Dziennik Urzędowy
Unii Europejskiej

PL
Seria C

C/2026/6005

12.11.2026

KOMUNIKAT KOMISJI

Wytyczne dotyczące wdrażania obowiązków w zakresie przejrzystości w odniesieniu do niektórych systemów AI na podstawie art. 50 rozporządzenia (UE) 2024/1689 („akt w sprawie AI”)

(C/2026/6005)

Spis treści

Strona

1. INFORMACJE OGÓLNE I CELE 3
2. PRZEGLĄD OBOWIĄZKÓW W ZAKRESIE PRZEJRZYSTOŚCI I ZAGADNIEŃ HORYZONTALNYCH 3
   - 2.1. Przegląd obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI 3
   - 2.2. Uzasadnienie i cele 4
   - 2.3. Podmioty odpowiedzialne na podstawie art. 50 aktu w sprawie AI 5
   - 2.4. Wyłączenia z zakresu stosowania aktu w sprawie AI 7
     - 2.4.1. Podmioty stosujące działające w kontekście czysto osobistej działalności pozazawodowej 7
     - 2.4.2. Badania i rozwój 8
     - 2.4.3. Systemy AI udostępniane na podstawie bezpłatnych licencji otwartego oprogramowania 8
   - 2.5. Wzajemne powiązania z zakazanymi praktykami, wymogami dotyczącymi systemów AI wysokiego ryzyka i kompetencjami w zakresie AI 9
   - 2.6. Wzajemne powiązania z przepisami mającymi zastosowanie do modeli lub systemów AI ogólnego przeznaczenia 9
3. ART. 50 UST. 1 AKTU W SPRAWIE AI: PRZEJRZYSTOŚĆ W ODNIESIENIU DO INTERAKTYWNYCH SYSTEMÓW AI 9
   - 3.1. Główne elementy, pojęcia i związane z nimi obowiązki w zakresie przejrzystości 10
     - 3.1.1. Systemy AI przeznaczone do wchodzenia w bezpośrednią interakcję z osobami fizycznymi 10
     - 3.1.2. Obowiązek w zakresie informowania wynikający z art. 50 ust. 1 aktu w sprawie AI 12
   - 3.2. Wyjątki od obowiązku w zakresie informowania na podstawie art. 50 ust. 1 aktu w sprawie AI 14
     - 3.2.1. Wyjątek dotyczący oczywistej interakcji z systemem AI 14
     - 3.2.2. Wyjątek dotyczący systemów AI, których wykorzystywanie jest dozwolone na mocy prawa do celów ścigania przestępstw 16
   - 3.3. Wzajemne powiązania z innymi aktami prawnymi Unii 17
4. ART. 50 UST. 2 AKTU W SPRAWIE AI: OZNACZANIE I WYKRYWANIE TREŚCI WYGENEROWANYCH LUB ZMANIPULOWANYCH PRZEZ AI 18
   - 4.1. Główne elementy i pojęcia, o których mowa w art. 50 ust. 2 aktu w sprawie AI 18
     - 4.1.1. Systemy AI generujące treści syntetyczne lub manipulujące nimi 18
     - 4.1.2. Rodzaje treści syntetycznych objęte zakresem stosowania 19
     - 4.1.3. Treści nieobjęte zakresem stosowania art. 50 ust. 2 aktu w sprawie AI 20
   - 4.2. Obowiązek oznaczania i zapewnienia wykrywalności określony w art. 50 ust. 2 aktu w sprawie AI 21
     - 4.2.1. Obowiązek oznaczania 21
     - 4.2.2. Obowiązek zapewnienia wykrywalności 22
     - 4.2.3. Zgodność z wymogami dotyczącymi rozwiązań technicznych: kryteria skuteczności, interoperacyjności, solidności i niezawodności 22
   - 4.3. Wyjątki od obowiązków wynikających z art. 50 ust. 2 aktu w sprawie AI 24
   - 4.4. Wzajemne powiązania z innymi aktami prawnymi Unii 25

5. ART. 50 UST. 3 AKTU W SPRAWIE AI: SYSTEMY ROZPOZNAWANIA EMOCJI I SYSTEMY KATEGORYZACJI BIOMETRYCZNEJ 26

5.1. Główne elementy, pojęcia i związane z nimi obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 3 aktu w sprawie AI 27

5.1.1. Pojęcie systemu rozpoznawania emocji 27

5.1.2. Pojęcie systemu kategoryzacji biometrycznej 27

5.1.3. Obowiązek informacyjny wynikający z art. 50 ust. 3 aktu w sprawie AI 27

5.2. Wyjątek dotyczący zastosowań do celów ścigania przestępstw 28

5.3. Wzajemne powiązania z innymi aktami prawnymi Unii 28

6. ART. 50 UST. 4: OZNAKOWANIE MATERIAŁÓW TYPU DEEPFAKE I NIEKTÓRYCH PUBLIKACJI TEKSTOWYCH 28

6.1. Główne elementy i pojęcia dotyczące materiałów typu deepfake i związane z nimi obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 4 aktu w sprawie AI 28

6.1.1. Pojęcie „deepfake” 29

6.1.2. Obowiązek ujawniania informacji na podstawie art. 50 ust. 4 akapit pierwszy aktu w sprawie AI 31

6.1.3. Przejrzystość w odniesieniu do treści typu deepfake o charakterze artystycznym, twórczym, satyrycznym, fikcyjnym lub analogicznym 32

6.1.4. Wyjątek na potrzeby ścigania przestępstw 34

6.1.5. Wzajemne powiązania z innymi aktami prawnymi Unii 34

6.2. Główne elementy i pojęcia dotyczące tekstu wygenerowanego lub zmanipulowanego przez AI oraz związane z tym obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 4 aktu w sprawie AI 35

6.2.1. Tekst publikowany w celu informowania opinii publicznej o sprawach leżących w interesie publicznym 36

6.2.2. Obowiązek ujawniania informacji na podstawie art. 50 ust. 4 akapit drugi aktu w sprawie AI ... 37

6.2.3. Wyjątek od obowiązku w zakresie przejrzystości w odniesieniu do tekstu podlegającego weryfikacji przez człowieka lub kontroli redakcyjnej i odpowiedzialności redakcyjnej 37

6.2.4. Wyjątek na potrzeby ścigania przestępstw 38

6.2.5. Wzajemne powiązania z innymi aktami prawnymi Unii 39

7. WYMOGI HORYZONTALNE MAJĄCE ZASTOSOWANIE DO INFORMACJI PRZEKAZYWANYCH NA PODSTAWIE ART. 50 UST. 5 AKTU W SPRAWIE AI 39

7.1. Przekazywanie informacji w jasny i wyraźny sposób 39

7.2. Pierwsza interakcja lub pierwsze stosowanie 40

7.3. Zgodność z mającymi zastosowanie wymogami dostępności 40

8. EGZEKWOWANIE ART. 50 AKTU W SPRAWIE AI 40

8.1. Skutki przestrzegania kodeksu praktyk uznanego za odpowiedni 41

8.2. Organy nadzoru rynku 42

8.3. Kary 42

8.4. Rozpoczęcie stosowania 42

9. PRZEGLĄD I AKTUALIZACJA WYTYCZNYCH 43

1. **INFORMACJE OGÓLNE I CELE**

(1) Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany szeregu rozporządzeń i dyrektyw („akt w sprawie AI") weszło w życie 1 sierpnia 2024 r. ([1]). W akcie w sprawie AI ustanowiono zharmonizowane przepisy dotyczące wprowadzania do obrotu, oddawania do użytku i wykorzystywania sztucznej inteligencji („AI") w Unii ([2]). Jego celem jest promowanie innowacji w dziedzinie AI i szerszego upowszechnienia godnej zaufania sztucznej inteligencji, przy jednoczesnym zapewnieniu wysokiego poziomu ochrony zdrowia, bezpieczeństwa i praw podstawowych w Unii, w tym demokracji i praworządności.

(2) W akcie w sprawie AI zastosowano podejście oparte na analizie ryzyka i podzielono systemy AI według czterech różnych kategorii ryzyka, z których jedna to systemy AI stwarzające ryzyko dla przejrzystości, które podlegają obowiązkom określonym w art. 50 aktu w sprawie AI. Te obowiązki w zakresie przejrzystości mają zastosowanie dwa lata po wejściu w życie aktu w sprawie AI, tj. od 2 sierpnia 2026 r.

(3) Zgodnie z art. 96 ust. 1 lit. d) aktu w sprawie AI niniejsze wytyczne Komisji mają pełnić rolę praktycznych wskazówek pomagających właściwym organom, a także dostawcom systemów AI i podmiotom stosującym systemy AI w spełnianiu obowiązków w zakresie przejrzystości wynikających z art. 50 aktu w sprawie AI w sposób spójny, skuteczny, proporcjonalny i jednolity.

(4) Przy opracowywaniu niniejszych wytycznych uwzględniono uwagi różnych zainteresowanych stron zebrane podczas szeroko zakrojonych konsultacji zorganizowanych przez Komisję oraz uwagi państw członkowskich zgłaszane na forum Rady ds. Sztucznej Inteligencji, a także uwagi zainteresowanych stron dotyczące projektu wytycznych, który opublikowano na potrzeby konsultacji.

(5) Niniejsze wytyczne nie mają charakteru wiążącego. Wszelkiej autorytatywnej wykładni aktu w sprawie AI może ostatecznie dokonywać wyłącznie Trybunał Sprawiedliwości Unii Europejskiej („TSUE").

2. **PRZEGLĄD OBOWIĄZKÓW W ZAKRESIE PRZEJRZYSTOŚCI I ZAGADNIEŃ HORYZONTALNYCH**

2.1. **Przegląd obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI**

(6) W art. 50 aktu w sprawie AI określono cztery obowiązki w zakresie przejrzystości, z których każdy ma zastosowanie do różnych rodzajów systemów AI lub ich wyników.

Przepis	Rodzaj systemu AI/ wyniku generowanego przez system AI	Obowiązek w zakresie przejrzystości	Wyjątki lub specjalne systemy
Art. 50 ust. 1	Systemy AI wchodzące w bezpośrednią interakcję z osobami fizycznymi	Dostawcy muszą opracować i zaprojektować system AI w taki sposób, aby zainteresowane osoby fizyczne były informowane o tym, że wchodzą w interakcję z systemem AI.	Wyjątki: jeżeli 1) widać wyraźnie, że interakcja została wygenerowana przez AI lub 2) wykorzystywanie systemu jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, chyba że system udostępnia się ogółowi społeczeństwa na potrzeby składania zawiadomień o popełnieniu przestępstwa.

([1]) Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji (akt w sprawie AI) (Dz.U. L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj).

([2]) Art. 1 aktu w sprawie AI.

Przepis	Rodzaj systemu AI/ wyniku generowanego przez system AI	Obowiązek w zakresie przejrzystości	Wyjątki lub specjalne systemy
Art. 50 ust. 2	Systemy AI generujące treści w postaci syntetycznych dźwięków, obrazów, wideo lub tekstu lub manipulujące tymi treściami	Dostawcy muszą zapewnić, aby wyniki systemu AI zostały oznakowane w formacie nadającym się do odczytu maszynowego i były wykrywalne – za pomocą rozwiązań technicznych, które są skuteczne, interoperacyjne, solidne i niezawodne – jako sztucznie wygenerowane lub zmanipulowane.	Wyjątki: jeżeli 1) system AI pełni funkcję wspomagającą w zakresie standardowej edycji lub nie zmienia w istotny sposób danych wejściowych lub ich semantyki lub 2) wykorzystywanie systemu AI jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców.
Art. 50 ust. 3	Systemy AI do rozpoznawania emocji lub kategoryzacji biometrycznej	Podmioty stosujące systemy AI muszą informować osoby fizyczne, wobec których systemy te są stosowane, o fakcie ich stosowania.	Wyjątek: jeżeli wykorzystywanie systemu AI jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców.
Art. 50 ust. 4	Systemy AI generujące treści typu deepfake lub teksty publikowane w celu informowania opinii publicznej o sprawach leżących w interesie publicznym lub manipulujące takimi treściami lub tekstami	Podmioty stosujące systemy AI muszą ujawnić, że treści zostały sztucznie wygenerowane lub zmanipulowane.	Wyjątki: jeżeli 1) wykorzystywanie systemu AI jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców lub 2) publikacja tekstu została poddana weryfikacji przez człowieka lub kontroli redakcyjnej i podlega odpowiedzialności redakcyjnej. Specjalny system ujawniania informacji ma zastosowanie do treści typu deepfake stanowiących część utworu lub programu o artystycznym, twórczym, fikcyjnym, satyrycznym lub analogicznym charakterze.

(7) Każdy z tych czterech obowiązków przeanalizowano w sekcjach 3–6 niniejszych wytycznych. Ponadto art. 50 ust. 5 aktu w sprawie AI stanowi, że informacje, o których mowa w ust. 1–4, są przekazywane zainteresowanym osobom fizycznym w jasny i wyraźny sposób, najpóźniej w momencie pierwszej interakcji lub pierwszego stosowania, oraz muszą spełniać mające zastosowanie wymogi dostępności. W sekcji 7 doprecyzowano te wymogi.

(8) Poszczególne obowiązki w zakresie przejrzystości określone w art. 50 aktu w sprawie AI mogą mieć zastosowanie łącznie do (wyników) pojedynczego systemu AI, co może wiązać się z odpowiedzialnością różnych podmiotów (dostawców lub podmiotów stosujących). Na przykład jeżeli system AI generuje obrazy w ramach bezpośredniej interakcji z osobą fizyczną, dostawca systemu może być zobowiązany do przestrzegania zarówno art. 50 ust. 1, jak i art. 50 ust. 2 aktu w sprawie AI. W przypadku gdy system ten jest również w stanie generować obrazy stanowiące treści typu deepfake lub publikować teksty dotyczące spraw leżących w interesie publicznym, podmiot stosujący może być dodatkowo zobowiązany do przestrzegania art. 50 ust. 4 aktu w sprawie AI.

2.2. **Uzasadnienie i cele**

(9) Jak wyjaśniono w motywach 132–136 aktu w sprawie AI, celem obowiązków w zakresie przejrzystości jest ograniczenie ryzyka podawania się za inną osobę, wprowadzania w błąd, podawania informacji wprowadzających w błąd, manipulacji na dużą skalę i oszustw oraz ograniczenie potencjalnego negatywnego wpływu na procesy demokratyczne i zaufanie społeczne spowodowanego treściami lub interakcjami generowanymi przez AI lub zmanipulowanymi przez AI. Informowanie osób fizycznych o tym, że prowadzą interakcję z systemem AI oraz korzystają z treści wygenerowanych przez AI, pomoże im w podejmowaniu świadomych decyzji, a jednocześnie przyczyni się do zagwarantowania zaufania i ochrony integralności ekosystemu informacyjnego. Przejrzystość jest nie tylko technicznym wymogiem zgodności, ale również warunkiem wstępnym umożliwiającym osobom fizycznym rozpoznawanie interakcji i treści opartych na AI lub stosowania systemów rozpoznawania emocji lub kategoryzacji biometrycznej. Zapewnienie wykorzystania systemów interaktywnej i generatywnej AI w sposób przejrzysty i godny zaufania umożliwi również utrzymanie autonomii osobistej i wartości europejskich, godności ludzkiej i innych praw podstawowych (np. wolności myśli, prywatności, równości i niedyskryminacji, ochrony konsumentów, własności intelektualnej itp.), przy jednoczesnym zachowaniu wolności wypowiedzi i wiarygodności treści medialnych.

2.3. **Podmioty odpowiedzialne na podstawie art. 50 aktu w sprawie AI**

(10) Zgodnie z art. 3 pkt 3 aktu w sprawie AI **dostawcy** to osoby fizyczne lub prawne, organy publiczne, agencje lub inne podmioty, które rozwijają systemy AI lub zlecają rozwój systemów AI oraz które – odpłatnie lub nieodpłatnie – pod własną nazwą lub własnym znakiem towarowym wprowadzają je do obrotu na unijnym rynku lub oddają je do użytku. Dostawcy muszą przestrzegać odpowiednich obowiązków w zakresie przejrzystości określonych w art. 50 ust. 1 i 2 aktu w sprawie AI, niezależnie od tego, czy mają siedzibę lub znajdują się w Unii czy w państwie trzecim ([3]). Dostawcy mający siedzibę lub znajdujący się poza Unią również podlegają aktowi w sprawie AI, jeżeli wyniki ich systemu AI są wykorzystywane w Unii ([4]). Przypadkowe, nieprzewidywalne lub nieuprawnione wykorzystywanie przez użytkowników na dalszym etapie łańcucha nie powinno jednak samo w sobie skutkować nałożeniem obowiązków na takich dostawców systemów interaktywnej lub generatywnej AI z państw trzecich, które to systemy nie są wprowadzane do obrotu ani oddawane do użytku w Unii, bez uszczerbku dla uprawnień właściwych organów nadzoru rynku do uznania ich wyników za niezgodne z aktem w sprawie AI.

(11) Dostawcy muszą zapewnić, aby ich systemy interaktywnej lub generatywnej AI cechowały się przejrzystością uwzględnioną już na etapie projektowania oraz działania systemu, oraz muszą zapewnić spełnienie wszystkich odpowiednich obowiązków w zakresie przejrzystości określonych w art. 50 ust. 1, 2 i 5 aktu w sprawie AI najpóźniej w momencie wprowadzenia tych systemów do obrotu lub oddania ich do użytku. Wszystkie osoby fizyczne wchodzące w interakcję z systemem AI lub mające styczność z jego wynikami w objętym zakresem obszarze muszą być skutecznie i konsekwentnie informowane przez cały cykl życia systemu AI wprowadzonego do obrotu lub oddanego do użytku przez tego dostawcę (niezależnie od tego, kim są podmioty stosujące system lub inni użytkownicy systemu).

> Na przykład: przedsiębiorstwo udostępnia aplikację generatywnej lub interaktywnej AI (np. chatbota, generatora obrazów, agenta AI) na rynku unijnym pod własną nazwą lub pod własnym znakiem towarowym konsumentom, użytkownikom profesjonalnym lub innym podmiotom prawnym, którzy mogą bezpośrednio wykorzystywać system AI do interakcji lub generowania treści (w tym poprzez włączenie systemu do swoich istniejących procesów i odpowiednie korzystanie z niego bez konieczności modyfikacji lub wprowadzenia dodatkowych komponentów). Przedsiębiorstwo to jest dostawcą odpowiedzialnym za spełnienie obowiązków w zakresie przejrzystości określonych w art. 50 ust. 1 lub 2 aktu w sprawie AI, niezależnie od tego, czy system AI jest udostępniany nieodpłatnie czy odpłatnie, oraz niezależnie od tego, czy dostawca ma siedzibę lub znajduje się w Unii czy w państwie trzecim.
>
> Innym przykładem dostawcy jest przedsiębiorstwo lub inna organizacja (np. organ publiczny), które opracowały interaktywny system AI (np. chatbota) we własnym zakresie i oddają go do użytku w Unii na własny użytek oraz pod własną nazwą lub pod własnym znakiem towarowym.
>
> Jeżeli przedsiębiorstwo wykorzystuje już istniejący system generatywnej AI wprowadzony do obrotu przez innego dostawcę i modyfikuje ten system (np. za pomocą nowych danych treningowych), który następnie oddaje do użytku pod własną nazwą lub pod własnym znakiem towarowym, przedsiębiorstwo to staje się dostawcą nowego systemu bez uszczerbku dla odpowiedzialności dostawcy pierwotnego systemu AI za tenże system.

(12) Podmioty stosujące to osoby fizyczne lub prawne, organy publiczne, agencje lub inne podmioty wykorzystujące systemy AI, nad którymi sprawują kontrolę, chyba że systemy AI są wykorzystywane w ramach osobistej działalności pozazawodowej ([5]). „Kontrolę" nad systemem AI należy rozumieć jako przyjęcie odpowiedzialności za decyzję o wdrożeniu tego systemu oraz za sposób jego faktycznego wykorzystania (w tym jego wyników). Niekoniecznie wymaga to kontroli technicznej nad funkcjonowaniem systemu AI, o ile podmiot stosujący podejmuje decyzję, do jakich celów i w jaki sposób będzie korzystać z systemu AI (w tym w zdecentralizowanych procesach pracy i strukturach korporacyjnych grupy). Podmioty stosujące uczestniczące w złożonych łańcuchach wartości produkcji i dystrybucji treści powinny wprowadzić proporcjonalne środki w celu zapewnienia, aby oznakowanie treści, które te podmioty zastosowały zgodnie z art. 50 ust. 4 aktu w sprawie AI, było wyświetlane w sposób jasny i rozpoznawalny dla docelowych i przewidywalnych odbiorców w momencie pierwszego stosowania zgodnie z art. 50 ust. 5 aktu w sprawie AI (np. za pośrednictwem warunków umownych z partnerami dystrybuującymi treści, ustawień doświadczenia użytkownika (UX) i interfejsów, które mają być wyświetlane).

([3]) Art. 2 ust. 1 lit. a), art. 3 ust. 3, 9 i 11 aktu w sprawie AI. Aby uzyskać dalsze wytyczne dotyczące pojęć „wprowadzanie do obrotu", „oddawanie do użytku" i „wykorzystywanie", zob. wytyczne Komisji w sprawie zakazanych praktyk w zakresie sztucznej inteligencji, C(2025) 5052 oraz zawiadomienie Komisji „Niebieski przewodnik – wdrażanie unijnych przepisów dotyczących produktów 2022", sekcja 2 (Dz.U. C 247, 29.6.2022, s. 1).

([4]) Art. 2 ust. 1 lit. c) oraz motyw 22 aktu w sprawie AI.

([5]) Art. 3 ust. 4 aktu w sprawie AI.

(13) Podmioty stosujące podlegają obowiązkom w zakresie przejrzystości określonym w art. 50 ust. 3 i 4 aktu w sprawie AI, jeżeli mają siedzibę lub znajdują się w Unii lub jeżeli mają siedzibę lub znajdują się w państwie trzecim, natomiast wyniki systemu AI są wykorzystywane w Unii ([6]). W takim przypadku obowiązki w zakresie przejrzystości mają zastosowanie do podmiotów mających siedzibę lub znajdujących się poza Unią, w przypadku gdy sam podmiot stosujący przewiduje rozpowszechnianie i wykorzystywanie wyników AI w Unii (tj. poprzez zlecanie dystrybucji lub zezwalanie na dystrybucję na terytorium Unii, w tym poprzez publikowanie treści typu deepfake w ogólnodostępnym internecie). Podmioty stosujące z państw trzecich nie są jednak związane obowiązkami w zakresie przejrzystości, jeżeli treści systemu AI docierają do odbiorców w Unii za pośrednictwem kanałów, których nie można przewidzieć i które są poza ich kontrolą.

(14) W przypadku gdy podmiot stosujący system AI jest osobą prawną, pod której kontrolą system jest wykorzystywany (np. agencja reklamowa), poszczególnych pracowników, którzy działają zgodnie z instrukcjami i pod kontrolą tej osoby prawnej (np. animatorów cyfrowych, projektantów stron internetowych, twórców treści, dziennikarzy), nie należy uznawać za odrębne podmioty stosujące ten system. Osoba prawna pozostaje podmiotem stosującym, jeżeli angażuje osoby trzecie (np. wykonawców, freelancerów) w obsługę systemu w jej imieniu, na jej odpowiedzialność i pod jej kontrolą.

Na przykład: podmiot medialny mający siedzibę w Unii, którego redaktorzy wykorzystują system AI do wspomagania relacjonowania bieżących wydarzeń za pomocą generowanych przez AI publikacji tekstowych dotyczących spraw leżących w interesie publicznym, zamieszczanych na ogólnodostępnej stronie internetowej, jest podmiotem stosującym objętym zakresem aktu w sprawie AI. Podobnie przedsiębiorstwo mające siedzibę w państwie trzecim, które wykorzystuje system AI do generowania treści typu deepfake z udziałem osoby powszechnie znanej w reklamie wyświetlanej w Unii, również jest podmiotem stosującym objętym zakresem aktu w sprawie AI.

Natomiast przedsiębiorstwo, które jedynie zleca agencji reklamowej stworzenie reklamy oraz nie podejmuje decyzji ani nie sprawuje kontroli nad tym, czy i w jaki sposób agencja reklamowa wykorzystuje AI w procesie tworzenia reklamy, nie jest podmiotem stosującym.

(15) Podmioty (np. dostawcy i podmioty stosujące) mogą pełnić jednocześnie więcej niż jedną rolę w odniesieniu do systemu AI. Na przykład przedsiębiorstwo może opracować wewnętrzny system generatywnej AI i wykorzystywać go do generowania treści typu deepfake. W takim przypadku przedsiębiorstwo kwalifikuje się zarówno jako dostawca, jak i podmiot stosujący, który musi wypełnić obowiązki dotyczące oznaczania nadającego się do odczytu maszynowego na podstawie art. 50 ust. 2 oraz spełnić wymogi dotyczące jasnego i dostrzegalnego oznaczania treści typu deepfake na podstawie art. 50 ust. 4 aktu w sprawie AI.

(16) Natomiast inne podmioty (na przykład dostawcy usług hostingu, w tym platformy internetowe, lub nadawcy), których rola ogranicza się do rozpowszechniania lub przekazywania wygenerowanych lub zmanipulowanych przez AI treści utworzonych przez osoby trzecie lub które otrzymują treści wygenerowane lub zmanipulowane przez AI lub mają styczność z takimi treściami i nie sprawują kontroli nad wykorzystaniem systemu AI do celów generowania treści lub manipulowania treściami przez AI, nie są podmiotami stosującymi w rozumieniu aktu w sprawie AI. Nie ponosząc odpowiedzialności na mocy aktu w sprawie AI, podmioty te mogą jednak odgrywać w łańcuchu wartości ważną rolę w utrzymaniu zaufania do ekosystemu informacyjnego i jego integralność. Zdecydowanie zachęca się je zatem do zachowania oznakowania i etykietowania wdrożonego zgodnie z art. 50 aktu w sprawie AI oraz do wdrożenia proporcjonalnych i technicznie wykonalnych środków umożliwiających osobom fizycznym wykrywanie treści wygenerowanych lub zmanipulowanych przez AI, bez uszczerbku dla odpowiedzialności dostawców generatywnych systemów AI i podmiotów stosujących systemy generatywnej AI ani dla jakichkolwiek obowiązków lub odpowiedzialności wynikających z rozporządzenia (UE) 2022/2065 (akt o usługach cyfrowych) ([7]) lub innego mającego zastosowanie prawa Unii lub prawa krajowego.

([6]) Art. 2 ust. 1 lit. b) i c) aktu w sprawie AI.

([7]) Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych (Dz.U. L 277 z 27.10.2022, s. 1).

(17) Inne podmioty działające w ramach swoich obowiązków zawodowych i rozpowszechniające treści (nawet jeśli nie są bezpośrednio zaangażowane na podstawie art. 50 aktu w sprawie AI) również zachęca się do wdrożenia proporcjonalnych i odpowiednich środków, tak aby osoby fizyczne będące ostatecznymi odbiorcami tych treści były skutecznie informowane i były w stanie wykryć fakt wygenerowania lub zmanipulowania tych treści przez AI. Ponadto strony te mogą być zobowiązane do oznaczania odpowiednich materiałów typu deepfake lub publikacji tekstowych dotyczących spraw leżących w interesie publicznym zgodnie z innymi obowiązującymi przepisami prawnymi lub deontologicznymi (np. oznakowanie nadawcy i pokazywanie w ramach programu telewizyjnego materiałów typu deepfake stworzonych przez osobę prywatną). Sam fakt, że podmiot oznacza treści lub wykorzystuje oznakowane treści, nie oznacza jednak, że kwalifikuje się on jako „podmiot stosujący" w rozumieniu aktu w sprawie AI. Akt w sprawie AI nie rozszerza odpowiedzialności prawnej tych podmiotów na mocy mających zastosowanie ram prawnych ani nie wpływa na nią, jeżeli podmioty te nie kwalifikują się jednocześnie jako dostawcy systemów AI lub podmioty stosujące systemy AI.

2.4. **Wyłączenia z zakresu stosowania aktu w sprawie AI**

(18) W art. 2 aktu w sprawie AI przewidziano szereg ogólnych wyłączeń z zakresu stosowania aktu w sprawie AI, które zostały już wyjaśnione w wytycznych Komisji w sprawie zakazanych praktyk w zakresie sztucznej inteligencji ([8]). W związku z tym w niniejszej sekcji omówiono jedynie te aspekty, które są istotne z punktu widzenia obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI.

2.4.1. *Podmioty stosujące działające w kontekście czysto osobistej działalności pozazawodowej*

(19) Zgodnie z art. 2 ust. 10 aktu w sprawie AI przedmiotowego rozporządzenia nie stosuje się do obowiązków podmiotów stosujących będących osobami fizycznymi, które korzystają z systemów AI w ramach czysto osobistej działalności pozazawodowej ([9]). Za działalność „zawodową" należy uznać wszelką działalność, w ramach której osoba fizyczna regularnie uzyskuje korzyść ekonomiczną lub jest w inny sposób zaangażowana w działalność specjalistyczną, gospodarczą, handlową, zawodową lub niezależną. Określenie „czysto osobista" doprecyzowuje pozazawodowy charakter działalności, co oznacza, że działania danej osoby powinny mieć charakter zarówno osobisty, jak i pozazawodowy, aby osoba ta mogła zostać objęta zakresem tego wyjątku. Wszelkie wykorzystanie przez osoby fizyczne działające w imieniu lub z upoważnienia podmiotu stosującego działającego w ramach swoich obowiązków zawodowych będzie wchodzić w zakres obowiązków w zakresie przejrzystości określonych w art. 50 ust. 3 i 4 aktu w sprawie AI ciążących na podmiocie stosującym prowadzącym działalność zawodową.

Przykłady osobistego, pozazawodowego wykorzystania systemów AI wyłączonego z zakresu stosowania rozporządzenia:

— osoby fizyczne wykorzystujące system AI do generowania treści typu deepfake dotyczących członków ich gospodarstwa domowego,

— osoby fizyczne wykorzystujące system AI we własnym imieniu do generowania tekstu na różne tematy (np. uczniowie wykorzystujący AI do pisania prac domowych),

— osoby fizyczne wykorzystujące system AI we własnym imieniu do generowania treści typu deepfake rozpowszechnianych w mediach społecznościowych.

(20) Wyłączenie przewidziane w art. 2 ust. 10 aktu w sprawie AI ma wyłącznie zastosowanie do obowiązków spoczywających na podmiotach stosujących, w przypadku gdy podmioty te wykorzystują system do czysto osobistej działalności pozazawodowej. System jako taki pozostaje objęty zakresem aktu w sprawie AI w odniesieniu do obowiązków dostawców wprowadzających system do obrotu lub oddających go do użytku oraz podmiotów stosujących system do użytku profesjonalnego. Wyjątek ten pozostaje również bez uszczerbku dla stosowania innych odpowiednich przepisów prawa Unii i prawa krajowego.

([8]) Zob. sekcja 2.5 wytycznych Komisji w sprawie zakazanych praktyk w zakresie sztucznej inteligencji ustanowionych rozporządzeniem (UE) 2024/1689 (akt w sprawie AI), Bruksela, C(2025) 5052 final z 29.07.2025.

([9]) Z definicji podmiotu stosującego wykluczeni są również użytkownicy zaangażowani w taką działalność (zob. sekcja 2.3 powyżej).

Na przykład: dostawca systemu AI wykorzystywanego do generowania treści typu deepfake w celu umieszczenia ich na osobistej kartce z życzeniami bożonarodzeniowymi powinien oznaczyć te treści typu deepfake w sposób nadający się do odczytu maszynowego zgodnie z obowiązkiem w zakresie przejrzystości określonym w art. 50 ust. 2 aktu w sprawie AI. Jednocześnie wykorzystywanie systemu przez osoby fizyczne na użytek w pełni osobisty w ramach działalności innej niż zawodowa nie będzie objęte obowiązkiem podmiotu stosującego określonym w art. 50 ust. 4 aktu w sprawie AI.

2.4.2. *Badania i rozwój*

(21) Art. 2 ust. 6 aktu w sprawie AI przewiduje wyłączenie dotyczące systemów AI lub modeli AI, w tym ich wyników, specjalnie opracowanych i oddanych do użytku wyłącznie do celów badań naukowych i rozwojowych. Wyłączenie to pozostaje bez uszczerbku dla mających zastosowanie norm etycznych w dziedzinie badań naukowych oraz unijnych przepisów dotyczących ochrony danych i innych przepisów. W kontekście art. 50 aktu w sprawie AI wyłączenie to obejmuje nie tylko systemy interaktywnej AI, ale również wyniki systemów generatywnej AI (w tym treści typu deepfake) wykorzystywanych w kontekście badań naukowych. Jeżeli jednak systemy te są oddawane do użytku lub ich wyniki są wykorzystywane do celów innych niż wyłącznie do celów badań naukowych i rozwoju, nadal należy przestrzegać odpowiednich obowiązków w zakresie przejrzystości, o których mowa w art. 50 aktu w sprawie AI.

Na przykład: jeżeli badacze opracowują i wprowadzają do obrotu lub oddają do użytku interaktywny system AI i chcą zbadać, czy osoby fizyczne są w stanie odróżnić systemy AI od ludzi podczas wchodzenia w interakcję głosową z danym systemem AI, nie muszą dopełniać odpowiednich obowiązków w zakresie przejrzystości wynikających z art. 50 ust. 1 i 2 aktu w sprawie AI. Jeżeli jednak ten sam interaktywny system AI jest również oddawany do użytku do celów innych niż badania i rozwój, należy przestrzegać odpowiednich obowiązków w zakresie przejrzystości, o których mowa w art. 50 aktu w sprawie AI.

(22) Ponadto art. 2 ust. 8 aktu w sprawie AI wyłącza z zakresu stosowania działalność badawczą, testową i rozwojową dotyczącą systemów AI lub modeli AI przed wprowadzeniem ich do obrotu lub oddaniem ich do użytku. Wyłączenie to nie obejmuje testów w warunkach rzeczywistych (w piaskownicach regulacyjnych w zakresie AI lub poza nimi). Ponadto po wprowadzeniu systemu AI do obrotu lub oddaniu go do użytku w następstwie takich badań i testów powinien on mimo wszystko spełniać odpowiednie obowiązki w zakresie przejrzystości wynikające z art. 50 aktu w sprawie AI ([10]).

2.4.3. *Systemy AI udostępniane na podstawie bezpłatnych licencji otwartego oprogramowania*

(23) Zgodnie z art. 2 ust. 12 aktu w sprawie AI systemy AI udostępniane na podstawie bezpłatnych licencji otwartego oprogramowania nie są objęte zakresem stosowania aktu w sprawie AI, pod warunkiem że nie są objęte zakazami określonymi w art. 5 aktu w sprawie AI ani obowiązkami w zakresie przejrzystości określonymi w art. 50 aktu w sprawie AI lub nie są sklasyfikowane jako systemy wysokiego ryzyka ([11]). Oznacza to, że dostawcy otwartych systemów AI i podmioty stosujące otwarte systemy AI objęte zakresem stosowania art. 50 aktu w sprawie AI nadal muszą zapewnić przestrzeganie odnośnych obowiązków w zakresie przejrzystości, które na nich spoczywają.

(24) Z kolei dostawcy elementów AI na bezpłatnej licencji otwartego oprogramowania, które obejmują oprogramowanie, dane, modele, narzędzia, usługi lub procesy na potrzeby personalizacji i integracji z systemem AI i które same w sobie (częściowo lub w całości) nie stanowią systemu AI, nie podlegają obowiązkom w zakresie przejrzystości określonym w art. 50 aktu w sprawie AI. Dostawców takich systemów zachęca się jednak do projektowania takich modeli, elementów i usług w sposób opisany w pkt 27 poniżej, o ile jest to proporcjonalne i wykonalne.

([10]) Motyw 25 aktu w sprawie AI.

([11]) W motywie 102 aktu w sprawie AI stwierdzono, że udostępnianie oprogramowania i danych na podstawie bezpłatnej licencji otwartego oprogramowania „umożliwia ich ogólne upowszechnianie i zezwala użytkownikom na swobodny dostęp do nich, ich wykorzystywanie, zmianę i ich redystrybucję lub ich zmienionych wersji".

2.5. Wzajemne powiązania z zakazanymi praktykami, wymogami dotyczącymi systemów AI wysokiego ryzyka i kompetencjami w zakresie AI

(25) W motywie 137 aktu w sprawie AI doprecyzowano, że spełnienie obowiązków w zakresie przejrzystości wynikających z art. 50 aktu w sprawie AI w odniesieniu do konkretnego systemu AI nie można interpretować jako wskazania, że wykorzystanie tego systemu AI lub jego wyników jest zgodne z prawem na podstawie aktu w sprawie AI. W szczególności systemy AI podlegające obowiązkom określonym w art. 50 aktu w sprawie AI mogą wchodzić w zakres stosowania art. 5 aktu w sprawie AI, a tym samym mogą być zakazane (np. stosowanie systemów rozpoznawania emocji w miejscu pracy lub instytucjach edukacyjnych) ([12]). Ponadto systemy AI objęte zakresem stosowania art. 50 aktu w sprawie AI można również sklasyfikować jako systemy wysokiego ryzyka ([13]) i mogą one podlegać odpowiednim wymogom i obowiązkom dotyczącym takich systemów. Zgodność z art. 50 aktu w sprawie AI nie zastępuje zatem obowiązków wynikających z innych przepisów aktu w sprawie AI, nie ogranicza tych obowiązków ani nie wpływa na nie ([14]). Jednocześnie systemy AI mogą również wchodzić w zakres stosowania art. 50 aktu w sprawie AI, nie będąc sklasyfikowane jako systemy wysokiego ryzyka zgodnie z art. 6 aktu w sprawie AI. Wymogi dotyczące kompetencji w zakresie AI określone w art. 4 aktu w sprawie AI mają również zastosowanie zarówno do dostawców systemów AI objętych zakresem stosowania art. 50 aktu w sprawie AI, jak i do podmiotów je stosujących.

2.6. Wzajemne powiązania z przepisami mającymi zastosowanie do modeli lub systemów AI ogólnego przeznaczenia

(26) Art. 50 aktu w sprawie AI ma zastosowanie do niektórych systemów AI, w tym systemów AI ogólnego przeznaczenia ([15]), takich jak chatboty wspierające bezpośrednią interakcję z osobami fizycznymi lub systemy AI zdolne do generowania treści syntetycznych (w tym treści typu deepfake) lub manipulowania nimi ([16]).

(27) Art. 50 aktu w sprawie AI nie stosuje się wprost do modeli AI ogólnego przeznaczenia. Jeżeli jednak system interaktywnej lub generatywnej AI, który wchodzi w zakres stosowania art. 50 ust. 1 lub 2 aktu w sprawie AI, opiera się na modelu AI ogólnego przeznaczenia dostarczanym przez tego samego dostawcę, środki w zakresie przejrzystości określone w niniejszych wytycznych można również wdrożyć na poziomie modelu (zob. również sekcja 4 poniżej). Ponadto, chociaż dostawcy innych modeli AI ogólnego przeznaczenia i innych modeli generatywnej AI nie są objęci zakresem stosowania art. 50 aktu w sprawie AI, również ich zachęca się do wdrożenia odpowiednich środków w zakresie przejrzystości na poziomie modelu w celu identyfikacji interaktywnych systemów AI oraz oznaczania treści wygenerowanych i zmanipulowanych przez AI. Takie podejście ma szczególnie tę zaletę, że ułatwia dostawcom systemów AI niższego szczebla wypełnianie obowiązków wynikających z art. 50 ust. 1 i 2 aktu w sprawie AI ([17]) oraz przyczynia się do utrzymania zaufania do ekosystemu informacyjnego i jego integralności. W przypadku modeli AI ogólnego przeznaczenia stwarzających ryzyko systemowe środki bezpieczeństwa na poziomie modelu służące identyfikacji interakcji z AI i oznaczaniu treści generowanych lub zmanipulowanych przez AI mogą, w stosownych przypadkach, stanowić część szerszego zestawu środków ograniczających ryzyko, które dostawcy mogą wprowadzić w ramach obowiązku ograniczania ryzyka systemowego zgodnie z art. 55 ust. 1 lit. b) aktu w sprawie AI.

3. ART. 50 UST. 1 AKTU W SPRAWIE AI: PRZEJRZYSTOŚĆ W ODNIESIENIU DO INTERAKTYWNYCH SYSTEMÓW AI

(28) Art. 50 ust. 1 aktu w sprawie AI jest skierowany do dostawców systemów AI wchodzących w bezpośrednią interakcję z osobami fizycznymi, którzy muszą projektować i rozwijać swoje systemy w taki sposób, aby osoby fizyczne, których to dotyczy, były informowane o tym, że prowadzą interakcję z systemem AI. Jak wyjaśniono w motywie 132 aktu w sprawie AI, celem tego obowiązku informacyjnego jest umożliwienie tym osobom fizycznym podejmowania świadomych decyzji dotyczących wyników systemu, uniknięcie nadmiernego polegania na takich systemach oraz umożliwienie tym osobom fizycznym odpowiedniego dostosowania ich zaufania do treści i interakcji.

([12]) Więcej informacji na temat art. 5 aktu w sprawie AI można znaleźć w wytycznych Komisji w sprawie zakazanych praktyk w zakresie sztucznej inteligencji określonych w rozporządzeniu (UE) 2024/1689 (akt w sprawie AI), C(2025) 5052 final.

([13]) Zob. art. 6 aktu w sprawie AI.

([14]) W szczególności w art. 26 ust. 11 aktu w sprawie AI na podmioty stosujące systemy AI wysokiego ryzyka, które to podmioty podejmują decyzje lub uczestniczą w podejmowaniu decyzji dotyczących osób fizycznych, nałożono obowiązek w zakresie przejrzystości, który dotyczy informowania osób fizycznych o tym, że jest w stosunku do nich wykorzystywany system AI wysokiego ryzyka. Te obowiązki w zakresie przejrzystości stanowią uzupełnienie obowiązków określonych w art. 50 aktu w sprawie AI, ponieważ dotyczą różnych informacji, które należy przekazać osobom fizycznym, ale można je połączyć w ramach jednego powiadomienia.

([15]) W art. 3 ust. 66 aktu w sprawie AI zdefiniowano system AI ogólnego przeznaczenia jako system AI oparty na modelu AI ogólnego przeznaczenia, który to system może służyć różnym celom, nadający się zarówno do bezpośredniego wykorzystania, jak i do integracji z innymi systemami AI.

([16]) Art. 50 ust. 2 aktu w sprawie AI zawiera wyraźne odniesienie do systemów AI ogólnego przeznaczenia.

([17]) Motyw 133 aktu w sprawie AI.

3.1. **Główne elementy, pojęcia i związane z nimi obowiązki w zakresie przejrzystości**

(29) Obowiązek w zakresie przejrzystości określony w art. 50 ust. 1 aktu w sprawie AI ma zastosowanie do dostawców systemów AI, w przypadku gdy system jest przeznaczony do wchodzenia w bezpośrednią interakcję z osobami fizycznymi, a interakcja ta nie wchodzi w zakres żadnego z następujących wyjątków: (1) to, że interakcja prowadzona jest z systemem AI, jest oczywiste z punktu widzenia osoby fizycznej, która jest dostatecznie poinformowana, uważna i ostrożna, lub (2) wykorzystywanie systemu jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, chyba że system udostępnia się ogółowi społeczeństwa na potrzeby składania zawiadomień o popełnieniu przestępstwa.

3.1.1. *Systemy AI przeznaczone do wchodzenia w bezpośrednią interakcję z osobami fizycznymi*

(30) Aby system wchodził w zakres stosowania art. 50 ust. 1 aktu w sprawie AI, muszą wystąpić cztery elementy: (i) system musi być systemem AI, (ii) przeznaczonym do wchodzenia w interakcję, (iii) w sposób bezpośredni, (iv) z osobami fizycznymi.

(i) **System AI**: system musi spełniać kryteria zawarte w definicji systemu AI ([18]), która wyklucza proste mechanizmy automatycznych odpowiedzi nieoparte na AI (np. tradycyjne wiadomości o nieobecności, oparte na regułach mechanizmy szybkich odpowiedzi na wiadomości).

(ii) **Przeznaczony do wchodzenia w interakcję:** system AI musi być przeznaczony do wchodzenia w interakcję z osobami fizycznymi. Interakcja wiąże się z możliwością dwukierunkowej wymiany informacji lub działań między osobami fizycznymi a systemem AI o autentycznie konwersacyjnym lub responsywnym charakterze. Interakcja może występować w różnych formach (np. dźwiękowych, wizualnych i fizycznych). Osoby fizyczne powinny mieć możliwość przekazywania danych (np. tekstu pisemnego lub innych form treści, głosu lub działań fizycznych) do systemu AI (np. za pośrednictwem technicznych środków komunikacji, w tym interfejsów internetowych, przez telefon lub za pomocą poczty elektronicznej, lub bezpośrednio), natomiast system AI powinien być w stanie reagować na wyniki kontekstowe (w tym wszelkiego rodzaju treści lub działania) i odwrotnie. Interakcja może obejmować każdy rodzaj wyniku lub formatu czytelnego i zrozumiałego dla ludzi. Interakcja nie musi być inicjowana przez człowieka; interakcja może również obejmować treści generowane przez AI lub działania inicjowane przez AI, które wywołują reakcję człowieka. Interakcja może być jednorazową wymianą (np. pojedyncze zapytanie, po którym następuje jedna odpowiedź) lub może być rozciągnięta w czasie (interakcja wielokrotna). Systemów, które jedynie biernie gromadzą dane i nie są w stanie prowadzić wymiany z osobami fizycznymi (np. automatyczna kontrola dostępu oparta na rozpoznawaniu twarzy) lub tylko jednorazowo gromadzą informacje zwrotne (np. filtry spamu), nie uznaje się za wchodzące w interakcję z osobami fizycznymi.

(iii) **Bezpośrednia interakcja:** interakcja między systemem AI a osobami fizycznymi musi być bezpośrednia. Zazwyczaj wiąże się to z interakcją w czasie rzeczywistym lub zbliżonym do rzeczywistego (w tym poprzez zdolność systemu do pisania i wysyłania wiadomości do osób fizycznych lub innej interakcji ze środowiskiem fizycznym lub wirtualnym dostrzegalnym dla osób fizycznych). Bezpośrednia interakcja wyklucza pośrednią lub moderowaną interakcję międzyludzką, w przypadku której osoba fizyczna ma styczność z wynikami AI bez bezpośredniej interakcji z systemem (na przykład jeżeli pracownicy działu obsługi klienta korzystają z narzędzi wsparcia opartych na AI, wspomagających komunikowanie się z osobami fizycznymi, lub jeżeli wyniki AI nie są udostępniane danej osobie przez sam system AI, lecz przez inną osobę rozpowszechniającą treści). Sama możliwość interwencji lub przeglądu wyników systemu AI przez człowieka nie powinna być jednak wykorzystywana do obchodzenia wymogu dotyczącego ujawniania informacji określonego w art. 50 ust. 1 aktu w sprawie AI. Przypadki produktów i usług łączących odpowiedzi wygenerowane przez AI z treściami opracowanymi przez człowieka wchodzą w zakres stosowania art. 50 ust. 1 aktu w sprawie AI i wymagają ujawnienia w ramach interakcji konkretnych wyników wygenerowanych przez AI, chyba że te wyniki AI zostały odpowiednio zweryfikowane i przesłane przez ludzi występujących w roli głównych podmiotów wchodzących w interakcję z osobami fizycznymi.

(iv) **Z osobami fizycznymi:** w art. 50 ust. 1 aktu w sprawie AI za kryterium przyjęto, że interakcja systemu AI musi przebiegać z osobami fizycznymi. Osoby te mogą być profesjonalnymi podmiotami stosującymi, innymi użytkownikami (również korzystającymi z systemu do działalności czysto osobistej) lub innymi osobami korzystającymi z systemu w ich imieniu. Systemy AI, które działają w zamkniętym środowisku fizycznym (takim jak część konfiguracji maszyn przemysłowych), wywołania typu maszyna-maszyna w zapleczu systemowym między systemami AI (w tym agentami AI), których wyniki nie mają z założenia docierać do osoby fizycznej, lub środowiska wirtualne bez bezpośredniego kontaktu z osobami fizycznymi są wyłączone z zakresu przedmiotowego obowiązku.

([18]) Zob. art. 3 pkt 1 aktu w sprawie AI. Dalsze wytyczne dotyczące pojęcia systemu AI znajdują się w wytycznych Komisji dotyczących definicji systemu sztucznej inteligencji, C(2025) 5053 final z 29.07.2025.

(31) Agenci AI są objęci art. 50 ust. 1 aktu w sprawie AI, jeżeli są w stanie wchodzić w interakcję z osobami wydającymi im instrukcje lub z innymi osobami fizycznymi w ramach wykonywania zadań (np. dokonywania rezerwacji, zarządzania korespondencją, negocjowania lub zawierania umów, dokonywania zakupów itp.). Agentów AI należy projektować i rozwijać w taki sposób, aby ujawniali oni zarówno swój sztuczny charakter, jak i osobę, w imieniu której działają, z uwzględnieniem potrzeby przejrzystości pochodzenia i delegowania uprawnień oraz odpowiedzialności za konsekwencje ich działań. Obejmuje to przypadki, w których agenci AI są w stanie wchodzić w interakcję w złożonych architekturach z udziałem wielu agentów AI z innymi agentami AI, którzy wchodzą w bezpośrednią interakcję z osobami fizycznymi w ramach wykonywania zadania. Dostawca agenta AI może jednak nie zawsze być w stanie zidentyfikować z wyprzedzeniem poszczególne przypadki bezpośredniej interakcji agenta AI z innymi osobami fizycznymi po tym, jak agent zostanie wdrożony w różnych kontekstach. W przypadku gdy dostawca nie jest w stanie wiarygodnie ustalić, czy agent AI będzie wchodził w bezpośrednią interakcję z osobą fizyczną przed wprowadzeniem agenta AI do obrotu lub oddaniem go do użytku, agent powinien zostać zaprojektowany na poziomie architektury i poinstruowany, aby ujawniał się jako taki w każdej sytuacji, w której istnieje uzasadnione prawdopodobieństwo, że agent może wchodzić w interakcję z osobą fizyczną (np. gdy osoba fizyczna reprezentuje podmiot prawny lub działa na jego rzecz). Ponadto agenci AI powinni również ujawniać się osobom wydającym im instrukcje na kluczowych etapach (np. w momencie udzielania zezwolenia, sprawozdawczości, walidacji itp., w tym gdy wdrażany agent AI otrzymuje lub przetwarza wyniki wygenerowane przez inne systemy AI, a nie bezpośrednio przez osobę fizyczną, lub też opiera się na takich wynikach) oraz przy każdej nowej interakcji.

Przykłady systemów AI przeznaczonych do wchodzenia w bezpośrednią interakcję z osobami fizycznymi, objętych zakresem stosowania art. 50 ust. 1 aktu w sprawie AI:

— asystenci głosowi wspomagani AI, chatboty/agenci konwersacyjni w różnych kontekstach (np. usługi publiczne, obsługa klienta, zarządzanie skargami, handel elektroniczny, finanse, opieka zdrowotna, edukacja itp.),

— infolinia AI służąca do zgłaszania incydentów, oszustw lub dokonywania innych zgłoszeń, roboty/koboty (humanoidalne), wirtualni towarzysze oparci na AI; roboty-zwierzęta do towarzystwa; awatary AI (np. w środowiskach rzeczywistości wirtualnej),

— boty AI w sieciach i mediach społecznościowych, agenci wspomagający kodowanie i inni agenci AI zdolni do bezpośredniej interakcji z osobami fizycznymi (w tym w złożonych architekturach z udziałem wielu agentów).

3.2.1. **Przykłady systemów AI, które nie wchodzą w bezpośrednią interakcję z osobami fizycznymi, nieobjętych zakresem stosowania art. 50 ust. 1 aktu w sprawie AI:**

— tradycyjne roboty przemysłowe wspomagane AI, które działają w zamkniętym środowisku przemysłowym i nie są przeznaczone do interakcji z ludźmi (np. roboty do montażu układów elektronicznych),

— systemy rekomendacji wspomagane AI, filtry spamu, narzędzia do automatycznego tłumaczenia lub transkrypcji,

— wyszukiwanie i pozyskiwanie informacji wspomagane AI bez generowania lub modyfikowania treści,

— techniki automatycznego uzupełniania tekstu i kodu z wykorzystaniem AI, systemy uwierzytelniania/rozpoznawania biometrycznego, systemy wspomagające podejmowanie decyzji działające w zapleczu systemowym, w przypadku których użytkownik widzi jedynie wynik AI bez możliwości bezpośredniej interakcji z systemem,

— oparte na AI systemy konserwacji predykcyjnej i optymalizacji w fabrykach i zastosowaniach przemysłowych, wizualizacja produktów wspomagana AI (np. wirtualna przymierzalnia, wirtualna aranżacja wnętrz, wizualizacja produktów), wspomagane AI narzędzia do planowania tras oparte na jednorazowym zapytaniu, transkrypcja rozmów z klientami przetwarzana na potrzeby analizy zachowań itp.

3.1.2. *Obowiązek w zakresie informowania wynikający z art. 50 ust. 1 aktu w sprawie AI*

(32) W art. 50 ust. 1 aktu w sprawie AI zobowiązano dostawców systemów AI wchodzących w bezpośrednią interakcję z osobami fizycznymi do zapewnienia, aby ich systemy AI (w tym agenci AI i systemy AI ogólnego przeznaczenia) były projektowane i rozwijane w taki sposób, aby zainteresowane osoby fizyczne były informowane o tym, że prowadzą interakcję z systemem AI.

(33) Chociaż mechanizm informowania powinien być wbudowany w projektowanie i rozwój systemu oraz w sposób interakcji systemu z osobami fizycznymi, stosowna informacja powinna być przekazywana podczas działania systemu AI, a najpóźniej w momencie pierwszej interakcji z osobą fizyczną (zgodnie z wymogami określonymi w art. 50 ust. 5 aktu w sprawie AI, zob. sekcja 7 poniżej). Dostawcy są odpowiedzialni za zapewnienie, aby po wprowadzeniu systemu AI do obrotu lub oddaniu go do użytku osoby fizyczne wchodzące w interakcję z systemem AI były skutecznie i spójnie informowane przez cały cykl życia systemu AI wprowadzonego do obrotu lub oddanego do użytku przez tego dostawcę.2.2

(34) **Format ujawniania:** w art. 50 ust. 1 aktu w sprawie AI nie określono żadnej konkretnej techniki ujawniania. Dostawcy mogą zatem wybrać dowolną odpowiednią technikę, o ile spełnione są dwa wymogi: (i) wymogi określone w art. 50 ust. 5 aktu w sprawie AI zostały spełnione w odniesieniu do przekazywania informacji w jasny i wyraźny sposób, najpóźniej w momencie pierwszej interakcji (zob. sekcja 7 poniżej); oraz (ii) uwzględnia się cechy osób fizycznych należących do grup szczególnie wrażliwych ze względu na wiek (np. dzieci lub osoby starsze) lub niepełnosprawność – w zakresie, w jakim system AI ma wchodzić w interakcję, lub gdy istnieje uzasadnione prawdopodobieństwo interakcji z osobami fizycznymi należącymi do tych grup (np. również wówczas, gdy istnieje uzasadnione prawdopodobieństwo interakcji systemu AI z ogółem społeczeństwa) ([19]). Informacje muszą być również przekazywane w formacie dostępnym dla osób fizycznych z niepełnosprawnościami, jak określono w art. 50 ust. 5 i motywie 132 aktu w sprawie AI (zob. sekcja 7). W przypadku gdy system może wchodzić w interakcję z dziećmi, powiadomienia te muszą być przyjazne dziecku, dostosowane do wieku, łatwe do zrozumienia i łatwo dostępne dla wszystkich dzieci, w tym dzieci z niepełnosprawnościami lub dodatkowymi potrzebami w zakresie dostępności. Zaleca się, aby w miarę dostępności i w stosownych przypadkach dostawcy stosowali znormalizowane praktyki w zakresie ujawniania i oznaczania w celu zapewnienia skuteczności i spójności powiadomień oraz zwiększenia świadomości użytkowników.

(35) **Treść powiadomienia:** w art. 50 ust. 1 aktu w sprawie AI zobowiązano dostawców interaktywnych systemów AI do projektowania i rozwijania tych systemów w taki sposób, aby wszystkie osoby fizyczne wchodzące w interakcję z takimi systemami AI były wyraźnie informowane, że druga strona interakcji ma sztuczny charakter i nie jest człowiekiem. Jeżeli dostawca może racjonalnie przewidzieć, że system AI może wchodzić w interakcję z osobami fizycznymi należącymi do wyżej wymienionych grup szczególnie wrażliwych, informacje te należy odpowiednio dostosować i przekazać w przystępnym formacie, zgodnie z wymogami mającego zastosowanie prawa Unii, aby zapewnić również pełne zrozumienie tych informacji przez te osoby fizyczne ([20]).

(36) Powiadomienia powinny zawierać informacje na temat sztucznego źródła interakcji w sposób odpowiedni do kontekstu, w którym ma miejsce interakcja. Należy tego dokonać w sposób zapewniający odpowiednie i skuteczne informowanie zainteresowanych osób, unikając wszelkiego ryzyka wprowadzenia w błąd lub manipulacji przed interakcją i w jej trakcie, przy jednoczesnym zapewnieniu wykonalności i proporcjonalności obowiązku informacyjnego. Przykłady przekazywania informacji obejmują powiadomienia w formie tekstowej (np. chatbot, który rozpoczynając rozmowę, zaznacza, że wykorzystuje technologię AI), środki wizualne (np. wygenerowana przez agenta AI wiadomość e-mail wysłana do osoby fizycznej, z oznaczeniem AI na górze), środki dźwiękowe (np. asystent głosowy, który na początku sesji informuje, że działa w oparciu o AI), ujawnianie identyfikatorów AI i danych uwierzytelniających (np. agenci AI, którzy ujawniają swoją tożsamość AI w możliwym do zweryfikowania zakresie) ([21]).

([19]) Zob. motyw 132 aktu w sprawie AI. Nie wymaga to, aby system był ukierunkowany na te grupy szczególnie wrażliwe. Jeżeli system ma wchodzić w interakcję z dowolnym członkiem społeczeństwa, uwzględnia się również członków grup szczególnie wrażliwych, których dostawcy powinni uwzględnić przy wdrażaniu środków z zakresu powiadamiania.

([20]) Mające zastosowanie wymogi dostępności znajdują się w sekcji 7.3.

([21]) Na przykład elektroniczne poświadczenia atrybutów ustanowione na podstawie rozporządzenia (UE) nr 910/2014 i udostępnione w europejskich portfelach tożsamości cyfrowej ustanowionych na podstawie tego samego rozporządzenia, a także proponowane europejskie portfele biznesowe mogą zapewnić bezpieczne i skuteczne środki identyfikacji agentów AI. W portfelach można przechowywać elektroniczne poświadczenia, które weryfikują tożsamość, atrybuty i zezwolenia agenta AI, oraz zarządzać nimi, umożliwiając tym samym płynne i wiarygodne ujawnianie osobom fizycznym.

(37) **Format powiadomień:** jako najlepszą praktykę zaleca się projektowanie systemu AI z wykorzystaniem mechanizmów powiadamiania stosujących rozwiązania multimodalne i gwarantujących dostępność. Takie powiadomienia mogą łączyć różne techniki w celu zwiększenia zrozumienia przez użytkowników, w stosownych przypadkach i w zależności od kontekstu oraz docelowych i przewidywalnych grup odbiorców oraz dostępnych zasobów po stronie dostawców, w szczególności małych i średnich przedsiębiorstw (MŚP). Przykłady odpowiednich technik mogą obejmować:

— **przekazywanie informacji w formie tekstowej** (w oparciu o interfejs użytkownika): widoczne oznaczenia lub banery w prostym języku (np. „Wchodzisz w interakcję z systemem AI") oraz powitania na etapie pierwszej tury interakcji z chatbotem, które można łączyć, w stosownych przypadkach, z nieznikającymi znacznikami lub oznaczeniami widocznymi przez cały czas trwania interakcji, proporcjonalnie do konkretnego kontekstu i ryzyka. Ponadto zaleca się umieszczenie odnośnych informacji w pobliżu interfejsu interakcji (np. w pobliżu pola wejścia/wyjścia) i stosowanie uproszczonych sformułowań, w szczególności w odniesieniu do użytkowników o niższych umiejętnościach cyfrowych i dzieci,

— **przekazywanie informacji w formie dźwiękowej:** w kontekście komunikacji głosowej lub telefonicznej – jednoznaczne komunikaty głosowe na początku interakcji (np. „Rozmawiasz z asystentem AI") w połączeniu, w stosownych przypadkach, z okresowymi przypomnieniami w trakcie dłuższych interakcji, w szczególności w przypadku przerw lub zmiany roli systemu AI w toku korzystania przez użytkownika z danej usługi lub danego systemu. Charakterystyczne sygnały dźwiękowe (np. sygnały tonowe lub dźwiękowe sygnały informacyjne (ang. *earcons*)) mogą być pomocne w identyfikacji, zwłaszcza dla użytkowników słabowidzących, ale nie uznaje się ich samych w sobie za wystarczające,

— **sygnały wizualne/graficzne:** projektowanie systemu AI z wykorzystaniem stale widocznych ikon, znaków wodnych, kolorowych ramek lub rozpoznawalnych symboli „AI" w celu uzupełnienia przekazywania informacji w formie tekstowej. Zaleca się stosowanie znormalizowanych wskaźników wizualnych, również w różnych usługach oferowanych przez tego samego dostawcę, ponieważ takie wskaźniki mogą znacznie zmniejszyć obciążenie poznawcze i zwiększyć rozpoznawalność,

— **kombinacje multimodalne:** połączenie komunikatów tekstowych, dźwiękowych i wizualnych (np. chatbot wyświetlający oznaczenie, a jednocześnie przekazujący informację tekstową na etapie pierwszej tury interakcji) w celu zapewnienia dostępności i zwiększenia jasności dla całego przekroju grup użytkowników.

(38) Niektóre techniki nie są konieczne, a gdy są stosowane samodzielnie, są niewystarczające do skutecznego wypełnienia obowiązku w zakresie przejrzystości określonego w art. 50 ust. 1 i 5 aktu w sprawie AI, w szczególności:

— ujawnianie informacji wyłącznie w regulaminie, za pośrednictwem adresów URL lub w dokumentacji (takie formy informowania mogą stanowić uzupełnienie, choć nie zastępują, przekazywania informacji w kontekście danej interakcji),

— oznakowania nadające się do odczytu maszynowego (np. metadane lub znaki wodne), które nie są dostrzegalne dla użytkowników w momencie interakcji. Nie ma to wpływu na adekwatność takich znaków w kontekście zapewnienia zgodności z art. 50 ust. 2 aktu w sprawie AI (zob. sekcja 4),

— niejasne lub niejednoznaczne sygnały (np. ogólne odniesienia do „asystenta") lub antropomorficzne formy prezentacji, które mogą wprowadzać użytkowników w błąd,

— ogólnikowe informacje, które nie są wystarczająco specyficzne dla wyników systemu AI i interakcji z takim systemem (np. na platformie oferującej różne usługi lub indywidualne elementy ogólna informacja, taka jak „Usługi na tej stronie internetowej wykorzystują AI", jest niewystarczająca),

— opisy techniczne lub opisy odnoszące się do zdolności systemu: sformułowania odnoszące się wyłącznie do podstawowych technologii (np. „system ten wykorzystuje LLM") bez wyjaśnienia funkcji lub implikacji systemu dla użytkownika i sztucznego pochodzenia systemu.

(39) Ujawnianie informacji powinno być proporcjonalne i dostosowane do poszczególnych kontekstów i grup docelowych systemu AI, aby zachować skuteczność, przyjazność dla użytkownika i uniknąć, w zakresie, w jakim jest to możliwe, zakłóceń w doświadczeniu użytkownika lub efektu przyzwyczajenia (tzw. „ślepota banerowa").

(40) Jednorazowe, wyraźne powiadomienie przed pierwszą interakcją systemu AI z konkretną osobą fizyczną prawdopodobnie będzie wystarczające w większości przypadków do wypełnienia obowiązku określonego w art. 50 ust. 1 aktu w sprawie AI. W pewnych kontekstach związanych z wyższym ryzykiem konieczne mogą być jednak okresowe przypomnienia i powiadomienia z uwzględnieniem kontekstu, aby zagwarantować, że użytkownik jest stale świadomy charakteru interakcji. Dotyczy to w szczególności systemów AI wchodzących w interakcję z osobami wymagającymi szczególnego traktowania (np. dziećmi, osobami starszymi, osobami z niepełnosprawnościami i osobami o niższym poziomie umiejętności czytania i pisania), długotrwałych i ewoluujących interakcji w kontekstach wrażliwych lub immersyjnych (np. w których użytkownicy mogą wyrażać lub doświadczać stresu emocjonalnego lub emocjonalnej bezbronności lub być narażeni na ryzyko zachowań o charakterze uzależniającym), w przypadku których istnieje zwiększone ryzyko wprowadzenia użytkowników w błąd (np. doradztwo finansowe, ubezpieczenia, pomoc prawna, doradztwo zdrowotne i rozpatrywanie skarg) lub kształtowania więzi emocjonalnych lub stanów zależności (np. wirtualni towarzysze oparci na AI), a także gdy istnieje potrzeba zapewnienia świadomości i kontroli działań agentów AI. Dostawcy muszą również projektować i rozwijać system AI z myślą o zapewnieniu ujawniania stosownych informacji we wszystkich sytuacjach, w których systemowi AI zadawane są pytania dotyczące jego charakteru lub źródła interakcji lub w których na podstawie wymiany informacji z osobą fizyczną można racjonalnie założyć, że osoba ta może zostać wprowadzona w błąd co do faktu, że drugą stroną interakcji jest system AI, lub może nie mieć co do tego pewności.

3.2. **Wyjątki od obowiązku w zakresie informowania na podstawie art. 50 ust. 1 aktu w sprawie AI**

(41) W art. 50 ust. 1 aktu w sprawie AI przewidziano dwa wyjątki od obowiązku w zakresie informowania w odniesieniu do interaktywnych systemów AI, które to wyjątki omówiono poniżej.

3.2.1. *Wyjątek dotyczący oczywistej interakcji z systemem AI*

(42) Aby móc skorzystać z pierwszego wyjątku, dostawcy będą musieli ocenić i wykazać: (i) ewidentnie sztuczny charakter interakcji z (ii) osobą fizyczną, która musi być dostatecznie poinformowana, uważna i ostrożna, z uwzględnieniem okoliczności i kontekstu korzystania.

(43) Norma ta opiera się na ugruntowanym unijnym prawie ochrony konsumentów w odniesieniu do pojęcia „przeciętnego konsumenta" i należy interpretować ją w sposób spójny, przy jednoczesnym uwzględnieniu szczególnych i zróżnicowanych kontekstów i zastosowań systemów AI wchodzących w bezpośrednią interakcję z osobami fizycznymi oraz celów określonych w art. 50 ust. 1 aktu w sprawie AI.

(44) Sprecyzowanie profilu takiej „przeciętnej" osoby fizycznej składa się z dwóch etapów. Po pierwsze dostawcy powinni wziąć pod uwagę odbiorców, z którymi system AI ma wchodzić w interakcję i z którymi z dużym prawdopodobieństwem będzie wchodzić w interakcję. Po drugie dostawca powinien przeanalizować, na ile dostatecznie poinformowany, uważny i ostrożny jest przeciętny członek tej grupy odbiorców, z uwzględnieniem okoliczności i kontekstu korzystania. Przeprowadzając tę ocenę, należy wziąć pod uwagę kilka czynników:

— dostawcy powinni wziąć pod uwagę docelową grupę odbiorców systemu AI, a także szerszą potencjalną grupę odbiorców, którą można racjonalnie przewidzieć. Dotyczy to w szczególności przypadków, w których ogół społeczeństwa ma łatwy dostęp do systemu AI lub gdy system działa w wirtualnej lub fizycznej przestrzeni publicznej i jest w stanie wchodzić w interakcję z osobami fizycznymi obecnymi w tej przestrzeni. Odbiorcy, którzy mogą mieć styczność z systemem AI, nie zawsze należą do grup docelowych,

— dostawcy powinni wziąć pod uwagę potencjalnie zróżnicowany skład grupy docelowej i racjonalnie przewidywalnej grupy odbiorców zgodnie z motywem 132 aktu w sprawie AI. W przypadku gdy do tej grupy odbiorców mogą należeć osoby z niepełnosprawnościami, osoby starsze lub małoletni, oczekiwany poziom poinformowania, uważności i ostrożności przeciętnego członka takiej grupy odbiorców będzie niższy w porównaniu z poziomem u przeciętnego członka grupy odbiorców, która nie obejmuje tych kategorii osób. Jeżeli natomiast interaktywny system AI jest dostępny wyłącznie dla profesjonalnej lub wyspecjalizowanej grupy docelowej, mogłoby to mieć pozytywny wpływ na oczekiwany poziom wiedzy, uważności i ostrożności, a przeciętnego członka tej grupy można uznać za lepiej poinformowanego oraz bardziej uważnego i ostrożnego,

— w ramach szerszych okoliczności i szerszego kontekstu korzystania dostawcy powinni wziąć pod uwagę poziom umiejętności cyfrowych i kompetencji w zakresie AI wśród docelowych i racjonalnie przewidywalnych odbiorców oraz istotne czynniki społeczne, kulturowe i językowe.

(45) **Oczywisty charakter interakcji**: po ustaleniu hipotetycznej, stosunkowo dobrze poinformowanej, uważnej i ostrożnej osoby fizycznej dostawcy powinni ocenić, czy dla takiej osoby będzie oczywiste, że wchodzi ona w bezpośrednią interakcję z systemem AI. Wyjątek ten należy interpretować w sposób zawężający, ponieważ pozbawia on osoby fizyczne ochrony i prawa do bycia informowanymi za pomocą jasnych i wyraźnie wyodrębnionych informacji o tym, że wchodzą one w interakcję z systemami AI. Ogólna świadomość konsumentów i innych osób fizycznych co do istnienia systemów AI (w tym chatbotów i agentów) nie oznacza, że rozpoznają je w interakcjach. Przy ocenianiu, czy charakter interakcji jest oczywisty, należy wziąć pod uwagę kilka istotnych czynników:

— przewidywany charakter interakcji: w przypadkach, w których system AI wchodzi w fizyczną interakcję z osobami fizycznymi, istotne elementy mogą obejmować widoczność elementów mechanicznych (co zwiększyłoby oczywistość) lub stopień, w jakim system AI wiernie odtwarza swój naturalny, ludzki odpowiednik (co zmniejszyłoby oczywistość). W przypadku interakcji o charakterze wizualnym lub dźwiękowym na ocenę tę mogą mieć wpływ takie elementy, jak wzorce pisania lub mowy, ton głosu (głos robota a głos brzmiący jak głos prawdziwego człowieka), projekt interfejsu użytkownika (np. zdjęcie profilowe chatbota przedstawiające człowieka) oraz zdolność do zaawansowanej spersonalizowanej interakcji,

— skład grupy docelowej i przewidywalnej grupy odbiorców: jeżeli interaktywny system AI jest dostępny wyłącznie dla odbiorców profesjonalnych lub specjalistycznych, a słabsze grupy społeczne nie powinny potencjalnie mieć z nim styczności, mogłoby to zwiększyć możliwą oczywistość sztucznego charakteru interakcji u przeciętnego członka danej grupy odbiorców. Jeżeli natomiast dostęp do systemu AI mają ogół społeczeństwa lub konsumenci, w tym również osoby wymagające szczególnego traktowania (np. dzieci, osoby starsze, osoby z niepełnosprawnością poznawczą, fizyczną lub umysłową lub osoby o niższych kompetencjach w zakresie AI), dla których charakter interakcji jest mniej oczywisty, nie można powoływać się na wyjątek w celu skutecznej ochrony tych użytkowników wymagających szczególnego traktowania,

— należy również wziąć pod uwagę fakt, że osobom fizycznym coraz trudniej jest ustalić, czy ma miejsce interakcja z systemem AI, czy też z człowiekiem. Ze względu na korzystanie z awatarów i agentów, ogólną anonimowość interakcji w internecie oraz zdolność systemów AI do pisania i wchodzenia w interakcję w sposób typowy dla człowieka wyjątek dotyczący „oczywistości" powinien być ograniczony do przypadków, w których nie ma prawie żadnych wątpliwości co do charakteru interakcji przeciętnej osoby należącej do docelowej i racjonalnie przewidywalnej grupy odbiorców.

Przykłady oczywistości, do których nie ma zastosowania obowiązek w zakresie przejrzystości określony w art. 50 ust. 1 aktu w sprawie AI:

— chatboty do wspomagania programowania i przeglądu kodu oparte na AI, dostępne wyłącznie dla profesjonalnych programistów, którzy ze względu na swoją wiedzę fachową i kontekst korzystania z tych narzędzi powinni, z dużą dozą prawdopodobieństwa, zdawać sobie sprawę, że nie wchodzą w interakcję z człowiekiem, i mogą łatwo rozpoznać, że podpowiedzi są generowane przez system AI;

— wewnętrzny asystent dla pracowników, przeznaczony dla odpowiednio przeszkolonego personelu, który potrafi korzystać ze sztucznej inteligencji i wie, że wykorzystuje systemy AI do wewnętrznych celów organizacyjnych (np. kadrowych, prawnych, związanych z zamówieniami publicznymi, zgodnością z przepisami lub wsparciem informatycznym);

— interaktywne systemy AI przeznaczone wyłącznie do wykorzystania przez odpowiednio przeszkolonych pracowników służby zdrowia w celu wsparcia diagnostyki medycznej i proponowania odpowiedniego leczenia;

— AI działająca w otoczeniu wbudowana w sprzęt gospodarstwa domowego z funkcjami ograniczonymi do pomocy i obsługi urządzenia gospodarstwa domowego;

— interakcje z postaciami niezależnymi (NPC) wykorzystującymi AI w grach wideo dla jednego gracza, w przypadku gdy charakter gry wideo jasno wskazuje, że żadna inna osoba fizyczna nie może uczestniczyć w grze i nie może wchodzić w interakcję z użytkownikiem, a charakter interakcji oparty na AI jest również oczywisty dla wszystkich użytkowników (w tym osób wymagających szczególnego traktowania).

Przykłady, które nie spełniają wyjątku dotyczącego oczywistego charakteru interakcji i do których ma zastosowanie obowiązek w zakresie przejrzystości określony w art. 50 ust. 1 aktu w sprawie AI:

— oparty na AI robot-zwierzę do towarzystwa, wyglądający bardzo podobnie do prawdziwego zwierzęcia i zaprojektowany tak, aby naśladować typową interakcję między człowiekiem a zwierzęciem, co prowadzi do niejasności co do tego, czy interakcja odbywa się z systemem AI, czy nie;

— systemy AI wbudowane w środowiska immersyjne (np. rzeczywistość wirtualną lub rozszerzoną) wykorzystujące realistyczne awatary lub głosy przypominające ludzi, w przypadku których użytkownicy, a w szczególności dzieci, osoby starsze lub osoby z niepełnosprawnościami, mogą nie być w stanie łatwo odróżnić interakcji z człowiekiem od interakcji ze sztuczną inteligencją;

— chatboty AI wbudowane w platformy internetowe lub narzędzia wsparcia (centra pomocy), za pośrednictwem których użytkownicy wchodzą w bezpośrednią interakcję z systemem AI i otrzymują wyniki AI (np. odpowiedzi na zapytania lub inne treści wygenerowane przez AI), które mogą postrzegać jako wygenerowane przez człowieka.

3.2.2. *Wyjątek dotyczący systemów AI, których wykorzystywanie jest dozwolone na mocy prawa do celów ścigania przestępstw*

(46) Dostawcy interaktywnych systemów AI są zwolnieni z obowiązku w zakresie przejrzystości określonego w art. 50 ust. 1 aktu w sprawie AI, jeżeli wykorzystywanie tych systemów jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, z zastrzeżeniem odpowiednich zabezpieczeń w zakresie praw i wolności osób trzecich.

(47) Wyrażenie „dozwolone na mocy prawa" należy rozumieć jako obejmujące prawo Unii, a także przepisy prawa krajowego państw członkowskich przyjęte zgodnie z prawem Unii. Prawo zezwalające na wykorzystanie systemu AI musi jasno określać cele i okoliczności dozwolonego wykorzystania oraz przewidywać odpowiednie zabezpieczenia w celu ochrony praw i wolności osób trzecich. Biorąc pod uwagę neutralne pod względem technologicznym podejście określone w akcie w sprawie AI, nie ma konieczności prawnego identyfikowania lub autoryzowania konkretnych systemów AI, jeżeli wykorzystanie systemu AI jest zasadniczo przewidziane w ramach prawnie dozwolonej działalności i jest zgodne z mającymi zastosowanie ramami prawnymi.

(48) Aby można było objąć system AI tym wyjątkiem, jego celem musi być wykrywanie przestępstw, zapobieganie im, prowadzenie postępowań przygotowawczych w ich sprawie lub ściganie ich sprawców (np. agent AI działający pod przykryciem, systemy AI wykorzystywane pod kontrolą człowieka przez organy ścigania do prowadzenia czynności dochodzeniowych w internecie lub do komunikacji z osobami, które mogą być zaangażowane w działalność przestępczą). Wyjątek ten nie ogranicza się do wykorzystywania takich systemów AI wyłącznie przez organy ścigania zdefiniowane w art. 3 pkt 48 aktu w sprawie AI, ale może również obejmować systemy interaktywnej (lub generatywnej) AI wykorzystywane przez inne unijne lub krajowe organy publiczne, o ile ich wykorzystywanie jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, z zastrzeżeniem odpowiednich zabezpieczeń w zakresie praw i wolności osób trzecich. Przedmiotowe wyłączenie nie wymaga, aby system AI był w stanie prowadzić czynności dochodzeniowe w sposób niezależny. Jeżeli dany system może być wykorzystywany zarówno do celów egzekwowania prawa, jak i do innych zwykłych celów, możliwość zastosowania wyłączenia należy oceniać w świetle konkretnego użycia systemu. Obowiązki w zakresie przejrzystości nadal mają zastosowanie w przypadku, gdy ten sam system AI jest wykorzystywany do celów innych niż dozwolone na mocy prawa wykrywanie przestępstw, zapobieganie im, prowadzenie postępowań przygotowawczych w ich sprawie lub ściganie ich sprawców. Nadal należy zapewnić przejrzystość osobom fizycznym, gdy wchodzą one w interakcję z systemem w kontekście jego zwykłych celów.

(49) Wyjątek ten nie ma zastosowania, jeżeli (i) system AI jest publicznie dostępny oraz (ii) system oferuje osobom fizycznym funkcję zgłaszania przestępstw. Takie systemy mogą być wykorzystywane, ale nadal podlegają obowiązkowi w zakresie przejrzystości na podstawie art. 50 ust. 1 aktu w sprawie AI, w związku z czym osoby fizyczne muszą być informowane o tym, że wchodzą w interakcję z takimi systemami.

Dla większej jasności wyjątek ten dotyczy wyłącznie wynikającego z art. 50 ust. 1 obowiązku informowania osób fizycznych o tym, że wchodzą one w interakcję z systemem AI ([22]). Pozostaje to bez uszczerbku dla mającego zastosowanie prawa Unii i prawa krajowego, w tym przepisów proceduralnych dotyczących dokumentacji, ujawniania, dopuszczalności i oceny dowodów.

([22]) Dotyczy to również podobnych wyjątków wprowadzonych z myślą o działaniach organów ścigania przewidzianych w art. 50 ust. 2–4 aktu w sprawie AI.

Przykłady systemów AI udostępnianych publicznie do celów zgłaszania przestępstw, które to systemy nie są objęte zwolnieniem i wchodzą w zakres stosowania art. 50 ust. 1 aktu w sprawie AI:

— chatboty policyjne oparte na AI, umieszczone na oficjalnych stronach internetowych policji lub w aplikacjach mobilnych, które umożliwiają osobom fizycznym zgłaszanie przestępstw lub skarg. Systemy te zadają użytkownikom pytania według ustalonego schematu, gromadzą istotne informacje i dokonują selekcji zgłoszeń, a następnie przekazują je do oceny funkcjonariuszom (ludziom),

— infolinie wykorzystujące AI, prowadzone przez policję lub inny organ ścigania do celów obsługi ogólnych zapytań obywateli,

— wspomagane przez AI infolinie lub portale cyfrowe do zgłaszania nadużyć finansowych prowadzone przez instytucje finansowe lub organy publiczne, za pośrednictwem których użytkownicy mogą zgłaszać podejrzenia popełnienia przestępstw finansowych. System AI wchodzi w interakcję z użytkownikiem w celu gromadzenia szczegółowych informacji, kategoryzacji zgłoszenia i priorytetowego traktowania spraw, natomiast osoby prowadzące dochodzenie zatwierdzają informacje przed podjęciem jakichkolwiek działań,

— wirtualni asystenci wykorzystywani do zbierania zeznań świadków, w przypadku gdy ofiary lub świadkowie mogą przekazywać informacje za pośrednictwem interfejsu opartego na AI. Takie systemy AI mogą obsługiwać wielojęzyczne dane wejściowe, funkcje dostępności i ustrukturyzowane przedkładanie dowodów, ale ostateczna ocena i czynności dochodzeniowe pozostają w gestii funkcjonariuszy.

3.3. Wzajemne powiązania z innymi aktami prawnymi Unii

(50) Obowiązek w zakresie przejrzystości wynikający z art. 50 ust. 1 aktu w sprawie AI ma zastosowanie bez uszczerbku dla obowiązujących unijnych przepisów dotyczących ochrony konsumentów, ochrony danych, cyfryzacji lub niedyskryminacji, które mogą wymagać ujawniania innych informacji, które nie dotyczą konkretnie faktu, że źródłem interakcji jest AI. W szczególności na mocy dyrektywy 2005/29/WE (dyrektywa o nieuczciwych praktykach handlowych) ([23]) zakazuje się działań lub zaniechań wprowadzających w błąd w odniesieniu do głównych cech produktu lub usługi. Ponadto dyrektywa 2011/83/WE („dyrektywa w sprawie praw konsumentów") ([24]) zobowiązuje przedsiębiorców do udzielania jasnych i zrozumiałych informacji przed zawarciem umowy na temat głównych cech towarów lub usług. W przypadku gdy usługa opiera się na AI (np. chatbot dostępny w modelu subskrypcyjnym, aplikacja towarzysząca oparta na AI lub wirtualna usługa coachingu), funkcja AI może kwalifikować się jako istotna cecha, którą należy ujawnić, zanim konsument zostanie związany umową. Obowiązki informacyjne wynikające z unijnego prawa ochrony konsumentów mają zastosowanie niezależnie od tego, czy interakcję uznaje się za „oczywistą" zgodnie z art. 50 ust. 1 aktu w sprawie AI. Obowiązki te mają również charakter uzupełniający wobec ogólnych obowiązków informacyjnych w zakresie ochrony konsumentów, natomiast obowiązek powiadamiania wynikający z aktu w sprawie AI ma bardziej kontekstowy charakter i jego celem jest zapewnienie stałej świadomości wszystkich osób fizycznych na temat bezpośredniej interakcji z systemem AI co najmniej przed rozpoczęciem interakcji i przez cały czas jej trwania.

(51) Akt o usługach cyfrowych również zawiera wymogi dotyczące przejrzystości systemów rekomendacji działających na platformach internetowych, w tym wymóg wyjaśnienia głównych parametrów plasowania. Obowiązki te stanowią uzupełnienie obowiązków w zakresie przejrzystości określonych w art. 50 ust. 1 aktu w sprawie AI, ponieważ mają inny zakres i cel oraz dotyczą systemów, które co do zasady nie wchodzą w bezpośrednią interakcję z użytkownikami w rozumieniu art. 50 ust. 1 aktu w sprawie AI.

(52) Jeżeli chodzi o obowiązki informacyjne wobec osób, których dane dotyczą, wynikające z unijnego prawa o ochronie danych, obowiązek określony w art. 50 ust. 1 aktu w sprawie AI służy innemu celowi niż te obowiązki i jako taki nie ma wpływu na te obowiązki informacyjne.

([23]) Dyrektywa 2005/29/WE Parlamentu Europejskiego i Rady z dnia 11 maja 2005 r. dotycząca nieuczciwych praktyk handlowych stosowanych przez przedsiębiorstwa wobec konsumentów na rynku wewnętrznym (Dz.U. L 149 z 11.6.2005, s. 2).

([24]) Dyrektywa Parlamentu Europejskiego i Rady 2011/83/UE z dnia 25 października 2011 r. w sprawie praw konsumentów (Dz.U. L 304 z 22.11.2011, s. 64).

(53) Podobnie obowiązek w zakresie przejrzystości wynikający z art. 50 ust. 1 aktu w sprawie AI ma zastosowanie bez uszczerbku dla prawa Unii i prawa krajowego dotyczącego udzielania informacji osobom w kontekście postępowania karnego (np. dyrektywy 2012/13/UE lub dyrektywy 2012/29/UE zmienionej dyrektywą (UE) 2026/1472) ([25]).

4. ART. 50 UST. 2 AKTU W SPRAWIE AI: OZNACZANIE I WYKRYWANIE TREŚCI WYGENEROWANYCH LUB ZMANIPULOWANYCH PRZEZ AI

(54) W art. 50 ust. 2 aktu w sprawie AI zobowiązano dostawców systemów AI generujących treści syntetyczne do wdrożenia rozwiązań technicznych, które spełniają określone wymogi jakościowe dotyczące oznakowania nadającego się do odczytu maszynowego i wykrywania wyników ich systemów AI. Celem jest umożliwienie osobom fizycznym odróżnienia treści wygenerowanych lub zmanipulowanych przez AI od innych treści (na przykład treści utworzonych przez ludzi) oraz zweryfikowania ich źródła, a tym samym przyczynienie się do zwiększenia integralności i zaufania do ekosystemu informacyjnego.

(55) Beneficjentami tego obowiązku w zakresie przejrzystości są zatem wszystkie osoby fizyczne, które mogą mieć styczność z treściami wygenerowanymi lub zmanipulowanymi przez AI, a także kluczowe podmioty, które odgrywają ważną rolę dla zaufania do ekosystemu informacyjnego i jego integralności, takie jak właściwe organy nadzoru rynku i inne właściwe organy, niezależni badacze, organizacje społeczeństwa obywatelskiego, media, zaufane podmioty sygnalizujące, obrońcy praw podstawowych, dostawcy platform internetowych i wyszukiwarek internetowych, posiadacze praw oraz inni twórcy i dystrybutorzy treści itp.

4.1. Główne elementy i pojęcia, o których mowa w art. 50 ust. 2 aktu w sprawie AI

(56) Art. 50 ust. 2 aktu w sprawie AI ma zastosowanie, jeżeli spełnionych jest łącznie kilka warunków:

(i) system musi kwalifikować się jako system AI;

(ii) system AI musi być w stanie generować treści syntetyczne lub nimi manipulować;

(iii) treści muszą być przedstawione w jednej lub kilku z następujących form: dźwięk, obraz, wideo lub tekst;

(iv) system AI nie może być objęty żadnym z następujących wyjątków: (1) system AI pełni funkcję wspomagającą w zakresie standardowej edycji; (2) system AI nie zmienia w istotny sposób danych wejściowych dostarczonych przez podmiot stosujący ani ich semantyki; lub (3) wykorzystywanie systemu AI jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców.

4.1.1. *Systemy AI generujące treści syntetyczne lub manipulujące nimi*

(57) Art. 50 ust. 2 aktu w sprawie AI ma zastosowanie do systemów AI ([26]), w tym systemów AI ogólnego przeznaczenia, generujących treści w postaci syntetycznych dźwięków, obrazów, wideo lub tekstu. Art. 50 ust. 2 aktu w sprawie AI obejmuje podzbiór systemów, które spełniają definicję systemu AI zawartą w art. 3 pkt 1 aktu w sprawie AI i które generują takie wyniki. Art. 50 ust. 2 aktu w sprawie AI obejmuje również systemy AI wykorzystywane do manipulowania treściami.

(58) Generowanie treści odnosi się do generowania materiałów syntetycznych przez system AI (np. na podstawie polecenia człowieka), na przykład generowania obrazu syntetycznego lub piosenki. W praktyce oznacza to, że art. 50 ust. 2 aktu w sprawie AI może mieć zastosowanie do systemów generatywnej AI stworzonych do realizacji wąskiego celu, jakim jest generowanie konkretnych wyników. Jednocześnie art. 50 ust. 2 aktu w sprawie AI ma również zastosowanie do systemów AI, które mogą służyć wielu celom, w tym do systemów AI ogólnego przeznaczenia, i które mogą generować różne rodzaje treści, a także do agentowych systemów AI, o ile generują one treści w postaci syntetycznych dźwięków, obrazów, wideo lub tekstu.

([25]) Dyrektywa Parlamentu Europejskiego i Rady 2012/13/UE z dnia 22 maja 2012 r. w sprawie prawa do informacji w postępowaniu karnym (Dz.U. L 142 z 1.6.2012, s. 1). Dyrektywa Parlamentu Europejskiego i Rady 2012/29/UE z dnia 25 października 2012 r. ustanawiająca normy minimalne w zakresie praw, wsparcia i ochrony ofiar przestępstw (Dz.U. L 315 z 14.11.2012, s. 57).

([26]) Zgodnie z definicją w art. 3 pkt 1 aktu w sprawie AI.

(59) Z kolei manipulacja treścią odnosi się do już istniejących treści (syntetycznych lub niesyntetycznych), które są zmieniane przez system AI (np. do istniejącego obrazu lub nagrania głosowego, które są poddawane manipulacji przez system AI zgodnie z instrukcjami człowieka). Aby treści wchodziły w zakres stosowania art. 50 ust. 2 aktu w sprawie AI, wystarczy, aby były generowane przez AI lub poddawane manipulacji wykraczającej poza standardową edycję na podstawie wyjątków określonych w art. 50 ust. 2 aktu w sprawie AI (zob. sekcja 4.3). Ponadto w art. 50 ust. 2 aktu w sprawie AI nie wymaga się, aby treści były wyłącznie generowane lub poddawane manipulacji przez AI. Treści, które są mieszane z materiałami stworzonymi przez człowieka, również kwalifikują się jako treści syntetyczne, jeżeli zostały zmanipulowane lub wygenerowane w jednym z rodzajów treści, o których mowa w art. 50 ust. 2 aktu w sprawie AI.

4.1.2. *Rodzaje treści syntetycznych objęte zakresem stosowania*

(60) Art. 50 ust. 2 aktu w sprawie AI zawiera wyczerpujący wykaz rodzajów treści syntetycznych objętych obowiązkiem w zakresie przejrzystości: dźwięk, obraz, wideo lub tekst. Obejmuje to również treści multimodalne, tj. treści składające się z połączenia tych rodzajów. Do celów art. 50 ust. 2 aktu w sprawie AI poszczególne rodzaje treści należy interpretować w świetle praktycznych i technologicznych zmian, które mogą z czasem ewoluować. Na obecnym etapie należy je rozumieć w następujący sposób:

— tekst odnosi się do treści symbolicznej składającej się z wyodrębnionych jednostek – znaków lub liczb uporządkowanych w określonej sekwencji, które mogą być odczytywane i interpretowane pod kątem semantycznym przez człowieka,

— obrazy odnoszą się do statycznych reprezentacji przestrzennych kodujących informacje wizualne możliwe do zobaczenia przez człowieka,

— dźwięk odnosi się do zmiennego w czasie sygnału kodującego fale dźwiękowe, które mogą być odbierane za pomocą słuchu przez człowieka. Może to obejmować mowę, muzykę instrumentalną lub inne sygnały dźwiękowe,

— wideo odnosi się do czasowej sekwencji obrazów, która może być zsynchronizowana z dźwiękiem i jest możliwa do zobaczenia przez człowieka. W takim przypadku należy oznaczyć zarówno treści wideo, jak i treści dźwiękowe oraz zapewnić ich wykrywalność, jak określono poniżej.

(61) Systemy AI generujące obrazy 3D, wideo 3D i dźwięki 3D lub manipulujące nimi, jak również systemy AI generujące rzeczywistość wirtualną, rozszerzoną lub mieszaną wchodzą w zakres stosowania art. 50 ust. 2 aktu w sprawie AI. „Rzeczywistość wirtualna" obejmuje technologie AI, które umożliwiają użytkownikom doświadczanie środowiska wygenerowanego komputerowo symulującego fizyczną obecność i interakcję z nim oraz umożliwiają użytkownikom interakcję w czasie rzeczywistym z tym środowiskiem. Ponieważ rzeczywistość wirtualna i rozszerzona składa się z czasowej sekwencji obrazów, opcjonalnie zsynchronizowanej z dźwiękiem, która odzwierciedla zmianę wizualną w czasie, można ją uznać za treści „wideo". Odpowiednim „wynikiem", który należy oznaczyć, jest ostateczna forma poszczególnych aktywów wygenerowanych przez AI, ze szczególnymi wyjątkami dotyczącymi efemerycznych sesji interaktywnych w czasie rzeczywistym, o których mowa w pkt 88 poniżej. Nie wymaga się oznaczania artefaktów cyfrowych wykorzystywanych do generowania obrazów 3D, nagrań wideo 3D i treści audio 3D, takich jak chmura punktów i siatka 3D.

(62) Cyfrowe bliźniaki, które stanowią wirtualne odwzorowanie osób, obiektów fizycznych lub systemów, wykorzystujące wiele rodzajów danych, również wchodzą w zakres stosowania art. 50 ust. 2 aktu w sprawie AI, jeżeli zostały wygenerowane i zmanipulowane przez AI i przyjmują formę dźwięku, obrazu, wideo lub tekstu, chyba że wchodzą w zakres wyjątków i kwalifikują się jako „zastosowania przemysłowe" lub „zastosowania między przedsiębiorstwami", o których mowa w pkt 87 poniżej, lub jako treści ulotne w czasie rzeczywistym, o których mowa w pkt 88 poniżej.

(63) Art. 50 ust. 2 aktu w sprawie AI może mieć również zastosowanie do agentów AI, jeżeli agent AI podejmuje działanie, którego wynikiem są treści wygenerowane lub zmanipulowane przez AI, dostrzegalne dla osób fizycznych w formie dźwięku, obrazu, wideo lub tekstu. Wyniki te muszą być zatem oznaczone i wykrywalne w sposób opisany poniżej. Natomiast wszelkie działania agenta AI, które nie obejmują treści wygenerowanych lub zmanipulowanych przez AI (np. pośrednie etapy przetwarzania, takie jak rozumowanie i łańcuch myśli), lub działania wygenerowane przez AI, które z założenia nie mają być bezpośrednio dostrzegalne dla osób fizycznych (np. żądanie sieciowe lub działanie w przeglądarce), nie kwalifikują się jako treści syntetyczne objęte zakresem stosowania art. 50 ust. 2 aktu w sprawie AI i nie muszą być zgodne z obowiązkami w zakresie oznaczania i wykrywania.

4.1.3. *Treści nieobjęte zakresem stosowania art. 50 ust. 2 aktu w sprawie AI*

(64) Treści wygenerowane przez system AI oparty na prostym przetwarzaniu danych, które nie zostały konkretnie wygenerowane lub zmanipulowane przez AI, na przykład wyrenderowana klatka, nie wchodzą w zakres stosowania art. 50 ust. 2 aktu w sprawie AI.

(65) Ponadto wyniki systemu AI, które jedynie odtwarzają istniejące treści lub umożliwiają prezentację lub uporządkowanie już istniejących treści, nie są objęte art. 50 ust. 2 aktu w sprawie AI. Może to dotyczyć np. list odtwarzania utworów muzycznych lub innych systemów rekomendacji, które wybierają lub plasują istniejące treści wyłącznie na podstawie preferencji lub aktywności użytkownika, nie tworząc żadnych nowych treści ani nie manipulując nimi. To samo dotyczy systemów AI wykorzystywanych w wewnętrznych procesach analitycznych, które pozyskują i porządkują dane, ale ich nie podsumowują.

(66) To samo dotyczy również danych wejściowych i wyjściowych systemu AI, które stanowią jedynie dane z monitorowania i zapisy danych ze środowisk fizycznych lub wirtualnych, które to dane nie są generowane ani manipulowane przez AI (np. przez roboty lub inne czujniki oparte na AI), lub transmisje danych przez systemy AI bez żadnych zmian. Obejmuje to różnorodne dane wejściowe i wyjściowe systemów AI, w szczególności w warunkach przemysłowych, takie jak dane produkcyjne dla robotów monitorujących dane; dane dotyczące zużycia rejestrowane przez inteligentne liczniki wykorzystujące AI, pomiary częstotliwości i napięcia w sieci; oraz rejestrowanie danych dotyczących lokalizacji GPS z pojazdów itp.

(67) Ponadto, ponieważ celem obowiązku w zakresie przejrzystości określonego w art. 50 ust. 2 aktu w sprawie AI jest umożliwienie ludziom odróżnienia treści wygenerowanych lub zmanipulowanych przez AI, tak aby przeciwdziałać ryzyku wprowadzania w błąd i manipulacji oraz zapewnić integralność ekosystemu informacyjnego i zaufanie do niego, przepis ten nie dotyczy wszelkich treści niezwiązanych z tym celem, które to treści nie są dostrzegalne lub nie są przeznaczone do interpretacji, weryfikacji lub podejmowania działań przez osoby fizyczne.

(68) W szczególności należy uznać, że następujące treści nie wchodzą w zakres stosowania art. 50 ust. 2 aktu w sprawie AI:

— wyniki wygenerowane w formie krótkiej sekwencji cyfr, symboli lub liter (np. pojedyncze słowa, podpisy pod obrazami, teksty alternatywne, etykiety UI i grafiki wielkości ikon oraz inne etykiety danych),

— kod źródłowy rozumiany jako treść zapisana w języku programowania, skryptowym, znaczników, zapytań lub konfiguracji, która ma być interpretowana, kompilowana lub wykonywana przez system komputerowy, niezależnie od tego, czy składa się ona ze znaków tekstowych. Może to również obejmować komentarze w języku naturalnym i informacje kontekstowe, które stanowią integralną część kodu źródłowego. Podobnie z zakresu stosowania wyłączone są zestawy do opracowywania oprogramowania (SDK), strukturalny język zapytań (SQL), infrastruktura jako kod oraz „YAML Ain't Markup Language" (YAML), format JavaScript Object Notation (JSON), schematy konfiguracji, skrypty, specyfikacje nadające się do odczytu maszynowego, interfejsy programowania aplikacji (API i biblioteki oprogramowania),

— wyniki systemu AI, które mają być przekazywane wyłącznie między maszynami i przetwarzane automatycznie bez styczności z człowiekiem, na przykład komunikacja między agentami, sygnały antyspamowe, o ile nie są dostrzegalne dla osób fizycznych,

— wyniki, które są wykorzystywane wyłącznie w środowiskach o obiegu zamkniętym w procesach przemysłowych i procesach rozwoju produktów (na przykład do produkcji filmów, animacji, gier lub reklam), chyba że są one końcowym produktem systemu AI w formie wygenerowanych lub zmanipulowanych przez AI treści w postaci tekstu, dźwięków, obrazów lub wideo. W takich przypadkach oznaczony i wykrywalny musi być wyłącznie ostateczny wynik procesu.

4.2. **Obowiązek oznaczania i zapewnienia wykrywalności określony w art. 50 ust. 2 aktu w sprawie AI**

(69) W art. 50 ust. 2 aktu w sprawie AI nałożono na dostawców systemów generatywnej AI objętych zakresem stosowania tego przepisu obowiązek w zakresie przejrzystości obejmujący dwa odrębne, ale z natury wzajemnie powiązane elementy. Po pierwsze dostawcy muszą zapewnić, aby wyniki systemu AI były oznaczone w formacie nadającym się do odczytu maszynowego. Po drugie dostawcy muszą zapewnić, aby wyniki były wykrywalne jako sztucznie wygenerowane lub zmanipulowane treści. W odniesieniu do każdego wdrożonego rozwiązania w zakresie oznaczania dostawcy powinni zapewnić dostępność odpowiednich środków wykrywania (jak określono poniżej), aby umożliwić osobom fizycznym mającym styczność z treściami i innym odpowiednim podmiotom identyfikację i odróżnienie treści wygenerowanych sztucznie przez AI od innych treści. Rozwiązanie techniczne należy rozumieć jako kombinację technik oznaczania i środków wykrywania, które dostawca wdrożył w celu wypełnienia obowiązku w zakresie przejrzystości na podstawie art. 50 ust. 2 aktu w sprawie AI.

(70) Każdy z dwóch wyżej wymienionych elementów musi zostać spełniony, aby skutecznie osiągnąć cele obowiązków w zakresie przejrzystości określonych w art. 50 ust. 2 aktu w sprawie AI. Spełnienie tylko jednego elementu (np. w odniesieniu do nadających się do odczytu maszynowego oznakowań wyników bez dostępnych środków ich wykrywania) nie wystarczy do zapewnienia zgodności z tym przepisem ([27]).

4.2.1. *Obowiązek oznaczania*

(71) Zakres obowiązku oznaczania ogranicza się do stosowania oznaczeń w formacie nadającym się do odczytu maszynowego. Format nadający się do odczytu maszynowego oznacza, że oznaczenia są ustrukturyzowane w sposób umożliwiający aplikacjom komputerowym ich łatwą identyfikację, rozpoznawanie i ekstrakcję bez interwencji człowieka. Nie wyklucza się wykorzystywania dostrzegalnych oznaczeń i etykiet jako środka uzupełniającego, w stosownych przypadkach, w celu ułatwienia podmiotom stosującym wywiązania się z obowiązku oznaczania treści typu deepfake zgodnie z art. 50 ust. 4 aktu w sprawie AI.

(72) Aby wywiązać się z obowiązku oznaczania, dostawcy mogą stosować jedną technikę oznaczania lub kombinację technik, o ile ich całościowe rozwiązanie techniczne umożliwia odczyt maszynowy i spełnia wymogi skuteczności, interoperacyjności, solidności i niezawodności w stopniu wymaganym prawem, jak przewidziano w art. 50 ust. 2 zdanie drugie aktu w sprawie AI (zob. sekcja 4.2.3 poniżej).

(73) W motywie 133 aktu w sprawie AI przedstawiono przykłady takich technik, które obejmują znaki wodne, identyfikację metadanych, metody kryptograficzne służące do potwierdzania pochodzenia i autentyczności treści, metody rejestracji zdarzeń, odciski cyfrowe lub inne techniki oraz połączenie takich technik. Chociaż metody potwierdzania pochodzenia i autentyczności wymieniono w motywie 133 aktu w sprawie AI, dostawcy nie są zobowiązani do rejestrowania ani przechowywania ewidencji pełnego łańcucha pochodzenia zawierającej informacje na temat pochodzenia i modyfikacji treści lub wszelkie inne istotne twierdzenia dotyczące historii treści. Takie metody ewidencjonowania pochodzenia mogą być jednak również wykorzystywane do zapewnienia zgodności z art. 50 ust. 2 aktu w sprawie AI i mogą służyć osobom fizycznym pomocą w odróżnianiu treści wygenerowanych lub zmanipulowanych przez AI od treści autentycznych.

(74) Dostawcy mogą wdrożyć rozwiązanie w zakresie oznaczania wyników na różnych etapach łańcucha wartości (np. post hoc po wygenerowaniu treści na poziomie systemu AI, na poziomie bazowego modelu AI lub jako element wbudowany w proces wnioskowania systemu AI). Mogą oni polegać na rozwiązaniu w zakresie oznaczania wdrożonym przez dostawcę modelu na wcześniejszym etapie łańcucha lub osobę trzecią dostarczającą rozwiązanie (np. jako standard otwarty lub usługa specjalistyczna), o ile takie rozwiązanie w zakresie oznaczania jest zgodne z art. 50 ust. 2 aktu w sprawie AI. Korzystanie z tych rozwiązań pozostaje bez uszczerbku dla odpowiedzialności dostawcy systemu AI za wykazanie zgodności z art. 50 ust. 2 aktu w sprawie AI.

([27]) W motywie 133 aktu w sprawie AI doprecyzowano, że należy zobowiązać dostawców tych systemów do wbudowania rozwiązań technicznych, które umożliwiają oznakowanie w formacie nadającym się do odczytu maszynowego i wykrywanie, że wyniki zostały wygenerowane lub zmanipulowane przez system AI, a nie przez człowieka. Motyw 135 aktu w sprawie AI odnosi się do obowiązku zapewnienia wykrywalności.

4.2.2. *Obowiązek zapewnienia wykrywalności*

(75) Dostawcy systemów AI objętych zakresem stosowania art. 50 ust. 2 aktu w sprawie AI muszą zapewnić, aby wyniki ich systemów były wykrywalne jako treści wygenerowane lub zmanipulowane przez AI. Oznacza to, że dostawca jest zobowiązany do zapewnienia, aby środki wykrywania były dostępne dla osób, które mogą mieć styczność z takimi treściami. Ponadto zgodnie z art. 50 ust. 5 aktu w sprawie AI takie rozwiązania w zakresie wykrywania powinny umożliwiać generowanie wyników w formacie czytelnym dla człowieka, wskazujących, czy treści zostały wygenerowane lub zmanipulowane przez AI.

(76) Narzędzie do wykrywania jest mechanizmem wykrywającym, czy treści zostały wygenerowane lub zmanipulowane przez AI, zwykle identyfikującym znaczniki techniczne lub podpisy, które weryfikują pochodzenie treści. Aby zapewnić pełną interoperacyjność, dostawcy muszą polegać na publicznie dostępnych branżowych standardowych rozwiązaniach w zakresie wykrywania, które umożliwiają każdej osobie trzeciej wdrożenie mechanizmu wykrywania, którego uruchamianie powinno optymalnie być możliwe na urządzeniu cyfrowym, a także umożliwiają płynną integrację z niezależnymi, odizolowanymi sieciami. W przypadku gdy takie standardowe rozwiązania nie są dostępne, w szczególności na początkowym etapie wdrażania art. 50 ust. 2 aktu w sprawie AI w odniesieniu do technologii opatrywania treści znakiem wodnym, dostawca może polegać na swoim własnym rozwiązaniu w zakresie wykrywania lub na rozwiązaniu w zakresie wykrywania opracowanym przez osobę trzecią lub wspólnym rozwiązaniu w zakresie wykrywania, o ile rozwiązania te zapewniają interoperacyjność z rozwiązaniami w zakresie wykrywania stosowanymi przez innych dostawców (zob. sekcja 4.2.3). Możliwość ta powinna być ograniczona w czasie do czasu opracowania norm zharmonizowanych i znormalizowanego, niezależnego od dostawcy interoperacyjnego rozwiązania w zakresie wykrywania, które będzie bezpieczne, będzie chronić prywatność i będzie działać lokalnie.

(77) Art. 50 ust. 5 aktu w sprawie AI stanowi ponadto, że informacje, o których mowa w art. 50 ust. 2 aktu w sprawie AI, należy przekazywać osobom fizycznym mającym styczność z treściami w jasny i wyraźny sposób, najpóźniej w momencie pierwszej interakcji lub pierwszego stosowania. Te informacje to w szczególności wynik wykrywania, który wskazuje, czy treści zostały wygenerowane lub zmanipulowane przez AI, a nie same treści wygenerowane lub zmanipulowane przez AI. „W momencie pierwszego stosowania (pierwszej styczności)" należy rozumieć jako moment, w którym osoby fizyczne pragną zweryfikować pochodzenie konkretnego elementu treści i uzyskać dostęp do rozwiązania w zakresie wykrywania, aby uzyskać wyniki wykrywania.

(78) Jeżeli dostawca systemu AI jest zależny od innego podmiotu zapewniającego rozwiązanie w zakresie wykrywania lub dostęp do niego, musi zapewnić, aby rozwiązanie to było zgodne z art. 50 ust. 2 i 5 aktu w sprawie AI oraz aby te inne podmioty wyświetlały wyniki wykrywania zgodnie z art. 50 ust. 5 aktu w sprawie AI (zob. sekcja 7).

4.2.3. *Zgodność z wymogami dotyczącymi rozwiązań technicznych: kryteria skuteczności, interoperacyjności, solidności i niezawodności*

(79) Art. 50 ust. 2 zdanie drugie aktu w sprawie AI stanowi, że rozwiązania techniczne w zakresie oznaczania i wykrywania muszą być skuteczne, interoperacyjne, solidne i niezawodne. Wymogi te należy rozumieć w następujący sposób:

— **skuteczność** odnosi się do zdolności rozwiązania technicznego wdrożonego przez dostawców do wykrywania stosowanych przez nich oznakowań i umożliwiania osobom fizycznym odróżnienia treści sztucznie wygenerowanych lub zmanipulowanych przez system AI danego dostawcy, a tym samym przyczyniania się do zwiększenia zaufania do ekosystemu informacyjnego i jego integralności,

— **niezawodność** odnosi się do zdolności rozwiązania technicznego do dokładnego identyfikowania i odróżniania treści wygenerowanych lub zmanipulowanych przez AI od innych treści w warunkach nominalnych, w odniesieniu do różnorodnych treści wygenerowanych lub zmanipulowanych przez system AI dostawcy,

— **solidność** odnosi się do zdolności rozwiązania technicznego do dokładnego identyfikowania i odróżniania treści wygenerowanych lub zmanipulowanych przez AI pochodzących z systemu AI danego dostawcy od innych treści w zróżnicowanych warunkach, co obejmuje zarówno powszechne modyfikacje, jak i ataki kontradyktoryjne,

— **interoperacyjność** odnosi się do zdolności różnych rozwiązań technicznych w zakresie oznaczania i wykrywania do niezakłóconego działania w wielu systemach, kontekstach, między różnymi podmiotami oraz w ramach różnych wdrożeń technicznych w celu umożliwienia wykrywania treści wygenerowanych lub zmanipulowanych przez AI, niezależnie od techniki lub technik oznaczania stosowanych przez dostawców różnych systemów AI.

(80) Dostawcy muszą zapewnić, aby połączenie rozwiązań technicznych w zakresie oznaczania i wykrywania w sposób całościowy spełniało wszystkie wyżej wymienione wymogi w wymaganym prawem stopniu, z uwzględnieniem ograniczeń i komplementarności poszczególnych wdrożonych rozwiązań. Rozwiązania techniczne muszą spełniać wszystkie wymogi w zakresie, w jakim jest to technicznie wykonalne, uwzględniając przy tym specyfikę i ograniczenia różnych rodzajów treści, koszty wdrażania oraz powszechnie uznany stan wiedzy technicznej, czego odzwierciedleniem mogą być odpowiednie normy techniczne.

(81) Rozwiązania „technicznie wykonalne" w rozumieniu art. 50 ust. 2 aktu w sprawie AI to rozwiązania, które można wdrożyć w warunkach objętych zakresem stosowania tego przepisu, z wykorzystaniem obecnie dostępnych technologii, metod i praktyk inżynieryjnych, w ramach danej architektury technicznej i danego środowiska operacyjnego. Dostawca nie jest zobowiązany do stosowania rozwiązania technicznego, które nie zostało jeszcze opracowane lub nie jest dostępne na rynku lub którego wdrożenie jest technicznie niewykonalne. Wykonalność techniczna jest pojęciem obiektywnym, które nie zależy od konkretnych zasobów i zdolności poszczególnych dostawców.

(82) W art. 50 ust. 2 aktu w sprawie AI zobowiązano dostawców do wdrożenia rozwiązań technicznie wykonalnych, które są zgodne z powszechnie uznanym stanem wiedzy technicznej, w celu spełnienia czterech wymogów jakościowych dotyczących skuteczności, solidności, niezawodności i interoperacyjności.

(83) „Stan wiedzy technicznej" należy rozumieć jako rozwinięty stan zdolności technicznych w danym momencie w odniesieniu do produktów, procesów i usług, oparty na odpowiednich skonsolidowanych ustaleniach naukowych i technologicznych oraz na doświadczeniu i uznany za dobrą praktykę technologiczną ([28]). Stan wiedzy technicznej niekoniecznie musi oznaczać najnowsze badania naukowe znajdujące się nadal w fazie eksperymentalnej lub niewystarczająco dojrzałe technologicznie. Dostawcy muszą stale dostosowywać swoje rozwiązania w zakresie oznaczania i wykrywania w sposób terminowy i proporcjonalny w miarę rozwoju technologii i stanu wiedzy technicznej.

(84) W świetle art. 50 ust. 7 aktu w sprawie AI dostawcy mogą opierać się na kodeksie praktyk uznanym za odpowiedni zgodnie z tym przepisem w celu wykazania zgodności z obowiązkami w zakresie wykrywania i oznaczania, o których mowa w art. 50 ust. 2 i 5 aktu w sprawie AI, który to kodeks jest zgodny ze stanem wiedzy technicznej. Więcej informacji na temat skutków przestrzegania takiego kodeksu oraz nadzoru nad innymi podmiotami stosującymi, które go nie przyjęły, znajduje się w sekcji 8.1 poniżej.

(85) Można również uwzględnić koszty wdrożenia niektórych rozwiązań technicznych w zakresie oznaczania i wykrywania. W niektórych przypadkach, jak opisano w poniższych punktach, koszty mogą być niewspółmierne do niewielkich korzyści o ograniczonej wartości pod względem umożliwienia ludziom odróżnienia treści wygenerowanych lub zmanipulowanych przez AI oraz zapewnienia integralność ekosystemu informacyjnego i zaufania do niego.

(86) Mogą zatem istnieć wąsko zdefiniowane przypadki, w których rozwiązanie techniczne oparte na mniej solidnych oznaczeniach metadanych będą wystarczające do celów zgodności z art. 50 ust. 2 aktu w sprawie AI, z uwzględnieniem specyfiki wyników systemu AI, jeżeli z natury wiążą się one z niższym ryzykiem wprowadzenia w błąd, manipulacji lub negatywnego wpływu na ekosystem informacyjny. Dotyczy to w szczególności przypadków, w których system generatywnej AI jest wbudowany w produkty fizyczne generujące wyniki w technicznie kontrolowanym i zamkniętym środowisku, które ma głównie informacyjny charakter (np. system AI wbudowany w systemy nawigacyjne w pojazdach), w zakresie, w jakim w produkcie zintegrowano skuteczne środki techniczne zapobiegające wydostaniu się wyników poza środowisko produktu, na przykład poprzez ich przechwytywanie, eksportowanie i rozpowszechnianie na zewnątrz.

([28]) Zob. załącznik II do decyzji wykonawczej Komisji C(2025) 3871 final w sprawie wniosku o normalizację do Europejskiego Komitetu Normalizacyjnego oraz do Europejskiego Komitetu Normalizacyjnego Elektrotechniki w odniesieniu do systemu AI wysokiego ryzyka na potrzeby rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 oraz uchylającej decyzję wykonawczą C(2023) 3215 final.

(87) Mogą również występować ograniczone przypadki „zastosowań przemysłowych AI" lub „zastosowań między przedsiębiorstwami" (z wyłączeniem publicznych systemów AI i systemów AI skierowanych do konsumentów), w których zasada proporcjonalności i ograniczone korzyści dla celów przejrzystości oznaczają, że nie jest wymagane oznakowanie i wykrywanie wyników wygenerowanych lub zmanipulowanych przez AI. Ma to zastosowanie tylko wtedy, gdy łącznie spełnione są następujące wymogi:

(i) wynik wygenerowany przez system AI ma charakter ściśle techniczny, na przykład projekty inżynieryjne, procesy produkcji przemysłowej, instrukcje techniczne, wynik wygenerowany w wyniku procesów predykcyjnej konserwacji systemu w warunkach przemysłowych, wewnętrznych procesów dokumentacji lub etapów i procesów produkcji przed sfinalizowaniem wyniku i udostępnieniem go innym osobom zewnętrznym lub opinii publicznej;

(ii) wynik wygenerowany przez system AI ma być dostrzegalny i przetwarzany wyłącznie przez ograniczoną z góry określoną liczbę osób fizycznych działających w ramach swoich obowiązków zawodowych w organizacji dostawcy i podmiotu stosującego;

(iii) wynik wygenerowany przez system AI nie może być przeznaczony do udostępniania poza przedsiębiorstwem ani do wykorzystania przez osoby zewnętrzne, przy zastosowaniu odpowiednich zabezpieczeń w celu uniknięcia racjonalnie przewidywalnego niewłaściwego wykorzystania (np. izolacja chmury, kontrole oparte na rolach).

(88) Generowanie treści w czasie rzeczywistym, które mają charakter ulotny i są natychmiast wykorzystywane, bez dalszej rejestracji, przechowywania lub rozpowszechniania (np. w grach wideo, zastosowaniach w zakresie rzeczywistości wirtualnej), może również zostać wyłączone, jeżeli oznakowanie nie jest technicznie wykonalne, a osoby mające styczność z treściami są informowane o tym, że treści te zostały wygenerowane lub zmanipulowane przez AI (np. informowanie użytkownika bezpośrednio w środowisku operacyjnym, powiadomienia na poziomie sesji).

4.3. **Wyjątki od obowiązków wynikających z art. 50 ust. 2 aktu w sprawie AI**

(89) W art. 50 ust. 2 aktu w sprawie AI przewidziano trzy wyraźne wyjątki od obowiązków w zakresie przejrzystości określonych w tym przepisie.

(90) Pierwszy wyjątek dotyczy systemów AI w zakresie, w jakim pełnią one funkcję wspomagającą w zakresie standardowej edycji. Standardową edycję należy rozumieć jako proces przygotowywania istniejących treści do publikacji lub rozpowszechniania (np. drobne edycje mające na celu poprawę czytelności i gramatyki, jakości i formatu), który nie obejmuje tworzenia nowych treści. Standardowa edycja ma na celu zapewnienie, aby treści były m.in. wolne od oczywistych błędów technicznych lub gramatycznych, dostosowane do mających zastosowanie wymogów dotyczących układu graficznego, prezentacji, formatowania lub dostępności lub zgodne z praktykami sektorowymi. Edycja wykracza poza standardową edycję, jeżeli treść zostaje zmieniona w istotny sposób (zmiany merytoryczne, zmiany strukturalne itp.), co wpływa na jej znaczenie, styl lub zamierzony przekaz.

(91) Drugi wyjątek ma zastosowanie w przypadku, gdy system AI nie zmienia w istotny sposób danych wejściowych przekazywanych przez podmiot stosujący ani ich semantyki. Zmianę należy uznać za istotną, jeżeli dane wejściowe lub ich semantyka zostały w znacznym stopniu zmanipulowane przez system AI podczas procesu generowania jego wyników. Musi to opierać się na ocenie istotnych czynników, takich jak format, rodzaj treści medialnych, styl i zmiany w treści, co wpływa na jej znaczenie, styl lub zamierzony przekaz. Wymaga to każdorazowo indywidualnej oceny konkretnego przypadku.

(92) Jeżeli system AI może być wykorzystywany zarówno do generowania treści lub manipulowania treścią, jak i do nieznacznych i drobnych zmian danych wejściowych, obowiązek w zakresie przejrzystości określony w art. 50 ust. 2 aktu w sprawie AI nie będzie miał zastosowania do treści zmienionych w sposób nieznaczny i nieistotny.

Przykłady standardowej edycji i drobnych zmian objętych wyjątkiem przewidzianym w art. 50 ust. 2 aktu w sprawie AI:

— korekta gramatyczna i sprawdzanie pisowni, korekta językowa i drobne poprawki stylistyczne, które nie zmieniają istoty, znaczenia, stylu lub przekazu tekstu, tłumaczenia wygenerowane przez AI,

— formatowanie, konwersje formatu, kompresja techniczna, redukcja lub usuwanie zakłóceń dla uzyskania lepszej wyrazistości bez zmiany znaczenia lub istoty treści,

— drobne kadrowanie, drobna regulacja lub korekta kolorów, rozjaśnienie lub przyciemnienie, zaostrzenie w celu zwiększenia wyrazistości lub inne standardowe korekty redakcyjne i techniczne,

— usunięcie śladów kurzu spowodowanych zabrudzeniem soczewek lub czujników, usunięcie efektu czerwonych oczu wywołanego użyciem lampy błyskowej; – usunięcie i zasłonięcie tła widocznego w oryginalnym pliku, pikselowanie lub zamazywanie twarzy,

— zmiana skali klipu wideo, kompresja zakresu dynamicznego i korekcja dźwięku; ograniczona stabilizacja wideo,

— standardowe dostosowania koloru i kontrastu, niewielkie dostosowania szybkości odtwarzania, drobne korekty poziomu horyzontu obrazu, stosowanie filtrów pikseli w celu wzmocnienia niektórych części obrazu lub stosowanie kolorowych map do obrazów w skali szarości, uzupełnianie obrazu na obrzeżach, konwersja obrazu lub wideo z czarno-białego na kolorowe i odwrotnie,

— wypełnianie pikseli w celu dostosowania proporcji obrazu lub zwiększenia przejrzystości, automatyczne przejścia wideo, inne nieistotne modyfikacje do celów kosmetycznych i technicznych,

— techniczne przetwarzanie lub rekonstrukcja obrazów medycznych w wyrobach medycznych (np. adnotacje, ramki, kontury, okręgi, mapy cieplne),

— transkrypcja rozmów; treści wygenerowane przez AI, które jedynie przekształcają autentyczne dane wejściowe wprowadzone przez człowieka za pomocą technologii wspomagających, umożliwiając osobom z niepełnosprawnościami komunikację (np. komunikacja wspomagająca i alternatywna (AAC) lub spersonalizowane głosy neuronowe (CNV)), ponieważ nie zmieniają znaczenia treści pod kątem semantycznym.

Przykłady zmian semantycznych, które wymagają oznakowania na podstawie art. 50 ust. 2 aktu w sprawie AI:

— streszczenia tekstu wygenerowane przez AI; parafrazowanie lub redagowanie tekstu, które zmienia styl, strukturę i znaczenie, wykraczając poza zwykłą gramatyczną i nieznaczną korektę stylistyczną,

— usuwanie, zastępowanie lub wstawianie obiektów lub osób w istniejących obrazach i nagraniach wideo w sposób zmieniający znaczenie i istotę treści; zamiana twarzy lub istotna modyfikacja twarzy,

— realistyczna synteza mowy konkretnej osoby lub generowanie realistycznych materiałów wideo przedstawiających zdarzenia, które nie miały miejsca; zmiana sylwetki lub koloru skóry osoby; ekstremalne zmiany rozjaśnienia, przyciemnienia, koloru i kontrastu, które zmieniają znaczenie, intencję i przekaz treści,

— tworzenie złożonych obrazów lub klipów wideo, które modyfikują sposób przedstawienia osób, obiektów, zdarzeń lub faktów, oraz wszelkie inne istotne zmiany treści.

(93) Trzeci wyjątek ma zastosowanie do systemu generatywnej AI, którego wykorzystanie jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców. Wyjaśnienia przedstawione w pkt 46–48 powyżej w odniesieniu do tego wyjątku mają zastosowanie również w tym przypadku.

4.4. **Wzajemne powiązania z innymi aktami prawnymi Unii**

(94) Obowiązki w zakresie oznaczania i wykrywania określone w art. 50 ust. 2 aktu w sprawie AI koncentrują się na sposobie, w jaki treść stworzono, i na jej sztucznym pochodzeniu, a nie na twórcy treści. Wszelkie rozwiązania w zakresie oznaczania i wykrywania stosowane przez dostawców systemów AI muszą zatem być zgodne z mającymi zastosowanie unijnymi przepisami o ochronie danych, w tym z zasadami i obowiązkami w zakresie ochrony danych, takimi jak obowiązek w zakresie uwzględniania ochrony danych w fazie projektowania i w zakresie domyślnej ochrony danych, minimalizacja danych oraz zasady bezpieczeństwa i poufności. Na przykład do celów oznaczania i wykrywania informacje dotyczące twórcy treści nie powinny być przetwarzane, a wszelkie inne przetwarzane dane osobowe powinny być usuwane, gdy tylko cel wykrywania zostanie osiągnięty, gdy należy zapewnić bezpieczeństwo i poufność itp.

(95) W motywie 136 aktu w sprawie AI podkreślono ponadto szczególne znaczenie obowiązków w zakresie przejrzystości wynikających z art. 50 ust. 1, 2 i 4 aktu w sprawie AI dla ułatwienia skutecznego wdrożenia aktu o usługach cyfrowych. Dotyczy to w szczególności spoczywającego na dostawcach bardzo dużych platform internetowych („VLOP") lub bardzo dużych wyszukiwarek internetowych („VLOSE") obowiązku starannego identyfikowania, analizowania, oceniania i ograniczania ryzyka systemowego, które może wynikać z rozpowszechniania treści sztucznie wygenerowanych lub zmanipulowanych, w szczególności ryzyka faktycznych lub przewidywalnych negatywnych skutków dla procesów demokratycznych, dyskursu obywatelskiego i procesów wyborczych, w tym poprzez dezinformację (art. 34 i 35 aktu o usługach cyfrowych). Oznaczenia nadające się do odczytu maszynowego mogą umożliwić tym dostawcom wykrywanie treści wygenerowanych lub zmanipulowanych przez AI oraz zapewnienie odpowiedniego oznakowania i ujawniania, zmniejszając w ten sposób ryzyko dezinformacji generowanej przez AI w ramach ich usług.

(96) Obowiązki w zakresie oznaczania i etykietowania treści wygenerowanych lub zmanipulowanych przez systemy AI na podstawie art. 50 ust. 2 i 4 aktu w sprawie AI pozostają bez uszczerbku dla określonego w art. 16 ust. 6 aktu o usługach cyfrowych obowiązku rozpatrywania przez dostawców usług hostingu zgłoszeń dotyczących nielegalnych treści otrzymanych na podstawie art. 16 ust. 1 aktu o usługach cyfrowych ([29]). W szczególności oznaczanie lub etykietowanie stosowane do treści wygenerowanych lub zmanipulowanych przez AI nie powinno wpływać na ocenę i decyzję w sprawie niezgodności z prawem konkretnych treści na podstawie innych ram regulacyjnych. Ocena ta powinna być dokonywana wyłącznie w odniesieniu do przepisów regulujących zgodność treści z prawem w myśl rzeczonych ram ([30]). Na przykład jeżeli oznaczony obraz typu deepfake zostanie oznakowany jako potencjalny materiał przedstawiający niegodziwe traktowanie dzieci w celach seksualnych (CSAM) lub naruszający prawo znaków towarowych lub prawo autorskie, zastosowanie oznaczenia nie ma wpływu na potencjalną niezgodność treści z prawem na podstawie tych ram regulacyjnych. Ocena ta powinna być dokonywana wyłącznie na podstawie odpowiednich mających zastosowanie przepisów (np. prawa karnego, przepisów dotyczących znaków towarowych lub praw autorskich).

(97) Natomiast oznakowanie nadające się do odczytu maszynowego wbudowane w treści zgodnie z art. 50 ust. 2 aktu w sprawie AI może pomóc dostawcom VLOP i VLOSE w wypełnianiu ich obowiązków wynikających z art. 34 i 35 aktu o usługach cyfrowych w zakresie identyfikacji, analizy i oceny ryzyka systemowego, które może wynikać z rozpowszechniania treści sztucznie wygenerowanych lub zmanipulowanych, w szczególności ryzyka faktycznych lub przewidywalnych negatywnych skutków dla procesów demokratycznych, dyskursu obywatelskiego i procesów wyborczych, w tym poprzez stosowanie dezinformacji ([31]), oraz w zakresie wprowadzania racjonalnych, proporcjonalnych i skutecznych środków zmniejszających to ryzyko, dostosowanych do konkretnego zidentyfikowanego ryzyka systemowego. Dostawcy VLOP i VLOSE mogą również kwalifikować się jako dostawcy systemów AI, jeżeli spełniają odpowiednie warunki określone w akcie w sprawie AI, jeżeli włączają systemy AI do swoich usług (zob. również sekcja 6.1.5).

(98) Dostawców platform internetowych, wyszukiwarek internetowych i innych odpowiednich podmiotów rozpowszechniających treści w łańcuchu wartości zachęca się w szczególności do zachowania i umożliwienia identyfikacji oznaczeń wdrożonych przez dostawców zgodnie z art. 50 ust. 2 aktu w sprawie AI, tak aby osoby fizyczne mające styczność z tymi treściami mogły zostać poinformowane o ich pochodzeniu.

5. **ART. 50 UST. 3 AKTU W SPRAWIE AI: SYSTEMY ROZPOZNAWANIA EMOCJI I SYSTEMY KATEGORYZACJI BIOMETRYCZNEJ**

(99) W art. 50 ust. 3 aktu w sprawie AI nałożono na podmioty stosujące systemy rozpoznawania emocji i systemy kategoryzacji biometrycznej obowiązek poinformowania osób fizycznych, wobec których systemy te są stosowane, o fakcie ich stosowania.

(100) W motywie 132 aktu w sprawie AI wyjaśniono, że celem tego obowiązku jest zapewnienie, aby osoby fizyczne (w tym osoby z niepełnosprawnościami) były świadome, że stosowane są wobec nich systemy rozpoznawania emocji i kategoryzacji biometrycznej, które mogą naruszać ich prywatność. Obowiązek ten ma zatem zastosowanie niezależnie od tego, czy osoby te mają styczność z takimi systemami w czasie rzeczywistym, czy też systemy te są stosowane ex post.

([29]) Zob. również motyw 11 aktu w sprawie AI.
([30]) Zob. również motyw 137 aktu w sprawie AI.
([31]) Motyw 136 aktu w sprawie AI.

5.1. **Główne elementy, pojęcia i związane z nimi obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 3 aktu w sprawie AI**

5.1.1. *Pojęcie systemu rozpoznawania emocji*

(101) W art. 3 pkt 39 aktu w sprawie AI „system rozpoznawania emocji" zdefiniowano jako „system AI służący do rozpoznawania emocji lub zamiarów osób fizycznych na podstawie danych biometrycznych tych osób, lub wyciągania wniosków odnośnie do tych emocji lub zamiarów". Równie istotne w kontekście art. 50 ust. 3 aktu w sprawie AI są szczegółowe wyjaśnienia tego pojęcia zawarte w wytycznych Komisji dotyczących klasyfikacji systemów AI wysokiego ryzyka, w tym przedstawione w nich praktyczne przykłady.

(102) Ponieważ wszystkie systemy rozpoznawania emocji klasyfikuje się jednocześnie jako systemy wysokiego ryzyka, to o ile stosowanie tych systemów nie jest zakazane na podstawie art. 5 ust. 1 lit. f) aktu w sprawie AI (co dotyczy miejsca pracy lub instytucji edukacyjnych), obowiązek w zakresie przejrzystości określony w art. 50 ust. 3 aktu w sprawie AI powinien mieć zastosowanie w połączeniu z pozostałymi zabezpieczeniami i wymogami mającymi zastosowanie do systemów AI wysokiego ryzyka.

5.1.2. *Pojęcie systemu kategoryzacji biometrycznej*

(103) W art. 3 pkt 40 aktu w sprawie AI „system kategoryzacji biometrycznej" zdefiniowano jako „system AI służący do przypisywania osób fizycznych do określonych kategorii na podstawie danych biometrycznych tych osób, oprócz przypadków, gdy taki system pełni funkcję pomocniczą w stosunku do innej usługi komercyjnej i jest bezwzględnie konieczny z obiektywnych względów technicznych". W kontekście art. 50 ust. 3 aktu w sprawie AI równie istotne są szczegółowe wyjaśnienia tego pojęcia zawarte w wytycznych Komisji dotyczących klasyfikacji systemów AI wysokiego ryzyka, w tym jego konkretne elementy, a także praktyczne przykłady określone w tych wytycznych.

(104) Co ważne, o ile nie jest to zakazane na podstawie art. 5 ust. 1 lit. g) aktu w sprawie AI, art. 50 ust. 3 aktu w sprawie AI ma zastosowanie do wszelkich systemów kategoryzacji biometrycznej (np. klasyfikacji według wieku lub płci na podstawie danych biometrycznych), niezależnie od tego, czy są one objęte zakresem systemów AI wysokiego ryzyka na podstawie aktu w sprawie AI.

5.1.3. *Obowiązek informacyjny wynikający z art. 50 ust. 3 aktu w sprawie AI*

(105) **Zakres informacji, które należy przekazać:** podmioty stosujące systemy rozpoznawania emocji lub systemy kategoryzacji biometrycznej są zobowiązane do poinformowania zainteresowanych osób fizycznych, wobec których systemy te są stosowane, o fakcie ich stosowania. Akt w sprawie AI nie wymaga podawania informacji o powodach stosowania systemu ani o tym, czy system przetwarza również informacje do celów innych niż rozpoznawanie emocji lub kategoryzacja biometryczna, które to informacje muszą być ujawniane na mocy unijnych przepisów o ochronie danych (zob. sekcja 5.3).

(106) **Adresaci informacji:** w art. 50 ust. 3 aktu w sprawie AI nałożono wymóg, aby podmioty stosujące system AI informowały o tym fakcie wszystkie osoby fizyczne, wobec których system AI jest stosowany, w tym dzieci.

(107) **Sposoby przekazywania informacji:** w akcie w sprawie AI nie określono konkretnego sposobu przekazywania informacji, z tym wyjątkiem, że art. 50 ust. 5 aktu w sprawie AI wymaga, aby informacje były przekazywane w jasny i wyraźny sposób oraz zgodnie z mającymi zastosowanie wymogami dostępności (odpowiadającymi potrzebom osób fizycznych z niepełnosprawnościami). Sposób informowania osób fizycznych, wobec których system jest stosowany, może zależeć od miejsca wdrożenia (np. wirtualna platforma gamingowa, sklep stacjonarny, stacja kolejowa), możliwych adresatów (np. dzieci, osoby starsze, osoby z niepełnosprawnościami, klienci), czasu trwania i rodzaju stosowania (np. długotrwałe, jednorazowe, ciągłe, przerywane) oraz ewentualnego istnienia związku między podmiotem stosującym a osobami, wobec których system jest stosowany (np. jeżeli istnieje kanał komunikacji z osobą fizyczną). W zależności od przypadku użycia informacje mogą być przekazywane na piśmie, za pomocą standardowych ikon (również w przypadku prezentacji elektronicznej), ustnie lub za pomocą kombinacji takich sposobów.

Przykładowe sposoby informowania osób fizycznych:

— centralnie umieszczone okno wyskakujące z komunikatem powitalnym wyświetlanym przed rozpoczęciem gry komputerowej, informującym o nagrywaniu twarzy gracza w celu rejestracji jego emocji,

— widoczna informacja przy każdym wejściu do sali wystawowej, zawierająca komunikat, że przy wejściu do pomieszczenia rejestrowany jest wizerunek twarzy zwiedzających w celu przypisania ich do określonej grupy wiekowej.

(108) **Termin przekazania informacji**: informacje muszą zostać przekazane najpóźniej w momencie pierwszej interakcji z systemem AI lub pierwszego stosowania systemu AI zgodnie z art. 50 ust. 5 aktu w sprawie AI. W stosownych przypadkach nie wyklucza się wcześniejszego przekazania wymaganych informacji.

5.2. Wyjątek dotyczący zastosowań do celów ścigania przestępstw

(109) Przedmiotowy obowiązek informacyjny nie ma zastosowania do systemów rozpoznawania emocji i systemów kategoryzacji biometrycznej, których wykorzystywanie jest dozwolone z mocy prawa do celów wykrywania przestępstw, przeciwdziałania im i prowadzenia postępowań przygotowawczych w ich sprawie, z zastrzeżeniem odpowiednich zabezpieczeń w zakresie praw i wolności osób trzecich oraz zgodnie z prawem Unii. Wyjaśnienia przedstawione w pkt 46–48 powyżej w odniesieniu do tego wyjątku mają zastosowanie również w tym przypadku z jedną różnicą. Podczas gdy pozostałe wyjątki określone w art. 50 aktu w sprawie AI wymagają wyraźnego przepisu prawa zezwalającego na stosowanie systemów AI bez konieczności zachowania zgodności z wymogami dotyczącymi przejrzystości, art. 50 ust. 3 aktu w sprawie AI umożliwia takie stosowanie, jeżeli jest ono dozwolone na mocy przepisów prawnych regulujących uprawnienia organów ścigania do wykorzystywania systemów rozpoznawania emocji i kategoryzacji biometrycznej do wykrywania przestępstw, przeciwdziałania im lub prowadzenia postępowań przygotowawczych w ich sprawie bez informowania zainteresowanych osób, z zastrzeżeniem odpowiednich zabezpieczeń.

5.3. Wzajemne powiązania z innymi aktami prawnymi Unii

(110) Obowiązek informacyjny wynikający z art. 50 ust. 3 aktu w sprawie AI sam w sobie nie sprawia, że wykorzystywanie systemu rozpoznawania emocji i kategoryzacji biometrycznej jest zgodne z prawem. Nie uzasadnia on również inwazyjnych lub dyskryminujących zastosowań, które mogą być niezgodne z prawem i zakazane na mocy art. 5 aktu w sprawie AI lub innych przepisów prawa Unii. Podmioty stosujące muszą przestrzegać obowiązku informacyjnego określonego w art. 50 ust. 3 aktu w sprawie AI, w dodatku do wszelkich mających zastosowanie wymogów informacyjnych wynikających z unijnych przepisów o ochronie danych. W niektórych przypadkach podmioty stosujące mogą rozważyć dodanie powiadomień wymaganych na podstawie art. 50 ust. 3 aktu w sprawie AI do informacji przekazywanych osobom, których dane dotyczą, na podstawie unijnych przepisów o ochronie danych.

6. ART. 50 UST. 4: OZNAKOWANIE MATERIAŁÓW TYPU DEEPFAKE I NIEKTÓRYCH PUBLIKACJI TEKSTOWYCH

(111) W art. 50 ust. 4 aktu w sprawie AI ustanowiono dwa odrębne obowiązki dla podmiotów stosujących systemy generatywnej AI, dotyczące informowania w jasny i wyraźny sposób o (i) materiałach typu deepfake oraz (ii) tekście wygenerowanym lub zmanipulowanym przez AI, publikowanym w celu informowania opinii publicznej o sprawach leżących w interesie publicznym, z wyjątkiem określonych przypadków. Te obowiązki w zakresie oznaczania mają zastosowanie dodatkowo i bez uszczerbku dla obowiązków w zakresie oznaczania nadającego się do odczytu maszynowego i wykrywania na podstawie art. 50 ust. 2 aktu w sprawie AI mających zastosowanie do dostawców systemów AI generujących treści w postaci syntetycznych obrazów, wideo, dźwięków lub tekstu lub manipulujących nimi.

6.1. Główne elementy i pojęcia dotyczące materiałów typu deepfake i związane z nimi obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 4 aktu w sprawie AI

(112) Art. 50 ust. 4 akapit pierwszy aktu w sprawie AI ma zastosowanie, jeżeli spełnionych jest kilka warunków:

(i) system musi być systemem AI;

(ii) system musi być wykorzystywany przez podmioty stosujące do celów zawodowych (zob. sekcje 2.3 i 2.4.1);

(iii) system musi generować obrazy, treści dźwiękowe lub treści wideo, które stanowią materiały typu deepfake, lub manipulować nimi;

(iv) system AI nie może być objęty wyjątkiem dotyczącym wykorzystywania dozwolonego na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców.

6.1.1. *Pojęcie „deepfake”*

(113) W art. 3 pkt 60 aktu w sprawie AI „deepfake” zdefiniowano jako „wygenerowane przez AI lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia, które odbiorca mógłby niesłusznie [mylnie] uznać za autentyczne lub prawdziwe”. Różnicę między manipulacją a generowaniem wyjaśniono w sekcji 4.1 powyżej. W celu doprecyzowania pojęcia deepfake poniżej przeanalizowano cztery łączne kryteria zawarte w tej definicji: (i) przypominanie (ii) istniejących (iii) osób, przedmiotów, miejsc, podmiotów lub zdarzeń (iv), które odbiorca mógłby niesłusznie (mylnie) uznać za autentyczne lub prawdziwe.

(i) **Przypominanie:** art. 3 pkt 60 aktu w sprawie AI stanowi, że wygenerowane lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo muszą „przypominać” istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia, aby stanowiły materiały typu deepfake. W motywie 134 aktu w sprawie AI dodano, że takie podobieństwo powinno być „łudzące”. Wygenerowane lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo łudząco przypominają osoby, przedmioty, miejsca, podmioty lub zdarzenia, jeżeli istnieje wysoki poziom podobieństwa między treściami typu deepfake a daną osobą, danym przedmiotem, miejscem, podmiotem lub zdarzeniem (w tym którymkolwiek z ich rozpoznawalnych elementów) symulowanymi przez treści typu deepfake. Treść nie musi być identyczna z osobą, przedmiotem, miejscem, podmiotem lub zdarzeniem. Oceny, czy poziom podobieństwa jest łudzący, dokonuje – indywidualnie dla każdego przypadku – podmiot stosujący na podstawie obiektywnego porównania symulowanej osoby lub symulowanego przedmiotu, miejsca, podmiotu lub zdarzenia z treściami typu deepfake, w oparciu m.in. o zakres, w jakim cechy charakterystyczne lub wyróżniające są reprezentowane przez treści typu deepfake.

(ii) **Istniejące:** art. 3 pkt 60 aktu w sprawie AI stanowi, że wygenerowane lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo muszą przypominać „istniejące” osoby, przedmioty, miejsca, podmioty lub zdarzenia, aby stanowiły materiały typu deepfake. Aby skutecznie osiągnąć cel art. 50 ust. 4 aktu w sprawie AI, jakim jest ograniczenie ryzyka podawania się za inną osobę, wprowadzania w błąd, dezinformacji, manipulacji i oszustw, kryterium „istnienia” oznacza, że wygenerowane lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo powinny przypominać realistyczne osoby, przedmioty, miejsca, podmioty lub zdarzenia. W związku z tym wystarczy, aby symulowane osoby, przedmioty, miejsca, podmioty lub zdarzenia przypominały kogoś lub coś, co istnieje, może istnieć lub mogłoby istnieć w rzeczywistości, aby można je było uznać za materiały typu deepfake, ponieważ takie treści mogą być głównym źródłem wprowadzania w błąd, manipulacji i dezinformacji. Natomiast symulowane osoby, przedmioty, miejsca, podmioty lub zdarzenia, które na przykład zaprzeczają prawom przyrody lub fizyki lub przedstawiają formy życia, które nie są powszechnie akceptowane w biologii (takie jak ludzie latający bez pomocy urządzeń mechanicznych, smoki lub słonie kierujące pojazdami) i nie mogą wprowadzać w błąd, uznaje się za nierealistyczne i w związku z tym nie wchodzą one w zakres obowiązku dotyczącego przejrzystości.

(iii) **Osoby, przedmioty, miejsca, podmioty lub zdarzenia:** art. 3 pkt 60 aktu w sprawie AI stanowi, że wygenerowane lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo powinny przypominać istniejące „osoby, przedmioty, miejsca, podmioty lub zdarzenia”, aby stanowiły materiały typu deepfake. „Osoby” należy rozumieć jako realistyczne istoty ludzkie (w tym cyfrowe repliki prawdziwych osób, realistyczne awatary ludzkie lub osoby wygenerowane przez AI oraz cechy osobiste lub formy ekspresji, takie jak obraz, głos, zachowanie, kreacje itp.). „Przedmioty” należy rozumieć jako realistyczne, nieożywione przedmioty materialne, w tym budynki, dzieła sztuki, maszyny, towary konsumpcyjne itp. „Miejsca” należy rozumieć jako realistyczne lokalizacje. „Podmioty” należy rozumieć jako realistyczne istoty inne niż ludzkie, lecz ożywione, w tym zwierzęta lub inne biologiczne formy życia. „Zdarzenia” należy rozumieć jako realistyczne sceny lub sytuacje, w które mogą być zaangażowane osoby, przedmioty, miejsca i podmioty (np. przywoływanie zdarzeń historycznych lub przedstawianie usług profesjonalnych lub konsumenckich).

(iv) **Wywoływanie u człowieka mylnego wrażenia autentyczności lub prawdziwości:** jeżeli chodzi o czwarte kryterium, art. 3 pkt 60 aktu w sprawie AI stanowi, że odbiorca może mylnie uznać wygenerowane przez AI lub zmanipulowane przez AI obrazy, treści dźwiękowe lub treści wideo, które przypominają istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia, „za autentyczne lub prawdziwe”. Kryterium to odnosi się do zasadniczej cechy treści deepfake i jej zdolności do potencjalnego wprowadzenia w błąd lub zmylenia osoby co do autentyczności lub prawdziwości treści ([32]). Autentyczność treści odnosi się do tego, czy treść jest rzeczywiście tym, czym wydaje się być pod względem jej źródła lub procesu tworzenia (w tym np. zaangażowania prawdziwych ludzi lub zwierząt oraz ich faktycznego zachowania lub działania, faktycznego wyglądu lub wykorzystania przedmiotów lub świadczenia usług, dokładnego rozwoju lub przebiegu zdarzenia). Prawdziwość odnosi się do wiarygodności treści (np. poprawności merytorycznej elementów przedstawionych w materiałach typu deepfake).

([32]) W związku z tym kryterium to należy rozumieć niezależnie i odrębnie od pojęcia wprowadzenia w błąd, o którym mowa w dyrektywie 2005/29/WE dotyczącej nieuczciwych praktyk handlowych stosowanych przez przedsiębiorstwa wobec konsumentów na rynku wewnętrznym (dyrektywa o nieuczciwych praktykach handlowych), https://eur-lex.europa.eu/eli/dir/2005/29/oj/pol.

(114) To, czy „odbiorca mógłby niesłusznie [mylnie] uznać [treści] za autentyczne lub prawdziwe" należy oceniać jako całość, biorąc pod uwagę poziom podobieństwa, potencjalny merytoryczny przekaz treści, zamierzone i przewidywalne konteksty wdrażania, środowisko, w którym treści są prezentowane, oraz zamierzony i racjonalnie przewidywalny skład grupy docelowej i oczekiwania odbiorców. Ocena ta jest jednak obiektywna i nie uzależnia klasyfikacji treści jako deepfake od tego, czy zamiarem podmiotu stosującego jest wprowadzenie w błąd lub zmylenie osób fizycznych mających styczność z takimi treściami. Jeżeli chodzi o kontekst wdrażania i oczekiwania odbiorców, oznacza to, że jeżeli treści są wykorzystywane w konkretnych kontekstach wdrażania, w których odbiorcy, do których w zamierzeniu kierowane są treści i których można racjonalnie przewidzieć, nie oczekują, że treści będą autentyczne lub prawdziwe, nie można uznać, że treści wygenerowane lub zmanipulowane przez AI *mylnie* wydają się autentyczne lub prawdziwe, nawet jeżeli w rzeczywistości są nieautentyczne lub nieprawdziwe. W takich przypadkach czwarte kryterium definicji deepfake nie zostanie spełnione. Na przykład generowanie przez AI scen w tle lub manipulowanie nimi przez AI, stosowanie efektów specjalnych, a także techniczna obróbka wstępna i końcowa w ramach standardowych procesów produkcji filmów prawdopodobnie nie sprawią, że treści będą *mylnie* postrzegane przez odbiorców jako autentyczne lub prawdziwe. Z drugiej strony wykorzystanie AI do generowania istotnych elementów, które mają wpływ na postrzeganie przez odbiorców, lub do manipulowania tymi elementami, prawdopodobnie to sprawi (np. aktorzy w pełni wygenerowani przez AI, cyfrowe repliki aktorów żyjących lub zmarłych, cyfrowe odmładzanie oraz symulowane kreacje aktorskie, w tym ich fragmenty, lub nieautentyczne lub nieprawdziwe przedstawienie osób, przedmiotów, miejsc lub zdarzeń w filmach dokumentalnych). W odniesieniu do obrazów i treści wideo uznaje się, że wysoki stopień fotorealizmu zwiększa prawdopodobieństwo uznania takich treści za deepfake (ponieważ prawdopodobnie przypominają one istniejące osoby, przedmioty, miejsca, podmioty lub zdarzenia), ale sam fotorealizm nie jest decydujący dla oceny. Należy jeszcze ocenić, czy takie treści są w stanie wprowadzić osobę w błąd lub zmylić ją co do autentyczności lub prawdziwości treści, biorąc pod uwagę między innymi zamierzony i przewidywalny kontekst wdrożenia.

(115) W przeciwieństwie do art. 50 ust. 1 aktu w sprawie AI ocena mylnego wrażenia autentyczności lub prawdziwości w odbiorze człowieka nie powinna opierać się na hipotetycznej „przeciętnej" osobie, która może mieć styczność z danymi treściami (zob. sekcja 3.2.1), ale powinna należycie uwzględniać możliwy zróżnicowany skład racjonalnie przewidywalnej grupy odbiorców, którzy mogą mieć styczność z treściami typu deepfake. Należyte uwzględnienie składu grupy odbiorców jest szczególnie ważne, jeżeli można racjonalnie przewidzieć, że treści typu deepfake mogą w pewnym momencie trafić do dzieci, osób starszych lub innych grup osób o niższych umiejętnościach cyfrowych i kompetencjach w zakresie AI lub o niższym poziomie wiedzy ogólnej, ponieważ osoby te można łatwiej zmylić lub wprowadzić w błąd co do autentyczności lub prawdziwości treści. W takich przypadkach wywoływanie u tej części przewidywalnej grupy odbiorców mylnego wrażenia autentyczności lub prawdziwości treści może wystarczyć do uznania jej za deepfake. Od podmiotów stosujących nie należy jednak oczekiwać, aby brały pod uwagę potencjalne dalsze rozpowszechnianie treści typu deepfake przez osoby trzecie w kręgach wykraczających poza grupę odbiorców, którą podmiot stosujący jest w stanie racjonalnie przewidzieć, uwzględniając zamierzone kanały dystrybucji (np. jeżeli treści typu deepfake są wyświetlane wyłącznie w części strony internetowej przeznaczonej dla subskrybentów lub w newsletterze firmowym, nie oznacza to, że podmiot stosujący powinien domyślnie przyjąć, że treści będą szeroko rozpowszechniane publicznie).

(116) Podobnie wspierana przez AI manipulacja nieistotnymi merytorycznymi lub technicznymi aspektami wcześniej istniejących treści może mieć niewielkie znaczenie dla dokonywanej przez odbiorcę oceny autentyczności lub prawdziwości treści, nie sprawiając, że powstałe w ten sposób treści staną się deepfake. Może to obejmować na przykład edytowanie szczegółów dotyczących tła (np. usunięcie przechodniów), regulację oświetlenia, dostosowanie parametrów dźwięku, korektę kolorów, redukcję lub usunięcie hałasu, poprawę dostępności lub kompresję plików, korekty kosmetyczne i ulepszenia. To, czy manipulacja będzie należeć do tej kategorii, zależy od kontekstu i wpływu na postrzeganie autentyczności lub prawdziwości treści przez odbiorców w danym przypadku. Na przykład wspomagana AI korekta kolorów, rozszerzenie istniejących treści na tło, dostosowanie lub zmiana tła do celów stricte estetycznych, kompozycje i aranżacje istniejących produktów lub zmiana skali obrazów stosowane w reklamach produktów lub na ich opakowaniach mogą mieć jedynie niewielki wpływ na postrzeganie przez daną osobę autentyczności i prawdziwości reklamy i produktu. Z drugiej strony szeroko zakrojona, oparta na AI edycja podstawowych szczegółów obrazów dziennikarskich poza standardowymi praktykami technicznymi i redakcyjnymi może negatywnie wpłynąć na postrzeganie przez daną osobę autentyczności i prawdziwości treści.

Przykłady materiałów typu deepfake w rozumieniu art. 3 pkt 60 aktu w sprawie AI:

— zmanipulowany przez AI obraz przedstawiający scenę z udziałem dwóch prawdziwych, profesjonalnych piłkarzy stojących przed budynkiem przypominającym stadion piłkarski,

— wygenerowane przez AI nagranie audio ze sklonowanymi głosami stałych prezenterów podcastu prasowego oraz zaproszonych rozmówców, dyskutujących o bieżących wydarzeniach,

— wygenerowane przez AI nagranie wideo przedstawiające osobę przypominającą polityka, wygłaszającą przemówienie przed publicznością,

— wygenerowane przez AI nagranie wideo przedstawiające wygenerowany przez AI wizerunek znanego influencera w kontekście reklamowym lub promocyjnym,

— wygenerowane przez AI nagranie wideo przedstawiające realistycznego syntetycznego awatara prezesa przedsiębiorstwa, który składa podziękowania pracownikom za ich pracę i wyniki przedsiębiorstwa w ubiegłym roku,

— wygenerowany przez AI obraz produktu w reklamie lub na opakowaniu, który może wpływać na postrzeganie go przez odbiorców i wprowadzać w błąd co do rzeczywistego wyglądu, cech lub zastosowania produktu (np. przedstawiający produkt w sposób odbiegający od jego rzeczywistego wyglądu – jako bardziej atrakcyjny lub o wyższej jakości niż w rzeczywistości).

Przykłady, które nie stanowią materiałów typu deepfake w rozumieniu art. 3 pkt 60 aktu w sprawie AI:

— wygenerowany przez AI obraz przedstawiający sfinksa lecącego nad Wieżą Eiffla,

— wygenerowane przez AI nagranie wideo przedstawiające myszy dyskutujące w ludzkim języku o najlepszym gatunku sera jako element kampanii reklamowej producenta serów,

— klonowanie głosu przez AI na potrzeby postaci fikcyjnych (np. w audiobookach, grach lub animacjach), gdy odbiorca nie jest wprowadzany w błąd co do tożsamości narratorów,

— zmanipulowana przez AI audycja radiowa dostosowująca techniczne parametry dźwięku (np. normalizująca poziomy głośności, redukująca szumy, kompresująca dźwięk) bez zmiany rzeczywistych słów wypowiadanych przez rozmówców lub sposobu ich wypowiedzi,

— wygenerowany przez AI rysunek na podstawie istniejącego obrazu wydarzenia historycznego,

— materiał wideo przedstawiający prezentera programu popularnonaukowego, który idzie przez teren górzysty i opowiada o topnieniu lodowców, co ilustruje zmodyfikowana przez AI animacja tła, przedstawiającą proces cofania się lodowców od ich historycznych poziomów do stanu obecnego,

— fikcyjne środowiska wygenerowane przez AI (np. fikcyjne lasy i zamki) w grach wideo,

— filmy, w których prawdziwi, realnie istniejący aktorzy (niezmanipulowani przez AI) grają na tle wygenerowanym przez AI (np. przedstawiającym fikcyjny obszar prerii, zabytkowe budynki starożytnych miast itp.),

— rzeczywisty produkt (np. samochód) pokazany w reklamie na tle i w otoczeniu wygenerowanym przez AI, o ile reklama ta nie wprowadza odbiorców w błąd co do rzeczywistego wyglądu produktu oraz jego cech charakterystycznych i zastosowania.

6.1.2. *Obowiązek ujawniania informacji na podstawie art. 50 ust. 4 akapit pierwszy aktu w sprawie AI*

(117) W art. 50 ust. 4 akapit pierwszy aktu w sprawie AI zobowiązano podmioty stosujące systemy AI, które generują obrazy, treści audio lub wideo stanowiące treści typu deepfake lub które manipulują takimi obrazami lub treściami, do ujawniania, że „treści te zostały sztucznie wygenerowane lub zmanipulowane". W motywie 134 aktu w sprawie

AI doprecyzowano, że podmioty stosujące wykorzystujące systemy AI do generowania treści typu deepfake lub do manipulowania nimi powinny jasno i wyraźnie ujawnić – poprzez odpowiednie oznakowanie wygenerowanych przez AI treści i ujawnienie ich sztucznego pochodzenia – że treści te zostały sztucznie wygenerowane lub zmanipulowane. Metody oznakowania lub ujawniania stosowane zgodnie z art. 50 ust. 4 akapit pierwszy aktu w sprawie AI powinny być zrozumiałe i dostrzegalne dla osób fizycznych (np. w formie widocznych lub słyszalnych oznaczeń), bez konieczności stosowania jakichkolwiek konkretnych narzędzi technicznych lub wykonywania specjalnych działań. W związku z tym podmioty stosujące nie mogą polegać na oznakowaniu nadającym się do odczytu maszynowego wbudowanym w treści przez dostawcę na podstawie art. 50 ust. 2 aktu w sprawie AI, ponieważ oznaczenia te nie są od razu jasne i wyraźne dla osób fizycznych, które mają styczność z treściami typu deepfake.

(118) W świetle art. 50 ust. 7 aktu w sprawie AI podmioty stosujące mogą przestrzegać kodeksu praktyk uznanego za odpowiedni zgodnie z art. 50 ust. 7 aktu w sprawie AI w celu wykazania zgodności z tym obowiązkiem ujawniania informacji. Więcej informacji na temat skutków przestrzegania takiego kodeksu oraz nadzoru nad innymi podmiotami stosującymi, które go nie przyjęły, znajduje się w sekcji 8.1 poniżej.

6.1.3. *Przejrzystość w odniesieniu do treści typu deepfake o charakterze artystycznym, twórczym, satyrycznym, fikcyjnym lub analogicznym*

(119) W art. 50 ust. 4 akapit pierwszy aktu w sprawie AI przewidziano złagodzony obowiązek w zakresie przejrzystości w odniesieniu do treści typu deepfake stanowiących część utworu lub programu mającego wyraźnie charakter artystyczny, twórczy, satyryczny, fikcyjny lub analogiczny. W przypadku takich treści obowiązek w zakresie przejrzystości ogranicza się do ujawnienia treści typu deepfake w odpowiedni sposób, który nie utrudnia wyświetlania utworu lub korzystania z niego.

(120) **Odpowiednie kategorie treści:** aby skorzystać ze złagodzonego obowiązku ujawniania informacji w celu zapewnienia przejrzystości, treści typu deepfake (obraz, treści dźwiękowe lub wideo) powinny należeć do przynajmniej jednej z następujących kategorii treści: utwory lub programy mające charakter artystyczny, twórczy, satyryczny, fikcyjny lub analogiczny. To, czy deepfake należy do przynajmniej jednej z tych kategorii, wymaga indywidualnej oceny przez podmiot stosujący. Do celów art. 50 ust. 4 aktu w sprawie AI kategorie te rozumie się w następujący sposób:

— przez utwory artystyczne należy rozumieć utwory stworzone na potrzeby sztuki, w tym muzykę, utwory kinematograficzne i sztuki wizualne,

— przez utwory twórcze należy rozumieć utwory, które wymagają twórczych wyborów. Utworów motywowanych głównie względami funkcjonalnymi lub technicznymi nie można uznać za utwory twórcze,

— przez utwory satyryczne należy rozumieć utwory, które mają na celu krytykowanie społeczeństwa, polityki, biznesu lub osób publicznych za pomocą technik humorystycznych (w tym ironii, sarkazmu, kpiny, pastiszu itp.),

— przez utwory fikcyjne należy rozumieć utwory kulturowe, które przedstawiają osoby, przedmioty, miejsca, podmioty lub zdarzenia w fikcyjnej, ale wiarygodnej scenerii,

— przez utwory analogiczne należy rozumieć utwory, które posiadają wspólne cechy podstawowe lub podobne cechy ekspresyjne lub funkcjonalne z powyższymi kategoriami, lecz nie wpisują się idealnie w żadną z nich.

(121) Powyższe kategorie treści mogą mieć również zastosowanie do programów, które należy rozumieć jako odrębną całość w układzie lub katalogu audycji przygotowanym przez dostawcę usług medialnych, która pod względem formy i treści jest porównywalna z rozpowszechnianiem telewizyjnym ([33]). Odniesienie do „utworów" w kontekście art. 50 ust. 4 aktu w sprawie AI nie przesądza o tym, czy treści kwalifikują się do ochrony na mocy prawa Unii dotyczącego praw autorskich i pokrewnych.

([33]) Zob. podobnie art. 1 lit. b) dyrektywy o audiowizualnych usługach medialnych.

(122) **Wyraźnie:** fakt, że deepfake należy do jednej (lub większej liczby) z powyższych kategorii treści, powinien być wyraźny dla osób fizycznych mających z nim styczność. Oznacza to, że kategorie te należy interpretować ściśle, co jest uzasadnione tym, że stosuje się do nich łagodniejszy system obowiązków informacyjnych ze względu na potencjalny wpływ na wolność wypowiedzi oraz wolność sztuki i nauki, natomiast nadal należy odpowiednio uwzględnić ryzyko dezinformacji, manipulacji i wprowadzania w błąd. W związku z tym treści, których charakter jest potencjalnie niejasny lub niejednoznaczny dla odbiorców, są wyłączone z zakresu tego łagodniejszego systemu obowiązków w zakresie przejrzystości. Czynniki istotne dla oceny wyraźnie artystycznego, twórczego, satyrycznego lub fikcyjnego charakteru obejmują: (i) czy treść przedstawia formaty lub style charakterystyczne dla kategorii treści (np. ironia lub wyolbrzymianie na potrzeby utworów satyrycznych, niektóre style sztuki, prezentowanie charakterystycznych motywów, technik lub estetyki itp.); (ii) kontekst, w jakim treści są prezentowane (np. czy platforma, nośnik lub miejsce, w których treści są prezentowane, udostępniane lub rozpowszechniane, są związane z wykorzystaniem artystycznym, twórczym, satyrycznym lub fikcyjnym); oraz (iii) oczekiwania odbiorców (np. film, środowisko gier lub scena rzeczywistości wirtualnej). Ponadto warunek, zgodnie z którym treści powinny „wyraźnie" należeć do jednej z powyższych kategorii treści, z zakresu wspomnianych kategorii treści wyklucza treści, które mają charakter wyłącznie informacyjny lub handlowy i są one rozpoznawalne jako takie (np. relacjonowanie wydarzeń). W tym względzie niektóre rodzaje treści (np. reklamy lub filmy dokumentalne) zawierające materiały typu deepfake można uznać za utwory o wyraźnie twórczym lub fikcyjnym charakterze w pewnych szczególnych sytuacjach, ale nie w innych, ponieważ oceny dokonuje się indywidualnie dla konkretnego przypadku. W przypadku gdy deepfake łączy w sobie wiele cech (np. charakter informacyjny i twórczy), charakter informacyjny powinien zawsze być kryterium nadrzędnym, oznaczając konieczność zapewnienia zgodności ze standardowymi wymogami dotyczącymi oznakowania.

(123) **Ujawnienie w odpowiedni sposób, który nie utrudnia wyświetlania lub korzystania z utworu:** treści typu deepfake, które stanowią część utworów lub programów o wyraźnie artystycznym, twórczym, satyrycznym, fikcyjnym lub analogicznym charakterze, nie są wyłączone z obowiązku w zakresie przejrzystości określonego w art. 50 ust. 4 akapit pierwszy aktu w sprawie AI. Podmioty stosujące nadal muszą ujawniać fakt wygenerowania lub zmanipulowania treści przez AI, ale mogą to zrobić w odpowiedni sposób, który nie utrudnia wyświetlania lub korzystania z utworu. W motywie 134 aktu w sprawie AI doprecyzowano, że takie odpowiednie ujawnienie nie powinno utrudniać normalnego wykorzystania i użytkowania, przy jednoczesnym zachowaniu użyteczności i jakości utworu. Określenie, które środki ujawniania informacji należy uznać za odpowiednie, wymaga oceny indywidualnej, przy czym podmioty stosujące mogą wziąć pod uwagę wszystkie istotne czynniki (w tym charakter utworów, odbiorców, kontekst itp.). W każdym przypadku podmioty stosujące muszą przestrzegać art. 50 ust. 5 aktu w sprawie AI (zob. sekcja 7). W świetle art. 50 ust. 7 aktu w sprawie AI podmioty stosujące mogą przestrzegać kodeksu praktyk uznanego za odpowiedni na podstawie tego przepisu w celu wykazania zgodności z tym obowiązkiem ujawniania informacji. Więcej informacji na temat skutków przestrzegania takiego kodeksu oraz nadzoru nad innymi podmiotami stosującymi, które go nie przyjęły, znajduje się w sekcji 8.1 poniżej.

(124) **Odpowiednie zabezpieczenia w zakresie praw i wolności osób trzecich:** w motywie 134 aktu w sprawie AI doprecyzowano, że stosowanie złagodzonego obowiązku w zakresie przejrzystości oraz wykorzystywanie treści artystycznych, twórczych, satyrycznych lub fikcyjnych typu deepfake musi podlegać „odpowiednim zabezpieczeniom w zakresie praw i wolności osób trzecich". W związku z tym podmioty stosujące muszą zapewnić odpowiednią ochronę i poszanowanie praw i wolności osób trzecich (w tym np. ich prawa do ochrony danych osobowych lub praw własności intelektualnej) przy tworzeniu lub publikowaniu treści typu deepfake, nawet jeżeli treści te mają charakter artystyczny, twórczy, satyryczny lub fikcyjny. Stosowanie złagodzonego obowiązku w zakresie przejrzystości nie może stanowić uzasadnienia dla nieprzestrzegania praw podstawowych osób fizycznych lub podmiotów praw wynikających z prawa Unii dotyczącego własności intelektualnej lub unijnych przepisów o ochronie danych ([34]).

Przykłady treści typu deepfake, które są utworami o charakterze twórczym, satyrycznym, fikcyjnym lub analogicznym:

— utwory artystyczne/fikcyjne: filmy (lub zwiastuny filmowe) zawierające wygenerowane lub zmanipulowane przez AI treści typu deepfake przedstawiające cyfrowo odmłodzonych, żyjących aktorów bądź cyfrowe repliki zmarłych aktorów lub innych osób, dostępne w kinach lub na platformach streamingowych,

— utwory artystyczne/twórcze: wygenerowana przez AI muzyka w dowolnym gatunku, naśladująca indywidualny styl żyjących artystów,

([34]) Dotyczy to również publikowanych tekstów informujących społeczeństwo o sprawach leżących w interesie publicznym.

— utwory satyryczne/fikcyjne: zmanipulowany przez AI obraz przedstawiający żyjącego polityka w sytuacji, która wyraźnie ma na celu humorystyczne skrytykowanie określonych decyzji politycznych podjętych lub popieranych przez tę osobę,

— analogiczne utwory twórcze/fikcyjne: wygenerowana przez AI sceneria gry z wykorzystaniem technologii deepfake do symulacji prawdziwych, żyjących osób.

Przykłady treści typu deepfake, które nie są utworami o charakterze artystycznym, twórczym, satyrycznym, fikcyjnym lub analogicznym:

— zmanipulowane przez AI nagranie wideo w stylu telezakupów, zawierające symulację deepfake osób reklamujących produkt w wygenerowanej przez AI scenerii; nagranie przedstawia praktyczne użycie produktu przez symulowanych konsumentów w celu zachęcenia widzów do zakupu produktu,

— wygenerowany przez AI obraz przedstawiający celebrytów, sugerujący ich udział w działaniach, które nigdy się nie wydarzyły, bez żadnego celu fikcyjnego, satyrycznego lub analogicznego,

— wygenerowane przez AI nagranie wideo przedstawiające realistycznego, syntetycznego influencera testującego sponsorowany rzeczywisty produkt, koncentrujące się wyłącznie na wyświetlaniu funkcji produktu,

— wygenerowane przez AI nagranie wideo przedstawiające realistyczne sceny z Holokaustu, udostępniane na ogólnodostępnych platformach mediów społecznościowych.

6.1.4. *Wyjątek na potrzeby ścigania przestępstw*

(125) Jeżeli wykorzystanie treści typu deepfake (np. wygenerowanego przez AI nagrania wideo symulującego zdarzenia, które nigdy nie miały miejsca) jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, podmioty stosujące są zwolnione z obowiązku w zakresie przejrzystości określonego w art. 50 ust. 4 akapit pierwszy aktu w sprawie AI. Wyjaśnienia przedstawione w pkt 46–48 powyżej w odniesieniu do podobnego wyjątku określonego w art. 50 ust. 1 aktu w sprawie AI mają zastosowanie również w odniesieniu do tego przepisu.

6.1.5. *Wzajemne powiązania z innymi aktami prawnymi Unii*

(126) Obowiązek w zakresie przejrzystości wynikający z art. 50 ust. 4 akapit pierwszy aktu w sprawie AI będzie powiązany z obowiązkami dostawców VLOP i VLOSE w zakresie identyfikacji i ograniczania ryzyka systemowego zgodnie z art. 34 i 35 aktu o usługach cyfrowych, w szczególności z art. 35 ust. 1 lit. k) aktu o usługach cyfrowych ([35]). W tym ostatnim przepisie wśród możliwych środków zmniejszających ryzyko, które dostawcy VLOP i VLOSE mogą przyjąć, wymieniono środki zapewniające, aby „informację – bez względu na fakt, czy stanowi ona wygenerowany lub zmanipulowany obraz, treść dźwiękową lub treść wideo, które łudząco przypominają istniejące osoby, obiekty, miejsca lub inne podmioty lub zdarzenia, przez co osoba będąca ich odbiorcą mylnie uznaje je za autentyczne lub prawdziwe – można było rozróżnić dzięki widocznym oznaczeniom, gdy prezentuje się ją na ich interfejsach internetowych". Ponadto dostawcy VLOP i VLOSE mogą również zapewnić łatwą w użyciu funkcję, która umożliwia odbiorcom usługi wskazanie takich informacji. Chociaż ten przepis aktu o usługach cyfrowych wydaje się podobny do art. 50 ust. 4 aktu w sprawie AI, ma on charakter uzupełniający i różni się w dwóch istotnych aspektach:

— **różny zakres przedmiotowy stosowania**: w porównaniu z art. 50 ust. 4 aktu w sprawie AI art. 35 ust. 1 lit. k) aktu o usługach cyfrowych obejmuje szerszy zakres treści, które mogą mylnie wydawać się autentyczne lub zgodne z prawdą, niezależnie od technologii wykorzystywanych do tworzenia treści (tj. przepis ten obejmuje treści, które zostały wygenerowane lub zmanipulowane przy użyciu systemów AI lub jakiegokolwiek innego narzędzia),

([35]) Więcej informacji na temat wzajemnych powiązań między art. 50 aktu w sprawie AI a aktem o usługach cyfrowych można znaleźć również w sekcji 4.4.

— **różny zakres podmiotowy stosowania**: w art. 50 ust. 4 aktu w sprawie AI zobowiązano podmioty stosujące systemy AI, które generują treści typu deepfake lub nimi manipulują, do oznaczania takich treści, natomiast art. 35 ust. 1 lit. k) aktu o usługach cyfrowych ma zastosowanie do dostawców VLOP i VLOSE, którzy rozpowszechniają treści lub umożliwiają ich odnalezienie w drodze wyszukiwania. Na przykład podmioty stosujące systemy AI zintegrowane z VLOP na podstawie aktu w sprawie AI można uznać za „odbiorców usługi" na podstawie art. 3 lit. b) aktu o usługach cyfrowych ([36]), jeżeli wykorzystują VLOP do rozpowszechniania treści wygenerowanych za pomocą odpowiedniego systemu AI, natomiast VLOP można uznać za dostawcę systemu AI, jeżeli spełnione są warunki określone w art. 3 pkt 3 aktu w sprawie AI. W przypadku gdy dostawcy VLOP lub VLOSE udostępniają takim podmiotom stosującym narzędzia oznakowania umożliwiające im oznaczanie ich treści typu deepfake zgodnie z art. 50 ust. 4 aktu w sprawie AI (tj. jasne i wyraźne ujawnienie pochodzenia treści jako wygenerowanych przez AI), podmioty te mogą stosować takie narzędzia w celu wypełnienia spoczywającego na nich obowiązku w zakresie przejrzystości wynikającego z tego przepisu w kontekście wykorzystywanych VLOP lub VLOSE. Zapewnienie takiej funkcji pozostaje bez uszczerbku dla spoczywającej na podmiotach stosujących na mocy aktu w sprawie AI odpowiedzialności za wypełnianie obowiązków w zakresie oznaczania wynikających z art. 50 ust. 4 aktu w sprawie AI. Ponadto dostawcy VLOP i VLOSE będą również uznawani za podmioty stosujące system AI na podstawie aktu w sprawie AI, jeżeli wykorzystują system AI, nad którym sprawują kontrolę, do własnych celów zawodowych (np. tworzenie materiałów wizualnych do wykorzystania w materiałach marketingowych).

(127) Jeżeli materiał typu deepfake zawiera informacje dotyczące bezpośrednio lub pośrednio możliwej do zidentyfikowania osoby fizycznej (która pozostaje przy życiu), podmioty stosujące jako administratorzy na mocy unijnych przepisów o ochronie danych będą również musiały przestrzegać odpowiednich obowiązków wynikających z unijnych przepisów o ochronie danych ([37]). Obejmuje to na przykład stosowanie odpowiedniej podstawy prawnej przetwarzania danych osobowych, informacji dotyczących przetwarzania danych odnoszących się do przedstawionych osób, których dane dotyczą, itp.

(128) Podmioty stosujące systemy AI, które generują treści typu deepfake, będą ponadto zobowiązane do zapewnienia zgodności z odpowiednimi unijnymi przepisami dotyczącymi własności intelektualnej (takimi jak unijne prawo znaków towarowych i unijne prawo dotyczące praw autorskich lub pokrewnych), jeżeli ich treści typu deepfake dotyczą przedmiotu objętego ochroną na mocy tych przepisów. Zastosowanie oznaczenia lub ikony do treści typu deepfake jako takie w żaden sposób nie przesądza o tym, czy dane treści kwalifikują się do ochrony prawem autorskim, co podlega ocenie zgodnie z mającym zastosowanie unijnym prawem autorskim.

(129) W przypadku gdy materiały typu deepfake ukazują rzeczywiste osoby, podmioty stosujące będą ponadto zobowiązane do przestrzegania odpowiednich przepisów chroniących dobra osobiste w odniesieniu do wizerunku lub głosu. Obowiązek w zakresie przejrzystości określony w art. 50 ust. 4 akapit pierwszy aktu w sprawie AI nie oznacza, że wygenerowane lub zmanipulowane przez AI treści typu deepfake, które są szkodliwe i niezgodne z prawem na mocy mającego zastosowanie prawa Unii lub prawa krajowego (przepisy regulujące wprowadzającą w błąd reklamę lub prawo karne, np. materiały przedstawiające niegodziwe traktowanie dzieci w celach seksualnych i intymne zdjęcia udostępniane bez zgody), mogą być generowane i rozpowszechniane (zob. również sekcja 4.4).

6.2. Główne elementy i pojęcia dotyczące tekstu wygenerowanego lub zmanipulowanego przez AI oraz związane z tym obowiązki w zakresie przejrzystości na podstawie art. 50 ust. 4 aktu w sprawie AI

(130) Art. 50 ust. 4 akapit drugi aktu w sprawie AI ma zastosowanie, jeżeli zostaną spełnione pewne warunki:

(i) system musi być systemem AI;

(ii) system jest wykorzystywany do generowania tekstu lub manipulowania tekstem (zob. sekcja 4.2.1) publikowanym w celu informowania opinii publicznej o sprawach leżących w interesie publicznym;

(iii) wykorzystanie systemu AI nie wchodzi w zakres żadnego z wyjątków, które mają zastosowanie wówczas, gdy (1) treści wygenerowane przez AI zostały poddane procesowi weryfikacji przez człowieka lub kontroli redakcyjnej oraz jeżeli za publikację treści odpowiedzialność redakcyjną ponosi osoba fizyczna lub prawna lub (2) wykorzystanie jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców.

([36]) Motyw 2 aktu o usługach cyfrowych.

([37]) Aby uzyskać bardziej szczegółowe wyjaśnienia, zob. wspólne wytyczne Komisji i EROD dotyczące wzajemnych powiązań między aktem w sprawie AI a unijnymi przepisami o ochronie danych (w przygotowaniu).

6.2.1. *Tekst publikowany w celu informowania opinii publicznej o sprawach leżących w interesie publicznym*

(131) W celu doprecyzowania zakresu stosowania art. 50 ust. 4 akapit drugi aktu w sprawie AI, poniżej przeanalizowano trzy elementy: (i) tekst publikowany, (ii) w celu informowania opinii publicznej, (iii) o sprawach leżących w interesie publicznym.

(i) **Tekst publikowany:** tekst wygenerowany lub zmanipulowany przez sztuczną inteligencję powinien być opublikowany. Oznacza to, że tekst powinien być dostępny dla nieokreślonej, odpowiednio dużej liczby niepowiązanych ze sobą potencjalnych czytelników, jednocześnie lub sukcesywnie, niezależnie od tego, czy dostęp ten jest odpłatny czy nie (np. w przypadku subskrypcji) ([38]). Tekstu nie uznaje się natomiast za opublikowany, jeżeli dostęp jest ograniczony do konkretnych osób należących do zamkniętej, prywatnej grupy (np. małej, zamkniętej grupy w komunikatorze internetowym lub jeżeli grupa jest zbyt mała lub nieznacząca). Przykłady tekstów nieuznawanych za opublikowane obejmują np. prywatną korespondencję interpersonalną (do celów zawodowych) lub teksty bądź komunikaty wewnątrz organizacji (np. publikacje w wewnętrznych sieciach firmy).

(ii) **Informowanie opinii publicznej:** opublikowany tekst, wygenerowany lub zmanipulowany przez AI, powinien mieć na celu informowanie opinii publicznej. Oznacza to, że tekst powinien mieć na celu przekazywanie wiedzy, opinii lub faktów. Natomiast krótkie teksty, które nie przekazują w istotny sposób wiedzy, opinii lub faktów, nie mogą być uznane za informujące opinię publiczną.

(iii) **O sprawach leżących w interesie publicznym:** tekst powinien informować opinię publiczną „o sprawach leżących w interesie publicznym". Ogólnie rzecz biorąc, takie sprawy należy rozumieć jako obejmujące kwestie istotne dla ogółu społeczeństwa, zarówno na szczeblu lokalnym, krajowym, unijnym, jak i międzynarodowym, i wymagające debaty lub kontroli publicznej. W związku z tym należy uznać, że teksty dotyczą spraw leżących w interesie publicznym, jeżeli obejmują tematy związane z polityką i procesami demokratycznymi, administracją publiczną i usługami publicznymi, sądownictwem i egzekwowaniem prawa, ochroną praw podstawowych, bezpieczeństwem publicznym, zdrowiem publicznym, ochroną środowiska, bezpieczeństwem konsumentów oraz wszelkimi zmianami gospodarczymi, finansowymi, politycznymi, naukowymi lub kulturowymi, które mogą być przedmiotem debaty publicznej. Sprawy, które można uznać za leżące w interesie publicznym, mogą zmieniać się w czasie i w zależności od kontekstu.

Przykłady tekstów publikowanych w celu informowania opinii publicznej o sprawach leżących w interesie publicznym na podstawie art. 50 ust. 4 aktu w sprawie AI, które to teksty należy oznaczyć jako sztucznie wygenerowane lub zmanipulowane, chyba że podlegają one weryfikacji przez człowieka lub kontroli redakcyjnej (zob. sekcja 6.2.3):

— wygenerowane przez AI streszczenie artykułu napisanego przez człowieka zamieszczonego na stronie internetowej gazety, w którym omówiono niedawną decyzję rady miasta,

— zmanipulowane przez AI fragmenty artykułu zamieszczonego na portalu o tematyce lifestylowej, w którym porównano wpływ różnych diet na konkretną chorobę u kobiet w średnim wieku,

— zmanipulowane przez AI raporty firmowe, publikowane na stronie internetowej spółki notowanej na rynku regulowanym, zawierające informacje dla inwestorów,

— wygenerowany przez AI komunikat umieszczony na profilu instytutu meteorologicznego w mediach społecznościowych, ostrzegający obywateli o burzowej pogodzie i informujący o odpowiednich środkach ostrożności.

Przykłady tekstów nieobjętych zakresem stosowania art. 50 ust. 4 aktu w sprawie AI, które nie są publikowane w celu informowania opinii publicznej o sprawach leżących w interesie publicznym:

— wygenerowane przez AI powieści fantasy,

— zmanipulowany przez AI tekst, który jest częścią reklamy lub opisu produktu przedsiębiorstwa (niezawierający żadnych oświadczeń związanych np. ze zdrowiem, bezpieczeństwem konsumentów lub zrównoważonym rozwojem),

— podsumowanie wiadomości wygenerowane przez chatbota, dostępne wyłącznie dla użytkownika, który mu to polecił,

— zmanipulowany przez AI tekst stworzony przez konsultanta na potrzeby porady dla klienta na temat środków, jakie należy zastosować w celu zapewnienia zgodności regulacyjnej z obowiązującymi przepisami.

([38]) Zob. analogicznie sprawa C-89/04, Mediakabel, EU:C:2005:348, pkt 33; C-192/04, Lagardère, EU:C:2005:475, pkt 31; C-306/05, SGAE, EU:C:2006:764, pkt 37–40; C-527/15, Stichting Brein, EU:C:2017:300, pkt 43–44 i przytoczone tam orzecznictwo; sprawa C-135/23, GEMA, EU:C:2024:526, pkt 38 i przytoczone tam orzecznictwo).

6.2.2. *Obowiązek ujawniania informacji na podstawie art. 50 ust. 4 akapit drugi aktu w sprawie AI*

(132) W art. 50 ust. 4 akapit drugi aktu w sprawie AI zobowiązano podmioty stosujące systemy AI, które generują tekst publikowany w celu informowania opinii publicznej o sprawach leżących w interesie publicznym lub manipulują takim tekstem, do ujawnienia, że taki tekst został sztucznie wygenerowany lub zmanipulowany. W motywie 134 aktu w sprawie AI doprecyzowano, że podobnie jak w przypadku treści typu deepfake podmioty stosujące realizujące publikacje tekstowe objęte zakresem tego przepisu powinny jasno i wyraźnie ujawnić – poprzez odpowiednie oznakowanie wyniku AI i ujawnienie, że źródłem jest AI – że treści te zostały sztucznie wygenerowane lub zmanipulowane. Zgodnie z wymogami dotyczącymi treści typu deepfake metody oznakowania lub ujawniania (w tym klauzule o wyłączeniu odpowiedzialności) stosowane zgodnie z art. 50 ust. 4 akapit drugi aktu w sprawie AI również powinny być zrozumiałe i dostrzegalne dla osób fizycznych (np. w formie widocznych lub słyszalnych oznaczeń), bez konieczności stosowania jakichkolwiek konkretnych narzędzi technicznych lub wykonywania specjalnych działań. W świetle art. 50 ust. 7 aktu w sprawie AI podmioty stosujące mogą przestrzegać kodeksu praktyk uznanego za odpowiedni na podstawie tego przepisu w celu wykazania zgodności z tym obowiązkiem ujawniania informacji. Więcej informacji na temat skutków przestrzegania takiego kodeksu oraz nadzoru nad podmiotami stosującymi, które go nie przyjęły, znajduje się w sekcji 8.1 poniżej.

6.2.3. *Wyjątek od obowiązku w zakresie przejrzystości w odniesieniu do tekstu podlegającego weryfikacji przez człowieka lub kontroli redakcyjnej i odpowiedzialności redakcyjnej*

(133) W art. 50 ust. 4 akapit drugi aktu w sprawie AI przewidziano wyjątek od obowiązku w zakresie przejrzystości określonego w tym przepisie, jeżeli spełnione są łącznie dwa warunki: (i) tekst wygenerowany lub zmanipulowany przez AI musi zostać poddany weryfikacji przez człowieka lub kontroli redakcyjnej oraz (ii) osoba prawna lub fizyczna musi ponosić odpowiedzialność redakcyjną za publikację. W stosownych przypadkach podmioty stosujące mogą opierać się na odpowiednich mających zastosowanie normach zawodowych lub deontologicznych w celu wykazania zgodności z tymi wymogami, jak określono poniżej.

i. Tekst podlegający weryfikacji przez człowieka lub kontroli redakcyjnej

(134) Pierwszy warunek zastosowania wyjątku od art. 50 ust. 4 akapit drugi aktu w sprawie AI wymaga, aby tekst wygenerowany lub zmanipulowany przez AI został poddany weryfikacji przez człowieka lub kontroli redakcyjnej. Weryfikacja przez człowieka oznacza celowe sprawdzenie istoty treści przez co najmniej jedną osobę fizyczną posiadającą odpowiednią wiedzę i zawodowy osąd w odniesieniu do przedmiotu kontroli (np. wzajemna ocena akademicka lub łańcuchy walidacji zawodowej). Sprawdzenie prawdziwości treści jest minimalnym wymogiem, który powinien być częścią tej weryfikacji. Kontrola redakcyjna oznacza kontrolę sprawowaną w praktyce przez odpowiedzialny podmiot redakcyjny (np. redaktora naczelnego) nad treścią, który to podmiot jest uprawniony do zatwierdzenia, zmodyfikowania lub odrzucenia istoty treści w oparciu o podstawy merytoryczne (w tym weryfikację informacji i zapewnienie wiarygodności źródeł). Pozostaje to bez uszczerbku dla istniejących procedur weryfikacyjnych i redakcyjnych oraz norm zawodowych mających zastosowanie do dostawców usług medialnych.

(135) Powierzchowne, wyłącznie formalne lub proceduralne kontrole (np. sprawdzanie pisowni lub korekta gramatyczna), samo istnienie polityki redakcyjnej, zautomatyzowane procesy weryfikacji lub pobieżne zatwierdzenie redakcyjne bez istotnego zaangażowania recenzenta lub jednostki redakcyjnej nie mogą być uznane za spełniające warunki weryfikacji przez człowieka lub kontroli redakcyjnej do celów tego wyjątku.

(136) W przypadku gdy systemy AI są wykorzystywane do modyfikacji, uzupełniania lub przeformułowania treści po zatwierdzeniu redakcyjnym, powstałe w ten sposób treści należy traktować jako wygenerowane lub zmanipulowane przez AI do celów art. 50 ust. 4 aktu w sprawie AI. Każda istotna ingerencja AI mająca miejsce po przeprowadzeniu weryfikacji przez człowieka lub procesu kontroli redakcyjnej spowoduje zatem unieważnienie wyjątku.

(137) W świetle art. 50 ust. 7 aktu w sprawie AI podmioty stosujące mogą przestrzegać kodeksu praktyk uznanego za odpowiedni na podstawie tego przepisu w celu wykazania zgodności z warunkiem weryfikacji przez człowieka lub kontroli redakcyjnej (w szczególności przez podmioty stosujące, które nie są dostawcami usług medialnych podlegającymi istniejącym normom zawodowym lub deontologicznym w zakresie redakcji) określonym w tym przepisie. Więcej informacji na temat skutków przestrzegania takiego kodeksu oraz nadzoru nad podmiotami stosującymi, które go nie przyjęły, znajduje się w sekcji 8.1.

ii. *Odpowiedzialność redakcyjna*

(138) Drugi warunek zastosowania wyjątku od art. 50 ust. 4 akapit drugi aktu w sprawie AI wymaga, aby osoba prawna lub fizyczna ponosiła odpowiedzialność redakcyjną za publikację treści. Oznacza to, że osoba ta musi ponosić ostateczną odpowiedzialność prawną za publikację treści, w tym za weryfikację przez człowieka lub kontrolę redakcyjną (np. osoba fizyczna, rada redakcyjna lub wydawnictwo). Aby zapewnić odpowiedzialność publiczną i zaufanie publiczne oraz zgodnie z obowiązującymi normami zawodowymi w branży mediów, tożsamość i dane kontaktowe osoby prawnej, osoby fizycznej lub podmiotu odpowiedzialnego za redakcję powinny zostać podane do wiadomości publicznej w łatwo dostępnym miejscu (jeśli nie udostępniono ich dotychczas w inny sposób). Informacje te można udostępnić w internecie poprzez umieszczenie ich na przykład w regulaminie strony internetowej lub w innych informacjach prawnych skierowanych do użytkownika. Poza internetem takie informacje można zawrzeć w stopce redakcyjnej lub metryce publikacji.

Przykłady tekstów podlegających weryfikacji przez człowieka lub kontroli redakcyjnej przez osobę prawną lub fizyczną ponoszącą odpowiedzialność redakcyjną:

— zmanipulowany przez AI artykuł umieszczony w gazecie lub wygenerowane przez AI streszczenie artykułu, który został poddany kontroli redakcyjnej przez właściwego redaktora naczelnego, podczas gdy odpowiedzialność redakcyjną ponosi osoba prawna będąca wydawcą tej gazety,

— zmanipulowany przez AI blog akademicki, który został poddany wewnętrznej wzajemnej ocenie i za który odpowiedzialność redakcyjną ponosi odpowiednie centrum badawcze zarządzające blogiem,

— wygenerowane przez AI ostrzeżenia dotyczące bezpieczeństwa publicznego, zatwierdzone przez urzędnika publicznego przed ich rozpowszechnieniem wśród obywateli, za które odpowiada odpowiednia agencja publiczna ds. ochrony ludności,

— wygenerowane przez AI sprawozdania w zakresie zrównoważonego rozwoju, opublikowane na stronie internetowej spółki giełdowej i poddane weryfikacji przez człowieka – specjalistę w danej dziedzinie (np. specjalistę ds. zgodności z przepisami),

— wspomagane AI tłumaczenie artykułu napisanego przez człowieka, które to tłumaczenie zostało zweryfikowane przez człowieka.

Przykłady tekstów, które nie spełniają wymogu, który uzależnia możliwość skorzystanie z wyjątku od weryfikacji przez człowieka lub kontroli redakcyjnej:

— strona internetowa, na której wygenerowane przez AI artykuły dotyczące polityki Unii są zamieszczane bez jakiejkolwiek celowej weryfikacji przez człowieka lub kontroli redakcyjnej,

— wygenerowane przez AI artykuły, które są poddawane weryfikacji i edycji przez inny system AI i w przypadku których redaktor przeprowadza jedynie powierzchowną kontrolę gramatyczną przed publikacją,

— wygenerowana przez AI, wydana we własnym zakresie książka na temat zmiany klimatu, udostępniona na platformie handlu elektronicznego, która to książka nie została poddana żadnej weryfikacji przez właściwą osobę fizyczną lub prawną (ani przez platformę).

6.2.4. *Wyjątek na potrzeby ścigania przestępstw*

(139) Jeżeli wykorzystywanie systemów AI do generowania tekstu publikowanego w celu informowania opinii publicznej o sprawach leżących w interesie publicznym bez weryfikacji przez człowieka (np. wygenerowane przez AI ostrzeżenie policji dotyczące zagrożenia terrorystycznego, opublikowane w mediach społecznościowych) jest dozwolone na mocy prawa w celu wykrywania przestępstw, zapobiegania im, prowadzenia postępowań przygotowawczych w ich sprawie lub ścigania ich sprawców, podmioty stosujące są zwolnione z obowiązku w zakresie przejrzystości określonego w art. 50 ust. 4 aktu w sprawie AI. Wyjaśnienia przedstawione w pkt 46–48 powyżej w odniesieniu do podobnego wyjątku określonego w art. 50 ust. 1 aktu w sprawie AI mają zastosowanie również w odniesieniu do tego przepisu.

6.2.5. *Wzajemne powiązania z innymi aktami prawnymi Unii*

(140) W unijnym prawie medialnym odpowiedzialność redakcyjna jest kluczowym pojęciem zdefiniowanym w art. 2 pkt 8 rozporządzenia (UE) 2024/1083 (europejski akt o wolności mediów) ([39]). W tym kontekście „odpowiedzialność redakcyjna" oznacza „sprawowanie faktycznej kontroli zarówno nad wyborem audycji lub publikacji prasowych, jak i nad sposobem ich zestawienia, do celów świadczenia usługi medialnej, niezależnie od odpowiedzialności za świadczoną usługę na podstawie prawa krajowego". W niektórych sytuacjach definicja ta może pokrywać się z pojęciem odpowiedzialności redakcyjnej stosowanym w art. 50 ust. 4 akapit drugi aktu w sprawie AI, w szczególności gdy podmioty stosujące system AI kwalifikują się również jako dostawcy usług medialnych na podstawie europejskiego aktu o wolności mediów. Odpowiedzialność redakcyjną w rozumieniu art. 50 ust. 4 aktu w sprawie AI należy zatem interpretować zgodnie z istniejącym dorobkiem prawnym w dziedzinie mediów, niezależnie od faktu, że pozostaje ona odrębnym pojęciem, które może mieć zastosowanie również w szerszym kontekście i do innych podmiotów stosujących, o ile przyjmują one odpowiedzialność redakcyjną za publikację tekstu. Dostawcy usług medialnych mogą opierać się na swoich istniejących procesach i normach redakcyjnych, aby skorzystać z wyjątku przewidzianego w art. 50 ust. 4 akapit drugi aktu w sprawie AI.

7. **WYMOGI HORYZONTALNE MAJĄCE ZASTOSOWANIE DO INFORMACJI PRZEKAZYWANYCH NA PODSTAWIE ART. 50 UST. 5 AKTU W SPRAWIE AI**

(141) Zgodnie z art. 50 ust. 5 aktu w sprawie AI informacje, które mają być przekazywane zgodnie z art. 50 ust. 1–4 aktu w sprawie AI, powinny być przekazywane zainteresowanym osobom fizycznym w jasny i wyraźny sposób, najpóźniej w momencie pierwszej interakcji lub pierwszego stosowania (pierwszej styczności). Informacje te powinny również spełniać mające zastosowanie wymogi dostępności. Zainteresowane osoby fizyczne w tym kontekście obejmują osoby wchodzące w bezpośrednią interakcję z systemem AI (art. 50 ust. 1 aktu w sprawie AI), aktywnych lub biernych użytkowników oraz inne osoby, wobec których stosowane są treści syntetyczne wygenerowane lub zmanipulowane przez AI (z bezpośrednią interakcją z systemem AI lub bez takiej interakcji (art. 50 ust. 2 i 4 aktu w sprawie AI)) oraz osoby, wobec których stosowane są systemy rozpoznawania emocji lub kategoryzacji biometrycznej (art. 50 ust. 3 aktu w sprawie AI).

7.1. **Przekazywanie informacji w jasny i wyraźny sposób**

(142) Informacje uznaje się za przekazane w jasny sposób, jeżeli są zauważalne, łatwe do zrozumienia i dostępne dla danej osoby fizycznej (w tym dla osób z niepełnosprawnościami). Informacje uznaje się za przekazane w wyraźny sposób, jeżeli można je łatwo zidentyfikować jako odrębne od innych informacji i od środowiska, w którym prezentowane są treści. Muszą one być również łatwo zrozumiałe dla odbiorców, w tym dla określonych grup, takich jak dzieci lub osoby z niepełnosprawnościami, jeżeli osoby te należą do grupy odbiorców, co do których można przewidzieć, że zetkną się z danymi treściami lub wejdą w interakcję z systemem ([40]). Aby przekazać informacje w jasny i wyraźny sposób, odpowiednie informacje można przedstawić między innymi w ramach interakcji (na podstawie art. 50 ust. 1 aktu w sprawie AI) lub odpowiednich treści (na podstawie art. 50 ust. 2 i 4 aktu w sprawie AI). Informacji nie uznaje się za przekazane w jasny i wyraźny sposób, jeżeli osoby fizyczne mogą je łatwo pominąć lub przeoczyć w normalnych warunkach stosowania lub interakcji (np. gdy są zawarte wyłącznie jako część instrukcji obsługi lub ukryte pod warstwami opcji menu w interfejsie internetowym lub gdy są zapisane w warunkach użytkowania, których użytkownicy zazwyczaj nie czytają).

([39]) Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1083 z dnia 11 kwietnia 2024 r. w sprawie ustanowienia wspólnych ram dla usług medialnych na rynku wewnętrznym i zmiany dyrektywy 2010/13/UE (europejski akt o wolności mediów) (Dz.U. L, 2024/1083, 17.4.2024, ELI: http://data.europa.eu/eli/reg/2024/1083/oj).

([40]) Gdy dzieci są odbiorcami docelowymi treści, informacje te muszą być: (i) przyjazne dla dziecka, dostosowane do wieku, łatwe do zrozumienia i łatwo dostępne dla wszystkich dzieci, w tym osób z niepełnosprawnościami lub dodatkowymi potrzebami w zakresie dostępności; (ii) przedstawione w sposób łatwy do zrozumienia oraz możliwie najprostszy i zwięzły; (iii) przedstawione w sposób łatwy do zweryfikowania i zapewniający natychmiastowy i intuicyjny dostęp w momentach, w których stają się istotne; (iv) przedstawione w językach urzędowych państwa członkowskiego, w którym świadczona jest usługa; (v) angażujące dla dzieci. Może to wymagać również wykorzystania grafik, filmów wideo lub postaci bądź innych technik; (vi) podawane dzieciom stopniowo i rozciągnięte w czasie w celu zmaksymalizowania ilości informacji zapamiętanych przez użytkownika.

7.2. **Pierwsza interakcja lub pierwsze stosowanie**

(143) W art. 50 ust. 5 aktu w sprawie AI wymaga się, aby informacje były przekazywane zainteresowanym osobom fizycznym najpóźniej w momencie pierwszej interakcji lub pierwszego stosowania (pierwszej styczności). Pierwsza interakcja lub pierwsza styczność odnosi się nie tylko do pierwszej osoby fizycznej wchodzącej w interakcję lub mającej styczność z systemem AI lub wynikami działania tego systemu, ale również do każdej późniejszej pierwszej interakcji i pierwszej styczności każdej kolejnej osoby fizycznej z systemem AI lub wynikami działania tego systemu. Jeżeli chodzi o pierwszą interakcję z systemem AI, informacje powinny być przekazywane co najmniej raz na początku sesji interaktywnej, z uwzględnieniem szczególnych kontekstów, o których mowa w pkt 36 i 40 powyżej. W odniesieniu do treści wygenerowanych lub zmanipulowanych przez AI objętych zakresem stosowania art. 50 ust. 2 i 4 aktu w sprawie AI obowiązek informacyjny ma zastosowanie do każdego wyniku działania systemu AI w odniesieniu do każdej osoby fizycznej mającej styczność z danymi treściami. Należy to rozumieć jako każdy moment, w którym istnieje uzasadnione prawdopodobieństwo, że osoba fizyczna będzie miała styczność z wynikami działania systemu AI i dostrzeże informacje, z uwzględnieniem specyfiki ujawnienia w przypadku utworów o wyraźnie twórczym i artystycznym charakterze oraz innych utworów (zob. sekcja 6.1.3). Ponadto podmioty stosujące mogą informować osoby fizyczne przed pierwszą faktyczną interakcją lub stycznością (np. na początku treści zawierających materiały typu deepfake). Jeżeli jednak można racjonalnie przewidzieć, że osoby mogą zacząć odbierać treści nie od ich samego początku, wówczas ujawnienie jedynie na początku treści nie informuje tych osób w odpowiedni sposób i w miarę możliwości powinno zostać uzupełnione ujawnieniem informacji ponownie w dalszej części treści. Można to wdrożyć w stosownych przypadkach, z uwzględnieniem celu obowiązków w zakresie przejrzystości, jakim jest zapewnienie skutecznego udzielania informacji i znaczącej świadomości użytkowników w coraz bardziej immersyjnych i realistycznych środowiskach AI, w których użytkownicy mają coraz większe trudności z rozróżnieniem pochodzenia i prawdziwości treści oraz interakcji.

Przykłady przekazywania informacji przy pierwszej interakcji lub styczności na podstawie art. 50 ust. 5 aktu w sprawie AI:

— Pierwsza interakcja: przy rozpoczynaniu rozmowy z chatbotem lub przy wchodzeniu w fizyczną interakcję z systemem AI.

— Pierwsza styczność: na początku filmu wideo przedstawiającego deepfake, na początku wygenerowanej lub zmanipulowanej przez AI publikacji tekstowej poruszającej sprawy leżące w interesie publicznym lub w przypadku styczności z wygenerowanymi przez AI materiałami typu deepfake podczas przewijania treści w mediach społecznościowych.

7.3. **Zgodność z mającymi zastosowanie wymogami dostępności**

(144) Dostawcy i podmioty stosujące systemy AI objęte zakresem poszczególnych obowiązków w zakresie przejrzystości wymienionych w art. 50 ust. 1–4 aktu w sprawie AI muszą znać mające zastosowanie wymogi dostępności (np. na podstawie dyrektywy 2016/2102 ([41]) i dyrektywy 2019/882 ([42])) i muszą ocenić, czy ich produkty, usługi lub treści muszą być zgodne z tymi przepisami. W takim przypadku muszą one zapewnić dostępność informacji przekazywanych zgodnie z art. 50 aktu w sprawie AI zgodnie z mającymi zastosowanie wymogami dostępności. Art. 50 aktu w sprawie AI nie nakłada odrębnych ani dodatkowych wymogów dostępności.

8. **EGZEKWOWANIE ART. 50 AKTU W SPRAWIE AI**

(145) Niniejsza sekcja ma na celu doprecyzowanie: (i) skutków przestrzegania kodeksu praktyk w odniesieniu do wypełniania obowiązków w zakresie przejrzystości w odniesieniu do treści wygenerowanych przez AI; (ii) roli właściwych organów nadzorczych w egzekwowaniu art. 50 aktu w sprawie AI; (iii) sankcji mających zastosowanie w przypadku naruszenia tych obowiązków; oraz (iv) momentu rozpoczęcia stosowania obowiązków określonych w art. 50 aktu w sprawie AI.

([41]) Dyrektywa Parlamentu Europejskiego i Rady (UE) 2016/2102 z dnia 26 października 2016 r. w sprawie dostępności stron internetowych i mobilnych aplikacji organów sektora publicznego (Tekst mający znaczenie dla EOG) (Dz.U. L 327 z 2.12.2016, s. 1).

([42]) Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/882 z dnia 17 kwietnia 2019 r. w sprawie wymogów dostępności produktów i usług (Tekst mający znaczenie dla EOG), PE/81/2018/REV/1 (Dz.U. L 151 z 7.6.2019, s. 70).

8.1. **Skutki przestrzegania kodeksu praktyk uznanego za odpowiedni**

(146) Dostawcy systemów AI i podmioty stosujące systemy AI objęte zakresem stosowania art. 50 ust. 2 i 4 aktu w sprawie AI mogą wykazać zgodność z odpowiednimi dla nich obowiązkami w zakresie przejrzystości w odniesieniu do treści wygenerowanych przez AI na podstawie tych przepisów, przestrzegając kodeksu praktyk uznanego za odpowiedni zgodnie z art. 50 ust. 7 aktu w sprawie AI ([43]). Taki kodeks nie zastępuje aktu w sprawie AI ani niniejszych wytycznych. Uzupełnia on je, zapewniając jedyne uznane w całej Unii ramy praktyczne dla dostawców systemów generatywnej AI i podmiotów je stosujących w celu wykazania zgodności z tymi obowiązkami, niezależnie od ich siedziby, działalności i właściwych organów nadzoru rynku.

(147) Chociaż dostawcy i podmioty stosujące mogą również wykazać zgodność z tymi obowiązkami za pomocą odpowiednich alternatywnych środków, przestrzeganie kodeksu praktyk, który uznaje się za odpowiedni, jest prostym, przewidywalnym i pewnym pod względem prawnym sposobem wykazania zgodności. W przypadku sygnatariuszy kodeksu praktyk uznanego za odpowiedni zgodnie z art. 50 ust. 7 aktu w sprawie AI ([44]) Komisja i właściwe organy nadzoru rynku skoncentrują swoje działania nadzorcze na ocenie, czy sygnatariusze ci przestrzegali kodeksu praktyk i wdrożyli zawarte w nim środki. Ponieważ dostawcy i podmioty stosujące, którzy przyjęli wspomniany kodeks, będą w przejrzysty sposób informować o środkach, jakie wdrażają w celu zapewnienia zgodności z art. 50 ust. 2, 4 i 5 aktu w sprawie AI, zyskają oni zwiększone zaufanie ze strony Komisji, innych właściwych organów nadzoru rynku i innych zainteresowanych stron (w tym ogółu społeczeństwa). Każde odstępstwo przez sygnatariuszy kodeksu praktyk od zapisów kodeksu, który uznano za odpowiedni, spowoduje utratę przez tych dostawców i przez te podmioty stosujące korzyści w postaci łatwiejszego wykazywania zgodności w tym zakresie.

(148) Oczekuje się, że dostawcy i podmioty stosujące, którzy nie są sygnatariuszami kodeksu praktyk uznanego za odpowiedni zgodnie z art. 50 ust. 7 aktu w sprawie AI, wykażą za pomocą innych odpowiednich środków, w jaki sposób wypełnili swoje obowiązki wynikające z art. 50 ust. 2, 4 i 5 aktu w sprawie AI. Ponadto od takich dostawców i podmiotów stosujących oczekuje się wyjaśnienia, w jaki sposób wdrażane przez nich środki zapewniają wypełnienie ich obowiązków wynikających z aktu w sprawie AI. Powinni oni na przykład przeprowadzić analizę luk, w której porównano by wdrożone przez nich środki ze środkami określonymi w kodeksie praktyk uznanym za odpowiedni. Ponieważ właściwe organy (w tym Urząd ds. AI) będą miały mniejszą wiedzę na temat sposobu, w jaki dostawcy i podmioty stosujące, którzy nie są sygnatariuszami kodeksu praktyk, zapewniają wypełnianie swoich obowiązków wynikających z art. 50 ust. 2, 4 i 5 aktu w sprawie AI, prawdopodobnie będą potrzebowały bardziej szczegółowych informacji podczas monitorowania zgodności. Na przykład dostawcy, którzy nie przyjęli kodeksu praktyk, mogą również otrzymywać większą liczbę wniosków o udzielenie informacji i wniosków o dostęp w celu oceny skuteczności, interoperacyjności, solidności i niezawodności rozwiązań technicznych wdrożonych w celu zapewnienia zgodności z art. 50 ust. 2 aktu w sprawie AI. Podmioty stosujące również mogą otrzymywać takie wnioski w odniesieniu do swoich praktyk oznakowania na podstawie art. 50 ust. 4 aktu w sprawie AI.

(149) Właściwe organy mogą uwzględnić zobowiązania wdrożone zgodnie z kodeksem praktyk uznanym za odpowiedni jako czynnik łagodzący przy ustalaniu kwoty grzywien, w zależności od konkretnych okoliczności ([45]).

(150) Jeżeli kodeks praktyk nie zostanie uznany za odpowiedni zgodnie z art. 50 ust. 7 aktu w sprawie AI, Komisja może przyjąć akt wykonawczy określający wspólne zasady wdrażania obowiązków określonych w art. 50 ust. 2, 4 i 5 aktu w sprawie AI, które miałyby zastosowanie do wszystkich odpowiednich dostawców systemów AI objętych zakresem stosowania tych przepisów i podmiotów stosujących te systemy ([46]).

([43]) Zob. odpowiednie informacje na temat Kodeksu praktyk dotyczący przejrzystości treści wygenerowanych przez AI | Kształtowanie cyfrowej przyszłości Europy opublikowanego 10 czerwca 2026 r.

([44]) Zob. opinia Komisji w sprawie oceny kodeksu praktyk dotyczącego przejrzystości treści wygenerowanych przez AI | Kształtowanie cyfrowej przyszłości Europy.

([45]) Art. 99 ust. 7 lit. e) aktu w sprawie AI.

([46]) Art. 50 ust. 7 aktu w sprawie AI.

8.2. **Organy nadzoru rynku**

(151) Organy nadzoru rynku wyznaczone przez państwa członkowskie, Urząd ds. AI ([47]) i Europejski Inspektor Ochrony Danych ([48]) są odpowiedzialni za nadzorowanie i egzekwowanie przepisów dotyczących systemów AI wchodzących w zakres ich kompetencji, w tym obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI. Takie egzekwowanie przepisów odbywa się w ramach systemu nadzoru rynku i zgodności produktów ustanowionego rozporządzeniem (UE) 2019/1020 i aktem w sprawie AI. Uprawnienia organów nadzoru rynku w zakresie egzekwowania przepisów w odniesieniu do systemów AI określono w tych aktach. Organy te mogą podejmować działania w zakresie egzekwowania prawa w odniesieniu do obowiązków wymienionych w art. 50 aktu w sprawie AI z własnej inicjatywy lub na podstawie skargi, przy czym skargę ma prawo złożyć każda osoba, na którą oddziałuje system AI, lub każda inna osoba fizyczna lub prawna mająca podstawy, by uznać, że zostały naruszone przepisy ([49]). Obowiązki organów nadzoru rynku wynikające z aktu w sprawie AI pozostają bez uszczerbku dla uprawnień i zadań innych organów nadzorczych lub organów ds. praw podstawowych wynikających z ich własnych ram prawnych w sytuacjach, w których zastosowanie ma zarówno art. 50 aktu w sprawie AI, jak i te ramy (np. ochrona danych, ochrona konsumentów).

8.3. **Kary**

(152) Dostawcy i podmioty stosujące, którzy nie spełniają mających zastosowanie obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI, mogą zostać ukarani grzywną w wysokości do 15 000 000 EUR lub – jeżeli sprawcą naruszenia jest przedsiębiorstwo – w wysokości do 3 % jego całkowitego rocznego światowego obrotu z poprzedniego roku obrotowego, w zależności od tego, która z tych kwot jest wyższa. Instytucje, organy i jednostki organizacyjne UE, które naruszają obowiązki w zakresie przejrzystości, mogą podlegać administracyjnym karom pieniężnym w wysokości do 750 000 EUR. W przypadku małych i średnich przedsiębiorstw (MŚP), w tym przedsiębiorstw typu start-up, kara pieniężna nie może przekraczać powyższych wartości procentowych lub kwoty, w zależności od tego, która z tych wartości jest niższa. Kary pieniężne muszą być skuteczne, proporcjonalne i odstraszające oraz uwzględniać interesy MŚP i małych spółek o średniej kapitalizacji, w tym przedsiębiorstw typu start-up, oraz ich sytuację ekonomiczną. Przy podejmowaniu decyzji o ewentualnym nałożeniu administracyjnej kary pieniężnej oraz przy podejmowaniu decyzji o jej wysokości w każdym indywidualnym przypadku właściwe organy powinny ocenić wszystkie istotne okoliczności danej sytuacji oraz, w stosownych przypadkach, szereg czynników, które wpływają na wagę naruszenia (np. charakter, waga i czas trwania naruszenia oraz jego konsekwencje, nieumyślny lub zamierzony charakter, inne okoliczności obciążające lub łagodzące, stopień współpracy z organami nadzoru rynku itp.).

8.4. **Rozpoczęcie stosowania**

(153) Zgodnie z art. 113 aktu w sprawie AI art. 50 aktu w sprawie AI stosuje się od dnia 2 sierpnia 2026 r. Wymaga to, aby wszystkie systemy AI objęte zakresem stosowania wprowadzane do obrotu lub oddawane do użytku w Unii były w tym dniu zgodne z tym przepisem, niezależnie od daty ich wprowadzenia do obrotu lub oddania do użytku ([50]). Rozporządzenie zmieniające akt w sprawie AI (akt zbiorczy dotyczący AI), które zostało niedawno przyjęte przez prawodawcę Unii, przewiduje ukierunkowaną zasadę praw nabytych wyłącznie w odniesieniu do obowiązków w zakresie oznaczania i wykrywania określonych w art. 50 ust. 2 aktu w sprawie AI w odniesieniu do systemów generatywnej AI wprowadzonych do obrotu lub oddanych do użytku przed dniem 2 sierpnia 2026 r. Zapewnia ono dostawcom tych istniejących systemów okres przejściowy na dostosowanie ich systemów do dnia 2 grudnia 2026 r. Systemy, które są częściowo interaktywne, a częściowo generatywne, mogą skorzystać z tego okresu przejściowego wyłącznie w odniesieniu do obowiązku oznakowania na podstawie art. 50 ust. 2 aktu w sprawie AI, natomiast od dnia 2 sierpnia 2026 r. należy zapewnić zgodność z obowiązkiem ujawniania informacji w odniesieniu do systemów AI wchodzących w bezpośrednią w interakcję z osobami fizycznymi.

([47]) Jako organ nadzoru rynku w odniesieniu do systemów AI opartych na modelach AI ogólnego przeznaczenia dostarczanych przez tego samego dostawcę (art. 75 ust. 1 aktu w sprawie AI). Obejmuje to obowiązek mający zastosowanie do dostawców na podstawie art. 50 ust. 1 i 2 aktu w sprawie AI. W odniesieniu do art. 50 ust. 3 i 4 aktu w sprawie AI Urząd ds. AI będzie właściwy tylko wtedy, gdy dostawca danego modelu i systemu AI ogólnego przeznaczenia jest zarazem podmiotem stosującym ten system.

([48]) Jako organ nadzoru rynku dla instytucji, agencji i organów UE.

([49]) Art. 85 aktu w sprawie AI.

([50]) Specjalna zasada praw nabytych określona w art. 111 ust. 2 aktu w sprawie AI ma zastosowanie do systemów AI wysokiego ryzyka wprowadzonych do obrotu przed datą rozpoczęcia stosowania przepisów dotyczących wysokiego ryzyka wyłącznie w odniesieniu do zgodności tych systemów z wymogami i obowiązkami dotyczącymi systemów AI wysokiego ryzyka. Jeżeli system AI wysokiego ryzyka podlega również co najmniej jednemu obowiązkowi w zakresie przejrzystości na mocy art. 50 aktu w sprawie AI, ta specjalna zasada praw nabytych nie ma zastosowania, ponieważ stosowanie poszczególnych obowiązków w zakresie wysokiego ryzyka i przejrzystości na podstawie art. 50 ma charakter kumulatywny, a zasada praw nabytych jest uzasadniona i ograniczona wyłącznie w odniesieniu do zgodności z przepisami dotyczącymi wysokiego ryzyka.

(154) Wygenerowane lub zmanipulowane przez AI wyniki objęte zakresem stosowania art. 50 ust. 2 aktu w sprawie AI oraz treści typu deepfake objęte zakresem stosowania art. 50 ust. 4 akapit pierwszy aktu w sprawie AI, które wygenerowano lub zmanipulowano przed dniem 2 sierpnia 2026 r., nie muszą być oznaczane ani etykietowane retroaktywnie. To samo dotyczy tekstów poruszających sprawy leżące w interesie publicznym objętych zakresem stosowania art. 50 ust. 4 akapit drugi aktu w sprawie AI, które zostały wygenerowane lub zmanipulowane przez AI i opublikowane przed dniem 2 sierpnia 2026 r. Jeżeli natomiast teksty wygenerowane lub zmanipulowane przed dniem 2 sierpnia 2026 r. zostaną opublikowane w tym dniu lub później, muszą zostać oznaczone. Biorąc pod uwagę cel aktu w sprawie AI, jakim jest zapewnienie przejrzystości oraz zwiększenie zaufania do ekosystemu informacyjnego i jego integralności, podmioty stosujące systemy AI i inne podmioty, które są w posiadaniu wcześniej istniejących nieoznaczonych treści typu deepfake lub rozpowszechniają takie treści, zachęca się jednak do podjęcia takiego działania, przy czym nie oczekuje się od nich podejmowania niewspółmiernych wysiłków w tym celu (takich jak audyt wcześniej istniejących baz danych dotyczących treści lub modyfikacja już wydrukowanych opakowań produktów).

9. **PRZEGLĄD I AKTUALIZACJA WYTYCZNYCH**

(155) Niniejsze wytyczne stanowią pierwszą interpretację obowiązków w zakresie przejrzystości określonych w art. 50 aktu w sprawie AI wraz z praktycznymi przykładami. Komisja dokona przeglądu niniejszych wytycznych, gdy tylko będzie to konieczne, biorąc pod uwagę praktyczne doświadczenia zdobyte podczas wdrażania obowiązków w zakresie przejrzystości oraz tempo zmian technologicznych, społecznych i regulacyjnych w tej dziedzinie. Obejmuje to również wszelkie istotne doświadczenia z działań egzekucyjnych w ramach nadzoru rynku oraz wykładnie art. 50 aktu w sprawie AI dokonywane przez TSUE. Podczas takiego przeglądu Komisja może podjąć decyzję o wycofaniu lub zmianie niniejszych wytycznych. Komisja zachęca dostawców systemów AI i podmioty je stosujące, krajowe organy nadzoru rynku za pośrednictwem Europejskiej Rady ds. Sztucznej Inteligencji, forum doradcze ds. sztucznej inteligencji, społeczność naukową i organizacje społeczeństwa obywatelskiego do wniesienia wkładu w ten proces poprzez udział w przyszłych konsultacjach publicznych.