Załącznik nr 15. Lista weryfikacyjna stosowania etyki w zakresie sztucznej inteligencji

Lista weryfikacyjna stosowania etyki w zakresie sztucznej inteligencji:

I. **UDZIAŁ I NADZÓR CZŁOWIEKA**

Podstawowe prawa

1. Czy przeprowadziłeś ocenę wpływu systemu na prawa podstawowe, w szczególności w obszarach, gdzie wpływ ten mógłby być negatywny?
2. Czy system będzie współdziałał z decyzjami użytkowników (końcowych) (np. rekomenduje działania/decyzje, przedstawia opcje)?
   A. Czy system może wpływać na autonomię człowieka poprzez zakłócanie procesu podejmowania decyzji przez (końcowego) użytkownika w niezamierzony sposób?
   B. Czy rozważyłeś, czy system powinien przekazywać użytkownikom (końcowym), że decyzja, porada czy wynik jest wynikiem decyzji algorytmicznej*?*

Udział człowieka

3. Czy rozważyłeś podział zadań między system i ludzi w obszarze znaczących zadań oraz zaplanowałeś odpowiedni nadzór i kontrolę systemu przez ludzi?
   A. Czy system sztucznej inteligencji wzmocni lub zwiększy możliwości człowieka?
   B. Czy zaplanowałeś zabezpieczenia, aby zapobiec nadmiernej pewności siebie lub nadmiernemu poleganiu operatora na systemie sztucznej inteligencji w procesach pracy?

Nadzór człowieka

4. Czy rozważyłeś odpowiedni poziom kontroli człowieka dla systemu AI ?
   A. Czy możesz opisać poziom ludzkiej kontroli lub zaangażowania?
   B. Czy zaplanowałeś mechanizmy i środki w celu zapewnienia kontroli lub nadzoru przez ludzi?
5. Czy będzie istniał samouczący się lub autonomiczny system sztucznej inteligencji? Jeśli tak, czy przewidziałeś bardziej szczegółowe mechanizmy kontroli i nadzoru?
   A. Jakie mechanizmy wykrywania i odpowiedzi zaplanowano tak, aby ocenić, czy coś może pójść w niewłaściwym kierunku?

II. **STABILNOŚĆ TECHNICZNA I BEZPIECZEŃSTWO**

Odporność na atak i bezpieczeństwo

6. Czy oceniłeś potencjalne formy ataków, na które system AI może być podatny?
   A. Czy brałeś pod uwagę różne rodzaje i charakter luk w zabezpieczeniach, takich jak zanieczyszczenie danych, infrastruktura fizyczna, ataki komputerowe?
   B. Czy zaplanowałeś środki lub systemy zapewniające integralność i odporność systemu AI na potencjalne ataki?

Plan awaryjny i ogólne bezpieczeństwo

7. Czy przewidziałeś, jakie szkody nastąpiłyby, gdyby system sztucznej inteligencji dokonał niedokładnych prognoz?

Niezawodność i odtwarzalność

8. Czy zaplanowałeś strategię monitorowania i testowania tego, czy system sztucznej inteligencji spełnia cele, zadania i zamierzone aplikacje?
   A. Czy przewidziałeś metody weryfikacji systemu w celu pomiaru i zapewnienia różnych aspektów jego niezawodności i odtwarzalności?

B. Czy zaplanowałeś wprowadzenie procesów opisujących, kiedy system sztucznej inteligencji zawiedzie przy pewnych typach ustawień?

C. Czy zaplanowałeś udokumentowanie i realne wykorzystanie tych procesów w celu testowania i weryfikacji niezawodności systemów AI?

III. **PRYWATNOŚĆ I ZARZĄDZANIE DANYMI**

Poszanowanie prywatności i ochrony danych

9. Czy zaplanowałeś mechanizm umożliwiający użytkownikom zgłaszanie problemów związanych z prywatnością lub ochroną danych w procesach zbierania danych (na potrzeby szkolenia i eksploatacji) i przetwarzania danych w systemie AI?
10. Czy oceniłeś rodzaj i zakres danych w swoich zbiorach danych (np. Czy zawierają one dane osobowe)?
11. Czy rozważyłeś sposoby rozwijania systemu AI lub trenowania modelu bez lub z minimalnym wykorzystaniem potencjalnie wrażliwych danych czy danych osobowych?
12. Czy zaplanowałeś wbudowanie mechanizmów zarządzania danymi osobowymi?
13. Czy zaplanowałeś kroki w celu zwiększenia prywatności, takie jak szyfrowanie, anonimizacja i agregacja?
14. Jeśli istnieje inspektor ochrony danych (IOD) w Twojej organizacji to, czy zaplanowałeś zaangażowanie takiej osoby na wczesnym etapie procesu powstawania systemu?

Jakość i integralność danych

15. Czy system będzie dostosowany do odpowiednich norm (na przykład ISO, IEEE) lub powszechnie przyjętych protokołów zarządzania danymi i nadzorem nad nimi?
16. Czy ustanowiłeś mechanizmy nadzoru dotyczące gromadzenia, przechowywania, przetwarzania i wykorzystywania danych?
17. Czy oceniłeś zakres, w jakim kontrolujesz jakość wykorzystanych zewnętrznych źródeł danych?
18. Czy zaplanowano wdrożenie procesów zapewniających jakość i integralność danych? W jaki sposób weryfikujesz, czy Twoje zbiory danych nie zostały naruszone lub zhakowane?
19. Czy przewidziałeś metodę weryfikacji wyników pracy systemu AI pod kątem zapobiegania stronniczości?

Dostęp do danych

20. Czy oceniłeś, kto może uzyskać dostęp do danych użytkowników i w jakich okolicznościach?
    A. Czy upewniłeś się, że te osoby które mogą uzyskać dostęp do danych posiadają niezbędne kwalifikacje?
    B. Czy zaplanowałeś mechanizm nadzoru umożliwiający rejestrowanie, kiedy, gdzie, jak, przez kogo i w jakim celu uzyskano dostęp do danych?
    C. Czy przewidziałeś procedurę na wypadek uzyskania nieuprawnionego dostępu do danych?

IV. **TRANSPARENTNOŚĆ**

Wytłumaczalność

21. Czy oceniłeś w jakim stopniu można zrozumieć decyzje, a co za tym idzie wyniki pracy systemu AI?
22. Czy od początku projektowałeś system AI z myślą o możliwości interpretacji?

    A. Czy zaplanowałeś przetestowanie najprostszego i najbardziej interpretowalnego modelu możliwego do zastosowania w programie?

    B. Czy zaplanowałeś ocenę swoich danych treningowych i testowych? Czy możesz to zmienić i zaktualizować z biegiem czasu?

    C. Czy zaplanowałeś ocenę, czy po opracowaniu i nauczeniu modelu możesz zbadać możliwość interpretacji jego wyników i czy masz dostęp do wewnętrznego przepływu pracy modelu?

Komunikacja

23. Czy zaplanowałeś mechanizmy informowania użytkowników (końcowych) o zasadach działania systemu AI?

    A. Czy zaplanowałeś przekazanie tego w jasny i zrozumiały sposób docelowej grupie użytkowników?

    B. Czy zaplanowałeś procesy, które uwzględniają opinie użytkowników i wykorzystanie ich do dostosowania systemu?

V. RÓŻNORODNOŚĆ, BRAK DYSKRYMIANCJI I UCZCIWOŚĆ

Unikanie nieuczciwych uprzedzeń

24. Czy ustaliłeś strategię lub zestaw procedur, aby uniknąć tworzenia lub wzmacniania nieuczciwych uprzedzeń w systemie AI, zarówno pod względem wykorzystania danych wejściowych, jak i konstrukcji algorytmu?

    A. Czy oceniłeś możliwe ograniczenia wynikające z wykorzystanych danych?

    B. Czy uwzględniłeś różnorodność i reprezentatywność użytkowników w zbiorach danych?

    C. Czy zaplanowałeś zbadanie i wykorzystanie dostępnych narzędzi technicznych, aby poprawić zrozumienie danych, modelu i działania systemu?

    D. Czy zaplanowałeś wdrożenie procesów testowania i monitorowania potencjalnych uprzedzeń podczas opracowywania i wdrażania systemu?

25. W odniesieniu do powyższego, czy zaplanowałeś mechanizm, który pozwala użytkownikom zgłaszać problemy związane z uprzedzeniem, dyskryminacją czy niewłaściwym działaniem systemu AI?

    A. Czy ustaliłeś jasne kroki i sposoby komunikowania się, tzn. jak i komu można zgłaszać takie kwestie?

    B. Czy oprócz użytkowników (końcowych) uwzględniono innych, na których potencjalnie pośrednio wpływa system sztucznej inteligencji?

Dostępność i uniwersalny design

26. Czy zapewniłeś, że system AI uwzględni szeroki zakres indywidualnych preferencji i umiejętności?

    A. Czy oceniłeś, czy system sztucznej inteligencji może być używany przez osoby ze specjalnymi potrzebami lub niepełnosprawnościami lub osoby zagrożone wykluczeniem?

Udział zainteresowanych stron

27. Czy rozważałeś mechanizm obejmujący udział różnych interesariuszy w rozwoju i wykorzystaniu systemu?

VI. ODPOWIEDZIALNOŚĆ

Audyt

28. Czy zaplanowałeś mechanizmy ułatwiające audyt systemu, takie jak zapewnienie identyfikowalności i rejestrowania procesów i wyników systemu AI?

29. Czy w aplikacjach wpływających na prawa podstawowe (w tym w aplikacjach o znaczeniu krytycznym dla bezpieczeństwa) zaplanowałeś warunek, że system AI może podlegać niezależnemu audytowi?