

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?
 - C. W przypadku bota do czatu lub innego systemu komunikacji, czy użytkownicy końcowi będący ludźmi są świadomi, że współpracują z robotem?

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 na systemie sztucznej inteligencji w procesach pracy?

Nadzór człowieka

4. Czy rozważyłeś odpowiedni poziom kontroli człowieka dla systemu SI ?
 - A. Czy możesz opisać poziom ludzkiej kontroli lub zaangażowania?
 - B. Kto jest „człowiekiem nadzorującym” i jakie są momenty lub narzędzia interwencji człowieka?
 - C. Czy zaplanowałeś mechanizmy i środki w celu zapewnienia kontroli lub nadzoru przez ludzi?
 - D. Czy zaplanowałeś jakieś środki, aby umożliwić audyt i rozwiązać problemy związane z zarządzaniem autonomią SI?
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?
 - B. Czy zapewniono procedurę bezpiecznego przerwania operacji w razie potrzeby? Czy ta procedura całkowicie lub częściowo przerywa proces czy też przekazuje kontrolę człowiekowi?

II. STABILNOŚĆ TECHNICZNA I BEZPIECZEŃSTWO

Odporność na atak i bezpieczeństwo

6. Czy oceniłeś potencjalne formy ataków, na które system SI 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 SI na potencjalne ataki?

Plan awaryjny i ogólne bezpieczeństwo

7. Czy przewidziałeś wystarczający plan awaryjny na wypadek ataków lub innych nieoczekiwanych sytuacji (na przykład techniczne procedury przełączania lub wcześniejsza prośba o kontakt z człowiekiem przed procedowaniem postępowania)?
8. Czy oceniłeś, czy istnieje prawdopodobieństwo, że system SI może działać ze szkodą dla użytkowników lub osób trzecich? Czy oceniłeś prawdopodobieństwo, potencjalne szkody, wpływ na odbiorców i wagę tych szkód?
 - A. Czy rozważyłeś zasady odpowiedzialności i ochrony konsumenta i wziąłeś je pod uwagę?
 - B. Czy wziąłeś pod uwagę potencjalny wpływ lub zagrożenie bezpieczeństwa dla środowiska lub zwierząt?
 - C. Czy w analizie ryzyka uwzględniono, czy problemy z bezpieczeństwem lub siecią, takie jak zagrożenia cyberbezpieczeństwa, mogą stwarzać zagrożenie dla bezpieczeństwa lub szkody spowodowane niezamierzonym działaniem systemu SI?
9. Czy przewidziałeś, jakie szkody nastąpiłyby, gdyby system sztucznej inteligencji dokonał niedokładnych prognoz?
10. Czy zaplanowałeś sposoby pomiaru, czy Twój system podaje niedopuszczalną ilość niedokładnych prognoz?
11. Czy zaplanowałeś działania w celu zwiększenia dokładności systemu?

Niezawodność i odtwarzalność

12. 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 proces w celu testowania i weryfikacji niezawodności systemów SI?
 - D. Czy zaplanowałeś mechanizmy komunikacji, aby zapewnić (końcowym) użytkownikom niezawodność systemu?

III. PRYWATNOŚĆ I ZARZĄDZANIE DANYMI

Poszanowanie prywatności i ochrony danych

13. Czy zaplanowałeś mechanizm umożliwiający innym zgłaszanie problemów związanych z prywatnością lub ochroną danych w procesach zbierania danych (na potrzeby szkolenia i eksploatacji) i przetwarzania danych w systemie SI?
14. Czy oceniłeś rodzaj i zakres danych w swoich zbiorach danych (np. Czy zawierają one dane osobowe)?

15. Czy rozważyłeś sposoby rozwijania systemu SI lub trenowania modelu bez lub z minimalnym wykorzystaniem potencjalnie wrażliwych danych czy danych osobowych?
16. Czy zaplanowałeś wbudowanie mechanizmów zarządzania danymi osobowymi?
17. Czy zaplanowano kroki w celu zwiększenia prywatności, takie jak szyfrowanie, anonimizacja i agregacja?
18. Jeśli istnieje inspektor ochrony danych (IOD), czy zaplanowano zaangażowanie takiej osoby na wczesnym etapie procesu powstawania systemu?

Jakość i integralność danych

19. Czy system będzie dostosowany do odpowiednich norm (na przykład ISO, IEEE) lub powszechnie przyjętych protokołów zarządzania danymi i nadzorem nad nimi?
20. Czy ustanowiłeś mechanizmy nadzoru dotyczące gromadzenia, przechowywania, przetwarzania i wykorzystywania danych?
21. Czy oceniłeś zakres, w jakim kontrolujesz jakość wykorzystanych zewnętrznych źródeł danych?
22. Czy zaplanowano wdrożenie procesów zapewniających jakość i integralność danych? W jaki sposób weryfikujesz, czy Twoje zbiory danych nie zostały naruszone lub zhakowane?

Dostęp do danych

- A. Czy oceniłeś, kto może uzyskać dostęp do danych użytkowników i w jakich okolicznościach?
- B. Czy upewniłeś się, że te osoby są wykwalifikowane, że wymagany jest ich dostęp do danych oraz że mają niezbędne kompetencje do zrozumienia szczegółów polityki ochrony danych?
- C. Czy zaplanowałeś mechanizm nadzoru umożliwiający rejestrowanie, kiedy, gdzie, jak, przez kogo i w jakim celu uzyskano dostęp do danych?

IV. TRANSPARENTNOŚĆ

Wytłumaczalność

23. Czy oceniłeś:
 - A. W jakim stopniu można zrozumieć decyzje, a co za tym idzie wynik systemu SI?
 - B. W jakim stopniu decyzja systemu wpływa na procesy decyzyjne podmiotu wdrażającego?
 - C. Dlaczego ten konkretny system musi zostać wdrożony w tym konkretnym obszarze?
24. Czy od początku projektowałeś system SI z myślą o możliwości interpretacji?
 - A. Czy zaplanowałeś przetestowanie najprostszego i najbardziej interpretowalnego modelu możliwego do zastosowania w programie?
 - B. Czy zaplanowałeś ocenę, czy możesz przeanalizować swoje dane treningowe i testowe? Czy możesz to zmienić i zaktualizować z biegiem czasu?
 - C. Czy zaplanowałeś ocenę, czy po przeszkoleniu i opracowaniu modelu możesz zbadać możliwość interpretacji i czy masz dostęp do wewnętrznego przepływu pracy modelu?

Komunikacja

25. Czy zaplanowałeś poinformowanie użytkowników końcowych, że wchodzi w interakcję z systemem SI, a nie z innym człowiekiem?
26. Czy zaplanowano mechanizmy informowania użytkowników (końcowych) o zasadach działania systemu SI?

- A. Czy zaplanowano przekazanie tego w jasny i zrozumiały sposób docelowej grupie użytkowników?
 - B. Czy zaplanowano procesy, które uwzględniają opinie użytkowników i wykorzystanie ich do dostosowania systemu?
27. Czy zaplanowałeś wyjaśnienie celu systemu SI i kto lub co może skorzystać z produktu / usługi?
- A. Czy określiłeś scenariusze użytkownika produktu i zaplanowałeś jasne przekazanie tych scenariuszy, aby upewnić się, że są one zrozumiałe i odpowiednie dla zamierzonych użytkowników?
28. Czy przewidziałeś jasne określenie i zakomunikowanie cech, ograniczeń i potencjalnych wady systemu SI?
- A. w przypadku rozwoju systemu: dla kogokolwiek kto będzie wdrażał go jako produktu lub usługę?
 - B. w przypadku wdrażania systemu: dla użytkownika (końcowego) czy konsumenta?

V. RÓŻNORODNOŚĆ, BRAK DYSKRYMIANCIJA I UCZCIWOŚĆ

Unikanie nieuczciwych uprzedzeń

29. Czy ustaliłeś strategię lub zestaw procedur, aby uniknąć tworzenia lub wzmocnienia nieuczciwych uprzedzeń w systemie SI, 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 do testowania i monitorowania potencjalnych uprzedzeń podczas opracowywania i wdrażania użytkownika systemu?
30. W odniesieniu do powyższego, czy zaplanowałeś mechanizm, który pozwala innym zgłaszać problemy związane z uprzedzeniem, dyskryminacją czy niewłaściwym działaniem systemu SI?
- A. Czy ustaliłeś jasne kroki i sposoby komunikowania się - jak i komu można podnosić takie kwestie?
 - B. Czy oprócz użytkowników (końcowych) uwzględniono innych, na których potencjalnie pośrednio wpływa system sztucznej inteligencji?

Dostępność i uniwersalny design

31. Czy zapewniłeś, że system SI 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? Jak zostało to zaprojektowane w systemie i jak będzie weryfikowane?
 - B. Czy zapewniłeś, że informacje o systemie SI są dostępne również dla użytkowników technologii wspomagających?
 - C. Czy zaangażowałeś tę społeczność lub konsultowałeś się z tą społecznością lub zaplanowałeś takie działania podczas fazy rozwoju systemu SI?
32. Czy wziąłeś pod uwagę wpływ swojego systemu SI na potencjalnych odbiorców?
- A. Czy oceniłeś, czy zespół zaangażowany w tworzenie systemu SI jest reprezentatywny w świetle docelowych użytkowników systemu
 - B. Czy oceniłeś, czy mogą istnieć osoby lub grupy, na które system może mieć nieproporcjonalnie gorszy wpływ?
 - C. Czy otrzymałeś opinie od innych zespołów lub grup, reprezentujących różne środowiska i doświadczenia?

Udział zainteresowanych stron

33. Czy rozważyłaś mechanizm obejmujący udział różnych interesariuszy w rozwoju i wykorzystaniu systemu?

VI. WPŁYW SPOŁECZNY I ŚRODOWISKOWY

Zrównoważona i przyjazna dla środowiska sztuczna inteligencja

34. Czy ustanowiłaś mechanizmy pomiaru wpływu rozwoju, wdrożenia i wykorzystania systemu SI na środowisko (na przykład rodzaj energii używanej przez centra danych)?
35. Czy zapewniłaś środki zaradcze w celu zmniejszenia wpływu twojego systemu SI na środowisko?

Wpływ społeczny

36. W przypadku, gdy system SI oddziałuje bezpośrednio z ludźmi:
- A. Czy oceniłaś, czy system sztucznej inteligencji zachęci ludzi do rozwijania przywiązania i empatii wobec systemu?
 - B. Czy zapewniłaś, że system sztucznej inteligencji wyraźnie zasygnalizuje, że jego interakcja społeczna jest symulowana i że nie ma on zdolności „rozumienia” i „czucia”?
37. Czy upewniłaś się, że społeczne skutki systemu SI są dobrze rozumiane? Na przykład, czy oceniłaś czy istnieje ryzyko utraty pracy lub obniżenia kwalifikacji siły roboczej? Jakie kroki podjęto, aby zminimalizować takie ryzyko?

Społeczeństwo i demokracja

38. Czy oceniłaś szerszy społeczny wpływ korzystania z systemu SI poza jego wpływem na indywidualnego (końcowym) użytkownika, np. potencjalnie pośrednio dotknięci interesariusze?

VII. ODPOWIEDZIALNOŚĆ

Audyt

39. Czy zaplanowałaś mechanizmy ułatwiające audyt systemu, takie jak zapewnienie identyfikowalności i rejestrowania procesów i wyników systemu SI?
40. Czy w aplikacjach wpływających na prawa podstawowe (w tym w aplikacjach o znaczeniu krytycznym dla bezpieczeństwa) zaplanowałaś warunek, że system SI może podlegać niezależnemu audytowi?

Minimalizowanie i zgłaszanie negatywnego wpływu

41. Czy przewidziałaś jakiegokolwiek zewnętrzne wytyczne lub wprowadziłaś procesy audytu w celu nadzorowania etyki i odpowiedzialności, oprócz inicjatyw wewnętrznych?
42. Czy ustanowiłaś lub zaplanowałaś procesy dla stron trzecich (np. dostawców, konsumentów, dystrybutorów / sprzedawców) lub pracowników do zgłaszania potencjalnych luk, zagrożeń lub uprzedzeń w systemie SI?