MAGDALENA DZIEDZIC

Przeciwdziałanie dezinformacji w kontekście wybranych regulacji Aktu o usługach cyfrowych oraz Aktu o Sztucznej Inteligencji

Combating Disinformation in the Context of Selected Regulations of the Digital Services Act and the Artificial Intelligence Act

Abstract

Disinformation is a phenomenon that is appearing more and more often in public debate and which, at the same time, poses a significant threat to the functioning of the state and society. In the European Union, actions have been taken for many years to counteract disinformation on the Internet. These actions consist of both regulatory initiatives (e.g., the Digital Services Act) and self-regulatory ones, such as codes of conduct for combating disinformation developed by international organizations and industry associations. Disinformation is one of the most important challenges to democratic processes, leading to a growing lack of trust in the media and government sources. The concept of disinformation itself is difficult to define because it is necessary to find a balance between guaranteeing an individual the right to freedom of speech and the need to impose restrictions aimed at combating false or distorted information. Currently, this difficulty is additionally influenced by the widespread use of generative AI.

SŁOWA KLUCZOWE: dezinformacja, sztuczna inteligencja, treści nielegalne, treści szkodliwe, cyfryzacja

KEYWORDS: disinformation, artificial intelligence, illegal content, harmful content, digitalisation

MAGDALENA DZIEDZIC – doktor nauk prawnych, Europejska Wyższa Szkoła Prawa i Administracji, ORCID – 0000-0003-1197-6917, e-mail: magdalenadziedzic@02.pl

HTTPS://DOI.ORG/10.36128/PRIW.VI53.1103

1 Wstęp

Dezinformacja to zjawisko pojawiające się coraz częściej w debacie publicznej, stanowiące realne zagrożenie dla funkcjonowania państwa i społeczeństwa. W erze szybkiego rozwoju technologii, dezinformacja pojawia się jako masowe i nieuniknione zjawisko. Nowe technologie mogą niejednokrotnie być wykorzystywane, w szczególności w mediach społecznościowych, do szerzenia dezinformacji w niespotykanej dotychczas skali, z bezprecedensową szybkością i precyzją, prowadząc do tworzenia spersonalizowanych sfer informacji, które wzmacniają przekaz kampanii dezinformacyjnych[1]. Przybierając rozmaite formy, manipuluje ona odbiorcami, a w konsekwencji wyrządza realne szkody. Dezinformacja w sposób negatywny oddziałuje na podziały społeczne, niepokoje społeczne, może wywoływać panikę i konflikty, oraz obniżać bezpieczeństwo publiczne i bezpieczeństwo narodowe. Wysoce realne stają się obecnie zagrożenia związane z zakłócaniem procesów demokratycznych za pomocą dezinformacji[2]. Internet w sposób istotny zwiększył ilość i różnorodność informacji dostępnych dla obywateli, oraz bardzo zmienił sposób, w jaki uzyskują oni dostęp do informacji i korzystają z nich. Media internetowe stają się obecnie podstawowym źródłem informacji zwłaszcza dla młodszego pokolenia użytkowników.

W Unii Europejskiej już od wielu lat podejmowane są działania ukierunkowane na przeciwdziałanie dezinformacji w Internecie, a samą walkę z dezinformację dostrzeżono jako cel regulacyjny na długo przed przyjęciem Aktu o usługach cyfrowych[3] czy Aktu o Sztucznej Inteligencji[4]. W komunikacie Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów z 26.4.2018 r. zatytułowanym „Zwalczanie dezinformacji w Internecie: podejście europejskie" zauważono, że wpływ dezinformacji różni się w zależności od społeczeństwa, zależnie od poziomu wykształcenia, kultury demokratycznej, zaufania do instytucji, integracyjnego charakteru systemów wyborczych, roli pieniędzy w procesach politycznych oraz nierówności społecznych

1 Komunikat Komisji z 26.4.2018 r.. https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52018DC0236. [dostęp: 11.11.2024].

2 https://www.euractiv.com/section/disinformation/news/disinformation-campaigns-likely-to-undermine-eu-elections-experts-say/. [dostęp: 11.11.2024].

3 Dz.Urz. L Nr 277 z 27.10.2022 r., s. 1 ze zm.; dalej jako: Akt o usługach cyfrowych (dalej jako DSA).

4 Dz.Urz. L Nr 362 z 12.7.2024 r., s.1 ze zm.; dalej jak Akt o sztucznej inteligencji (dalej jako AIA).

i gospodarczych. W dłuższej perspektywie przeciwdziałanie dezinformacji odniesie skutek tylko wtedy, gdy będzie mu towarzyszyła wyraźna wola polityczna, by wzmocnić zbiorową odporność, wspierając tym samym zachowania demokratyczne i poszanowanie wartości europejskich. Dezinformacja jest olbrzymim i niedrogim – a często zyskownym – narzędziem wpływów. Dotychczas większość znanych przypadków dotyczyła artykułów, niekiedy uzupełnionych autentycznymi zdjęciami lub treściami audiowizualnymi wyjętymi z kontekstu. Obecnie dostępna jest jednak nowa, tania i łatwa w użyciu technologia pozwalająca na tworzenie fałszywych obrazów i treści audiowizualnych (tzw. *deep fakes*), stwarzająca bardziej skuteczne możliwości manipulowania opinią publiczną.

Z kolei 16 czerwca 2022 r. przedstawiony został tzw. udoskonalony kodeks postępowania w zakresie zwalczania dezinformacji[5]. Nowa, ulepszona wersja kodeksu została przyjęta w rezultacie przedstawienia przez Komisję w maju 2021 r. wytycznych dla sygnatariuszy, w których zwróciła się o dopracowanie kodeksu we wszystkich obszarach. Dokumenty powyższe mają charakter samoregulacyjny. Udoskonalony kodeks postępowania zawiera 44 zobowiązania i 128 konkretnych środków, w ramach których sygnatariusze zobowiązali się do podjęcia działań na kilku polach, takich jak: demonetyzacja rozpowszechniania dezinformacji; zapewnienie przejrzystości reklamy politycznej; wzmocnienie pozycji użytkowników; zacieśnienie współpracy z podmiotami weryfikującymi fakty; oraz zapewnienie naukowcom lepszego dostępu do danych. Kodeks postępowania w zakresie zwalczania dezinformacji jest pierwszym tego rodzaju narzędziem, za pomocą którego odpowiednie podmioty w branży uzgodniły – po raz pierwszy w 2018 r. – standardy samoregulacji w celu zwalczania dezinformacji.

Dezinformacja stanowi ponadto broń w wojnie informacyjnej. W obliczu łatwego dostępu do publikowania w internecie fałszywe informacje oddziałują na postawy i zachowania obywateli i polityków. Niejednokrotnie celami dezinformacji stają się również tradycyjne media i instytucje państwowe, a sama dezinformacja może stanowić narzędzie agresywnej polityki państw, czy nawet element wojny hybrydowej, o czym mówi się przede wszystkim w kontekście działań rosyjskich związanych z agresją

5 Por. https://digital-strategy.ec.europa.eu/pl/library/2022-strengthened-code-practice-disinformation. [dostęp: 28.11.2023]; dalej jako: Udoskonalany kodeks postępowania w zakresie zwalczania dezinformacji.

Rosji na Ukrainę[6]. Kampanie dezinformacyjne prowadzone przez państwa trzecie mogą stanowić element zagrożeń hybrydowych dla bezpieczeństwa wewnętrznego, w tym procesów wyborczych, w szczególności w połączeniu z cyberatakami[7]. Walka informacyjna została w sposób zdecydowany potraktowana jako jeden z obszarów rosyjskiej doktryny wojskowej[8].

2 Pojęcie dezinformacji

Przeciwdziałanie dezinformacji w Internecie, jako cel regulacyjny prawodawcy europejskiego, został dostrzeżony jeszcze przed przyjęciem Aktu o usługach cyfrowych. W styczniu 2018 r. Komisja Europejska powołała Niezależną Grupę Ekspertów Wysokiego Szczebla w kwestii *fake news* oraz dezinformacji online. Niezależna Grupa Ekspertów Wysokiego Szczebla opracowała raport z marca 2018 r., zatytułowany *A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation*[9]. Zgodnie z definicją przyjętą w raporcie, „dezinformację" (w tym online) należy odróżnić po pojęcia „fake news"[10]. W konsekwencji, dezinformacja ujęta została jako „false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit"[11], tj. jako wszelkie formy fałszywych, niedokładnych lub wprowadzających w błąd informacji stworzonych, prezentowanych i promowanych w celu intencjonalnego wyrządzenia szkody publicznej lub dla zysku. Nie obejmuje ona kwestii wynikających z tworzenia i rozpowszechniania w internecie nielegalnych treści (w szczególności zniesławienia, mowy nienawiści, podżegania do przemocy), które podlegają regulacyjnym środkom zaradczym na mocy prawa UE lub prawa

6 Katarzyna Chałubińska-Jentkiewicz, Monika Nowikowska, *Prawo mediów* (Warszawa: C.H. Beck, 2022), 120.

7 We „Wprowadzeniu" (pkt 1) do Komunikatu Komisji z 26.4.2018 r. wskazuje się m.in. na wzmożone działania Rosji w zakresie kampanii dezinformacyjnych.

8 https://www.rusemb.org.uk/press/2029. [dostęp: 11.11.2024].

9 Por. https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en, s. 10-11. [dostęp: 12.11.2024].

10 Ibidem, 10-11.

11 Ibidem, 10.

krajowego. Nie obejmuje również innych form celowego, ale niewprowadzającego w błąd zniekształcania faktów, takich jak satyra i parodia[12].

Jednocześnie w tym samym raporcie Niezależna Grupa Ekspertów Wysokiego Szczebla zaleca podejście wielowymiarowe do problemu dezinformacji online, oparte na szeregu powiązanych i wzajemnie wzmacniających się elementach, tzw. 5 filarach: zwiększenia przejrzystości informacji online, obejmującego odpowiednie i zgodne z prywatnością udostępnianie danych o systemach, które umożliwiają ich obieg online; promowanie umiejętności korzystania z mediów i informacji w celu przeciwdziałania dezinformacji i pomocy użytkownikom w poruszaniu się po środowisku mediów cyfrowych; opracowywania narzędzi umożliwiających użytkownikom i dziennikarzom radzenie sobie z dezinformacją i wspieranie pozytywnego zaangażowania w szybko rozwijające się technologie informacyjne; ochrony różnorodności i zrównoważonego rozwoju europejskiego ekosystemu mediów informacyjnych oraz promowania ciągłych badań nad wpływem dezinformacji w Europie w celu oceny podejmowanych środków[13].

W komunikacie Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów z 26.4.2018 r. „Zwalczanie dezinformacji w Internecie: podejście europejskie" zauważono, że dezinformację w internecie rozprzestrzenia się w trojaki sposób, a mianowicie: „oparty na algorytmach", „oparty na reklamach" oraz „oparty na technologii"[14]. W raporcie tym dezinformację zdefiniowano jako „[d]ezinformację należy rozumieć jako możliwe do zweryfikowania nieprawdziwe lub wprowadzające w błąd informacje, tworzone, przedstawiane i rozpowszechniane w celu uzyskania korzyści gospodarczych lub wprowadzenia w błąd opinii publicznej, które mogą wyrządzić szkodę publiczną". W Komunikacie wskazuje się, że „szkoda publiczna w przypadku dezinformacji polega na zagrożeniach dla demokratycznych procesów politycznych i kształtowania polityki oraz dla dóbr politycznych, które obejmują bezpieczeństwo, ochronę zdrowia obywateli Unii Europejskiej czy środowisko naturalne"[15]. Komisja w swoim Komunikacie przedstawiła nadrzędne zasady i cele, które mają pomóc w działaniom zwalczającym dezinformację w sieci, m.in.: poprawę

12 Ibidem, 11.

13 Ibidem, 5.

14 Por. pkt 2.2. Komunikatu z 26.4.2018 r. „Kontekst i główne przyczyny dezinformacji", 6.

15 Por. pkt 2.1 Komunikatu z 26.4.2018 r. „Zakres i przyczyny dezinformacji w internecie", 4.

przejrzystości jeżeli chodzi o pochodzenie, sposoby tworzenia, finansowania, rozpowszechniania i ukierunkowywania informacji w celu umożliwienia obywatelom oceny treści, do których uzyskują dostęp w internecie; promowanie różnorodności informacji w celu umożliwienia obywatelom podejmowania świadomych decyzji opartych na krytycznym myśleniu; wspieranie wiarygodności informacji poprzez wskazywanie, czy można im ufać, w szczególności z pomocą zaufanych podmiotów sygnalizujących; oraz opracowanie rozwiązań integracyjnych, skuteczne długofalowe rozwiązania wymagają zwiększenia wiedzy o problemie, lepszej umiejętności korzystania z mediów, oraz współpracy organów publicznych, platform internetowych, reklamodawców, zaufanych podmiotów sygnalizujących, dziennikarzy i grup medialnych[16].

Ze względu na coraz większą popularność tego zjawiska dezinformacja jest na różne sposoby kategoryzowana. Raport Rady Europy zatytułowany *Information Disorder: Toward an interdisciplinary framework for research and policy making* wyróżnia trzy kategorie dezinformacji we współczesnym świecie: „dis-information", czyli informację nieprawdziwą mającą na celu wyrządzenie krzywdy jakiejś osobie, organizacji, grupie społecznej lub państwu; „misinformation", czyli informację nieprawdziwą, ale stworzoną bez intencji skrzywdzenia kogokolwiek; oraz „mal-information", czyli informację opartą na rzeczywistości, której wykorzystanie ma na celu wyrządzenie krzywdy jakiejś osobie, organizacji lub państwu[17].

Na gruncie polskich przepisów prawnych brak jest legalnej definicji dezinformacji. Pojęcie to definiowane jest przez doktrynę. Przez dezinformację rozumie się tworzenie i rozpowszechnianie wprowadzających w błąd lub fałszywych informacji w celu wyrządzenia szkody wizerunkowi kraju wybranego za cel[18]. Niezależnie od przybranej formy, dezinformację zdefiniować można jako proces, który polega na celowym, błędnym informowaniu lub jako sytuację, w której brakuje rzetelnych informacji. Rozwój technologiczny sprawia, że zjawisko dezinformacji w ostatnich latach rozprzestrzenia się i wywiera coraz większy wpływ na społeczeństwo, politykę i prawo[19].

16 Por. pkt 2.2. Komunikatu z 26.4.2018 r. „Kontekst i główne przyczyny dezinformacji", 7.

17 Raport Rady Europy, *Information Disorder; Toward an interdisciplinary framework for research and policymaking*, październik 2017, 20. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for--research-and-policy-making.html. [dostęp: 12.11.2024].

18 Robert Brzeski, *Dezinformacja* (Warszawa, 2011), 107.

19 Ewa Kurowska-Tober, Franciszek Ignacy Fortuna, Dezinformacja a ochrona danych osobowych" *Prawo Nowych Technologii*, nr 1 (2024): 9-24.

3 Dezinformacja w akcie o usługach cyfrowych

Akt o usługach cyfrowych został opublikowany 27 października 2022 r. W świetle jego art. 93 ust. 2 rozporządzenie to będzie stosowane od 17 lutego 2024 r., jakkolwiek niektóre jego przepisy stosuje się od 16 listopada 2022 r. Zgodnie z podstawowym założeniem wyrażonym w motywie 1 aktu, transformacja cyfrowa i wzmożone korzystanie z tych usług przyniosły również nowe zagrożenia i wyzwania dla indywidualnych odbiorców danej usługi, przedsiębiorstw i całego społeczeństwa[20]. Jak zauważono w motywie 2, państwa członkowskie coraz częściej wprowadzały przepisy krajowe dotyczące spraw objętych zakresem stosowania rozporządzeniem lub rozważały wprowadzenie takich przepisów, nakładając w szczególności wymogi w zakresie należytej staranności, obowiązujące dostawców usług pośrednich w kwestii tego, w jaki sposób podmioty te powinny podchodzić do problemu nielegalnych treści, dezinformacji online i innych zagrożeń społecznych. Takie rozbieżne przepisy krajowe miały negatywny wpływ na rynek wewnętrzny, utrudniając harmonizację, co uzasadniało objęcie powyższej problematyki (w tym w zakresie dezinformacji online) materią rozporządzenia. W związku z tym ustanowiono mechanizmy reagowania przez dostawców usług pośrednich na informacje o nielegalnych treściach oraz określono odpowiedzi na nakazy kierowane przez właściwe krajowe organy sądowe lub administracyjne[21]. Podobnie w motywie 9 Aktu o usługach cyfrowych wskazuje się na zasadność nieprzyjmowania czy utrzymywania (w razie ich obowiązywania), dodatkowych przepisów krajowych zawierających wymogi odpowiadające wymogom rozporządzenia, w tym odnośnie do dezinformacji online. Takie regulacje miałyby negatywnie wpływać na zharmonizowanie przepisów objętych materią rozporządzenia. Oznacza to, że przyjęte rozporządzenie – Akt o usługach cyfrowych – ma na celu pełną harmonizację przepisów znajdujących zastosowanie do usług pośrednich na rynku wewnętrznym, również dla przeciwdziałania rozpowszechnianiu dezinformacji oraz zapewnienia bezpiecznego i przewidywalnego środowiska internetowego.

W motywie 69 Aktu o usługach cyfrowych kwestia dezinformacji pojawiła się w kontekście platform internetowych i ich dostawców. Dotyczy to sytuacji, gdy reklamy zamieszczane na platformach są skonstruowane w oparciu o techniki targetowania lub inne techniki manipulacyjne, które

20 Pkt 1 DSA.
21 Pkt 2 Preambuły DSA.

mogą przyczyniać się do rozwijania kampanii dezinformacyjnych lub dyskryminujących niektóre grupy społeczne. Zgodnie z motywem 83 duże platformy i wyszukiwarki internetowe mogą być jedną z kategorii ryzyk dotyczących faktycznego lub przewidywalnego negatywnego wpływu działań i kampanii dezinformacyjnych. Takie platformy i wyszukiwarki mogą stanowić potencjalną przestrzeń do organizowania skoordynowanych działań i kampanii dezinformacyjnych, w szczególności dotyczących zdrowia publicznego, a także mogą stymulować uzależnienia behawioralne u odbiorców usług[22]. W motywie 104 Aktu, problematyka dezinformacji pojawiła się w kontekście kodeksów postępowań, które, zgodnie z założeniem prawodawcy unijnego, przyczyniać się mają do realizacji jego postanowień w drodze samoregulacji, z kolei samą dezinformację lub działania manipulacyjne, jako przykład możliwego negatywnego wpływu ryzyka systemowego na społeczeństwo i demokrację, zalicza się do obszarów, które powinny być uwzględniane przy tworzeniu takich kodeksów. Jak zauważono, w przypadku dezinformacji przestrzeganie i stosowanie danego kodeksu postępowania przez bardzo dużą platformę internetową lub bardzo dużą wyszukiwarkę internetową można uznać za odpowiedni środek zmniejszający ryzyko. Ponadto, w motywie 106 Aktu o usługach cyfrowych zwrócono uwagę na szczególną rolę Ulepszonego kodeksu postępowania w zakresie zwalczania dezinformacji z 16 czerwca 2022 r. jako podstawę działań o charakterze samoregulacyjnym na poziome Unii Europejskiej[23].

Co istotne, prawodawca unijny nie zdecydował się na wprowadzenie definicji pojęcia dezinformacji w Akcie o usługach cyfrowych. Na gruncie rozporządzenia posłużono się natomiast pojęciem „nielegalnych treści" i „treści szkodliwych". Zgodnie z motywem 12, pojęcie „nielegalnych treści" należy zdefiniować szeroko, tak by obejmowało informacje dotyczące nielegalnych treści, produktów, usług i działań. Należy je rozumieć w szczególności jako odnoszące się do informacji, niezależnie od ich formy, które zgodnie z obowiązującym prawem są albo same w sobie nielegalne, takich jak nielegalne nawoływanie do nienawiści lub treści o charakterze terrorystycznym i niezgodne z prawem treści dyskryminujące, albo które stają się nielegalne na mocy obowiązujących przepisów ze względu na fakt, iż odnoszą się one do nielegalnych działań. W doktrynie zwraca się uwagę,

22 Maciej Kubiak, Michał Majcher, „Prywatyzacja oraz komercjalizacja przestrzeni wymiany myśli i publicznej debaty jako wyzwanie regulacyjne dla przeciwdziałania dezinformacji" *Prawo Nowych Technologii*, nr 1 (2024): 31.

23 Aleksandra Auleytner, Marcin J. Stępień, „Dezinformacja a Akt o usługach cyfrowych" *Prawo Nowych Technologii*, nr 3 (2023): 142.

że brak zdefiniowania pojęcia dezinformacji wydaje się być pewną wadą tego, jak wskazuje się jednego „z najważniejszych w ostatnich dwudziestu latach aktów prawnych z zakresu prawa Internetu". Jakkolwiek motywem dla takiego rozwiązania było zaliczenie dezinformacji do „szkodliwych", aczkolwiek nie „nielegalnych" treści[24]. Takie podejście ustawodawcy europejskiego wynikało z obawy przed znaczną ingerencją w wolność słowa z uwagi na nieostre znaczenie treści szkodliwych. W konsekwencji, ustalenie znaczenia dezinformacji pozostawiono praktyce i doktrynie. Przy ustaleniu tego znaczenia pomocne mogą okazać się zatem definicje dezinformacji wynikające z innych aktów unijnych, tj. zaproponowana przez Niezależną Grupę Ekspertów Wysokiego Szczebla w swoim raporcie z 2018 r., czy ustalona w samoregulacyjnym Udoskonalanym kodeksie postępowania w zakresie zwalczania dezinformacji.

Jeszcze na etapie prac nad Aktem o usługach cyfrowych, na gruncie uzasadnienia wniosku zawierającego projekt rozporządzenia, przedstawiano propozycje wyróżnienia także „treści szkodliwych". Jednak, jak zauważono wówczas, wśród zainteresowanych stron panuje zasadnicza zgoda, że treści szkodliwe (choć nie, lub przynajmniej niekoniecznie, nielegalne) nie powinny być zdefiniowane w akcie o usługach cyfrowych i nie powinny podlegać obowiązkom usunięcia, gdyż jest to delikatny obszar o poważnych konsekwencjach dla ochrony wolności wypowiedzi[25]. W efekcie treści takie, choć niebędące nielegalnymi, jakkolwiek mogące być uznane za szkodliwe, to z uwagi na ich niejednoznaczny charakter i wzgląd na zachowanie swobody wypowiedzi, nie powinny podlegać konieczności usunięcia w odróżnieniu od treści nielegalnych. Jednocześnie postulowano przy tym niedefiniowanie treści szkodliwych. Zgodnie ze stanowiskiem prezentowanym w polskiej doktrynie, zaliczenie dezinformacji do „treści szkodliwych", tj. takich, które formalnie nie są nielegalne, ale z innych powodów mogą zostać uznane za nieetyczne i społecznie niepożądane (np. dezinformacja, nagość, przemoc, rasizm, ksenofobia), należy uznać za trafne, tym bardziej jeżeli uwzględnić postanowienia wynikające ze wspomnianego wcześniej motywu 104 Aktu o usługach cyfrowych[26].

24 Ibidem, 142.

25 Ibidem, 144.

26 Xawery Konarski, „Unijny Akt o Usługach Cyfrowych – cele uchwalenia, zakres stosowania oraz najważniejsze obowiązki dostawców usług pośrednich" *Prawo Nowych Technologii*, nr 3 (2022): 37.

4 Dezinformacja w Akcie o sztucznej inteligencji

Akt został przyjęty przez Parlament Europejski w środę 13 marca 2024 r., następnie 12 lipca 2024 r. został opublikowany w Dzienniku Urzędowym, a wszedł w życie 1 sierpnia 2024 r. Akt 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. W konsekwencji podmioty podlegające pod Akt ws. sztucznej inteligencji muszą być w stanie w pełni się do niego dostosować w ciągu najbliższych 2 lat, z dodatkowym zapasem jednego roku w zakresie obowiązków dotyczących systemów AI wysokiego ryzyka. Podstawowym pojęciem zdefiniowanym w Akcie ds. sztucznej inteligencji jest system AI. Został on zdefiniowany w Akcie jako oprogramowanie opracowane przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I (tj. system maszynowy zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu[27]), które może – dla danego zestawu celów określonych przez człowieka – generować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję. 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*[28].

Z punktu widzenia przeciwdziałania dezinformacji podstawowe znaczenie ma zakres podmiotowy rozporządzenia, który obejmuje m.in. dostawców wprowadzających do obrotu lub oddających do użytku systemy sztucznej inteligencji w Unii, niezależnie od tego, czy dostawcy ci mają siedzibę w Unii, czy w państwie trzecim; użytkowników systemów sztucznej inteligencji, którzy znajdują się w Unii; oraz dostawców i użytkowników

27 Marlena Czapska, „Akt o sztucznej inteligencji a walka z dezinformacją – przegląd wybranych obowiązków dostawców i podmiotów stosujących systemy AI" *Prawo Nowych Technologii*, nr 1 (2024): 47.

28 Ibidem, 142.

systemów sztucznej inteligencji, którzy znajdują się w państwie trzecim, jeżeli wyniki działania systemu są wykorzystywane w Unii[29]. Zgodnie z definicjami przyjętymi w AIA, dostawca oznacza osobę fizyczną lub prawną, organ publiczny, agencję lub inny podmiot, które opracowują system sztucznej inteligencji lub zlecają jego opracowanie w celu wprowadzenia go do obrotu lub oddania go do użytku pod własną nazwą handlową lub własnym znakiem towarowym – odpłatnie lub nieodpłatnie[30]; natomiast użytkownik oznacza osobą fizyczną lub prawną, organ publiczny, agencję lub inny podmiot, które korzystają z systemu sztucznej inteligencji pod swoją kontrolą, z wyjątkiem sytuacji, gdy system sztucznej inteligencji jest wykorzystywany w ramach osobistej działalności pozazawodowej[31]. Obowiązki wynikające z Aktu w sprawie sztucznej inteligencji zależą od tego, którego z wyżej wymienionych podmiotów dotyczą, oraz od tego jaki poziom ryzyka generowany jest przez dany rodzaj systemu AI. Zgodnie z postanowieniami Aktu, podejście do systemów AI opiera się na analizie ryzyka, w związku z czym, wyróżnia się cztery poziomy ryzyka: po pierwsze, ryzyko nieakceptowalne (tzw. zakazane praktyki w zakresie sztucznej inteligencji), co wyraża się w ustaleniu katalogu zakazanych systemów AI[32]; po drugie, systemy wysokiego ryzyka, których klasyfikacja była kontrowersyjna, gdyż generować ona może istotne obciążenia dla podmiotów; ten obszar działania systemów AI jest wysoce regulowany, gdyż może mieć negatywny wpływ na zdrowie i bezpieczeństwo ludzi, ich prawa podstawowe lub środowisko, zwłaszcza kiedy systemy zawiodą lub będą niewłaściwie wykorzystywane; systemy sztucznej inteligencji należące do tej klasy ryzyka muszą spełniać szereg określonych wymogów, aby mogły zostać wprowadzone na rynek i eksploatowane w UE[33]; po trzecie, ograniczone ryzyko, wyrażające się w ustaleniu obowiązków w zakresie transparentności w odniesieniu do pewnych rodzajów systemów AI; oraz po czwarte, tego typu systemy AI nie będą podlegały szczególnym obowiązkom.

Systemy wysokiego ryzyka zdają się być najbardziej narażone na zagrożenie działaniami dezinformacyjnymi, w związku z czym to właśnie te systemy, oczywiście nie uwzględniając systemów ryzyka nieakceptowalnego, charakteryzują się najbardziej rygorystycznymi regulacjami.

29 Art. 2 pkt 1 AIA.
30 Art. 3 pkt 2 AIA.
31 Art. 3 pkt 4 AIA.
32 Art. 5 AIA.
33 Art. 6 AIA.

Mając na uwadze zasady klasyfikacji systemów wysokiego ryzyka, należy wskazać, że dany system sztucznej inteligencji zostanie uznany za system wysokiego ryzyka, bez względu na to, czy zostanie on wprowadzony do obrotu lub oddany do użytku niezależnie od produktów, o których mowa poniżej, jeżeli spełnione są następujące warunki: 1) system sztucznej inteligencji jest przeznaczony do wykorzystywania jako związany z bezpieczeństwem element produktu objętego unijnym prawodawstwem harmonizacyjnym wymienionym w załączniku II lub sam jest takim produktem; produkt, którego związanym z bezpieczeństwem elementem jest system sztucznej inteligencji, lub sam system sztucznej inteligencji jako produkt podlegają – na podstawie unijnego prawodawstwa harmonizacyjnego wymienionego w załączniku II – ocenie zgodności przeprowadzanej przez osobę trzecią w celu wprowadzenia tego produktu do obrotu lub oddania go do użytku. Dodatkowo, za systemy wysokiego ryzyka uznaje się również systemy sztucznej inteligencji, o których mowa w załączniku III do Aktu, tj. systemy działające w określonych w tym załączniku obszarach, takich jak systemy sztucznej inteligencji przeznaczone do stosowania w celu zdalnej identyfikacji biometrycznej osób fizycznych „w czasie rzeczywistym" i „post factum"; systemy sztucznej inteligencji przeznaczone do stosowania jako związane z bezpieczeństwem elementy procesów zarządzania i obsługi ruchu drogowego oraz zaopatrzenia w wodę, gaz, ciepło i energię elektryczną; systemy sztucznej inteligencji przeznaczone do kształcenia i szkoleń zawodowych, w tym stosowania w celu podejmowania decyzji o dostępie do instytucji edukacyjnych i instytucji szkolenia zawodowego lub nadawania osobom przydziału do tych instytucji; systemy sztucznej inteligencji przeznaczone do zatrudnienia i zarządzania pracownikami, w tym wykorzystania w celu rekrutacji lub wyboru osób fizycznych, w szczególności w przypadku informowania o wakatach, selekcji lub filtrowania podań o pracę, oceny kandydatów w trakcie rozmów kwalifikacyjnych lub testów; systemy sztucznej inteligencji służące do zapewnienia dostępu do podstawowych usług prywatnych oraz usług i świadczeń publicznych, a także korzystanie z nich: systemy sztucznej inteligencji służące do ścigania przestępstw, do zarządzania migracją, azylem i kontrolą graniczną, oraz systemy dla organów sądowych służące pomocą w badaniu i interpretacji stanu faktycznego i przepisów prawa oraz w stosowaniu prawa do konkretnego stanu faktycznego[34].

34 Załącznik III do AIA.

Z punktu widzenia przeciwdziałania dezinformacji istotne znaczenie ma ustanowienie, wdrożenie, udokumentowanie i utrzymanie odpowiedniego systemu zarządzenia ryzykiem w odniesieniu do systemów sztucznej inteligencji wysokiego ryzyka[35]. System ten obejmuje ciągły proces realizowany przez cały cykl życia systemu sztucznej inteligencji wysokiego ryzyka, wymagający regularnej, systematycznej aktualizacji. Składa się on m.in. po pierwsze, z obowiązku identyfikacji i analizy znanego i dającego się przewidzieć ryzyka związanego z każdym systemem sztucznej inteligencji wysokiego ryzyka; po drugie, należy oszacować i ocenić ryzyko, jakie może wystąpić podczas wykorzystywania systemu sztucznej inteligencji wysokiego ryzyka zgodnie z jego przeznaczeniem i w warunkach dającego się racjonalnie przewidzieć niewłaściwego wykorzystania; w następnej kolejności należy ocenić inne mogące wystąpić ryzyka na podstawie analizy danych zebranych z systemu monitorowania po wprowadzeniu do obrotu; w dalszej kolejności należy zaadoptować odpowiednie środki zarządzania ryzykiem wynikające z powszechnie uznawanego stanu techniki. Mając na uwadze najodpowiedniejsze środki zarządzania ryzykiem wskazać należy na, że obejmują one m.in. po pierwsze, obowiązek eliminacji lub wyłączenia ryzyka w możliwie największym stopniu poprzez odpowiedni projekt systemu i proces jego opracowywania; po drugie, obowiązek w stosownych przypadkach podjęcia odpowiednich środków służących ograniczeniu i kontroli ryzyka, którego nie można wyeliminować; po trzecie natomiast, obowiązek dostarczenia odpowiednich informacji, w szczególności w odniesieniu do systemu wysokiego ryzyka. W przypadku systemów wysokiego ryzyka zagrożenie takie mogłoby przykładowo obejmować możliwość wykorzystania systemu do stworzenia dezinformacji, co skutkowałoby koniecznością ich usunięcia[36].

Kolejnym obowiązkiem ustanowionym przez prawodawcę europejskiego w Akcie o sztucznej inteligencji, jaki w pozytywny sposób wpłynąć może na przeciwdziałanie dezinformacji, jest obowiązek ludzkiego nadzoru[37]. Jak zauważono, systemy sztucznej inteligencji wysokiego ryzyka projektuje się i przygotowuje w taki sposób, w tym poprzez przyjęcie odpowiednich narzędzi interfejsu człowiek-maszyna, aby osoby fizyczne były w stanie je skutecznie nadzorować w okresie wykorzystywania systemu sztucznej inteligencji wysokiego ryzyka. Nadzór ze strony człowieka ma pozwolić

35 Art. 9 AIA.
36 Czapska, *Akt o sztucznej inteligencji*, 49.
37 Art. 14 AIA.

uniknąć ryzyka dla zdrowia, bezpieczeństwa lub praw podstawowych lub mitygować takie ryzyko, które może się pojawić, gdy system sztucznej inteligencji wysokiego ryzyka jest wykorzystywany zgodnie z jego przeznaczeniem lub w warunkach dającego się racjonalnie przewidzieć niewłaściwego wykorzystania, w szczególności gdy takie ryzyko utrzymuje się pomimo stosowania innych wymogów wskazanych w Akcie. Środki nadzoru powinny odpowiadać poziomowi ryzyka i autonomii oraz okoliczności wykorzystywania systemu sztucznej inteligencji wysokiego ryzyka i muszą być zapewnione za pomocą co najmniej jednego z następujących rodzajów środków: po pierwsze, określonych i wbudowanych, jeżeli jest to technicznie wykonalne, w system sztucznej inteligencji wysokiego ryzyka przez dostawcę przed wprowadzeniem systemu do obrotu lub oddaniem go do użytku; po drugie, określonych przez dostawcę przed wprowadzeniem systemu sztucznej inteligencji wysokiego ryzyka do obrotu lub oddaniem go do użytku i które to środki nadają się do wdrożenia przez użytkownika.

Kolejnym obowiązkiem przewidzianym w Akcie, mającym na celu przeciwdziałanie dezinformacji, jest wymóg w zakresie przejrzystości, odnoszący się do określonych rodzajów systemów AI, które nie są systemami wysokiego ryzyka, ale wciąż są poddane regulacjom Aktu[38]. W pierwszej kolejności zauważono, że dostawcy mają obowiązek zapewnienia, aby systemy sztucznej inteligencji zaprojektowane do wchodzenia w interakcję z osobami fizycznymi opracowywano w taki sposób, aby osoby fizyczne były informowane o tym, że prowadzą interakcję z systemem sztucznej inteligencji, chyba że okoliczności i kontekst korzystania z systemu jednoznacznie na to wskazują. Wskazano tutaj na jedno wyłączenie, gdyż obowiązek ten nie ma zastosowania do systemów sztucznej inteligencji zatwierdzonych z mocy prawa do celów wykrywania przestępstw, przeciwdziałania przestępstwom, prowadzenia dochodzeń/śledztw w związku z przestępstwami i ścigania ich sprawców, chyba że systemy te udostępnia się ogółowi społeczeństwa na potrzeby składania zawiadomień o popełnieniu przestępstwa. Podobnie użytkownicy systemów rozpoznawania emocji lub systemów kategoryzacji biometrycznej mają obowiązek poinformowania osób fizycznych, wobec których systemy te są stosowane, o fakcie ich stosowania. Kolejny obowiązek mający na celu przeciwdziałanie dezinformacji dotyczy generatywnej sztucznej inteligencji. Użytkownicy systemu sztucznej inteligencji, w tym systemów AI ogólnego przeznaczenia, który generuje obrazy, treści dźwiękowe lub treści wideo, które

38 Art. 52 AIA.

łudząco przypominają istniejące osoby, obiekty, miejsca lub inne podmioty lub zdarzenia, lub który tymi obrazami i treściami manipuluje, przez co osoba będąca ich odbiorcą mogłaby niesłusznie uznać je za autentyczne lub prawdziwe („deepfake"), mają obowiązek ujawnienia, że dane treści zostały wygenerowane lub zmanipulowane przez system sztucznej inteligencji. W kontekście przeciwdziałania dezinformacji bardzo istotne jest zapewnienie użytkownikowi możliwości odróżnienia treści stworzonych przez człowieka od tych sztucznie wygenerowanych.

5 Zakończenie

Akt o usługach cyfrowych oraz Akt o sztucznej inteligencji w założeniu ustawodawcy unijnego mają stanowić istotny instrument w kontekście zwalczania dezinformacji. W celu przeciwdziałania dezinformacji podejmowane są obecnie wysiłki, takie jak: weryfikacja informacji, kampanie uświadamiające, inicjatywy oparte na współpracy międzyinstytucjonalnej czy badania lub programy globalne[39]. Wiele z tych czynności obejmuje dementowanie fałszywych treści oraz edukowanie użytkowników[40]. Są to ważne i potrzebne inicjatywy, których skuteczność może przynieść rezultaty w dłuższej perspektywie czasowej, a nie bezpośrednio w okresie intensyfikacji działań dezinformacyjnych. Akt ws. sztucznej inteligencji zawiera szereg obowiązków, które zapewne przyczynią się do przeciwdziałania dezinformacji. Przepisy te zmierzają do uregulowania systemów AI, których stosowanie może wiązać się z wysokim ryzykiem, jak również na przejrzystość w zakresie oznaczania treści wygenerowanych lub zmanipulowanych przez AI. Działania te powinny doprowadzić do ograniczenia rozprzestrzeniania się i oddziaływania dezinformacji na społeczeństwo.

39 Klaudia Rosińska, „Dezinformacja – zagrożenia dla polskiego społeczeństwa w okresie wyborczym" *Prawo Nowych Technologii*, nr 1 (2024): 21.

40 Klaudia Rosińska, „Metody zwalczania szkodliwego wpływu fake newsów", [w:] Oblicza fake newsa. Perspektywy naukowych analiz zjawiska fałszywych wiadomości (Warszawa: Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego, 2021), 165-182.

Bibliografia

Auleytner Aleksandra, Marcin J. Stępień, „Dezinformacja a Akt o usługach cyfrowych" *Prawo Nowych Technologii*, nr 3 (2023): 139-143.

Brzeski Robert, *Dezinformacja.* Warszawa 2011.

Chałubińska-Jentkiewicz Katarzyna, „Dezinformacja jako akt agresji w cyberprzestrzeni" *Cybersecurity and Law*, Nr 5 (2021): 9-24.

Chałubińska-Jentkiewicz Katarzyna, Monika Nowikowska, *Prawo mediów*. Warszawa: C.H. Beck, 2022.

Czapska Marlena, „Akt o sztucznej inteligencji a walka z dezinformacją – przegląd wybranych obowiązków dostawców i podmiotów stosujących systemy AI" *Prawo Nowych Technologii*, nr 1 (2024): 47-49.

Konarski Xawery, „Unijny Akt o Usługach Cyfrowych – cele uchwalenia, zakres stosowania oraz najważniejsze obowiązki dostawców usług pośrednich" *Prawo Nowych Technologii*, nr 3 (2022): 33-34.

Kubiak Maciej, Michał Majcher, „Prywatyzacja oraz komercjalizacja przestrzeni wymiany myśli i publicznej debaty jako wyzwanie regulacyjne dla przeciwdziałania dezinformacji" *Prawo Nowych Technologii*, nr 1 (2024): 27-35.

Kurowska-Tober Ewa, Franciszek Ignacy Fortuna, „Dezinformacja a ochrona danych osobowych" *Prawo Nowych Technologii*, nr 1 (2024): 9-24.

Rosińska Klaudia, „Dezinformacja – zagrożenia dla polskiego społeczeństwa w okresie wyborczym" *Prawo Nowych Technologii*, nr 1 (2024): 17-22.

Rosińska Klaudia, „Metody zwalczania szkodliwego wpływu fake newsów", [w:] *Oblicza fake newsa. Perspektywy naukowych analiz zjawiska fałszywych wiadomości*. 165-182. Warszawa: Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego, 2021.