praktyczne. Oznacza bowiem, że przedmiotem regulacji nie jest abstrakcyjny i trudny do zdefiniowania koncept[14], ale realne elementy otaczającej nas rzeczywistości. Jest to rozwiązanie przyjęte już w pierwszej, przedstawionej w 2021 r., wersji aktu. Należy zakwalifikować je jako słuszne, pozwala bowiem stworzyć definicję spełniającą założenia niezbędne w kontekście nowych technologii, mianowicie odporność na zmiany techniczne i uniwersalność, przy zachowaniu odpowiedniego stopnia klarowności i precyzyjności. Artykuł 3 pkt 1 aktu w sprawie sztucznej inteligencji określa system sztucznej inteligencji jako „system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne".

Wskazana definicja niweluje mankamenty obecne w pierwszej wersji regulacji. Prawodawca zrezygnował z hybrydowego charakteru definicji, opierającego się na wskazaniu przesłanek ogólnych oraz zamkniętego katalogu technik determinujących uznanie danego systemu za kwalifikujący się do kategorii sztucznej inteligencji. Wykaz technik zawarty w załączniku do aktu miałby być aktualizowany w drodze aktów delegowanych przyjmowanych przez Komisję w sytuacji pojawiania się nowych rozwiązań technologicznych. Szczególnie problematyczne było wykorzystywane uprzednio pojęcie „oprogramowanie" oraz zakres wymienianych w załączniku części technik i podejść, m.in. „metody statystyczne" czy „techniki wyszukiwania", które mogły być bardzo szeroko rozumiane. W obecnej wersji definicji system sztucznej inteligencji nie jest już „oprogramowaniem", ale „systemem maszynowym". Zmiana ta prawidłowo kalibruje zakres przedmiotowy regulacji. Wiemy bowiem, że współczesne inkarnacje sztucznej inteligencji to nie tylko *software*. Definicja systemu AI opiera się na definicji wykorzystywanej przez OECD, co sprzyja spełnieniu postulatu pewności prawa i uzyskaniu uznania na poziomie międzynarodowym. Co ważne, prawodawca wprost akcentuje ten cel[15]. Uwzględnia również kluczową przesłankę pozwalającą odróżnić prostsze systemy od tych opartych na sztucznej inteligencji, mianowicie autonomiczność działania. System AI wykazuje zdolność do samouczenia się, a także operowania w pewnym stopniu niezależne od zaangażowania ze strony człowieka[16].

Systemem AI w rozumieniu definicji z aktu w sprawie sztucznej inteligencji będzie zarówno autonomiczny dron, wirtualny asystent w smartfonie, system zarządzania łańcuchem dostaw w branży spedycyjnej czy zaawansowany, wykorzystywany w edukacji chatbot. Stanowi to ilustrację szerokości zakresu regulacji, ale także skali wyzwania stojącego przed prawodawcą. Akt o sztucznej inteligencji jest regulacją, którą możemy określić jako „generalną". Oznacza to, że prawodawca unijny aspiruje do stworzenia „konstytucji AI", funkcjonującej jako *lex generalis* w odniesieniu do sztucznej inteligencji.

3. Zakazane praktyki

Artykuł 5 ust. 1 aktu w sprawie sztucznej inteligencji stanowi, że „zakazuje się następujących praktyk w zakresie AI". To, jak należy rozumieć pojęcie „praktyki w zakresie AI", nie zostało wyjaśnione ani w motywach rozporządzenia, ani w „słowniczku" z art. 3. Dlaczego zatem prawodawca posługuje się tym pojęciem? Odpowiedź na to pytanie wyznacza charakterystyka i rola najwyższego stopnia ryzyka w piramidzie stanowiącej podstawę klasyfikacji ryzyka w akcie o sztucznej inteligencji. Termin „praktyki" odnosi się do sposobu, w jaki systemy AI są wykorzystywane,

12 Komisja Europejska, *AI Act*, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (dostęp: 21.05.2024 r.).

13 *Deepfake* oznacza wygenerowane przez AI lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia, które odbiorca mógłby niesłusznie uznać za autentyczne lub prawdziwe.

14 Na temat multidyscyplinarności i złożoności dyskursu wokół pojęcia sztucznej inteligencji zob. S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Upper Saddle River 2010.

15 Motyw 12 aktu w sprawie sztucznej inteligencji.

16 Motyw 12 aktu w sprawie sztucznej inteligencji.

co jest kluczowe w kontekście regulacji. Zakazane mogą być określone sposoby wykorzystania technologii, które są uznane za niebezpieczne lub skrajnie nieetyczne, niezależnie od samego systemu. Przykładowo, pewne zastosowania technologii rozpoznawania twarzy mogą być zabronione, ale nie cała technologia rozpoznawania twarzy jako taka. Określone praktyki mogą być uznane za potencjalnie naruszające prawa podstawowe, nawet jeśli sam system AI jest neutralny z technicznego punktu widzenia.

Prezentowaną interpretację zdaje się potwierdzać dalsze brzmienie przepisu, na które składa się zamknięty katalog owych zakazanych praktyk. Poszczególne punkty przejmują uniwersalną konstrukcję w postaci określenia konkretnych, zakazanych czynności wyrażanych w postaci zwrotu: „wprowadzania do obrotu, oddawania do użytku lub wykorzystywania systemu AI" (art. 5 ust. 1 lit. a, b, c), „wprowadzania do obrotu, oddawania do użytku w tym konkretnym celu" (art. 5 ust. 1 lit. d, e, f, g) lub przedmiotowego opisu technologii opartej na AI. Wyjątkiem w tym zakresie jest art. 5 ust. 1 lit. h aktu w sprawie sztucznej inteligencji, w którym prawodawca posługuje się jedynie terminem „wykorzystanie", następnie wskazując przesłanki determinujące sytuację, w której owo wykorzystanie stanowi praktykę zakazaną.

Wynika z tego wniosek, że intencją prawodawcy jest precyzyjne sformułowanie, jakie konkretne działania związane z określonymi rodzajami systemów sztucznej inteligencji mają zostać zakazane. Unijny prawodawca, działając w świadomości stosowania najdalej ingerującego w rozwój oraz wykorzystanie technologii opartych na AI środka regulacyjnego, którym jest zakaz, projektował treść art. 5 aktu w sprawie sztucznej inteligencji, starając się maksymalnie precyzyjnie określić zakres zakazu. W ten sposób wypełnił zawartą w art. 1 aktu normę służącą zachowaniu balansu pomiędzy rozwojem technologicznym a ochroną wskazanych w przepisie dóbr. Jakich praktyk zakazuje art. 5 akt w sprawie sztucznej inteligencji? Ich przekrój jest bardzo szeroki, dotyczą m.in. scoringu społecznego (art. 5 ust. 1 lit. c), praktyk służących manipulacji spełniających przesłanki określone w art. 5 ust. 1 lit. a i b, wybranych praktyk bazujących na systemach kategoryzacji biometrycznej (art. 5 ust. 1 lit. g). Nie ulega wątpliwości, że zakazane praktyki z art. 5 będą przedmiotem interpretacji. Przykładem w tym zakresie jest przywoływany już art. 5 ust. 1 lit. a aktu w sprawie sztucznej inteligencji. Wyróżnia on dwie łączne przesłanki, które musi wypełnić dana praktyka, aby zostać zakwalifikowana jako niedopuszczalna. Po pierwsze, jest to stosowanie technik podprogowych będących poza świadomością danej osoby lub celowych technik manipulacyjnych lub wprowadzających w błąd. Po drugie, ma to doprowadzić do dokonania 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. Analogicznie, w wielu innych kontekstach, Trybunał definiował praktyki sprowadzające się do manipulacji. Przykładami w tym zakresie są orzeczenia w sprawach C-453/10, Pereničová i Perenič[17], czy C-435/11, CHS Tour Services[18], dotyczące ochrony konsumentów.

4. Systemy wysokiego ryzyka

System sztucznej inteligencji jest uznawany na gruncie aktu w sprawie sztucznej inteligencji za system wysokiego ryzyka, jeśli spełnia łącznie dwie przesłanki określone w art. 6 ust. 1. Po pierwsze, musi być przeznaczony do wykorzystania jako element związany z bezpieczeństwem produktu, który jest objęty unijnym prawodawstwem harmonizacyjnym wymienionym w załączniku I do aktu, lub sam system AI musi być takim produktem (art. 6 ust. 1 lit. a). Oznacza to, że system AI musi odgrywać kluczową rolę w zapewnieniu bezpieczeństwa produktu, który podlega regulacjom unijnym określonym w załączniku I. Przykładami takich produktów mogą być medyczne urządzenia diagnostyczne, systemy zarządzania infrastrukturą krytyczną czy zabawki[19]. Zakresem omawianego przepisu zostaną także objęte autonomiczne pojazdy, gdzie niezawodność i bezpieczeństwo działania AI mają bezpośredni wpływ na zdrowie i życie ludzi. W załączniku I do aktu w sprawie sztucznej inteligencji wymienione zostało rozporządzenie w sprawie wymogów dotyczących homologacji typu pojazdów silnikowych i ich przyczep oraz układów, komponentów i oddzielnych zespołów technicznych przeznaczonych do tych pojazdów, w odniesieniu do ich ogólnego bezpieczeństwa oraz ochrony osób znajdujących się w pojeździe i niechronionych uczestników ruchu drogowego[20], które formułuje w art. 11 wymogi szczegółowe dotyczące pojazdów zautomatyzowanych i w pełni zautomatyzowanych.

Po drugie, produkt, którego związanym z bezpieczeństwem elementem jest system AI lub sam system AI jako odrębny produkt, musi podlegać ocenie zgodności przez stronę trzecią przed jego wprowadzeniem do obrotu lub oddaniem do użytku (art. 6 ust. 1 lit. b). Ocena zgodności przez stronę trzecią oznacza, że niezależny organ ocenia, czy dany produkt spełnia określone wymogi bezpieczeństwa i normy techniczne, zanim zostanie on udostępniony na rynku lub użyty w praktyce. Proces oceny ma na celu zapewnienie, że produkt jest bezpieczny, niezawodny i zgodny z obowiązującymi regulacjami prawnymi, co jest szczególnie istotne w przypadku produktów, których nieprawidłowe działanie mogłoby prowadzić do poważnych konsekwencji dla zdrowia, bezpieczeństwa lub praw podstawowych osób fizycznych.

Opisana powyżej konstrukcja nie wyczerpuje aspiracji prawodawcy w zakresie szerokości zakresu przedmiotowego przepisów dotyczących systemów wysokiego ryzyka. Ich dopełnieniem są dalsze ustępy art. 6, poczynając od ust. 2, który stanowi,

17 Wyrok TS z 15.03.2012 r., C-453/10, Jana Pereničová i Vladislav Perenič przeciwko SOS financ, spol. s r. o., EU:C:2012:144, pkt 47. „Powoduje ona lub może spowodować podjęcie przez konsumenta decyzji dotyczącej transakcji, której inaczej by nie podjął".

18 Wyrok TS z 19.09.2013 r., C-435/11, CHS Tour Services GmbH przeciwko Team4 Travel GmbH, EU:C:2013:574, pkt 42. „Może spowodować podjęcie przez konsumenta decyzji dotyczącej transakcji, której nie podjąłby w braku istnienia takiej praktyki".

19 W załączniku I do aktu w sprawie sztucznej inteligencji znajduje się 20 aktów prawa wtórnego UE (11 rozporządzeń i 9 dyrektyw).

20 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2019/2144 z 27.11.2019 r. w sprawie wymogów dotyczących homologacji typu pojazdów silnikowych i ich przyczep oraz układów, komponentów i oddzielnych zespołów technicznych przeznaczonych do tych pojazdów, w odniesieniu do ich ogólnego bezpieczeństwa oraz ochrony osób znajdujących się w pojeździe i niechronionych uczestników ruchu drogowego, zmieniające rozporządzenie Parlamentu Europejskiego i Rady (UE) 2018/858 oraz uchylające rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 78/2009, (WE) nr 79/2009 i (WE) nr 661/2009 oraz rozporządzenia Komisji (WE) nr 631/2009, (UE) nr 406/2010, (UE) nr 672/2010, (UE) nr 1003/2010, (UE) nr 1005/2010, (UE) nr 1008/2010, (UE) nr 1009/2010, (UE) nr 19/2011, (UE) nr 109/2011, (UE) nr 458/2011, (UE) nr 65/2012, (UE) nr 130/2012, (UE) nr 347/2012, (UE) nr 351/2012, (UE) nr 1230/2012 i (UE) 2015/166 (Dz.Urz. UE L 325, s. 1).

że oprócz systemów AI wysokiego ryzyka określonych w ust. 1, za systemy wysokiego ryzyka uznaje się również systemy AI wymienione w załączniku III. Komisja może modyfikować załącznik III do aktu w sprawie sztucznej inteligencji, uwzględniając zasady sformułowane w art. 7 aktu. Zaimplementowany został zatem mechanizm podobny do tego, z którego zrezygnowano w przypadku przepisów dotyczących definicji systemu sztucznej inteligencji. Prawodawca określa przesłanki ogólne, do których dołączony został załącznik zawierający konkretny wykaz, który w założeniu może być dynamicznie aktualizowany w zależności od potrzeb wynikających ze zmian technologicznych czy potrzeb społecznych. Załącznik w tym wypadku pełni nieco inną funkcję niż w przypadku definicji systemów SI, ponieważ nie służy precyzowaniu definicji ogólnej – uzupełnia i poszerza zakres przepisu. Jego rolę można porównać do elastycznego i niezbędnego wypełniacza, który służy temu, żeby zbudowana przez prawodawcę konstrukcja okazała się szczelna i stabilna.

Załącznik III do aktu w sprawie sztucznej inteligencji zawiera osiem kategorii systemów SI wysokiego ryzyka. Kategorie obejmują systemy w następujących obszarach: zdalna identyfikacja i kategoryzacja biometryczna, rozpoznawanie emocji; zarządzanie infrastrukturą krytyczną; decyzje edukacyjne i oceny w instytucjach edukacyjnych; rekrutacja i zarządzanie pracownikami; ocena kwalifikowalności do usług publicznych, ocena zdolności kredytowej, ustalanie ryzyka ubezpieczeniowego, zarządzanie zgłoszeniami alarmowymi; wsparcie organów ścigania; zarządzanie migracją, azylem i kontrolą granic; wspomaganie wymiaru sprawiedliwości i procesów demokratycznych.

Warto zwrócić uwagę na zakwalifikowanie jako systemów cechujących się wysokim ryzykiem tych kategorii i zastosowań, co do których dysponujemy udokumentowanymi przypadkami w zakresie generowania dyskryminujących wyników. Ilustracją w tym zakresie mogą być przypadki dotyczące zatrudniania. Akt w sprawie o sztucznej inteligencji klasyfikuje jako cechujące się wysokim ryzkiem systemy przeznaczone do rekrutacji i wyboru kandydatów, podejmowania decyzji wpływających na warunki i charakter stosunku pracy, awansu lub rozwiązania umowy, a także monitorowania i oceny wydajności pracowników. Charakterystyka omawianego katalogu wyraźnie świadczy o świadomości prawodawcy w zakresie przeszłych spraw dotyczących wykorzystania systemów AI i ryzyka dla praw podstawowych, które generują.

Nie każdy system AI należący do grup wskazanych w załączniku III będzie kwalifikowany jako system wysokiego ryzyka na gruncie aktu w sprawie o sztucznej inteligencji. Prawodawca przewidział bowiem w tym zakresie wyjątki. System AI nie jest klasyfikowany jako system wysokiego ryzyka, pod warunkiem że nie stwarza on istotnego zagrożenia dla zdrowia, bezpieczeństwa ani praw podstawowych osób fizycznych (art. 6 ust. 3). Szczególnie dotyczy to sytuacji, gdy system nie wywiera znaczącego wpływu na wynik procesu decyzyjnego (art. 6 ust. 3). Warunki, które umożliwiają takie wyłączenie, obejmują: przeznaczenie systemu do realizacji ściśle określonego zadania proceduralnego; wykorzystanie systemu do poprawy wyniku uprzednio zakończonej czynności ludzkiej; zastosowanie systemu do detekcji wzorców decyzyjnych lub odchyleń od nich bez intencji zastąpienia lub wpływania na wcześniejszą ludzką ocenę bez odpowiedniej weryfikacji; oraz użycie systemu do zadań przygotowawczych w kontekście oceny istotnej dla przypadków użycia wymienionych w załączniku III (art. 6 ust. 3 lit. a–d). Niezależnie od powyższych wyłączeń każdy system AI wspomniany w załączniku III, który dokonuje profilowania osób fizycznych, zawsze będzie uznawany za system wysokiego ryzyka (art. 6 ust. 3).

Na jakiej podstawie prawodawca dokonał selekcji systemów znajdujących się w katalogu z załącznika III do aktu w sprawie sztucznej inteligencji? Celem ustalenia bardziej uniwersalnych przesłanek, którymi kierował się w tym zakresie prawodawca, należy odwołać się do art. 7 aktu w sprawie sztucznej inteligencji, w którym określono tryb zmian w załączniku III. Analiza art. 7 pozwala dokonać rekonstrukcji bardziej szczegółowych kryteriów, którymi kierował się prawodawca przy tworzeniu katalogu systemów wysokiego ryzyka załącznika III. Skoro bowiem określony system powinien zostać zbadany pod kątem przesłanek sformułowanych w art. 7 w razie wystąpienia intencji jego dołączenia do załącznika III, to z wysokim prawdopodobieństwem można stwierdzić, że kryteria te były kluczowe na etapie prowadzenia procesu legislacyjnego. Jedną z przesłanek, które muszą zostać spełnione, aby system AI został dodany do załącznika III, jest to, że „stwarza ryzyko szkody dla zdrowia i bezpieczeństwa lub ryzyko niepożądanego wpływu na prawa podstawowe i ryzyko to jest równoważne ryzyku szkody lub niepożądanego wpływu, jakie stwarzają systemy AI wysokiego ryzyka wymienione już w załączniku III, lub jest od niego większe" (art. 7 ust. 1 lit. b). Oznacza to, że Komisja nie może „obniżyć standardu" ochrony poprzez uzupełnienie katalogu systemów wysokiego ryzyka o mniej inwazyjny system. Czy można ów poziom zmierzyć? Prawodawca w tym względzie formułuje kryteria szczegółowe, które powinny zostać uwzględnione przez Komisję w przypadku takiej oceny. Warto wśród nich wymienić zakres i przeznaczenie danego systemu AI, zakres autonomiczności systemu, charakter przetwarzanych przez system danych, historię działania danego systemu (w szczególności szkody, które już zostały wyrządzone przez jego działanie), to, czy wykorzystanie systemu AI nie tworzy sytuacji dysproporcji sił pomiędzy człowiekiem a podmiotem wykorzystującym system AI, oraz to, czy istnieje możliwość rezygnacji z bycia objętym działaniem systemu (art. 7 ust. 2). Wymienione kryteria wskazują na podmiotowy aspekt ochrony, skupiony na człowieku i prawach podstawowych. Spełnia w ten sposób założenie antropocentrycznej regulacji sztucznej inteligencji. Racjonalnym postulatem uzupełniającym funkcjonalnie zakres przepisu byłoby dodanie kryteriów związanych z ochroną środowiska, wpływem systemów sztucznej inteligencji na demokrację czy inne wartości z art. 2 Traktatu o Unii Europejskiej[21].

System został dopełniony w naturalny sposób wymogami, które system wysokiego ryzyka musi spełniać, które tworzą tzw. system zarządzania ryzykiem (art. 8–15). Na przykład, art. 11 aktu w sprawie sztucznej inteligencji nakłada obowiązek opracowania i utrzymywania szczegółowej dokumentacji technicznej, która pozwala na ocenę zgodności systemu z obowiązującymi przepisami, a art. 13 wymaga, aby systemy były zaprojektowane w sposób zapewniający przejrzystość działania, a użytkownicy muszą być informowani o tym, że korzystają z systemu sztucznej inteligencji. Artykuł 14 aktu w sprawie sztucznej inteligencji podkreśla konieczność zapewnienia odpowiedniego nadzoru ludzkiego nad działaniem systemu, co ma na celu minimalizację ryzyka i umożliwienie interwencji

21 Wersja skonsolidowana Dz.Urz. UE C 202 z 2016 r., s. 13.

w sytuacjach krytycznych. Istotnym dla praktyki rozwiązaniem przyjętym w akcie jest wprowadzenie w art. 27 obowiązku przeprowadzenia obligatoryjnego procesu oceny skutków dla praw podstawowych dla niektórych systemów wysokiego ryzyka. Co interesujące, prawodawca różnicuje w tym zakresie obowiązek przeprowadzania procesu oceny ze względu na publiczny lub prywatny charakter podmiotu stosującego. Obowiązuje podwyższony standard w tym zakresie dla podmiotów prawa publicznego lub podmiotów prywatnych świadczących usługi publiczne. Podmioty prywatne zobligowane są do przeprowadzania oceny wyłącznie w odniesieniu do dwóch rodzajów systemów: przeznaczonych do wykorzystywania do celów oceny zdolności kredytowej osób fizycznych lub ustalenia ich scoringu kredytowego, także tych dotyczących oceny ryzyka i ustalania cen w odniesieniu do osób fizycznych w przypadku ubezpieczenia na życie i ubezpieczenia zdrowotnego (załącznik III pkt 5 lit. b i c). Wydaje się, że zakres powinien być szerszy i objąć choćby systemy wykorzystywane w celach rekrutacyjnych czy wykorzystujące technologię zdalnej identyfikacji biometrycznej. Dla porównania, regulacja oceny skutków dla ochrony danych z RODO[22] nie zawiera ograniczeń podmiotowych, a przesłanki dotyczące przeprowadzania samej oceny mają charakter przedmiotowy.

5. Podsumowanie

Dalszych badań niewątpliwie wymagają m.in. relacje pomiędzy aktem w sprawie sztucznej inteligencji a innymi regulacjami w prawie unijnym, w szczególności RODO czy DSA (Digital Services Act). Akt w sprawie sztucznej inteligencji stanowi istotny krok w regulacji dynamicznie rozwijającej się technologii, jednak jego wejście w życie napotyka na szereg wyzwań związanych z interpretacją przepisów i zakresem regulacji. W szczególności, choć definicja systemu AI przyjęta w akcie jest bardziej precyzyjna niż we wcześniejszych wersjach, to aktualna regulacja może wywołać problemy interpretacyjne wynikające z wielowarstwowego i nieprecyzyjnego podejścia do kategoryzacji ryzyka wykorzystywanych w akcie. Wykładni będzie wymagać także omówione pojęcie praktyki w kontekście systemów objętych niedopuszczalnym ryzykiem, podobnie jak wiele innych sformułowań, np. pojęcie szkody. Wątpliwości budzi również zakres poszczególnych kategorii ryzyka, ich wzajemne relacje. Istotną trudność interpretacyjną może sprawiać kwalifikowanie konkretnych systemów AI do poszczególnych kategorii ryzyka wytyczanych przez akt w sprawie sztucznej inteligencji. Ponadto wątpliwości budzi również przyjęty w rozporządzeniu zakres poszczególnych kategorii ryzyka, ze względu na pominięcie wskazanych wcześniej systemów, które wydają się być inwazyjne dla dóbr chronionych przez prawodawcę za pomocą aktu. Rekomendowane jest dalsze monitorowanie i ewentualne modyfikacje regulacji, aby zapewnić jej adekwatność i skuteczność, zwłaszcza w kontekście ochrony praw podstawowych oraz minimalizacji ryzyka związanego z wykorzystaniem sztucznej inteligencji.

Unia Europejska od początku prac nad regulacją sztucznej inteligencji kieruje się antropocentrycznym paradygmatem, który ma cechować się priorytetową rolą człowieka i jego praw oraz stworzenia takiego środowiska, w którym będą one najpełniej chronione. Warto dodatkowo zauważyć, że jeśli rozwiązanie proponowane przez UE zostanie przyjęte i będzie efektywne, może ono stać się globalnym standardem, oddziałując na zewnątrz między innymi poprzez efekt brukselski[23].

Abstract

Jakub Kozłowski
The author is a doctoral student at the Department of European Law, Jagiellonian University in Kraków, Poland (ORCID: https://orcid.org/0000-0003-2889-1426).

Artificial Intelligence Act in Practice – Characteristics of EU Risk-Based Regulation

Keywords: *artificial intelligence, artificial intelligence system, high-risk systems, fundamental rights*

The article aims at analysing key aspects of EU regulation of artificial intelligence. The author begins with a discussion of the definition of an artificial intelligence system in the Artificial Intelligence Act and the impact of this definition on the practice of applying the regulation. Then, the current definition is compared with those in previous versions of the Act, highlighting adjustments aimed at enhancing its precision and resilience to technological changes. The article characterises the Act as a risk-based regulation, explaining how the EU legislator classifies AI systems according to the level of threat they may pose. The analysis also includes an interpretation of the prohibited practices concept under the regulation and a detailed description of high-risk systems, pointing out the challenges related to the application of these provisions and the potential legal consequences.

Bibliografia/References

Bradford A., *The Brussels Effect: How the European Union Rules the World*, New York 2020.
Mahler T., *Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal,* „Nordic Yearbook of Law and Informatics" 2021.
Mazur J., *Unia Europejska wobec rozwoju sztucznej inteligencji: proponowane strategie regulacyjne a budowanie jednolitego rynku cyfrowego*, „Europejski Przegląd Sądowy" 2020/9.
Russell S., Norvig P., *Artificial Intelligence. A Modern Approach*, Upper Saddle River 2010.

22 Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z 27.04.2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych) (Dz.Urz. UE L 119, s. 1) – dalej RODO. W art. 35 ust. 1 i 3 tego rozporządzenia prawodawca formułuje przesłanki ogólne: „Jeżeli dany rodzaj przetwarzania – w szczególności z użyciem nowych technologii – ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, administrator przed rozpoczęciem przetwarzania dokonuje oceny skutków planowanych operacji przetwarzania dla ochrony danych osobowych" (ust. 1) i uzupełnia katalogiem sytuacji, w których dokonanie oceny skutków dla ochrony danych jest obowiązkowe (ust. 3).

23 A. Bradford, *The Brussels Effect: How the European Union Rules the World*, New York 2020, s. 1–2.