|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Nazwa dokumentu:***  ***„***[***Projekt dokumentu zawierającego rekomendacje użycia generatywnej sztucznej inteligencji w urzędach***](https://www.gov.pl/web/krmc/projekt-dokumentu-zawierajacego-rekomendacje-uzycia-generatywnej-sztucznej-inteligencji-w-urzedach)***” - wnioskodawca Minister Cyfryzacji.*** | | | | | |
| **Lp.** | **Organ wnoszący uwagi** | **Jednostka redakcyjna, do której wnoszone są uwagi** | **Treść uwagi** | **Propozycja zmian zapisu** | **Odniesienie do uwagi** |
| **1.** | **MSWiA** | **Wstęp**  str. 2  Chcąc zaadresować tę kwestię Ministerstwo Cyfryzacji przygotowało ten poradnik.  Mając świadomość tempa rozwoju technologii AI, wytyczne będą poddawane regularnemu przeglądowi. | Poradnik, rekomendacje, wytyczne – pod rozwagę uspójnienie terminologii i posługiwanie się jednym terminem na potrzeby dokumentu. Może tym, który został wskazany w tytule tj. rekomendacje.  Ponadto pod rozwagę dodanie, że nie mają one wiążącego charakteru, o ile tak jest. |  |  |
| **2.** |  | **W skrócie: o czym pamiętać korzystając z GenAI w celach służbowych**  str. 3 | Zanim umożliwi się pracownikom skorzystanie  z rozwiązań AI, to w pierwszej kolejności powinny one zostać poddane weryfikacji przez specjalistów w dziedzinie AI i systemowo zostać zarekomendowane dla administracji np.  w postaci konkretnych programów, produktów, usług itp.  Nie należy przerzucać odpowiedzialności na najniższy szczebel – „rekomendując” jedynie dane narzędzie.  Większość dostępnych obecnie generatywnych modeli AI działa w chmurze. Także w chmurach nie podlegających jurysdykcji polskiej, czy jurysdykcji państwa członkowskiego Unii Europejskiej. Zwykle chmury te nie są także  w stanie zagwarantować prawa właściciela do swojej informacji. Należy mieć na uwadze, że  w przypadku korzystania z takich rozwiązań urząd administracji publicznej traci prawa do wytworzonych w ten sposób informacji. Informacji, które mogą być informacjami istotnymi dla realizacji zadań statutowych instytucji administracji publicznej.  W kontekście tworzenia zapytań do publicznych LLM, proponuje się uwzględnienie w wyliczeniu kategorii informacji niepodlegających upublicznieniu, które mogą dotyczyć m.in. np:  szczegółów lub podatności infrastruktury IT podmiotu  elementów polityk bezpieczeństwa podmiotu lub treści korespondencji wewnętrznej podmiotu (w kontekście redagowania treści przez AI), | „Nie należy wprowadzać do takich narzędzi informacji lub dokumentów, które:   * są niejawne lub zawierają dane wrażliwe. * zawierają informacje urzędowe będące w fazie przygotowawczej, a tym samym nie przeznaczone do upublicznienia; * naruszają przepisy o ochronie danych osobowych. * **ujawnione mogą doprowadzić do podniesienia poziomu ryzyk związanych z bezpieczeństwem informacji i cyberbezpieczeństwem.”.** |  |
| **3.** |  | **Jak Generatywna Sztuczna Inteligencja może pomóc w pracy?**  str. 4 | Czy lista bezpiecznych programów GenAI będzie dostępna, w szczególności tych, które pozwalają na uzyskanie poniższych propozycji zastosowania GenAI?  W jaki konkretnie sposób pracownik ma sprawdzić czy wygenerowane pomysły nie są błędne, stronnicze lub dyskryminujące? Czy zostanie przygotowane jakieś wsparcie jak to sprawdzać? Czy pracownik ma się zdać na własną ocenę, intuicję?  Warto zwrócić uwagę na problem stronniczości AI, który w głównej mierze powstaje na etapie projektowania i tworzenia. Pracownik na etapie stosowania może nawet nie być świadom tego, że wynik zapytania wyszedł stronniczy, błędny lub dyskryminujący. Dlatego rozwiązania, które zostaną zaproponowane pracownikom powinny zostać zweryfikowane przez specjalistów  w dziedzinie AI. |  |  |
| **4.** |  | **Jak prowadzić konstruktywną „rozmowę” z GenAI**  **Precyzyjny kontekst**  str. 5-6 | Przykład: *„Poinformuj GenAI, że jest się pracownikiem konkretnej instytucji…”.*  Kontekst oczywiście jest ważny dla tego typu rozwiązań, ale w ramach różnego rodzaju szkoleń zwraca się uwagę na to, aby pracownicy  z rozsądkiem podchodzili do ujawniania informacji o sobie oraz o miejscu pracy. Wynika to z rosnącej liczby zagrożeń związanych  z wyłudzeniami danych, oszustwami, phishingu, vishingu. Może zatem dojść do sytuacji w której będziemy mu rekomendowali sprzeczne zalecenia.  Dopóki nie zostanie opracowana zweryfikowana lista programów/rozwiązań/narzędzi, z której będą mogli korzystać pracownicy administracji, to nie możemy sobie tym zakresie pozwolić na dowolność. W gruncie rzecz większość z tych rozwiązań, to rozwiązania komercyjne, co do których pracownicy nie mają wiedzy w jaki sposób przetwarzane są w nich informacje, gdzie  i czy w ogóle przechowywane oraz do czego potencjalnie mogą zostać wykorzystane np.  w wyniku cyberataku na dostawcę takiego rozwiązania.  W związku z powyższym proponuje się zmianę treści porad dotyczących pisania promptów  w zakresie rekomendowania zachowania maksymalnej anonimowości w zapytaniach kierowanych do publicznych GenAI, dotyczących danych osobowych jak i informacji o instytucji.  Część modeli językowych jest uczona na podstawie wprowadzonych zapytań co może skutkować zbieraniem wewnętrznych danych  o Podmiocie przez osoby nieupoważnione, cyberprzestępców lub obce służby. | W przykładzie proponuje się wykreślenie zdania „Poinformuj GenAI, że jest się pracownikiem konkretnej instytucji, np. kuratorium oświaty.” |  |
| **5.** |  | **Jak prowadzić konstruktywną „rozmowę” z GenAI**  **Określenie wymaganego tonu i stylu?**  str. 6 | *„Można także poprosić o wygenerowanie mniej formalnego emaila na zadany temat, kierowanego do współpracowników. W zależności od zadania, do polecenia można dołączyć (wkleić) przyjęty w danym urzędzie tekstowy szablon dokumentu lub przykład struktury i języka tekstu, na bazie którego tworzony jest materiał.”*  W tym kontekście pod rozwagę należy poddać określenie katalogu spraw/wątków, w których tego typu pomoc mogłaby mieć zastosowanie, bez narażania się pracownika na udostępnienia wrażliwych informacji do takiego narzędzia/programu/usługi.  Pod rozwagę poddaje się jednak wskazanie jakie szablony dokumentów mogą być „dołączone” – wklejone. |  |  |
| **6.** |  | **O czym należy pamiętać, korzystając z GenAI w celach służbowych**  str. 7 | „*Wchodząc w interakcję z wybranym narzędziem GenAI należy zachować podstawowe* [*zasady higieny cyfrowej*](https://www.gov.pl/web/baza-wiedzy/cyberbezpieczenstwo)*.”*  Pod rozwagę poddaje się wskazanie zasad higieny cyfrowej. |  |  |
| **7.** |  | **O czym należy pamiętać, gdy wchodzi się w interakcję z GenAI - zwłaszcza w celach służbowych?**  **Ochrona danych**  str. 7-8 | *„Korzystając z jednego z popularnych ogólnodostępnych modeli językowych,* ***nie mamy kontroli nad danymi****, które są mu przekazywane.”*  Pod rozwagę poddaję się nie skupianie się tylko na popularnych rozwiązaniach, bo ten rynek szybko się rozwija.  Zgodnie z Art. 2 pkt 1 Dyrektywy Parlamentu Europejskiego i Rady (UE) 2016/943 z dnia 8 czerwca 2016 r. w sprawie ochrony niejawnego know-how i niejawnych informacji handlowych (tajemnic przedsiębiorstwa) przed ich bezprawnym pozyskiwaniem, wykorzystywaniem i ujawnianiem pod pojęciem "tajemnica przedsiębiorstwa" rozumie się „informacje, które spełniają wszystkie następujące wymogi: a) są poufne w tym sensie, że jako całość lub w szczególnym zestawie i zbiorze ich elementów nie są ogólnie znane lub łatwo dostępne dla osób  z kręgów, które zwykle zajmują się tym rodzajem informacji; b) mają wartość handlową dlatego, że są objęte tajemnicą; c) poddane zostały przez osobę, która zgodnie z prawem sprawuje nad nimi kontrolę, rozsądnym, w danych okolicznościach, działaniom dla utrzymania ich  w tajemnicy”.  Jest to zakres znacznie szerszy, niż zakres informacji opisanych w rekomendacjach, jako dane wymagające ochrony.  Tymczasem, podczas korzystania z rozwiązań GenAI może dość do przekazywania informacji spełniających wymogi "tajemnicy przedsiębiorstwa" jako elementów zapytań. Nie musi to tylko dotyczyć informacji urzędowych będących w fazie przygotowawczej, jak wskazują to rekomendacje. | Pod rozwagę modyfikacja na: „Korzystając  z ogólnodostępnych modeli językowych….”  Wskazane jest, by rozszerzony został katalog informacji uznanych, za informacje istotne dla realizacji działań statutowych instytucji administracji publicznej, udostępniane bez ograniczeń pracownikom instytucji lub na podstawie porozumień o zachowaniu poufności, które powinny podlegać ochronie.  Brak jest także wskazania wymaganej prawem ochrony dla danych stanowiących dane referencyjne krajowych rejestrów, które są określone w odrębnych przepisach (w tym np. dane  o kluczowym znaczeniu dla bezpieczeństwa publicznego). |  |
| **8.** |  | **Halucynacje**  str. 8 | *„Dlatego* ***zawsze należy weryfikować informacje uzyskane od modeli językowych.”***  Większość urzędów nie posiada specjalistów  w zakresie AI, więc będzie to zalecenie trudne do realizacji. Chyba, że zostanie administracji zalecone jakieś narzędzie do np. factchecingu.  ***„****Dlatego zawsze należy weryfikować materiały uzyskane z GenAI pod* ***względem ich reprezentatywności i bezstronności.”***  Pracownik na etapie stosowania może nawet nie być świadom tego, że wynik zapytania wyszedł stronniczy. Dlatego rozwiązania które zostaną zaproponowane pracownikom powinny zostać zweryfikowane przez specjalistów w dziedzinie AI. |  |  |
| **9.** |  | **Manipulacja**  str. 8 | *„Dlatego zawsze należy weryfikować zgodność z oryginałem, nie przyczyniać się do rozpowszechniania dezinformacji i* ***oznaczać, że stworzone treści są wygenerowane za pomocą sztucznej inteligencji.”***  To zbyt ogólne zalecenia. Jakimi narzędziami na to być sprawdzane? Jak ma taki materiał być oznaczony? Powinno to zostać określone odgórnie i jednolicie w ramach administracji publicznej. |  |  |
| **10.** |  | **Rekomendacje użycia GenAI w zależności od modelu dostępu**  str. 9-10 | Proponuje się rekomendować rozwiązania  on-premise  (wdrożenia lokalne). Ewentualny dostęp do chmury z dedykowanym dostępem będzie możliwy na zasadach i warunkach określonych  w Standardach Cyberbezpieczeństwa Chmur Obliczeniowych (SCOO) v.1.00.  Komercyjne usługi GenAI w chmurze bez dedykowanego dostępu stanowią zagrożenie udostępniania poufnych danych ze względu na specyfikę korzystania z tego typu narzędzi, zgodnie z uwagą pkt 2. |  |  |
| **11.** | **MSWiA** | **Rozdział: „Cyberbezpieczeństwo”**  str. 12 | Proponuje się rozważenie dopisania do punktu 2 w celu informacyjnym również innych ataków ukierunkowanych na AI i LLM opisanych przez organizacje Mitre.org zajmującą się cyberbezpieczeństwem, albo wskazanie poniższego źródła. Matryca: https://atlas.mitre.org/matrices/ATLAS | Środowisko modelu powinno być zabezpieczone w celu uniemożliwienia ataków typu reverse-engineering, data-poisoning czy DDoS, także ataków ukierunkowanych na LLM opisanych przez organizację Mitre.org <https://atlas.mitre.org/matrices/ATLAS>. |  |
| **12.** |  | **System godny zaufania** str. 12 | Użyto sformułowania „user experience”.  Do wyjaśnienia – urzędowym językiem jest j. polski a nie j. angielski |  |  |
| **13.** | **MSWiA** | **Rozdział „Rekomendacje technologiczne”**Str. 12 | Proponuje się dodanie zalecenia stosowania modeli z popularnych i uznanych w środowisku open-source źródeł (np. Ollama.com, pllum.org.pl itp.). Modele przygotowane przez twórców o niskiej reputacji lub cyerprzestepców mogą propagować dezinformację lub generować niecenzuralne treści.  Do wyjaśnienia pojęcia użyte w języku angielskim (glass-box, black-box) – urzędowym językiem jest j. polski a nie j. angielski. | „Należy używać modeli na licencji typu open-source zamiast technologii zamkniętych, przy czym należy używać modeli z popularnych i uznanych w społeczności open-source źródeł. |  |