Niniejszy szablon jest częścią serii „Mapa drogowa odpowiedzialnego AI dla NGOsów” stworzonej przez Ayşegül Güzel przy okazji projektu AI for Social Change realizowanego przez TechSoup w ramach programu Digital Activism Program przy wsparciu Google.org.

Pełna seria jest dostępna na platformie Hive Mind.
Mapa drogowa jest przewodnikiem dla organizacji społeczeństwa obywatelskiego po każdym etapie odpowiedzialnego wdrażania sztucznej inteligencji—począwszy od opisania obecnego krajobrazu AI, zdefiniowania zasad i struktur zarządzania oraz udokumentowania polityki, po wdrażanie praktyk AI poprzez ocenę ryzyka, planowanie łagodzenia skutków i bieżący nadzór.

Jak korzystać z tego szablonu
Pobierz kopię i pracuj nad nią zgodnie z opisem podanym w towarzyszących artykułach dostępnych na platformie Hive Mind.
Po wykorzystaniu tego szablonu **podziel się z nami swoją opinią TUTAJ** - pozwoli nam to ulepszać udostępniane zasoby.

SZABLON POLITYKI AI

Praktyczny przewodnik po tworzeniu polityki AI

Wersja 1.0

Ostatnia aktualizacja: [Data]

Jak korzystać z tego szablonu do opracowania polityki AI

Niniejszy dokument jest narzędziem, które ma pomóc ci w stworzeniu praktycznej i skutecznej polityki AI dla twojej organizacji. Niezależnie od tego, czy prowadzisz wspólne warsztaty, czy samodzielnie kierujesz zespołem, ten szablon daje ci niezbędną strukturę i wskazówki.

Jak rozumieć poszczególne elementy

- **Pytania przewodnie** *(pisane kursywą)*: Stanowią one sedno tego szablonu. Nie są to pytania sprawdzające, lecz wskazówki mające na celu pobudzenie krytycznego myślenia i dyskusji na temat tego, co jest najważniejsze dla twojej organizacji.

- **Przykładowa treść** (w kolorze niebieskim): Te treści służą jako tekst wypełniający lub punkt wyjścia. Możesz je dowolnie dostosować, przeredagować lub całkowicie usunąć. Mają one na celu przybliżenie, jak mogłaby wyglądać deklaracja zasad.

Zalecany przez nas trzystopniowy proces

Wykonaj poniższe kroki, aby przejść od pustego szablonu do gotowej polityki.

Krok 1: Przygotowanie i dyskusja

Najpierw zbierz odpowiednią grupę osób (może to być twój główny zespół, kierownicy projektów, a na początek nawet tylko ty sam). Zarezerwuj czas na omówienie pytań przewodnich zawartych w tym szablonie.

Cel: Najpierw skup się na dyskusji, a nie pisaniu idealnych zdań. Zapisz swoje wstępne przemyślenia i najważniejsze decyzje dotyczące każdej sekcji. Nie czuj się zobowiązany do udzielenia odpowiedzi na każde pytanie – skup się na tych, które są najbardziej istotne dla twojej organizacji w tym momencie.

Wskazówki ułatwiające – Krok 1:

Kto powinien być obecny na spotkaniu? Staraj się uwzględnić różnorodne role, nie tylko opinie kierownictwa. Osoby, które na co dzień korzystają z narzędzi AI często dostrzegają zagrożenia i możliwości, których kierownictwo nie dostrzega. Jeżeli w spotkaniu może wziąć udział wolontariusz lub pracownik pierwszej linii, zaproś go.

Zacznij od ujawnienia napięć, a nie od pisania tekstu. Przed otwarciem tego szablonu warto rozważyć rozpoczęcie sesji od prostych pytań: „Co najbardziej fascynuje cię w wykorzystaniu sztucznej inteligencji w naszej pracy?” „Co cię martwi?” Zacznij od nadziei i obaw – to pozwoli ujawnić prawdziwe problemy i stworzyć atmosferę bezpieczeństwa psychicznego sprzyjającą szczerej rozmowie.

Mów językiem uczestników spotkania. Gdy czyjeś zdanie trafnie oddaje jakąś zasadę lub granicę, zapisz je dokładnie tak, jak zostało wypowiedziane. Polityka sformułowana przez zespół jego własnymi słowami staje się dokumentem, który jest własnością tego zespołu.

Zapewnienie większego uporządkowania procesu: W niektórych organizacjach sprawdza się podejście etapowe, w którym na początku organizowana jest sesja diagnostyczna mająca na celu zidentyfikowanie obecnych zastosowań AI oraz otoczenia regulacyjnego, następnie warsztaty poświęcone zasadom, których celem jest określenie kwestii niepodlegających

negocjacjom, po czym wykorzystuje się ten szablon jako narzędzie do wspólnego tworzenia. Nie jest to konieczne, ale stanowi opcję dla organizacji, które przed przystąpieniem do tworzenia projektu polityki chcą zdobyć solidniejsze podstawy.

Etap 2: Tworzenie projektu polityki

Gdy już zgromadzisz notatki z dyskusji, przystąp do sporządzenia projektu polityki. Przejdź przez szablon sekcja po sekcji i zastąp tekst przykładowy zaznaczony na niebiesko własnymi odpowiedziami oraz deklaracjami programowymi.

Cel: Przełóż swoje notatki z dyskusji na jasny i prosty język, który zrozumie każdy w twojej organizacji. Twoim celem jest przygotowanie pełnego projektu „Wersji 1".

Etap 3: Przegląd, finalizacja i udostępnienie

Żadnej polityki nie należy tworzyć w próżni. Gdy masz już gotowy projekt, udostępnij go szerszemu gronu zainteresowanych osób — całemu zespołowi, kluczowym wolontariuszom lub doradcom — w celu uzyskania ich opinii.

Cel: Zbierz opinie, aby upewnić się, że wytyczne są praktyczne i zrozumiałe dla osób, które będą z nich korzystać na co dzień.

Działanie: Udostępniając projekt, podaj jasny termin zgłoszenia uwag (np. „Proszę o przesłanie uwag do dnia [data]"). Po uwzględnieniu zgłoszonych uwag sporządź ostateczną wersję polityki i oficjalnie przekaż ją wszystkim pracownikom organizacji.

Uwaga dotycząca zakresu: Zacznij od generatywnej AI

Większość organizacji zaczyna od stworzenia polityki dotyczącej wykorzystania generatywnej AI i jest to właściwy punkt wyjścia. Nie musisz od samego początku zajmować się wszystkimi aspektami sztucznej inteligencji. W miarę pojawianie się wyników kolejnych ocen ryzyka (zob. artykuł nr 4 z tej serii) przeprowadzanych przez twoją organizację, polityka będzie w naturalny sposób rozszerzać się, obejmując nowe narzędzia i przypadki użycia. Pozwól sobie zostawić niektóre fragmenty pierwszej wersji w mniej dopracowanej formie. Celem jest stworzenie użytecznej wersji wstępnej, a nie kompletnej wersji ostatecznej.

Pamiętaj, że polityka AI to żywy dokument. Będzie ewoluować wraz z postępem technologicznym i zmieniającymi się potrzebami twojej organizacji. Najważniejsze to rozpocząć rozmowę i stworzyć wersję podstawową.

1. Wprowadzenie i cel

W tej sekcji nakreślone zostają ramy całej polityki w oparciu o twoje podejście do sztucznej inteligencji na podstawowej misji i celów strategicznych organizacji.

- *W jaki konkretny sposób generatywna AI może wspierać realizację celów strategicznych naszej organizacji oraz jej bieżącą działalność?*
- *W jaki sposób możemy zadbać o to, by wykorzystanie AI wspierało naszą podstawową misję i wartości?*
- *Jaki jest główny powód wprowadzenia tej polityki właśnie teraz i jakie kluczowe cele mamy nadzieję dzięki niej osiągnąć?*
- *Jak będzie wyglądał sukces, gdy ta polityka zostanie już skutecznie wdrożona w całym naszym zespole?*

2. Zakres

Jasna polityka wymaga jasno określonych granic. W tym miejscu należy dokładnie określić, kogo ta polityka dotyczy, zarówno w organizacji, jak i poza nią.

- *Kto w naszej organizacji powinien przestrzegać tych wytycznych? Określmy dokładnie role, od wolontariuszy po główny zespół.*
- *Czy mamy jakichś partnerów zewnętrznych, wykonawców lub współpracowników, od których również oczekujemy przestrzegania niniejszej polityki podczas współpracy z nami?*
- *W jaki sposób wykorzystywanie sztucznej inteligencji przez naszą organizację wpływa na ludzi i społeczności, którym służymy? Czy przy kształtowaniu tej polityki uwzględniliśmy ich punkt widzenia, choćby w nieformalny sposób?*

3 . Powiązane polityki i dalsze odniesienia

Niniejsza polityka nie istnieje w próżni. Ta sekcja zapewni jej zgodność z innymi istotnymi wytycznymi organizacyjnymi oraz z otoczeniem regulacyjnym, w którym działa twoja organizacja.

- *W jaki sposób niniejsza polityka AI wiąże się z innymi ważnymi dokumentami, które już posiadamy, np. Kodeksem postępowania czy Polityką prywatności?*
- *Czy są jeszcze jakieś inne kluczowe polityki, z którymi warto się zapoznać, aby uzyskać pełny obraz naszych zasad?*
- *Jakie przepisy ochrony danych obowiązują w jurysdykcjach, w których prowadzimy działalność? (np. RODO w Europie, POPIA w RPA, krajowe przepisy ochrony danych)*
- *Czy istnieją jakieś przepisy branżowe lub wymagania podmiotów finansujących, które mają wpływ na to, w jaki sposób możemy wykorzystywać sztuczną inteligencję? (np. wymogi dotyczące zautomatyzowanego podejmowania decyzji, lokalizacji danych lub transparentności w zakresie sztucznej inteligencji)*
- *Jeżeli prowadzimy działalność w wielu krajach, jak radzić sobie w sytuacjach, gdy przepisy różnią się w poszczególnych jurysdykcjach?*

4. Definicje

Aby mieć pewność, że wszyscy mówią tym samym językiem, w tej sekcji zdefiniowane zostają stosowane w niniejszej polityce terminy związane ze sztuczną inteligencją.

- *Jakie kluczowe pojęcia z zakresu sztucznej inteligencji użyte w niniejszym dokumencie należy zdefiniować, aby wszyscy rozumieli je takie samo?*

5. Zarządzanie AI

Ta sekcja określa przewodnie wartości w zakresie korzystania ze sztucznej inteligencji oraz wyjaśnia, kto jest odpowiedzialny za ich przestrzeganie.

Nasze zasady AI

W tym miejscu określisz podstawowe wartości, które będą stanowić fundament każdej decyzji dotyczącej wykorzystania sztucznej inteligencji.

- *Jakie są podstawowe zasady etyczne, którymi będziemy się kierować w wykorzystywaniu sztucznej inteligencji? Powinny one być inspirowane zarówno wartościami naszej organizacji, jak uznanymi najlepszymi praktykami.*
- *W odniesieniu do każdej wybranej przez nas zasady określmy w prosty i jasny sposób, co ona oznacza w naszej codziennej pracy.*
- *W jaki sposób nasze zasady odzwierciedlają potrzeby i obawy społeczności, którym służymy, a nie tylko nasze wewnętrzne działania? Czego ludzie, z którymi*

współpracujemy chcieliby się dowiedzieć na temat tego, jak wykorzystujemy sztuczną inteligencję?

Kto jest osobą odpowiedzialną? (Nasze zasady zarządzania AI)

W tej części zostaje sprecyzowane, kto jest właścicielem niniejszej polityki i do kogo należy się zwracać w razie wątpliwości. Zapewnia to jasny podział odpowiedzialności oraz wsparcie dla całego zespołu. *W wielu organizacjach odpowiedź na pytanie „Kto jest odpowiedzialny?" nie musi oznaczać powołania formalnej komisji. Można to zrobić prosto: „Główny zespół odpowiada za analizę nowych narzędzi oraz aktualizację niniejszych wytycznych, a [Imię i nazwisko] jest główną osobą kontaktową w razie jakichkolwiek pytań". Celem jest przejrzystość, a nie zawiłość.*

- *Kto jest główną osobą odpowiedzialną za tę politykę i kto odpowiada za jej aktualizację?*
- *Do kogo lub do jakiego zespołu należy się zwrócić w razie pytań dotyczących korzystania z narzędzi AI lub interpretacji niniejszych wytycznych?*
- *Jak wygląda proces weryfikacji i zatwierdzania nowych narzędzi AI do użytku zespołowego i kto jest za ten proces odpowiedzialny?*

Sekcja 6 Zastosowanie naszych zasad w praktyce: Nasze systemy i wsparcie

Ta sekcja przedstawia praktyczne rozwiązania będące wsparciem twojego zespołu i zapewniające odpowiedzialne korzystanie ze sztucznej inteligencji.

Rejestr AI

Aby zapewnić przejrzystość i bezpieczeństwo, będziesz prowadzić scentralizowaną listę wszystkich zatwierdzonych narzędzi AI.

- *W jaki sposób będziemy śledzić, z jakich narzędzi AI korzysta nasz zespół?*
- *Jakie kluczowe informacje powinniśmy zamieścić w opisie każdego narzędzia znajdującego się na naszej liście zatwierdzonych narzędzi? (np. jego przeznaczenie, główne zagrożenia oraz do kogo zwrócić się o pomoc).*
- *Kto będzie odpowiedzialny za aktualizację tej listy i jak często powinniśmy dokonywać jej przeglądu, aby zapewnić jej aktualność?*
- *Jaki jest kolor sygnalizacji świetlnej (🔴 czerwony / 🟡 żółty / 🟢 zielony) dla każdej kombinacji narzędzia i przypadku użycia przewidzianej w naszym rejestrze? Pamiętaj: to*

samo narzędzie może mieć różny status w zależności od przypadku użycia i wrażliwości danych.

Twój Rejestr AI (rozpoczęty w artykule nr 1) działa najlepiej, gdy zawiera kolor sygnalizacji świetlnej zostaje przypisany do poszczególnych przypadków użycia, a nie do poszczególnych narzędzi. Na przykład ChatGPT może otrzymać światło 🟢 zielone na potrzeby przygotowania publicznego wpisu na blogu, ale 🔴 czerwone na potrzeby sporządzania streszczeń poufnych notatek. Podczas opracowywania sekcji dotyczącej sygnalizacji świetlnej w niniejszej polityce należy ponownie przeanalizować treść Rejestru – te dwa elementy powinny się wzajemnie uzupełniać.

Poziomy wrażliwości danych (dostosuj do swojego kontekstu):

Podczas tworzenia Rejestru i przypisywania kolorów w systemie sygnalizacji świetlnej warto sklasyfikować dane związane z każdym przypadkiem użycia:

- **Publiczne:** Brak obaw związanych z ujawnieniem informacji (np. opublikowane raporty, treści na stronie internetowej)
- **Wewnętrzni:** Wyłącznie do użytku organizacyjnego (np. notatki służbowe, plany projektów)
- **Poufne:** Ograniczony dostęp, poufne informacje organizacyjne (np. listy darczyńców, dokumentacja finansowa)
- **Bardzo wrażliwy:** Dane osobowe, informacje o beneficjentach lub treści objęte ochroną prawną (np. akta spraw, informacje medyczne, zeznania osób ubiegających się o azyl)

Jakie poziomy wrażliwości danych stosuje twoja organizacja i w jaki sposób przekładają się one na decyzje podejmowane przez was w ramach systemu sygnalizacji świetlnej?

Znajomość AI

Naszym celem jest wyposażenie naszego zespołu w wiedzę niezbędną do pewnego i bezpiecznego korzystania ze sztucznej inteligencji.

Zapoznanie z generatywną AI

- *Jakiego poziomu podstawowej wiedzy oczekujemy od wszystkich osób korzystających ze sztucznej inteligencji w naszej organizacji? O jakich kluczowych kompetencjach, zagrożeniach i ograniczeniach powinny one wiedzieć?*

- *W jaki sposób pomożemy naszemu zespołowi zdobyć tę wiedzę? Jakie szkolenia, przewodniki lub materiały udostępnimy, aby promować bezpieczne i odpowiedzialne korzystanie ze sztucznej inteligencji?*

Zrozumienie, co potrafią, a czego nie potrafią wstępnie wytrenowane modele bazowe

- *Jak możemy zapewnić, by członkowie zespołu korzystający z bardziej zaawansowanych narzędzi AI dobrze rozumieli konkretne mocne i słabe strony tej technologii?*
- *Jak będziemy na bieżąco śledzić zmiany i informować nasz zespół o ewolucji tych zaawansowanych modeli AI?*

Ocena narzędzi generatywnej AI pochodzących od innych podmiotów

Wraz z pojawianiem się nowych, interesujących narzędzi AI potrzebujemy spójnego sposobu ich oceny, zanim zostaną one wdrożone przez zespół. Dzięki temu mamy pewność, że są one bezpieczne, skuteczne i zgodne z naszymi wartościami.

- *Kiedy znajdujemy nowe narzędzie AI, w jaki sposób oceniamy, czy jego użycie jest bezpieczne i odpowiedzialne?*
- *Na jakie kluczowe „sygnały ostrzegawcze" lub zagrożenia powinniśmy zwracać uwagę? (np. kwestie dotyczące ochrony danych, bezpieczeństwa lub potencjalnej stronniczości).*
- *Jakie niezbędne środki bezpieczeństwa lub kontrole należy przeprowadzić przed dodaniem nowego narzędzia do naszego zatwierdzonego Rejestru AI?*
- *W szczególności, co powinniśmy wiedzieć na temat sposobu przetwarzania danych przez dane narzędzie? Czy dana firma wykorzystuje nasze dane do trenowania swoich modeli?*
- *Pomijając kwestie bezpieczeństwa, jak możemy ocenić, czy nowe narzędzie jest rzeczywiście przydatne i dobrze odpowiada potrzebom naszej organizacji?*

Zarządzanie ryzykiem i reagowanie na problemy

Kluczowym elementem odpowiedzialnego korzystania ze sztucznej inteligencji jest zrozumienie związanych z nią zagrożeń oraz opracowanie jasnego planu na wypadek, gdyby sprawy nie potoczyły się zgodnie z oczekiwaniami. W tej sekcji omówimy, jak zamierzamy sobie z tym poradzić jako zespół.

Sugestia: Dla wielu organizacji prostym i skutecznym rozwiązaniem jest zarządzanie ryzykiem w odniesieniu do konkretnego narzędzia zamiast tworzenia skomplikowanego systemu zarządzania ryzykiem. Można to zrobić dodając do naszego Rejestru AI kolumnę „Główne

zagrożenia i sposoby ich ograniczania". Sprawi to, że proces zarządzania ryzykiem będzie działaniem praktycznym i ciągłym.

- *Jakie są największe zagrożenia dla naszej organizacji związane z wykorzystaniem generatywnej AI? (np. wycieki danych osobowych, nieścisłości w naszych raportach, niezamierzona stronniczość w naszych treściach).*
- *W przypadku każdego zatwierdzanego przez nas narzędzia, jakie proste i praktyczne działania możemy podjąć, aby zminimalizować związane z nim konkretne zagrożenia?*
- *Jak wygląda nasza procedura w przypadku napotkania nowego lub nieprzewidzianego problemu związanego z narzędziem AI?*

Zgłaszanie incydentów

- *Jeżeli zauważysz, że narzędzie AI generuje stronnicze, błędne lub szkodliwe wyniki, albo jeżeli masz obawy dotyczące bezpieczeństwa, komu należy to zgłosić?*
- *Jak najlepiej zgłosić problem (e-mail? prywatna wiadomość?), żeby można było szybko się nim zająć?*
- *Jakie kluczowe informacje należy zawrzeć w zgłoszeniu, aby pomóc nam zrozumieć i rozwiązać dany problem?*

Zbieranie opinii na temat skutków użycia generatywnej AI

- *Oprócz zgłaszania problemów, jak wypracować prosty sposób, dzięki któremu członkowie zespołu będą mogli dzielić się opiniami, pomysłami lub wątpliwościami dotyczącymi narzędzi AI, z których korzystamy?*
- *W jaki sposób wykorzystamy te opinie, aby podejmować trafniejsze decyzje dotyczące tego, z jakich narzędzi korzystamy i w jaki sposób je wykorzystujemy?*
- *Jak wygląda nasz proces analizy poważnych zdarzeń, mający na celu wyciągnięcie wniosków i zapobieganie ich powtórzeniu się?*

7 . Sygnalizacja drogowa: Wytyczne dotyczące dopuszczalnego użytkowania

Ta sekcja to twój podstawowy „zbiór przepisów". Odpowiada na pytanie, które zadaje sobie każdy członek zespołu: „Czy mogę tego użyć?" Model sygnalizacji świetlnej zastępuje paraliżującą niepewność prostą i praktyczną procedurą kontrolną.

- *Jak najlepiej wykorzystać sztuczną inteligencję do usprawnienia naszej pracy i realizacji naszej misji?*

- *Jak ustalimy, które narzędzia AI będą dopuszczone do użytku przez nasz zespół? (Jest to powiązane z naszym Rejestrem AI).*
- *Jakie są jasno określone granice stosowania sztucznej inteligencji? Gdzie powinniśmy ustalić, że korzystanie z AI jest w naszej organizacji zabronione?*

🔴 Czerwone światło — Stop (czego NIE WOLNO robić)

Czerwone światła to granice niepodlegające negocjacjom, bezwzględne linie wynikające bezpośrednio z twoich zasad. Oto zadania, do realizacji których twoja organizacja nigdy użyje sztucznej inteligencji, niezależnie od narzędzia czy kontekstu. Przy opracowywaniu swojej polityki zacznij od tego miejsca: jeśli możesz określić tylko jedną rzecz, to niech będą to „czerwone światła".

W celu ochrony organizacji, jej wartości oraz społeczności, którym służy, następujące sposoby wykorzystania sztucznej inteligencji są surowo zabronione.

- **Ochrona informacji poufnych:** Nigdy nie wprowadzaj poufnych, zastrzeżonych ani osobistych danych swojej organizacji do żadnego publicznego lub niezatwierdzonego systemu AI. Dotyczy to takich narzędzi, jak bezpłatna wersja ChatGPT. Traktuj wszelkie informacje wprowadzane do zewnętrznego narzędzia AI tak, jakbyś je publikował w Internecie. Zasada ta dotyczy bez wyjątku wszystkich danych wrażliwych, w tym danych wolontariuszy i partnerów, dokumentacji finansowej oraz wewnętrznych dokumentów strategicznych.
- **Przestrzeganie zasad bezpieczeństwa i kontroli dostępu:** Nie wykorzystuj sztucznej inteligencji do prób obejścia zabezpieczeń ani uzyskania dostępu do danych, do których nie masz uprawnień. Wykorzystywanie sztucznej inteligencji do jakichkolwiek nielegalnych, oszukańczych lub nieuprawnionych celów, takich jak hakowanie lub pozyskiwanie informacji zastrzeżonych jest surowo zabronione.
- **Wymóg nadzoru ze strony człowieka w przypadku ważnych decyzji:** Nie wykorzystuj sztucznej inteligencji do podejmowania ostatecznych decyzji, które mają skutki prawne, finansowe lub inne istotne konsekwencje dla życia ludzi, bez ich bezpośredniej weryfikacji i zatwierdzenia przez człowieka. Na przykład nie można wykorzystywać sztucznej inteligencji do pełnej automatyzacji decyzji dotyczących zatrudnienia, zatwierdzania kredytów czy wyboru beneficjentów.
- **Dbałość o etyczną i uczciwą komunikację:** Nie wykorzystuj sztucznej inteligencji do tworzenia treści naruszających Kodeks postępowania, takich jak mowa nienawiści lub nękanie. Ponadto, nie wykorzystuj sztucznej inteligencji do oszukiwania lub wprowadzania w błąd innych ludzi, na przykład poprzez tworzenie „deepfake'ów",

fałszywych wiadomości lub oszukańczych komunikatów. Wszelka komunikacja wspomagana przez sztuczną inteligencję musi być zgodna z prawdą.

- **Poszanowanie własności intelektualnej:** Nie wykorzystuj sztucznej inteligencji do tworzenia ani udostępniania treści naruszających prawa autorskie, znaki towarowe lub licencje. Dotyczy to zlecania sztucznej inteligencji wykonania kopii konkretnego, chronionego źródła lub wykorzystywania kodu wygenerowanego przez sztuczną inteligencję, który nie jest objęty odpowiednią licencją. W razie wątpliwości skonsultuj się z odpowiednim zespołem lub unikaj korzystania z wyników.
- **Przestrzeganie wszelkich wymogów prawnych i regulacyjnych:** Nie wolno wykorzystywać sztucznej inteligencji w sposób, który naruszałby obowiązujące przepisy prawa lub regulacje. Jeśli twoja praca podlega określonym przepisom (np. przepisom ochrony danych, takim jak RODO), nie możesz wykorzystywać sztucznej inteligencji w sposób, który je narusza. Nie korzystaj z niezatwierdzonych rozwiązań AI w celu obejścia procedur zgodności.
- **Stwierdzanie i zgłaszanie błędów AI:** Niedopuszczalne jest świadome korzystanie z systemu AI zawierającego poważne błędy (takie jak znacząca stronniczość lub niedokładność) bez podjęcia działań zmierzających do ich usunięcia. Nie wolno ukrywać faktu korzystania ze sztucznej inteligencji, gdy wymagana jest transparentność. Jeżeli zauważysz, że dane narzędzie AI działa nieprawidłowo, nie ignoruj tego. Przestań wykorzystywać je do realizacji danego zadania i natychmiast zgłoś problem.

Każde naruszenie tych zakazów stanowi poważną sprawę i może skutkować podjęciem działań dyscyplinarnych. Zawsze zalecamy zachowanie szczególnej ostrożności. Jeżeli potencjalne zastosowanie sztucznej inteligencji budzi wątpliwości lub wydaje się ryzykowne, należy założyć, że jest ono niedozwolone i zasięgnąć porady.

🟡 Żółte światło — zatrzymaj się i zapytaj (działaj ostrożnie)

Rzeczywiste zarządzanie realizowane jest właśnie w sytuacjach, w których zapala się żółte światło. Są to sytuacje, w których można skorzystać ze sztucznej inteligencji, ale tylko po skonsultowaniu się z odpowiednią osobą lub zastosowaniu określonych zabezpieczeń. Żółte światło nie oznacza „nie", tylko „sprawdźmy to razem". Z biegiem czasu, wraz z pogłębianiem się wiedzy twojego zespołu, liczba konkretnych scenariuszy będzie rosła.

W tych scenariuszach sztuczna inteligencja stanowi przydatne narzędzie, ale twoja ocena sytuacji i ostrożność są absolutnie niezbędne. Podczas wykonywania tych czynności należy zachować ostrożność i przestrzegać poniższych zasad bezpieczeństwa.

- **Sceptyczne podejście do narzędzi zewnętrznych:** Przede wszystkim chroń swoje dane. Jeżeli korzystasz z usług sztucznej inteligencji oferowanych przez podmiot

zewnętrzny (zwłaszcza bezpłatnych), zachowaj szczególną ostrożność i wprowadzaj wyłącznie informacje, które nie są wrażliwe i mają charakter publiczny. Nawet w przypadku zatwierdzonych narzędzi, krytycznie analizuj uzyskane wyniki pod kątem błędów lub stronniczości. Nigdy nie traktuj zewnętrznej sztucznej inteligencji jako ostatecznego autorytetu.

- **Sprawdzanie wszystkiego pod kątem zgodności z faktami:** Wyniki generowane przez sztuczną inteligencję mogą być błędne, zwodnicze lub całkowicie zmyślone („halucynacje"). Nie ufaj im ślepo. Traktuj odpowiedzi wygenerowane przez sztuczną inteligencję jako wstępną wersję, a nie ostateczny fakt. Zanim wykorzystasz lub udostępnisz jakiekolwiek twierdzenia, dane statystyczne lub informacje krytyczne masz obowiązek ich sprawdzenia w oparciu o wiarygodne źródła. *Krótko mówiąc: Sztuczna inteligencja popełnia błędy. Twoim zadaniem jest je wyłapać.*
- **Redagowanie pod kątem jakości, stylu i wartości:** Zanim wykorzystasz jakiekolwiek treści wygenerowane przez sztuczną inteligencję w celach służbowych, dokonaj ich dokładnego przeglądu. Zwróć uwagę na błędy merytoryczne, nieodpowiedni język lub ton, który nie pasuje do stylu twojej organizacji. Nigdy nie przekazuj dalej ani nie publikuj tekstów ani obrazów wygenerowanych przez sztuczną inteligencję bez ich sprawdzenia przez człowieka. W przypadku wewnętrznych wersji roboczych warto je odpowiednio opisać (np. „Pierwsza wersja sporządzona przy pomocy sztucznej inteligencji"), aby zachęcić do zgłaszania uwag i zapewnić transparentność.
- **Uważne wdrażanie nowych narzędzi:** Dodając nowe narzędzie AI lub wtyczkę do istniejącego procesu pracy, najpierw wykonaj okres próbny. Przetestuj to rozwiązanie na małą skalę, aby sprawdzić, jak wpływa na twoje dane i procesy. Uważnie obserwuj wyniki i bądź gotowy do wyłączenia narzędzia w razie pojawienia się jakichkolwiek problemów.
- **Ostrożne podejście w kontaktach z odbiorcami zewnętrznymi:** Zachowuj ostrożność korzystając ze sztucznej inteligencji do komunikacji z odbiorcami (np. poprzez chatbota). Zawsze informuj, że jest to system oparty na sztucznej inteligencji. Dopilnuj, aby odbiorcy mieli jasną i prostą możliwość skontaktowania się z człowiekiem. Uważaj, aby sztuczna inteligencja nie składała obietnic w imieniu organizacji. Dokonuj regularnych przeglądów dzienników interakcji, aby wykryć wszelkie nieodpowiednie odpowiedzi.
- **Zachowanie świadomości zmian zachodzących z upływem czasu:** Narzędzie AI, które dziś działa bezbłędnie, jutro może stać się mniej wiarygodne w związku ze zmianami w modelu lub danych. Jeżeli zauważysz, że dane narzędzie popełnia więcej błędów niż dotychczas lub jest używane w nowym kontekście, zgłoś to. Wszyscy użytkownicy powinni zachować czujność i zgłaszać takie nieprawidłowości, abyśmy mogli się nimi zająć.
- **Oddzielenie użytku prywatnego i służbowego:** Jeśli korzystasz z osobistego narzędzia AI do wykonywania zadań służbowych, nadal musisz przestrzegać niniejszych

zasad. Uważaj, aby w tych narzędziach nie mieszać swoich danych osobowych z danymi organizacji. W miarę możliwości korzystaj z oficjalnych zasobów AI udostępnianych przez organizację, ponieważ zostały one skonfigurowane tak, aby chronić twoje dane.

Podsumowując, zawsze kieruj się zdrowym rozsądkiem i zachowaj krytyczne podejście. Sztuczna inteligencja to potężne narzędzie, ale nie jest idealna. Wykorzystując swoją wiedzę i zachowując ostrożność, możesz czerpać z niej korzyści, unikając jednocześnie pułapek.

🟢 Zielone światło — działaj (zatwierdzone zastosowania)

Zielone światło daje twojemu zespołowi pewność siebie potrzebną do wykonywania rutynowych działań bez potrzeby pytania o zgodę. Dotyczy ono zastosowań zatwierdzonych zgodnie z ustalonymi wytycznymi. Chodzi o jasność, ludzie powinni wiedzieć, co mogą robić, a nie tylko tego, czego nie mogą.

Zachęcamy do korzystania z zatwierdzonych narzędzi AI w następujący sposób, pod warunkiem że zawsze będziesz przestrzegać zasad przewodnich i stosować środki ostrożności określone w niniejszej polityce.

- **Zwiększenie wydajności:** Wykorzystuj sztuczną inteligencję do wykonywania czasochłonnych zadań, takich jak streszczanie długich raportów, tworzenie rutynowych wiadomości e-mail czy porządkowanie informacji. Dzięki temu zyskujesz więcej czasu na zajęcie się zadaniami o charakterze bardziej strategicznym.
- **Pobudzenie kreatywności:** Wykorzystuj generatywną AI jako swojego partnera twórczego. Świetnie nadaje się do burzy mózgów nad pomysłami na kampanie, tworzenia pierwszych szkiców treści lub wizualizacji koncepcji projektowych. Pamiętaj, że sztuczna inteligencja daje zarys projektu, a ty nadajesz mu ostateczny szlif i wnosisz swoje spostrzeżenia.
- **Głębszy wgląd:** Wykorzystuj sztuczną inteligencję do analizy danych, wykrywania trendów lub rozpoznawania wzorców. Chociaż sztuczna inteligencja potrafi wskazać istotne informacje, do interpretacji wyników i podjęcia ostatecznych decyzji potrzebna jest twoja ludzka ocena sytuacji.
- **Zwiększenie zaangażowania (przy zachowaniu nadzoru):** Jeśli korzystasz z sztucznej inteligencji, np. chatbota, do udzielania użytkownikom odpowiedzi na podstawowe pytania, zawsze wyraźnie zaznaczaj, że jest to sztuczna inteligencja. W każdej złożonej lub delikatnej sprawie rozmowa musi zostać przekazana do człowieka.
- **Rozwój umiejętności:** Zachęcamy do eksperymentowania ze sztuczną inteligencją w celu zdobywania nowej wiedzy – niezależnie od tego, czy chodzi o zgłębianie nowego tematu, czy też zapoznanie z zagadnieniem technicznym. Pamiętaj tylko, aby korzystać

z danych przykładowych lub informacji ogólnodostępnych, a nigdy z poufnych danych twojej organizacji.

- **Wprowadzanie innowacji w bezpiecznym środowisku („piaskownice”):** Jeżeli masz świetny pomysł na wykorzystanie sztucznej inteligencji, zachęcamy do jego wypróbowania. Za zgodą przełożonego możesz przeprowadzać niewielkie eksperymenty w kontrolowanym środowisku i z wykorzystaniem danych testowych. Dzięki temu możesz odkrywać nowe możliwości bez podejmowania niepotrzebnego ryzyka.

Złota zasada: Każda sytuacja wymaga twojej ludzkiej czujności, krytycznego myślenia i zdrowego rozsądku. Sztuczna inteligencja to potężne narzędzie, które służy ci pomocą, ale to ty zawsze podejmujesz ostateczną decyzję. Dokładnie sprawdź wynik, upewnij się, że jest zgodny z Twoimi wartościami i weź odpowiedzialność za końcowy efekt pracy.

Istotny niuans: System sygnalizacji świetlnej odnosi się do kombinacji *narzędzie + przypadek użycia + wrażliwość danych*, a nie do samego narzędzia. To samo narzędzie może mieć zielone światło dla przygotowania publicznego wpisu na blogu, ale czerwone dla sporządzenia podsumowania poufnych notatek dotyczących sprawy. W razie wątpliwości potraktuj to jako sytuację niejednoznaczną i skonsultuj się z organem zarządzającym.

Zasady ruchu drogowego: Wytyczne dotyczące wszelkich zastosowań sztucznej inteligencji

Oprócz systemu sygnalizacji świetlnej, wszelkie zastosowania AI podlegają jeszcze tym ustaleniom, nawet w przypadku zielonego światła. Są to wspólne zobowiązania twojego zespołu dotyczące odpowiedzialnego wykonywania codziennych obowiązków.

Oryginalność i własność intelektualna

- *Jakie konkretne przepisy dotyczące własności intelektualnej należy znać kiedy korzysta się ze sztucznej inteligencji?*
- *Czy w naszej jurysdykcji obowiązują jakieś konkretne przepisy (np. dotyczące praw autorskich), które powinniśmy wziąć pod uwagę?*

Treści generowane przez sztuczną inteligencję powstają na podstawie istniejących dzieł, więc nie są w pełni oryginalne. Należy pamiętać o ewentualnych kwestiach związanych z prawami autorskimi. Na przykład obraz lub fragment kodu wygenerowany przez sztuczną inteligencję może być bardzo podobny do materiału objętego ochroną. Zawsze traktuj treści generowane przez sztuczną inteligencję jako pierwszy szkic, a nie gotowy, oryginalny produkt. W przypadku ważnych materiałów, wynik należy sprawdzić pod kątem oryginalności. W razie wątpliwości bezpieczniej jest tworzyć samodzielnie.

Etyczne stosowanie i przestrzeganie przepisów

- *Jakich obowiązujących przepisów lub wewnętrznych zasad (takich jak nasz Kodeks postępowania) musimy zawsze przestrzegać podczas korzystania ze sztucznej inteligencji?*
- *Jak możemy zadbać o to, by nasze wykorzystanie sztucznej inteligencji zawsze odzwierciedlało podstawowe wartości i zasady etyczne naszej organizacji?*
- *Czego wymagają od nas „Warunki korzystania z usługi" dotyczące narzędzi AI, z których korzystamy?*

Dbałość o bezpieczeństwo naszych danych

- *Jakie konkretne rodzaje danych należy uznać za niedostępne dla publicznych narzędzi AI? (np. dane osobowe wolontariuszy, poufne informacje dotyczące partnerów, wewnętrzne dokumenty finansowe).*
- *Jakie środki bezpieczeństwa stosujemy, aby zapewnić ochronę danych wykorzystywanych w zatwierdzonych narzędziach AI?*

Nie wprowadzaj danych poufnych ani osobowych do żadnego publicznie dostępnego narzędzia AI, chyba że zostało ono wyraźnie zatwierdzone do tego celu. Publiczne systemy AI potrafią uczyć się na podstawie wprowadzanych przez ciebie danych. Jeżeli chcesz poruszyć delikatny temat, zamień konkretne dane na symbole zastępcze (np. zamiast prawdziwego nazwiska użyj sformułowania „Klient A"). Zasada ta dotyczy również przesyłania plików.

Zapewnienie jakości: weryfikacja faktów i nadzór ludzki

- *Jak w odpowiedzialny sposób włączyć do naszej pracy treści generowane przez sztuczną inteligencję?*
- *Jakie procedury obowiązują w naszej firmie w zakresie weryfikacji wyników AI przez człowieka przed ich wykorzystaniem?*
- *Jeżeli wykryjemy błędy lub stronniczość w treściach generowanych przez sztuczną inteligencję, w jaki sposób je korygujemy?*

Generatywna AI może generować nieprawdziwe lub całkowicie zmyślone informacje („halucynacje"). Nie zakładaj, że wynik jest poprawny tylko dlatego, że brzmi przekonująco. Za poprawność ostatecznej treści zawsze odpowiadasz ty. Wszelkie dane statystyczne, twierdzenia i inne istotne informacje zawsze weryfikuj odwołując się do wiarygodnego źródła.

Transparentność: kiedy i jak ujawniać fakt wykorzystania sztucznej inteligencji

- *Kiedy powinniśmy informować innych, że do stworzenia danej treści wykorzystano sztuczną inteligencję?*
- *Czy istnieją konkretne sposoby, w jakie powinniśmy to ujawniać, zarówno wewnątrz firmy, jak i na zewnątrz?*

Nasze wytyczne:

Użytek wewnętrzny: W trosce o transparentność wobec współpracowników warto zaznaczać, kiedy przy tworzeniu treści korzystano z pomocy sztucznej inteligencji. Sprzyja to tworzeniu zdrowej kultury kontroli. (np. dodaj adnotację: „Opracowano przy pomocy ChatGPT, wymaga weryfikacji faktów.")

Użytek zewnętrzny: W przypadku wszelkich treści publikowanych na zewnątrz (np. raportów lub postów w mediach społecznościowych) upewnij się, że spełniają one twoje wysokie standardy jakości. Surowego, nieedytowanego tekstu wygenerowanego przez sztuczną inteligencję nie należy nigdy przesyłać partnerom zewnętrznym ani udostępniać publicznie.

8. Zmiany i aktualizacje polityki

Polityka AI nie powinna być statycznym dokumentem, który tylko zbiera kurz. W tej sekcji omówiono sposób, w jaki będziecie wspólnie dbać o to, by niniejsza polityka pozostała użytecznym, aktualnym źródłem informacji, które ewoluuje wraz z waszymi potrzebami i rozwojem technologii.

- *Jaki jest najlepszy sposób zbierania bieżących opinii na temat tej polityki od całego naszego zespołu oraz społeczności?*
- *W jaki sposób zadbamy o to, by te cenne opinie posłużyły do wprowadzenia praktycznych ulepszeń i zapewniły aktualność polityki?*
- *Kto będzie odpowiedzialny za kierowanie procesem przeglądu i jak często powinniśmy formalnie aktualizować niniejszą politykę (np. raz w roku)?*
- *Kiedy niniejsza polityka oficjalnie wejdzie w życie?*

9. Egzekwowanie polityki

Skuteczność polityki zależy od tego, jak dobrze potrafisz ją stosować. W tej sekcji omówiono, jak wdrożyć te wytyczne w sposób sprawiedliwy, wspierający i skuteczny. Pracujmy razem, aby znaleźć odpowiedzi na te kluczowe pytania.

Edukowanie i szkolenie zainteresowanych stron

- *Jak zadbamy o to, by wszyscy – od nowych wolontariuszy po główny zespół – znali i rozumieli tę politykę?*
- *Jakie są najskuteczniejsze sposoby realizacji tego szkolenia? (np. obowiązkowa sesja w ramach wdrożenia do pracy?) Krótki filmik? Krótki przewodnik?)*
- *Jakie są najważniejsze informacje, które każdy powinien znać?*

Praktyczny pomysł: Warto rozważyć stworzenie skróconej, jednostronicowej karty informacyjnej, która w zwięzły sposób przedstawi system sygnalizacji świetlnej i przepisy ruchu drogowego – takiej, którą można trzymać na biurku lub przypiąć jako wiadomość na wspólnym kanale. Często to właśnie ten dokument, a nie pełna treść polityki, jest faktycznie używany na co dzień. Karta informacyjna może obejmować czerwone światła, logikę działania sygnalizacji świetlnej, osoby, z którymi należy się skontaktować w razie pytań, a także najważniejsze przepisy ruchu drogowego. W niektórych organizacjach szkolenia dla współpracowników prowadzi osoba z pierwszej linii, która dobrze rozumie realia codziennej pracy. Forma powinna być dostosowana do specyfiki organizacji: omówienie podczas spotkania zespołu, krótki film lub dyskusje oparte na scenariuszach.

Informowanie o polityce

- *W jaki sposób powinniśmy oficjalnie ogłosić i rozpowszechnić tę politykę, aby dotarła ona do całej naszej społeczności?*
- *Gdzie zostanie umieszczona ta polityka, aby każdy mógł ją łatwo znaleźć i się do niej odwołać? (np. na naszej witrynie internetowej? W folderze udostępnionym na Dysku Google? W podręczniku dla wolontariuszy?)*
- *Jak zamierzamy informować o przyszłych aktualizacjach lub zmianach w polityce?*

Wspólna odpowiedzialność i wsparcie

- *Jak możemy stworzyć kulturę wspólnej odpowiedzialności za przestrzeganie niniejszej polityki, zamiast polegania wyłącznie na kontroli odgórnej?*
- *Jakie proste i nieinwazyjne środki kontrolne możemy wprowadzić, aby zapewnić przestrzeganie tej polityki? (np. kierownicy projektów sprawdzający kluczowe dokumenty? Delikatne przypomnienia podczas spotkań zespołu?)*
- *Skąd będziemy wiedzieć, czy polityka się sprawdza, albo czy niektóre jej elementy są dla ludzi trudne do realizacji?*

Reagowanie na naruszenia polityki (sprawiedliwe i konstruktywne)

- *Jak powinniśmy postępować w sytuacji nieumyślnego błędu i naruszenia polityki? (Naszym celem powinno zawsze być edukowanie, a nie karanie).*
- *Jak powinno wyglądać sprawiedliwe i konstruktywne postępowanie w sytuacjach, gdy polityka jest świadomie lub wielokrotnie lekceważona?*
- *Jakie działania należałoby podjąć w przypadku poważnego, umyślnego nadużycia, zwłaszcza jeśli zagraża to bezpieczeństwu naszych danych, naszej reputacji lub zaufaniu społeczności?*

10. Kontakt

- *Z kim najlepiej skontaktować się w przypadku pytań, uwag lub wątpliwości dotyczących niniejszej polityki?*
- *Podajmy tutaj ich nazwiska i dane kontaktowe tych osób, aby każdy mógł łatwo uzyskać potrzebną pomoc.*

11. Jeżeli możesz dziś zrobić tylko jedną rzecz

Nie każda organizacja jest w stanie przejść przez cały ten szablon za jednym razem, ale to nic złego. Jeżeli potrzebujesz punktu wyjścia, zrób cztery następujące rzeczy:

1. **Określ 3–5 swoich „czerwonych świateł”** (kwestii, które nie podlegają negocjacji). Czego twoja organizacja nigdy nie zrobi przy pomocy sztucznej inteligencji?
2. **Wskaż właściciela polityki.** Jedną osobę. Nazwisko, a nie funkcję.
3. **Powiedz członkom zespołu, do kogo mają kierować pytania.** Kanał, e-mail, osoba – po prostu daj znać.
4. **Zaznacz w kalendarzu termin przeglądu.** Za sześć miesięcy. Zrób to już dziś.

W ten sposób w ciągu 30 minut przygotujesz funkcjonalne ramy polityki. Wszystko inne może się rozwijać od tego punktu.

Niniejszy szablon polityki AI należy do serii „Mapa drogowa odpowiedzialnego AI dla NGOsów” opracowanej przy okazji projektu AI for Social Change realizowanego przez TechSoup w ramach programu Digital Activism Program przy wsparciu Google.org. Wszystkie materiały są udostępniane na licencji Creative Commons Attribution 4.0 International.

O autorce

Ayşegül Güzel jest architektem odpowiedzialnego zarządzania sztuczną inteligencją, który pomaga organizacjom realizującym misję przekształcić lęk przed sztuczną inteligencją w godne zaufania systemy. Jej kariera łączy kierownicze przywództwo społeczne, w tym założenie Zumbara, największej na świecie sieci banków czasu, z prowadzeniem technicznej praktyki AI jako certyfikowany audytor AI i były analityk danych. Prowadzi organizacje przez cały proces transformacji zarządzania AI i przeprowadza techniczne audyty AI. Wykłada na uczelni ELISAVA i prowadzi międzynarodowe odczyty poświęcone podejściu do technologii zorientowanemu na człowieka. Dowiedz się więcej na stronie https://aysegulguzel.info lub zasubskrybuj jej biuletyn AI of Your Choice na stronie https://aysegulguzel.substack.com.