UNIWERSYTET ZIELONOGÓRSKI

DYSKURS PRAWNICZY I ADMINISTRACYJNY 2/2026 ISSN 2657-926X

Martyna Kaczmarczyk
Uniwersytet Warmińsko-Mazurski w Olsztynie
ORCID: 0000-0001-6169-9466
martyna.kaczmarczyk@uwm.edu.pl

Między innowacją a bezpieczeństwem: o zasadach regulujących europejski system sztucznej inteligencji

Słowa kluczowe: sztuczna inteligencja, prawa podstawowe, cyberbezpieczeństwo, prawo Unii Europejskiej

Streszczenie. Dynamiczny rozwój technologii sztucznej inteligencji stwarza nowe możliwości innowacji, jednocześnie rodząc istotne wyzwania prawne, etyczne i społeczne. W odpowiedzi na te wyzwania Unia Europejska wprowadziła kompleksowe ramy regulacyjne mające na celu zapewnienie, że systemy sztucznej inteligencji są projektowane i wykorzystywane w sposób bezpieczny, przejrzysty i godny zaufania. Niniejszy artykuł analizuje zasady przewodnie regulujące funkcjonowanie sztucznej inteligencji w europejskim środowisku prawnym, ze szczególnym uwzględnieniem przepisów zawartych w AI Act. W opracowaniu omówiono podejście oparte na analizie ryzyka, które klasyfikuje systemy AI w zależności od ich potencjalnego wpływu na bezpieczeństwo, prawa podstawowe oraz interes społeczny. Szczególną uwagę poświęcono kluczowym zasadom regulacyjnym, takim jak przejrzystość działania systemów, nadzór człowieka, odpowiedzialność podmiotów tworzących i wykorzystujących AI, jakość danych oraz zarządzanie ryzykiem. W artykule przedstawiono również obowiązki nakładane na dostawców i twórców systemów wysokiego ryzyka, w tym wymagania dotyczące dokumentacji technicznej, procedur oceny zgodności oraz monitorowania systemów po ich wdrożeniu. Przeprowadzona analiza wskazuje, że europejskie podejście regulacyjne dąży do znalezienia równowagi między wspieraniem innowacji technologicznych a ochroną praw podstawowych i budowaniem zaufania społecznego do nowych technologii. Ustanowienie jasnych standardów prawnych dla projektowania i stosowania systemów sztucznej inteligencji ma sprzyjać odpowiedzialnemu rozwojowi AI, jednocześnie ograniczając potencjalne zagrożenia wynikające z wykorzystania zaawansowanych technologii algorytmicznych.

Between innovation and security: about principles governing the European artificial intelligence system

Keywords: artificial intelligence, fundamental rights, cybersecurity, European Union law

Summary. Rapid development of artificial intelligence technologies has created new opportunities for innovation while simultaneously raising significant legal, ethical and societal concerns. In response to these challenges European Union has introduced a comprehensive regulatory framework aimed at ensuring that artificial intelligence systems are developed in a safe, transparent and trustworthy manner. Article analyzes guiding principles governing artificial intelligence within the European regulatory environment, with particular reference to the provisions of the AI Act. Study examines the risk-based approach adopted by the regulation, which classifies AI systems according to their potential impact on fundamental rights, safety and societal interests. Special attention is devoted to

key regulatory principles such as transparency, human oversight, accountability, data quality and risk management. Article also discusses the obligations imposed on developers and providers of high-risk AI systems, including requirements related to documentation, compliance assessment and post-market monitoring. Analysis demonstrates that the European approach seeks to balance technological innovation with the protection of fundamental rights and public trust. By establishing clear legal standards for the design and deployment of AI systems, the regulation aims to foster responsible innovation while minimizing potential harms associated with the use of advanced algorithmic technologies. Article concludes that the European regulatory model may serve as an important reference point for global discussions on the governance of artificial intelligence.

Wprowadzenie

Rozwój technologii cyfrowych w ostatnich latach doprowadził do znaczącego wzrostu roli systemów sztucznej inteligencji w wielu obszarach życia społecznego, gospodarczego oraz administracyjnego. Rozwiązania oparte na sztucznej inteligencji znajdują coraz szersze zastosowanie między innymi w medycynie, finansach, edukacji, transporcie, przemyśle czy administracji publicznej. Systemy te wspierają procesy decyzyjne, analizę dużych zbiorów danych oraz automatyzację wielu zadań, które dotychczas wymagały bezpośredniego udziału człowieka. W konsekwencji sztuczna inteligencja staje się jednym z kluczowych czynników transformacji cyfrowej współczesnych społeczeństw oraz gospodarek.

Jednocześnie dynamiczny rozwój technologii sztucznej inteligencji rodzi liczne wyzwania o charakterze prawnym, etycznym i społecznym. Coraz częściej podnoszone są kwestie związane z przejrzystością działania algorytmów, odpowiedzialnością za decyzje podejmowane przez systemy autonomiczne, ochroną danych osobowych czy ryzykiem występowania uprzedzeń algorytmicznych. W kontekście rosnącej autonomii systemów AI pojawiają się również pytania dotyczące możliwości skutecznego nadzoru nad ich funkcjonowaniem oraz zapewnienia zgodności ich działania z podstawowymi wartościami demokratycznego państwa prawa, w tym z poszanowaniem praw podstawowych jednostki.

W odpowiedzi na te wyzwania instytucje europejskie podjęły działania zmierzające do stworzenia spójnych i kompleksowych ram prawnych regulujących rozwój oraz wykorzystanie technologii sztucznej inteligencji. Jednym z najważniejszych elementów tych działań jest przyjęcie aktu prawnego znanego jako AI Act[1]. Regulacja ta stanowi pierwszą na świecie próbę całościowego uregulowania funkcjonowania systemów sztucznej inteligencji na poziomie ponadnarodowym. Jej głównym celem jest stworzenie ram prawnych umożliwiających bezpieczny

[1] Parlament Europejski i Rada Unii Europejskiej. Rozporządzenie (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (AI Act), Dz.Urz. UE L 1689, 12.07.2024.

i odpowiedzialny rozwój technologii AI, przy jednoczesnym wspieraniu innowacyjności i konkurencyjności gospodarki europejskiej.

Podstawą przyjętego modelu regulacyjnego jest podejście oparte na analizie ryzyka, w ramach którego systemy sztucznej inteligencji klasyfikowane są w zależności od poziomu potencjalnego zagrożenia, jakie mogą stwarzać dla bezpieczeństwa użytkowników, praw podstawowych jednostki oraz interesu publicznego. W zależności od stopnia ryzyka regulacja przewiduje zróżnicowany zakres obowiązków nakładanych na podmioty projektujące, rozwijające i wdrażające systemy AI. Szczególnie rygorystyczne wymagania dotyczą systemów wysokiego ryzyka, które mogą mieć istotny wpływ na sytuację prawną lub społeczną jednostki, np. w obszarze zatrudnienia, edukacji, ochrony zdrowia czy usług finansowych[2].

Istotnym elementem europejskiego podejścia do regulacji sztucznej inteligencji jest także promowanie tzw. godnej zaufania sztucznej inteligencji (*trustworthy AI*), której funkcjonowanie powinno opierać się na zasadach przejrzystości, bezpieczeństwa, odpowiedzialności oraz nadzoru człowieka[3]. W tym kontekście regulacja ma na celu nie tylko ograniczenie potencjalnych zagrożeń związanych z wykorzystaniem technologii AI, lecz również budowanie społecznego zaufania do nowych technologii oraz wspieranie ich odpowiedzialnego wdrażania w różnych sektorach gospodarki.

Celem niniejszego artykułu jest przedstawienie i analiza zasad przewodnich regulujących funkcjonowanie systemów sztucznej inteligencji w europejskim porządku prawnym. Szczególna uwaga zostanie poświęcona kluczowym zasadom regulacyjnym wynikającym z przepisów wspomnianego aktu prawnego, takim jak przejrzystość działania systemów AI, nadzór człowieka nad procesami decyzyjnymi, odpowiedzialność podmiotów rozwijających i wdrażających technologie sztucznej inteligencji, a także wymogi dotyczące jakości danych oraz zarządzania ryzykiem. Analiza ta ma na celu ukazanie, w jaki sposób europejski model regulacyjny stara się równoważyć potrzebę wspierania innowacji technologicznych z koniecznością zapewnienia bezpieczeństwa, ochrony praw podstawowych oraz budowania zaufania społecznego do systemów sztucznej inteligencji.

Podejście oparte na ryzyku

Jednym z kluczowych elementów europejskiego modelu regulacyjnego jest podejście oparte na analizie ryzyka, stanowiące fundament regulacji zawartej w AI Act[4].

[2] AI Act, art. 8-15.

[3] Komisja Europejska, *Communication on Fostering a European Approach to Trustworthy AI*, COM (2019) 168 final, 2019, https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-trustworthy-ai [dostęp: 21.01.2026].

[4] AI Act, art. 6.

Podejście to zakłada, że zakres obowiązków prawnych oraz poziom kontroli nad systemami sztucznej inteligencji powinny być uzależnione od stopnia zagrożenia, jakie dany system może stwarzać dla bezpieczeństwa, praw podstawowych jednostki oraz interesu publicznego.

Istotą podejścia opartego na ryzyku jest klasyfikacja systemów sztucznej inteligencji według poziomu potencjalnego ryzyka ich zastosowania. W ramach przyjętego modelu wyróżnia się cztery podstawowe kategorie systemów AI: systemy stwarzające niedopuszczalne ryzyko, systemy wysokiego ryzyka, systemy o ograniczonym ryzyku oraz systemy o minimalnym ryzyku[5]. Najbardziej restrykcyjne regulacje dotyczą systemów uznanych za stwarzające niedopuszczalne ryzyko, których stosowanie zostało w zasadzie zakazane ze względu na możliwość poważnego naruszenia praw podstawowych, np. poprzez manipulowanie zachowaniem użytkowników lub stosowanie niektórych form biometrycznej identyfikacji w przestrzeni publicznej.

Szczególną kategorię stanowią systemy wysokiego ryzyka, które mogą mieć istotny wpływ na sytuację prawną lub społeczną jednostki. Do tej grupy zaliczane są m.in. systemy wykorzystywane w procesach rekrutacyjnych, systemy oceny zdolności kredytowej, rozwiązania stosowane w sektorze ochrony zdrowia czy w zarządzaniu infrastrukturą krytyczną. W odniesieniu do tych systemów regulacja przewiduje szereg obowiązków dla podmiotów rozwijających i wdrażających te systemy. Obejmują one między innymi konieczność przeprowadzania analizy ryzyka, zapewnienia odpowiedniej jakości danych wykorzystywanych do trenowania modeli, prowadzenia szczegółowej dokumentacji technicznej oraz zapewnienia nadzoru człowieka nad działaniem systemu[6].

Systemy o ograniczonym ryzyku podlegają przede wszystkim wymogom transparentności. W praktyce oznacza to obowiązek informowania użytkowników o tym, że mają do czynienia z systemem sztucznej inteligencji, co ma szczególne znaczenie w przypadku *chatbotów* czy systemów generujących treści[7]. Natomiast systemy o minimalnym ryzyku, takie jak filtry antyspamowe czy algorytmy rekomendacyjne, mogą być stosowane bez dodatkowych wymogów regulacyjnych, ponieważ uznaje się, że nie stwarzają one istotnego zagrożenia dla użytkowników lub społeczeństwa[8].

---

[5] *Ibidem*.

[6] A. Pajdo, *Zakres obowiązków podmiotów odpowiedzialnych za systemy sztucznej inteligencji wysokiego ryzyka w świetle aktu o SI – Rozporządzenia 2024/1689*, „Roczniki Nauk Prawnych" 2025, nr 2(35), s. 47.

[7] AI Act, art. 50; S. Westerstrand, *Fairness in AI systems development: EU AI Act compliance and beyond*, „Information and Software Technology" 2025, vol. 187, s. 7.

[8] AI Act, art. 53; M. Ebers, *Truly Risk-based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*, „European Journal of Risk Regulation" 2024, vol. 16, is. 2, s. 6-7.

Podejście oparte na analizie ryzyka stanowi próbę znalezienia równowagi pomiędzy potrzebą zapewnienia bezpieczeństwa a koniecznością wspierania rozwoju innowacji technologicznych. Zamiast wprowadzać jednolite, restrykcyjne regulacje dla wszystkich systemów sztucznej inteligencji, model ten umożliwia elastyczne dostosowanie poziomu regulacji do rzeczywistego stopnia zagrożenia. Dzięki temu możliwe jest jednoczesne promowanie rozwoju technologii AI oraz ograniczanie potencjalnych negatywnych skutków ich wykorzystania. W konsekwencji podejście oparte na ryzyku stanowi jeden z najważniejszych elementów współczesnej regulacji sztucznej inteligencji w Europie. Jego skuteczność będzie jednak zależeć od praktycznego wdrożenia przepisów oraz od dalszego rozwoju badań naukowych analizujących wpływ tych regulacji na rozwój technologii oraz ochronę praw jednostki.

Wymóg transparentności

W miarę jak systemy sztucznej inteligencji stają się coraz bardziej powszechne w codziennym życiu i w kluczowych sektorach gospodarki, rośnie znaczenie zasady transparentności w regulacji ich funkcjonowania. Transparentność stanowi jeden z fundamentalnych filarów europejskiego modelu regulacyjnego, określonego w AI Act, i ma na celu zapewnienie, że użytkownicy i interesariusze systemów AI są świadomi sposobu ich działania, a także potencjalnych konsekwencji podejmowanych decyzji[9].

Wymóg transparentności obejmuje kilka kluczowych aspektów. Przede wszystkim systemy sztucznej inteligencji powinny informować użytkowników, gdy mają do czynienia z AI, co jest szczególnie istotne w przypadku interaktywnych narzędzi, takich jak *chatboty*, systemy rekomendacyjne czy generatory treści. Brak takiej informacji może prowadzić do wprowadzania użytkowników w błąd oraz utrudniać świadome korzystanie z technologii. Transparentność wymaga także udostępnienia informacji dotyczących źródeł danych, metod analitycznych i kryteriów podejmowania decyzji przez algorytmy, co umożliwia audyt i ocenę zgodności systemu z zasadami etycznymi i prawnymi[10].

W kontekście systemów wysokiego ryzyka transparentność staje się nie tylko narzędziem informacyjnym, ale także mechanizmem odpowiedzialności i nadzoru. Umożliwia to podmiotom regulacyjnym, użytkownikom i ekspertom technicznym ocenę, czy systemy AI działają zgodnie z obowiązującymi standardami bezpieczeń-

[9] K. Vladimirova, *The Principle of Transparency Under The Artificial Intelligence Act*, „Law Journal of New Bulgarian University" 2024, nr 2 (vol. 20), s. 44.

[10] M. Wyszomirska, *Implementation of the AI Act (EU) into the Polish legal system, its impact on this system and on selected strategic sectors*, „Safety & Fire Technology" 2025, nr 2(66), s. 120.

stwa, jakości danych i ochrony praw podstawowych. Transparentność przyczynia się również do budowania zaufania społecznego do technologii, co jest niezbędne dla szerszego wdrażania innowacyjnych rozwiązań w sektorach takich jak zdrowie, edukacja czy finanse.

W literaturze akademickiej wymóg transparentności systemów AI dopiero zaczyna być analizowany w kontekście praktycznego wdrażania regulacji i oceny ich skuteczności. Badacze zwracają uwagę na wyzwania związane z równoważeniem pełnej przejrzystości z ochroną własności intelektualnej, złożonością algorytmów oraz ryzykiem nadużyć informacyjnych[11]. Pomimo wyzwań transparentność pozostaje jednym z kluczowych elementów odpowiedzialnego rozwoju sztucznej inteligencji, zapewniającym zarówno ochronę użytkowników, jak i możliwość prowadzenia badań nad wpływem AI na społeczeństwo.

Podsumowując, wymóg transparentności w regulacji systemów AI nie jest jedynie formalnym obowiązkiem prawnym, lecz fundamentem budowania zaufania, nadzoru i odpowiedzialności w ekosystemie sztucznej inteligencji. Prawidłowe wdrożenie jest niezbędne dla zapewnienia, że rozwój technologii będzie przebiegał w sposób etyczny, bezpieczny i zgodny z wartościami demokratycznymi.

Nadzór człowieka

Kolejnym kluczowym elementem europejskiego podejścia regulacyjnego do sztucznej inteligencji, określonego w AI Act, jest zasada nadzoru człowieka (*human oversight*)[12]. Wymóg ten wynika z potrzeby zapewnienia, że decyzje podejmowane przez systemy AI, zwłaszcza te o wysokim ryzyku, są kontrolowane i weryfikowane przez człowieka. Celem jest ochrona praw podstawowych, ograniczenie ryzyka błędnych lub dyskryminujących decyzji oraz zwiększenie odpowiedzialności podmiotów wykorzystujących technologię[13].

Nadzór człowieka może przybierać różne formy w zależności od charakteru systemu i jego zastosowania. W przypadku systemów wysokiego ryzyka, np. w medycynie, finansach czy rekrutacji, nadzór obejmuje możliwość interwencji człowieka w procesie decyzyjnym, weryfikację wyników generowanych przez AI oraz, w razie potrzeby, wyłączenie systemu. Mechanizmy te mają na celu nie tylko zapobieganie

[11] M. Namysłowska *et al.*, *O etycznych, prawnych i społecznych konsekwencjach stosowania systemów sztucznej inteligencji w państwach członkowskich Unii Europejskiej. Uwagi na tle projektu rozporządzenia w sprawie sztucznej inteligencji*, „Przegląd Sejmowy" 2023, nr 6(179), s. 66-67.

[12] AI Act, art. 14.

[13] J. Kulesza, *Europe's Regulatory Sovereignty in the Age of Artificial Intelligence. A Human-Centric Response to the US-China Technological Rivalry*, „Sprawy Międzynarodowe" 2025, t. 78, nr 4, s. 177 i nast.

szkodom, ale również umożliwienie wykrywania błędów algorytmicznych i monitorowania wpływu systemów AI na prawa jednostki[14].

Wymóg nadzoru człowieka wiąże się ściśle z innymi zasadami regulacyjnymi, takimi jak transparentność i odpowiedzialność. Aby człowiek mógł skutecznie kontrolować system, musi mieć dostęp do informacji o sposobie działania algorytmu, kryteriach podejmowania decyzji oraz jakości danych wykorzystanych do trenowania modelu. W literaturze naukowej temat nadzoru człowieka w AI jest nadal w początkowej fazie badań, zwłaszcza w kontekście jego praktycznej realizacji w różnorodnych sektorach gospodarki i administracji publicznej. Wyzwania obejmują m.in. określenie optymalnego stopnia interwencji człowieka, uniknięcie nadmiernego obciążenia operacyjnego oraz minimalizowanie ryzyka automatyzacji decyzji bez odpowiedniego nadzoru[15].

Nadzór człowieka jest fundamentem odpowiedzialnego stosowania systemów sztucznej inteligencji. Zapewnia równowagę między automatyzacją procesów a ochroną praw jednostki, wzmacnia zaufanie społeczne do technologii oraz umożliwia skuteczne monitorowanie i kontrolę działania AI. Jego prawidłowe wdrożenie stanowi kluczowy warunek bezpieczeństwa, etycznego stosowania i legalności systemów sztucznej inteligencji w środowisku europejskim.

Jakość danych

Innym ważnym elementem skutecznej i odpowiedzialnej regulacji systemów sztucznej inteligencji, określonej w AI Act, jest wymóg zapewnienia wysokiej jakości danych wykorzystywanych do trenowania, testowania i działania modeli AI[16]. Dane stanowią fundament rzetelnego funkcjonowania systemów sztucznej inteligencji, ponieważ decyzje podejmowane przez algorytmy są bezpośrednio zależne od jakości informacji, na których zostały oparte. Błędne, niekompletne lub stronnicze dane mogą prowadzić do decyzji niezgodnych z prawem, dyskryminujących lub potencjalnie szkodliwych dla jednostek i społeczeństwa.

Regulacja europejska kładzie nacisk na to, aby dane używane w systemach AI były dokładne, reprezentatywne, kompletne i wolne od uprzedzeń. Oznacza to konieczność stosowania zaawansowanych procedur weryfikacji danych, oczyszczania zbiorów danych z błędów i anomalii, a także monitorowania jakości danych w czasie rzeczywistym. Uwzględnienie różnorodności demograficznej, społecznej i geograficznej w zbiorach danych jest istotne nie tylko dla poprawności działania

[14] M. Ebers, *op. cit.*, s. 16-17.

[15] B. Kaczmarek-Templin, *Sztuczna inteligencja (AI) i perspektywy jej wykorzystania w postępowaniu przed sądem cywilnym*, „Studia Prawnicze. Rozprawy i Materiały" 2022, nr 2(31), s. 74-75.

[16] AI Act, art. 10.

algorytmów, lecz także dla ograniczenia ryzyka powstawania uprzedzeń algorytmicznych, które mogłyby prowadzić do nierównego traktowania użytkowników lub grup społecznych[17].

Wysoka jakość danych ma także fundamentalne znaczenie dla przejrzystości i kontroli systemów AI. Tylko systemy o danych wysokiej jakości pozwalają na skuteczną kontrolę nad algorytmami oraz identyfikację potencjalnych błędów i ryzyk. Transparentność źródeł danych oraz sposobu ich przetwarzania umożliwia organom nadzoru, ekspertom technicznym i użytkownikom weryfikację zgodności działania systemu z obowiązującymi standardami prawnymi i etycznymi. W kontekście systemów wysokiego ryzyka, takich jak narzędzia wykorzystywane w medycynie, edukacji czy finansach, zapewnienie wysokiej jakości danych jest warunkiem umożliwiającym nadzór człowieka nad procesem decyzyjnym i minimalizowanie ryzyka szkód dla jednostek.

Literatura naukowa w obszarze jakości danych w AI jest wciąż w fazie rozwoju. Badania wskazują, że głównym wyzwaniem nie jest jedynie zapewnienie poprawności i kompletności danych, lecz także wykrywanie ukrytych uprzedzeń i asymetrii wynikających z historycznych nierówności społecznych lub specyfiki procesów zbierania danych[18]. Problemem pozostaje także brak jednolitych standardów oceny jakości danych oraz metodologii audytu stosowanych w różnych sektorach gospodarki i administracji publicznej. W praktyce twórcy systemów AI muszą balansować między pełną przejrzystością a ochroną własności intelektualnej, złożonością algorytmów oraz ochroną danych osobowych[19].

Zasada jakości danych jest również ściśle powiązana z innymi kluczowymi wymaganiami regulacyjnymi, takimi jak transparentność, nadzór człowieka i zarządzanie ryzykiem. System oparty na niskiej jakości danych ogranicza skuteczność nadzoru, utrudnia identyfikację uprzedzeń oraz zwiększa ryzyko podejmowania decyzji niezgodnych z prawem lub wartościami społecznymi. Dlatego zapewnienie wysokiej jakości danych jest niezbędne nie tylko dla prawidłowego działania systemów AI, lecz także dla budowania społecznego zaufania do technologii oraz odpowiedzialnego rozwoju innowacji w Unii Europejskiej.

Jakość danych stanowi fundament odpowiedzialnego i bezpiecznego stosowania systemów sztucznej inteligencji. Zapewnienie odpowiedniego standardu danych pozwala na rzetelne podejmowanie decyzji przez algorytmy, ogranicza ryzyko dyskryminacji, wspiera transparentność i umożliwia skuteczny nadzór człowieka.

---

[17] K. Sienkiewicz-Małyjurek, *Możliwości i problemy zastosowania sztucznej inteligencji w zarządzaniu kryzysowym,* „Bezpieczeństwo. Teoria i Praktyka" 2024, nr 1, s. 44-47.

[18] R. Rejmaniak, *Bias in Artificial Intelligence Systems*, „Białostockie Studia Prawnicze" 2021, nr 3 (vol. 26), s. 26-29.

[19] M. Namysłowska *et al.*, *op. cit.*, s. 50.

Niewiele jest jeszcze badań akademickich w tym obszarze, co wskazuje na potrzebę dalszych analiz i opracowania standardów jakości danych, które mogłyby być stosowane w praktyce przez twórców systemów AI oraz organy regulacyjne w różnych sektorach życia społecznego i gospodarczego.

Bezpieczeństwo i zarządzanie ryzykiem

Bezpieczeństwo i zarządzanie ryzykiem stanowią fundament odpowiedzialnego rozwoju i stosowania systemów sztucznej inteligencji. W regulacji europejskiej zagadnienia te zajmują centralne miejsce, szczególnie w kontekście systemów wysokiego ryzyka, których działanie może mieć istotny wpływ na prawa jednostki, bezpieczeństwo publiczne oraz interes społeczny[20]. Skuteczna implementacja regulacji z zakresu AI wymaga nie tylko innowacyjnych rozwiązań technologicznych, lecz także systematycznego zarządzania potencjalnymi zagrożeniami i zapewnienia bezpieczeństwa użytkowników.

Podejście oparte na bezpieczeństwie w AI obejmuje wszystkie etapy cyklu życia systemu – począwszy od projektowania, rozwoju, aż po wdrożenie i eksploatację. Na etapie projektowania konieczne jest przewidywanie możliwych scenariuszy zagrożeń oraz opracowywanie mechanizmów minimalizujących ryzyko awarii, błędów algorytmicznych czy niezamierzonych konsekwencji decyzji systemu. W fazie rozwoju i testowania systemu kluczowe znaczenie mają procedury walidacji, testowania odporności systemu na błędy oraz symulacje scenariuszy krytycznych. Po wdrożeniu systemu istotne jest natomiast monitorowanie jego działania w czasie rzeczywistym, identyfikacja potencjalnych nieprawidłowości oraz szybka reakcja w przypadku wystąpienia incydentów[21].

Zarządzanie ryzykiem w regulacji AI polega na systematycznej identyfikacji, ocenie i minimalizacji potencjalnych zagrożeń związanych z funkcjonowaniem systemu. AI Act wprowadza podejście oparte na klasyfikacji ryzyka, które pozwala dostosować poziom wymogów regulacyjnych do potencjalnego wpływu systemu na bezpieczeństwo i prawa jednostki. Systemy wysokiego ryzyka są objęte szczegółowymi obowiązkami dotyczącymi dokumentacji technicznej, audytu algorytmów, jakości danych oraz mechanizmów nadzoru człowieka, co pozwala na ograniczenie prawdopodobieństwa powstania szkód lub naruszeń praw podstawowych[22].

---

[20] AI Act, art. 9.

[21] D. Skoczylas, *Sztuczna inteligencja a dobrostan człowieka i ochrona praw podstawowych – rozważania na gruncie aktu w sprawie sztucznej inteligencji*, „Studia Prawnoustrojowe" 2025, nr 1(67), s. 357-358.

[22] AI Act, art. 11-12.

Bezpieczeństwo systemów AI wiąże się ściśle z innymi zasadami regulacyjnymi, takimi jak jakość danych, transparentność oraz nadzór człowieka. Niedokładne lub stronnicze dane mogą prowadzić do błędnych decyzji algorytmów. Brak transparentności ogranicza możliwość monitorowania działania systemu, a brak nadzoru człowieka zwiększa ryzyko niekontrolowanych skutków decyzji automatycznych. Dlatego skuteczne zarządzanie ryzykiem wymaga szerokiego podejścia, integrującego wszystkie powyższe aspekty oraz regularnej oceny ryzyka w trakcie całego cyklu życia systemu sztucznej inteligencji.

W literaturze przedmiotu dopiero zaczyna być szerzej analizowane zagadnienie bezpieczeństwa i zarządzania ryzykiem w kontekście praktycznej implementacji regulacji AI. Wskazuje się m.in. na potrzebę opracowania standardów audytu bezpieczeństwa, wytycznych w zakresie monitorowania systemów oraz metod oceny skuteczności działań minimalizujących ryzyko[23]. Szczególne wyzwania dotyczą dynamicznie rozwijających się modeli AI, takich jak generatywne systemy uczenia maszynowego, w przypadku których tradycyjne mechanizmy kontroli mogą być niewystarczające[24].

Podsumowując, bezpieczeństwo i zarządzanie ryzykiem stanowią nieodłączny element odpowiedzialnego wykorzystania sztucznej inteligencji. Wdrażanie skutecznych mechanizmów w tym zakresie jest niezbędne dla ochrony użytkowników i społeczeństwa, budowania zaufania do technologii oraz zapewnienia zgodności z europejskimi wartościami prawnymi i etycznymi. Przyszłe badania w tym obszarze będą kluczowe dla opracowania standardów i wytycznych, które umożliwią harmonijne połączenie innowacji technologicznych z bezpieczeństwem i ochroną praw jednostki.

Odpowiedzialność i dokumentacja

Odpowiedzialność za działania systemów sztucznej inteligencji oraz dokumentacja stosowania AI stanowią istotny wymóg prawny stawiany przez europejską regulację[25]. Zasady te mają na celu zapewnienie, że twórcy, dostawcy i użytkownicy systemów AI ponoszą odpowiedzialność za działanie technologii, a jednocześnie umożliwiają audyt, kontrolę i weryfikację ich funkcjonowania. W dobie rosnącej autonomii algorytmów odpowiedzialność prawna i dokumentacja stają się narzędziami ochrony użytkowników oraz budowania zaufania społecznego do nowych

[23] A. Pajdo, *op. cit.*, s. 49.
[24] C. Gutierrez, A. Aguirre, R. Uuk, C. Boine, M. Franklin, *A Proposal for a Definition of General Purpose Artificial Intelligence Systems*, „Digital Society" 2023, vol. 2, art. 36, s. 2-4.
[25] AI Act, art. 13-14.

technologii, a także mechanizmem ograniczania ryzyka prawnego i finansowego dla podmiotów rozwijających AI.

Podstawowym wymogiem w zakresie odpowiedzialności jest możliwość jednoznacznej identyfikacji podmiotów odpowiedzialnych za projektowanie, rozwój i wdrażanie systemów sztucznej inteligencji. Regulacja AI Act szczególnie podkreśla obowiązki dotyczące systemów wysokiego ryzyka, które mogą znacząco wpływać na prawa jednostki, bezpieczeństwo publiczne lub interes społeczny. Twórcy i dostawcy systemów zobowiązani są do wdrożenia procedur weryfikacji, testowania i monitorowania algorytmów oraz do prowadzenia szczegółowej dokumentacji technicznej[26]. Dokumentacja ta umożliwia audyt systemu, analizę ryzyka oraz przypisanie odpowiedzialności w przypadku błędów lub incydentów.

Dokumentacja techniczna obejmuje szeroki zakres informacji, w tym: architekturę systemu, wykorzystane metody uczenia maszynowego, szczegóły dotyczące zbiorów danych (źródła, jakość, reprezentatywność), wyniki testów i walidacji systemu, procedury monitorowania oraz mechanizmy zarządzania ryzykiem. W przypadku systemów wysokiego ryzyka zebrany materiał powinien również dokumentować interakcję człowieka z systemem, strategie nadzoru oraz procedury reagowania na sytuacje krytyczne[27]. Dzięki temu możliwe jest nie tylko identyfikowanie przyczyn ewentualnych błędów, ale także weryfikacja, czy system spełnia wymagania prawne i etyczne.

W literaturze zagadnienia odpowiedzialności i dokumentacji w AI dopiero zaczynają być szerzej analizowane[28]. Złożoność współczesnych systemów AI, w tym modeli generatywnych i autonomicznych, stwarza nowe wyzwania w zakresie przypisywania odpowiedzialności. W praktyce często trudne jest jednoznaczne określenie, który podmiot – twórca algorytmu, dostawca danych, integrator systemu czy użytkownik końcowy – odpowiada za skutki decyzji podejmowanych przez system. Ponadto dynamiczny charakter uczenia się algorytmów i zmienność danych utrudniają prowadzenie stałej dokumentacji oraz ocenę ryzyka w czasie rzeczywistym.

Odpowiedzialność i dokumentacja są ściśle powiązane z innymi zasadami regulacyjnymi, takimi jak transparentność, nadzór człowieka, jakość danych i bezpieczeństwo. Bez rzetelnej dokumentacji trudno zapewnić skuteczny nadzór człowieka,

---

[26] J. Laux, H. Ruschemeier, *Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI*, „European Journal of Risk Regulation" 2025, nr 16, s. 1522-1523.

[27] F. Doshi-Velez, M. Kortz, *Accountability of AI Under the Law: The Role of Explanation*, Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper 2017, s. 11-12.

[28] Zob. M. Świerczyński, Z. Więckowski, *Liability for damages caused by artificial intelligence systems – main challenges to be addressed by the European Union conflict-of-laws regulations*, „Prawo w Działaniu" 2023, t. 54, s. 200-216.

identyfikację uprzedzeń w danych oraz ocenę bezpieczeństwa systemu. Właściwa dokumentacja umożliwia również organom regulacyjnym kontrolę zgodności systemów z przepisami, a także służy jako podstawa do wprowadzenia mechanizmów odpowiedzialności cywilnej lub administracyjnej. W tym kontekście dokumentacja i odpowiedzialność pełnią zarówno funkcję prewencyjną, jak i kontrolną, zmniejszając ryzyko nieetycznych lub niebezpiecznych działań AI.

Odpowiedzialność i dokumentacja są niezbędnymi elementami bezpiecznego stosowania systemów sztucznej inteligencji. Zapewniają mechanizmy weryfikacji, audytu i kontroli, umożliwiają przypisanie odpowiedzialności podmiotom tworzącym i wdrażającym systemy AI, a także wspierają budowanie zaufania społecznego do technologii. W kontekście unijnej regulacji zasady te pozwalają łączyć innowacyjność z bezpieczeństwem, ochroną praw jednostki i zgodnością z wartościami demokratycznymi. Dalsze badania naukowe w tym obszarze są niezbędne dla opracowania standardów dokumentacji, audytu i mechanizmów odpowiedzialności, które będą mogły być stosowane w praktyce przez twórców systemów AI, organy regulacyjne oraz użytkowników końcowych, tworząc spójny i bezpieczny ekosystem sztucznej inteligencji.

Specjalne zasady dla modeli ogólnego przeznaczenia

Modele ogólnego przeznaczenia (ang. *General Purpose AI*, GPAI) stanowią jeden z najbardziej innowacyjnych i dynamicznie rozwijających się obszarów współczesnej sztucznej inteligencji[29]. W odróżnieniu od systemów wyspecjalizowanych, które zostały zaprojektowane do realizacji określonych zadań, modele GPAI charakteryzują się dużą elastycznością i możliwością adaptacji do różnych zastosowań w wielu sektorach gospodarki, administracji publicznej i życia codziennego. Ta wszechstronność niesie ze sobą zarówno ogromny potencjał innowacyjny, jak i zwiększone ryzyko wystąpienia nieprzewidzianych skutków, co wymaga wprowadzenia specjalnych zasad regulacyjnych.

Europejska regulacja sztucznej inteligencji przewiduje dla modeli ogólnego przeznaczenia wiele wymogów, które mają na celu zapewnienie ich bezpiecznego i odpowiedzialnego stosowania[30]. Ze względu na możliwość wykorzystywania modeli GPAI w wielu różnych kontekstach, kluczowe jest przewidywanie potencjalnych scenariuszy ich użycia oraz wprowadzenie mechanizmów ograniczających ryzyko. W praktyce oznacza to, że twórcy i dostawcy modeli muszą prowadzić

[29] I. Triguero, D. Molina, J. Poyatos, J. Del Ser, F. Herrera, *General Purpose Artificial Intelligence Systems (GPAIS): Properties, Definition, Taxonomy, Open Challenges and Implications*, „Information Fusion" 2024, vol. 103, s. 2-3.

[30] AI Act, rozdz. II-III.

szczegółową analizę ryzyka obejmującą różnorodne konteksty zastosowań, a także przewidywać możliwe interakcje systemu z innymi technologiami i procesami decyzyjnymi.

Wśród najważniejszych wymogów dla modeli ogólnego przeznaczenia wyróżnia się: rozszerzony nadzór człowieka, kompleksową dokumentację techniczną, szczegółową ocenę ryzyka oraz zapewnienie jakości danych[31]. Rozszerzony nadzór człowieka oznacza, że decyzje podejmowane przez modele GPAI muszą podlegać kontroli człowieka w każdym istotnym obszarze zastosowania, a mechanizmy interwencji muszą umożliwiać szybkie reagowanie na błędy lub nieprzewidziane skutki działania systemu. Dokumentacja techniczna obejmuje informacje o architekturze modelu, metodach uczenia, zbiorach danych, możliwościach adaptacyjnych, ograniczeniach systemu oraz procedurach monitorowania i reagowania na incydenty[32].

Bezpieczeństwo i przejrzystość stanowią szczególnie istotny aspekt stosowania modeli GPAI. Ich uniwersalność zwiększa ryzyko błędów lub nadużyć, dlatego dokumentacja i transparentność działania modelu są niezbędne dla zapewnienia odpowiedzialności podmiotów wdrażających technologię oraz dla ochrony użytkowników. Transparentność umożliwia audyt działania systemu, identyfikację potencjalnych błędów oraz ocenę wpływu decyzji algorytmów na prawa jednostki. W literaturze naukowej zwraca się uwagę, że brak właściwych mechanizmów nadzoru i monitorowania może prowadzić do poważnych konsekwencji prawnych, społecznych i etycznych, w tym do dyskryminacji, naruszenia prywatności lub niekontrolowanego wpływu na procesy decyzyjne[33].

Specjalne zasady dla modeli GPAI są również ściśle powiązane z innymi kluczowymi wymaganiami regulacyjnymi, takimi jak jakość danych, nadzór człowieka, zarządzanie ryzykiem i odpowiedzialność. Modele ogólnego przeznaczenia muszą być projektowane w sposób umożliwiający identyfikację i minimalizowanie ryzyka wynikającego z błędnych lub stronniczych danych, a także umożliwiać audyt i monitorowanie systemu w różnych zastosowaniach. Integracja tych zasad pozwala tworzyć systemy AI, które są nie tylko innowacyjne, ale także bezpieczne, etyczne i zgodne z wartościami europejskimi.

Zagadnienie regulacji prawnej modeli ogólnego przeznaczenia dopiero zaczyna być szerzej analizowane w literaturze przedmiotu. Badacze podkreślają, że elastyczność i uniwersalność GPAI sprawiają, że tradycyjne mechanizmy nadzoru i audytu mogą być niewystarczające. Istnieje potrzeba opracowania nowych metod oceny ryzyka, standardów dokumentacji oraz wytycznych dotyczących nadzoru człowieka,

[31] AI Act, art. 14-15.
[32] J. Laux, H. Ruschemeier, *op. cit*., s. 1525.
[33] M. Świerczyński, Z. Więckowski, *op. cit*., s. 209-211.

które uwzględniają wieloaspektowe zastosowania tych modeli[34]. Dalsze badania naukowe w tym obszarze będą kluczowe dla zrozumienia, w jaki sposób zapewnić odpowiedzialne stosowanie GPAI w dynamicznie zmieniającym się środowisku technologicznym.

Modele ogólnego przeznaczenia reprezentują obszar sztucznej inteligencji o ogromnym potencjale innowacyjnym, jak również podwyższonym poziomie ryzyka. Specjalne zasady przewidziane w AI Act – obejmujące rozszerzony nadzór człowieka, szczegółową dokumentację, ocenę ryzyka i jakość danych – mają na celu zapewnienie bezpieczeństwa, odpowiedzialności i zgodności z prawami jednostki. Skuteczne wdrożenie tych zasad wymaga ścisłej współpracy między twórcami systemów, organami regulacyjnymi i użytkownikami końcowymi, a także bieżącego monitorowania i oceny skutków stosowania GPAI. W kontekście dynamicznego rozwoju technologii dalsze badania naukowe są niezbędne do opracowania skutecznych metod nadzoru, audytu i regulacji, które pozwolą w pełni wykorzystać potencjał modeli ogólnego przeznaczenia w sposób bezpieczny, etyczny i zgodny z wartościami europejskimi.

Podsumowanie

Rozwój technologii sztucznej inteligencji stanowi jedno z najważniejszych wyzwań współczesnych systemów prawnych i społecznych. Rosnąca skala zastosowań systemów AI w różnych sektorach gospodarki oraz życia publicznego powoduje konieczność tworzenia odpowiednich mechanizmów regulacyjnych, które pozwolą na bezpieczne i odpowiedzialne wykorzystywanie technologii. W tym kontekście szczególne znaczenie ma inicjatywa legislacyjna Unii Europejskiej w postaci AI Act, której celem jest ustanowienie kompleksowych ram prawnych dla projektowania, wdrażania oraz wykorzystywania systemów sztucznej inteligencji.

Przeprowadzona analiza wskazuje, że europejski model regulacyjny opiera się przede wszystkim na podejściu opartym na analizie ryzyka, które umożliwia dostosowanie zakresu obowiązków regulacyjnych do potencjalnego wpływu danego systemu AI na prawa podstawowe jednostki, bezpieczeństwo oraz interes publiczny. Kluczowe znaczenie mają w tym zakresie zasady przejrzystości działania systemów sztucznej inteligencji, nadzoru człowieka nad procesami decyzyjnymi, odpowiedzialności podmiotów rozwijających i wdrażających technologie AI, a także wymogi dotyczące jakości danych i zarządzania ryzykiem. Zasady te mają na celu stworzenie ram umożliwiających rozwój technologii przy jednoczesnym ograniczeniu potencjalnych zagrożeń związanych z ich wykorzystaniem.

[34] P. Chyc, *Europejskie regulacje prawne dotyczące funkcjonowania sztucznej inteligencji*, „Prawo i Więź" 2025, nr 6(59), s. 207-209.

Należy jednak podkreślić, że problematyka prawnej regulacji sztucznej inteligencji oraz zasad funkcjonowania systemów AI wciąż znajduje się na stosunkowo wczesnym etapie rozwoju w literaturze przedmiotu. Ze względu na dynamiczny charakter postępu technologicznego oraz relatywnie niedawne przyjęcie europejskich regulacji prawnych, zagadnienia te dopiero zaczynają być szerzej analizowane w badaniach akademickich. W konsekwencji wiele kwestii związanych z praktycznym stosowaniem nowych regulacji, ich skutecznością oraz wpływem na rozwój innowacji pozostaje nadal otwartych i wymaga dalszych pogłębionych badań.

Można zatem stwierdzić, że europejskie podejście do regulacji sztucznej inteligencji stanowi istotny krok w kierunku budowy odpowiedzialnego i bezpiecznego systemu prawnego z zakresu nowych technologii. Jednocześnie dalszy rozwój refleksji naukowej nad zasadami regulującymi funkcjonowanie systemów AI będzie odgrywał kluczową rolę w ocenie skuteczności przyjętych rozwiązań prawnych oraz w kształtowaniu przyszłych kierunków rozwoju regulacji w tym obszarze.

Literatura

Chyc P., *Europejskie regulacje prawne dotyczące funkcjonowania sztucznej inteligencji*, „Prawo i Więź" 2025, nr 6 (59).

Doshi-Velez F., Kortz M., *Accountability of AI Under the Law: The Role of Explanation*, Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper 2017.

Ebers M., *Truly Risk-based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*, „European Journal of Risk Regulation" 2024, vol. 16, is. 2.

Gutierrez C., Aguirre A., Uuk R., Boine C., Franklin M., *A Proposal for a Definition of General Purpose Artificial Intelligence Systems*, „Digital Society" 2023, vol. 2, article 36.

Kaczmarek-Templin B., *Sztuczna inteligencja (AI) i perspektywy jej wykorzystania w postępowaniu przed sądem cywilnym*, „Studia Prawnicze. Rozprawy i Materiały" 2022, nr 2(31).

Kulesza J., *Europe's Regulatory Sovereignty in the Age of Artificial Intelligence. A Human-Centric Response to the US-China Technological Rivalry*, „Sprawy Międzynarodowe" 2025, t. 78, nr 4.

Laux J., Ruschemeier H., *Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI*, „European Journal of Risk Regulation" 2025, nr 16.

Namysłowska M. *et al.*, *O etycznych, prawnych i społecznych konsekwencjach stosowania systemów sztucznej inteligencji w państwach członkowskich Unii Europejskiej. Uwagi na tle projektu rozporządzenia w sprawie sztucznej inteligencji*, „Przegląd Sejmowy" 2023, nr 6(179).

Pajdo A., *Zakres obowiązków podmiotów odpowiedzialnych za systemy sztucznej inteligencji wysokiego ryzyka w świetle aktu o SI – Rozporządzenia 2024/1689*, „Roczniki Nauk Prawnych" 2025, nr 2(35).

Rejmaniak R., *Bias in Artificial Intelligence Systems*, „Białostockie Studia Prawnicze" 2021, nr 3 (vol. 26).

Skoczylas D., *Sztuczna inteligencja a dobrostan człowieka i ochrona praw podstawowych – rozważania na gruncie aktu w sprawie sztucznej inteligencji*, „Studia Prawnoustrojowe" 2025, nr 1(67).

Sienkiewicz-Małyjurek K., *Możliwości i problemy zastosowania sztucznej inteligencji w zarządzaniu kryzysowym*, „Bezpieczeństwo. Teoria i Praktyka" 2024, nr 1.

Świerczyński M., Więckowski Z., *Liability for damages caused by artificial intelligence systems – main challenges to be addressed by the European Union conflict-of-laws regulations*, „Prawo w Działaniu", 2023, t. 54.

Triguero I., Molina D., Poyatos J., Del Ser J., Herrera F., *General Purpose Artificial Intelligence Systems (GPAIS): Properties, Definition, Taxonomy, Open Challenges and Implications*, „Information Fusion" 2024, vol. 103.

Vladimirova K., *The Principle of Transparency Under The Artificial Intelligence Act*, „Law Journal of New Bulgarian University" 2024, nr 2 (vol. 20).

Westerstrand S., *Fairness in AI systems development: EU AI Act compliance and beyond*, „Information and Software Technology" 2025, vol. 187.

Wyszomirska M., *Implementation of the AI Act (EU) into the Polish legal system, its impact on this system and on selected strategic sectors*, „Safety & Fire Technology" 2025, nr 2(66).