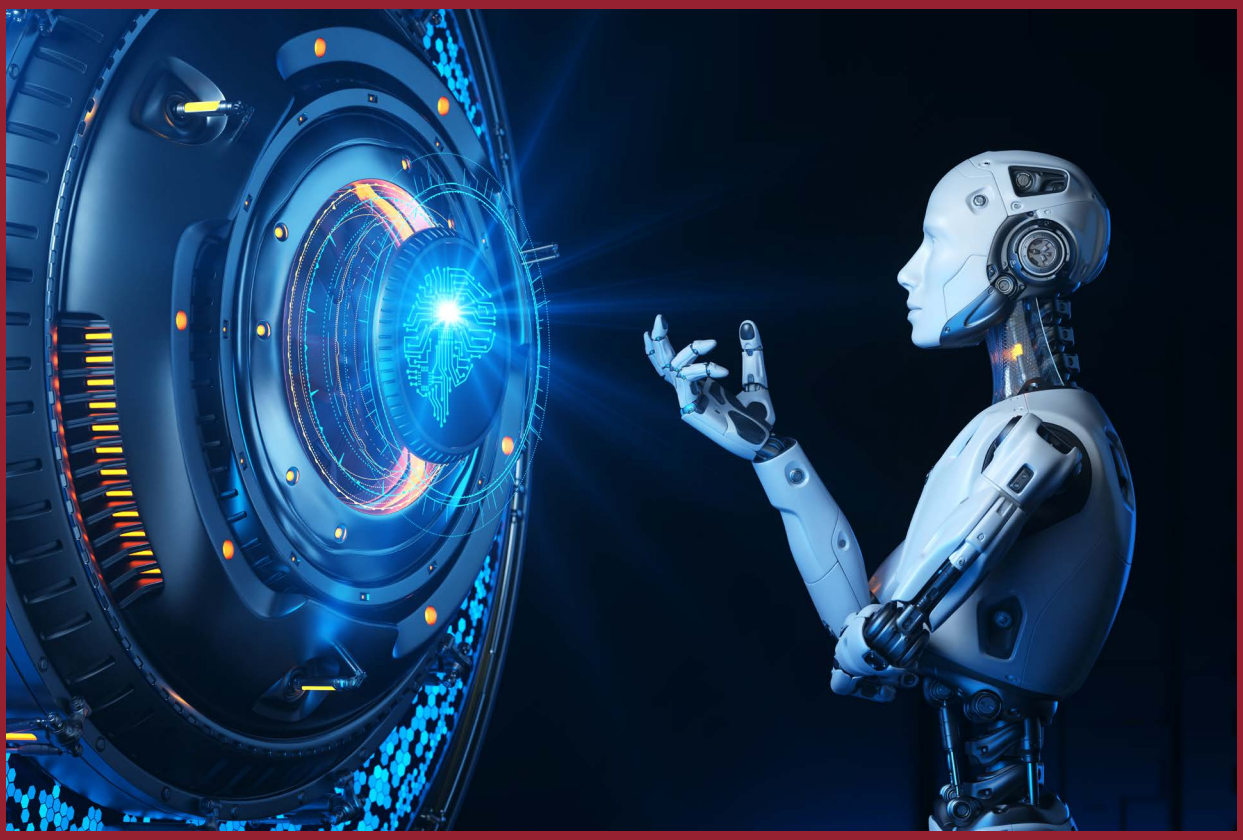
sieć obywatelska
WATCHDOG

European
Artificial Intelligence
& Society Fund

CZY ADMINISTRACJA PUBLICZNA KORZYSTA Z AI W SPOSÓB ODPOWIEDZIALNY I PRZEJRZYSTY?

STAN NA NA 21.04.2026



Działamy na rzecz Jawności!

Czy administracja publiczna korzysta z AI w sposób odpowiedzialny i przejrzysty?

Autor: Michał Zemełka
Korekta i redakcja: Martyna Bójko
Skład: Autofokus

sieć obywatelska

WATCHDOG

Spis treści

Wprowadzenie

Prezentujemy omówienie wyników **drugiej fazy badania** odpowiedzialnego wykorzystania sztucznej inteligencji przez administrację publiczną (z wynikami I fazy możesz zapoznać się tutaj). Spośród uprzednio zidentyfikowanych przypadków wytypowaliśmy potencjalnie szczególnie interesujące do pogłębionego badania, w ramach którego sprawdzaliśmy poziom przejrzystości w odniesieniu do systemów AI stosowanych w administracji publicznej. Zastanawiało nas przy tym, na ile poszczególne organy spełniają bądź są gotowe na spełnienie przyszłych wymogów wynikających z AI Act . Badanie obejmowało wysyłkę wniosków o informację oraz odniesienie otrzymanych odpowiedzi do informacji ogólnodostępnych. Zakres i charakter danych, o jakie wnosiliśmy w kontekście systemów AI wykorzystywanych w danej instytucji, wynikały zarówno z przepisów AI Act (przykładowo identyfikacja poziomu ryzyka systemu AI jest niezbędna, żeby wiedzieć, jakie przepisy w danym przypadku znajdują zastosowanie), jak i danych zbieranych w słoweńskim rejestrze systemów AI, już wcześniej przygotowanym przez organizację *Danes je nov dan*, jednego z partnerów w projekcie.

W ramach międzynarodowego projektu realizowanego wspólnie z K-Monitor oraz Danes Je Nov Dan podejmujemy działania na rzecz zwiększenia przejrzystości wykorzystania sztucznej inteligencji (AI) i systemów algorytmicznych (ADM) w administracji publicznej. Projekt koncentruje się na monitorowaniu sposobów wdrażania i używania tych technologii przez instytucje publiczne, analizie konkretnych przypadków ich zastosowania oraz opracowywaniu rekomendacji dotyczących przejrzystości i społecznej kontroli nad procesami decyzyjnymi wspieranymi przez technologie cyfrowe.

Niniejszy raport jest wynikiem monitoringu prowadzonego w ramach projektu i przedstawia ustalenia dotyczące wykorzystywania AI oraz systemów algorytmicznych przez administrację publiczną. Istotnym elementem projektu są również wspólne działania rzecznicze oraz budowanie sieci współpracy organizacji obywatelskich zainteresowanych tematyką odpowiedzialnego i transparentnego wykorzystywania AI przez państwo. Naszym celem jest wzmacnianie społecznej kontroli nad technologiami, które coraz częściej wpływają na funkcjonowanie administracji i decyzje podejmowane wobec obywateli.

Wniosek o informację

Szanowni Państwo!
W ramach międzynarodowego projektu badawczego AI Transparency in Public Administration in the CEE region and beyond zbieramy informacje dotyczące stosowania sztucznej inteligencji przez podmioty publiczne.
Polska, tak jak inne kraje Unii Europejskiej, znajduje się obecnie w procesie wdrażania przepisów, wynikających z rozporządzenia zwanego Artificial Intelligence Act (AI Act). Regulacja ta nakłada określone obowiązki na podmioty publiczne, stosujące systemy informatyczne wspierane sztuczną inteligencją - tym większe, im wyższe potencjalne ryzyko związane z ich stosowaniem. Naszym celem jest zbadanie poziomu przejrzystości w odniesieniu do ww. systemów oraz potencjalną gotowość organów do spełniania wymogów ww. rozporządzenia, które zaczną obowiązywać w nieodległej przyszłości.
W związku z powyższym i na podstawie ustawy o dostępie do informacji publicznej, wnosimy o **wypełnienie podlinkowanej ankiety osobno** w odniesieniu do wszystkich systemów informatycznych i aplikacji, które są wykorzystywane pośrednio lub bezpośrednio przez organ, jednostki mu podległe lub na zlecenie organu i w których stosowane są elementy szeroko rozumianej **sztucznej inteligencji**, a jednocześnie pośrednio lub bezpośrednio wspierają **procesy decyzyjne**.

1. Nazwa systemu AI
2. Przeznaczenie systemu - krótko, np. rozpoznawanie twarzy.
3. Opis systemu - do czego ogólnie cały system służy, jak działa, jakie procesy wspiera?
4. Użycie sztucznej inteligencji - do czego w tym systemie i we wspieranych przez system procesach są używane komponenty oparte na sztucznej inteligencji?
5. Kwalifikacja ryzyka - czy przeprowadzono kwalifikację ryzyka w odniesieniu do poziomów określonych w AI Act? Prosimy o wskazanie jednej właściwej odpowiedzi:
   a) Nie przeprowadzono
   b) Ryzyko minimalne
   c) Ryzyko ograniczone
   d) Ryzyko wysokie
   e) Ryzyko niedopuszczalne
6. Ocena skutków dla ochrony danych (DPIA) - czy przeprowadzono ocenę skutków dla ochrony danych (DPIA), wynikającą z przepisów RODO? Jeśli odpowiedź brzmi tak, to wnosimy o udostępnienie publicznego linku.
7. Ocena skutków dla praw podstawowych (FRIA) - czy przeprowadzono ocenę skutków dla praw podstawowych (FRIA)? Jeśli odpowiedź brzmi tak, to wnosimy o udostępnienie publicznego linku. Zdajemy sobie sprawę, że oficjalny unijny kwestionariusz nie został jeszcze ogłoszony, jednak być może skorzystali Państwo z nieoficjalnych lub zastosowali własną metodę. Prosimy o wskazanie i ewentualne uzupełnienie jednej właściwej odpowiedzi:
   a) Nie dotyczy, bo zidentyfikowany dla systemu poziom ryzyka tego nie wymaga.
   b) Nie, ponieważ nie ogłoszono jeszcze kwestionariusza.
   c) Nie, z innych powodów...
   d) Tak, link:...
8. Podręcznik użytkownika / instrukcja - czy użytkownicy systemu dysponują jakimś rodzajem podręcznika lub instrukcji? Jeśli tak, to wnosimy o podanie linku lub linków do publicznego zasobu z tego typu dokumentami.
9. Publiczna wiedza o systemie - czy wedle wiedzy organu w publicznych zasobach internetu dostępne są jakiekolwiek wzmianki, opisy, dokumenty związane z systemem inne, niż wskazane w poprzednich pytaniach? Jeśli tak, to wnosimy o podanie linków.
10. Modele ogólnego przeznaczenia (GPAI) - czy system korzysta z modeli AI ogólnego przeznaczenia? Jeśli tak, to wnosimy o wyliczenie nazw modeli i ich producentów.
11. Dostawca systemu
12. Koszty wytworzenia i wdrożenia - całkowite łączne koszty brutto ponoszone po stronie organu
13. Koszty utrzymania - koszty brutto ponoszone po stronie organu w ujęciu rocznym
14. Początek obowiązywania licencji
15. Koniec obowiązywania licencji

Uściślenie
Oryginalna wersja wniosku zamiast pytań zawierała link do formularza ankiety online (zakładaliśmy, że tak może być wygodniej), jednak niektóre instytucje odmówiły udzielenia odpowiedzi w takiej formie (do czego jeszcze wrócimy). Wtedy dosyłaliśmy im pytania brzmiące tak, jak zaprezentowano powyżej. W jednym przypadku - konkretnie chodzi o Ośrodek Przetwarzania Informacji - ograniczyliśmy wniosek do systemów wytworzonych przy udziale pytanej instytucji.

Kogo pytaliśmy i czy udzielono informacji?

Wnioski wysłaliśmy do poniższych **52** instytucji. Oznaczamy, czy dana instytucja odpowiedziała ostatecznie na nasze pytania w kontekście przynajmniej jednego z używanych przez nią lub na jej zlecenie systemów AI. Czerwonym kolorem wyróżnione są nazwy instytucji, które - jak się wydaje - zignorowały wniosek.

	Nazwa	Czy ujawniono informacje o AI
1.	Agencja Restrukturyzacji i Modernizacji Rolnictwa	TAK
2.	Gdyńskie Centrum Informatyki	TAK
3.	Główny Urząd Geodezji i Kartografii	TAK
4.	Kancelaria Sejmu RP	TAK
5.	Komenda Wojewódzka Policji w Białymstoku	TAK
6.	Komenda Wojewódzka Policji w Szczecinie	TAK
7.	Miejskie Wodociągi i Kanalizacja w Bydgoszczy - Sp. z o.o.	TAK
8.	Ministerstwo Finansów	TAK
9.	Ministerstwo Rozwoju i Technologii	TAK
10.	Regionalna Dyrekcja Lasów Państwowych w Pile	TAK
11.	Sąd Rejonowy w Strzelcach Opolskich	TAK
12.	Uniwersytet Szczeciński	TAK
13.	Urząd Gminy Lipnica Wielka	TAK
14.	Urząd Gminy Rokietnica	TAK
15.	Urząd Gminy w Kaźmierzu	TAK
16.	Urząd Miasta Katowice	TAK
17.	Urząd Miasta Krakowa	TAK
18.	Urząd Miasta Poznania	TAK
19.	Urząd Miasta Warszawa	TAK
20.	Urząd Miasta Włocławek	TAK
21.	Urząd Miasta Zgierza	TAK
22.	Urząd Miejski Gminy Rawicz	TAK
23.	Urząd Miejski w Boguchwale	TAK
24.	Urząd Miejski w Grodzisku Mazowieckim	TAK
25.	Urząd Ochrony Konkurencji i Konsumentów	TAK
26.	Urząd Patentowy Rzeczypospolitej Polskiej	TAK
27.	Wojewódzki Inspektorat Ochrony Roślin i Nasiennictwa w Opolu	TAK
28.	Zakład Ubezpieczeń Społecznych	TAK
29.	Akademia Wychowania Fizycznego Józefa Piłsudskiego w Warszawie	NIE
30.	Dowództwo Komponentu Wojsk Obrony Cyberprzestrzeni	NIE
31.	Instytut Badawczy Leśnictwa	NIE
32.	Komenda Wojewódzka Policji w Katowicach	NIE
33.	Miejskie Wodociągi i Kanalizacja Spółka z o.o. w Koszalinie	NIE
34.	Ministerstwo Cyfryzacji	NIE
35.	Ministerstwo Klimatu i Środowiska	NIE
36.	Ministerstwo Nauki i Szkolnictwa Wyższego	NIE
37.	Ministerstwo Obrony Narodowej	NIE
38.	Ministerstwo Spraw Zagranicznych	NIE
39.	Ośrodek Przetwarzania Informacji	NIE
40.	Sąd Okręgowy w Katowicach	NIE
41.	Sąd Okręgowy w Rybniku	NIE
42.	Starostwo Powiatowe w Mikołowie	NIE
43.	Straż Miejska w Opolu	NIE
44.	Uniwersytet Komisji Edukacji Narodowej w Krakowie	NIE
45.	Urząd Miasta Siemianowice Śląskie	NIE
46.	Urząd Miasta Stalowej Woli	NIE
47.	Urząd Miasta Szczecin	NIE
48.	Urząd Miasta Torunia	NIE
49.	Urząd Miejski w Białymstoku	NIE
50.	Urząd Miejski w Chojnicach	NIE
51.	Urząd Miejski w Dąbrowie Górniczej	NIE
52.	Urząd Miejski Wrocławia	NIE

Odwrót od jawności

Druga faza badania – zakres i przebieg

Po badaniu przeprowadzonym **latem 2025** roku, obejmującym blisko 5,5 tysiąca instytucji, w którym pytaliśmy bardziej ogólnie o korzystanie z AI, wytypowaliśmy **52** podmioty do pogłębionej analizy. **1 października 2025** roku skierowaliśmy do wybranych instytucji wniosek zawierający bardziej szczegółowe pytania dotyczące wykorzystywania systemów AI. We wniosku celowo nie wskazywaliśmy, o jakie konkretnie systemy informatyczne nam chodzi - pomimo że to one właśnie były głównym kryterium selekcji kandydatów do tej fazy badania - ponieważ chcieliśmy sprawdzić, czy wybrane instytucje będą z nami tak samo szczere, jak poprzednio. Przeczucie nas nie myliło: pomimo że pytaliśmy *de facto* o te same systemy, okazało się, że uzyskanie odpowiedzi będzie tym razem znacznie trudniejsze. Konieczne było ponaglanie instytucji, ponowne przesyłanie pytań w innej formie i podejmowanie dodatkowych prób kontaktu. Ostatecznie odpowiedzi odnoszących się do używanych systemów AI udzieliło jedynie **28** instytucji, czyli niewiele więcej niż połowa zapytanych — dokładnie 54%.

Gdzie zniknęły systemy AI?

W pozostałych przypadkach mieliśmy do czynienia między innymi z milczeniem lub zasłanianiem się tajemnicą, do czego jeszcze wrócimy. Ponadto jednak pojawiło się zjawisko szczególnie zastanawiające: otóż systemy, które jeszcze latem 2025 roku — według deklaracji samych instytucji — **były systemami AI**, jesienią 2025 roku w odpowiedziach tych samych instytucji **przestały nimi być**! Dla porządku trzeba odnotować teoretycznie możliwe wyjaśnienie: część tych rozwiązań mogła po prostu przestać być używana. Żadna z instytucji nie zadeklarowała jednak tego wprost (najbliżej było Ministerstwo Klimatu i Środowiska). Wyjątkiem w zakresie konsekwencji w nieudzielaniu informacji było Ministerstwo Nauki i Szkolnictwa Wyższego, które w obu badaniach nie deklarowało wykorzystywania systemów AI, do czego jeszcze wrócimy.

Jak już wspomniano, wiele odpowiedzi sprowadzało się do stwierdzenia, że dana instytucja nie korzysta z systemów AI. W przypadkach szczególnie dla nas istotnych konfrontowaliśmy te deklaracje z odpowiedziami z poprzedniej fazy badania, w których te same instytucje wskazywały konkretne wykorzystywane systemy AI. Wnosiliśmy wówczas o wyjaśnienie, dlaczego rozwiązania te nagle przestały być uznawane za systemy AI, ewentualnie prosiliśmy o odpowiedź na nasze pytania niezależnie od przyjętej klasyfikacji danego systemu. Takie działania przynosiły różne rezultaty. Jedno pozostawało jednak niezmienne: korespondencja była często zarazem ciekawa i pouczająca. Poszczególne przypadki szerzej omówimy w dalszej części raportu.

Problemy z formą wniosku

Trzeba zaznaczyć, że zastosowana przez nas inicjalnie forma ankiety online (przesyłana w postaci linku w wiadomości przewodniej) — choć ustawowo dopuszczalna — mogła dla części instytucji wydawać się problematyczna z perspektywy bezpieczeństwa. Zgodnie z art. 14 ust. 1 ustawy o dostępie do informacji publicznej, udostępnianie informacji publicznej na wniosek

następuje w sposób i w formie zgodnych z wnioskiem. Jednocześnie ust. 2 ww. artykułu stanowi, że jeżeli informacja publiczna nie może zostać udostępniona w sposób lub w formie określonych we wniosku, podmiot obowiązany do jej udostępnienia powinien pisemnie poinformować wnioskodawcę o przyczynach takiego stanu rzeczy oraz wskazać, w jaki sposób lub w jakiej formie informacja może zostać udostępniona niezwłocznie.

Dlatego gdy otrzymywaliśmy sygnały, że formularz może stanowić problem, przesyłaliśmy pytania z ankiety e-mailem, prosząc o odpowiedź tym samym kanałem. W kilku przypadkach robiliśmy to nawet z własnej inicjatywy, mimo braku wyraźnej informacji o trudnościach z wypełnieniem formularza. Nie zmienia to jednak ogólnej oceny sytuacji. Nasz pierwotny wniosek o udzielenie odpowiedzi za pośrednictwem formularza online nie usprawiedliwia ani całkowitego braku odpowiedzi, ani deklaracji braku systemów AI, ani odmowy udzielenia informacji. Tymczasem właśnie z takimi reakcjami spotkaliśmy się w **24 z 52** przypadków. Warto przypomnieć, że niemal każda z tych instytucji jeszcze cztery miesiące wcześniej deklarowała wykorzystywanie systemów AI — i to na tyle interesujących, że uznaliśmy za zasadne dopytanie o szczegóły. Co więc zmieniło się przez ten czas?

Opieszałość w odpowiedzi

Pierwotne wnioski w tej sprawie wysłaliśmy **1 października 2025** roku. Mimo podstawowego 14-dniowego terminu na odpowiedź, korespondencja — często po licznych ponagleniach — spływała jeszcze **pod koniec 2025**, a ostatnią - jak do tej pory - odpowiedź otrzymaliśmy pod koniec **lutego 2026** roku. Była to odpowiedź Ministerstwa Spraw Zagranicznych — odmowna. Ze względu jednak na to, że pierwotnie wnioskowaliśmy o odpowiedź poprzez formularz online, co mogło budzić pewne wątpliwości po stronie adresatów, nie będziemy szczegółowo wskazywać instytucji, które przekroczyły ustawowy termin odpowiedzi.

Uzasadnienia zmiany stanowiska

ZUS

W przypadku Zakładu Ubezpieczeń Społecznych instytucja przedstawiła rozbudowane stanowisko dotyczące rozumienia pojęcia systemu AI, zakończone konkluzją, że w Zakładzie nie są stosowane systemy spełniające definicję wynikającą z AI Act. W odpowiedzi na to napisaliśmy, że niezależnie od przyjętej kwalifikacji oczekujemy odpowiedzi na pytania zawarte we wniosku w odniesieniu do narzędzia opartego na modelu predykcyjnym, wykorzystywanego przez organ do systemowej analizy danych z wystawionych zaświadczeń lekarskich o czasowej niezdolności do pracy. O tym, że taki system AI jest wykorzystywany w organie, wiedzieliśmy z odpowiedzi ZUS-u udzielonej w pierwszej fazie badania. W efekcie ZUS dwukrotnie przedłużał termin udzielenia odpowiedzi, najpierw do 5 grudnia 2025 r., a następnie do 19 grudnia 2025 r. Ostatecznie instytucja odniosła się do wskazanego systemu predykcyjnego, zwanego - jak się okazało - *SOS AI*. Jednocześnie podtrzymała swoje nowe stanowisko, że **rozwiązanie to nie stanowi jednak systemu AI**, przedstawiając w tym zakresie obszerne uzasadnienie.

Ministerstwa

Ministerstwo Klimatu i Środowiska w pierwszym etapie korespondencji (tj. jeszcze w czerwcu 2025) wskazało, że wykorzystuje system informatyczny *Testportal*, oparty na sztucznej

inteligencji w zakresie generowania pytań do testów selekcyjnych w procesie rekrutacji. Z przekazanych informacji wynikało, że system służy do tworzenia treści pytań testowych na podstawie dostarczonych materiałów źródłowych, takich jak akty prawne, zarządzenia czy dane statystyczne. Dlatego w kolejnej fazie badania, po dłuższej wymianie korespondencji (ponaglanie, wyjaśnianie kwestii formalnych, itd., itp.), zwróciliśmy się o udzielenie informacji w odniesieniu do systemu *Testportal*. W odpowiedzi Ministerstwo niespodziewanie stwierdziło, że nie dysponuje i nie używa systemu o tej nazwie, nie odnosząc się przy tym do przeszłości. Nie wiemy więc, co jest źródłem niespójności w odpowiedziach Ministerstwa.

W przypadku Ministerstwa Nauki i Szkolnictwa Wyższego, po otrzymaniu odpowiedzi, w której urząd nie odniósł się do znanego nam skądinąd *Jednolitego Systemu Antyplagiatowego* (dalej *JSA*), zapytaliśmy o powody jego pominięcia. W odpowiedzi Ministerstwo wyjaśniło, że nie uwzględniło *JSA*, ponieważ nie wykorzystuje go w procesie decyzyjnym, ani bezpośrednio, ani pośrednio. Wskazano również, że system jest udostępniany podmiotom szkolnictwa wyższego i nauki, natomiast dostępna w jego ramach funkcja analizy wspartej AI ma charakter opcjonalny. Warto przy tym zauważyć, że wytwórcą tego systemu jest Ośrodek Przetwarzania Informacji, od którego nie odnotowaliśmy żadnej odpowiedzi.

Ministerstwo Spraw Zagranicznych początkowo wskazało, że nie korzysta z systemów spełniających definicję systemów AI w rozumieniu AI Act. W odpowiedzi na to zwróciliśmy uwagę, że w odrębnej korespondencji z 6 czerwca 2025 r. Ministerstwo informowało o wdrożeniu narzędzi wykorzystujących mechanizmy sztucznej inteligencji do bieżącej pracy w niektórych obszarach, w tym w zakresie problematyki sankcyjnej. Po skierowaniu pytania doprecyzowującego Ministerstwo poinformowało o konieczności przeprowadzenia dodatkowej analizy technicznej oraz formalnoprawnej i zapowiedziało udzielenie odpowiedzi w późniejszym terminie. Następnie termin załatwienia sprawy był przedłużany, a w toku postępowania skierowano również wezwanie do uzupełnienia braków formalnych. Ostatecznie sprawa zakończyła się odmową udzielenia informacji, jako *objętych tajemnicą dyplomatyczną, bo ich ujawnienie mogłoby szkodzić polityce zagranicznej Rzeczypospolitej Polskiej i naruszać jej wizerunek międzynarodowy*. Warto przy tym odnotować przewlekłość postępowania: od złożenia wniosku 1 października 2025 r. do wydania odmowy 20 lutego 2026 r. **upłynęło niemal pięć miesięcy**, mimo że chodziło o podstawowy tryb dostępu do informacji publicznej, a nie o postępowanie odwoławcze.

Sądy

Sąd Okręgowy w Rybniku poinformował, że nie posiada systemu AI i nie stosuje sztucznej inteligencji. W odpowiedzi na to zwróciliśmy uwagę, że we wcześniejszej korespondencji z 27 czerwca 2025 r. organ wymieniał rozwiązania takie jak *SAWA* oraz *LEX Kara łączna* w kontekście systemów wspomaganych sztuczną inteligencją i w związku z tym zapytaliśmy, czy oznacza to, że żadne z tych rozwiązań, pomimo deklaracji z czerwca, nie zawiera jednak elementów kwalifikujących oprogramowanie do którejkolwiek z kategorii systemów AI zdefiniowanych w AI Act. W odpowiedzi Sąd podtrzymał swoje wcześniejsze stanowisko, stwierdzając krótko, że nie posiada systemu AI i nie stosuje sztucznej inteligencji. Skąd się wzięła w takim razie na przestrzeni kilku miesięcy różnica w interpretacji? Nie wiemy, Skąd wzięła się zatem różnica w interpretacji w ciągu zaledwie kilku miesięcy?

Sąd Okręgowy w Katowicach podobnie oświadczył, że nie wykorzystuje żadnych systemów wspieranych lub w całości opartych na szeroko rozumianej sztucznej inteligencji, które pośrednio lub bezpośrednio wspierałyby procesy decyzyjne. W odpowiedzi przyznaliśmy, że systemy *Wirtualny Agent i Czatbot z funkcją Wirtualnego Asystenta*, o których wiedzieliśmy z czerwcowej korespondencji, faktycznie nie wpływają na procesy decyzyjne po stronie organu, jednak niezależnie od tego prosimy o udzielenie odpowiedzi w ich kontekście, bo nadal są to systemy AI. Ponadto zastanawialiśmy się w naszym piśmie, czy wykorzystywany w Sądzie do przetwarzania mowy na tekst program *Newton Dictate 5*, nie wpływa pośrednio na decyzje. Wszak błędy generowane na etapie transkrypcji mogą pośrednio wpływać na dalszą pracę z przygotowanym materiałem, a tym samym na treść podejmowanych rozstrzygnięć. W związku z tym poprosiliśmy o udzielenie odpowiedzi również w odniesieniu do tego systemu. Na to dostaliśmy pismo, w którym Sąd stwierdza, że udzielił już odpowiedzi na wniosek i że nie ma potrzeby uzupełniania udzielonej informacji. Ponadto zostaliśmy pouczeni, że w przypadku wielokrotnie ponawianych wniosków o udostępnienie tej samej informacji przez tego samego wnioskodawcę, na adresacie wniosku nie spoczywa obowiązek jej udzielenia po raz kolejny. Nie był to jednak ten sam wniosek, ponieważ zmieniliśmy jego zakres.

Samorządy

Urząd Miejski w Chojnicach odpisał, że wypełnianie ankiet nie pozostaje w związku z przedmiotem i trybem udostępniania informacji publicznej, a zatem nie stanowi informacji publicznej. Jednocześnie, na marginesie, urząd poinformował, że gmina nie korzysta z systemów AI w rozumieniu definicji zawartej w AI Act. W odpowiedzi zwróciliśmy uwagę na wykorzystywanie przez urząd programu *WPF Asystent*, służącego do przygotowywania dokumentów związanych z *Wieloletnią Prognozą Finansową* i zapytaliśmy, czy rozwiązanie to zostało uznane za niespełniające kryteriów żadnej z kategorii systemów AI określonych w AI Act. W odpowiedzi urząd podtrzymał stanowisko, że *WPF Asystent* nie kwalifikuje się jako system sztucznej inteligencji, powołując się przy tym na informację uzyskaną od dostawcy oprogramowania, zgodnie z którą w systemie *ePublink WPF*, dawniej *WPF Asystent*, nie są wykorzystywane mechanizmy oparte na sieciach neuronowych, uczeniu maszynowym, modelach językowych ani innych technologiach zaliczanych do szeroko rozumianej sztucznej inteligencji. Stoi to w sprzeczności zarówno z odpowiedzią tego samego urzędu w pierwszej fazie badania, jak i z odpowiedzią z Boguchwały (aczkolwiek w tym przypadku nie mamy pewności, czy chodzi o tę samą aplikację).

Urząd Miejski Gminy Rawicz początkowo poinformował, że po analizie przepisów AI Act nie stwierdzono, aby systemy informatyczne wykorzystywane przez urząd spełniały definicję systemów AI. W odpowiedzi zwróciliśmy uwagę, że we wcześniejszej korespondencji organ wskazywał na wykorzystywanie systemu monitoringu miejskiego wyposażonego w funkcję inteligentnej analizy obrazu, obejmującą rozpoznawanie tablic rejestracyjnych oraz rozpoznawanie obiektów, takich jak człowiek, rowerzysta i pojazd. Po przedłużeniu terminu odpowiedzi urząd doprecyzował, że system opiera się na kamerach Dahua, które

rozpoznają i zapisują numery rejestracyjne pojazdów, natomiast funkcja odróżniania kategorii *pojazd/osoba* nie prowadzi do generowania dodatkowych danych ani do identyfikacji osób. Organ wskazał zarazem, że w jego ocenie rozwiązanie to najprawdopodobniej nie kwalifikuje się jako system wysokiego ryzyka w rozumieniu AI Act, ponieważ nie służy do zdalnej identyfikacji biometrycznej ani nie stanowi komponentu bezpieczeństwa w zarządzaniu ruchem drogowym lub infrastrukturą krytyczną. Jednocześnie urząd poinformował, że informacje o tym systemie zostały ostatecznie wprowadzone do ankiety.

W przypadku Urzędu Miasta Stalowej Woli, mimo potwierdzenia odczytania korespondencji, nie udzielono odpowiedzi na skierowany wniosek. Po pewnym czasie wystosowaliśmy przypomnienie, a następnie wezwanie do niezwłocznego udzielenia odpowiedzi, wskazując, że wniosek dotyczy informacji odnoszących się do automatycznej analizy sytuacji nietypowych prowadzonej przez system monitoringu miejskiego. Również to pismo zostało odczytane, jednak organ ponownie nie udzielił odpowiedzi.

Urząd Miejski w Dąbrowie Górniczej początkowo wskazał, że na dzień udzielenia odpowiedzi urząd oraz jednostki podległe nie stosują systemów, które bezpośrednio lub pośrednio wspierałyby procesy decyzyjne z wykorzystaniem technologii AI. W odpowiedzi na to zwróciliśmy uwagę na system *eRecruiter*, którego niektóre elementy mogą wykorzystywać sztuczną inteligencję do wspierania pracy w obszarze analizy formularzy rekrutacyjnych i CV kandydatów, a także do tworzenia ogłoszeń o pracę oraz pytań rekrutacyjnych. Organ wyjaśnił, że system jest wykorzystywany wyłącznie jako narzędzie do obsługi napływu ofert oraz komunikacji z kandydatami i członkami komisji rekrutacyjnych, natomiast moduły AI, takie jak preselekcja czy rekomendacje, pozostają w przyjętej konfiguracji nieaktywne i nie uczestniczą w ocenie kandydatów. Wskazano również, że decyzje o dopuszczeniu do kolejnych etapów rekrutacji oraz o wyborze kandydata podejmuje wyłącznie komisja rekrutacyjna, w związku z czym system *eRecruiter* nie został ujęty w odpowiedzi jako rozwiązanie wspierające procesy decyzyjne.

W przypadku Urzędu Miejskiego w Białymstoku początkowo wskazano, że elementy szeroko rozumianej sztucznej inteligencji nie są wykorzystywane pośrednio ani bezpośrednio przez systemy stosowane przez Urząd oraz nie wspierają realizowanych w nim procesów decyzyjnych. W odpowiedzi na to zwróciliśmy uwagę na znany z pierwszej fazy badania system klasy *RPA*, wspomagający Biuro Zarządzania Efektywnością Energetyczną w analizie faktur, pytając, czy rozwiązanie to zostało uznane za niespełniające kryteriów żadnej z kategorii systemów AI określonych w AI Act, mimo jego potencjalnego wpływu na proces decyzyjny. Urząd podtrzymał swoje wcześniejsze stanowisko i wyjaśnił, że stosowany system nie spełnia przesłanek definicji systemu AI zawartej w AI Act. W uzasadnieniu wskazano, że rozwiązanie to, należące do klasy *RPA*, nie jest autonomiczne, nie wykazuje zdolności adaptacyjnych, a także nie jest zdolne do wnioskowania ani generowania na podstawie danych wejściowych wyników takich jak predykcje, treści, zalecenia lub decyzje mogące wpływać na środowisko fizyczne lub wirtualne. W konsekwencji organ uznał, że system ten nie kwalifikuje się jako system AI w rozumieniu rozporządzenia.

W przypadku Urzędu Miasta Siemianowice Śląskie, wobec braku odpowiedzi na wniosek także po uprzednim przypomnieniu, skierowaliśmy kolejne pismo wzywające do niezwłocznego udzielenia informacji publicznej dotyczącej systemów wykorzystujących elementy sztucznej inteligencji. W piśmie wskazaliśmy w szczególności na potrzebę udzielenia odpowiedzi odrębnie w odniesieniu do wskazanych w pierwszej fazie badania systemów wspierających analizę nagrań monitoringu miejskiego *Milestone XProtect Incident Manager*, *Milestone XProtect Rapid REVIEW* oraz *XProtect License Plate Recognition*. Również na to wystąpienie nie odnotowaliśmy odpowiedzi.

W przypadku Urzędu Miasta Szczecina urząd początkowo poinformował, że nie używa systemów AI w rozumieniu definicji zawartej w AI Act. W odpowiedzi na to zwróciliśmy uwagę na *System Monitoringu Miejskiego* z analizą obrazu, pytając, czy został on uznany za niespełniający kryteriów żadnej z kategorii systemów AI określonych w rozporządzeniu, mimo jego potencjalnego wpływu na procesy decyzyjne. Dodatkowo zapytaliśmy o tzw. *Wirtualną ankieterkę*, wykorzystaną na potrzeby konsultacji społecznych prowadzonych w związku z pracami nad Strategią Rozwoju Miasta, prosząc o odniesienie się również do tego rozwiązania. W odpowiedzi urząd wyjaśnił, że system monitoringu wykazuje pewne cechy systemu AI, ponieważ jest systemem maszynowym i na podstawie danych wejściowych przekazuje wyniki mogące wpływać na środowisko fizyczne, jednak w jego ocenie nie spełnia podstawowego kryterium wnioskowania. Wskazano, że system realizuje wyłącznie sztywno zapisane algorytmy i nie wykazuje zdolności adaptacyjnych, wobec czego nie został uznany za system AI w rozumieniu AI Act. Odnosząc się natomiast do *Wirtualnej ankieterki*, urząd podkreślił, że narzędzie to miało charakter edukacyjno-promocyjny, nie wpływało bezpośrednio na decyzje podmiotu opracowującego strategię i nie służyło agregowaniu wniosków. Zaznaczono również, że było to jednorazowe, prototypowe przedsięwzięcie przygotowane przez wykonawcę zewnętrznego w celu zachęcenia mieszkańców do udziału w konsultacjach społecznych. Jednocześnie urząd dopuścił możliwość, że w przyszłości podobne rozwiązania mogłyby zostać rozwinięte w kierunku pozyskiwania wniosków.

Uczelnie

W przypadku Uniwersytetu Komisji Edukacji Narodowej w Krakowie po wcześniejszej wymianie korespondencji, w której instytucja deklarowała brak wykorzystywania systemów AI, zwróciliśmy się z prośbą o wyjaśnienie, czy Uniwersytet nie korzysta z *Jednolitego Systemu Antyplagiatowego*, czy też korzysta, ale uznał, że system ten nie należy do kategorii rozwiązań wykorzystujących sztuczną inteligencję. Jednocześnie zawnioskowaliśmy, żeby w przypadku korzystania z JSA wypełniono naszą ankietę w kontekście tego systemu, niezależnie od przyjętej przez Uniwersytet klasyfikacji. Od tego momentu nie dostaliśmy już żadnej odpowiedzi.

Podobnie przebiegała korespondencja z Akademią Wychowania Fizycznego Józefa Piłsudskiego w Warszawie. Uczelnia najpierw poinformowała, że obecnie nie stosuje systemów informatycznych ani aplikacji wspieranych sztuczną inteligencją, które pośrednio lub bezpośrednio wspierają procesy decyzyjne. W odpowiedzi wskazaliśmy, że z wcześniej uzyskanych od nich informacji wynikało, iż na zlecenie Filii AWF w Białej Podlaskiej opracowano aplikację *Test FUS*, opartą na zaawansowanych technologiach przetwarzania obrazu i uczenia maszynowego. W związku z tym zwróciliśmy się o udzielenie odpowiedzi na pierwotny wniosek w odniesieniu do tego rozwiązania. Na to również nie otrzymaliśmy odpowiedzi.

Straż Miejska

W przypadku Straży Miejskiej w Opolu instytucja wskazała, że nie korzysta i nie wykorzystuje w swojej pracy sztucznej inteligencji. W odpowiedzi zwróciliśmy uwagę, że we wcześniejszej korespondencji z 7 lipca 2025 r. organ w kontekście systemów AI deklarował wykorzystanie programu *Genetec* do obsługi obrazu z kamer monitoringu, w związku z czym poprosiliśmy o doprecyzowanie, czy program ten nie jest już obecnie wykorzystywany, czy też nie są w nim używane funkcje wspierane sztuczną inteligencją. W odpowiedzi Straż Miejska podtrzymała swoje wcześniejsze stanowisko, ograniczając się do ponownego stwierdzenia, że nie korzysta i nie wykorzystuje w swojej pracy sztucznej inteligencji. Również w tym przypadku nie dowiedzieliśmy się, dlaczego system był uznawany za system AI w czerwcu, a w październiku już nie.

Co wynika z udostępnionych informacji?

Niezależnie od powyższego, ostatecznie udało się uzyskać mniej lub bardziej wyczerpujące informacje na temat 37 systemów AI wykorzystywanych przez 28 instytucji publicznych. Poniżej omawiamy obraz, jaki wyłania się z uzyskanych odpowiedzi.

Główny wniosek

Analiza wyników badania w kontekście przepisów AI Act oraz - szerzej - ochrony praw człowieka ujawnia **szereg wyzwań** dotyczących przejrzystości, kwalifikacji ryzyka i gotowości instytucji do nowych obowiązków regulacyjnych.

Widać, że polskie instytucje publiczne zaczynają identyfikować użycie AI, ale przejrzystość jest nadal reaktywna, nierówna i mocno zależna od kontekstu. To problem, bo AI Act opiera się na założeniu, że AI ma być technologią zorientowaną na człowieka, zgodną z Kartą praw podstawowych, demokracją i praworządnością. W szerszym standardzie przejrzystość oznacza nie tylko udzielenie odpowiedzi na wniosek, ale też dokumentowanie użycia AI, udostępnianie informacji osobom dotkniętym, możliwość zakwestionowania decyzji oraz informowanie, że ktoś wchodzi w interakcję z systemem AI.

Ujawnione zastosowania systemów wspomaganych AI

Badane systemy są wykorzystywane w bardzo różnych celach, ale da się wyróżnić kilka dominujących grup. Najliczniejsze grupy stanowią **systemy analizy obrazu i rozpoznawania obiektów oraz asystenci tekstowi i głosowi służący do obsługi informacyjnej.**

Najwyraźniej zaznacza się grupa systemów związanych z obrazem i monitoringiem. Obejmuje ona zarówno klasyczne zastosowania rozpoznawcze, takie jak automatyczne rozpoznawanie tablic rejestracyjnych, jak i bardziej złożone systemy analizy obrazu: monitoring kąpieliska, oznaczanie ludzi i pojazdów na nagraniach, analiza zdjęć RTG czy automatyczna klasyfikacja materiałów zdjęciowych i wideo. W tej grupie przezna-

czenie systemu jest zwykle opisane konkretnie i operacyjnie.

Drugą dużą grupą są chatboty, voiceboty i wirtualni asystenci. Ich przeznaczenie najczęściej dotyczy udzielania informacji, obsługi zapytań, kierowania rozmów do właściwego konsultanta albo rezerwacji wizyt. Widać więc, że instytucje publiczne używają AI nie tylko do analizy danych, ale także jako narzędzie kontaktu z obywatelem lub użytkownikiem usług publicznych.

Trzecia istotna grupa to systemy pracujące na dokumentach, tekście, mowie i informacjach prawnych. Należą tu narzędzia do transkrypcji wypowiedzi, ekstrakcji danych ze skanów, wykrywania danych osobowych, analizy faktur, klasyfikacji dokumentów patentowych czy wykrywania niedozwolonych postanowień umownych. W tych przypadkach AI pełni funkcję wspomagającą pracę urzędniczą, analityczną lub ekspercką.

Ogólny wniosek jest taki, że w badanym zbiorze **administracja wykorzystuje systemy AI głównie do trzech celów: rozpoznawania i analizy obrazu, automatyzacji kontaktu informacyjnego z użytkownikiem oraz przetwarzania dokumentów lub treści tekstowych**. Mniej liczne są zastosowania związane z predykcją, sterowaniem procesami, cyberbezpieczeństwem czy ogólną produktywnością biurową.

Shadow AI

Warto podkreślić, że to badanie, jako skierowane do instytucji, nie objęło zjawiska określanego jako ***shadow AI*, czyli nieautoryzowanego lub niekontrolowanego wykorzystywania narzędzi sztucznej inteligencji przez pracowników poza oficjalnymi systemami i politykami organizacji**. Najczęściej przyjmuje ono postać wspierania się popularnymi czatami, jak np. *ChatGPT*, bez wiedzy przełożonych, a więc również bez stosownego nadzoru i odpowiednich zabezpieczeń proceduralnych. Zjawisko to może prowadzić do ryzyk związanych z nieautoryzowanym wpływem dużego modelu językowego na podejmowane decyzje, czy też bezpieczeństwem danych. O skali zjawiska można spekulować na podstawie wspomnianego w omówieniu pierwszej fazy badania raportu, z którego wynika, że połowa urzędników przyznaje się do korzystania z AI do celów służbowych a 59% przypadków użycia AI w tygodniu poprzedzającym badanie dotyczyło wyszukiwania faktów lub informacji*, które pomagają w wykonywaniu pracy*.

Kwalifikacja systemu, kwalifikacja ryzyka

Zebrane informacje dotyczą 37 systemów, z czego w 19 przypadkach wskazano, że nie przeprowadzono kwalifikacji ryzyka. I to jest kluczowy sygnał ostrzegawczy: bez kwalifikacji ryzyka instytucja nie będzie wiedziała, czy powinna w danym kontekście przestrzegać obowiązków wynikających przykładowo z art. 26–27 AI Act, między innymi czy musi informować osoby, rejestrować użycie systemu, przeprowadzić FRIA albo uwzględnić nadzór człowieka w procesach wspieranych przez dany system.

Przykładowo nie przeprowadzono analizy ryzyka takich systemów jak *eRecruiter* (bo moduł AI nie włączony - a jak go włączą? a jak się włączy automatycznie po jakiejś aktualizacji?), czy *SOS AI* (bo nie uznano go za system AI). Podobnie wydaje się dziać w przypadku korzystania z oprogramowania z serii LEX, którego twórcy podążając za trendem dodają różne funkcje wspomagane sztuczną inteligencją.

Niepokojącym zjawiskiem jest przy tym **tendencja niektórych podmiotów do zaprzeczania, że ich rozwiązania są *systemami AI* w rozumieniu AI Act**, co może służyć uniknięciu rygorystycznych obowiązków. Przykładowo ZUS twierdzi, że jego system *SOS AI* do typowania zwolnień lekarskich nie jest oparty na AI, lecz jest *narzędziem analitycznym*, mimo że wykorzystuje modele predykcyjne do oceny prawdopodobieństwa nadużyć. A przypomnijmy definicję systemu AI z art. 3 pkt 1 AI Act: *System maszynowy zaprojektowany do działania z różnym poziomem autonomii, który po wdrożeniu może wykazywać zdolność adaptacji. Na podstawie otrzymanych danych wejściowych wnioskuje, jak generować wyniki (takie jak **predykcje**, treści, zalecenia lub decyzje), które mogą wpływać na środowiska fizyczne lub wirtualne*. ZUS bardzo się nagimnastykował, żeby uzasadnić, dlaczego SOS AI nie mieści się w tej definicji, jednak w naszym przekonaniu autorzy AI Act zdecydowanie mieli na myśli również takie zastosowania. Takie podejście uderza w przejrzystość, ponieważ uniemożliwia obywatelom zrozumienie, w jaki sposób algorytmy wpływają na ich sytuację prawną (tutaj prawo do zasiłku).

I dalej, w odniesieniu do 12 systemów ryzyko uznano za minimalne, w 5 za ograniczone, a tylko w 1 za wysokie. Przy czym AI Act klasyfikuje jako wysokiego ryzyka m.in. systemy w obszarach biometrii, infrastruktury krytycznej, zatrudnienia, podstawowych usług publicznych, ścigania przestępstw i wymiaru sprawiedliwości. W praktyce oznacza to, że część odpowiedzi wymagałaby dokładniejszej weryfikacji, bo z opisu systemu wynika, że ryzyko może być większe, niż zadeklarowano.

Przykładowo systemy dotyczące monitoringu wizyjnego, ANPR, kontroli zwolnień lekarskich, rekrutacji, świadczeń, bezpieczeństwa publicznego czy infrastruktury wodnej mogą nie być automatycznie wysokiego ryzyka, ale ich kontekst użycia jest dokładnie tym, który AI Act uznaje za wrażliwy z punktu widzenia praw człowieka. W przypadku świadczeń publicznych wysokim ryzykiem mogą być systemy służące do oceny kwalifikowalności, przyznawania, ograniczania, cofania lub żądania zwrotu świadczeń; w zatrudnieniu — systemy do selekcji kandydatów lub oceny osób; w ściganiu — systemy wspierające ocenę dowodów czy profilowanie.

Ujawniono tylko jeden system, którego ryzyko w klasyfikacji AI Act zidentyfikowano jako wysokie i jest to, co ciekawe, używany w Agencji Restrukturyzacji i Modernizacji Rolnictwa *System Identyfikacji Roślin AI*. System wykorzystuje algorytmy uczenia maszynowego do klasyfikacji zdjęć polowych według rozpoznanego gatunku lub typu uprawy. Wspiera procesy oceny i weryfikacji danych terenowych wykorzystywanych przez Agencję Restrukturyzacji i Modernizacji Rolnictwa w ramach obsługi świadczeń. Oznacza to, że wpływa na udzielane świadczenia, co zapewne jest przyczyną, dla której ryzyko zostało uznane za wysokie (zgodnie z AI Act, art. 6 ust. 2 w powiązaniu z załącznikiem III pkt 5 lit. a).

Ocena skutków dla ochrony danych (DPIA)

Zgodnie z przepisami RODO (które AI Act uzupełnia, a nie zastępuje), to administrator danych – czyli podmiot, który decyduje o celach i sposobach przetwarzania danych osobowych – jest zobowiązany do przeprowadzenia oceny skutków dla ochrony danych tzw. DPIA. W ekosystemie AI podmiotem tym jest za-

zwyczaj podmiot stosujący system AI. Stąd częstą odpowiedzią, tłumaczącą niedopełnienie tego obowiązku było zaprzeczenie przetwarzania danych osobowych przez dany system.

I o ile w przypadku dajmy na to *Systemu Identyfikacji Roślin AI* przypuszczalnie tak właśnie jest w istocie, o tyle w przypadku chatbotów, czy voicebotów używanych do komunikacji z obywatelami, założenie że żaden użytkownik nigdy sam z siebie nie ujawni danych osobowych tak samo naturalnie, jakby to zrobił w rozmowie z urzędnikiem, jest nierealistyczne. Ponadto system może przecież przetwarzać dane pośrednie pozyskane w procesie interakcji (i prawdopodobnie to robi w mniejszym lub większym zakresie, chociażby ze względów audytowych), które również mogą być danymi osobowymi (np. adres IP, dane o urządzeniu, dane behawioralne).

Czy więc brak przeprowadzenia DPIA w takich przypadkach nie jest jednak niezgodny z RODO i może stanowić naruszenie praw osób, których dane dotyczą? Czy system, który służy do składania ofert pracy na ogłoszenia o naborze w formie elektronicznej nie przetwarza danych osobowych? Co z monitoringiem obiektów typu osoba na obszarze kąpieliska? Co z systemami rozpoznającymi tablice rejestracyjne? A mamy szereg tego typu przypadków, gdzie jednocześnie zadeklarowano brak przeprowadzenia DPIA, z czego w części z nich zapewniano jednocześnie o braku przetwarzania danych osobowych (sic!).

Przejdźmy teraz do omówienia sytuacji, w których zadeklarowano przeprowadzenie DPIA. Otóż takich przypadków było... całe osiem, z czego tylko w **trzech** instytucja była uprzejma udostępnić treść (ocena ich adekwatności, chociaż warta przeprowadzenia, wykracza poza zakres niniejszego raportu). Oznacza to, że w praktyce audyt społeczny DPIA w odniesieniu do systemów AI jest niemożliwy często nawet tam, gdzie taką ocenę przeprowadzono, a obywatele nie wiedzą w jaki sposób i w jakim zakresie ich dane osobowe są przetwarzane, jakie zidentyfikowano ryzyka i jakie zastosowano środki zaradcze.

Podsumowując. Co się dzieje z naszym wizerunkiem przetwarzanym przez coraz bardziej zaawansowane systemy monitoringu? Dokąd trafiają i jak są przetwarzane odczyty tablic rejestracyjnych? Jakie dane behawioralne są zbierane przez chatboty i w jaki sposób są wykorzystywane? To są pytania, na które nie mamy odpowiedzi, bo instytucje rzadko udostępniają treści DPIA, o ile w ogóle została przeprowadzona.

Ocena wpływu na prawa podstawowe (FRIA)

Zgodnie z art. 27 AI Act, który ma zacząć obowiązywać od 2 sierpnia 2026 roku (o ile nie zostanie to zmienione w ramach pakietu *AI Omnibus*, nad którym trwają prace) podmioty publiczne stosujące systemy wysokiego ryzyka będą miały obowiązek przeprowadzenia oceny wpływu na prawa podstawowe, tzw. FRIA, przed jego wdrożeniem. Tymczasem prawie żadna instytucja tego nie przeprowadziła, tłumacząc to niższym poziomem ryzyka, brakiem oficjalnego unijnego kwestionariusza lub wskazując inne powody. Na brak kwestionariusza wskazała między innymi ARiMR w odniesieniu do *Systemu Rozpoznawania Roślin AI*, który okazał się jedynym zgłoszonym w badaniu systemem, zadeklarowanym jako wysokiego ryzyka.

W dwóch przypadkach zadeklarowano przeprowadzenie FRIA, pomimo że system nie został sklasyfikowany jako wysokiego ryzyka. Pierwsza sytuacja, w której ryzyka wcale nie przeprowadzono, ale zrobiono FRIA, dotyczy systemu AXON, wspierającego miejski monitoring w Rokietnicy. System służy do rozpoznawania zdarzeń takich, jak wypadek, osoba leżąca, zatrzymanie ruchu, czy rozpoznawanie numerów tablic rejestracyjnych, a komponent AI wspiera analizę behawioralną. Niestety Urząd Gminy Rokietnica uznał ten dokument za niepodlegający udostępnieniu, co czyni audyt społeczny niemożliwym do przeprowadzenia, a powodów tej decyzji nie podano. Drugim przypadkiem jest *Wirtualny Urzędnik* (tekstowy i głosowy) Ministerstwa Rozwoju i Technologii. W obu tych przypadkach urzędy przypuszczalnie uznały, że mimo braku klasyfikacji systemu jako wysokiego ryzyka, przeprowadzenie FRIA jest zasadne ze względu na potencjalny wpływ systemu na prawa podstawowe, co samo w sobie jest dobrą, godną naśladowania praktyką.

Ocenę ww. wirtualnego urzędnika jako jedyną nam udostępnioną, zacytujemy w całości. Wprawdzie nie spełnia ona wszystkich wymogów ustanowionych wspomnianym już art. 27, jednak z uwagi na *ograniczone ryzyko* systemu, którego dotyczy, należy ją traktować jako opcjonalną, a w związku z tym niezobligowaną do literalnej zgodności z przepisami.

Ocena wpływu na prawa podstawowe (FRIA) dla Wirtualnego Urzędnika w Ministerstwie Rozwoju i Technologii

1. **Nadzór człowieka i wytłumaczalność**
   - Model nie działa autonomicznie w krytycznym podejmowaniu decyzji.
   - Operatorzy ludzcy, tacy jak agenci usług IT, zachowują kontrolę nad wynikami generowanymi przez AI.
   - Odpowiedzi mogą być audytowane, przeglądane i modyfikowane przez agentów przed wdrożeniem w krytycznych procesach.
2. **Stronniczość i sprawiedliwość**
   - Model został wytrenowany przy użyciu starannie wyselekcjonowanych zbiorów danych, zapewniając, że nie propaguje niesprawiedliwych uprzedzeń w swoich odpowiedziach.
   - Mechanizm ciągłego monitorowania pozwala klientom na walidację wyników i dostrajanie odpowiedzi zgodnie z wewnętrznymi politykami zgodności.
3. **Audytowalność i raportowanie zgodności**
   - System zapewnia logowanie i monitorowanie dla organizacji, które wymagają ścieżek audytu dla zgodności z ramami zarządzania AI.
   - Na żądanie wszystkie logi mogą zostać wyłączone, co jest zgodne ze ścisłymi zasadami prywatności w fazie projektowania (privacy-by-design).

W szczególności, wobec całości zamówienia w ramach projektu, dostawca spełnia następujące wymagania:

- Zamawiający posiada opcje monitorowania wolumenów interakcji konsultanta w ramach rozwiązań wspieranych modelem LLM i ma możliwość wyłączenia opcji korzystania z modelu LLM dla użytkownika po przekroczeniu limitów wykorzystania interakcji (rozumianych jako zapytania, wyświetlenia odpowiedzi)
- Wszystkie transkrypty, logiki oraz dane użytkowników pozostają na serwerach kontrolowanych przez Wykonawcę.
- W chwili przesłania zapytania do modelu AI, dane trafiają do modelu obsługiwanego przez Wykonawcę, gdzie następuje analiza treści i generowanie odpowiedzi zwracanej do interfejsu systemu wywołującego (CRM) Zamawiającego.

- Dane nie są używane do trenowania modeli AI.
- Dane nie są udostępniane podmiotom trzecim, klient pozostaje właścicielem wszystkich swoich danych przez cały proces.

Podręczniki i instrukcje do systemów

Publiczna dostępność podręczników, instrukcji lub materiałów użytkownika dla badanych systemów AI jest ograniczona. Spośród 37 pozycji tylko w 9 przypadkach podano publiczny link do instrukcji, dokumentacji, centrum pomocy lub materiałów producenta. W większości przypadków instytucje albo odpowiedziały, że instrukcji nie ma, albo wskazały, że dokumentacja istnieje, ale ma charakter wewnętrzny, techniczny, podlega ograniczeniom licencyjnym lub jest dostępna wyłącznie dla uprawnionych użytkowników. To utrudnia ocenę, czy użytkownicy systemów są odpowiednio przygotowani do korzystania z narzędzi AI, zwłaszcza w zakresie rozumienia ograniczeń, typowych błędów, zasad nadzoru człowieka czy procedur eskalacji.

Ogólny wniosek jest taki, że instrukcje użytkownika systemu AI rzadko funkcjonują jako element przejrzystości w przestrzeni publicznej. Nawet gdy materiały są dostępne, często dotyczą ogólnej obsługi produktu, nie zaś odpowiedzialnego korzystania z systemu w konkretnym wdrożeniu administracji publicznej. W efekcie obywatel lub badacz ma ograniczoną możliwość sprawdzenia, w jakim zakresie AI faktycznie wspiera pracę urzędnika oraz na ile dokumentacja jest kompletna pod względem bezpiecznego i właściwego korzystania z danego systemu.

Publiczna wiedza o systemie

Przy omawianiu odpowiedzi na pytanie związane z publiczną wiedzą o systemie AI należy podkreślić, że zapytane podmioty mogły nie mieć wiedzy na ten temat, a nawet jeśli, to zasadniczo nie były zobowiązane do udzielenia tego typu informacji.

W każdym razie spośród 37 systemów, w 12 przypadkach podano co najmniej jeden publiczny adres URL prowadzący do wzmianki, opisu, dokumentu, strony producenta, materiału prasowego albo strony instytucji. W pozostałych przypadkach odpowiedź najczęściej sprowadza się do tego, że organ nie posiada informacji o takich zasobach.

Wśród odpowiedzi z linkami widać kilka różnych typów źródeł. Czasem są to oficjalne strony instytucji publicznych lub informacje o wdrożeniu, np. przy krakowskim monitoringu, systemie SIS-RZ w Bydgoszczy, Portalu Głosowym Warszawy czy eMCeK Ministerstwa Finansów. W innych przypadkach są to raczej strony producentów lub dostawców, np. Microsoft Copilot, eXtracToolbox, InteliWISE, Cellebrite, eRecruiter czy ForProgress. Są też przykłady źródeł zewnętrznych, takich jak artykuł medialny i interpelacja sejmowa dotyczące aplikacji OKO.

Pod względem jakości odpowiedzi są bardzo nierówne. Niektóre linki rzeczywiście prowadzą do informacji o konkretnym wdrożeniu albo przynajmniej o usłudze wykorzystywanej przez organ. Inne mają charakter ogólny: opisują produkt, producenta lub klasę rozwiązania, ale niewiele mówią o tym, jak konkretnie system działa w danej instytucji publicznej, jakie procesy wspiera, jakie są jego ograniczenia i jakie decyzje lub czynności administracyjne są z nim powiązane.

Warto też zauważyć, że część odpowiedzi nie realizuje w pełni prośby o linki. Policja w Białymstoku wskazuje, że publicznie dostępne są jedynie dokumenty zamówieniowe, ale nie podaje adresów. UOKiK stwierdza, że nie gromadzi takich danych, choć są one ogólnodostępne i możliwe do samodzielnego pozyskania. Zgierz odpowiada, że informacja o wdrożeniu znajduje się na stronie miasta, ale również bez linku.

Ogólny wniosek jest taki, że publiczne ślady informacji o systemach AI istnieją jedynie punktowo i często są rozproszone. Najczęściej nie tworzą spójnego, łatwo dostępnego obrazu systemu używanego przez administrację. Tam, gdzie linki podano, zwykle pokazują one raczej komunikat, stronę promocyjną, opis produktu lub wzmiankę medialną niż pełny opis funkcjonowania systemu w konkretnym organie. W rezultacie publiczna wiedza o tych systemach pozostaje fragmentaryczna i zależna od przypadkowych źródeł, a nie od systematycznej przejrzystości po stronie instytucji.

Wykorzystanie modeli ogólnego przeznaczenia (GPAI)

Korzystanie z modeli ogólnego przeznaczenia nie jest szczególnie częste w badanym zbiorze. Spośród 37 systemów tylko w 8 przypadkach instytucje wskazały konkretne modele lub rodziny modeli GPAI wraz z producentem. Dotyczy to m.in. chatbotów w Gdyni i Zgierzu, *GeoChatbota* GUGiK, *Copilota* w Gdyni i Krakowie czy konwertera mowy na tekst w Kancelarii Sejmu. Przy czym tylko w jednym przypadku wskazano rodzime rozwiązania (niekoniecznie typu GPAI w tym akurat przypadku), a w jeszcze jednym poinformowano o planach wdrożenia PLLuM (tej rodzinie modeli poświęcamy dalej osobny rozdział).

Warto zatem zwrócić uwagę na to, że wspomniane wyżej modele (poza może dwoma ostatnimi) są generalnie udostępniane w postaci usług chmurowych, a więc bez możliwości uruchomienia w bezpiecznym, lokalnym środowisku. Oznacza to, że treści wpisywane przez użytkownika, a więc również potencjalnie wrażliwe dane (być może po mniej lub bardziej skutecznej anonimizacji, tego nie wiemy) bywają wysyłane nie tylko na zagraniczne serwery, ale potencjalnie nawet poza obszar Unii Europejskiej. Związane z tym ryzyka miał właśnie ograniczać PLLuM.

Najbardziej wyraźną grupę stanowią odpowiedzi negatywne. W 21 przypadkach organ odpowiada po prostu *Nie*. Dotyczy to zarówno systemów monitoringu, analizy obrazu, obsługi faktur, kontroli dostępu, voicebotów, chatbotów, jak i systemów klasyfikacyjnych. W dwóch kolejnych przypadkach odpowiedzi wskazują, że system nie korzysta z GPAI, lecz z rozwiązań własnych lub dedykowanych: *Cellebrite* ma bazować na modelach wbudowanych przez producenta, a *eXtracToolbox* na komponentach opracowanych przez wyspecjalizowany instytut i zaimplementowanych do systemu dziedzinowego.

W odpowiedziach pozytywnych dominuje OpenAI. Pojawiają się modele lub produkty: *GPT*, *GPT-4*, *GPT-4o*, *GPT-5*, *ChatGPT* oraz — przy Kancelarii Sejmu — *ggml-large-v3-turbo-q8_0*. Część odpowiedzi wiąże GPAI z usługą *Microsoft*

Copilot, wskazując, że korzysta ona z modeli OpenAI oraz rozwiązań własnych Microsoftu, w tym *Prometheus*. W przypadku Gdyni dodatkowo pojawia się informacja o planowanym wdrożeniu polskiego modelu PLLuM.

Osobno warto potraktować odpowiedź Krakowa dotyczącą pilotażowego monitoringu, gdzie wskazano biblioteki *Ultralytics YOLO*. Organ odpowiedział, że system korzysta z modeli ogólnego przeznaczenia, ale kwalifikacja ta jest mniej oczywista niż w przypadku dużych modeli językowych. *YOLO* to rozwiązanie z obszaru detekcji obiektów i analizy obrazu, więc odpowiedź pokazuje szersze, nie zawsze jednolite rozumienie pojęcia GPAI przez instytucje.

Istotna jest też grupa odpowiedzi niepełnych. Niektóre organy piszą, że nie posiadają informacji o nazwach lub wersjach modeli używanych w produktach komercyjnych, odsyłają do dostawcy albo nie podają nazwy modelu. To dotyczy m.in. *eRecruitera*, *Mycroft BIP Scanner*, produktów z rodziny *LEX*, *Cisco Webex* i *FortiGate/FortiGuard*. W takich przypadkach nie da się ustalić z samej odpowiedzi, czy system rzeczywiście korzysta z GPAI, czy organ po prostu nie ma wiedzy o tej warstwie używanego narzędzia.

Ogólny wniosek jest taki, że instytucje najczęściej deklarują brak korzystania z GPAI, ale tam, gdzie takie modele się pojawiają, są to głównie rozwiązania dużych dostawców, przede wszystkim OpenAI i Microsoft. Jednocześnie objawia się tutaj kolejny problem z przejrzystością: część organów nie zna nazw modeli wykorzystywanych w komercyjnych systemach albo nie potrafi ich wskazać. To utrudnia ocenę zależności od dostawców, zgodności z regulacjami oraz ryzyk związanych z używaniem modeli ogólnego przeznaczenia w sektorze publicznym.

Oznaczanie decyzji

Art. 26 ust. 11 AI Act

(...) **podmioty stosujące** systemy wysokiego ryzyka, (...) które to podmioty podejmują decyzje lub uczestniczą w podejmowaniu decyzji dotyczących osób fizycznych, **informują osoby fizyczne** o tym, że jest w stosunku do nich wykorzystywany system AI wysokiego ryzyka.

Art. 86 ust. 1 AI Act

Każda osoba, na którą AI ma wpływ, będąca przedmiotem decyzji podjętej przez podmiot stosujący na podstawie wyników systemu AI wysokiego ryzyka (...), która to decyzja wywołuje skutki prawne lub w podobny sposób oddziałuje na tę osobę na tyle znacząco, że uważa ona, iż ma to niepożądany wpływ na jej zdrowie, bezpieczeństwo lub prawa podstawowe, ma prawo uzyskania od podmiotu stosującego **jasnego i merytorycznego wyjaśnienia roli** tego systemu AI w procedurze podejmowania decyzji oraz głównych elementów podjętej decyzji.

Wiele wskazuje na to, że o ile instytucje publiczne w Polsce sięgają po narzędzia oparte na AI w celach mających mniejszy lub większy wpływ na podejmowane decyzje, o tyle raczej unikają mierzenia się z konsekwencjami. Najprościej nie uznać systemu za system AI, albo zapewniać o ścisłym nadzorze człowieka nad procesem (tzw. *human-in-the-loop*) i już instytucja czuje się zwolniona z obowiązku informowania o użyciu AI oraz innych obowiązków wynikających z przepisów tak aktualnych, jak i wchodzących w życie w nieodległej przyszłości. Prawdopodobnie z taką sytuacją mamy do czynienia w przypadku ZUS.

Organ utrzymuje, że *narzędzie analityczne oparte o działanie modelu predykcyjnego, które jest wykorzystywane przez organ w systemowej analizie danych z wystawionych zaświadczeń lekarskich o czasowej niezdolności do pracy, nie zastępuje działalności, w tym decyzji podejmowanych przez lekarzy orzecznictwa lekarskiego Zakładu Ubezpieczeń Społecznych, a stanowią rozwiązania wspomagające ww. proces przez nich prowadzony, przede wszystkim na podstawie analizy i merytorycznej oceny stanu faktycznego sprawy, z uwzględnieniem medycznych przesłanek podjęcia kontroli, za który to proces decyzyjny biorą i ponoszą pełną odpowiedzialność*. Niemniej jednocześnie organ *potwierdza wykorzystanie rozwiązań, które noszą cechy szeroko rozumianej, opisanej w odpowiedzi z dnia 22 października 2025 r., sztucznej inteligencji sensu largo, m.in. w ramach analizy służącej wstępnemu typowaniu zaświadczeń lekarskich do kontroli, prowadzonej w oparciu o rozszerzony zakres dostępnych danych.*

Czyli nawet jeśli ostateczną decyzję podejmuje człowiek, to podejmuje ją na podstawie preselekcji dokonanej przez system predykcyjny AI. Jeśli więc system predykcyjny zawiera błędy i jego typowanie jest uprzedzone, w związku z czym przykładowo pomija jakąś specyficzną grupę bardzo podejrzanych przypadków, to może to pozostawać niezauważone przez długi czas, ponieważ nadzór człowieka wkracza na późniejszym etapie procesu, służąc potwierdzeniu słuszności wytypowanych sytuacji. Ponadto na ile staranny jest ten nadzór? Na ile jest uprzedzony faktem, że przypadek został wytypowany przez AI? Doświadczenie pokazuje, że ludzie wykazują dużą skłonność do ufania sztucznej inteligencji. Badania dot. *automation bias* pokazują, że człowiek nadzorujący system może zbyt łatwo przyjmować rekomendację AI, zwłaszcza pod presją czasu, przy niskiej wiedzy technicznej albo gdy system wygląda na kompetentny (patrz np. metanaliza 35 badań). Zresztą, skoro system nie został zakwalifikowany przez organ jako system AI, to jeśli rola człowieka w procesie w przyszłości będzie malała i sprowadzi się do *klepania* decyzji podjętych przez system lub nawet zostanie całkowicie pominięta, to z punktu widzenia osób oskarżanych o nadużycia będzie to niezauważone, ponieważ decyzje, które otrzymują nie są, jak się zdaje, oznaczone wyraźnie jako podjęte z użyciem AI.

Jedynym systemem wysokiego ryzyka, jaki badane instytucje wskazały, jest System Rozpoznawania Roślin AI stosowany przez Agencję Restrukturyzacji i Modernizacji Rolnictwa. *System wykorzystuje algorytmy uczenia maszynowego do klasyfikacji zdjęć polowych według rozpoznanego gatunku lub typu uprawy. Wspiera procesy oceny i weryfikacji danych terenowych wykorzystywanych przez Agencję Restrukturyzacji i Modernizacji Rolnictwa w ramach **obsługi świadczeń***. Nie wiemy, czy idzie za tym jakaś informacja na decyzjach o świadczeniach trafiających do świadczeniobiorców, która wskazuje na zautomatyzowanie przynajmniej części procesu. Jednak przewaga nad poprzednią sytuacją jest taka, że skoro organ sam zaklasyfikował swój system, jako system wysokiego ryzyka, to znaczy, że jest gotowy wziąć za to odpowiedzialność i zmierzyć się z wynikającymi z tego obowiązkami. Jeśli powstaną rejestry systemów AI (co - niezależnie od regulacji AI Act - postulujemy na poziomie krajowym), to taki system się w nich znajdzie, co otworzy pole do objęcia społecznym nadzorem. Jest to ważne, bo przecież nie da się kontrolować czegoś, o którego istnieniu nie ma się pojęcia.

Oznaczanie treści

Jeszcze we wrześniu 2024 Ministerstwo Cyfryzacji w zamieszczonych na stronie internetowej poradach pisało: *Jeżeli chcesz udostępnić treść wygenerowaną przez GenAI oznacz ją*. I dalej, co prawda bardziej w obawie o potencjalne naruszenia prawa autorskiego, niż w kontekście przejrzystości i prawa do informacji, ale: *W przypadku wykorzystania wyników w dalszej pracy, zawsze informuj, że treści, które przekazujesz, były stworzone lub przetworzone z wykorzystaniem narzędzi GenAI. Oznacz fragmenty wprost zaczerpnięte z tego narzędzia, np. wskazując w przypisie „Treść została wygenerowana przez GenAI [tu wstaw nazwę narzędzia i datę wygenerowania]" Jeżeli na bazie informacji pozyskanych z Gen AI tworzysz własne treści zaznacz: „Ten dokument powstał z wykorzystaniem GenAI [tu wstaw nazwę narzędzia i datę wygenerowania]"* (snapshot na webarchive). Porady te, chociaż nie wyczerpują w pełni zagadnienia, pozostają w duchu artykułu 50 AI Act, który poświęcony jest obowiązkom oznaczania treści generowanych z użyciem AI.

Natomiast opublikowana niedawno, tj. 4 marca 2026 roku 99-stronicowa wersja Przewodnika po sztucznej inteligencji dla administracji publicznej pomimo że znacznie bardziej obszerna i szczegółowa, niż ww. zestaw porad, to jednak stosunkowo ogólnie odnosi się akurat do obowiązków oznaczania treści generowanych z użyciem AI, nakładanych przez art. 50 AI Aktu. Właściwie związane wprost z oznaczaniem są bodaj tylko dwa fragmenty: *Zapewnij przejrzystość wobec mieszkańców. Należy informować mieszkańców o wykorzystywaniu systemów AI w danym procesie (...) Informacje o stosowaniu rozwiązań opartych na AI (w szczególności chatbotów) powinny być przekazywane mieszkańcom w sposób jasny i zrozumiały.* Można więc odnieść wrażenie, że waga transparentności oznaczania u decydentów znacznie spadła.

Tyle mówi teoria. Bardzo trudno jest ją zweryfikować, ponieważ nie sposób ustalić, jaki odsetek dokumentów wytwarzanych przez administrację powstaje przy mniejszym lub większym udziale sztucznej inteligencji, tym bardziej że przeważająca część takich zastosowań należy do tzw. *shadow AI*, czyli zjawiska nieformalnego wspierania się zewnętrznymi narzędziami AI przez pracowników administracji. Dysponujemy odmownymi decyzjami najprawdopodobniej przygotowanymi z mniejszym lub większym wykorzystaniem AI, ponieważ zawierają halucynowane przepisy, co jest zjawiskiem znacznie częściej przydarzającym się modelom językowym, niż urzędnikom - opisaliśmy to przy okazji raportu z pierwszej fazy badania, jednak trudno na podstawie kilku przypadków, które wpadły nam w ręce, wyrokować o skali zjawiska.

Dzięki niniejszemu badaniu dowiedzieliśmy się o jednej, względnie pewnej (o ile nie doszło do jakiegoś nieporozumienia) sytuacji, w której AI użyto do przygotowania dokumentu, ale nie zadbano o czytelną dla człowieka informację o tym. Otóż istnieje coś takiego, jak wieloletnia prognoza finansowa - obligatoryjny dokument planistyczny polskich samorządów. Prawidłowo przygotowana wieloletnia prognoza finansowa umożliwia trafne przewidywanie zarówno dochodów bieżących, jak i majątkowych, wspiera efektywne zarządzanie wydatkami z uwzględnieniem zadłużenia jednostki samorządu terytorialnego oraz pomaga ustalić właściwe poziomy nakładów na realizację poszczególnych inwestycji i projektów. Do przygotowania tego ważnego dokumentu gmina Boguchwała (a także, zgodnie z odpowiedziami w pierwszej fazie, gmina Chojnice i Starostwo Powiatowe w Lidzbarku Warmińskim) używa oprogramowania o nazwie takiej samej, jak akronim tworzonego dokumentu, czyli *WPF*, wykorzystującego sztuczną inteligencję ogólnego przeznaczenia (w odpowiedzi urzędu podano angielski akronim GPAI). Sprawdziliśmy trzy uchwały Rady Miejskiej w Boguchwale dotyczące ww. dokumentu - jego uchwalenia lub zmian (VIII.78.2024 z 30 grudnia 2024, X.108.2025 z 27 lutego 2025 i XX.203.2025 z 23 grudnia 2025). Ani w treści uchwał, ani w załącznikach zawierających właściwą treść WPF, nie dostrzegliśmy żadnej wzmianki o wsparciu AI w przygotowaniu dokumentu.

Co ciekawe, o całej tej sytuacji dowiedzieliśmy się przypadkowo, ponieważ do drugiej fazy badania wytypowaliśmy

Boguchwałę, nic nie wiedząc o tym, że akurat tam stosowane jest oprogramowanie *WPF* - wytypowaliśmy tę gminę, bo w pierwszej fazie napisali, że w ramach projektu *Smartcity* zainstalowano kamerę, która analizuje zachowania tłumu pod kątem wychwytywania potencjalnie niebezpiecznych zagrożeń i chodziło nam o zebranie dodatkowych informacji o tym właśnie rozwiązaniu. Tymczasem, ku naszemu zaskoczeniu, Boguchwała odpowiedziała w drugiej fazie badania na szczegółowe pytania w odniesieniu do systemu wspierającego przygotowanie wieloletniej prognozy finansowej, a z kolei o systemie monitoringu zagrożeń nie napisali ani słowa.

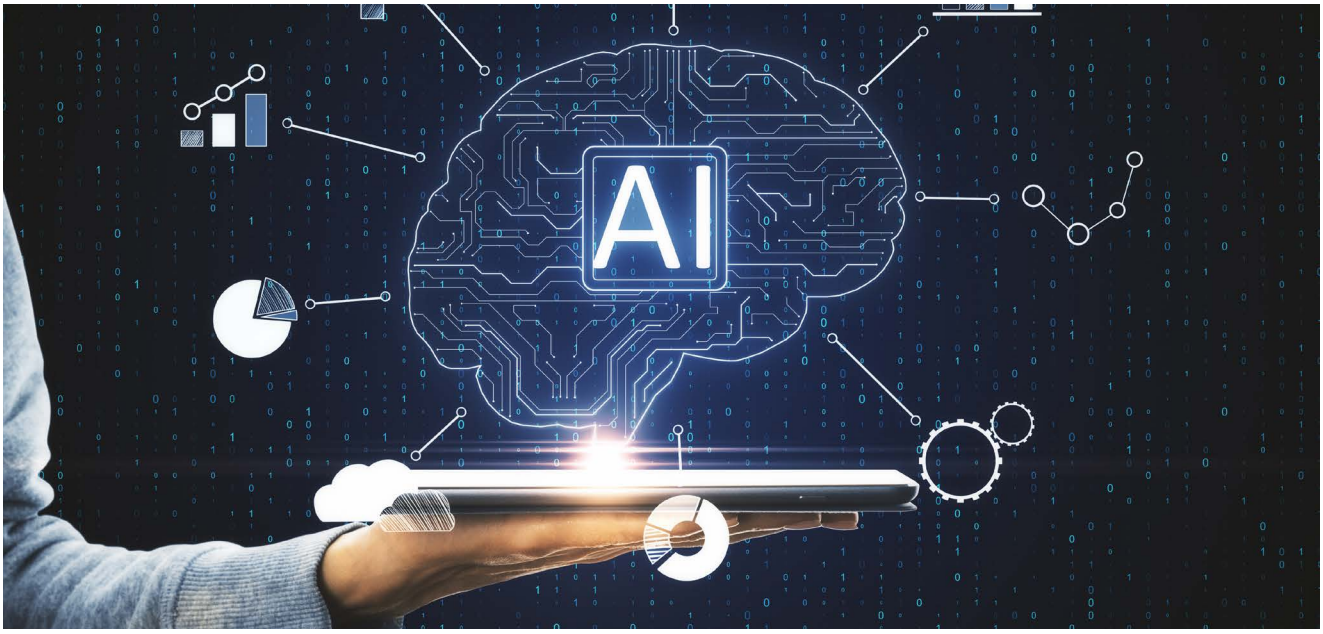
Ponadto od jakiegoś czasu (najstarszy snapshot na webarchive z 17 stycznia 2025) istnieje na rynku przynajmniej jedno narzędzie reklamowane jako wspierające generowanie innego, ważnego dokumentu jednostek samorządu terytorialnego, mianowicie raportu o stanie gminy. Narzędzie nazywa się *Raport_AI* i obiecuje przygotowanie raportu na podstawie danych i przy użyciu między innymi generatywnej sztucznej inteligencji. Na podstawie informacji na stronie dostawcy tego oprogramowania, a także ogólnej wzmianki we wstępie do Raportu o stanie gminy Kielce 2024 o podmiotach, które brały udział w procesie przygotowania tego dokumentu, można zakładać, że powstał z użyciem AI. Mimo to dokument nie zawiera stosownego oznaczenia a Kielce odpowiedziały przecząco na pierwsze pytanie wniosku w pierwszej fazie badania (*Czy w organie lub na jego zlecenie są wykorzystywane systemy informatyczne, które automatycznie przetwarzają dane wejściowe (np. tekst, obraz, dźwięk, dane liczbowe), aby generować raporty…*).

Wirtualni urzędnicy

Specyficznym przypadkiem, który wydaje się być oczywisty w kontekście oznaczania bez względu na zidentyfikowany poziom ryzyka, są wirtualni urzędnicy, którzy występują najczęściej w dwóch postaciach (póki co): chatbotów i voicebotów. Kwestia transparentności nie podlega tutaj dyskusji - tego typu rozwiązania są albo mniej lub bardziej wyraźnie oznaczone (np. *Ania – Wirtualny Asystent stron Gminy Zgierz* w tytule okna chatu, czy *Uwaga: GeoChatbot oparty jest o AI* małymi literami pod polem do wpisywania wiadomości na Geoportalu), albo przedstawiają się w sposób raczej nie pozostawiający wątpliwości, że z czymkolwiek mamy do czynienia, nie jest to człowiek (*Witaj! Jestem Wirtualnym Urzędnikiem Miasta Gdyni*, albo *Dzień dobry! Z tej strony e-Konsultantka miasta Poznań*).

> Niektóre z rozwiązań typu chatbot oferują przełączenie na rozmowę z człowiekiem, bez utraty kontekstu. W ramach weryfikacji tego mechanizmu w dwóch przypadkach, pytaliśmy konsultanta, co widzi za oknem. Nie doczekawszy odpowiedzi, dodawaliśmy, że pytamy dla upewnienia się, że rozmawiamy z człowiekiem. W obu przypadkach przyszła odpowiedź, że tak, ale ostatecznie bez informacji, co konsultant widzi za oknem. Musimy zatem wierzyć na słowo, że faktycznie doszło do kontaktu z człowiekiem. Wniosek dodatkowy: warunki pracy konsultantów być może wymagają odrębnego badania.

Polish Large Language Universal Model - PLLuM



O projekcie PLLuM Ministerstwo Cyfryzacji pisało w poprzedniej fazie badania: *Dla uwzględnienia potrzeb administracji publicznej, w szczególności bezpiecznego przetwarzania danych (w granicach kraju, zgodnie z regulacjami europejskimi) w 2024 r. powstała ze środków publicznych rodzina modeli PLLuM (Polish Large Language Universal Model). Na bazie tych modeli będziemy tworzyć narzędzia AI.* Dalej Ministerstwo zapowiadało inteligentnego asystenta urzędniczego, który *będzie wspierał urzędników w przygotowaniu prostych odpowiedzi na powtarzające się pytania, w redakcji dłuższych wypowiedzi i w analizie tekstu, a także prowadził przez procedury obowiązujące w urzędach*, co ziściło się jeszcze przed powstaniem tego raportu, do czego niedługo wrócimy.

Modele PLLuM przygotowano jako rodzinę polskich dużych

modeli językowych, których celem było lepsze dostosowanie sztucznej inteligencji do języka, kultury i realiów instytucjonalnych Polski. Podstawą prac był bardzo duży korpus tekstów, przede wszystkim polskojęzycznych, uzupełniony danymi angielskimi oraz wybranymi językami słowiańskimi i bałtyckimi. Dane pochodziły z wielu typów źródeł, m.in. tekstów internetowych, prawnych, urzędowych, naukowych, medialnych, literackich i specjalistycznych.

Szczególną wagę przywiązano do legalności i jakości danych. Zanim teksty trafiły do treningu, analizowano ich status prawny, porządkowano metadane, usuwano duplikaty, filtrowano treści niskiej jakości oraz sprawdzano poprawność techniczną plików. Dane były też klasyfikowane tematycznie, aby zapewnić możliwie szerokie pokrycie różnych dziedzin życia społecznego, administracji, nauki i kultury. Dzięki temu model miał uczyć się nie tylko języka ogólnego, lecz także stylów i pojęć ważnych w polskim kontekście.

Proces treningu przebiegał etapami. Najpierw modele uczono ogólnego rozumienia języka na dużych zbiorach tekstów. Następnie dostrajano je na specjalnym korpusie instrukcji, czyli przykładach poleceń i oczekiwanych odpowiedzi. Ten etap miał nauczyć modele praktycznego odpowiadania na pytania, streszczania, klasyfikowania, prowadzenia dialogu, tłumaczenia czy wykonywania prostych zadań wymagających rozumowania. Autorzy testowali różne strategie treningu, w tym trenowanie modeli od podstaw oraz dalsze uczenie istniejących modeli, ostatecznie kładąc duży nacisk na podejście bardziej efektywne obliczeniowo.

Osobnym elementem było zadbanie o alignment, czyli dopasowanie zachowania modelu do oczekiwań użytkowników, zasad bezpieczeństwa i norm etycznych. W tym celu przygotowano korpus preferencji, w którym ludzie oceniali i porównywali różne odpowiedzi modeli pod kątem poprawności, pomocności, bezpieczeństwa i adekwatności. Na tej podstawie modele uczono wybierania lepszych odpowiedzi. Dodatkowo zastosowano warstwę zabezpieczeń, która analizowała zarówno pytania użytkowników, jak i odpowiedzi modelu, wykrywając m.in. treści szkodliwe, dane osobowe, agresję czy próby obejścia zabezpieczeń.

Ewaluacja modeli była wielowymiarowa. Nie ograniczała się do jednego rankingu, lecz obejmowała testy automatyczne, oceny ekspertów, porównania dokonywane przez użytkowników oraz osobne badania bezpieczeństwa. Sprawdzano m.in. poprawność językową, faktyczność, zwięzłość, przydatność odpowiedzi, odporność na szkodliwe polecenia oraz zdolność pracy z dokumentami w systemach typu RAG. Ważnym elementem była też ocena przez rodzimych użytkowników języka polskiego, ponieważ pozwalała wykrywać problemy niewidoczne w typowych anglojęzycznych benchmarkach, takie jak nienaturalna polszczyzna czy nieodpowiedni styl komunikacji.

Szczegółowe informacje o rodzinie modeli PLLuM można znaleźć w publikacji naukowej **PLLuM: A Family of Polish Large Language Models**, która opisuje architekturę, proces treningu, zestaw danych i wyniki ewaluacji tych modeli.

Wytworzona z publicznych środków rodzina modeli PLLuM wydaje się znajdować rozmaite zastosowania w różnych instytucjach w realnych wdrożeniach. Przykładem jest wirtualny asystent wspierający użytkownika w aplikacji *mObywatel*, który został uruchomiony w grudniu 2025 roku. Asystent ten wykorzystuje model PLLuM do udzielania odpowiedzi na pytania dotyczące usług publicznych, procedur administracyjnych i innych zagadnień związanych z funkcjonowaniem administracji publicznej. Jego celem jest ułatwienie obywatelom dostępu do informacji i usług publicznych poprzez interakcję w naturalnym języku.

Podejmowane są także próby wykorzystania PLLuMa jako ułatwienia w przeglądaniu zasobów Biuletynu Informacji Publicznej, np. w **Gdyni**. Z kolei w **Poznaniu** wdrożenie miało charakter wspomagający pracę samych urzędników. Wykorzystano tam opartą na PLLuM aplikację ShpaRAG (wykorzystującą mechanizm Retrieval-Augmented Generation). Narzędzie to działa na miejskiej infolinii, pełniąc funkcję asystenta pracownika – na bieżąco analizuje obszerne bazy wiedzy urzędu i podpowiada urzędnikom poprawne i szybkie odpowiedzi na zapytania dzwoniących mieszkańców. Inny przykład to wyszukiwarka semantyczna/RAG dla **pracowników banku**, oparta na modelach PLLuM, do szybkiego dostępu do procedur i regulacji. Z kolei wbudowany w rozwiązania dla przedsiębiorstw oparty na PLLuM **ChatERP** umożliwia swobodną komunikację z systemami realizującymi różne zadania przy użyciu języka naturalnego.

> Wszystkie zagrożenia związane z wykorzystaniem dużych modeli językowych pozostają w mocy w przypadku modeli PLLuM, jednak dzięki trenowaniu ich na polskich zasobach wysokiej jakości, można oczekiwać, że niektóre z tych zagrożeń są w mniejszym lub większym stopniu osłabione.

Rekomendacje

Na podstawie wyników i wniosków z badania można wyprowadzić szereg rekomendacji, które będą się różnić w zależności od przekonań i celów osoby je formułującej. Poniżej znajdują się dwa postulaty, które wydają nam się najważniejsze i bez których realizacji wszelkie regulacje wokół AI związane z przejrzystością i ochroną praw obywateli będą kompletnie nieskuteczne. Dostrzegamy, że te sytuacje są w pewnym stopniu zaadresowane w AI Act, jednak przeprowadzone przez nas badania jasno pokazują, że jest to stopień dalece niewystarczający.

1. Utworzenie publicznego rejestru zastosowań systemów AI w administracji publicznej

Proponujemy mówić o rejestrze zastosowań AI, a nie wyłącznie o rejestrze systemów AI. Z perspektywy obywateli, organizacji społecznych, mediów i organów kontrolnych kluczowe jest bowiem nie tylko to, jakie narzędzia zostały zakupione lub wdrożone, lecz przede wszystkim to, w jakich procesach administracyjnych są wykorzystywane i w jakim zakresie wpływają na sytuację obywateli.

Taki rejestr powinien stanowić jedno z podstawowych narzędzi realizacji zasady przejrzystości i rozliczalności AI w sektorze publicznym. Jego celem nie powinno być tworzenie kolejnego obowiązku sprawozdawczego dla administracji, lecz zapewnienie, że rozwój AI w państwie będzie możliwy do społecznej, instytucjonalnej i eksperckiej kontroli.

Doświadczenia z prowadzonych przez nas działań monitoringowych wskazują, że informacje o wykorzystywaniu AI przez podmioty publiczne są obecnie rozproszone, niejednolite i często możliwe do uzyskania dopiero w trybie wniosków o informację publiczną, o ile w ogóle. Taki model jest niewystarczający wobec skali wdrożeń AI w administracji. Obywatele nie są w stanie skutecznie kontrolować wykorzystania sztucznej inteligencji przez państwo, jeżeli najpierw muszą samodzielnie ustalać, czy i gdzie takie rozwiązania w ogóle są stosowane.

Rejestr zastosowań AI miałby również znaczenie porządkujące dla samej administracji. Ułatwiłby koordynację wdrożeń, ocenę ryzyk, wymianę dobrych praktyk, identyfikację rozwiązań powielanych w wielu instytucjach oraz monitorowanie zgodności z wymaganiami wynikającymi z prawa unijnego i krajowego.

Rekomendacja dla Ministerstwa Cyfryzacji

Utworzenie i utrzymywanie centralnego, publicznego rejestru zastosowań systemów sztucznej inteligencji w administracji publicznej, obejmującego co najmniej informacje pozwalające ustalić, przez jaki podmiot, w jakim obszarze, w jakim celu i wobec kogo dane rozwiązanie jest wykorzystywane.

Ponadto wpis do ww. rejestru powinien być warunkiem dopuszczalności wykorzystania danego zastosowania AI w administracji publicznej. Innymi słowy, organ publiczny nie powinien korzystać z systemu AI w procesie dotyczącym obywateli, jeżeli zastosowanie to nie zostało wcześniej ujawnione w rejestrze. W przypadku systemów mogących wpływać na sytuację prawną obywatela brak wpisu powinien stanowić co najmniej istotną wadę proceduralną oraz podstawę do oceny odpowiedzialności służbowej lub organizacyjnej osób odpowiedzialnych za wdrożenie.

2. Wprowadzenie obowiązku oznaczania użycia AI

Postulujemy także zasadę, zgodnie z którą wykorzystanie sztucznej inteligencji przez administrację publiczną powinno być widoczne dla obywatela zawsze wtedy, gdy AI uczestniczyła - w dowolnym zakresie i na dowolnym etapie - w przygotowaniu sprawy, dokumentu, pisma, rekomendacji, analizy lub rozstrzygnięcia. Obywatel powinien wiedzieć, kiedy administracja korzysta wobec niego z AI albo kiedy AI współuczestniczy w tworzeniu dokumentów i działań organu publicznego.

Taki obowiązek powinien obejmować nie tylko sytuacje, w których system formalnie podejmuje decyzję. W praktyce istotny wpływ na sprawę może mieć również system, który *jedynie* wspiera urzędnika: wskazuje sprawy do kontroli, porządkuje kolejność ich rozpatrywania, przygotowuje projekt pisma, generuje rekomendację, analizuje ryzyko albo podsumowuje materiał dowodowy. Dla obywatela różnica między *decyzją podjętą przez AI* a *decyzją przygotowaną z istotnym udziałem AI* jest znacznie mniej istotna niż fakt, że technologia wpłynęła na sposób potraktowania jego sprawy.

Oznaczanie użycia AI jest warunkiem realnej rozliczalności administracji. Bez takiej informacji obywatel nie wie, że może zapytać o zasady działania systemu, zakwestionować jego wpływ na sprawę, domagać się wyjaśnień lub dodatkowej weryfikacji przez człowieka.

Rekomendacja dla Ministerstwa Cyfryzacji

Wprowadzenie standardu oznaczania użycia sztucznej inteligencji w działaniach administracji publicznej, w szczególności w decyzjach, pismach, dokumentach urzędowych i komunikacji z obywatelami, jeżeli AI była wykorzystywana w sposób mogący mieć znaczenie dla merytoryki dokumentu, przebiegu sprawy lub sytuacji obywatela. Przy czym przesłanką do oznaczenia powinien być potencjalny wpływ na decyzje (zarówno urzędników, jak i tych podejmowanych przez obywateli) a nie poziom klasyfikacji ryzyka.

Link do tabeli zawierającej informacje o systemach AI zgłoszonych przez instytucje publiczne w ramach tego badania.

Źródła

- Zamieszczone poniżej linki prowadzą do naszych zasobów.
- Korespondencja w sprawach dot. wykorzystywania AI
- Zebrane informacje o systemach AI
- AI Act - Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 - z dnia 13 czerwca 2024 r.
- Generatywna sztuczna inteligencja w służbie pracowników administracji publicznej - pierwsze kroki
- Przewodnik po sztucznej inteligencji dla administracji publicznej
- PLLuM: A Family of Polish Large Language Models
- Ustawa o dostępie do informacji publicznej
- Exploring automation bias in human–AI collaboration: a review and implications for explainable AI
- AI w e-administracji publicznej – perspektywa urzędników i instytucji
- Podstawowe źródło: Odpowiedzialne wykorzystywanie sztucznej inteligencji - faza 2
- Opracowanie: Sieć Obywatelska Watchdog Polska, ostatnia aktualizacja: 21 kwietnia 2026.