RUCH PRAWNICZY, EKONOMICZNY I SOCJOLOGICZNY
Rok LXXXVIII – zeszyt 1 – 2026
https://doi.org/10.14746/rpeis.2026.88.1.05

PAULINA KUBERA[a]

ODPOWIEDZIALNOŚĆ ZA SZKODY WYNIKAJĄCE Z DZIAŁANIA SYSTEMÓW SZTUCZNEJ INTELIGENCJI

LIABILITY FOR HARM RESULTING FROM THE OPERATION OF ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) increasingly accompanies us in our professional and private lives. Alongside its undeniable benefits, it can also be a source of harm. The purpose of this article is to present the issue of legal liability for damage resulting from the operation of AI systems. The problem is significant because research indicates that liability for damage caused by AI constitute one of the key external obstacles to the adoption of artificial intelligence. The article identifies major challenges in determining liability, including: autonomy, constant adaptation, limited predictability, and lack of transparency. Using primarily the dogmatic-legal method, it outlines the general EU approach to regulating AI systems, which is based on risk analysis and management of the entire product life cycle, and presents the EU legal framework for liability for AI-related harm. The framework provides a two-track procedure for pursuing claims: under the liability of economic operators for defective products and under national tort liability systems based on the principle of fault. The article contributes to the discussion on the optimal regime of liability for harm caused by AI systems and evaluates the impact of these rules from the perspective of innovative enterprises and users. The study identifies key problem areas, including: partial discretion in classifying AI systems as high-risk, the regulatory focus on product safety risks with insufficient attention to other fundamental rights, and the lack of harmonization of national tort regimes resulting in different evidentiary standards adopted by national courts.

Keywords: artificial intelligence; high-risk AI systems; risk management; liability for harm; product defect

Sztuczna inteligencja (AI) coraz częściej towarzyszy nam w życiu zawodowym i prywatnym. Oprócz niewątpliwych dobrodziejstw, które oferuje, może być także źródłem szkód. Celem niniejszego artykułu jest przybliżenie kwestii odpowiedzialności prawnej za szkody wynikające z działania systemów sztucznej inteligencji. Problem jest doniosły, gdyż jak pokazują badania, kwestie odpowiedzialności za szkody wyrządzone przez AI stanowią jedną z kluczowych przeszkód zewnętrznych we wdrażaniu systemów sztucznej inteligencji. W artykule wskazano najważniejsze wyzwania w zakresie ustalania odpowiedzialności za szkody wyrządzone przez AI, takie jak: autonomiczność, ciągła adaptacja, ograniczona przewidywalność i brak przejrzystości. Wykorzystując głównie metodę dogmatycznoprawną, zarysowano podejście UE do regulacji systemów sztucznej inteligencji, które oparte jest na analizie ryzyka oraz zarządzaniu całym

[a] Poznan University of Technology, Poland / Politechnika Poznańska, Polska
paulina.kubera@put.poznan.pl, https://orcid.org/0000-0002-6246-6952

RPEiS 88(1), 2026: 87–106. *ISSN (Online) 2543-9170 ISSN (Print) 0035-9629*

cyklem życia produktu, oraz dokonano przeglądu unijnych ram prawnych odpowiedzialności za szkody wyrządzone przez AI. Ustanawiają one dwutorowy tryb dochodzenia roszczeń za szkody, które wynikły z działania systemów sztucznej inteligencji w ramach (1) odpowiedzialności podmiotów gospodarczych za produkty wadliwe oraz (2) krajowych systemów odpowiedzialności deliktowej opartej na zasadzie winy. Artykuł wnosi wkład w dyskusję nad optymalnym reżimem odpowiedzialności za szkody wyrządzone przez systemy sztucznej inteligencji, oceniając skutki regulacji z perspektywy innowacyjnych przedsiębiorstw i użytkowników. W opracowaniu zidentyfikowano w szczególności następujące obszary problemowe: częściową uznaniowość w klasyfikowaniu systemów AI jako wysokiego ryzyka, koncentrację regulacji na zagrożeniach związanych z bezpieczeństwem produktów przy jednoczesnym niedostatecznym uwzględnieniu ryzyk dla innych praw podmiotowych, a także brak harmonizacji krajowych reżimów odpowiedzialności deliktowej skutkujący odmiennymi standardami dowodowymi przyjmowanymi przez sądy krajowe.

Słowa kluczowe: sztuczna inteligencja; systemy AI wysokiego ryzyka; zarządzanie ryzykiem; odpowiedzialność odszkodowawcza; wadliwość produktu

I. WSTĘP

Sztuczna inteligencja coraz częściej towarzyszy nam w życiu zawodowym i prywatnym. Wkracza w różnorodne dziedziny, takie jak medycyna, transport czy przemysł. Szybko i trafnie opisuje badania RTG, wyniki tomografii komputerowej czy rezonansu magnetycznego, wspiera działanie pojazdów mechanicznych czy optymalizację procesów produkcyjnych i planowanie dystrybucji w łańcuchu dostaw. Dzięki zastosowaniu sztucznej inteligencji możemy cieszyć się nowymi możliwościami i usługami, które ułatwiają nam codzienne funkcjonowanie. Wirtualni asystenci przypominają o ważnych wydarzeniach, inteligentne termostaty, oświetlenie czy systemy bezpieczeństwa dostosowują się do naszych potrzeb i preferencji, pozwalając na oszczędność energii oraz zwiększenie naszego komfortu. Systemy sztucznej inteligencji (AI) są niejako wbudowane w określone produkty czy usługi i stanowią jedną całość lub są z nimi funkcjonalnie zintegrowane. Oprócz niewątpliwych dobrodziejstw, które oferują, mogą być jednak źródłem szkód.

Celem niniejszego artykułu jest przybliżenie kwestii odpowiedzialności prawnej za szkody wynikające z działania systemów sztucznej inteligencji. Problem jest doniosły, gdyż jak pokazują badania na temat stosowania technologii opartej na sztucznej inteligencji przeprowadzone w 2020 r. wśród europejskich przedsiębiorstw: „European enterprise survey on the use of technologies based on AI" (Komisja Europejska, 2020a) – kwestie odpowiedzialności za szkody wyrządzone przez sztuczną inteligencję stanowią jedną z najważniejszych przeszkód zewnętrznych we wdrażaniu systemów AI. Wynika to ze stopnia złożoności i skomplikowania systemów sztucznej inteligencji, które rodzą trudności w prześledzeniu procesu decyzyjnego danego systemu AI, a także nieprzejrzystość wielu algorytmów, które mogą powodować niepewność i utrudniać skuteczne egzekwowanie obowiązujących przepisów dotyczą-

cych bezpieczeństwa i ochrony praw podstawowych. Stąd zaistniała potrzeba wprowadzenia przepisów szczególnych mających na celu z jednej strony zwiększenie pewności prawa w zakresie narażenia zainteresowanych stron na odpowiedzialność, a z drugiej – ułatwienie dochodzenia roszczeń za szkody, które wynikły z działania systemów AI.

Artykuł składa się z sześciu części: część II – opisuje nowe wyzwania w zakresie ustalania odpowiedzialności za szkody wyrządzone przez systemy AI. Systemy te charakteryzują się określonym stopniem autonomiczności, czyli zdolnością do samodzielnego działania bez ingerencji człowieka, co rodzi dwa podstawowe pytania: pierwsze – związane jest z określeniem osób odpowiedzialnych za szkody, drugie – z przyjęciem optymalnego reżimu ich odpowiedzialności (na zasadzie winy czy surowszego, oderwanego od winy, podobnie jak w przypadku odpowiedzialności producentów za produkty niebezpieczne). Część III zarysowuje generalne podejście UE do regulacji systemów sztucznej inteligencji, które oparte jest na zarządzaniu ryzykiem oraz całym cyklem życia produktu, a IV analizuje unijne ramy prawne odpowiedzialności za szkody wyrządzone przez AI. Ustanawiają one dwutorowy tryb dochodzenia roszczeń za szkody, które wynikły z działania systemów sztucznej inteligencji: w ramach odpowiedzialności podmiotów gospodarczych za produkty wadliwe oraz w ramach krajowych systemów odpowiedzialności deliktowej opartych na zasadzie winy. Część V ukazuje perspektywę ekonomiczną zasad odpowiedzialności, w tym powiązanie wymogów bezpieczeństwa produktów i odpowiedzialności za produkt oraz ryzyka i odpowiedzialności. W podsumowaniu (VI) zawarto próbę oceny przyjętych i planowanych regulacji prawnych, tzn. czy udało się znaleźć równowagę między zapewnieniem skutecznej rekompensaty osobom poszkodowanym przez systemy AI a stworzeniem zachęt do innowacji i wdrażania systemów sztucznej inteligencji.

W artykule wykorzystano metodę analizy prawa oraz dokumentów programowych UE, w tym metody logiczno-językowe, metody argumentacyjne i hermeneutyczne oraz komparatystyczne. Artykuł wnosi wkład w dyskusję nad optymalnym reżimem odpowiedzialności za szkody wyrządzone przez systemy sztucznej inteligencji oraz stanowi próbę oceny skutków regulacji w zakresie odpowiedzialności za szkody wyrządzone przez systemy AI z punktu widzenia dostarczających je innowacyjnych przedsiębiorstw oraz ich użytkowników.

II. WYZWANIA W ZAKRESIE USTALANIA ODPOWIEDZIALNOŚCI ZA SZKODY WYRZĄDZONE PRZEZ AI

Systemy AI charakteryzują się postępującą autonomicznością, tj. ich funkcjonowanie jest coraz mniej zależne od kontrolującego je człowieka. Definicja systemu AI zawarta w rozporządzeniu PE i Rady (UE) z 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących

sztucznej inteligencji[1] wprost wskazuje na tę kluczową cechę systemu AI, jaką jest zdolność do wnioskowania, co ma odróżnić system AI od prostszych, bardziej tradycyjnych systemów oprogramowania lub założeń programistycznych, bądź systemów opartych na zasadach określonych wyłącznie przez człowieka w celu automatycznego wykonywania operacji. W przypadku systemów AI ważne są techniki umożliwiające wnioskowanie, takie jak mechanizmy uczenia maszynowego (na podstawie danych systemy uczą się, jak osiągnąć określone cele), czy podejścia oparte na logice i wiedzy (wnioskowanie na podstawie zakodowanej wiedzy lub symbolicznego przedstawienia zadania, które należy rozwiązać; zob. motyw 12 Aktu o AI). W myśl art. 3 par. 1 Aktu o AI „system AI oznacza 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".

Owa autonomiczność jest właśnie problematyczna z punktu widzenia przypisania odpowiedzialności prawnej za szkody wyrządzone przez systemy AI. W tym kontekście podkreślić należy, że Unia Europejska prezentuje podejście do sztucznej inteligencji oparte na humanocentryzmie, w którym człowiek stanowi punkt odniesienia. Człowiek jest zatem pierwotnym twórcą systemów opartych na AI, to człowiek korzysta z możliwości oferowanych przez systemy AI i to z punktu widzenia ryzyk i zagrożeń stwarzanych dla człowieka oceniane są systemy AI (Derave i in, 2022; Komisja Europejska, 2020b). Tym samym odrzuca się pogląd przemawiający za przyznaniem sztucznej inteligencji odrębnej podmiotowości prawnej (pogląd bodaj najbardziej dyskutowany w kontekście praw autorskich do utworów wytworzonych z udziałem systemów AI; np. Księżak i Wojtczak, 2021). Niemniej problem, kto jest odpowiedzialny za szkody wyrządzone przez system AI, jest dużo bardziej złożony niż w przypadku tradycyjnych produktów z uwagi na zakłócenie typowego podziału ról między producentem a użytkownikiem (zob. niżej, część V.2). Pojawić się może także tzw. problem wielu rąk (*problem of many hands* [PMH]), szczególnie w przypadku operacji na dużą skalę, gdzie działania mają charakter zbiorowy. W przypadku systemów AI wskazuje się na szerokie grono osób wpływających na ich funkcjonowanie: producenta, dostawcę, podmiot nadzorujący czy użytkownika tych systemów (de Bruyne i in., 2023; Michalak, 2019).

Oprócz wspomnianej wcześniej autonomiczności, tj. niezależności od ludzkich interwencji, połączonej z quasikognitywnymi i adaptacyjnymi zdolnościami AI, wśród najważniejszych ryzyk związanych z AI należy również wymienić brak przejrzystości oraz złożoność i stopień skomplikowania współczesnych jej systemów, które rodzą problemy w prześledzeniu procesu decyzyj-

[1] Rozporządzeniu PE i Rady (UE) z 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji, Dz. Urz. UE L 2024/1689, 12.7.2024 (dalej jako: Akt o AI).

nego. Wynik działania systemu AI w konkretnym przypadku może być trudny bądź prawie niemożliwy do przewidzenia (Yampolskiy, 2020). Powstaje zatem pytanie, komu przypisać odpowiedzialność, gdy wadliwe działanie systemów AI jest trudne do przewidzenia przez ich producentów, a obowiązki ich monitorowania przez użytkowników trudno jest zdefiniować. Zauważyć należy, że użytkownicy nie mają takiej kontroli nad autonomicznymi systemami sztucznej inteligencji jak w przypadku tradycyjnych produktów. Stąd mogą błędnie postrzegać ryzyko związane z korzystaniem ze sztucznej inteligencji (Buiten, 2024; Buiten et al., 2023).

Ponadto sztuczna inteligencja może rodzić szczególne zagrożenia. Tendencyjne, mało zróżnicowane dane treningowe wykorzystywane do uczenia się sztucznej inteligencji mogą skutkować rekomendacjami lub prognozami stronniczymi wobec niektórych grup, czyli prowadzić bądź utrwalać dyskryminację ze względu np. na rasę, płeć czy wiek (Hacker, 2023). Sztuczna inteligencja zwiększa także możliwość śledzenia codziennych poczynań ludzi, analizowania dużej ilości danych oraz znajdowania powiązań między nimi, a tym samym wymaga szczególnego podejścia w zakresie ochrony danych osobowych oraz prywatności (Rauccio, 2021).

III. PODEJŚCIE OPARTE NA ANALIZIE RYZYKA

Prawodawstwo unijne klasyfikuje systemy AI zgodnie z ryzykiem, jakie stwarzają one dla użytkowników. U podstaw tej klasyfikacji legło założenie, że im większy stopień ryzyka, tym wyższy stopień regulacji, nie wyłączając zakazu wykorzystywania niektórych systemów AI, jeśli ryzyko jest nieakceptowalne.

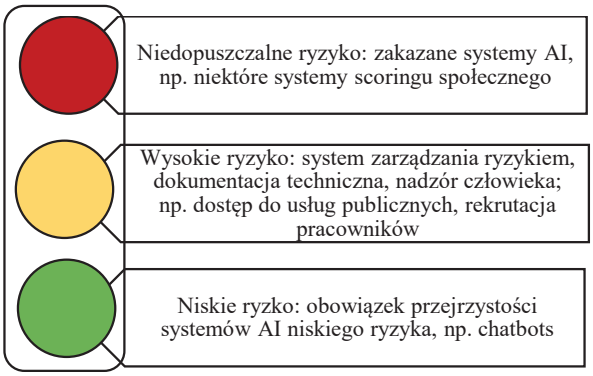
W myśl Aktu o AI: „»ryzyko« oznacza połączenie prawdopodobieństwa wystąpienia szkody oraz jej dotkliwości" (art. 6 par. 2). W odniesieniu do rodzaju ryzyka, które należy uwzględnić w ocenie danego systemu AI, obejmują one dwie podstawowe kategorie: (1) ryzyko dla zdrowia i bezpieczeństwa (*health and safety risks*), oraz (2) ryzyko dla praw podstawowych (*risks to fundamental rights*).

Zakaz wprowadzania do obrotu, oddawania do użytku i wykorzystywania objął systemy AI, które spełniają określone kryteria praktyk zakazanych. Zalicza się do nich m.in. stosowanie technik podprogowych, manipulacyjnych lub wprowadzających w błąd, które mogłyby być wykorzystywane w celach od subtelnego wpływania na zachowania konsumentów po manipulację polityczną. Wśród zakazanych praktyk znalazł się także scoring społeczny, czyli klasyfikacja osób fizycznych na podstawie ich zachowań społecznych lub cech osobistych, jeżeli prowadzi do krzywdzącego lub niekorzystnego traktowania niektórych osób lub grup społecznych, z uwagi na jego wykorzystanie w innych kontekstach społecznych niż te, w których pierwotnie je zebrano, bądź z uwagi na nieproporcjonalność do ich zachowania społecznego lub jego wagi (art. 5 Aktu o AI). Zasadniczo katalog zakazów objął te systemy AI, które mogłyby

godzić w godność i autonomię jednostki, wykorzystywać słabości określonych osób fizycznych i grup społecznych, naruszać prywatność i nieść ryzyko szeroko zakrojonego nadzoru i profilowania osób (Ren i Du, 2024).

Rysunek 1

Klasyfikacja systemów AI według ryzyka



Źródło: opracowanie własne.

Z kolei za systemy AI wysokiego ryzyka uznaje się w szczególności te systemy, które wymienione są w załączniku III do Aktu o AI. Zawiera on konkretne przeznaczenia systemów AI w takich newralgicznych obszarach, jak: biometria, infrastruktura krytyczna, kształcenie i szkolenie zawodowe, zatrudnienie, zarządzanie pracownikami i dostęp do samozatrudnienia, dostęp do podstawowych usług prywatnych i publicznych, ściganie przestępstw, zarządzanie migracją, azylem i kontrolą graniczną, sprawowanie wymiaru sprawiedliwości i procesów demokratycznych. Za systemy wysokiego ryzyka mogą być przykładowo uznane: systemy AI wykorzystywane do podejmowania decyzji o przyjęciu do instytucji edukacyjnych lub szkolenia zawodowego; rekrutacji osób fizycznych, w szczególności filtrowania podań o pracę i oceny kandydatów; czy wykorzystywane przez organy publiczne do kwalifikacji osób fizycznych do podstawowych usług publicznych. Decydujące znaczenie w kwalifikacji danego systemu AI jako systemu wysokiego ryzyka ma to, w jakim stopniu może on negatywnie wpływać na ochronę praw podstawowych wymienionych i zagwarantowanych w Karcie praw podstawowych UE[2]. A zatem jeśli dany system AI wprawdzie wykorzystywany jest w obszarach i do celów wyżej wymienionych, ale przeznaczony jest jedynie do: wykonania wąsko opisanego zadania określonej procedury; poprawienia wyniku czynności uprzednio zakończonej przez człowieka, wykrywania wzorców decyzji podejmowanych przez człowieka czy zadań przygotowawczych, nie traktuje się go jako systemu wysokiego ryzyka.

[2] Karta praw podstawowych Unii Europejskiej, Dz. Urz. UE C, 326/291, 26.10.2012.

Ponadto za systemy AI wysokiego ryzyka uznaje się systemy, które łącznie spełniają dwie przesłanki odnoszące się do zdrowia i bezpieczeństwa produktów. Systemy AI są przeznaczone do stosowania jako produkt lub element produktu, który ze względów bezpieczeństwa został objęty unijnym prawodawstwem harmonizacyjnym i w związku z tym podlega ocenie zgodności przez stronę trzecią w związku z wprowadzeniem go do obrotu lub oddaniem do użytku. (Dodać należy, że w sytuacji uznania, że system AI jest elementem produktu, istotne jest to, by system AI związany był właśnie z bezpieczeństwem produktu).

Konsekwencją uznania danego systemu AI za system wysokiego ryzyka jest szereg wymogów, wśród których należy wymienić m.in.: ustanowienie, wdrożenie, dokumentowanie i utrzymanie systemu zarządzania ryzykiem, który obejmuje cały cykl życia systemu AI i wymaga regularnego i systematycznego przeglądu i aktualizacji (art. 9 Aktu o AI); zapewnienie odpowiedniej jakości danych treningowych, walidacyjnych i testowych (art. 10 Aktu o AI); opracowanie dokumentacji technicznej systemu AI przed wprowadzeniem go do obrotu (lub oddaniem do użytku) oraz jej aktualizacja (art. 11 Aktu o AI); zapewnienie funkcji automatycznego rejestrowania zdarzeń w całym cyklu życia danego systemu AI (art. 12 i 19 Aktu o AI); zapewnienie przejrzystych informacji dla użytkowników, w tym stworzenie instrukcji obsługi zawierającej m.in. informacje umożliwiające interpretację wyników systemu i ich właściwe wykorzystanie (art. 13 Aktu o AI); projektowanie i opracowywanie systemów w sposób umożliwiający ich nadzorowanie przez osoby fizyczne (art. 14 Aktu o AI) oraz zapewnienie dokładności, solidności, cyberbezpieczeństwa systemu AI (odporność na błędy, redundancja, ograniczenie nieuprawnionego dostępu; art. 15 Aktu o AI). Analizując powyższe wymogi z punktu widzenia ich dotkliwości dla dostawców systemów AI wysokiego ryzyka, należy wskazać przede wszystkim na trudne do spełnienia wymogi odnoszące się do jakości zbiorów danych służących do trenowania, walidacji oraz testowania systemów AI. Mają być one adekwatne, wystarczająco reprezentatywne oraz w jak największym stopniu wolne od błędów i kompletne z punktu widzenia przeznaczenia. Muszą zatem charakteryzować się odpowiednimi właściwościami statystycznymi, uwzględniać cechy lub elementy charakterystyczne dla określonego środowiska geograficznego, kontekstowego, behawioralnego lub funkcjonalnego, w którym dany system AI ma być wykorzystywany. Stąd potrzeba wypracowania wytycznych oraz kodeksów dobrych praktyk, które zawierałyby uszczegółowienia tych wymogów. Z kolei pozytywnie należy ocenić możliwość przygotowania uproszczonej dokumentacji technicznej przez mikro, małe i średnie przedsiębiorstwa, która ma dostarczyć właściwym organom informacji niezbędnych do oceny zgodności systemu AI z wymogami Aktu o AI.

Natomiast wymogi dotyczące systemów AI, które nie stwarzają wysokiego ryzyka, obejmują przede wszystkim obowiązki w zakresie przejrzystości i informacji. Systemy AI przeznaczone do wchodzenia w interakcję z osobami fizycznymi powinny być tak zaprojektowane i rozwijane, by osoby fizyczne były świadome, że wchodzą w interakcję z maszyną. Na przykład podczas korzystania z systemów sztucznej inteligencji, takich jak chatboty, należy poin-

formować osobę fizyczną, że wchodzi w interakcję z chatbotem, a nie z drugim człowiekiem. Pomaga to uniknąć nieporozumień i nie nadużywa zaufania (Følstad i in., 2023; Skjuve in., 2019). Ponadto dostawcy muszą zapewnić możliwość identyfikacji treści generowanych przez sztuczną inteligencję (art. 50 Akt o AI; zob. także Knott i in., 2024).

W Akcie o AI uwzględniono także ryzyko systemowe, które może wyniknąć z działania systemów AI ogólnego przeznaczenia, w tym przede wszystkim dużych generatywnych modeli AI. Są one trenowane na dużej ilości danych z wykorzystaniem nadzoru własnego na dużą skalę. Wykazują znaczną ogólność i są w stanie kompetentnie wykonywać szeroki zakres różnych zadań, niezależnie od sposobu, w jaki zostały wprowadzane do obrotu. Można je zintegrować z różnymi systemami lub aplikacjami niższego szczebla. W związku z powyższym odznaczają się zdolnością dużego oddziaływania. Dostawcy wszystkich modeli AI ogólnego przeznaczenia są zobowiązani m.in. do ujawniania określonych informacji wszystkim tym dostawcom, którzy zamierzają zintegrować model AI ogólnego przeznaczenia ze swoimi systemami AI, w celu zapewnienia przejrzystości oraz lepszego zrozumienia możliwości i ograniczeń danego modelu AI. Powinni także stosować zasady zapewniające przestrzeganie prawa autorskiego przy trenowaniu swoich modeli. Ponadto gdy modele AI ogólnego przeznaczenia stwarzają ryzyko systemowe, w grę wchodzą dodatkowe wymogi w zakresie oceny i ograniczania ryzyka systemowego, zgłaszania poważnych incydentów, przeprowadzania najnowocześniejszych testów i oceny tych modeli oraz zapewnienia odpowiedniego poziomu cyberbezpieczeństwa.

IV. DWUTOROWY SYSTEM OCHRONY

Prawodawstwo unijne zakłada dwutorowy tryb dochodzenia odszkodowania za szkody, które wynikły z działania systemów sztucznej inteligencji. Zostały one zawarte w nowej dyrektywie PE i Rady (UE) 2024/2853 z 23 października 2024 r. w sprawie odpowiedzialności za produkty wadliwe i uchylenia dyrektywy Rady 85/374/EWG[3], która reguluje niezależną od winy odpowiedzialność producenta za określone rodzaje szkód doznanych głównie przez osoby fizyczne. Ma ona zastosowanie do produktów – rzeczy ruchomych, także oprogramowania, w tym systemów AI, wprowadzanych do obrotu lub oddawanych do użytku (z wyjątkiem darmowego i otwartego oprogramowania opracowanego lub dostarczonego poza działalnością handlową). Jej uzupełnienie miała stanowić propozycja dyrektywy PE i Rady w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji (AILD). Dotyczy ona krajowych roszczeń z tytułu odpowiedzialności opartej na zasadzie winy i miała służyć zapewnieniu odszkodowania niezależnie od rodzaju szkody i osoby poszkodowanego. Nie osiągnięto jednak

[3] Dyrektywa PE i Rady (UE) 2024/2853 z 23 października 2024 r. w sprawie odpowiedzialności za produkty wadliwe i uchylenia dyrektywy Rady 85/374/EWG, Dz. Urz. UE L, 2024/2853, 18.11.2024 (dalej jako: PLD).

konsensusu i propozycja AILD została odrzucona przynajmniej w najbliższej perspektywie czasowej. O tym, czego nie udało się przeforsować i jakie są tego konsekwencje, będzie mowa w dalszej części artykułu.

Natomiast 8 grudnia 2024 r. weszła w życie dyrektywa PLD (najważniejsze zmiany wprowadzone tą dyrektywą przedstawiono w tabeli 1). Jej zakres przedmiotowy oraz podmiotowy został rozszerzony w porównaniu do jej poprzedniczki – dyrektywy 85/374/EWG z 25 lipca 1985 r. w sprawie zbliżenia przepisów ustawowych, wykonawczych i administracyjnych Państw Członkowskich dotyczących odpowiedzialności za produkty wadliwe[4]. Jest to konsekwencją chęci dostosowania zasad odpowiedzialności za produkty wadliwe do wyzwań związanych z transformacją cyfrową i gospodarką o obiegu zamkniętym. W definicji „produktu" uwzględniono bowiem cyfrowe pliki produkcyjne oraz oprogramowanie, a także wskazano, kiedy powiązaną usługę należy uznać za część składową produktu. Tym samym rozstrzygnięto jednoznacznie kwestię, że system AI może stanowić zarówno samo oprogramowanie (*software based AI-system*), jak i oprogramowanie przechowywane w urządzeniu (*software embedded in hardware devices*). Jest ono produktem, niezależnie od sposobu jego dostarczenia lub użytkowania (motyw 13 dyrektywy PLD). Co się zaś tyczy pojęcia „szkoda" – zostało ono rozszerzone o szkody niematerialne, a także szkody w postaci zniszczenia lub uszkodzenia danych, takich jak pliki cyfrowe usunięte z dysku twardego, w tym koszty odzyskania lub przywrócenia tych danych, z zastrzeżeniem, że nie są one wykorzystywane do celów zawodowych (art. 6 ust. 1 dyrektywy PLD). Natomiast okoliczności, które uwzględnia się przy ocenie wadliwości produktu, zostały rozbudowane o: „wpływ na produkt wszelkich zdolności do dalszego uczenia się lub nabywania nowych cech po wprowadzeniu go do obrotu lub oddaniu do użytku" (art. 7c dyrektywy PLD); a także „dający się racjonalnie przewidzieć wpływ na produkt innych produktów, co do których można oczekiwać, że będą stosowane razem z produktem, w tym w drodze wzajemnego połączenia" (art. 7d dyrektywy PLD). Wynikają z tego dwie doniosłe konsekwencje. W świetle nowych regulacji istnieje możliwość uznania produktu za wadliwy z uwagi na luki w cyberbezpieczeństwie, jeśli produkt nie spełnia obowiązkowych wymogów w zakresie cyberbezpieczeństwa określonych w prawie UE lub prawie krajowym. W kontekście internetu rzeczy (IoT), w którym obiekty fizyczne (maszyny, produkty czy urządzenia) łączą się i wymieniają dane między sobą za pośrednictwem internetu – atak skierowany na jeden obiekt może również wpływać na inne urządzenia w tym samym systemie IoT, a przez to prowadzić do szkody znacznych rozmiarów. Po drugie, wadliwość może powstać po wprowadzeniu produktu do obrotu, gdy związana jest np. z aktualizacją oprogramowania. W konsekwencji wadliwość produktu rodząca odpowiedzialność producenta nie musi istnieć w momencie wprowadzania go do obrotu, ale może pojawić się później, jeśli produkt znajdował się pod jego kontrolą (zob. część V.2).

---

[4] Dyrektywa Rady 85/374/EWG z dnia 25 lipca 1985 r. w sprawie zbliżenia przepisów ustawowych, wykonawczych i administracyjnych Państw Członkowskich dotyczących odpowiedzialności za produkty wadliwe, Dz. Urz. UE L, 210, 7.8.1985.

W dyrektywie PLD rozszerzono także katalog podmiotów, które mogą być pociągnięte do odpowiedzialności za szkody wyrządzone przez produkt wadliwy. W sytuacji gdy producent produktu lub jego części składowej nie ma siedziby w UE, do odpowiedzialności może zostać pociągnięty importer lub upoważniony przedstawiciel producenta, a w ich braku – dostawca usług realizacji zamówień, którym jest osoba fizyczna lub prawna, która – nie będąc jego właścicielem – w ramach działalności handlowej świadczy co najmniej dwie z następujących usług: magazynowanie, pakowanie, adresowanie i wysyłanie produktu, z wyłączeniem usług pocztowych, doręczania paczek lub usług transportu towarów. Uwzględnienie dostawców usług realizacji zamówień w katalogu podmiotów odpowiedzialnych wynika z faktu, iż często pełnią oni wiele tych samych funkcji co importerzy, umożliwiając i ułatwiając dostęp do rynku unijnego dla produktów z państw trzecich, ale nie wpisują się łatwo w tradycyjne łańcuchy dostaw zgodnie z obowiązującymi ramami prawnymi.

Jeśli nie można zidentyfikować podmiotu gospodarczego spośród tych wymienionych wcześniej i mających siedzibę w UE, odpowiedzialność ponosi każdy dystrybutor wadliwego produktu (tj. osoba fizyczna lub prawna w łańcuchu dostaw, która udostępnia produkt na rynku, nie będąc jego producentem lub importerem), w sytuacji gdy osoba poszkodowana zwróci się do niego z wnioskiem o wskazanie producenta, importera, upoważnionego przedstawiciela producenta lub dostawcy realizacji zamówień mających siedzibę w UE, a dystrybutor ten nie wskaże ich (lub swojego dystrybutora mającego siedzibę w UE), w terminie jednego miesiąca od otrzymania wniosku. Co istotne, zasada ta odnosi się także do dostawców platformy internetowej, która umożliwia konsumentom zawieranie umów na odległość z przedsiębiorcami. Nowa regulacja ma zagwarantować, by poszkodowany mógł dochodzić rekompensaty, i to niezależnie od tego, czy produkt został wyprodukowany w UE (art. 8 dyrektywy PLD).

Odnosząc się zaś do zasad odpowiedzialności za szkodę wyrządzoną przez produkt wadliwy, należy wskazać przede wszystkim na wprowadzenie szeregu ułatwień w dochodzeniu roszczeń dla poszkodowanych. Zgodnie bowiem z ogólną zasadą to na poszkodowanym spoczywa ciężar dowodu takich okoliczności, jak: wadliwość produktu, poniesiona szkoda oraz istnienie związku przyczynowego między wadliwością produktu a poniesioną szkodą. W przypadku systemów AI ciężar dowodu spoczywający na poszkodowanym może okazać się wyjątkowo trudny i kosztowny, z uwagi na wcześniej wspomnianą autonomiczność systemów AI, ograniczoną przewidywalność i brak przejrzystości. Trudno jest mu bowiem ustalić np., który kod (źródłowy) lub które dane wejściowe wywołały szkodę. Stąd dyrektywa PLD zobowiązuje producenta do ujawnienia przed sądem niezbędnych informacji, którymi dysponuje, gdy poszkodowany przedstawi „fakty i dowody wystarczające do wykazania wiarygodności roszczenia o odszkodowanie" (art. 9 ust. 1). Wykonanie tego obowiązku ma się odbywać z poszanowaniem ochrony tajemnicy handlowej oraz poufności. Ponadto dyrektywa ustanawia, pod pewnymi warunkami, domniemanie wadliwości produktu oraz związku przyczynowego między wadliwością produktu a szkodą. I tak – wadliwości produktu domniemywa się, gdy: (1) pozwany (producent) nie dopełnia obowiązku ujawnienia odpowiednich dowodów, o których mowa powy-

żej, (2) powód (poszkodowany) wykaże, że produkt nie spełnia obowiązkowych wymogów bezpieczeństwa określonych prawem krajowym lub UE, oraz gdy (3) powód wykaże, że szkoda została spowodowana oczywistym nieprawidłowym działaniem produktu. Należy jednak zauważyć, że wykazanie niezgodności produktu z obowiązkowymi wymogami bezpieczeństwa oraz istnienia oczywistej nieprawidłowości jego działania wraz ze wzrostem złożoności technologicznej produktów staje się coraz trudniejsze i może wymagać angażowania wyspecjalizowanych ekspertów. Znowelizowana dyrektywa PLD nie definiuje pojęcia *oczywistej nieprawidłowości działania produktu*, ograniczając się w motywie 46 do przykładu wybuchu szklanej butelki podczas normalnego jej użytkowania, który to przykład z pewnością nie oddaje złożoności problemu w środowisku cyfrowym. W konsekwencji sądy krajowe kształtują to pojęcie kazuistycznie, co prowadzi do zwiększonej niepewności prawnej (Li i Schütte, 2023).

Kolejnym domniemaniem jest zakładanie związku przyczynowego pomiędzy wadliwością produktu a szkodą w przypadku ustalenia, że produkt jest wadliwy, a powstała szkoda odpowiada uszczerbkom zwykle powstającym przy danego rodzaju wadzie (art. 10 ust. 3 dyrektywy PLD).

Tabela 1

Nowe zasady odpowiedzialności za wadliwość produktu wprowadzone dyrektywą PLD

Zakres przedmiotowy	Zakres podmiotowy	Zasady odpowiedzialności za szkodę
Rozszerzenie definicji produktu o cyfrowe pliki produkcyjne oraz oprogramowanie	Uwzględnienie w katalogu podmiotów odpowiedzialnych dostawców usług realizacji zamówień, którzy choć pełnią wiele tych samych funkcji co importerzy, nie wpisują się łatwo w tradycyjne łańcuchy dostaw zgodnie z obowiązującymi ramami prawnymi.	Wprowadzenie obowiązku ujawnienia dowodów przez pozwanego (producenta) w sytuacji, gdy poszkodowany przedstawi „fakty i dowody wystarczające do wykazania wiarygodności roszczenia o odszkodowanie".
Rozszerzenie pojęcia szkody o szkody niematerialne	Przez pojęcie dystrybutora rozumie się także dostawców platformy internetowej, która umożliwia konsumentom zawieranie umów na odległość z przedsiębiorcami.	Wprowadzenie domniemania wadliwości produktu.
Rozszerzenie katalogu okoliczności branych pod uwagę przy ocenie wadliwości produktu o funkcje samouczenia się oraz powiązania między produktami, a także luki w cyberbezpieczeństwie produktu	Producent odpowiada za wadliwość produktu, która powstała także po wprowadzeniu go do obrotu, jeżeli produkt pozostawał pod jego kontrolą, np. gdy dostarcza aktualizację oprogramowania.	Wprowadzenie domniemywania związku przyczynowego pomiędzy wadliwością produktu a szkodą.

Źródło: opracowanie własne.

Wycofana przez Komisję Europejską dyrektywa UE w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji (AILD) miała na celu zapewnienie ochrony za szkody wynikające z działania systemów AI w przypadkach nieobjętych zakresem dyrektywy PLD przez zmniejszenie ciężaru dowodu w przypadku roszczeń odszkodowawczych dochodzonych w ramach krajowych systemów odpowiedzialności opartych na zasadzie winy. Podobnie jak w dyrektywie PLD, służyć temu miały: wprowadzenie obowiązku ujawnienia dowodów przez pozwanego oraz ustanowienie domniemań wzruszalnych. Ostatecznie jednak nie udało się jej przeforsować, m.in. pod wpływem nacisków firm technologicznych oraz coraz bardziej słyszalnych głosów opowiadających się za deregulacją technologii AI. W związku z czym spory dotyczące szkód wyrządzonych przez sztuczną inteligencję nieobjęte zakresem dyrektywy PLD będą rozstrzygane na podstawie przepisów krajowych ustanawiających różne standardy dowodowe.

V. PERSPEKTYWA EKONOMICZNA ZASAD ODPOWIEDZIALNOŚCI ZA SZKODY WYRZĄDZONE PRZEZ AI

1. Bezpieczeństwo i odpowiedzialność jako dwie strony medalu

Z założenia przyjęte zasady odpowiedzialności mają być środkiem minimalizowania ryzyka wypadków przez zachęcanie do stosowania przez podmiot odpowiedzialny odpowiednich środków zapobiegawczych, np. producenta do poprawy bezpieczeństwa produktu (Buiten, 2024). Problem wpływu zasad odpowiedzialności na zachęty do ograniczenia ryzyka wypadków i związanych z nim szkód jest szeroko dyskutowany w literaturze. Wskazuje się na istotną rolę zasad odpowiedzialności w kształtowaniu zachęt do zmniejszenia ryzyka wypadków (Hibiki, 2024). Ich oddziaływanie zależy przede wszystkim od typu odpowiedzialności – czy jest ona oparta na zasadzie ryzyka (*strict liability*) czy na zasadzie winy (*negligence liability*), a także od dostępności ubezpieczenia (np. Jacob, 2015; Shavel, 2005). W przypadku odpowiedzialności na zasadzie ryzyka dany podmiot jest zobowiązany do rekompensaty za szkody wynikające z jego działań i zaniechań niezależnie od winy, tj. czy dopuścił się zaniedbania. Natomiast odpowiedzialność na zasadzie winy ogranicza odpowiedzialność danego podmiotu tylko za te szkody, które powstały w sposób zawiniony przez dany podmiot. Przez zawinienie rozumie się takie zachowanie, które polega na naruszeniu ogólnie przyjętych standardów ostrożności. Zakłada się zatem istnienie standardu ostrożności, którego przestrzegania można było od danego podmiotu rozsądnie oczekiwać, a do którego się on nie dostosował, wyrządzając przez to szkodę.

Unijne zasady odpowiedzialności za szkody wynikające z działania systemów sztucznej inteligencji wprowadzają dychotomię odpowiedzialności za szkody wyrządzone przez AI: opartej na zasadzie winy (*negligence liability*) oraz odpowiedzialności niezależnej od winy (*strict liability*).

Podstawową zasadą odpowiedzialności w większości państw UE jest zasada winy. Uzasadnieniem wprowadzenia dla naruszyciela dużo dotkliwszego reżimu odpowiedzialności opartego na zasadzie ryzyka jest generowanie ponadstandardowego niebezpieczeństwa w ramach prowadzonych przez niego działań, a także podejmowanie pewnej działalności w celu osiągnięcia zysku i własnych korzyści. Taka odpowiedzialność przypisywana jest podmiotom gospodarczym (producentom i innym aktywnym uczestnikom łańcucha dostaw) wprowadzającym system AI do obrotu (lub oddającym go do użytku) w ramach prowadzonej przez nich działalności handlowej, jeżeli nie zapewnia on bezpieczeństwa, którego osoba ma prawo oczekiwać lub które jest wymagane na mocy prawa UE lub prawa krajowego (art. 7 par. 1 dyrektywy PLD). Regulacja ta ma na celu zapewnienie poszkodowanym – osobom fizycznym jak najszerszej ochrony prawnej oraz wiąże się z procesem obostrzania przesłanek i zakresu odpowiedzialności podmiotów profesjonalnych. Odpowiedzialność ta musi mieć jednak granice, w przeciwnym razie zasady odpowiedzialności niezależne od winy prowadziłyby do niepożądanych skutków w postaci zahamowania rozwoju nowych technologii, znacznego wzrostu kosztów produkcji, a co za tym idzie – wzrostu cen produktów (Kuźmicka-Sulikowska, 2011).

W pierwszej kolejności podkreślić jednak należy komplementarną rolę zasad odpowiedzialności za szkody wyrządzone przez produkt oraz wymogów dotyczących bezpieczeństwa produktów. Ani zasady odpowiedzialności, ani same wymogi dotyczące bezpieczeństwa produktów nie są w stanie wyeliminować ryzyka szkód spowodowanych przez produkt. Prawodawca nie dysponuje bowiem pełną informacją na temat ryzyka stwarzanego przez produkt, by wyeliminować groźbę wystąpienia wypadków. Poza tym wymogi szybko mogą stać się nieaktualne, ponieważ są formułowane raczej w sposób statyczny niż dynamiczny i aby były skuteczne, potrzebne jest ich zdecydowane egzekwowanie (Faure, 2014). W grę może wejść także lobbing interesów prywatnych, który producenci wykorzystują, aby wpływać na politykę regulacyjną na swoją korzyść (Maggi i Ossa, 2023). Z kolei zasady odpowiedzialności mogą zostać osłabione przez praktyczne trudności w dochodzeniu roszczeń, np. upływ czasu – gdy pomiędzy czynem niedozwolonym a ujawnieniem szkody mija długi okres, pojawiają się trudności dowodowe, w tym w zidentyfikowaniu źródła szkody (tak istotne z punktu widzenia systemów AI), czy dochodzi do niewypłacalności podmiotu odpowiedzialnego za szkodę. Stąd połączenie obu podejść jest wskazywane w literaturze jako najbardziej korzystne społecznie (Hiriart i in., 2004; Rouillon, 2008; Shavell, 1984).

Strategię taką przyjął także prawodawca unijny: z jednej strony zawarł w Akcie o AI oraz w innych sektorowych i horyzontalnych przepisach dotyczących bezpieczeństwa produktów regulacje mające na celu zapewnienie bezpieczeństwa i ochronę praw podstawowych przy opracowaniu, wdrażaniu i korzystaniu z systemów sztucznej inteligencji; z drugiej – ustanowił przepisy szczególne w zakresie zasad odpowiedzialności za szkody wynikłe z działania systemów AI. Obie regulacje mają się wzajemnie wzmacniać. Skuteczne przepisy w zakresie odpowiedzialności stanowią bowiem zachętę do przestrzega-

nia przepisów w zakresie bezpieczeństwa, by zapobiegać powstawaniu szkód w przyszłości. Przyczyniają się one również do egzekwowania wymogów bezpieczeństwa dotyczących systemów AI wysokiego ryzyka, ponieważ ich niespełnienie powoduje domniemanie wadliwości systemu, a tym samym zmniejszenie ciężaru dowodu i ułatwienie dochodzenia roszczeń (zob. część IV).

2. Współzależność ryzyka i odpowiedzialności

Teoria i praktyka sugerują, że odpowiedzialność powinna być przypisana temu, kto kontroluje i ma świadomość ryzyka związanego z produktem. Jest on bowiem w stanie najskuteczniej ograniczać ryzyko wystąpienia szkód. Jednak problem ten jest dużo bardziej złożony w odniesieniu do systemów AI niż tradycyjnych produktów z uwagi na zakłócenie typowego podziału ról między producentem a użytkownikiem systemu AI, a tym samym kontroli i świadomości ryzyka. Użytkownik może być *de facto* producentem (lub współproducentem) systemu AI, ponieważ może wpływać i modelować cechy i funkcje danego systemu, na przykład wprowadzając określone dane lub wyznaczając istotne parametry funkcjonowania systemu. Z drugiej strony w dotychczasowych regulacjach kluczowym momentem dla przypisania odpowiedzialności producentowi jest moment wprowadzenia produktu do obrotu, tj. wymaga się, aby wada produktu, która spowodowała szkodę, istniała w momencie wprowadzenia produktu do obrotu (art. 7 pkt b dyrektywy 85/374/EWG). Tymczasem technologie cyfrowe umożliwiają podmiotowi gospodarczemu kontrolę produktu już po wprowadzeniu go do obrotu lub oddaniu do użytku, gdy dostarcza on np. aktualizację oprogramowania. Dyrektywa PLD uwzględniła tę sytuację, przewidując odpowiedzialność producentów za wadliwość, która pojawiła się dopiero po wprowadzeniu produktu do obrotu lub oddania do użytku, gdy jest ona spowodowana oprogramowaniem bądź powiązanymi usługami pozostającymi pod kontrolą producenta, czy to w formie aktualizacji lub modernizacji, czy też algorytmów uczenia maszynowego. Oprogramowanie lub powiązane usługi uznaje się za znajdujące się pod kontrolą producenta, jeżeli są one przez niego dostarczane samodzielnie lub producent zezwala na ich dostarczanie przez stronę trzecią (art. 4 par. 5 dyrektywy PLD). Tym samym nastąpiło przesunięcie czasowej granicy odpowiedzialności producenta z momentu wprowadzenia produktu do obrotu lub oddania go do użytku (tj. wadliwość musiała istnieć w tym momencie) do momentu, w którym można uznać, że produkt znajdował się pod jego kontrolą. Aby jednak nadmiernie nie obarczać podmiotów gospodarczych ryzykiem, dyrektywa PLD przewiduje możliwość podniesienia „zarzutu ryzyka rozwoju". Producenci będą mogli zwolnić się z odpowiedzialności za wadliwość produktu, jeżeli udowodnią, że stan wiedzy naukowej i technicznej (ustalony na podstawie najbardziej zaawansowanego poziomu dostępnej obiektywnej wiedzy, a nie wiedzy danego podmiotu gospodarczego), w czasie gdy produkt znajdował się pod kontrolą producenta, nie pozwalał na stwierdzenie istnienia wadliwości. U podstaw tej przesłanki egzoneracyjnej legło przekonanie, że przyjęcie nieograniczonej odpowiedzialności w tym zakresie groziłoby za-

hamowaniem innowacji, z którą immanentnie wiąże się ryzyko popełniania trudnych do przewidzenia w danym momencie błędów[5].

Rozważając zaś kwestie kontroli i świadomości ryzyka z perspektywy użytkownika danego systemu AI, należy przede wszystkim podkreślić ukryty i złożony charakter zagrożeń generowanych przez sztuczną inteligencję. Zagrożenia te nie są bezpośrednio obserwowalne i wykraczają poza typowe rozumienie i postrzeganie otaczającej nas rzeczywistości. Z drugiej jednak strony trudno przecenić rolę użytkownika AI, poczynając od ochrony przed np. różnymi formami cyberataków, a kończąc po prostu na braku trafności działań AI względem mniej lub bardziej oczekiwanych i zdefiniowanych parametrów. Artykuł 13 Aktu o AI wskazuje na obowiązek zapewnienia przejrzystości i udostępniania informacji podmiotom stosującym. Zawiera on postulatywny obowiązek projektowania i rozwijania systemów AI „w sposób zapewniający wystarczającą przejrzystość ich działania, umożliwiającą podmiotom stosującym interpretację wyników systemu i ich właściwe wykorzystanie". Środkiem służącym realizacji wymogu przejrzystości systemów AI wysokiego ryzyka jest przede wszystkim instrukcja obsługi, która nie tylko powinna zawierać kompleksowe informacje, w szczególności o przeznaczeniu i właściwym użytkowaniu danego systemu AI, ale być dostępna i zrozumiała dla podmiotów stosujących, co wymaga od dostawcy systemu zwrócenia uwagi nie tylko na zawartość instrukcji, ale także na jej właściwą redakcję.

Warner i Solan (2021) podkreślają różnicę między „przejrzystością" (*transparency*) a „wytłumaczalnością" bądź „wyjaśnialnością" (*explainability*) systemów AI, wskazując że wytłumaczalność jest istotna dla przejrzystości, ale nie jest dla niej ani konieczna, ani wystarczająca. Podobnie Gawroński (2023) uznaje wyjaśnialność za pewien aspekt przejrzystości rozumianej jako „możliwość zrozumienia danego ciągu przyczynowo-skutkowego" (s. 27), czyli wiedzę, jak przebiega proces decyzyjny z udziałem danego systemu AI. Akt o AI nie wymaga, aby systemy AI były wytłumaczalne, przynajmniej w sensie wytłumaczalności racjonalnej, tj. by istniała możliwość śledzenia toku rozumowania danego systemu AI[6]. (W przypadku niektórych systemów AI jest to wręcz niemożliwe z uwagi na problem „czarnej skrzynki", tzn. brak śladu danych wejściowych, które by informowały o podejmowanych przez dany system decyzjach; Panigutti i in., 2023). Niemniej użytkownik systemu AI wysokiego ryzyka musi orientować się, jak działa dany system. Sugeruje to art. 14 Aktu o AI, który traktuje o środkach nadzoru ludzkiego. Wbudowane w interfejsy AI oraz instrukcje obsługi środki nadzoru mają umożliwiać ludzkiemu nad-

[5] Niemniej jednak państwa członkowskie w myśl art. 18 dyrektywy PLD będą mogły wyłączyć zarzut „ryzyka rozwoju", gdy uznają to wyłączenie za nadmiernie ograniczające ochronę osób fizycznych, powiadamiając o tym Komisję Europejską, która następnie poinformuje o tym pozostałe państwa członkowskie.

[6] Pojęcie „wytłumaczalność" (*explainability*) pojawia się tylko w motywie 27 Aktu o AI. Można przyjąć, że chodzi o wytłumaczalność empiryczną odwołującą się do efektywności działania systemu AI (Gwaroński, 2023). Wytłumaczalność empiryczna – w odróżnieniu od wytłumaczalności racjonalnej, która pozwala na zrozumienie logiki decyzji podejmowanych przez AI – odwołuje się do praktycznych spostrzeżeń opartych na danych, które są zgodne z obserwacjami ze świata rzeczywistego, np. czy decyzje podejmowane przez AI są zgodne z ludzką intuicją.

zorcy m.in.: zrozumienie możliwości i ograniczeń systemu AI, należyte monitorowanie jego działania, wykrywanie anomalii oraz podejmowanie środków zaradczych; pozostawanie świadomym potencjalnej tendencji do tzw. błędu automatyzacji, czyli nadmiernego polegania na wyniku wytworzonym przez AI; prawidłową interpretację wyniku systemu AI, biorąc pod uwagę dostępne narzędzia i metody interpretacji; podjęcie decyzji o niekorzystaniu z systemu AI lub w inny sposób zignorowanie, unieważnienie lub odwrócenie wyniku systemu AI; zatrzymanie systemu AI w stanie bezpiecznym. Regulacja zakłada więc pewną kontrolę podmiotu stosującego system AI wysokiego ryzyka, chociażby w zakresie wyników generowanych przez system. Co więcej, jeżeli producent może przewidzieć, do czego można wykorzystać jego system AI, nawet niezgodnie ze wskazanym w instrukcji przeznaczeniem, to powinien opisać tego konsekwencje (art. 14 par. 2 Aktu o AI).

W myśl art. 26 Aktu o AI podmioty stosujące systemy AI wysokiego ryzyka mają obowiązek przede wszystkim: korzystania z systemu AI zgodnie z załączoną do niego instrukcją obsługi; powierzania nadzoru nad systemem osobom fizycznym odpowiednio wykwalifikowanym; zapewnienia adekwatności danych wejściowych; monitorowania korzystania z systemu AI zgodnie z instrukcją obsługi pod kątem ryzyka dla zdrowia, bezpieczeństwa i praw podstawowych oraz informowania, w stosowanych przypadkach, dostawców i organy nadzoru rynku o stwierdzonym ryzyku i zaistniałych incydentach, a także przechowywania generowanych automatycznie przez system AI rejestrów zdarzeń. Co istotne, jeśli użytkownicy systemów AI dokonają istotnej modyfikacji w systemie poza kontrolą producenta, sami mogą zostać pociągnięci do odpowiedzialności za szkody wyrządzone przez ten system na podstawie art. 8 dyrektywy PLD, gdyż są wówczas uważani za producentów nowego produktu.

VI. UWAGI KOŃCOWE

Ryzyko odpowiedzialności odszkodowawczej w pewnym sensie stanowi cenę za wprowadzenie innowacyjnych rozwiązań technologicznych. Zbyt rygorystyczne reżimy odpowiedzialności za AI mogą wpływać hamująco na ich rozwój i wdrażanie (Bożek i Jakubiec, 2017; Świerczyński i Więckowski, 2023). Zrzucanie ryzyka zmian wyłącznie na przedsiębiorcę, który je wprowadza, mogłoby zniechęcić do innowacji, co wyraźnie nie leży w interesie publicznym. Z kolei zbyt łagodne – powodują brak zaufania do AI, rodzą zagrożenia dla ochrony prywatności, niedyskryminacji czy przejrzystości podejmowanych decyzji.

Oceniając zasady odpowiedzialności za szkody wynikłe z działania systemów AI, należy odnieść się przede wszystkim do trzech celów, które przyświecały wprowadzeniu szczególnych regulacji w tym zakresie, tj. (1) stworzenie zachęt do zapobiegania szkodom, (2) zapewnienie rekompensaty stronom poszkodowanym oraz (3) zapewnienie producentom, dostawcom i użytkownikom większej pewności prawa w zakresie ryzyka, jakie podejmują oni w trakcie swojej działalności.

Ogólnie rzecz biorąc, podejście prawodawcy unijnego do regulacji sztucznej inteligencji można scharakteryzować dwojako. Po pierwsze, jako powiązanie wymogów bezpieczeństwa (interwencja *ex ante*) z zasadami odpowiedzialności (interwencja *ex post*). Wymogi nałożone na systemy AI w Akcie o AI oraz w innych sektorowych i horyzontalnych przepisach dotyczących bezpieczeństwa produktów mają stanowić zachętę do stosowania przez dostawcę systemów AI odpowiednich środków zapobiegawczych, by minimalizować ryzyko powstania szkód w przyszłości. Po drugie, jako oparte na analizie ryzyka oraz zarządzaniu całym cyklem życia produktu. W konsekwencji prawodawstwo unijne klasyfikuje systemy AI zgodnie z ryzykiem, jakie stwarzają one dla użytkowników. Im większy stopień ryzyka, tym wyższy stopień regulacji. W tym kontekście należy zauważyć, że Akt o AI wyraźnie zakazuje używania i wprowadzania do obrotu systemów AI w odniesieniu do nielicznych zastosowań sztucznej inteligencji. Innymi słowy, systemy AI spełniające kryteria praktyk zakazanych stanowią niewielki ułamek rozwijanych technologii AI, a jedynie systemy AI zakwalifikowane jako systemy wysokiego ryzyka obwarowane są rozbudowanymi procedurami jakościowymi, jak m.in. system zarządzania ryzykiem, dokumentacja techniczna czy odpowiedni nadzór człowieka. Pozostałe systemy AI nie podlegają szczególnym wymogom regulacyjnym, z wyjątkiem obowiązków w zakresie przejrzystości. Ponadto po wielu rundach poprawek Akt o AI przyrównywany jest przez niektórych do szwajcarskiego sera, w którym jest „więcej powietrza niż twardej regulacyjnej materii" (Szymielewicz, 2024, s.64), ponieważ niemal wszystkie reguły obwarowane są wyjątkami, a do niemal każdej definicji znajdzie się zmiękczające ją zastrzeżenie. Lobby biznesowe w imię innowacji przewalczyło również uznaniowość w klasyfikowaniu systemów AI jako systemów wysokiego ryzyka. Z wyjątkiem systemów AI związanych z bezpieczeństwem produktów, to dostawca samodzielnie dokonuje oceny ryzyka wprowadzanych i wykorzystywanych przez siebie systemów AI. Z perspektywy prawnej więc nie do przeceniania będzie rola organów nadzoru rynku oraz tzw. organów notyfikujących, które w każdym państwie członkowskim mają być odpowiedzialne za opracowanie i stosowanie procedur oceny zgodności systemów AI wysokiego ryzyka z wymogami określonymi w Akcie o AI. Natomiast z perspektywy zarządczej – wypracowanie wspólnych standardów, w tym inicjatywy takie jak VAIR (*open vocabulary for AI risks*), które mogą pomóc w identyfikacji i dokumentacji ryzyk związanych z AI (a tym samym systemów AI wysokiego ryzyka), a także ułatwić dzielenie się wiedzą i interoperacyjność między podmiotami w łańcuchu wartości AI. Obiecujący kierunek badań w tej dziedzinie wyznaczają m.in. Golpayegani i in. (2023), którzy zaproponowali uproszczone i ustrukturyzowane podejście do identyfikacji systemów AI wysokiego ryzyka w kontekście praw podstawowych, które można by wykorzystać w audytach i w dochodzeniach w sprawie zgodności systemów AI z wymogami. Interesujące są także koncepcje rozszerzenia cyfrowego paszport produktu (Digital Product Passport [DPP]) na systemy AI (np. García-Gómez i in., 2024).

Jeśli zaś pomimo spełnienia wymogów dotyczących bezpieczeństwa systemu AI, lub w ich braku, szkody jednak powstaną – prawodawstwo unijne

przewiduje dwutorowy system dochodzenia roszczeń: w ramach odpowiedzialności opartej na zasadzie ryzyka podmiotów gospodarczych za produkty wadliwe (dyrektywa PLD) oraz w ramach krajowych systemów odpowiedzialności deliktowej opartych na zasadzie winy. W tym względzie należy odnotować, że zakres przedmiotowy i podmiotowy dyrektywy PLD został istotnie rozszerzony, w tym co do pojęcia samego produktu (może je stanowić samo oprogramowanie), szkody (także niematerialne) i wadliwości (przy ocenie wadliwości produktu bierze się pod uwagę funkcje samouczenia się oraz powiązania między produktami, a także luki w cyberbezpieczeństwie). Nie udało się jednak uchwalić dyrektywy w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji (AILD) będącej ważnym uzupełnieniem wspomnianej wyżej dyrektywy w sprawie odpowiedzialności za produkty wadliwe (PLD), choć w kwestiach proceduralnych oba akty w niewielkim stopniu ingerują w krajowe systemy prawne, ponieważ opierają się one przede wszystkim na wprowadzeniu obowiązku ujawnienia dowodów oraz wzruszalnych domniemań prawnych dotyczących wadliwości, winy i związku przyczynowego. Takie ukierunkowane, czy „punktowe," podejście i tak było krytykowane jako niewystarczające: na przykład Hacker (2023), określił podejście UE do uregulowania kwestii odpowiedzialności za AI wręcz jako „połowiczne" (*half-hearted*), a De Bruyne i in. (2023) wskazali na niezaadresowane przez prawodawstwo unijne wyzwania związane m.in. z pojęciem winy i standardu starannego działania (*duty of care*), które są warunkiem pociągnięcia do odpowiedzialności na zasadzie winy w odniesieniu do podmiotów, które nie są dostawcami systemów AI wysokiego ryzyka.

Podsumowując, można stwierdzić, że omawiane regulacje UE stanowią krok naprzód, ale wymagają dalszego udoskonalenia, aby w pełni uwzględnić złożoność szkód związanych ze sztuczną inteligencją. Warto w tym kontekście wskazać na dwa elementy, których obecnie nie przewidują omawiane regulacje, a których wprowadzenie mogłoby zwiększyć pewność prawną: górny limit odpowiedzialności oraz obowiązkowe ubezpieczenia dla podmiotów gospodarczych, które w przypadku mikro, małych i średnich przedsiębiorstw mogłyby być dotowane. Z jednej strony ułatwiłoby to podmiotom profesjonalnym zarządzanie ryzykiem generowanym przez AI, a poszkodowanym dałoby gwarancję rekompensaty. Należy mieć bowiem na uwadze, że odpowiedzialność niezależna od winy (*strict liability*), o której stanowi dyrektywa PLD, nie tylko jest środkiem internalizacji ryzyka związanego z AI, nijako wymuszającej uwzględnienie postępu technologicznego, ale także sprzyja zwiększeniu społecznej akceptacji dla sztucznej inteligencji (zob. Hacker 2023; Spindler, 2023).

Author contributions / Indywidualny wkład autora (CRediT): **Paulina Kubera** – 100% (Conceptualization / Konceptualizacja; Investigation / Przeprowadzenie badań; Writing – original draft / Pisanie – pierwszy szkic; Writing – review & editing / Pisanie – recenzja i edycja).

Conflict of interest / Konflikt interesów: The author declares no conflict of interest. / Autorka nie zgłosiła konfliktu interesów.

Funding / Finansowanie: The author declares no institutional funding. / Autorka oświadczyła, że nie korzystała z finansowania instytucjonalnego.
The use of AI tools / Wykorzystanie narzędzi AI: The author declares no use of AI tools. / Autorka oświadczyła, że nie korzystała z narzędzi AI.
Data availability / Dostępność danych: Not applicable. / Nie dotyczy.

References / Bibliografia

Bożek, B., i Jakubiec, M. (2017). On the legal responsibility of autonomous machines. *Artificial Intelligence and Law*, *25*(3), 293–304. https://doi.org/10.1007/s10506-017-9207-8

Buiten, M. (2024). Product liability for defective AI. *European Journal of Law and Economics*, *57*, 239–273. https://doi.org/10.1007/s10657-024-09794-z

Buiten, M., de Streel, A., i Peitz, M. (2023). The law and economics of AI liability. *Computer Law & Security Review*, *48*, 105794. https://doi.org/10.1016/j.clsr.2023.105794

De Bruyne, J., Dheu, O., i Ducuing, C. (2023). The European Commission's approach to extra-contractual liability and AI – An evaluation of the AI liability directive and the revised product liability directive. *Computer Law & Security Review*, *51*, 105894. https://doi.org/10.1016/j.clsr.2023.105894

Derave, C., Genicot, N., i Hetmanska, N. (2022). The risks of trustworthy artificial intelligence: The case of the European travel information and authorisation system. *European Journal of Risk Regulation*, *13*(3), 389–420. https://doi:10.1017/err.2022.5

Faure, M. G. (2014). The complementary roles of liability, regulation and insurance in safety management: Theory and practice. *Journal of Risk Research*, *17*(6), 689–707. https://doi.org/10.1080/13669877.2014.889199

Følstad, A., Larsen, A., i Bjerkreim-Hanssen, N. (2023). The human likeness of government chatbots – An empirical study from Norwegian municipalities. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14130 LNCS, 111–127. https://doi.org/10.1007/978-3-031-41138-0_8

García-Gómez, J., Blanes-Selva, V., i Doñate-Martínez, A. (2024). Proposing an AI passport as a mitigating action of risk associated to artificial intelligence in healthcare. W: J. Mantas, A. Hasman, G. Demiris, K. Saranto, M. Marschollek, T. Arvanitis, I. Ognjanović, A. Benis, P. Gallos, E. Zoulias i E. Andrikopoulou (red.), *Digital health and informatics innovations for sustainable health care systems* (s. 537–551). Proceedings of MIE 2024. IOS Press. https://doi.org/10.3233/SHTI240472

Gawroński, M. (2023). Wyjaśnialność SI w fazie wykorzystania – aspekty prawne i techniczne. W: *Projektowanie systemów SI zgodnie z RODO. Materiały pokonferencyjne* (s. 26–37). Urząd Ochrony Danych Osobowych.

Golpayegani, D., Pandit, H., i Lewis, D. (2023). To be high-risk, or not to be – semantic specifications and implications of the AI Act's high-risk AI applications and harmonised standards. W: *6th ACM Conference on Fairness, Accountability, and Transparency, FAccT 2023* (s. 905–915). https://dl.acm.org/doi/10.1145/3593013.3594050

Hacker, P. (2023). The European AI liability directives – Critique of a half-hearted approach and lessons for the future. *Computer Law & Security Review*, *51*, 105871. https://doi.org/10.1016/j.clsr.2023.105871

Hibiki, A. (2024). Precaution against environmental accidents and liability rules for damages. W: T. Arimura i A. Hibiki (red.), *Introduction to Environmental Economics and Policy in Japan* (s. 75–87). Springer.

Hiriart, Y., Martimort, D., i Pouyet, J. (2004). On the optimal use of ex ante regulation and ex post liability. *Economics Letters*, *84*(2), 231–235. https://doi.org/10.1016/j.econlet.2004.02.007

Jacob, J. (2015). Innovation in risky industries under liability law: The case of double-impact innovations. *Journal of Institutional and Theoretical Economics* (JITE) / *Zeitschrift für die Gesamte Staatswissenschaft*, *171*(3), 385–404.

Knott, A., Pedreschi, D., Jitsuzumi, T., i Bengio, Y.(2024). AI content detection in the emerging information ecosystem: New obligations for media and tech companies. *Ethics and Information Technology*, *6*(63). https://doi.org/10.1007/s10676-024-09795-1

Komisja Europejska. (2020a). Directorate-General for Communications Networks, Content and Technology, European enterprise survey on the use of technologies based on artificial intelligence – Final report. Publications Office. https://data.europa.eu/doi/10.2759/759368European

Komisja Europejska. (2020b). Biała księga w sprawie sztucznej inteligencji Europejskie podejście do doskonałości i zaufania. Bruksela, dnia 19.2.2020 r. COM(2020) 65 final.

Księżak P., i Wojtczak S. (2020). Prawo autorskie wobec sztucznej inteligencji (próba alternatywnego spojrzenia). *Państwo i Prawo*, *75*(2), 18–33.

Kuźmicka-Sulikowska, J. (2011). *Zasady odpowiedzialności deliktowej w świetle nowych tendencji w ustawodawstwie polskim*. Wolters Kluwer Polska.

Li, S., i Schütte, B. (2023). The proposal for a revised Product Liability Directive: The emperor's new clothes? *Maastricht Journal of European and Comparative Law*, *30*(5), 573–596. https://doi.org/10.1177/1023263X231216941

Maggi, G., i Ossa, R. (2023). The political economy of international regulatory cooperation. *American Economic Review*, *113*(8), 2168–2200. https://doi.org/10.1257/aer.20200780

Michalak, A. (2021). Projekt rozporządzenia Parlamentu UE o odpowiedzialności cywilnej za działania systemów sztucznej inteligencji – krok w dobrym kierunku czy niepotrzebne odstępstwo od zasad ogólnych? W: B. Fischer, A. Pązik i M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii* (s. 41–50). Wolters Kluwer Polska.

Panigutti, C., Hamon, R., Hupont, I., Llorca, D., Yela, D., Junklewitz, H., Scalzo, S., Mazzini, G., Sanchez, I., Garrido, J., i Gomez, E. (2023). The role of explainable AI in the context of the AI Act. W: *FAccT'23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (s. 1139–1150). https://doi.org/10.1145/3593013.3594069

Rauccio, C. (2021). Artificial intelligence and genomics: The data protection implications in the use of AI for genomic diagnostics. *European Journal of Privacy Law and Technologies*, *1*, 115–141.

Ren, Q., i Du, J. (2024). Harmonizing innovation and regulation: The EU Artificial Intelligence Act in the international trade context. *Computer Law & Security Review*, *54*, 106028. https://doi.org/10.1016/j.clsr.2024.106028

Rouillon, S. (2008). Safety regulation vs. liability with heterogeneous probabilities of suit. *International Review of Law and Economics*, *28*(2), 133–39. https://doi.org/10.1016/j.irle.2008.02.002

Shavell, S. (1984). A model of the optimal use of liability and safety regulation. *Rand Journal of Economics*, *15*, 271–280.

Shavell, S. (2005). Economics and liability for accidents: New Palgrave dictionary of economics (2nd Edition, 2008). *Harvard Law and Economics Discussion Paper*, *535*. https://ssrn.com/abstract=870565

Skjuve, M., Haugstveit, I., Følstad, A., i Brandtzaeg, P. (2019) Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, *15*(1), 30–54. https://doi.org/10.17011/ht/urn.201902201607

Spindler, G. (2023). Different approaches for liability of artificial intelligence – pros and cons – the new proposal of the EU Commission on liability for defective products and AI systems. *SSRN*. http://dx.doi.org/10.2139/ssrn.4354468

Szymielewicz, K. (2024). Nieśmiała kontrola. *Tygodnik Powszechny*. Wydanie specjalne, 3–4, 63–66.

Świerczyński, M., i Więckowski, W. (2023). Liability for damages caused by artificial intelligence systems – main challenges to be addressed by the European Union conflict-of-laws regulations. *Prawo w Działaniu*, *54*, 200–206. https://doi.org/10.32041/pwd.5407

Warner, R., i Sloan, R. (2021). Making artificial intelligence transparent: Fairness and the problem of proxy variables. *Criminal Justice Ethics*, *40*(1), 23–39. https://doi.org/10.1080/0731129X.2021.1893932

Yampolskiy, R. (2020). Unpredictability of AI: On the impossibility of accurately predicting all actions of a smarter agent. *Journal of Artificial Intelligence and Consciousness*, *7*(1), 109–118. https://doi.org/10.1142/S2705078520500034