Akt o sztucznej inteligencji a walka z dezinformacją – przegląd wybranych obowiązków dostawców i podmiotów stosujących systemy AI

*Martyna Czapska**

Artykuł zawiera przegląd obowiązków nałożonych w Akcie ds. sztucznej inteligencji[1] na dostawców i podmioty stosujące systemy AI odnoszących się do systemów AI wysokiego ryzyka, systemów AI ogólnego przeznaczenia oraz obowiązków w zakresie transparentności, które mogą służyć walce z dezinformacją.

Wprowadzenie

Globalna komunikacja, łatwość rozpowszechniania informacji za pomocą Internetu, szeroka dostępność narzędzi generatywnych opartych na sztucznej inteligencji[2], dynamiczna sytuacja geopolityczna – wszystko to tworzy dogodne warunki dla rozwoju dezinformacji. Rok 2024 upłynie pod znakiem wyborów, w tym do Parlamentu Europejskiego oraz wyborów prezydenckich w Stanach Zjednoczonych. Obawy o zakłócanie procesów demokratycznych za pomocą dezinformacji są realne i obecne[3]. O dezinformacji mówi się również jako o elemencie wojny hybrydowej, przede wszystkim w kontekście działań rosyjskich związanych z agresją Rosji na Ukrainę, ale nie tylko. W styczniu 2024 r. Europejska Służba Działań Zewnętrznych (EEAS) opublikowała swój drugi już raport dotyczący zewnętrznej manipulacji w przestrzeni informacyjnej i zagrożeń ingerencją[4]. Wynika z niego, że pomiędzy 1.12.2022 r. a 30.11.2023 r. doszło do 750 incydentów stanowiących próbę obcej dezinformacji, a najczęstszym celem ataków była Ukraina (160 incydentów), Polska była zaś na trzecim miejscu (33 incydenty)[5].

Na poziomie unijnym działania na rzecz walki z dezinformacją są podejmowane od kilku lat. W Komunikacie Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów w sprawie europejskiego działania na rzecz demokracji[6] **dezinformację** zdefiniowano jako treści fałszywe lub wprowadzające w błąd, które są rozpowszechniane z zamiarem wprowadzenia w błąd lub zapewnienia korzyści ekonomicznych lub politycznych i które mogą wyrządzić szkodę publiczną. Z kolei 16.6.2022 r. 34 sygnatariuszy podpisało wzmocniony kodeks postępowania w zakresie dezinformacji[7], stanowiący rezultat przeglądu kodeksu z 2018 r. Na potrzeby tego dokumentu „Dezinformację" zdefiniowano jako obejmującą informacje wprowadzające w błąd[8], dezinformację[9], działania w zakresie wywierania wpływu informacyjnego[10] oraz zewnętrzną ingerencję w przestrzeń informacyjną[11]. Akt ws. sztucznej inteligencji wymienia dezinformację jako jedno z ryzyk systemowych, jakie mogą stwarzać modele sztucznej inteligencji ogólnego przeznaczenia dla demokratycznych wartości i praw człowieka (motyw 110, motyw 136). Ochrona demokracji jest jednym z celów przyjęcia Aktu ws. sztucznej inteligencji, wskazanym już w jego art. 1.

Narzędziem prawnym na poziomie UE, które wysuwa się obecnie na pierwszy plan, jeśli chodzi o walkę z dezinformacją w sieci, jest Akt o usługach cyfrowych[12]. Motyw 120 Aktu

* Autorka jest radczynią prawną, senior associate w Baker McKenzie Krzyżowski i Wspólnicy sp. k., ORCID: 0000-0002-3765-8944.

[1] Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniające niektóre akty ustawodawcze Unii; dalej jako: Akt ws. sztucznej inteligencji, przyjęte przez Parlament Europejski 13.3.2024 r., https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_PL.html (dostęp: 1.5.2024 r.), sprostowanie z 17.4.2024 r., https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_PL.pdf (dostęp: 1.5.2024 r.), brak publikacji na dzień finalizacji niniejszego artykułu.

[2] Dalej również jako: AI.

[3] Zob. np. *K. Neubert*, Disinformation campaigns likely to undermine EU elections, experts say, 13.3.2024 r., https://www.euractiv.com/section/disinformation/news/disinformation-campaigns-likely-to-undermine-eu-elections-experts-say/ (dostęp: 18.3.2024 r.).

[4] Raport EEAS: https://www.eeas.europa.eu/sites/default/files/documents/2024/EEAS-2nd-Report%20on%20FIMI%20Threats-January-2024_0.pdf (dostęp: 30.3.2024 r.).

[5] *Ibidem*, s. 9.

[6] Zob. https://eur-lex.europa.eu/legal-content/PL/TXT/HTML/?uri=CELEX:52020DC0790 (dostęp: 30.3.2024 r.); dalej jako: EDAP.

[7] Komunikat Komisji Europejskiej z 16.6.2022 r.: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation (dostęp: 29.3.2024 r.).

[8] Zgodnie z definicją zawartą EDAP, „informacje wprowadzające w błąd są fałszywymi lub wprowadzającymi w błąd treściami udostępnianymi bez szkodliwego zamiaru, np. gdy ludzie dzielą się fałszywymi informacjami z przyjaciółmi i rodziną w dobrej wierze".

[9] Zgodnie z podaną wyżej definicją zawartą w EDAP.

[10] Zgodnie z definicją zawartą EDAP, „działania w zakresie wywierania wpływu informacyjnego odnoszą się do skoordynowanych wysiłków podejmowanych przez podmioty krajowe lub zagraniczne w celu wywarcia wpływu na grupę docelową przy użyciu szeregu środków wprowadzających w błąd, w tym ograniczania niezależnych źródeł informacji w połączeniu z dezinformacją".

[11] Zgodnie z definicją zawartą EDAP, zewnętrzna ingerencja w przestrzeń informacyjną, często prowadzona w ramach szerszej operacji hybrydowej, może być rozumiana jako represyjne i zwodnicze wysiłki – mające na celu zakłócenie swobodnego kształtowania się i wyrażania woli politycznej jednostek – podejmowane przez podmiot z zagranicy lub jego przedstawicieli.

[12] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z 19.10.2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (Akt o usługach cyfrowych), Dz.Urz. L Nr 277 z 27.10.2022 r., s. 1 ze zm.; dalej jako: AUC.

ws. sztucznej inteligencji wskazuje, że obowiązki, jakie zostaną nałożone na dostawców i podmioty stosujące określone systemy AI, by umożliwić wykrywanie i ujawnianie, że dane wyjściowe tych systemów są sztucznie wygenerowane lub zmanipulowane, są szczególnie istotne dla ułatwienia skutecznego wdrożenia AUC. Dotyczy to w szczególności obowiązków dostawców bardzo dużych platform internetowych lub bardzo dużych wyszukiwarek internetowych w zakresie identyfikacji i ograniczania ryzyka systemowego, które może wynikać z rozpowszechniania treści sztucznie wygenerowanych lub zmanipulowanych, w szczególności identyfikowania i ograniczania ryzyka rzeczywistego lub przewidywalnych negatywnych skutków dla procesów demokratycznych, dyskursu obywatelskiego i procesów wyborczych, w tym poprzez stosowanie dezinformacji.

Obowiązki nałożone na poszczególne podmioty objęte zakresem działania Aktu ws. sztucznej inteligencji będą służyły realizacji innych aktów prawnych, ale też same w sobie stanowią zestaw nowych wymogów, które te podmioty będą musiały spełnić, przyczyniając się tym samym do dalszego rozrostu obowiązków w zakresie *compliance*. Celem niniejszego artykułu jest dokonanie przeglądu wybranych obowiązków zawartych w Akcie ws. sztucznej inteligencji nakierowanych na walkę z dezinformacją.

Akt został przyjęty przez Parlament Europejski w środę 13.3.2024 r.[13] W chwili pisania niniejszego artykułu oczekiwał jeszcze na przyjęcie przez Radę oraz publikację w Dzienniku Urzędowym. Akt wejdzie w życie 20 dni po publikacji, a zacznie być w pełni stosowany 24 miesiące po wejściu w życie, z następującymi wyjątkami: zakazy niedozwolonych praktyk zaczną obowiązywać 6 miesięcy po wejściu Aktu w życie, przepisy dotyczące kodeksów postępowania – 9 miesięcy po wejściu w życie, przepisy dotyczące sztucznej inteligencji ogólnego przeznaczenia – 12 miesięcy po wejściu w życie oraz obowiązki dotyczące systemów sztucznej inteligencji wysokiego ryzyka – 36 miesięcy po wejściu Aktu w życie. Reasumując, podmioty podlegające pod Akt ws. sztucznej inteligencji muszą osiągnąć gotowość do jego pełnego stosowania w ciągu najbliższych 2 lat, z dodatkowym zapasem jednego roku w zakresie obowiązków dotyczących systemów AI wysokiego ryzyka.

Centralnym pojęciem zdefiniowanym w Akcie ds. sztucznej inteligencji jest **system AI**. Został on zdefiniowany w Akcie jako system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych, wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne[14].

Zakres podmiotowy Aktu ws. sztucznej inteligencji

Akt ws. sztucznej inteligencji ma szeroki **zakres podmiotowy**[15], który obejmuje:

1) dostawców wprowadzających do obrotu lub oddających do użytku systemy AI lub wprowadzających do obrotu modele AI ogólnego przeznaczenia w Unii, niezależnie od tego, czy dostawcy ci mają siedzibę lub znajdują się w Unii, czy w państwie trzecim;
2) podmioty stosujące systemy AI mające swoją siedzibę lub znajdujące się na terenie Unii;
3) dostawców systemów AI i podmiotów stosujących systemy AI mających siedzibę lub znajdujących się w państwie trzecim, w przypadku gdy wyniki wytworzone przez system AI są wykorzystywane w Unii;
4) importerów i dystrybutorów systemów AI, producentów produktu, którzy pod własną nazwą lub znakiem towarowym oraz wraz ze swoim produktem wprowadzają do obrotu lub oddają do użytku system AI, upoważnionych przedstawicieli dostawców, którzy nie mają siedziby w Unii oraz
5) osoby, na które AI ma wpływ i które znajdują się w Unii.

Akt ws. sztucznej inteligencji nie ma zastosowania m.in. do systemów AI, jeśli – i w zakresie, w jakim – wprowadzono je do obrotu, oddano do użytku lub są one wykorzystywane ze zmianami lub bez zmian, wyłącznie do celów wojskowych, obronnych lub do celów bezpieczeństwa narodowego, niezależnie od rodzaju podmiotu prowadzącego te działania, jak również Akt nie ma zastosowania do systemów AI, które nie zostały wprowadzone do obrotu ani oddane do użytku w Unii, a których wyniki są wykorzystywane w Unii wyłącznie do celów wojskowych, obronnych lub do celów bezpieczeństwa narodowego, niezależnie od rodzaju podmiotu prowadzącego te działania. Aktu ws. sztucznej inteligencji nie stosuje się także do systemów AI lub modeli AI, w tym ich wyników, rozwiniętych i oddanych do użytku wyłącznie do celów badań naukowych i rozwojowych.

Akt nie ma również zastosowania do obowiązków podmiotów stosujących będących osobami fizycznymi korzystającymi z systemów AI w ramach czysto osobistej działalności pozazawodowej.

Obowiązki wynikające z Aktu ws. sztucznej inteligencji są zróżnicowane w zależności od tego, którego z tych podmiotów dotyczą (a także w zależności od poziomu ryzyka generowanego przez dany rodzaj systemów AI). Akt ws. sztucznej inteligencji opiera się o podejście do systemów AI oparte na analizie ryzyka i wyróżnia w związku z tym cztery poziomy ryzyka:

1) ryzyko nieakceptowalne, co wyraża się w ustaleniu katalogu zakazanych systemów AI[16] (art. 5 Aktu);
2) wysokie ryzyko, wyrażone poprzez ustalenie kryteriów dla uznania systemu AI za system wysokiego ryzyka i wprowa-

[13] Komunikat prasowy Parlamentu Europejskiego: https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law (dostęp: 26.3.2024 r.).

[14] Art. 3 pkt 1 Aktu ws. sztucznej inteligencji.

[15] Art. 2 Aktu ws. sztucznej inteligencji.

[16] Jedną z zakazanych praktyk jest wprowadzanie do obrotu, oddawanie do użytku lub używanie systemu AI wykorzystującego techniki podprogowe wykraczające poza świadomość danej osoby lub techniki celowo manipulacyjne lub oszukańcze, których celem lub skutkiem jest istotne zniekształcenie zachowania osoby lub grupy osób, poprzez zauważalne ograniczenie ich zdolności do podejmowania świadomej decyzji, powodując w ten sposób podjęcie przez tę osobę decyzji, której w innym przypadku by nie podjęła, w sposób, który powoduje lub może spowodować, istotną szkodę dla tej osoby, innej osoby lub grupy osób.

dzenie szczegółowych i ścisłych wymogów regulacyjnych dla tego typu systemów;
3) ograniczone ryzyko, wyrażające się w ustaleniu obowiązków w zakresie transparentności w odniesieniu do pewnych rodzajów systemów AI oraz
4) minimalne ryzyko – tego typu systemy AI nie będą podlegały szczególnym obowiązkom, a jako przykłady podaje się gry wideo z obsługą AI lub filtry spamu[17].

Obowiązki dotyczące systemów wysokiego ryzyka

1. Określenie systemów AI wysokiego ryzyka

Najbardziej restrykcyjne regulacje, pomijając zakazane praktyki w zakresie AI, dotyczą **systemów AI wysokiego ryzyka**. System AI będzie uznany za system wysokiego ryzyka[18] bez względu na to, czy jest wprowadzany do obrotu lub oddawany do użytku niezależnie od produktów wskazanych w punktach (a) i (b) poniżej, jeżeli są spełnione oba następujące warunki:
1) system AI jest przeznaczony do wykorzystania jako związany z bezpieczeństwem element produktu objętego unijnym prawodawstwem harmonizacyjnym wymienionym w załączniku I do Aktu lub sam system AI jest takim produktem;
2) produkt, którego związanym z bezpieczeństwem elementem zgodnie z lit. a) jest system AI, lub sam system AI jako produkt podlegają – na podstawie unijnego prawodawstwa harmonizacyjnego wymienionego w załączniku I do Aktu – ocenie zgodności przez stronę trzecią w związku z wprowadzeniem tego produktu do obrotu lub oddaniem go do użytku.

Ponadto systemami AI wysokiego ryzyka są również systemy wskazane w załączniku III do Aktu, czyli systemy działające w określonych w tym załączniku obszarach takich jak zdalne systemy identyfikacji biometrycznej, infrastruktura krytyczna, edukacja i kształcenie zawodowe, zatrudnienie, zarządzanie pracownikami i dostęp do samozatrudnienia, dostęp i korzystanie z podstawowych usług prywatnych oraz podstawowych usług i świadczeń publicznych, organy ścigania, o ile wykorzystanie takich systemów jest dozwolone na mocy odpowiednich przepisów prawa unijnego lub krajowego, zarządzanie migracją, azylem i kontrolą graniczną, o ile stosowanie takich systemów jest dozwolone na mocy odpowiednich przepisów prawa unijnego lub krajowego, wymiar sprawiedliwości i procesy demokratyczne. W tej ostatniej kategorii jako systemy wysokiego ryzyka wskazano m.in. systemy AI przeznaczone do wykorzystywania do wpływania na wynik wyborów lub referendum lub na zachowanie osób fizycznych podczas głosowania w wyborach lub referendach. Nie obejmuje to systemów AI, na których działanie osoby fizyczne nie są bezpośrednio narażone, takich jak narzędzia wykorzystywane do organizowania, optymalizowania lub strukturyzowania kampanii politycznych z administracyjnego lub logistycznego punktu widzenia.

System AI wskazany w załączniku III nie zostanie jednak uznany za system wysokiego ryzyka, jeżeli nie stwarza znaczącego ryzyka szkody dla zdrowia, bezpieczeństwa lub praw podstawowych osób fizycznych, w tym poprzez brak znaczącego wpływu na wynik procesu decyzyjnego. Ma to miejsce, gdy spełniony jest którykolwek z następujących warunków:
1) system AI jest przeznaczony do wykonywania wąsko zakreślonego zadania proceduralnego;
2) system AI jest przeznaczony do poprawienia wyniku uprzednio zakończonej czynności wykonywanej przez człowieka;
3) system AI jest przeznaczony do wykrywania wzorców decyzyjnych lub odstępstw od wzorców podjętych uprzednio decyzji i nie ma na celu zastąpienia ani wywarcia wpływu na zakończoną uprzednio ocenę dokonaną przez człowieka – bez odpowiedniej weryfikacji przez człowieka lub
4) system AI jest przeznaczony do wykonywania zadań przygotowawczych do oceny istotnej do celów przypadków wykorzystania wymienionych w załączniku III.

Jednak system AI, o którym mowa w załączniku III do Aktu, uznaje się za system wysokiego ryzyka zawsze wtedy, gdy system AI dokonuje profilowania osób fizycznych.

Celowo niniejszy artykuł nie omawia wszystkich obowiązków związanych z systemami AI wysokiego ryzyka, skupiając się jedynie na wybranych, które mogą mieć znaczenie w walce z dezinformacją.

2. System zarządzania ryzykiem

Pierwsza grupa obowiązków to te odnoszące się do analizy i mitygowania ryzyka. W odniesieniu do systemów AI wysokiego ryzyka powinien zostać wdrożony udokumentowany i utrzymywany **system zarządzania ryzykiem**[19], czyli, w rozumieniu Aktu ws. sztucznej inteligencji, ciągły, iteracyjny proces planowany i realizowany przez cały cykl życia systemu AI wysokiego ryzyka, wymagający regularnego, systematycznego przeglądu i aktualizacji. System zarządzania ryzykiem ma się składać z czterech etapów:
1) identyfikacja i analiza znanych i dającego się racjonalnie przewidzieć ryzyka, jakie dany system AI wysokiego ryzyka może stwarzać dla zdrowia, bezpieczeństwa lub praw podstawowych, podczas jego stosowania zgodnie z przeznaczeniem;
2) oszacowanie i ocena ryzyka, jakie może wystąpić, gdy system AI wysokiego ryzyka jest wykorzystywany zgodnie z jego przeznaczeniem i w warunkach dającego się racjonalnie przewidzieć niewłaściwego wykorzystania;
3) ocena innego mogącego wystąpić ryzyka, w oparciu o analizę danych zebranych z systemu monitorowania po wprowadzeniu do obrotu, o którym mowa w art. 72 Aktu;
4) przyjęcie odpowiednich i ukierunkowanych środków zarządzania ryzykiem zaprojektowanych w celu przeciwdziałania ryzyku zidentyfikowanemu zgodnie z pkt 1).

Ryzyka, o których mowa powyżej, dotyczą wyłącznie tych ryzyk, które można stosownie ograniczyć lub wyeliminować poprzez rozwój lub zaprojektowanie systemu AI wysokiego ryzyka lub poprzez zapewnienie odpowiednich informacji technicznych. Akt zawiera szczegółowe wymagania co do zarządzania ryzykiem, w tym testowania systemu AI wysokiego ryzyka. Wdrażając system zarządzania ryzykiem, dostawcy muszą wziąć

[17] Por. komunikat Komisji Europejskiej *AI Act*, 6.3.2024 r., AI Act | Shaping Europe's digital future (europa.eu) (dostęp: 30.3.2024 r.).
[18] Art. 6 Aktu ws. sztucznej inteligencji.
[19] Art. 9 Aktu ws. sztucznej inteligencji.