



OCENA WPŁYWU ALGORYTMÓW

Systemów Sztucznej Inteligencji i Systemów Automatycznego
Podejmowania Decyzji

Propozycja dla sektora publicznego



Fundacja
Moje Państwo

Ocena wpływu algorytmów systemów sztucznej inteligencji i systemów automatycznego podejmowania decyzji

– propozycja dla sektora publicznego

Fundacja Moje Państwo jest organizacją działającą na rzecz rozwoju demokracji, otwartej i przejrzystej władzy publicznej oraz zaangażowania obywatelskiego. Wierzymy, że państwo powinno być sprawne i użyteczne dla ludzi. Tworzymy narzędzia, które ułatwiają korzystanie z jego zasobów i danych. Dbamy o kulturę wykorzystywania technologii przez państwo.

Projekt uzyskał wsparcie Europejskiego Funduszu AI, wspólnej inicjatywy Sieci Fundacji Europejskich (NEF). Wyłączna odpowiedzialność za projekt spoczywa na organizatorze (organizatorach), a treść niekoniecznie musi odzwierciedlać stanowisko Europejskiego Funduszu AI, NEF lub fundacji partnerskich Europejskiego Funduszu AI.

Autorzy

Tekst: **Magdalena Siwanowicz-Suska** Fundacja Moje Państwo,

Monika Kajalidis Fundacja Moje Państwo,

Narzędzie oceny wpływu algorytmów:

Magdalena Siwanowicz-Suska, Fundacja Moje Państwo

Monika Kajalidis, Fundacja Moje Państwo

Bartosz Paszcza

Konsultacje: **Karolina Iwańska**, The European Center for Not-for-Profit Law Stichting

Projekt graficzny i skład: **Jakub Waluchowski**

Font: Azo Sans by Rui Abreu

wersja PL

Podmiot publikujący: Fundacja Moje Państwo (Luty, 2023)

Autorskie prawa majątkowe do raportu przysługują Fundacji Moje Państwo.



Ocena wpływu algorytmów systemów sztucznej inteligencji i systemów automatycznego podejmowania decyzji

– propozycja dla sektora publicznego

1. WSTĘP

Celem dokumentu jest zwrócenie uwagi na sposób wdrażania systemów sztucznej inteligencji (AI) i automatycznego podejmowania decyzji (Automated Decision Making, ADM) w sektorze publicznym. Dokument przedstawia zarys koncepcji narzędzia oceny wpływu algorytmów (Algorithmic Impact Assessment, AIA). Raport zawiera też motywy, które przemawiają za stosowaniem tego rodzaju narzędzi w procesie zarządzania systemami AI/ADM. W dokumencie zawarto propozycję oceny wpływu algorytmów systemów AI/ADM, adresowaną do władz publicznych, zarówno na poziomie centralnym jak i samorządowym. Narzędzie jest punktem wyjścia do dalszej dyskusji o lepszym, bezpieczniejszym i bardziej efektywnym wdrażaniu systemów AI/ADM w państwie.

2. Technologia AI i ADM w sektorze publicznym w praktyce. Dziś i jutro.

Coraz więcej instytucji publicznych decyduje się przy wykonywaniu swoich zadań na korzystanie z systemów sztucznej inteligencji lub systemów automatycznego podejmowania decyzji. To trend, który może usprawnić działanie państwa i prowadzić do bardziej efektywnego i ekonomicznego zarządzania w publicznych instytucjach. Celem Unii Europejskiej jest zostanie strategicznym liderem w wykorzystaniu sztucznej inteligencji w sektorze publicznym. Ten zamiar został określony w przeglądzie skoordynowanego planu w sprawie sztucznej inteligencji z 2021 r.¹, w którym Komisja Europejska proponuje „uczynienie z sektora publicznego pioniera w zakresie stosowania sztucznej inteligencji”. Kierunek polegający na zwiększaniu udziału technologii w zarządzaniu procesami w państwie jest wyraźnie widoczny. Wymaga jednak wystąpienia szeregu zmian w sektorze publicznym, który powinien być właściwie przygotowany – organizacyjnie, finansowo, a także pod kątem posiadania odpowiednich kwalifikacji przez urzędników. Rozwijanie technologii AI/ADM w sposób bezpieczny, godny zaufania i zapewniający poszanowanie podstawowych praw i wolności człowieka będzie możliwe tylko pod warunkiem utworzenia odpowiedniego ekosystemu wewnątrz administracji publicznej.

Wizja nowoczesnych i innowacyjnych struktur publicznych, chociaż atrakcyjna, nie jest pozbawiona wad. Używanie systemów AI/ADM wiąże się z obawami o rozmycie odpowiedzialności za podejmowane rozstrzygnięcia, czy zwiększone pole do nadużyć. Niejasny sposób działania² i kontroli tych systemów, czy nieoczywiste ryzyko szkodliwego działania algorytmów dla społeczeństwa, które może

ujawnić się dopiero w dłuższej perspektywie – to tylko część kolejnych kwestii budzących wątpliwości przy korzystaniu z AI przez państwo. Wszystkie te kwestie nie są problemem, ale realnymi ryzykami, które wymagają odpowiedniej reakcji i zarządzania nimi na wczesnym etapie planowania wprowadzenia systemu w ramach konkretnego środowiska. W sektorze publicznym systemy automatycznego podejmowania decyzji powinny być objęte modelem zarządzania, który będzie uwzględniał skutki działania takiego systemu, w szczególności w kontekście praw obywatelskich. Taki model powinien już na wstępie zakładać sprawdzenie, czy zastosowanie danego systemu jest niezbędne w kontekście rozwiązywanego zagadnienia oraz proporcjonalne z punktu widzenia praw człowieka.

Długa lista wybranych przypadków używania AI przez sektor publiczny w Europie, utworzona przez Komisję Europejską, Wspólne Centrum Badań (JRC)³ pokazuje, że trend korzystania z tej technologii jest obecny, zarówno na poziomie władz krajowych, jak i lokalnych. Instytucje publiczne używają AI do obsługi spraw obywateli (stosowanie chatbotów), czy wykrywania rozbieżności w orzecznictwie, a także w bardziej kontrowersyjnych z punktu widzenia praw człowieka celach wykrywania poszukiwanych osób czy identyfikowania podejrzanych zachowań na lotniskach⁴. Wykorzystywanie algorytmów sztucznej inteligencji przez sektor publiczny ma zastosowanie od analizy danych z monitoringu jakości wody, po automatyczne podejmowanie decyzji w elektronicznych systemach deklaracji urzędów skarbowych (ten przykład w każdym roku jest częścią rzeczywistości wielu z nas).⁵

Wiele z wdrożonych w sektorze publicznym systemów skutecznie wspiera pracę administracji, ale pojawiają się też przypadki kontrowersyjnych lub nawet nielegalnych wdrożeń. Na liście utworzonej przez KE JRC są systemy, z których używania zrezygnowano po pewnym czasie od ich wdrożenia. Przykładem może

1. <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review> [dostęp: 5.12.2022 r.]

2. <https://mojepanstwo.pl/aktualnosci/773> [dostęp: 5.12.2022 r.]

3. European Commission, Joint Research Centre (JRC) (2021): Selected AI cases in the public sector. European Commission, Joint Research Centre (JRC) [Dataset] PID: <http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a> [dostęp: 5.12.2022 r.]

4. Ibidem.

5. Więcej informacji o ADM w Europie w raporcie AlgorithmWatch “Automating Society Report 2020” <https://automatingsociety.algorithm-watch.org/> [dostęp: 20.01.2023 r.]

być system prognostyczny o nazwie „Early Help Profiling System (EHPS)” używany w 2018 r. w hrabstwie Hackney (Londyn), który wykorzystywał dane z różnych źródeł do identyfikacji dzieci i rodzin, które mogłyby potrzebować dodatkowej pomocy ze strony gminnych pracowników socjalnych. System miał umożliwić wczesną interwencję w tych rodzinach, unikając wystąpienia kryzysu z kosztowniejszymi działaniami w późniejszym czasie. Charakter gromadzenia danych i ocena ryzyka były nieprzejrzyste, ponieważ dane były przetwarzane przez podmiot prywatny, co budziło obawy obywateli. Osoby, których dane wykorzystano, nie były o tym informowane. Dodatkowo utrzymywano w tajemnicy, które wskaźniki miały określać ryzyko, ponieważ za rozwój systemu odpowiadała prywatna firma i argumentowała, że informacje są wrażliwe z handlowego punktu widzenia. Gmina zrezygnowała z korzystania z systemu.^{6,7} Kontrowersje pojawiły się także wokół holenderskiego systemu do skuteczniejszego wykrywania oszustw socjalnych „SyRi – Detect welfare fraud”, wdrożonego na poziomie krajowym. System sam nie podejmował żadnych decyzji, ale wydawał zalecenia dla urzędników dotyczące prowadzenia dalszych postępowań. Różne organizacje zwracały uwagę, że używanie systemu powoduje zbyt wiele naruszeń prywatności oraz jest dyskryminujące wobec niezamożnych obywateli i osób należących do mniejszości etnicznych. Krytykowano również brak przejrzystości działania systemu i brak możliwości zapoznania się ze swoimi własnymi danymi przez osoby, których dotyczy sugestia systemu. W następstwie sprawy sądowej sąd holenderski zdecydował na początku 2020 r., że użycie SyRi jest niezgodne z art. 8 EKPC⁸, w związku z czym jego użycie zostało wycofane.⁹

Identyfikowanie ryzyka związanego z działaniem systemu AI/ADM w sektorze publicznym powinno się odbywać na jak najwcześniejszym etapie, a nie już po jego wdrożeniu – z bezpośrednimi skutkami dla obywateli, których dotyczą operacje systemu.

W interesie społeczeństwa jest używanie przez sektor publiczny narzędzi, które sprawią, że stosowanie systemów AI/ADM będzie bardziej przejrzyste i rozliczalne, a ryzyka będą możliwe do zidentyfikowania i zaadresowania przed uruchomieniem działania systemów. Rozbudowy wymaga bufor, który pomagałby administracji publicznej eliminować możliwe negatywne skutki działania systemów AI/ADM oraz weryfikować, czy dany system w ogóle powinien znaleźć zastosowanie do rozwiązania danego problemu.

3. Ocena wpływu algorytmów zarządzania systemami AI/ADM – wybrane przykłady narzędzi

Siłą w podejściu do każdej technologii wydaje się być różnorodność. Podążając w tym kierunku – systemy AI/ADM używane w sektorze publicznym nie powinny być tylko przejrzyste lub tylko dokładne, ale powinny uwzględniać jeszcze inne aspekty. Do zagospodarowania jest znacznie więcej czynników: etycznych, społecznych, organizacyjnych, prawnych, technologicznych, czasem środowiskowych. Takim systemom powinien towarzyszyć odpowiedni proces, który równoważyłby sposób jego działania przez cały cykl życia systemu, zmniejszał ryzyka związane z jego funkcjonowaniem i dbał, by technologia używana przez państwo nie była niebezpieczna dla człowieka.

Odpowiedzią na te wyzwania jest znalezienie modelu zarządzania systemami AI i ADM, który uwzględniałby wpływ działania systemu na otoczenie oraz pozwalałaby na odpowiednie reagowanie na skutki wprowadzenia systemu. Ważnym elementem tego modelu jest ocena wpływu algorytmów.

6. Ibidem. ID:146.

7. <https://www.hackneycitizen.co.uk/2019/10/30/town-hall-drops-pilot-programme-profiling-families-without-their-knowledge/> [dostęp: 5.12.2022 r.]

8. https://www.echr.coe.int/documents/convention_pol.pdf [dostęp: 20.01.2023 r.]

9. Ibidem. ID: 78.

Ocena wpływu algorytmów (Algorithmic Impact Assessment, **AIA**) jest narzędziem do oceny ryzyka wdrożenia konkretnego systemu **AI/ADM**. Ocena ma pomagać w lepszym zrozumieniu ryzyk związanych z automatycznym podejmowaniem decyzji oraz ma ułatwiać zarządzanie zidentyfikowanymi zagrożeniami. Ocena wpływu algorytmów pozwala uporządkować proces opracowania, wdrażania i utrzymywania systemu – ze szczególnym uwzględnieniem kontekstu, w którym system ma działać. Nie ma jednej ustalonej metody dokonywania takiej oceny. Stosowanie narzędzi typu **AIA** jest stosunkowo nowym trendem, ale widać, że potrzeba refleksji nad działaniem systemów podejmowania decyzji jest zauważalna. Aktualnie możemy obserwować sięganie po rozwiązania oparte na **AIA** zarówno w sektorze publicznym jak i w biznesie.

Unia Europejska

Aktualnie procedowany na poziomie UE projekt Aktu w sprawie sztucznej inteligencji (*AI Act*)¹⁰ zakłada stosowanie narzędzi w pewnym stopniu podobnych do **AIA** tj. „oceny zgodności”. W pierwotnej wersji aktu przedstawionej przez KE w kwietniu 2021 r. zawarto obowiązek dokonywania oceny zgodności wobec systemów wysokiego ryzyka. Ocena zgodności ma być procesem weryfikacji spełnienia wymagań dotyczących systemów **AI**, określonych projektowanej regulacji. Wymagania te są związane z ustanowieniem systemu zarządzania jakością, zarządzaniem danymi, dokumentacją techniczną, prowadzeniem dokumentacji, przejrzystością i dostarczaniem informacji użytkownikom, nadzorem ludzkim, dokładnością, solidnością i cyberbezpieczeństwem. Obowiązek przeprowadzania oceny zgodności nałożono wyłącznie na dostawców systemów (rozumianych jako podmioty, które opracowują dany system i wprowadzają go do obrotu lub użytku odpłatnie lub nieodpłatnie). Projektowana regulacja zakłada dwie opcje oceny zgodności (art. 43). Pierwsza z nich polega na wykonywaniu oceny przez samych dostawców.

Druga opcja zakłada ocenę zgodności z udziałem jednostki notyfikowanej. Dla większości systemów wysokiego ryzyka wskazanych w aneksie III przewidziano ocenę zgodności przeprowadzaną przez samych dostawców (ocena zgodności opierająca się na kontroli wewnętrznej). Sektor publiczny byłby adresem takiego obowiązku tylko w przypadku, gdyby był dostawcą systemu wysokiego ryzyka. Na przykład, gdyby zdecydował się na opracowanie i oddanie do użytku systemu sztucznej inteligencji przeznaczonego do podejmowania decyzji o przyznawaniu świadczeń socjalnych osobom fizycznym. Instytucja publiczna nie musiałaby dokonywać oceny zgodności, gdyby zamówiła system „na zewnątrz” od producenta systemów **AI**. Wówczas producent dokonałby oceny, a instytucja publiczna musiałaby po prostu na niej polegać. Taka konstrukcja oceny zgodności wiąże się jednak z poświęceniem niewielkiej uwagi¹¹ temu, co dzieje się na etapie używania systemów **AI** wysokiego ryzyka. Pewne nieprawidłowości, czy niechciane skutki związane z działaniem systemu **AI** mogą się pojawić dopiero na etapie jego wykorzystywania w określonym obszarze. Dostawcy w ramach oceny zgodności będą w stanie ocenić jedynie fragment kontekstu, w jakim może działać system wysokiego ryzyka. Sposób i cel wykorzystywania systemu sztucznej inteligencji ma kluczowe znaczenie dla wpływu jaki dany system może wywierać na otoczenie, w tym na prawa podstawowe. Dlatego należałoby zaadresować dodatkowe obowiązki dla użytkowników (wdrażających) systemów **AI** wysokiego ryzyka w ramach **AIA** w zakresie przestrzegania praw podstawowych.¹²

Sposób używania systemów **AI** przez władze publiczne będzie miał wpływ na sytuację wielu obywateli, którzy znajdą się w zasięgu oddziaływania wdrażanego rozwiązania. Dlatego sektor publiczny powinien dokonywać **AIA** przed rozpoczęciem używania systemu **AI** w procesach podejmowania decyzji wobec obywateli oraz publikować wyniki przeprowadzonej oceny.

10. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0012.02/DOC_1&format=PDF [dostęp: 12.12.2022 r.]

11. Projekt Aktu w sprawie sztucznej inteligencji przedstawiony w kwietniu 2021 r. przez KE zawiera w art. 29 pewne obowiązki dla użytkowników systemów sztucznej inteligencji wysokiego ryzyka. [dostęp: 12.12.2022 r.]

12. <https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/> [dostęp: 12.12.2022 r.]

Ta kwestia została zauważona w toku prac Parlamentu Europejskiego nad *AI Act*. W ramach przedstawionych poprawek zaproponowano wprowadzenie obowiązku stosowania oceny wpływu na prawa podstawowe przez użytkowników systemów wysokiego ryzyka, z uwzględnieniem specyficznych warunków dla sektora publicznego (np. AM 2079, AM 2081).¹³ Prace nad regulacją dalej są w toku, ale obowiązek stosowania *AIA* przez użytkowników systemów AI wysokiego ryzyka – w tym przez sektor publiczny – ma realną szansę stać się częścią regulacji. Należy zaznaczyć, że *AIA* powinny być używane nie tylko przez instytucje publiczne, ale także przez sektor prywatny.¹⁴

Rada Europy

Rada Europy (RE) pracuje nad pierwszym wiążącym międzynarodowym instrumentem prawnym w dziedzinie sztucznej inteligencji. Powołany przez Komitet Ministrów RE Komitet *Ad hoc* ds. Sztucznej Inteligencji (CAHAI)¹⁵, uzgodnił pod koniec 2021 r. możliwe elementy ram prawnych dotyczących sztucznej inteligencji, w oparciu o standardy Rady Europy dotyczące praw człowieka, demokracji i praworządności.¹⁶ Ostateczne zalecenia CAHAI stanowią podstawę negocjacji w Komitecie ds. Sztucznej Inteligencji (CAI).¹⁷

W ramach wypracowanych zaleceń CAHAI zarekomendował, by celem instrumentu było zapewnienie pełnej spójności z poszanowaniem praw człowieka, funkcjonowaniem demokracji i przestrzeganiem zasad prawa w zakresie opracowywania, projektowania i stosowania systemów sztucznej inteligencji, niezależnie od tego, czy działania te są podejmowane przez podmioty prywatne czy publiczne (par. 11). W zaleceniach pojawia się propozycja modelu oceny wpływu systemów sztucznej inteligencji na korzystanie z praw człowieka, funkcjonowanie demokracji i przestrzeganie praworządności (HUDERIA) (rozdział XII).

CAHAI zalecił, aby w HUDERIA uwzględniono co najmniej następujące główne etapy, pod warunkiem przeprowadzenia wstępnej oceny i uwzględnienia zaangażowania interesariuszy, w stosownych przypadkach:

- **Identyfikacja ryzyka:** Identyfikacja istotnych zagrożeń dla praw człowieka, demokracji i praworządności;
- **Ocena wpływu:** ocena wpływu, z uwzględnieniem prawdopodobieństwa i dotkliwości skutków dla tych praw i zasad;
- **Ocena zarządzania:** ocena ról i obowiązków podmiotów odpowiedzialnych, posiadaczy praw i interesariuszy we wdrażaniu i zarządzaniu mechanizmami łagodzenia skutków;
- **Łagodzenie i ocena:** Identyfikacja odpowiednich środków łagodzących i zapewnienie ciągłej oceny. (par. 50)

W zaleceniach podkreślono także, że należy zapewnić udział zainteresowanych stron w ocenie wpływu (par. 53).

13. <https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-IMCO-LIBE-Report-All-Amendments-14-June.pdf> [dostęp: 12.12.2022 r.]

14. <https://edri.org/wp-content/uploads/2022/05/Obligations-on-users-AIA-Amendments-17022022.pdf> [dostęp: 20.01.2023 r.]

15. <https://www.coe.int/en/web/artificial-intelligence/cahai> [dostęp: 23.01.2023 r.]

16. <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d> [dostęp: 23.01.2023 r.]

17. <https://www.coe.int/en/web/artificial-intelligence/cai#> [dostęp: 23.01.2023 r.]

Kanada

Praktycznym przykładem oceny wpływu algorytmów w sektorze publicznym jest **AIA** stosowane w Kanadzie. Kanada już w 2016 r. rozpoczęła prace nad tworzeniem koncepcji regulacji prawnej dotyczącej odpowiedzialnego korzystania z systemów AI w państwie.¹⁸ Dyrektywa w sprawie automatycznego podejmowania decyzji obowiązuje w Kanadzie od 2019 r. Akt prawny nakłada na podmioty publiczne objęte zakresem regulacji¹⁹ obowiązek:

- **przeprowadzenia AIA** przed wyprodukowaniem jakiegokolwiek zautomatyzowanego systemu decyzyjnego;
- **stosowanie odpowiednich wymagań** określonych w dyrektywie, określonych w AIA;
- **aktualizowanie AIA** w przypadku zmiany funkcjonalności systemu lub zakresu zautomatyzowanego systemu decyzji.
- **publikowanie ostatecznych wyników** algorytmicznych ocen wpływu w dostępnym formacie za pośrednictwem stron internetowych.²⁰

Kanadyjska regulacja definiuje **AIA** jako ramy pomagające instytucjom lepiej zrozumieć i zmniejszyć ryzyko związane z systemami zautomatyzowanych decyzji oraz zapewnić odpowiednie wymogi w zakresie zarządzania, nadzoru i sprawozdawczości/audytu, które najlepiej pasują do rodzaju projektowanej aplikacji.

Holandia

W sektorze publicznym w Europie także mamy przypadki stosowania rozwiązań typu **AIA**. W Holandii Ministerstwo Spraw Wewnętrznych i Stosunków z Królestwem opracowało ocenę skutków w kontekście praw podstawowych i algorytmów (FRAIA).^{21, 22} Ocena została przygotowana do stosowania przez władze publiczne. Narzędzie ma pomagać w mapowaniu zagrożeń dla praw człowieka związanych ze stosowaniem algorytmów oraz w podejmowaniu działań mających na celu przeciwdziałanie tym zagrożeniom. Stosowanie oceny skutków ma uniemożliwiać korzystanie z algorytmów, których konsekwencje działania nie są jeszcze jasne. Ma też zmniejszać ryzyko nieskuteczności systemów, a także ryzyko naruszeń praw obywatelskich.²³

Polska

Pod koniec grudnia 2020 r. w Polsce została przyjęta „*Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020*”.²⁴ Polityka, czy raczej strategia, opisuje działania, które Polska powinna wdrożyć i cele, które powinna osiągnąć w perspektywie krótkoterminowej (do 2023 r.), średnioterminowej (do 2027 r.) i długoterminowej (po 2027 r.). Cele mają służyć rozwojowi polskiego społeczeństwa, polskiej gospodarki i polskiej nauki w obszarze sztucznej inteligencji. Polska strategia **AI** także odnosi się do narzędzia oceny ryzyka w kontekście systemów sztucznej inteligencji. Po pierwsze, dokument zakłada pilotażowe wdrożenie w Polsce modelowych ocen ryzyka dla systemów wykorzystujących **AI** oraz modelowych wyjaśnień, skierowanych do odbiorców tych systemów, aby inspirować prace nad podobnymi standardami na poziomie UE.²⁵ Po drugie, jednym z celów krótkoterminowych

18. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html> [dostęp: 19.12.2022 r.]

19. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appA> (patrz art. 9 dyrektywy) [dostęp: 19.12.2022 r.]

20. Ibidem.

21. <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms> [dostęp: 19.12.2022 r.]

22. <https://www.uu.nl/en/news/dutch-house-of-representatives-endorses-mandatory-use-of-human-rights-and-algorithms-impact> [dostęp: 20.01.2023 r.]

23. <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms> [dostęp: 19.12.2022 r.]

24. <https://www.gov.pl/web/govtech/polityka-rozwoju-ai-w-polsce-przyjeta-przez-rade-ministrow--co-dalej> [dostęp: 19.12.2022 r.]

25. Ibidem.

polskiej strategii AI jest analiza etycznych skutków implementacji AI oraz oddziaływania systemów AI na sferę praw człowieka. Wśród narzędzi, które mają doprowadzić do jego realizacji wskazano wspieranie, w ramach prac nad regulacją AI:

- **zasad przetwarzania danych osobowych** wynikających z RODO (w szczególności zasady minimalizacji danych),
- **rzetelnej oceny ryzyka** dla systemów wykorzystujących AI i przeciwdziałania błędom w ich projektowaniu (np. algorytmicznego skrzywienia),
- **przejrzystości, rozliczalności i wyjaśnialności** systemów wykorzystujących AI,

w szczególności tych, które realizują zadania w sferze publicznej lub mają wpływ na sferę praw i wolności człowieka.

4. Propozycja oceny wpływu algorytmów dla sektora publicznego

Ocena wpływu algorytmów powoli staje się narzędziem wykorzystywanym przy wdrażaniu systemów AI/ADM w administracji publicznej i prawdopodobne, że wkrótce jej przeprowadzanie stanie się wiążącym wymogiem prawnym. Kierunek jest ten sam – chodzi o lepszą i bezpieczniejszą technologię używaną przez państwo. Nie ma przy tym jednej ustalonej metodologii przeprowadzania oceny wpływu algorytmów. Jest to nowe podejście, które wymaga podjęcia wielu decyzji dotyczących sposobu jej wdrożenia, zależnych od specyfiki potrzeb, na które ma odpowiadać system.

Ocena wpływu algorytmów jest jednym z komponentów dobrego zarządzania systemami AI/ADM przez sektor publiczny. Pozostaje jeszcze wiele innych elementów takich jak np. ocena adekwatności zastosowania systemu AI/ADM, edukacja i rozwijanie kompetencji urzędników²⁶, stosowanie odpowiednich trybów zamówień publicznych²⁷, czy audyty systemów AI/ADM.

W prezentowanym dokumencie proponujemy model oceny wpływu algorytmów, adresowany do sektora publicznego, zarówno na poziomie rządowym jak i samorządowym. Przedstawione narzędzie jest inspirowane holenderskim rozwiązaniem FRAIA²⁸ oraz kanadyjskim modelem oceny wpływu algorytmów²⁹. Narzędzie jest wyjściową propozycją, która może wymagać dostosowania (rozwięcia/przeformułowania) w zależności od potrzeb danego wdrożenia.

Model dzieli się na dwie części. Pierwsza z nich ma na celu ustalenie charakterystyki systemu AI/ADM za pomocą udzielenia odpowiedzi na pytania zawarte w formularzu. Druga część zmierza do przypisania systemowi poziomu wpływu i zastosowania rekomendacji przewidzianych dla danej grupy systemów.

Nie zawsze realizacja takiej oceny będzie łatwym zadaniem. Czasem nie sposób przewidzieć krótkoterminowych czy długoterminowych skutków działania systemów AI lub ADM. W wielu przypadkach ocena pozwoli jednak zidentyfikować ryzyka związane z funkcjonowaniem systemu w danym środowisku.

Kto powinien stosować zaproponowane rozwiązanie? W modelu idealnym ocena skutków algorytmizacji mogłaby być wykonywana przez niezależną publiczną instytucję, ale obecnie, w związku z brakiem takich wyspecjalizowanych jednostek publicznych,

26. <https://mojepanstwo.pl/pliki/zarzadzanie-ai-adm-sektor-publiczny.pdf> [dostęp: 19.12.2022 r.]

27. <https://mojepanstwo-public.s3.eu-central-1.amazonaws.com/publications/BkQLXRhU41fRlpBdxlZZUe3l6gJvQjID8SFjCpb6.pdf> [dostęp: 19.12.2022 r.]

28. <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms>

29. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>

AIA może być realizowana przez instytucje publiczne, które są zainteresowane wdrożeniem danego systemu AI/ADM.³⁰ Każdy system powinien mieć swojego „właściciela”, a struktura takiej „własności” powinna być jasno określona w strukturze administracji. W hiszpańskiej ustawie o systemie prawnym sektora publicznego już w 2016 roku pojawiła się intencja ustanowienia dedykowanej instytucji publicznej, która byłaby odpowiedzialna za nadzór, kontrolę i audyt systemów ADM w administracji publicznej.³¹ Niestety pojawiają się problemy z wdrożeniem tych przepisów w praktyce.³²

Praktyka polegająca na stosowaniu oceny wpływu algorytmów pozwala na zwiększenie przejrzystości, bo wprowadza obowiązek przemyślenia przez urzędników do jakich interakcji może doprowadzić zastosowanie systemu w konkretnym środowisku. Prowadzi też do wytworzenia konkretnych dokumentów, odnoszących się do sposobu podejmowania decyzji w państwie. W większości przypadków dokumenty te będą miały charakter informacji publicznej i co do zasady będą podlegały udostępnieniu na zasadzie prawa dostępu do informacji publicznej. Obywatele będą mieli więc możliwość dotarcia do informacji o systemie, ale także do informacji o tym w jaki sposób państwo zadbało o to, by system był bezpieczny dla obywateli. Rekomendowanym rozwiązaniem w kwestii dokumentów wytworzonych w ramach stosowania AIA jest ich proaktywne publikowanie w ramach publicznie dostępnej bazy on-line, zawierającej podstawowe informacje na temat używania systemów AI/ADM przez daną instytucję.

Rejestry informujące o używaniu systemów AI/ADM pozwalają zainteresowanym obywatelom uzyskać informację, że dana instytucja publiczna w ogóle korzysta z rozwiązania opartego o sztuczną inteligencję w wykonywaniu swoich zadań. Z drugiej

strony taki rejestr może być jednym z elementów zarządzania systemami AI/ADM dla urzędników danej jednostki publicznej. Rejestry sztucznej inteligencji funkcjonują już w niektórych europejskich miastach: przykładem jest tu fiński rejestr sztucznej inteligencji, który zawiera wykaz systemów AI używanych przez miasto Helsinki. Za pośrednictwem rejestru każdy może dokonać szybkiego przeglądu systemów sztucznej inteligencji miasta lub zbadać ich bardziej szczegółowe informacje.³³ Podobny rejestr systemów AI/ADM jest prowadzony przez miasto Amsterdam w Holandii.³⁴ Rekomendujemy publikowanie dokumentacji związanej z AIA, tym bardziej, że być może wkrótce będzie to wymóg prawny na gruncie AI Act, na który trzeba się będzie przygotować.

W raporcie na temat publicznych rejestrów AI³⁵, utworzonym we współpracy Helsinek i Amsterdamu zarekomendowano organizacjom publicznym, aby zapewniły przejrzystość za pośrednictwem rejestrów AI w takich aspektach swoich systemów algorytmicznych jak: cel i wpływ, odpowiedzialność, zbiory danych, przetwarzanie danych, niedyskryminacja, nadzór ludzki, ryzyko i środki ich ograniczania oraz możliwość wyjaśnienia. Zgadza się co do tego, że tworzenie tego rodzaju informacji we współpracy z dostawcami technologii AI i innymi partnerami oraz publikowanie tych informacji powinno stać się standardową procedurą we wszystkich publicznych projektach wykorzystujących systemy AI/ADM.



30. <https://mojepanstwo.pl/pliki/zarzadzanie-ai-adm-sektor-publiczny.pdf> [dostęp: 19.12.2022 r.]

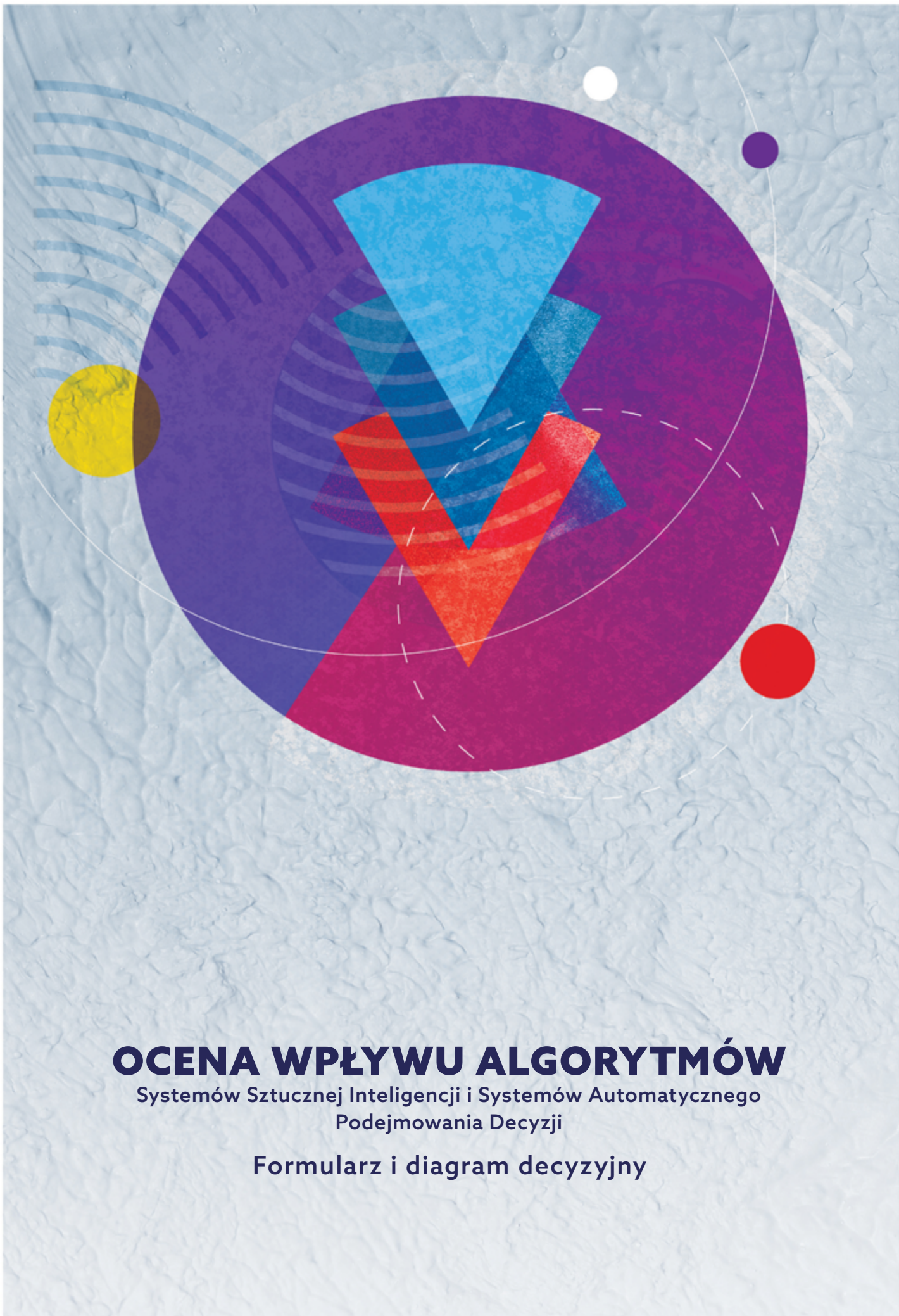
31. <https://www.boe.es/buscar/act.php?id=BOE-A-2015-10566> (art. 41) [dostęp: 19.12.2022 r.]

32. <https://privacyweek.it/event/assessing-the-ai-that-assesses-us-the-work-of-civil-society-and-research-in-setting-auditing-standards-for-the-public-sector/> [dostęp: 19.12.2022 r.]

33. <https://ai.hel.fi/en/ai-register/> [dostęp: 21.12.2022 r.]

34. <https://algorithmeregister.amsterdam.nl/en/ai-register/> [dostęp: 21.12.2022 r.]

35. <https://ai.hel.fi/en/get-to-know-ai-register/> [dostęp: 21.12.2022 r.]



OCENA WPŁYWU ALGORYTMÓW

Systemów Sztucznej Inteligencji i Systemów Automatycznego
Podjęcia Decyzji

Formularz i diagram decyzyjny



Nazwa narzędzia:

Kontakt do opiekuna merytorycznego narzędzia

Instytucja wiodąca i instytucje współpracujące:

Data sporządzenia

Osoba odpowiedzialna za narzędzia

Nr projektu wdrożenia

OCENA WPŁYWU ALGORYTMÓW FORMULARZ

1. JAKA JEST PODSTAWA PRAWNA AUTOMATYZACJI?

Na jakiej podstawie prawnej wdrażany jest algorytm do automatycznego podejmowania decyzji (dalej także jako „system”)

.....

.....

.....

.....

.....

2. JAKI PROBLEM JEST ROZWIĄZYWANY?

A Jakiego obszaru dotyczy automatyzacja?

- Zdrowie
- Interesy ekonomiczne
- Wsparcie społeczne
- Dostępność i mobilność
- Udzielanie zezwoleń i licencji
- Wymiar sprawiedliwości, policja, prokuratura
- Prawa obywatelskie
- Ochrona środowiska
- inne? Jakież?

.....

.....

.....

.....

.....





B Jakiego konkretnego procesu dotyczy automatyzacja?

.....

.....

.....

.....

.....

C Dlaczego Twój zespół chce wykorzystać algorytm do pomocy w podejmowaniu decyzji?

- Wyeliminowanie opóźnień w obecnym systemie
- Zwiększenie jakości decyzji
- Zmniejszenie kosztów
- Skrócenie terminów
- Zwiększenie dostępności usług dla obywatela
- inne? Jakież?

.....

.....

.....

.....

.....

D Czy i jak próbowano wdrażać zmiany procesu innymi metodami niż przy wykorzystaniu automatyzacji? Jeśli tak, dlaczego się nie powiodły?

.....

.....

.....

.....

.....

E Jakie wartości publiczne przemawiają za wykorzystaniem automatyzacji?

.....

.....

.....

.....

.....





4. W JAKI SPOSÓB I KIEDY NASTĄPI EWALUACJA EFEKTÓW NARZĘDZIA ORAZ JAKIE MIERNIKI ZOSTANĄ ZASTOSOWANE?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

5. CZY PODOBNE ROZWIĄZANIA AUTOMATYZACJI PROCESU ISTNIEJĄ W INNYCH JEDNOSTKACH ADMINISTRACJI LUB W INNYCH KRAJACH, W SZCZEGÓLNOŚCI KRAJACH CZŁONKOWSKICH OECD/UE?

Wymień znane Ci podobne rozwiązania.

.....

.....

.....

A Jeśli istnieją podobne rozwiązania, to czy były dla nich przygotowane podobne oceny skutków? Jeśli zostały opublikowane - wskaź źródło.

.....

.....

.....

.....

B Jeśli istnieją podobne rozwiązania, to czy były dla nich opracowane audyty/raporty z ewaluacji? Jeśli zostały opublikowane - wskaź źródło.

.....

.....

.....

.....



6. GŁÓWNE ELEMENTY NARZĘDZIA

A Jakie są główne funkcje narzędzia, które odbywają się w sposób automatyczny?

- Rozpoznawanie obrazów i przedmiotów
- Ocena ryzyka
- Analiza tekstu i mowy
- Wytwarzanie treści
- Optymalizacja procesu i automatyzacja pracy
- inne? Jakież?

B Czy proces automatyzacji będzie trudny do wyjaśnienia?

C Czy skutki decyzji podejmowanych przez system są odwracalne?



D Na jaki okres są podejmowane decyzje?

.....

.....

.....

.....

.....

7. OBSZARY ODDZIAŁYWANIA ALGORYTMU AUTOMATYCZNEGO PODEJMOWANIA DECYZJI

[powrót do diagramu](#)



A Czy, a jeśli tak to jak, decyzje będą oddziaływać na:

- Prawa człowieka i wolności obywatelskie?
- Zdrowie i dobrostan obywateli?
- Interesy ekonomiczne obywateli?
- Ekosystem i środowisko naturalne?

.....

.....

.....

.....

.....

B Czy rozważono potencjalne ryzyka wystąpienia niepożądanych rezultatów w związku z działaniem systemu i podjęto kroki zapobiegawcze?

Jeśli ryzyka i środki zapobiegawcze opisane zostały w innym dokumencie, załącz go.



**C** Uzupełnij tabelę poniżej.

- Ryzyka niepożądanego oddziaływania systemu na obszary wskazane w pkt. **7A**

- Źródło ryzyka

- Możliwe następstwa wystąpienia ryzyka, w tym wskazanie grup osób dotkniętych

- Ocena ryzyka (niskie/umiarkowane/wysokie)

- Podjęte środki zapobiegawcze

- Planowane działania w następstwie wystąpienia ryzyka

- D**
- Czy cele, które ma osiągnąć system są wystarczająco ważne, aby mogły uzasadnić podjęcie ryzyka wskazanego w tabeli powyżej (pkt
- 7C**
-)?

 T N

- E**
- Czy w związku z oddziaływaniem systemu na obszary wskazane w pkt.
- 7A**
- występuje ryzyko nie do zaakceptowania, przemawiające za rezygnacją z zastosowania systemu?

8. ODPOWIEDZIALNOŚĆ I ROZLICZALNOŚĆ PODEJMOWANYCH DECYZJI

- A.** Czy proces podejmowania decyzji przez system jest udokumentowany?
W tym, czy istnieje dokumentacja źródeł danych, ich reprezentatywności, potencjalnych skrzywień algorytmu (bias), poziomu dokładności i niezawodności algorytmu?

Jeśli tak, proszę ją załączyć w formie linku lub dokumentu.

.....
.....

- B.** Kto jest odpowiedzialny za opracowanie, wdrożenie i utrzymanie systemu?

.....
.....

- C.** Kto jest odpowiedzialny za podjętą w ramach procesu decyzję?

.....
.....

- D.** Czy proces pozwala na przedstawienie uzasadnienia dla podejmowanej decyzji?

T N

- E.** Czy system jest w pełni zautomatyzowany, czy też jest nadzorowany przez człowieka?

T N

- F.** W przypadku nadzoru człowieka nad systemem: czy osoba nadzorująca system ma kompetencje do zakwestionowania decyzji systemu?

T N

- G.** Czy przewidziany jest mechanizm pozyskiwania informacji zwrotnej od użytkowników w trakcie funkcjonowania systemu?

T N





H. Czy przewidziany jest mechanizm kwestionowania sugerowanej decyzji i/lub odwołania od decyzji przez obywatela?

 T N

I. Jakie zastosowano metody komunikacji z użytkownikami/osobami objętymi działaniem systemu, w przypadku nieprawidłowości?

.....

.....

.....

.....

J. Czy kod źródłowy systemu będzie ogólnodostępny?

 T N

K. Czy ogólnodostępne będą inne niż kod źródłowy informacje o systemie, np. jego cele, zasady działania?

 T N

L. Czy i w jaki sposób przewidziano audyty systemu?

.....

.....

.....

.....

M. Czy wyniki audytów będą udostępnione publicznie?

 T N

N. Czy istnieje dokumentacja zarządzania jakością wykorzystanych danych?

 T N

O. Czy istnieje dokumentacja dotycząca bezpieczeństwa, w tym cyberbezpieczeństwa, systemu?

 T N



9. DANE OSOBOWE?

A. Czy system zakłada wykorzystanie danych osobowych?

.....
.....
.....

B. Czy system zakłada wykorzystanie szczególnych kategorii danych osobowych w rozumieniu RODO?

.....
.....
.....

C. Kto kontroluje prawidłowość wykorzystania danych?

.....
.....
.....

D. Z jakich źródeł system będzie wykorzystywał dane osobowe?

.....
.....
.....

E. Czy system będzie bezpośrednio wykorzystywał dane pobrane z urzędzeń należących do obywateli bądź podmiotów gospodarczych? Jeśli tak, jakie i z jakich źródeł?

.....
.....
.....



10. PODMIOTY, NA KTÓRE ODDZIAŁUJE PROJEKT

GRUPA	WIELKOŚĆ	ŹRÓDŁO DANYCH	ODDZIAŁYWANIE
.....
.....
.....

11. CZY NARZĘDZIE BYŁO KONSULTOWANE Z INNYMI INSTYTUCJAMI PUBLICZNYMI ORAZ PODMIOTAMI ZEWNĘTRZNYMI?

Jeśli tak to z kim, w jakim zakresie i jakie były wnioski z konsultacji?

.....

.....

.....

.....

.....

12. KTO BĘDZIE WYKONYWAŁ NARZĘDZIE?

A Pracownicy instytucji

B Podmiot zewnętrzny

- Jeśli tak, w jakim trybie zostanie zamówione narzędzie?
(np. przetarg nieograniczony, zamówienie z wolnej ręki)?

.....

.....

C Narzędzie jest budowane na bazie dotychczas istniejących produktów. Jakich?

.....

.....



13. W JAKI SPOSÓB I KIEDY NASTĄPI EWALUACJA EFEKTÓW NARZĘDZIA ORAZ JAKIE MIERNIKI ZOSTANĄ ZASTOSOWANE?

.....

.....

.....

.....

.....

14. ZAŁĄCZNIKI

(istotne dokumenty źródłowe, badania, analizy itp.)

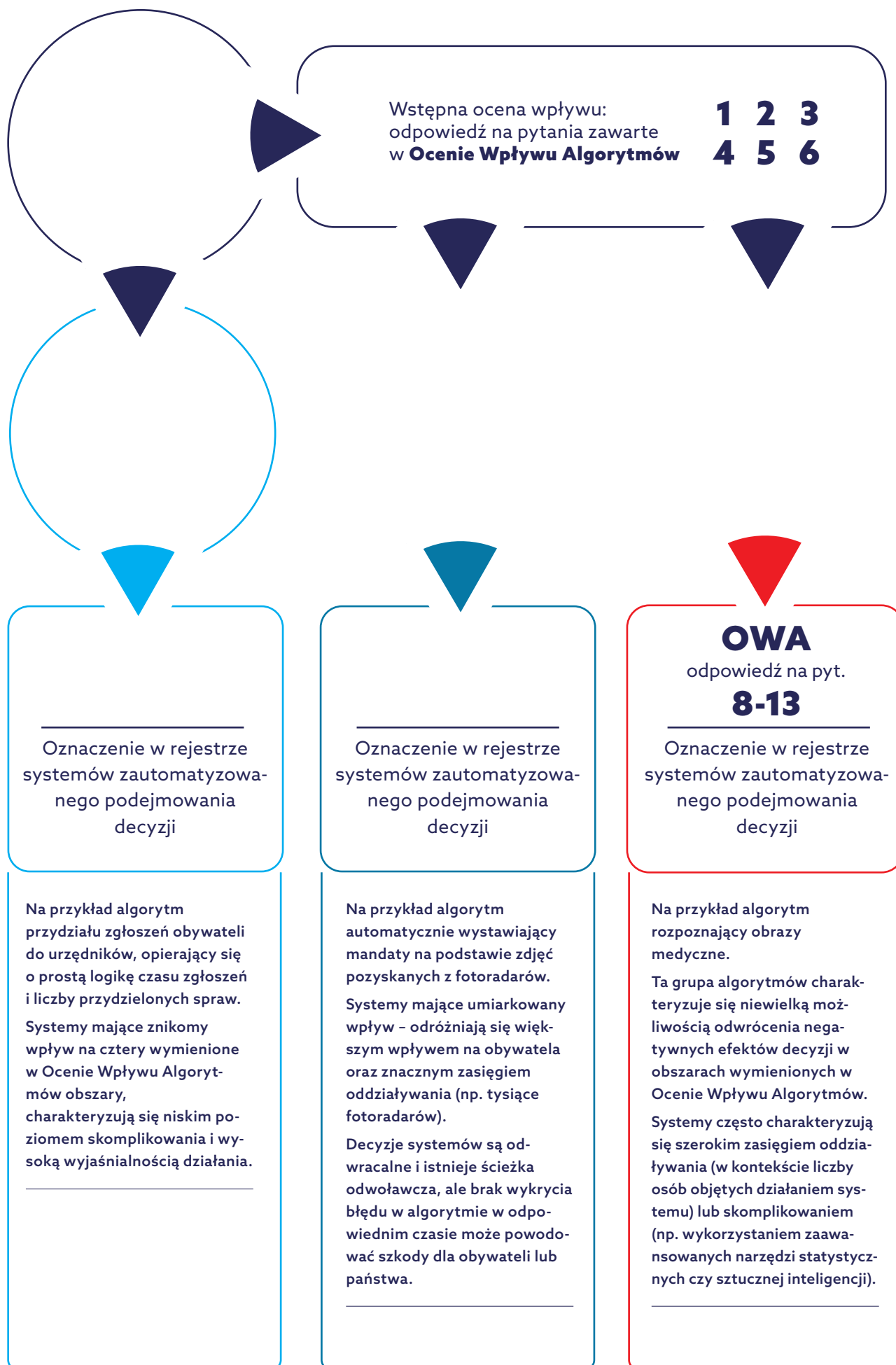
.....

.....

.....

.....

.....



POZIOM 1

POZIOM 2

POZIOM 3



MONITORING	Konieczność stałego monitorowania działania systemu i wyników jego pracy		
TWORZENIE I SZKOLENIE UŻYTKOWNIKÓW		Dokumentacja funkcjonalności i designu (jeśli możliwe) systemu	Dokumentacja designu i funkcjonalności Szkolenie użytkowników Weryfikacja szkolenia
TESTOWANIE	Konieczność przeprowadzenia i udokumentowania testów systemu przed wprowadzeniem go w życie.		
PILOTAŻ	Konieczność przeprowadzenia i udokumentowania wyników pilotażu z udziałem osób spoza grona wewnątrzinstytucjonalnych użytkowników systemu (urzędników). W zależności od systemu pilotaż powinien uwzględniać udział obywateli (np. w przypadku e-usługi), ekspertów (np. lekarzy - w przypadku algorytmu medycznego) lub innych użytkowników (np. policjantów czy aptekarzy - w przypadku systemów wspomagających dane branże).		
CZŁOWIEK W PROCESIE	Decyzje nie wymagają bezpośredniego zaangażowania użytkownika (np. urzędnika).		Proces powinien uwzględnić konkretne momenty, w których człowiek podejmuje decyzje; ostateczna decyzja podejmowana przez użytkownika
OGÓLNO-DOSTĘPNA INFORMACJA, WYRAŻONA W JĘZYKU ZROZUMIAŁYM DLA OBYWATELA	Wyjaśnienie znaczenia najczęstszych decyzji i odpowiedź na pytania (np. FAQ) oraz informacja o możliwości odwołania.	Informacja o działaniu systemu oraz możliwości odwołania. Wyjaśnienie znaczenia indywidualnej decyzji na żądanie obywatela.	Informacja o działaniu systemu oraz ogólnodostępna dokumentacja systemu. Wyjaśnienie znaczenia i podstaw każdej decyzji odmawiającej usług, ograniczającej prawa oraz poinformowanie o możliwości odwołania.
OGÓLNO-DOSTĘPNY AUDYT	Audyt narzędzia przeprowadzony przynajmniej przez jedną osobę/instytucję wskazaną poniżej.: <ul style="list-style-type: none"> • eksperta o odpowiednich kwalifikacjach z urzędu lokalnego lub centralnego, innego niż urząd wdrażający, naukowca lub eksperta pracującego w NGO z odpowiednim doświadczeniem; • niezależną organizację lub firmę z odpowiednim doświadczeniem; • certyfikowanego audytora - urzędowy organ utworzony w celu audytu algorytmów. Wraz z obowiązkiem publikacji wyników audytu.		Audyt narzędzia przeprowadzony przez przynajmniej dwie różne osoby/instytucje wymienione po lewej. Wraz z obowiązkiem publikacji wyników audytu.
ZATWIERDZENIE SYSTEMU DO DZIAŁANIA	min. Dyrektor departamentu		min. Podsekretarz stanu, zastępca prezydenta, wójta, burmistrza

ŹRÓDŁA

- European Commission, Joint Research Centre (JRC) (2021): Selected AI cases in the public sector. European Commission, Joint Research Centre (JRC) [Dataset] PID: <http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a> (dostęp: 5.12.2022 r.)
- <https://ai.hel.fi/en/ai-register/> (dostęp: 21.12.2022 r.)
- <https://algoritmeregister.amsterdam.nl/en/ai-register/> (dostęp: 21.12.2022 r.)
- <https://ai.hel.fi/en/get-to-know-ai-register/> (dostęp: 21.12.2022 r.)
- <https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-IMCO-LIBE-Report-All-Amendments-14-June.pdf> (dostęp: 12.12.2022 r.)
- <https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-IMCO-LIBE-Report-All-Amendments-14-June.pdf> (dostęp: 12.12.2022 r.)
- <https://www.coe.int/en/web/artificial-intelligence/cahai> (dostęp: 23.01.2023 r.)
- <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review> (dostęp: 5.12.2022 r.)
- https://www.echr.coe.int/documents/convention_pol.pdf (dostęp: 20.01.2023 r.)
- <https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/> (dostęp: 12.12.2022 r.)
- <https://edri.org/wp-content/uploads/2022/05/Obligations-on-users-AIA-Amendments-17022022.pdf> (dostęp: 20.01.2023 r.)
- https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0012.02/DOC_1&format=PDF (dostęp: 12.12.2022 r.)
- <https://mojepanstwo.pl/aktualnosci/773> (dostęp: 5.12.2022 r.)
- <https://privacyweek.it/event/assessing-the-ai-that-assesses-us-the-work-of-civil-society-and-research-in-setting-auditing-standards-for-the-public-sector/> (dostęp: 19.12.2022 r.)
- <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d> (dostęp: 23.01.2023 r.)
- <https://www.boe.es/buscar/act.php?id=BOE-A-2015-10566> (art. 41) (dostęp: 19.12.2022 r.)
- <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html> (dostęp: 19.12.2022 r.)
- <https://www.coe.int/en/web/artificial-intelligence/cai#> (dostęp: 23.01.2023 r.)
- <https://www.hackneycitizen.co.uk/2019/10/30/town-hall-drops-pilot-programme-profiling-families-without-their-knowledge/> (dostęp: 5.12.2022 r.)
- <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms> (dostęp: 19.12.2022 r.)
- <https://www.gov.pl/web/govtech/polityka-rozwoju-ai-w-polsce-przyjeta-przez-rade-ministrow--co-dalej> (dostęp: 19.12.2022 r.)
- <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appA> (patrz 6.1 dyrektywy) (dostęp: 19.12.2022 r.)
- <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appA> (patrz art. 9 dyrektywy) (dostęp: 19.12.2022 r.)
- <https://www.uu.nl/en/news/dutch-house-of-representatives-endorses-mandatory-use-of-human-rights-and-algorithms-impact> (dostęp: 20.01.2023 r.)
- Projekt Aktu w sprawie sztucznej inteligencji przedstawiony w kwietniu 2021 r. przez KE zawiera w art. 29 pewne obowiązki dla użytkowników systemów sztucznej inteligencji wysokiego ryzyka. (dostęp: 12.12.2022 r.)
- Więcej informacji o ADM w Europie w raporcie AlgorithmWatch "Automating Society Report 2020" <https://automatingsociety.algorithmwatch.org/> (dostęp: 20.01.2023 r.)



**Fundacja
Moje Państwo**