

Stanowisko Fundacji im. Julii Woykowskiej w konsultacjach publicznych dokumentu pn. „Polityki rozwoju sztucznej inteligencji w Polsce do 2030 roku” prowadzonych przez Ministerstwo Cyfryzacji. Przesłane drogą mailową dnia 1 lipca 2025 r.

1. Uwaga do całego dokumentu, szczególnie do obszaru: „Sztuczna inteligencja godna zaufania”

Rozwój sztucznej inteligencji jest zjawiskiem horyzontalnym, przenikającym wszystkie aspekty życia obywateli i obywaterek Polski, zarówno w aspekcie rodzinnym, społecznym czy zawodowym. W tym kontekście ważne jest tworzenie polityk publicznych, które z jednej strony zabezpieczają strategiczne wyzwania i ambicje rynkowe i polityczne, z drugiej jednak strony zwracają uwagę i tworzą odpowiednie mechanizmy dbałości o bezpieczeństwo obywateli i obywaterek, w tym w szczególności najbardziej narażonych na negatywne konsekwencje rozwoju technologicznego jakimi są dzieci i młodzież.

„Polityka rozwoju sztucznej inteligencji w Polsce do 2030 r.” przedstawia szeroką wizję wspierania rozwoju AI z wskazaniem na poszanowanie etyki i praw obywateli. Dokument koncentruje się głównie na stymulowaniu innowacji i korzyści z AI dla gospodarki, porusza również kwestie zaufania i bezpieczeństwa, szczególnie na poziomie technologicznym (szybkie reagowanie na rozprzestrzenianie dezinformacji i tworzenie dedykowanych temu narzędzi). Dokument wspomina też o grupach wrażliwych, w kontekście tworzenia dedykowanych im różnych narzędzi (filtry, kontrole rekomendacji i chatbotów) – sformułowane zapisy są jednak ogólne i nie odnoszą się wprost do największych zagrożeń związanych z rozwojem technologii jak m.in. deepfake. Pojęcia takie jak cyberprzemoc czy przemoc wobec kobiet nie pojawiają się explicite w tekście. Brak konkretnych zapisów oznacza, że ryzyka te nie są wyeksponowane ani szczegółowo przeanalizowane, a założenia systemu wdrażania dokumentu zdefiniowane w opisie obszaru „Sztuczna inteligencja godna zaufania” nie uwzględniają obecności instytucji kompetencyjnie odpowiedzialnych za przeciwdziałanie przemocy (Ministra ds. równego traktowania), ochronę praw i przeciwdziałanie dyskryminacji (Rzecznik Praw Dziecka, Rzecznik Praw Obywatelskich). Brakuje też wskazania odpowiedzialności platform cyfrowych za przeciwdziałanie takim zjawiskom.

W kontekście alarmujących danych dotyczących cyberprzemocy, kryzysów zdrowia psychicznego czy też wzrostu liczby prób samobójczych wśród młodzieży będących bezpośrednim następstwem cyberprzemocy konieczne wydaje się zwrócenie uwagi na te zjawiska w dokumencie.

Rozpowszechnianie deepfake’ów stało się jednym z najgroźniejszych narzędzi cyberprzemocy. Zjawisko to szczególnie dotyka kobiet i dzieci, często w formie przemocy obrazkowej o podłożu seksualnym. Szacuje się, że aż 96% deepfake’ów krążących w sieci to niekonsensualne materiały pornograficzne z wizerunkiem kobiet (Deepfakes and Harm to Women, J. Laffier, Journal of Digital Life and Learning, https://www.academia.edu/105972509/Deepfakes_and_Harm_to_Women).

Tego typu poniżenie o podłożu seksualnym wywołuje ogromne szkody psychiczne: ofiary doświadczają traumy, lęku, depresji, poczucia upokorzenia i utraty kontroli nad własnym życiem. Kobiety narażone na wykorzystanie ich wizerunku często cierpią na zespół stresu

pourazowego (PTSD), problemy z zaufaniem do ludzi oraz głębokie poczucie naruszenia godności. U nastolatków skutkuje to izolacją społeczną – ofiary wstydzą się wychodzić do rówieśników, często wycofują się z życia szkolnego, w skrajnych przypadkach może to doprowadzić do samobójstw.

Drugim ważnym aspektem w kontekście ochrony praw kobiet jest kwestia algorytmów. Algorytmy sztucznej inteligencji – szczególnie w mediach społecznościowych – odgrywają dziś kluczową rolę w kształtowaniu wizerunku kobiet. Badania pokazują, że algorytmy platform faworyzują treści seksualizujące kobiety, promując wyidealizowane, często nierealne obrazy kobiecego ciała (z filtrowaniem twarzy i sylwetki, poprawianiem urody itp.). oraz, że kobiety są niedoreprezentowane w wynikach wyszukiwania lub rekomendacjach eksperckich, co utrwala przekonanie o „braku autorytetu” kobiet w sferze publicznej.

W kontekście ochrony praw swoich obywateli i obywaterek, do których Polska jest zobowiązana na mocy konstytucji i innych aktów prawa europejskiego i krajowego zasadnym wydaje się rozpatrzenie następujących działań w kontekście dokumentu „Polityka rozwoju sztucznej inteligencji w Polsce do 2030 r.” oraz jego wdrażania:

1. Na poziomie dokumentu:

1) dodanie diagnozy/ wyciągu z diagnozy uwzględniającego również zagrożenia płynące z rozwoju AI lub dodanie do części obrazującej wyzwania zapisów dotyczących nie tylko wyzwań technologicznych, konkurencyjnych i biznesowych, ale i społecznych ze szczególnym uwzględnieniem cyberprzemocy, w tym deepfake’ów.

2) dodanie do obszaru „Sztuczna inteligencja godna zaufania” celu związanego wprost z Przeciwdziałaniem cyberprzemocy np. o pasującym do pozostałych brzmieniu „Ochrona przed cyberprzemocą” w ramach którego uwzględnione będą działania, mechanizmy, narzędzia, instytucje oraz środki na przeciwdziałanie przemocy. W tym kontekście kluczowe wydaje się: wyznaczenie koordynatora ds. przeciwdziałania cyberprzemocy, który miałby kompetencje horyzontalnego łączenia różnych obszarów i ich przedstawicieli – przede wszystkim systemu równego traktowania i przeciwdziałania przemocy (Ministra ds. równego traktowania), systemu sprawiedliwości (Ministerstwo Sprawiedliwości), edukacji (MEN), systemu wsparcia rodziny (MRPiPS), Rzecznika Praw Dziecka i Rzecznika Praw Obywatelskich, NASK, Ministerstwa Cyfryzacji, MSW (policja) itd.

Koordynator mógłby być umiejscowiony przy systemie przeciwdziałania przemocy, a jego powinien jednak dysponować swoim budżetem oraz uprawnieniami pozwalającymi na faktyczną koordynację działań innych jednostek, wymianę informacji. W zakres realizowanych działań mogłyby wchodzić m.in.:

- Opracowanie horyzontalnego dokumentu – planu wykonawczego na rzecz przeciwdziałania cyberprzemocy spójnego z pozostałymi dokumentami strategicznymi w tym Strategią Rozwoju AI czy Krajowym Planem Przeciwdziałania Przemocy czy Krajowym Planem na Rzecz Równego Traktowania,
- Koordynacja działań strategicznych, budżetów poszczególnych agend,
- Koordynacja dostosowywania przepisów prawa karnego oraz wzmocnienia ochrony osób doznających cyberprzemocy,
- Koordynacja współpracy z platformami i egzekwowania ich obowiązków ochrony osób doznających cyberprzemocy,

- Koordynacja i realizacja badań związanych z cyberprzemocą i cyberdyskryminacją, w tym zobowiązanie do prowadzenia audytów algorytmicznych pod kątem płci – zarówno w administracji publicznej, jak i w dużych serwisach cyfrowych działających w Polsce oraz wprowadzenie standardów etycznych dot. reprezentacji kobiet w systemach rekomendacji, filtrach upiększających, modelach AI generujących obrazy i teksty.
- Zwiększenie skali działań w zakresie edukacji i świadomego korzystania z Internetu w ramach edukacji medialnej w szkołach oraz kampanii społecznych oraz kampanii zero tolerancji dla deep fake'ów.

Punktem odniesienia czy też inspiracji mogłaby być instytucja powołana w Australii – eSafety Commisionner (<https://www.esafety.gov.au/about-us>), która skupiałaby kompetencje w zakresie ochrony użytkowników Internetu (w tym dzieci) i mogła szybciej niż zwykłe organy reagować na zgłoszenia o szkodliwych treściach.

2. Uwaga do obszaru „Sprawne państwo” - obszarem wartym uwzględnienia w politykach publicznych powinien być rozwój systemu monitorowania realizacji zadań publicznych i dostępu do danych generowanych w ramach sprawozdawczości publicznej. W nowoczesnym państwie, w którym powszechna staje się idea Evidence Based Policy kwestia jakości danych służących do monitoringu i ewaluacji powinna zostać zauważona.

Na dzień dzisiejszy w Polsce nie funkcjonuje system ujednoliconego podejścia do sprawozdawczości zadań realizowanych w różnych systemach (np. oświaty, polityki społecznej, równego traktowania). Przykładowo w systemie polityki społecznej generowane są ogromne ilości danych, które nie mają szansy być wykorzystane, ze względu na zmienność kształtu sprawozdań rok do roku (np. ocena zasobów pomocy społecznej), brak ujednoliconego podejścia do zbierania zakresu danych, w tym np. uwzględniania płci jako jednej z istotnych danych oraz sam kształt aplikacji (np. centralnej aplikacji statystycznej). Skutkuje to brakiem wiarygodnych informacji o odbiorcach zadań publicznych.

System polityki społecznej, równego traktowania, przeciwdziałania przemocy, edukacji również wymagają strategicznej interwencji w cyfryzacji zarządzania. Są to systemy, które na poziomie zarządczym i budżetowym generują ogromne koszty m.in. w budżecie centralnym i budżetach jednostek samorządu terytorialnego (ok. 20 % PKB to tzw. „wydatki socjalne” wg Eurostatu), jednak nie dysponują mechanizmami sprawnej oceny jakości zadań, racjonalności budżetów oraz kosztów zaniechania. Procesy te mogłyby zostać usprawnione poprzez odpowiednie wsparcie programistyczne oraz w zakresie sztucznej inteligencji, co powinno zostać uwzględnione w obszarze „Sprawne państwo”.