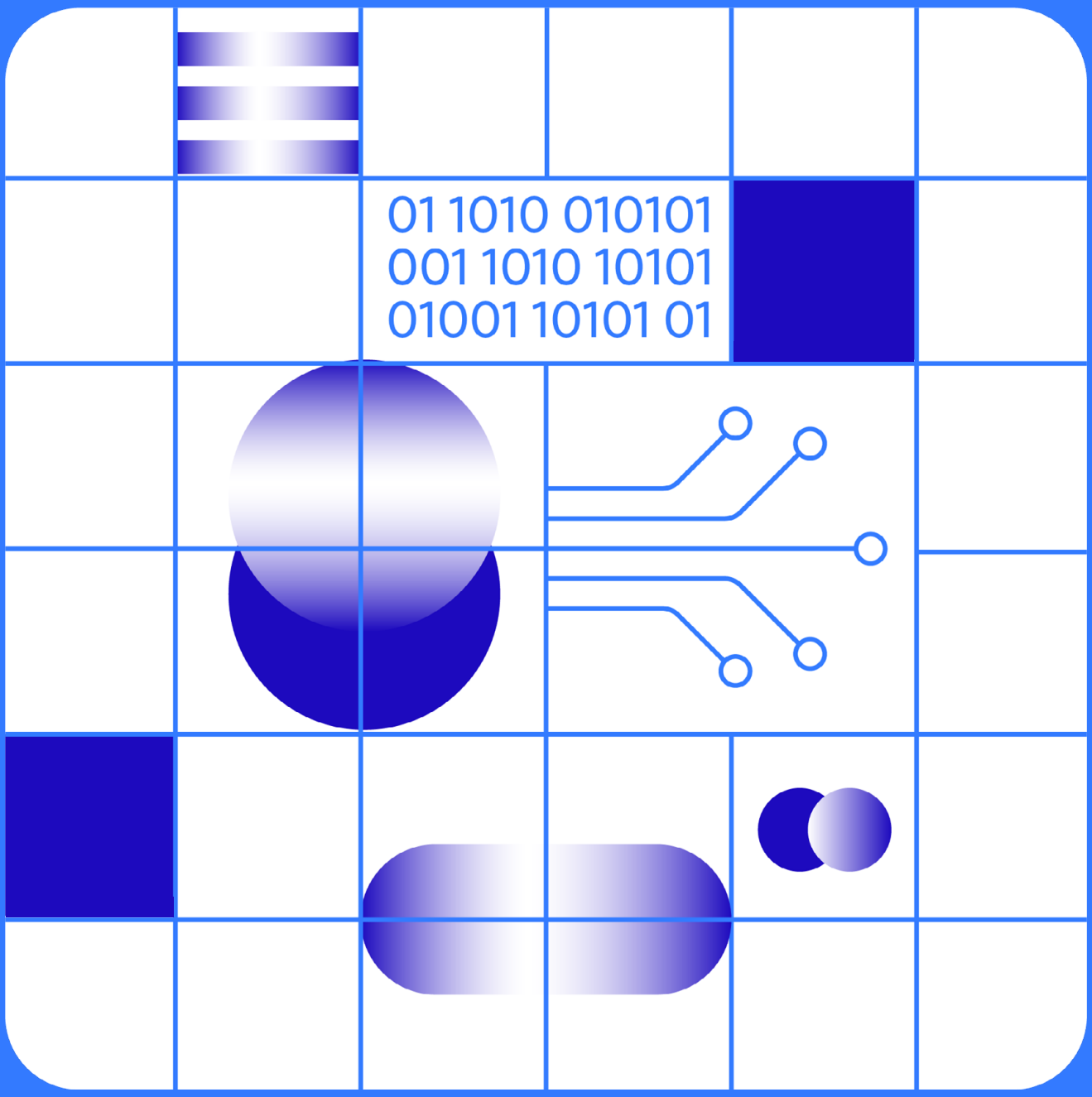
POLITYKA
INSIGHT | RESEARCH



Sztuczna inteligencja pod kontrolą

AUTOR

Ryszard Łuczyn
Polityka Insight

REDAKCJA

Anna Chyckowska

PROJEKT GRAFICZNY

Anna Olczak

Spis treści

Wstęp s. 3

Czym jest SI s. 4

Co zawiera AI Act s. 8

Obszary sporne s. 14

Partnerem raportu jest Fundacja Panoptykon.
Opracowanie jest bezstronne i obiektywne, partner nie miał wpływu na jego tezy ani wymowę.

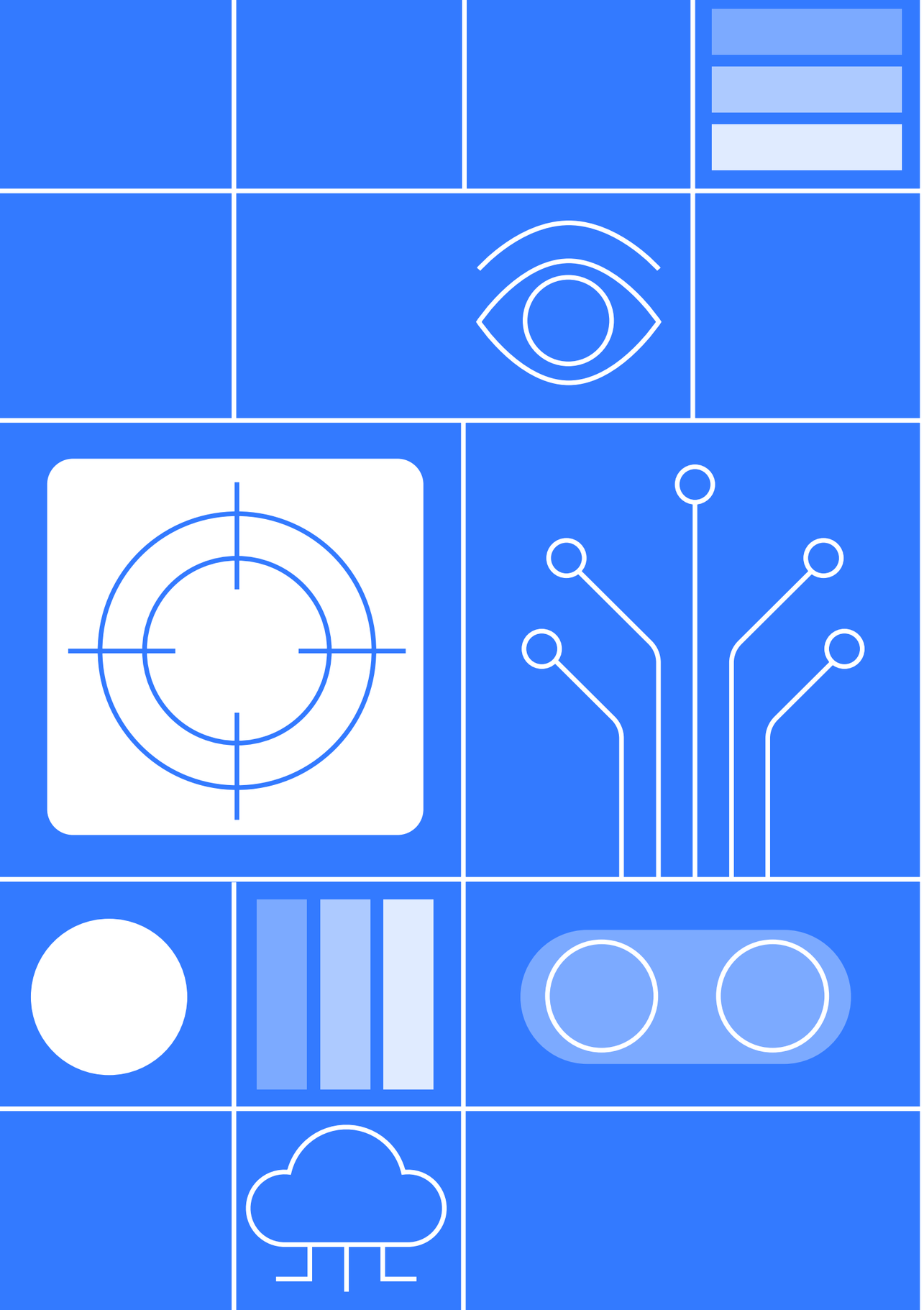
Warszawa, październik 2022 r.

FUNDACJA PANOPTYKON

FUNDACJA PANOPTYKON to polska organizacja pozarządowa, której celem jest ochrona podstawowych wolności wobec zagrożeń związanych z rozwojem współczesnych technik nadzoru nad społeczeństwem. Działalność Fundacji wpisuje się w szerszy nurt badania i reagowania na zjawisko „społeczeństwa nadzorowanego".

POLITYKA INSIGHT

POLITYKA INSIGHT to pierwsza w Polsce platforma wiedzy dla liderów biznesu, decydentów politycznych i dyplomatów. Działa od od 2013 r. i ma trzy linie biznesowe: wydaje serwisy analityczne dostępne w abonamentach (PI Premium, PI Finance i PI Energy), przygotowuje opracowania, prezentacje i szkolenia na zlecenie firm, administracji publicznej i organizacji międzynarodowych oraz organizuje debaty tematyczne i konferencje.
www.politykainsight.pl



W kwietniu 2021 r. Komisja Europejska zaprezentowała projekt Aktu w sprawie Sztucznej Inteligencji – AI Act (AIA). Jak podkreślali przedstawiciele Komisji, rozporządzenie to ma być pierwszą na świecie kompleksową regulacją dotyczącą sztucznej inteligencji. Prezentujący projekt komisarze ds. cyfrowych Margrethe Vestager i ds. wspólnego rynku Thierry Breton obiecywali wprowadzenie zasad gwarantujących etyczne i bezpieczne wykorzystanie sztucznej inteligencji, przy jednoczesnym ograniczeniu interwencji do tych obszarów, gdzie jest ona niezbędna i nie zaburza innowacji[1].

Planowane wprowadzenie AIA wpisuje się w politykę Komisji Europejskiej trwającej kadencji, która uczyniła gospodarkę cyfrową jednym ze swoich priorytetów. Oprócz sfery sztucznej inteligencji (SI) zaangażowanie Komisji jest widoczne w obszarze danych (Data Governance Act, Data Act i wspólne przestrzenie danych), praw użytkowników sieci (Digital Services Act), walki z praktykami monopolistycznymi (Digital Markets Act) czy sytuacji pracowników gospodarki platformowej (dyrektywa w sprawie poprawy warunków pracy za pośrednictwem platform internetowych).

Pierwsza wersja AIA była oparta na rekomendacjach grupy ekspertów wysokiego szczebla ds. SI oraz wydanej w lutym 2020 r. przez Komisję Białej Księdze w sprawie SI. Prace nad rozporządzeniem postępują. Mimo wysiłków francuskiej prezydencji w Radzie nie udało się doprowadzić do pełnego porozumienia między krajami członkowskimi; kompromisowy tekst zaprezentowała prezydencja czeska dopiero w lipcu tego roku. Europarlamentarne komisje LIBE i IMCO w kwietniu przedstawiły wspólny raport w sprawie AIA, do którego złożono ponad 3 tys. poprawek.

Zgodnie z planem pod koniec października komisje dyskutować będą o kompromisowych poprawkach. Niedługo później odbędą się głosowania w komisjach i na posiedzeniu plenarnym[2]. **Zbliża się więc szczególnie istotny moment w pracach nad unijną regulacją SI.** Dlatego Polityka Insight na zlecenie fundacji Panoptykon publikuje niniejszy raport. Ma on na celu przybliżenie tematyki SI i unijnych planów regulacyjnych. Liczymy na to, że przyczyni się do ożywienia konstruktywnej debaty w tym kluczowym momencie.

1 https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682 [dostęp: 6.8.2022].
2 https://www.europarl.europa.eu/cmsdata/248368/Timetables%20May%2022.pdf [dostęp: 6.8.2022].

Czym jest SI

Trudno dziś o osobę, która nie znałaby pojęcia „sztuczna inteligencja”. Buntem maszyn straszą scenarzyści filmowi, a producenci urządzeń AGD zapewniają, że SI czyni ich urządzenia lepszymi od konkurencyjnych. Tymczasem „sztuczna inteligencja” jako dziedzina wiedzy to pojęcie bardzo szerokie – nie ma jednej, powszechnie uzgodnionej definicji prawnej. Co do zasady opisuje się ją jako dziedzinę wiedzy obejmującą m.in. sieci neuronowe, robotykę i tworzenie modeli zachowań inteligentnych oraz programów komputerowych symulujących te zachowania[3]. Sytuacja jest nieco prostsza, jeśli chodzi o definicję „systemu SI”. Powszechnie – również w Polsce – stosowana jest definicja uzgodniona na forum OECD: SI to „system oparty na koncepcji maszyny, która może wpływać na środowisko, formułując zalecenia, przewidywania lub decyzje dotyczące zadanego zestawu celów”[4]. Systemy te:

otrzymują dane lub zbierają je za pomocą czujników.

przetwarzają je, wykorzystując wewnętrzne algorytmy.

na podstawie wyciągniętych wniosków podejmują działania w sferze realnej lub wirtualnej.

Aby dany system można było uznać za wykorzystujący SI, nie może on jedynie wykonywać dokładnych instrukcji zawartych w komputerowym kodzie – musi funkcjonować na podstawie samodzielnie przetwarzanych danych[5]. Pojęciem utożsamianym czasem z SI jest uczenie maszynowe, które, mówiąc precyzyjnie, jest sposobem jej wykorzystania. Gdy chcemy rozwiązać jakiś problem za pomocą uczenia maszynowego, oznacza to, że wykorzystujemy matematyczne modele do tego, by komputer sam – korzystając z dostarczonych danych – nauczył się, w jaki sposób to zrobić; nie dostarczamy mu gotowych rozwiązań.

3 *Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020*, Załącznik do uchwały nr 196 Rady Ministrów z dnia 28 grudnia 2020 r. (poz. 23), s. 65, Polityka_dla_rozwoju_sztucznej_inteligencji_w_Polsce_od_roku_2020.pdf [dostęp: 6.8.2022].
4 Ibidem.
5 https://panoptykon.org/sztuczna-inteligencja-non-fiction [dostęp: 6.8.2022].

ZASTOSOWANIA SI SĄ JUŻ DZIŚ BARDZO SZEROKIE. DZIELI SIĘ JE NA SIEDEM KATEGORII, ZWANYCH WZORAMI*

WZÓR**	OPIS
	Hiperpersonalizacja Tworzenie i dostosowywanie szczegółowego profilu danej osoby na podstawie zbieranych danych, a następnie wykorzystywanie go w systemach rekomendacyjnych (np. w serwisach streamingowych), marketingu czy systemach scoringu kredytowego.
	Rozpoznawanie Wykorzystywanie SI do rozpoznawania obiektów lub innych punktów danych w obrazach, nagraniach filmowych lub dźwiękowych, tekście lub innych materiałach. W praktyce systemy takie służą do rozpoznawania twarzy, przekładania zdjęć tekstu na edytowalny format czy analizy wyników badań medycznych.
	Rozmowa i interakcja z człowiekiem Wchodzenie przez systemy SI w interakcję z człowiekiem za pomocą interfejsu dźwiękowego, tekstowego czy wizualnego. Popularne wykorzystanie tego typu systemów to chatboty lub wirtualni asystenci, np. Alexa (Amazon) czy Siri (Apple).
	Analiza predykcyjna i podejmowanie decyzji Wykorzystywanie danych dotyczących wcześniejszych zdarzeń lub zachowań do przewidywania przyszłych trendów. Systemy bazujące na analizie predykcyjnej mogą być wykorzystywane w meteorologii czy przewidywaniu kryzysów zdrowotnych.
	Systemy skoncentrowane na celu Systemy te koncentrują się na rozwiązywaniu jednego typu problemu, wykorzystując m.in. uczenie maszynowe do zwiększania swojej efektywności dzięki metodom prób i błędów. Mogą być wykorzystywane np. do gier planszowych (AlphaGO), ale również do optymalizacji łańcuchów dostaw czy licytowania w czasie rzeczywistym.
	Systemy autonomiczne Są to systemy tworzone do wykonywania zadań lub wpływania na środowisko (realne lub wirtualne) z minimalnym udziałem człowieka lub całkowicie bez niego. Przykłady ich wykorzystania to boty czy pojazdy autonomiczne***.
	Wykrywanie wzorów i anomalii Zastosowanie SI do rozpoznawania w danych powtarzalnych schematów i odstępstw od nich. Tego typu systemu służą m.in. do wykrywania oszustw finansowych.

* https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf [dostęp: 7.8.2022].
** https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf [dostęp: 7.8.2022].
*** https://www.forbes.com/sites/cognitiveworld/2020/05/30/the-autonomous-systems-pattern-of-ai/ [dostęp: 7.8.2022].

Ograniczenia i kontrowersje wokół SI

Systemy SI są skuteczne w rozwiązywaniu problemów, które wymagają analizy dużej ilości danych i są możliwe do rozstrzygnięcia w ramach określonego schematu (modelu matematycznego). Jednak sposób ich działania – uzależniony od wstępnego określenia założeń systemu, jego architektury, wykorzystywanych danych czy braku ludzkiego nadzoru – może nieść za sobą nie tylko problemy w dziedzinie efektywności, lecz także zagrożenia dla osób, na które te systemy wpływają. W unijnych dokumentach wymienia się zagrożenia z zakresu dyskryminacji, systemu sprawiedliwości, prywatności, odpowiedzialności i wyjaśnialności decyzji oraz m.in. równego dostępu do usług publicznych[6].

System SI może być tylko tak dobry jak dane, na których opiera się jego funkcjonowanie. Przykładowo chatbot wytrenowany przy użyciu nawet ogromnej liczby wpisów encyklopedycznych nie będzie dobrze sobie radził z obsługą klienta posługującego się językiem codziennym. Aby systemy efektywnie funkcjonowały, muszą opierać się na odpowiednio dużym, różnorodnym, adekwatnym i uporządkowanym zbiorze danych. W przeciwnym razie zastosowanie znajdzie powtarzana przez specjalistów od SI zasada „garbage in, garbage out", czyli „śmieci wejdą, śmieci wyjdą". **Ogromne zapotrzebowanie na dane, a zatem ich wartość, wiąże się z tendencją firm technologicznych do naruszania sfery prywatności obywateli.** Potrzeba trenowania coraz częściej wykorzystywanych systemów opartych na uczeniu maszynowym będzie jednym z czynników podtrzymujących tę presję w przyszłości.

Nawet jednak adekwatne, powstałe w sposób nienaruszający prywatności i uporządkowane zbiory danych mogą zawierać pułapki prowadzące do jednego z najczęściej wspominanych zagrożeń związanych z SI – przechyłu algorytmicznego (*algorithmic bias*). **Odzwierciedlające realny świat zbiory będą zawierały dane będące efektem ludzkich decyzji; czasem opartych na historycznych podziałach, czasem błędnych, czasem obciążonych stereotypami.** Jeśli nie zostaną podjęte kroki zaradcze, obciążone takimi problemami dane wpłyną na sugestie i decyzje podejmowane przez system (którego architektura też zresztą może być źródłem problemów). Wśród przykładów przechyłu algorytmicznego na pierwszy plan wybijają się te dotyczące podziałów rasowych oraz nierówności płciowych. Efektem zastosowania „nieobiektywnych" danych może być więc (niepoparte innymi kryteriami) przyznawanie mężczyznom wyższych pensji czy zdolności kredytowej[7] lub klasyfikowanie czarnoskórych przestępców jako bardziej skłonnych do ponownego łamania prawa niż biali[8].

Dodatkowym utrudnieniem bywa to, że systemy SI często są zaprojektowane tak, by dostarczać rozwiązań bez jednoczesnego ujawniania kryteriów, jakie zaważyły na proponowanych rozstrzygnięciach (systemy oparte na uczeniu maszynowym korzystają z dostarczonych algorytmom danych, przetwarzają je i generują odpowiedzi, jednak procesy prowadzące do konkretnego rozwiązania mogą być zamknięte w „czarnej skrzynce"). Wyjaśnialne systemy są trudniejsze do zaprojektowania, droższe lub mniej dokładne. Minusem stosowania systemów opartych na „czarnej skrzynce" jest jednak to, że ich rozstrzygnięcia są niemożliwe do wyjaśnienia. Czasem „skrzynka" ta nie jest otwierana nie tylko z powodu problemów technicznych, lecz także względu na tajemnicę handlową czy obawy o próby obejścia reguł. Unijne prawo już teraz przewiduje ochronę użytkowników przed zagrożeniami związanymi z „czarnymi skrzynkami" algorytmów. Obowiązuje ona jednak tylko w ściśle określonym i dość wąskim zakresie.

6 https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf
7 European Commission, *Algorithmic discrimination in Europe*, s. 7, raport Janneke Gerards (Utrech University) i Raphaële Xenidis (University of Edinburgh, University of Copenhagen), Algorithmic discrimination in Europe – Publications Office of the EU (europa.eu) [dostęp: 10.8.2022].
8 https://panoptykon.org/sztuczna-inteligencja-non-fiction [dostęp: 10.8.2022].

Najważniejsze wymaganie wynika z RODO, według którego wyjaśnialne muszą być systemy podejmujące decyzje w sposób całkowicie automatyczny w oparciu o wykorzystanie danych osobowych, które wywierają znaczący lub prawny wpływ na odbiorcę decyzji[9]. Przy wąskiej interpretacji tego przepisu prawo do wyjaśnienia decyzji podjętej z wykorzystaniem SI znika, jeśli w procesie pojawi się człowiek, choćby jedynie zatwierdzał decyzję podjętą przez algorytm.

Unijne przepisy dotyczące odpowiedzialności za produkty wskazują na producenta jako odpowiedzialnego za wadliwe ich działanie[10]. **Jednak zasada „czarnej skrzynki", niechęć przedsiębiorców do dzielenia się szczegółami algorytmów, zmiany w funkcjonowaniu systemu wskutek wprowadzenia nowych danych czy interakcji z innymi systemami sprzyjają rozmyciu odpowiedzialności za funkcjonowanie systemów SI**. O ile bowiem w teorii odpowiedzialny jest producent, w praktyce trudno dowieść błędów, jakie powstały na etapie projektowania systemu.

SI jest grupą technologii o bardzo szerokich zastosowaniach i rosnącym znaczeniu. Mogą one jednak być wykorzystywane do szkodliwych działań (np. tworzenia deepfake'ów) i wciąż mają szereg nierozwiązanych ograniczeń. Oprócz już wskazanych problemów SI nie dysponuje „zdrowym rozsądkiem", ma problemy z uczeniem się na bieżąco (trzeba ponownie trenować modele z użyciem nowych danych) i odróżnianiem skutku od przyczyny, działań etycznych i nieetycznych[11]. Osoba dotknięta decyzją podjętą *de facto* przez system SI (nawet jeśli w procesie, zgodnie z unijnym prawem, uczestniczy człowiek) może mieć trudności z jej zrozumieniem, a tym bardziej wyjaśnieniem czy zakwestionowaniem.

Ze względu na wszystkie te ograniczenia wykorzystanie SI w celach takich jak projektowanie polityk publicznych, przewidywanie ludzkich zachowań czy podejmowanie decyzji istotnych dla dużych grup społecznych budzi duże kontrowersje. Krytycy dodatkowo podnoszą jeszcze jeden argument – stosowanie SI do podejmowania kontrowersyjnych lub bolesnych działań umożliwia ich właścicielom „chowanie się" za systemem, a więc unikanie odpowiedzialności. Za przykład może posłużyć holenderski skandal związany z zasiłkami na dzieci, który w styczniu 2021 r. doprowadził do dymisji rządu. W latach 2013–2019 tamtejsza administracja podatkowa bezpodstawnie domagała się od tysięcy rodziców zwrotu zasiłków na dzieci, w wielu przypadkach poważnie utrudniając ich sytuację finansową[12]. Jak ujawniono, decyzje podejmowane były na podstawie rozstrzygnięć systemu SI, który uznawał „niewłaściwą" narodowość za jeden z czynników uprawdopodobniających podjęcie śledztwa[13].

9 CERRE, Explaining the Black Box - when law controls AI, raport Alexandre'a de Streel (University of Namur) i Benoit Frenay (University of Namur),s. 7 https://cerre.eu/wp-content/uploads/2020/03/issue_paper_explaining_the_black_box_when_law_controls_ai.pdf
10 KE, *Biała księga w sprawie sztucznej inteligencji – Europejskie podejście do doskonałości i zaufania*, s. 15.
11 https://www.forbes.com/sites/robtoews/2021/06/01/what-artificial-intelligence-still-cant-do/ [dostęp: 10.8.2022].
12 https://www.cnbc.com/2021/01/15/dutch-government-resigns-after-childcare-benefits-scandal-.html [dostęp: 11.9.2022].
13 https://netzpolitik.org/2022/childcare-benefits-scandal-dutch-government-to-pay-million-euro-fine-over-racist-data-discrimination/ [dostęp: 11.9.2022].

Co zawiera AI Act

Zagrożeń związanych z różnymi zastosowaniami SI nie da się całkowicie wyeliminować. Celem Komisji jest więc ich ograniczenie, zwłaszcza w kontekście zastosowań szczególnie niebezpiecznych, związanych choćby z profilowaniem ludzi. AIA opiera się zatem na logice hierarchicznej, grupując systemy SI pod względem stopnia generowanego przezeń ryzyka. Wprowadza definicję systemu sztucznej inteligencji jako „oprogramowania opracowanego przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I, które może – dla danego zestawu celów określonych przez człowieka – generować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję". Wspomniany załącznik – którego zawartość może być zmieniona za pomocą aktu delegowanego Komisji – wymienia następujące techniki i podejścia:

mechanizmy uczenia maszynowego, w tym uczenie nadzorowane, uczenie się maszyn bez nadzoru i uczenie przez wzmacnianie, z wykorzystaniem szerokiej gamy metod, w tym uczenia głębokiego.

metody oparte na logice i wiedzy, w tym reprezentacja wiedzy, indukcyjne (logiczne) programowanie, bazy wiedzy, silniki inferencyjne i dedukcyjne, rozumowanie (symboliczne) i systemy ekspertowe.

podejścia statystyczne, estymację bayesowską, metody wyszukiwania i optymalizacji*.

* https://www.parp.gov.pl/component/content/article/76823:akt-w-sprawie-sztucznej-inteligencji-na-co-powinni-przygotowac-sie-dostawcy-i-uzytkownicy-ai-w-zwiazku-z-nowymi-unijnymi-przepisami [dostęp: 11.8.2022].

AIA dzieli systemy SI na cztery kategorie ryzyka: nieakceptowalne, wysokie, ograniczone (wiążące się z ryzykiem manipulacji) i minimalne. Wykorzystanie systemów cechujących się nieakceptowalnym poziomem ryzyka jest zakazane. Tworzenie i wykorzystanie systemów o wysokim stopniu ryzyka wiąże się z szeregiem obowiązków i ograniczeń. Są one znacznie szersze niż te, które dotyczą systemów z kategorii trzeciej, w przypadku których podstawowym obowiązkiem będzie informowanie użytkowników o tym, że stykają się z SI. Wykorzystanie systemów z kategorii minimalnego ryzyka nie będzie podlegało żadnym ograniczeniom. **Obostrzenia i obowiązki dotyczące systemów z każdej kategorii wciąż są przedmiotem negocjacji i mogą zostać przeformułowane, jednak fundamentalne założenia AIA najpewniej nie zostaną już zmienione.**

KATEGORIA RYZYK SYSTEMÓW SI	PRZYKŁADY	WYBRANE OBOWIĄZKI I OBOSTRZENIA
Nieakceptowalne	Systemy kredytu społecznego, systemy identyfikacji biometrycznej w czasie rzeczywistym	Całkowity zakaz
Wysokie	Systemy związane z infrastrukturą krytyczną, edukacją, egzekwowaniem prawa czy zatrudnieniem	> stworzenie systemu zarządzania ryzykiem > zapewnienie jakości danych treningowych > rejestracja systemu w unijnej bazie > monitorowanie funkcjonowania systemu przez cały cykl życia
Ograniczone	Chatboty, systemy przetwarzania obrazu	Konieczność informowania osób wchodzących w interakcję z systemem o wykorzystaniu SI
Minimalne	Filtry antyspamowe, gry komputerowe oparte na SI	Brak

Zakazane praktyki w dziedzinie SI wyliczone są w tytule drugim AIA. Są to:

Udostępnianie i wykorzystanie systemów opartych na technikach podprogowych w celu istotnego zniekształcania zachowania danej osoby „w sposób, który powoduje lub może powodować u niej lub u innej osoby szkodę fizyczną lub psychiczną”.

Udostępnianie i wykorzystanie systemów wykorzystujących „dowolne słabości określonej grupy osób ze względu na ich wiek, niepełnosprawność ruchową lub zaburzenie psychiczne w celu istotnego zniekształcenia zachowania osoby należącej do tej grupy w sposób, który powoduje lub może powodować u tej osoby lub u innej osoby szkodę fizyczną lub psychiczną”.

Udostępnianie lub wykorzystanie przez organy publiczne systemów podobnych do chińskiego systemu zaufania społecznego.

Wykorzystanie systemów zdalnej identyfikacji biometrycznej w czasie rzeczywistym w przestrzeni publicznej do celów egzekwowania prawa, z wyjątkiem zapobiegania poważnym zagrożeniom oraz poszukiwania ofiar przestępstw lub ich sprawców.

Lista całkowicie zakazanych zastosowań SI jest relatywnie mało pojemna (sposób klasyfikacji sprawia, że zakazy dotkną niewielu zastosowań AI), w przeciwieństwie do listy systemów wysokiego ryzyka. Jest ona zawarta w aneksie trzecim do AIA i zawiera systemy:

- identyfikacji biometrycznej
- związane z zarządzaniem i funkcjonowaniem infrastruktury krytycznej
- związane z edukacją i szkoleniem zawodowym, w tym rekrutacją do placówek edukacyjnych
- związane z zatrudnieniem i zarządzaniem zasobami ludzkimi
- związane z dostępem do usług prywatnych i publicznych, w tym systemy oceny zdolności kredytowej i zarządzeniem służbami ratunkowymi
- związane z egzekwowaniem prawa, w tym służące do oceny ryzyka popełnienia przestępstwa przez daną osobę, oceny stanów emocjonalnych, wykrywania deepfake’ów i oceny wartości dowodów w śledztwie
- przeznaczone do zarządzania migracją, dostępem do azylu i zarządzaniem granicami
- związane z wymiarem sprawiedliwości i procesami demokratycznymi

Powyższa lista może być zmieniona za pomocą aktu delegowanego Komisji. Oprócz objętych nią systemów, jako systemy wysokiego ryzyka, kwalifikować się też będą te, które wiążą się z bezpieczeństwem produktów objętych unijnym prawodawstwem harmonizacyjnym, wyszczególnionym w załączniku drugim do AIA, czyli przede wszystkim różnorakich pojazdów i związanych z nimi urządzeń czy infrastruktury (np. linii kolejowych).

Większość obowiązków związanych z systemami wysokiego ryzyka spoczywa na ich dostawcach, a więc dotyczą one projektowania, testowania, audytowania i certyfikowania systemów SI. Muszą być projektowane tak, by były maksymalnie przejrzyste dla użytkowników i możliwe do nadzorowania przez człowieka. Muszą automatycznie rejestrować zdarzenia związane ze swoim funkcjonowaniem, mieć szeroką dokumentację techniczną i instrukcję obsługi. AIA wymaga też, by systemy SI były projektowane tak, by przez cały cykl swojego życia funkcjonowały dokładnie, stabilnie i na odpowiednim poziomie cyberbezpieczeństwa (może to oznaczać np. stosowanie redundancji). Dostawca zobowiązany jest do stworzenia systemu zarządzania ryzykiem wykorzystywanego przez cały cykl życia SI i zapewnienia, by zbiory danych przeznaczone do trenowania systemów spełniały szereg wymagań co do jakości i reprezentatywności, w tym były testowane pod kątem ryzyka przechyłu algorytmicznego. System zarządzania ryzykiem ma być częścią wymaganego od dostawców systemu zarządzania jakością, obejmującego także m.in. procedury zgłaszania incydentów i komunikacji z organami nadzoru. Wymagania dotyczące zarządzania jakością mają zaś być proporcjonalne do wielkości dostawcy systemów SI.

Przed wprowadzeniem do obrotu systemu SI wysokiego ryzyka (a także w razie wprowadzenia do niego istotnych zmian) konieczne jest poddanie go procedurze oceny zgodności z wymaganiami zawartymi w AIA. W większości przypadków – wszystkich pozycji podanych na liście na str. 10–11 oprócz systemów identyfikacji biometrycznej – jest ona przeprowadzana samodzielnie przez dostawcę. Dostawca może samodzielnie dokonać oceny systemów identyfikacji biometrycznej, ale tylko w przypadku, gdy wykorzysta do tego unijne normy zharmonizowane lub stworzone przez Komisję Europejską wspólne specyfikacje, o ile jedne lub drugie istnieją. W przeciwnym razie musi zastosować bardziej skomplikowaną procedurę, zakładającą udział zewnętrznej jednostki oceniającej system. Prawo wyznaczania, notyfikowania, takich podmiotów będą miały organy wskazane przez państwa członkowskie. Dla każdego systemu wysokiego ryzyka dostawcy sporządzają deklarację zgodności z regulacjami; jednostki notyfikowane przyznają certyfikaty zgodności kontrolowanym przez siebie systemom.

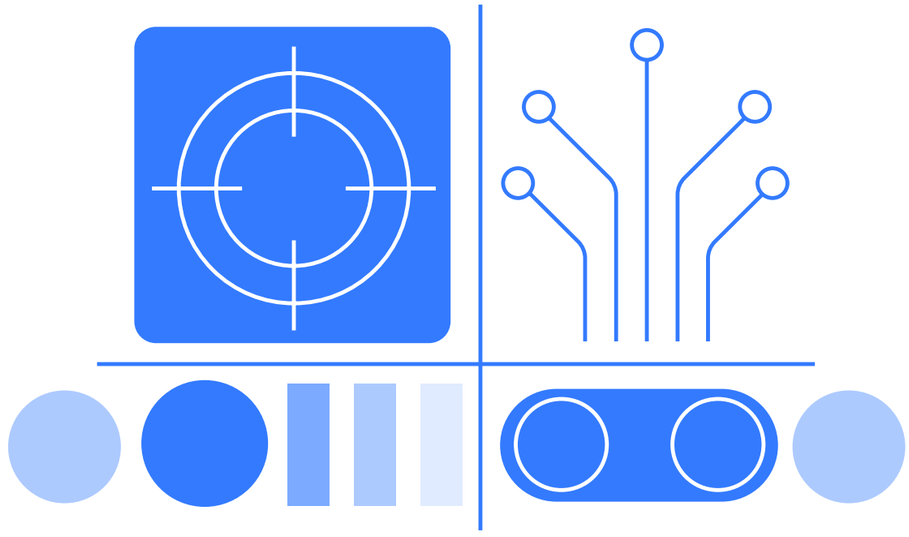
Przed oddaniem systemu do użytku dostawca musi go zarejestrować w prowadzonej przez Komisję bazie danych. Obowiązki dostawców nie kończą się jednak w tym momencie. Konieczne jest też stworzenie systemu monitorowania funkcjonowania danego produktu SI po wprowadzeniu go do obrotu, w ramach którego dostawca ma za zadanie gromadzić i analizować dane na temat funkcjonowania systemu. Jeśli dojdzie do poważnego incydentu związanego z systemem lub zacznie on działać w sposób zagrażający prawom podstawowym obywateli Unii, dostawcy zobowiązani są do informowania o tym krajowych organów nadzoru rynku. W razie potrzeby muszą też podejmować działania naprawcze. Jeśli organ nadzorczy stwierdzi nieprawidłowość, a operator danego systemu jej nie zlikwiduje, system będzie musiał być wycofany z rynku.

Znacznie mniej obowiązków ciąży na użytkownikach systemów Si wysokiego ryzyka. Muszą oni wykorzystywać systemy zgodnie z prawem i instrukcją obsługi, zapewniać adekwatność danych wprowadzanych do systemów, monitorować je i przechowywać rejestry zdarzeń na wypadek incydentów. Te zaś zobowiązują ich do zatrzymania systemu i zgłoszenia problemu dostawcy.

Niewielki zakres obowiązków ciąży również na dostawcach i użytkownikach systemów ograniczonego ryzyka. Mają oni jedynie zapewniać, by osoby stykające się z tymi systemami (np. chatbotami, systemami rozpoznawania emocji) lub ich wytworami (deepfake'i) były o tym fakcie informowane. AIA zachęca też do dobrowolnego tworzenia przez dostawców lub ich organizacje kodeksów postępowania związanych z ekologicznym podejściem do SI czy dostępnością systemów dla osób z niepełnosprawnościami.

Aby monitorować i koordynować przestrzeganie AIA, rozporządzenie zakłada stworzenie Europejskiej Rady ds. SI, do której należeć będą organy nadzorcze wyznaczone przez państwa członkowskie (w Polsce najpewniej UOKiK) i Europejski Inspektor Ochrony Danych. Za nieprzestrzeganie przepisów rozporządzenia, AIA przewiduje możliwość nałożenia bardzo wysokich kar. Stosowanie zakazanych praktyk w zakresie SI lub niewłaściwe obchodzenie się z danymi może kosztować dostawcę do 30 mln euro lub nawet do 6 proc. jego światowego rocznego obrotu. Inne naruszenia AIA są zagrożone karą do 20 mln euro lub 4 proc. światowego rocznego obrotu, a przekazywanie nieprawidłowych informacji jednostkom notyfikowanym lub organom nadzorczym – do 10 mln euro lub 2 proc. światowego rocznego obrotu.

AIA ma nie tylko zagwarantować, że systemy SI wykorzystywane w Unii będą bezpieczne i etyczne, ale też stymulować rozwój tej technologii w UE. Obawiając się więc przeregulowania, Komisja proponuje relatywnie wąski zakres regulacji rynku oraz wprowadzenie zachęt dla innowatorów. Mają nimi być piaskownice regulacyjne w obszarze SI, do których dostęp w pierwszej kolejności zapewniony będą miały startupy.



Obszary sporne

Jak zazwyczaj w przypadku regulacji o tak szerokim zakresie i znaczeniu, AIA wywołało reakcje ze strony licznych podmiotów dążących do zmian w dokumencie. W Parlamencie Europejskim złożono tysiące poprawek, a stolice długo spierały się w Radzie. Organizacje społeczne chcą rozszerzenia zakresu powstającego prawa, a biznes obawiają się przeregulowania i zbyt szerokich definicji. Poniżej prezentujemy **najważniejsze kontrowersje**, które będą musiały zostać rozstrzygnięte na etapie prac w Parlamencie Europejskim lub negocjacji w trilogu.

Kalendarz prac nad AIA

	DATA	WYDARZENIE
2021	21.04	Komisja Europejska przedstawiła projekt AIA
	6.08	Zakończyły się konsultacje publiczne
2022	20.04	Przedstawiciele komisji IMCO i LIBE przedstawili projekt raportu
	01.06	Grupy polityczne w PE zakończyły składanie poprawek do AIA
	15.06	Czeska prezydencja w Radzie przedstawiła propozycję kompromisową wersję AIA
	26.10	Komisje będą debatowały nad kompromisowymi poprawkami
	Jesień/zima	głosowanie w połączonych komisjach i na posiedzeniu plenarnym PE
		Finalna propozycja Rady
	Kolejne miesiące	Trilogi

Zakres stosowania AIA

Jedne z najgorętszych dyskusji budzi już samo to, jaka definicja systemów SI zostanie wykorzystana w rozporządzeniu. Organizacje biznesowe argumentują, że w obecnym brzmieniu, uwzględniającym m.in. techniki statystyczne czy systemy optymalizacji wyszukiwań, obejmie ona również systemy, w których SI nie jest lub nie musi być wykorzystywane, a jednocześnie stanowią one niewielkie ryzyko dla praw obywateli Unii. Grupujący firmy technologiczne European Tech Alliance (EUTA) proponuje wykorzystanie definicji wypracowanej przez grupę ekspertów wysokiego szczebla Komisji, która nie obejmuje wyszczególnienia podejść i technik wykorzystania SI[14]. Również w Parlamencie Europejskim pojawił się pomysł usunięcia listy technik i podejść, jednak bez zmiany zawartej w AIA definicji – chodzi o uczynienie regulacji bardziej odporną na zmiany technologiczne[15]. Z kolei europosłowie z komisji IMCO i LIBE sugerują objęcie definicją również tych systemów, w których cele nie są definiowane przez człowieka[16]. Jedna z najgoręcej dyskutowanych propozycji zmian w tym obszarze pojawiła się w Radzie. Dotyczy ona poszerzenia zakresu obowiązywania AIA na tzw. SI ogólnego przeznaczenia – takich, które mogą być zastosowane do różnego rodzaju zadań bez istotnych zmian w architekturze[17].

Relatywnie mało kontrowersji budzi założenie, że obowiązki dostawców i użytkowników systemów SI powinny zależeć od poziomu ryzyka (tzw. podejście oparte na ryzyku). Dyskusja skupia się natomiast na kwestiach bardziej szczegółowych, czyli zaklasyfikowaniu konkretnych typów systemów do poszczególnych kategorii i obowiązkach, jakie się z tym wiążą.

Organizacje społeczne zrzeszone w sieci EDRi zwracają uwagę, że ramy przyjęte przez Komisję zakładają ocenę ryzyka generowanego przez system, ale abstrahują od kontekstu, w jakim jest on wykorzystywany. Krytykują też fakt, że choć zgodnie z AIA można rozszerzyć listę systemów wysokiego ryzyka, to jedynie w ramach obecnych już w aneksie do rozporządzenia ośmiu obszarów. Bez nowelizacji przepisów nie będzie możliwe ani zwiększenie liczby obszarów, ani zmienienie listy zakazanych zastosowań SI oraz systemów ograniczonego ryzyka. Według organizacji społecznych należałoby stworzyć też dodatkową kategorię systemów nieuznanych za wysokiego ryzyka, lecz mających istotny wpływ na jednostki i w związku z tym poddanych zwiększonym obowiązkom przejrzystości (np. algorytmy regulacji cen).

Klasyfikacja systemów

Krytyka biznesu koncentruje się wokół sposobu klasyfikacji systemów do poszczególnych kategorii ryzyka. Również i w tym przypadku EUTA chciałaby powrotu do podejścia zarysowanego w unijnej Białej Księdze. Zakładało ono klasyfikację systemów w oparciu o dwa czynniki – obszar oraz szczegółowy kontekst ich zastosowania. Według EUTA, obecne podejście (nie biorące pod uwagę kontekstu) grozi bowiem objęciem nadmiernymi obostrzeniami licznych niegroźnych systemów. Odpowiedź na te obawy – przypominającą założenia z Białej Księgi – przygotowała czeska prezydencja w Radzie. Propozycja Czechów zakłada, że system nie będzie mógł być uznany za wysokiego ryzyka, jeśli nie będzie podejmował decyzji samodzielnie, bez ludzkiego nadzoru[18]. Takie podejście pozwalałoby jednak łatwo sztucznie „zaniżać" ryzyko dzięki wprowadzeniu formalnego zatwierdzania rozstrzygnięć systemów przez człowieka.

14 UE, *Biała księga w sprawie sztucznej inteligencji...*, op. cit., s. 19.
15 https://www.euractiv.com/section/digital/news/ai-regulation-filled-with-thousands-of-amendments-in-the-european-parliament/ [dostęp: 14.8.2022].
16 Raport IMCO i LIBE, s. 46. https://www.europarl.europa.eu/doceo/document/CJ40-PR-731563_EN.pdf
17 https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf [dostęp: 14.8.2022].
18 https://www.euractiv.com/section/digital/news/ai-act-czech-presidency-pushes-narrower-ai-definition-shorter-high-risk-list/ [dostęp: 18.8.2022].

W odniesieniu do zastosowań SI objętych całkowitym zakazem biznes domaga się przede wszystkim zwiększenia szczegółowości definicji, w tym poprzez oparcie ich na konkretnych przykładach. Ecommerce Europe chce m.in. dookreślenia, że zakaz stosowania „technik podprogowych" nie oznacza zakazu stosowania SI w marketingu czy personalizacji[19]. Organizacje społeczne z EDRi również dostrzegają problem z definicjami zakazanych systemów SI, ale mają inne propozycje. Chcą m.in. rozszerzenia zakazu stosowania systemów identyfikacji biometrycznej i poszerzenia listy zakazów o systemy rozpoznawania emocji czy przewidywania przestępstw. Podobne poprawki pojawiły się też w Parlamencie Europejskim.

Analogiczne argumenty obu stron słychać w dyskusji o systemach wysokiego ryzyka. Organizacje przedsiębiorców obawiają się, że obecny schemat klasyfikacji sprawi, że do tej kategorii trafią również „niegroźne" systemy AI. EUTA podaje jako przykład obszar pracy, gdzie znacznie utrudnione miałoby być stosowanie systemów SI do niegroźnego ogłaszania wakatów. Stosowanie niegroźnych systemów w potencjalnie ryzykownych obszarach jest dla biznesu argumentem za uwzględnieniem kontekstu w ocenie zagrożeń związanych z danym systemem. Z kolei organizacje społeczne z EDRi chcą, by za systemy wysokiego ryzyka były uznawane także te, które stosuje się w obszarach ochrony zdrowia i ubezpieczeń. Za kluczowe uważają też to, by Komisja mogła modyfikować listę obszarów bez nowelizacji AIA.

Obowiązki związane z systemami

Kolejnym obszarem spornym są obowiązki związane z tworzeniem, udostępnianiem i wykorzystaniem systemów SI, zwłaszcza wysokiego ryzyka. **Z perspektywy EDRi błędem jest obciążanie obowiązkami przede wszystkim dostawców systemów SI**. Kluczowym postulatem organizacji społecznych jest nałożenie na użytkowników systemów wysokiego ryzyka (czyli wdrażające je podmioty) obowiązku przeprowadzania oceny wpływu na prawa podstawowe (*fundamental rights impact assessment*). Miałaby ona wskazywać grupy, na które system będzie miał wpływ, zagrożenia dla praw jednostek i grup, dostępności dla osób z niepełnosprawnościami i wpływu na środowisko oraz środki zaradcze proponowane przez wdrażającego system SI. Zdaniem EDRi użytkownicy powinni być też zobowiązani do weryfikacji zgodności systemów z AIA przed ich wykorzystaniem.

Biznes ze sceptycyzmem odnosi się do zakresu obowiązków wprowadzanych przez nowe prawo – organizacje zrzeszone w EUTA podnoszą problemy biurokracji, skomplikowanych procedur kontrolnych i wysokich kosztów spełniania wymagań, problematycznych zwłaszcza dla mniejszych przedsiębiorstw[20]. Zwracają uwagę, że niektóre obowiązki będą niemal niemożliwe do spełnienia; powtarzającym się przykładem są wymogi dotyczące zbiorów danych treningowych, które mają być „adekwatne, reprezentatywne, wolne od błędów i kompletne". Dodają, że obecna wersja AIA pozwalałaby omijać część obowiązków poprzez trenowanie systemów poza jurysdykcją europejskiego prawa. Domagają się załatania tej luki, ostrzegając przed pogorszeniem pozycji europejskich firm względem amerykańskiej i azjatyckiej konkurencji[21]. Ta sama organizacja apeluje, by dostawcy systemów SI wysokiego ryzyka musieli rejestrować je w unijnej bazie danych tylko w sytuacji, gdy w ramach wewnętrznej oceny stwierdzą, że nie są w stanie odpowiednio zmitygować ryzyk, a systemy mogły być wprowadzane na rynek jeszcze przed otrzymaniem certyfikatu (by nie powodować opóźnień).

19 https://www.euractiv.com/section/digital/news/ai-act-czech-presidency-pushes-narrower-ai-definition-shorter-high-risk-list/ [dostęp: 18.8.2022].
20 https://www.digitaleurope.org/resources/digitaleuropes-initial-findings-on-the-proposed-ai-act/ [dostęp: 18.8.2022].
21 https://eutechalliance.eu/wp-content/uploads/2021/11/EUTA-AI-Position-Paper_Final.pdf [dostęp: 18.8.2022].

Jednocześnie organizacje społeczne argumentują, że rejestrowane powinny być nie tylko systemy, ale także przypadki ich wykorzystania przez poszczególne podmioty; to bowiem pozwoliłoby oceniać wpływ, jaki mogą mieć na konkretne grupy. EDRi, powołując się na konieczność zwiększenia transparentności, chce, by w bazie publikowano również kluczowe wnioski z oceny wpływu systemu SI na otoczenie oraz informacje o celu, jaki ów system realizuje, a sama baza była łatwa w wykorzystaniu i możliwa do maszynowego przeszukiwania.

Organizacje społeczne domagają się również rozszerzenia obowiązku informowania o styczności z SI na wszystkie systemy wysokiego ryzyka. Obecnie bowiem obowiązek ten jest wąski – o ile dana osoba będzie musiała zostać poinformowana o tym, że rozmawia z chatbotem opartym na SI, to już niekoniecznie o tym, że to SI zdecydowała o odrzuceniu jego podania o pracę. Jak podkreśla należący do EDRi Panoptykon, bez wprowadzenia tego typu powiadomienia niemożliwe będzie dochodzenie praw przez osoby, na które wpływają systemy SI[22]. Powiązanym postulatem organizacji społecznych jest właśnie wprowadzenie do AIA konkretnych, nowych praw dla osób dotkniętych działaniem systemów SI. Zdaniem NGO-sów AIA powinno jasno stanowić, że jednostki i grupy mają prawo do niebycia poddanymi wpływowi niebezpiecznych systemów, a także do zrozumiałego wyjaśnienia istotnych (w tym prawnie wiążących) i wpływających na nie decyzji podjętych z udziałem systemów SI. W przypadku naruszenia ich praw wprowadzanych na gruncie AIA powinny móc efektywnie odwołać się, a osoby i organizacje społeczne – złożyć skargę do organu nadzorczego (w imieniu pokrzywdzonych).

OBSZAR SPORNY	WYBRANE PROPONOWANE ROZWIĄZANIA
Definicja systemu SI	> zawężenie (np. w formie powrotu do definicji z Białej Księgi) > poszerzenie na systemy, w których cele nie są definiowane przez człowieka > objęcie definicją systemów SI ogólnego przeznaczenia
Klasyfikacja systemów	> uwzględnienie kontekstu zastosowania systemu w jego klasyfikacji > doprecyzowanie definicji dotyczących zakazanych zastosowań SI oraz SI wysokiego ryzyka > doprecyzowanie i rozwinięcie listy zakazanych zastosowań SI oraz listy obszarów wysokiego ryzyka
Obowiązki związane z systemami	> obciążenie użytkowników obowiązkiem wykonania oceny wpływu na prawa podstawowe i weryfikacji zgodności systemów z AIA przed ich wdrożeniem > ograniczenie poziomu skomplikowania procedur związanych z zapewnianiem zgodności systemów z prawem > zapewnienie, że obowiązki dotyczące systemów tworzonych poza UE będą obejmowały również ich trenowanie zgodnie z unijnym prawem
Transparentność i prawa osób dotkniętych przez SI	> zwiększenie zakresu informacji publikowanych w unijnej bazie danych > rozciągnięcie obowiązki informowania o styczności z SI na systemy wysokiego ryzyka > stworzenie mechanizmu skargowego dla osób i organizacji > ograniczenie liczby sytuacji, w których dostawcy muszą rejestrować system w unijnej bazie danych

22 https://panoptykon.org/regulacja-ai [dostęp: 19.8.2022].

Powyższe wątki obrazują zakres i kierunki debaty o treści AIA, ale nie wyczerpują wszystkich ważnych kwestii. Inne, nadal nierozwiązane, problemy dotyczą m.in. kwestii ekologii, dostosowania prawa do potrzeb osób z niepełnosprawnościami, zasady udostępniania (np. niezależnym badaczom) informacji o systemach SI oraz ich zastosowaniach, relacje AIA z innymi unijnymi aktami legislacyjnymi czy krajowe procedury postępowania w zakresie certyfikacji systemów. Przedstawienie wszystkich wątków debaty w tak wąskim materiale byłoby niemożliwe. Mamy jednak nadzieję, że nawet i on skutecznie zwróci przedstawicielom polskiej administracji, biznesu i mediów na potrzebę regulacji SI i kluczowy moment, w którym dziś się znajdujemy.