

Generatywna sztuczna inteligencja – rekomendacje dla pracowników administracji publicznej

Projekt do konsultacji

Wstęp

W ostatnich kilkunastu miesiącach obserwujemy dynamiczny wzrost dostępność narzędzi opartych o sztuczną inteligencję, które na podstawie poleceń użytkownika generują i przetwarzają tekst, obrazy, filmy i inne rodzaje danych. Te narzędzia określa się jako tzw. generatywną sztuczną inteligencję (dalej: GenAI), którą to nazwę stosuje się zamiennie z określeniami takimi jak duże modele językowe (Large Language Models, LLM).

Pojawienie się i upowszechnienie takich narzędzi z stwarza możliwość ich zastosowania do powtarzalnych zadań służbowych. Jednocześnie, ich używanie niesie ze sobą zagrożenia związane m.in. z przekazywaniem błędnych informacji, tworzeniem treści niezgodnych z prawem czy bezpodstawnym przetwarzaniem lub ujawnianiem danych osobowych, treści niejawnych lub utworów objętych prawami autorskimi.

Rosnąca popularność takich narzędzi sprawia, że w praktyce są coraz częściej wykorzystywane przez pracowników - w tym urzędników administracji publicznej. Chcąc zaadresować tę kwestię Ministerstwo Cyfryzacji przygotowało ten poradnik. Jego głównym celem, obok prezentacji tej technologii oraz jej użyteczności, wsparcie pracowników sektora publicznego w minimalizowaniu ryzyka niewłaściwego i potencjalnie szkodliwego wykorzystania GenAI.

Poradnik składa się z części, które opisują kolejno: definicje i podstawowe terminy, przykłady zastosowań, jak również wskazówki dla pracowników administracji rządowej dotyczące tego, w jaki sposób bezpiecznie korzystać z narzędzi GenAI dla usprawnienia bieżącej pracy. Ostatnia część jest najbardziej techniczna i opisuje rekomendacje użycia GenAI zależnie od modelu dostępu oraz dalsze technologiczne rekomendacje, które warto wziąć pod uwagę przy wyborze systemów sztucznej inteligencji do użycia i zakupienia do urzędu.

Mając świadomość tempa rozwoju technologii AI, wytyczne będą poddawane regularnemu przeglądowi. Po raz pierwszy odbędzie się to w I kwartale 2025 r.

W imieniu Departamentu Badań i Innowacji Ministerstwa Cyfryzacji prosimy o przesyłanie uwag, pytań lub sugestii na adres e-mail: sekretariat.DBI@cyfra.gov.pl

Czym jest generatywna sztuczna inteligencja?

Generatywna sztuczna inteligencja to technologia, która, za pomocą poleceń (ang. *prompt*), pozwala użytkownikom tworzyć nowe treści takie jak teksty, obrazy czy filmy. Algorytmy modeli generatywnej sztucznej inteligencji działają w oparciu o dostarczone na etapie tworzenia narzędzia dane w procesie nazywanym „trenowaniem”.

Najczęściej używane narzędzia generatywnej AI w odpowiedzi na polecenie generują teksty, obrazy albo filmy. Należy pamiętać, że jakość wygenerowanych na zadane pytanie treści jest zatem wprost zależna od jakości danych – nie tylko tych, na których dane narzędzie było trenowane, ale także tych wpisywanych jako polecenia przez użytkowników. Z tego powodu ważne jest, aby przed wydaniem takich komend, dane w nich zawarte były zweryfikowane pod kątem ich poprawności.

W skrócie: o czym pamiętać korzystając z GenAI w celach służbowych

Podobnie jak w przypadku innych narzędzi cyfrowych, to użytkownik będzie odpowiadał za sposób oraz skutki ich wykorzystania. Przy zachowaniu odpowiednich środków ostrożności, GenAI może pomóc wykonywać niektóre, rutynowe zadania służbowe. Aby te zastosowanie nie naruszało przepisów prawa konieczne jest jednak przestrzeganie podstawowych zasad.

Nie należy wprowadzać do takich narzędzi informacji lub dokumentów, które

- są niejawnie lub zawierają dane wrażliwe.
- zawierają informacje urzędowe będące w fazie przygotowawczej, a tym samym nie przeznaczone do upublicznienia;
- naruszają przepisy o ochronie danych osobowych.

Dodatkowo, przy otrzymaniu wyników wprowadzonych poleceń należy:

- Za każdym razem weryfikować ich rezultaty - narzędzia GenAI są podatne na stroniczość i dezinformację (tzw. *Halucynacje*).
- W przypadku wykorzystania wyników w dalszej pracy, zawsze informować, że treści, które są przekazywane były stworzone lub przetworzone z wykorzystaniem narzędzi GenAI (np. oznaczać cytatem fragmenty wprost zaczerpnięty z takiego narzędzia).

Jak Generatywna Sztuczna Inteligencja może pomóc w pracy?

GenAI to narzędzie z dużym potencjałem zastosowania, które przy zachowaniu należytej staranności może być pomocne w usprawnieniu realizowanych zadań. Poniżej przedstawione są przykłady bezpiecznego i zgodnego z zasadami zastosowania programów GenAI.

Przykład nr 1: Źródło inspiracji

Szukając pomysłu na nazwę, strukturę lub wstępną treść programu, projektu, dokumentu, artykułu, prezentacji, warsztatu, newslettera lub wpisu w kanałach społecznościowych, narzędzia GenAI mogą posłużyć jako przydatna pomoc na początku procesu twórczego.

W takich zadaniach narzędzia GenAI mogą ułatwić rozpoczęcie pracy. Trzeba pamiętać, że należy możliwie dokładnie opisać konkretne zadanie. W takim opisie warto uwzględnić nie tylko treść merytoryczną zapytania, ale także kontekst: styl, długość, charakterystykę grupy odbiorców.

Należy pamiętać, aby sprawdzać czy wygenerowane pomysły nie są błędne, stronnicze lub dyskryminujące. Jeżeli wynik działania systemu ma być przekazany dalej, wskazane jest oznaczenie, które treści zostały przygotowane ze wsparciem sztucznej inteligencji.

Przykład nr. 2: Streszczanie informacji

Generatywna sztuczna inteligencja dobrze radzi sobie ze streszczaniem obszernych materiałów, takich jak artykuły naukowe lub prasowe. Z pomocą GenAI można m.in. uzyskać szybką analizę, najistotniejsze tezy czy krótkie podsumowanie zagadnienia, o którym mowa jest w danej bazie danych lub dokumencie. Materiał zawierający streszczenia przeszukiwanych źródeł można np. dodać jako załącznik do materiału tezewego, lub jako materiał wstępny służący przygotowaniu komunikatu prasowego.

Należy pamiętać, aby zawsze weryfikować, czy streszczenie zgadza się z oryginałem bądź źródłem informacji. Wymaga to sprawdzenia czy nie pojawiły się żadne informacje, których nie ma w podanym materiale. Jeżeli dokument ma być przekazany dalej, należy również oznaczyć to, że został on przygotowany ze wsparciem sztucznej inteligencji.

Przykład 3: Wstępne rozpoznanie tematu

GenAI pomaga w zebraniu podstawowych informacji na nowy, nieznanый temat. Przykładowo, chcąc dowiedzieć się jakie są najnowsze trendy rynkowe w mało znanym obszarze technologii, jak wygląda historyczna perspektywa na pewien problem, w jaki sposób najlepiej skonstruować komunikat prasowy albo jak dany

problem adresowany jest w innych państwach lub gdzie znaleźć wymagane źródła - narzędzia GenAI mogą posłużyć jako przydatny, pierwszy krok.

GenAI może pomóc znaleźć potrzebne informacje, odpowiedzieć stosowane techniki podejścia do problemu, które można wykorzystać w dalszej analizie zagadnienia. Można także dopytać GenAI w jakich źródłach szukać dalszych informacji na ten temat i poprosić o dostosowanie źródeł do konkretnych potrzeb.

Należy pamiętać, że GenAI nie działa jak internetowa wyszukiwarka, a przez to jej odpowiedzi potrafią łączyć informacje prawdziwe z fałszywymi. Dlatego rezultaty zapytań należy traktować jako wstęp do dalszych analiz, a nie ostateczną odpowiedź.

Przykład 4: Tłumaczenie lub redakcja treści

Ze względu na ilość oraz łatwość zdobycia danych najpopularniejsze narzędzia GenAI były tworzone oraz uczone w przeważającej mierze w języku angielskim. Sprawia to, że funkcje przetwarzania tego języka są rozbudowane, a przez to mogą być bardziej pomocne w zadaniach takich jak tłumaczenia treści czy redakcja wiadomości.

Takie programy i narzędzia pomagają np. w tworzeniu propozycji odpowiedzi na wiadomości mailowe. Wpisując w pole komend treść zadania (np. „przygotuj szablon zaproszenia na konferencję”, „przetłumacz treść otrzymanej wiadomości”) można, w przypadku wątpliwości, uzyskać potwierdzenie, że dobrze zrozumieliśmy lub skonstruowaliśmy konkretną treść.

Należy pamiętać, że jeśli chcemy korzystać z narzędzi automatycznego tłumaczenia tekstów lub dokumentów istnieją dedykowane takim funkcjom rozwiązania wykorzystujące AI, które nie są modelami i systemami generatywnymi.

Jak poprowadzić konstruktywną „rozmowę” z GenAI

Generatywna sztuczna inteligencja już dziś jest wykorzystywana jako przydatne narzędzie w pracy umysłowej. Im bardziej precyzyjnie określi się zadanie do wykonania, tym lepszy powinien być ostateczny rezultat.

Poniżej lista praktycznych wskazówek, jak najefektywniej wydawać polecenia (ang. Prompt, prompting)

Precyzyjny kontekst

W promptach warto podawać kontekst problemu, aby GenAI mogła lepiej zrozumieć intencje zapytania i przygotować odpowiednią odpowiedź.

Komentarz [KF1]: Tutaj skończyłem, na dziś już odpadam.

@Krzyrkowska Pamela

Zmieniałem przede wszystkim formę - moim zdaniem forma "Ty, Ciebie" nie przejdzie ani na etapie resortu, ani urzędowych odbiorców. Dlatego zmieniłem wszystkie formy na bezosobowe (moim zdaniem to bardziej neutralne, a niekoniecznie biurokratyczne).

Na przykład: Poinformuj GenAI, że jest się pracownikiem konkretnej instytucji, np. kuratorium oświaty. Wyjaśniając, że zadaniem programu jest sprawdzenie w jaki sposób aktualnie obowiązujący system ocen w polskich szkołach pozwala na weryfikację rzeczywistych postępów ucznia w nauce. W kolejnych krokach zamiast ogólnych pytań (np. "Jak to działa?"), lepiej jest zapytać: „Jak aktualnie obowiązujący system nauczania w szkołach podstawowych w Polsce pozwala na weryfikację postępów uczniów?”.

Prosić o przykłady i podawać informacje dodatkowe

Jako, że modele GenAI uczyły się w większości na treściach tworzonych przez ludzi, w ich używaniu obowiązują podobne zasady, jak w komunikacji międzyludzkiej. Aby lepiej zrozumieć dany temat, warto zadawać pytania, prosić o podawanie konkretnych przykładów.

Na przykład: Chcąc dowiedzieć się więcej o zrównoważonym rolnictwie w Polsce, można zapytać np.: "przedstaw mi przykłady zrównoważonych praktyk w rolnictwie". Dla dodania kontekstu lepiej jest podzielić się z GenAI znanymi już przykładami. Pozwolą one ukierunkować GenAI w tworzeniu materiału bardziej odpowiadającego intencjom pytającego. Należy również informacje zwrotne – podtrzymywać wymianę informacji („rozmowę”). Jeżeli otrzymana z narzędzia nie jest satysfakcjonująca – np. ze względu na zbyt wysoki poziom ogólności, pomocna może być dodatkowa odpowiedź, taka jak np.: „Zaproponowane przykłady są zbyt mało konkretne. Podaj więcej informacji na temat użytych metod i opisz je w punktach”.

Eksperymentowanie i próbowanie różnych form poleceń często pozwala na osiągnięcie najlepszych rezultatów.

Na przykład: Dobrze jest otwarcie napisać, co w odpowiedzi GenAI nie spełnia oczekiwań. Przykładowo, można wskazać: „Materiał jest zbyt obszerny i napisany zbyt formalnym językiem. Przygotuj tekst nie dłuższy niż 1000 znaków, w tym spacje, zachowaj treść, ale opisz zagadnienie mniej formalnym językiem”.

Określenie wymaganego tonu i styl

Opisanie odbiorcy lub grupy odbiorców, do których kierowany jest oczekiwany materiał ma duże znaczenie. Ważne, by poinformować GenAI na przykład, że przygotowywana korespondencja jest korespondencją formalną, przewidzianą do dystrybucji w trybie obiegowym do urzędników administracji publicznej. Można także poprosić o wygenerowanie mniej formalnego emaila na zadany temat, kierowanego do współpracowników. W zależności od zadania, do polecenia można dołączyć (wkleić) przyjęty danym urzędzie tekstowy szablon dokumentu lub przykład struktury i języka tekstu, na bazie którego tworzony jest materiał.

Osadzenie modelu w roli

Dla polepszenia jakości wyników, można poprosić model GenAI, aby wnioskował i odpowiadał w pewien konkretny sposób. W ten sposób szansa, że będzie on generował treści pasujące do danej roli czy osoby wzrośnie.

Na przykład: Polecenie modelowi: „Odpowiadaj jak osoba bez specjalistycznej wiedzy z zakresu sztucznej inteligencji” pomoże stworzyć materiały zrozumiałe dla wybranej grupy odbiorców albo lepiej zrozumieć osobę do której kierowany jest przekaz.

Wymaganie podania wariantów

GenAI, jako system bazujący na rozpoznawaniu powtarzających się wzorców, bardzo dobrze radzi sobie z wymienianiem wariantów i proponowaniem alternatyw. Warto korzystać z tej funkcji – dzięki niej można wybierać najbardziej odpowiednie fragmenty z kilku otrzymanych wyników.

Na przykład: szukając dobrej nazwy dla nowego projektu na program dla przedsiębiorstw lub obywateli warto prosić o więcej jedną propozycję. Zadaniując program GenAI, można wskazać potrzebę podania np. 10 propozycji. Jeżeli propozycje nie są satysfakcjonujące, polecenie można doprecyzować: „Te propozycje są zbyt mało interesujące. Stwórz 10 propozycji, które lepiej trafią do młodego grona odbiorców”.

O czym należy pamiętać, korzystając z GenAI w celach służbowych

Modele generatywnej sztucznej inteligencji mają swoje ograniczenia. Wynikają one ze sposobu ich trenowania, rodzaju danych, na których się uczyły oraz ich architektury technicznej. Tak jak inne rozwiązania informatyczne, także modele GenAI bywają zawodne. Podobnie jak w przypadku innych narzędzi cyfrowych to użytkownicy odpowiadają za sposób i skutki korzystania z nich.

Wchodząc w interakcję z wybranym narzędziem GenAI należy zachować podstawowe zasady higieny cyfrowej. Należy być równie ostrożnym jak wtedy, gdy korzysta się ze zwykłej wyszukiwarki lub gdy zamieszcza się publiczne treści w mediach społecznościowych.

O czym należy pamiętać, gdy wchodzi się w interakcję z GenAI - zwłaszcza w celach służbowych?

Ochrona danych

Korzystając z jednego z popularnych ogólnodostępnych modeli językowych, **nie mamy kontroli nad danymi**, które są mu przekazywane. **Nie wiemy też do czego mogą być w przyszłości użyte przez twórcę systemu.** Ogólnodostępne w sieci modele językowe należy traktować analogicznie jak wyszukiwarkę online albo media społecznościowe.

Do ogólnodostępnych w sieci narzędzi GenAI nie należy wpisywać informacji, które:

- są niejawnie lub zawierają dane wrażliwe;

- zawierają wewnętrzne informacje urzędowe, będące w fazie przygotowawczej, a tym samym nie przeznaczone do upublicznienia;
- zawierają dane osobowe, których ujawnienie naruszyłoby przepisy o ochronie danych osobowych.

Zasady te mogą być mniej restrykcyjne w odniesieniu do dokumentów urzędowych jeśli GenAI jest częścią architektury informatycznej urzędu i objęty systemem cyberbezpieczeństwa instytucji.

Halucynacje

Ze względu na sposób trenowania modeli GenAI mają one tendencje do produkowania halucynacji, czyli generowania nieprawdziwych treści. Modele GenAI mogą tworzyć treści wyglądające bardzo realistycznie (np. opisy wydarzeń, które nigdy nie miały miejsca), podając nawet fałszywe źródła informacji dla tych faktów.

Dlatego **zawsze należy weryfikować informacje uzyskane od modeli językowych.**

Stronniczość

Modele GenAI, jak wszystkie modele sztucznej inteligencji są trenowane na pewnych danych, które nie zawsze są reprezentatywne. Modele GenAI mogą być stronnicze i dyskryminować ze względu na płeć, pochodzenie, rasę, wyznanie, wiek czy wiele innych cech. Jest to bezpośrednio powiązane z jakością wprowadzanych podczas treningu modelu danych, a czasem także związane jest z preferencjami twórcy.

Dlatego zawsze należy weryfikować materiały uzyskane z GenAI pod **względem ich reprezentatywności i bezstronności.**

Manipulacja

Narzędzia GenAI mają potencjał do generowania treści (tekst, audio, zdjęcia, wideo), które do złudzenia przypominają styl danej osoby (np. sposób wypowiedzi, głos czy wizerunek) albo jakiś obiekt, czy jego otoczenie (może to np. być zdjęcie osoby przedstawionej w jakiejś niezręcznej sytuacji).

Model może "chcieć" stworzyć fałszywy obraz chcąc jak najlepiej sprostać postawionym oczekiwaniom.

Dlatego zawsze należy weryfikować zgodność z oryginałem, nie przyczyniać się do rozpowszechniania dezinformacji i **oznaczać, że stworzone treści są wygenerowane za pomocą sztucznej inteligencji.**

Naruszenia własności intelektualnej

Modele językowe mogą przedstawić materiał podobny lub identyczny z treściami objętymi ochroną praw własności intelektualnej lub przemysłowej. Może być to efektem przypadku, ale także tego, że model był uczony na bazach danych zawierających dzieła objęte taką ochroną.

Jeżeli tworzymy publikację korzystając z pomocy GenAI, zawsze **należy sprawdzić czy dany utwór lub znak nie jest objęty ich ochroną. Jeśli tak - należy podać źródło i - jeśli to przewidziane prawnie - wnieść opłatę za wykorzystanie materiału lub w inny sposób spełnić wymagania prawne.**

Rekomendacje użycia GenAI w zależności od modelu dostępu

Modele sztucznej inteligencji, podobnie jak inne usługi informatyczne, mogą być dostępne w różnych modelach wdrożeniowych, na przykład w chmurze albo na lokalnej infrastrukturze. Zależnie od tego, jak dostępny jest dany model sztucznej inteligencji, powinniśmy decydować, czy i jak możemy go użyć w pracy oraz na co zwrócić uwagę. Poniżej przedstawiamy trzy główne scenariusze dostępu do modeli generatywnej sztucznej inteligencji wraz z potencjalnymi zagrożeniami i rekomendacjami wynikającymi ze sposobu wdrożenia modelu i dostępu do niego.

GenAI w chmurze bez dedykowanego dostępu

Narzędzia GenAI są często dostępne w sieci za darmo lub za drobną opłatą przez przeglądarkę albo aplikacje webowe czy na smartfonie. Choć taki model użycia jest atrakcyjny, ponieważ jest często bezpłatny i łatwo dostępny, należy pamiętać, że w przypadku jego użycia nie mamy kontroli nad danymi, które wysyłamy, ani nie wiemy, do czego będą one w przyszłości użyte przez twórcę systemu. Twórcy często informują w regulaminach, że prompty, rozmowy i wszystkie ich metadane mogą być użyte do dalszego trenowania modelu. Dostępne modele traktujemy analogicznie jak wyszukiwarkę online albo media społecznościowe. W szczególności:

1. Nie należy wpisywać do takich narzędzi żadnych wewnętrznych informacji urzędowych, które nie są publicznie dostępne dla wszystkich obywateli i obywaterek.
2. Nie należy wpisywać do takich narzędzi żadnych informacji niejawnych, kontrolowanych wrażliwych danych urzędowych ani kontrolowanych danych urzędowych (np. objętych RODO).
3. Nie powinno się do takich narzędzi wpisywać niczego, co nie powinno być publicznie dostępne, dla wszystkich.

GenAI w chmurze z dedykowanym dostępem

Dostawcy narzędzi GenAI często oferują swoim klientom możliwość dedykowanego dostępu w publicznej chmurze obliczeniowej. Taki model jest bezpieczniejszy niż użycie otwartego serwisu, jednak nadal dane wpisywane do modelu są przetwarzane poza urzędem i jego infrastrukturą obliczeniową. Korzystając z takiego typu wdrożenia, traktuj taki serwis jak każdą inną dedykowaną usługę chmurową w urzędzie. W szczególności:

1. Należy wpisywać tam tylko informacje, które mogą być przetwarzane w publicznej chmurze obliczeniowej, zgodnie z centralnymi rekomendacjami rządowymi.
2. Należy wpisywać tam tylko informacje jawne i nieobjęte kontrolą.
3. Nie należy wpisywać tam żadnych informacji niejawnych, kontrolowanych wrażliwych danych urzędowych ani kontrolowanych danych urzędowych.

GenAI na infrastrukturze urzędowej

Jednym z podstawowych modeli wdrożenia systemów informatycznych w administracji, w tym sztucznej inteligencji, jest utrzymywanie ich na urzędowej infrastrukturze obliczeniowej. Taki sposób wdrożenia często nazywa się wdrożeniem on-premise lub wdrożeniem lokalnym.

W takim modelu traktujemy serwis jak każdy inny dostępny na infrastrukturze wewnętrznej urzędu. Jeśli taki serwis jest odpowiednio zabezpieczony pod względem cyberbezpieczeństwa oraz zasad minimalnego dostępu do danych w urzędzie, daje nam to najwięcej swobody w pracy z danymi urzędowymi. W takim przypadku należy zwrócić uwagę jedynie na opisane powyżej rekomendacje związane z charakterystyką generatywnej sztucznej inteligencji.

Jaki wybrać i przygotować system?

Przy wyborze systemów sztucznej należy zwrócić szczególną uwagę na bezpieczeństwo i przejrzystość. Rekomendacje w tym zakresie są szczegółowo opisane w Akcie o sztucznej inteligencji¹. Akt należy stosować bezpośrednio, a przed uchwaleniem polskiej ustawy wdrażającej warto przygotować się do nowych standardów poprzez spełnienie poniższych dobrych praktyk w zakresie modeli AI.

Bezpieczeństwo i przejrzystość

1. Wszystkie systemy sztucznej inteligencji wykorzystujące dane osobowe lub inne dane o istotnym znaczeniu dla sektora publicznego powinny spełniać wymagania Aktu o sztucznej inteligencji dla systemów wysokiego ryzyka:
 - a) Zarządzanie ryzykiem przez cały cykl życia systemu,
 - b) Zarządzanie danymi (trening, walidacja, testowanie), procesem ich zbierania i ich skrzywieniem (bias),
 - c) Posiadanie aktualnej dokumentacji technicznej systemu i sprzętu,
 - d) Stałe monitorowanie modelu,
 - e) Działania naprawcze i informowanie w przypadku niezgodności.
2. Przed wdrożeniem model powinien być zweryfikowany pod względem stronniczości - czy nie dyskryminuje konkretnej grupy z jakiegokolwiek względu.
3. Przed wdrożeniem model powinien zostać poddany dokładnym testom wydajności, zgodności z celem i procesami w jednostce organizacyjnej.
4. W procedurze zamawiania rozwiązań AI (nie tylko w rozumieniu zamówień publicznych) można zastosować klauzule rekomendowanych przez Komisję Europejską (dokument w języku polskim: https://public-buyers-community.ec.europa.eu/sites/default/files/2023-10/AI_Procurement_Clauses_Template_NON_HIGH_RISK_PL.pdf).

Cyberbezpieczeństwo

1. Wszystkie systemy sztucznej inteligencji wykorzystujące dane osobowe lub inne dane zawierające informacje o istotnym znaczeniu dla sektora publicznego powinny mieć zapewnione warunki:

¹ Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji). *Dz.U. L*, 2024/1689, 12.7.2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>

- a) solidności technicznej i najwyższego poziomu cyberbezpieczeństwa,
 - b) zgodności z przepisami ustawy z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa²
2. Środowisko modelu powinno być zabezpieczone w celu uniemożliwienia ataków typu reverse-engineering, data-poisoning czy DDoS.
 3. Retrenowanie modelu powinno odbywać się tylko na uporządkowanych danych zweryfikowanych statystycznie.

System godny zaufania

1. Wszystkie systemy sztucznej inteligencji wykorzystujące dane osobowe lub inne dane zawierające informacje o istotnym znaczeniu dla sektora publicznego powinny mieć zapewnione warunki:
 - a) przejrzystości modelu,
 - b) efektywnego nadzoru człowieka.
2. Wdrożenie systemów AI powinno być poprzedzone konsultacjami oraz badaniami typu user experience przeprowadzonymi na grupie urzędników mających korzystać z tych systemów oraz ich docelowych użytkowników (np. obywateli, przedsiębiorstw, organizacji pozarządowych). Wnioski z konsultacji i badań powinny być uwzględnione przed końcowym wdrożeniem rozwiązania.
3. Domyślnie modele nie powinny być trenowane na danych osobowych.
4. Decyzje modelu powinny być szczegółowo wyjaśniane i powszechnie dostępne.

Rekomendacje technologiczne

1. Należy używać modeli na licencji typu open-source zamiast technologii zamkniętych.
2. Należy stosować zasadę minimalizacji skomplikowania algorytmicznego – powinna być używana najprostsza technologia adresująca konkretne zadanie.
3. Należy używać modeli typu glass-box zamiast modeli typu black-box.
4. Procesy używania i budowania systemów powinny być prowadzone zgodnie z istniejącymi i tworzącymi się normami, na przykład normą ISO/IEC 42001.

² Dz. U. z 2024 r. poz. 1077 z późn. zm.