



Współpraca międzynarodowa w oparciu o godną zaufania Sztuczną Inteligencję

Newsletter sekcji – wydanie 7

Październik 2026



Ministerstwo
Cyfryzacji

Tu tworzymy przyszłość

GRAi
GRUPA ROBOCZA
DS. SZTUCZNEJ INTELIGENCJI

Od redakcji

Szanowni Państwo,

Drodzy Przyjaciele i Przyjaciółki Sekcji,

To już 7 wydanie newslettera naszej Sekcji ds. współpracy międzynarodowej w oparciu o godną zaufania sztuczną inteligencję. Dzięki wsparciu Ministerstwa Cyfryzacji oraz zaangażowaniu Grupy Roboczej ds. Sztucznej Inteligencji (GRAI) tworzymy publikację, która ma na celu wspieranie wymiany wiedzy i budowanie międzynarodowego dialogu w tej dynamicznie rozwijającej się dziedzinie.

To wyjątkowe wydanie – w związku z zainteresowaniem innych sekcji naszej Grupy Roboczej, zaprosiliśmy do udziału w naszych pracach nowych autorów i autorki. Witamy na pokładzie i dziękujemy za wartościowy wkład!

W niniejszym numerze znajdziecie 11 tekstów autorów i autorek. Mamy nadzieję, że prezentowane tu teksty przybliżą naszym czytelnikom i czytelniczkom nowe zagadnienia, dzięki którym jeszcze lepiej będziemy w stanie poruszać się w świecie sztucznej inteligencji i innych pokrewnych zaawansowanych technologii.

Ponownie zachęcamy do włączania się w nasze prace wszystkich członków i członkinie GRAI – wierzymy, że właśnie tego typu zaangażowanie może przynosić satysfakcję i prowadzić do prawdziwej synergii działań¹.

Michał Jackowski

Urszula Uchmańska

¹ Newsletter przygotowali eksperci działający na zasadach pro publico bono w Grupie Roboczej ds. Sztucznej Inteligencji (GRAI) przy Ministerstwie Cyfryzacji w ramach sekcji Współpraca międzynarodowa w oparciu o godną zaufania Sztuczną Inteligencję.

Ani Rada Ministrów, ani żadna osoba działająca w imieniu Rady Ministrów nie ponosi odpowiedzialności za sposób wykorzystania zamieszczonych w niniejszym materiale informacji. Wyłącznie odpowiedzialność za treści zawarte w newsletterze ponoszą jego autorzy. Poglądy w nim wyrażone odzwierciedlają opinię autorów i w żadnym wypadku nie mogą być postrzegane jako oficjalne stanowisko Rady Ministrów ani jej poszczególnych członków.

Dokument może być kopiowany i wykorzystywany publicznie jedynie bez naruszania jego spójności. Prawa autorskie i majątkowe do materiałów wykorzystanych w raporcie, które pochodzą z obcych źródeł, należą do ich właścicieli.

Spis treści

Od redakcji	2
Czy przewodnik AI może być godny zaufania? Dziedzictwo kulturowe jako infrastruktura sztucznej inteligencji w turystyce	4
Regulacje dotyczące treści generowanych przez AI w kampaniach crowdfundingowych cz.2/2	7
Nie tylko Digital Omnibus - projekt uproszczonych środków ochrony danych osobowych w Chinach	10
AI w zastosowaniach krytycznych: dlaczego o wartości systemów wizyjnych decydują walidacja, powtarzalność i warunki operacyjne	14
AI Index 2026: technologia szybsza niż systemy, które ją oswajają	16
Czy można zaufać rozumowaniu AI? - Traceability modeli, chain of thought i nowe wyzwania dla AI Governance	18
Dokumentowanie realnego nadzoru człowieka nad systemami AI	21
Oznaczanie treści AI: czy techniczne watermarki wystarczą do walki z dezinformacją?	23
AI & Rynek Pracy – Fakty, Trendy, Prognozy	25
Legislacja AI w Polsce nabiera tempa - ustawa o systemach AI (USSI) już obowiązuje	29
Cyberobrona państwa oparta na agentach AI: czego Polska może nauczyć się z brytyjskiego projektu Cyber Shield i strategii cyfrowej NATO	35
Sekcja ds. współpracy międzynarodowej w oparciu o godną zaufania sztuczną inteligencję	38

Czy przewodnik AI może być godny zaufania? Dziedzictwo kulturowe jako infrastruktura sztucznej inteligencji w turystyce

Maciej Badowicz

Wykorzystanie sztucznej inteligencji w humanistyce cyfrowej pokrywa także istotny z ekonomicznego punktu widzenia obszar turystyki. Obecnie zasoby światowego dziedzictwa, bibliotek stają się podwalinami do tworzenia m.in. „inteligentnych tras turystycznych”.

Strategia cyfrowego dziedzictwa dla Europy

Unia Europejska rozwija od kwietnia 2025 roku AI Continent Action Plan. Opublikowany w ramach szerszej strategii cyfrowej UE, plan działania na kontynencie dotyczący AI przedstawia długoterminową wizję, w ramach której Europa przechodzi od eksperymentów do godnego zaufania wdrożenia sztucznej inteligencji. To łączy się z obszarem humanistyki cyfrowej, bo zasoby muzeów, archiwów, bibliotek i repozytoriów mogą stać się materiałem dla rzetelnych, źródłowych modeli i aplikacji dziedzictwa kulturowego i turystyki.

Ponadto nadal aktualna jest strategia UE Common European Data Space for Cultural Heritage 2025–2030, łącząca dane dziedzictwa z projektami wirtualnymi i europejską przestrzenią danych turystycznych. Europeana wprost wpisuje się w aktualne działania zmierzające do stworzenia kulturowego „kontynentu AI” w Europie.

Istotnym będzie także dostarczenie rozwiązań, które będą miały konkretne zastosowanie społeczne i gospodarcze. Takim mogą się stać źródłowo ugruntowany przewodniki AI – „inteligentne trasy turystyczne” – po historycznych miastach i obiektach, oparte na zweryfikowanych kolekcjach muzealnych, archiwalnych, kartograficznych i ikonograficznych, zintegrowane z europejską przestrzenią danych dziedzictwa kulturowego.

Ryzyka i rozwiązania

Jednak zauważalne będą niebezpieczeństwa, typowe dla wszystkich rozwiązań opartych o sztuczną inteligencję. W tym wypadku będą związane z „halucynacjami historycznymi” (tworzeniem przekonujących, ale fałszywych narracji o wydarzeniach, obiektach i tradycjach i innych).

Ponadto jeśli dane trafią do komercyjnych modeli bez mechanizmów atrybucji, kontroli i wzajemności, instytucje mogą utracić wpływ na sposób reprezentowania swoich zbiorów. Dlatego środowiska skupione wokół Europeany akcentuje wartości istotne dla wszystkich jednostek: użyteczność, wzajemność i wiarygodność.

Modele trenowane na zasobach najlepiej zdigitalizowanych mogą wzmacniać dominację największych języków, instytucji i narracji. Stąd siła m.in. Bielika, zbudowanego w oparciu o specyfikę polskiego kodu kulturowego.

W praktyce potrzebne będą:

- jednoznaczne metadane;
- informacja o prawach autorskich i warunkach ponownego wykorzystania;
- trwałe identyfikatory obiektów;
- dokumentacja pochodzenia danych;
- informacje o jakości skanów, transkrypcji i modeli 3D;
- mechanizmy kontroli, kto i w jakim celu korzysta z danych.

Podsumowanie

Najbardziej przyszłościowe rozwiązania w zakresie dziedzictwa kulturowego i humanistyki cyfrowej dla potrzeb wdrożenia rozwiązań społeczno – gospodarczych wymagają następujących działań:

- **Audyt zbiorów pod kątem AI-readiness**

Czy obiekty mają metadane, licencje, identyfikatory, powiązania ze słowami kluczowymi?

- **Budowa korpusów źródłowych**

Humanistyka cyfrowa powinna tworzyć kontrolowane korpusy, np. miejskie, regionalne, tematyczne, ikonograficzne, kartograficzne, biograficzne.

- **Modele z cytowaniem źródeł**

W edukacji, turystyce i muzeach warto preferować rozwiązania typu RAG, które odpowiadają na podstawie wskazanych kolekcji, a nie „z pamięci modelu”.

- **Rozwój kompetencji humanistyczno-technicznych**

Potrzebni będą kuratorzy danych, historycy cyfrowi, historycy danych i projektanci doświadczeń AI.

Dla instytucji dziedzictwa i środowiska humanistyki cyfrowej najważniejsza zmiana polega na tym, że jakość danych stanie się równie ważna jak sama digitalizacja. W nowym ekosystemie nie wystarczy mieć skany lub fotografie. Trzeba mieć dane, które mogą być bezpiecznie i sensownie użyte przez ludzi, wyszukiwarki i modele AI. Tylko wtedy możliwym będzie stworzenie dobrej jakości produktu, m.in. do wykorzystania w sektorze turystyki.

Źródła:

- [Komunikat Komisji Europejskiej na temat planu działania „AI Continent” wyznaczającego europejską mapę drogową rozwoju godnej zaufania sztucznej inteligencji.](#)
- [Dokument strategiczny Unii Europejskiej na lata 2025–2030 dotyczący wspólnej europejskiej przestrzeni danych dla dziedzictwa kulturowego.](#)
- [Plan biznesowy Fundacji Europeana na rok 2026 przedstawiający cele rozwoju cyfrowego dziedzictwa kulturowego.](#)

Regulacje dotyczące treści generowanych przez AI w kampaniach crowdfundingowych cz.2/2

Mateusz Hyla

Indiegogo – przeniesienie obowiązku na twórców

Indiegogo wymaga od twórców kampanii oznaczania jakiegokolwiek wykorzystania generatywnej sztucznej inteligencji poprzez wyraźne ujawnienie tego faktu na stronie projektu. Dotyczy to w szczególności opisów kampanii, materiałów wizualnych, filmów promocyjnych czy innych materiałów marketingowych oraz tworzenia lub rozwoju produktu oferowanego za pośrednictwem portalu.

Serwis wymaga, aby oznaczenie:

- wskazywało, które konkretne elementy treści zostały wygenerowane przy użyciu AI,
- było widoczne dla wspierających przed rozpoczęciem składania zobowiązań,
- zapewniało jasność i przejrzystość w prezentacji projektu.

Warto zauważyć, że portal nie wymusza konkretnej formy, miejsca ani sposobu informowania wspierających. Zapisy odnośnie wykorzystania AI nie znajdują się w Regulaminie, a w Wytocznych dla Twórców.

BackerKit – znaczne restrykcje w wykorzystaniu AI

BackerKit, według medialnych doniesień, całkowicie zablokował projekty wykorzystujące generatywne AI, jednak w rzeczywistości odnosi się to wyłącznie do elementów stworzonych w całości przez sztuczną inteligencję oraz projektów mających na celu tworzenie narzędzi generatywnego AI.

Serwis dopuszcza wykorzystywanie AI do edycji obrazów, poprawy składni i gramatyki oraz transkrypcji, tak długo jak treści stworzone przez człowieka są rozpoznawalne. Jednocześnie nie zezwala na wykorzystywanie umieszczonych w nim materiałów do trenowania modeli AI.

Pomimo tak znacznych restrykcji, nie zostały one wpisane w Regulaminie strony, ale w dokumencie "Zasady Crowdfundingu" oraz w dokumencie "Polityka AI".

Patreon – ograniczenie AI wyłącznie w treściach dla dorosłych

Serwis Patreon nie służy co prawda do jednorazowych zbiorów, ale do stałego pozyskiwania środków, często w zamian za nowe materiały. Z tego powodu zasady dotyczące crowdfundingu znajdują tu częściowe zastosowanie.

Portal wprowadził interesujący zakaz umieszczania wygenerowanych przez AI hiperrealistycznych treści w kategorii Adult/18+, o ile nie dotyczą one prawdziwych osób, które w sposób udokumentowany wyraziły swoją zgodę. Pozostałe wykorzystanie AI jest dozwolone – w tym ilustrowane dzieła w tej kategorii oraz hiperrealistyczne przedstawienia osób w ramach treści bezpiecznych dla wszystkich odbiorców.

Zakazane jest również udostępnianie generatorów, narzędzi i oprogramowania wykorzystujących AI do tworzenia materiałów o charakterze erotycznym, erotycznych chatbotów AI oraz innych wirtualnych doświadczeń erotycznych, a także materiałów edukacyjnych uczących, jak takie aplikacje tworzyć. Wszystkie wskazane zakazy zostały umiejscowione w Regulaminie.

Zrzutka – udostępnienie narzędzi AI do projektowania kampanii

Dla porównania warto przedstawić również polski portal Zrzutka.pl, który sam zaoferował twórcom zintegrowaną sztuczną inteligencję do generowania opisów i grafik. Serwis automatycznie i trwale oznacza takie treści jako wygenerowane przez AI.

Na twórcę projektu nałożono obowiązek zweryfikowania wygenerowanych materiałów – zarówno pod kątem zgodności z regulaminem, jak i innymi przepisami prawa. Co ciekawe, regulamin nie odnosi się w żaden sposób do treści AI wygenerowanych poza serwisem.

AI poza regulaminami

Z przeprowadzonego przeglądu portali crowdfundingowych wynika, że w większości przypadków zapisy dotyczące wykorzystania sztucznej inteligencji w projektach umieszczane są poza głównym regulaminem. Praktyka ta wpisuje się w branżowy trend tzw. „soft law”, w którym zasady nie mają charakteru prawnego, lecz są wiążące operacyjnie.

Takie rozwiązanie pozwala platformom elastycznie dostosowywać przepisy do zmieniającego się prawa oraz do dynamicznych oczekiwań społecznych. Dzięki temu mogą one również szybko reagować na nadużycia związane z wykorzystaniem AI i mają większą swobodę w ocenie naruszeń.

Z drugiej strony, taka praktyka sprawia, że portale nie muszą formalnie informować użytkowników o zmianach w polityce AI i jednocześnie minimalizują ryzyko prawne. Mogą w ten sposób moderować treści lub zawieszać konta bez ponoszenia formalnej odpowiedzialności umownej.

Źródła:

- [Regulamin i wytyczne platformy Indiego skierowane do twórców projektów crowdfundingowych.](#)
- [Artykuł w serwisie Polygon omawiający wprowadzenie zakazu stosowania sztucznej inteligencji na platformie BackerKit.](#)
- [Oficjalny wpis na blogu BackerKit przedstawiający politykę platformy wobec wykorzystania AI w kampaniach.](#)
- [Centrum pomocy BackerKit zawierające zasady postępowania obowiązujące w projektach finansowania społecznościowego.](#)
- [Artykuł pomocy technicznej serwisu Patreon wyjaśniający politykę dotyczącą AI dla twórców treści przez dorosłych \(18+\).](#)
- [Aktualizacja polityki serwisu Patreon z połowy 2024 roku opracowana na podstawie opinii społeczności twórców.](#)
- [Regulamin prawny i warunki korzystania z serwisu Patreon.](#)
- [Regulamin i warunki świadczenia usług serwisu zbiorów internetowych Zrzutka.pl.](#)

Nie tylko Digital Omnibus - projekt uproszczonych środków ochrony danych osobowych w Chinach

dr hab. prof. AKP Mariusz Krzysztofek

W dniu 22 lipca 2026 r. Chińska Administracja Cyberprzestrzeni (Cyberspace Administration of China - CAC) oraz Ministerstwo Bezpieczeństwa Publicznego ogłosiły przepisy dotyczące uproszczonych środków ochrony danych osobowych dla małoskalowych administratorów danych (*Provisions on Simplified Measures for Personal Information Protection by Small-Scale Personal Information Processors*, Order No. 25). Przepisy weszły w życie 1 września 2026 r.

Są ważnym, szczególnie że kolejnym, sygnałem liberalizacji chińskiej ustawy o ochronie danych osobowych - Personal Information Protection Law (PIPL), choć w odniesieniu do podmiotów o ograniczonej skali przetwarzania danych. Należy podkreślić, że nie prowadzą do obniżenia ustawowych standardów ochrony, lecz zmierzają do dostosowania wymogów dokumentacyjnych i proceduralnych do możliwości organizacyjnych mniejszych podmiotów. Odpowiadają na rosnące znaczenie startupów i MŚP dla innowacji i rozwoju gospodarki Chin.

To następna istotna liberalizacja chińskich przepisów o ochronie danych, poprzednia miała pozytywny wpływ również na firmy europejskie prowadzące działalność transgraniczną obejmującą Chiny i wymagającą transferów danych poza ten kraj. Była to zmiana wprowadzona przez przepisy Chińskiej Administracji Cyberprzestrzeni dotyczące promowania i regulowania transgranicznego przepływu danych (Regulation on Promoting and Regulating Cross-border Data Flows), które zostały ogłoszone i weszły w życie 22.03.2024 r., oraz uprościły transfery transgraniczne i zbliżyły je do modelu RODO.

Dodam, że nowe przepisy potwierdzają te uproszczone zasady (w art. 11), a dodatkowo jeżeli małoskalowy administrator jest obowiązany do przejścia oceny bezpieczeństwa dokonanej przez CAC zgodnie z art. 40 PIPL (np. ze względu na zamiar transferu poza Chiny danych osobowych na poziomach ilościowych określonych przez CAC w przepisach z 22.03.2024 r.) nowa regulacja zwiększa rolę regionalnego organu CAC w wydawaniu oceny, co ma na celu usprawnienie procedury. Jako „małoskalowego administratora danych” nowe przepisy definiują administratora przetwarzającego dane osobowe mniej niż 100 000 osób (art. 2). Próg liczby osób, których dane są przetwarzane, jako kryterium jest zbieżny z innymi chińskimi przepisami w obszarze ochrony danych, m.in. o transgranicznych transferach danych, natomiast różni się od kryteriów stosowanych w Unii Europejskiej dla „small mid-cap enterprises” w Digital Omnibus, gdzie uwzględnia się m.in. liczbę pracowników czy obrót. Jednak nowa regulacja nie określa ram czasowych dla progu 100 000 osób (np. w ujęciu rocznym), co może powodować trudności interpretacyjne i praktyczne.

Przepisy przewidują możliwość realizacji obowiązku informacyjnego przez udostępnienie informacji o przetwarzaniu danych w siedzibie przedsiębiorstwa, np. w formie plakatu w widocznym miejscu, przy wejściu lub stanowisku obsługi (w przypadku zbierania danych offline) albo włączenie ich do regulaminu świadczenia usług lub komunikatów w aplikacji (w przypadku zbierania danych online) – art. 4. To wyraźne odejście od obowiązku przedstawienia odrębnej polityki prywatności. Uproszczeniu,

jednak z wyjątkiem informacji przekazywanych dzieciom poniżej 14. roku życia, podlega również zakres informacji w ramach klauzuli informacyjnej.

Ponadto administrator może ograniczyć się do publicznego udostępnienia informacji o przetwarzaniu w łatwo dostępnej i możliwej do utrwalenia formie zamiast indywidualnego informowania każdego podmiotu danych o aspektach przetwarzania, pod warunkiem m.in., że przetwarza wyłącznie dane osobowe niezbędne do świadczenia swoich usług lub dostarczania produktów, nie przetwarza danych wrażliwych, nie udostępnia danych osobowych innym podmiotom ani nie upublicznia ich (art. 6).

Jednak to uproszczenie może nie znaleźć zastosowania w praktyce wielu firm, ponieważ przetwarzanie danych dotyczących płatności wiąże się z przetwarzaniem danych dotyczących rachunków finansowych, które art. 28 PIPL kwalifikuje jako dane wrażliwe, a ponadto przedsiębiorstwa funkcjonujące w grupach korporacyjnych w praktyce muszą przekazywać dane osobowe pomiędzy podmiotami grupy.

Uproszczeniu uległy również obowiązki informacyjne związane z przekazywaniem danych osobowych w następstwie połączenia lub podziału małoskalowego administratora. W przypadku usług świadczonych online obowiązek ten może zostać spełniony poprzez zamieszczenie komunikatu popup, wyświetlanego przez co najmniej 30 dni roboczych przed planowaną operacją.

Liberalizacją dla administratorów działających w ramach platform internetowych jest możliwość korzystania ze wspólnej polityki prywatności udostępnianej przez operatora platformy (art. 8). Dotyczy to więc m.in. podmiotów przetwarzających dane osobowe za pośrednictwem dużych platform, takich jak Douyin - chińskiego odpowiednika TikToka funkcjonującego wyłącznie na rynku krajowym czy Taobao, będącej kluczową platformą e-commerce w ekosystemie Alibaba. Skorzystanie z tego uproszczenia jest możliwe pod warunkiem, że operator platformy wykonuje obowiązki informacyjne wobec osób, których dane dotyczą, a przedsiębiorstwo nie przekazuje danych osobowych poza ekosystem platformy.

Ta zasada ma analogiczne zastosowanie również do działalności offline, mogą z niej skorzystać małoskalowi administratorzy prowadzący działalność np. w galeriach handlowych (art. 5). Publiczne ogłoszenie przez zarządcę obiektu w widocznym miejscu jednolitej polityki prywatności dla podmiotów prowadzących określonych typ działalności, oraz zgoda na jej stosowanie przez wprost wskazanego administratora, zwalnia go z obowiązku tworzenia własnej polityki prywatności.

Nowa regulacja wprowadza również ułatwienia dla małoskalowych administratorów danych w zakresie obowiązków audytowych i oceny skutków dla ochrony danych (art. 8, art. 13, art. 14, art. 15 i art. 17). Kluczowym rozwiązaniem jest uproszczony formularz samooceny zgodności, funkcjonalnie zbliżony do DPIA, który wystarczy przeprowadzać raz na pięć lat. Podmioty posiadające certyfikację z zakresu ochrony danych mogą zostać zwolnione z obowiązku audytów na okres obowiązywania certyfikatu.

Niezależnie od certyfikacji, małoskalowi administratorzy będą podlegać obowiązkowi audytowemu rzadziej niż pozostałe organizacje. A administratorzy działający w ramach platform internetowych będą mogli opierać się na audytach zgodności lub ocenach skutków dla ochrony danych przeprowadzonych przez operatora platformy, co zwalnia ich z konieczności samodzielnego wykonywania tych obowiązków.

Nowe przepisy przewidują bardziej liberalne i elastyczne podejście do egzekwowania przepisów wobec małoskalowych administratorów danych (art. 18 i 19). W szczególności odstąpienie od nakładania istotnych sankcji może mieć miejsce w przypadku naruszeń o nieznacznym charakterze, które nie spowodowały negatywnych skutków i zostały niezwłocznie usunięte, a także w sytuacji pierwszego naruszenia o ograniczonych skutkach, szybko naprawionego przez administratora. Jednocześnie organy nadzorcze zachowują możliwość stosowania innych środków administracyjnych, takich jak ostrzeżenia czy rozmowy wyjaśniające. Nowa regulacja wpisuje się w obserwowaną w chińskim prawie tendencję do promowania współpracy regulacyjnej i dobrowolnego compliance, przewidując możliwość łagodzenia odpowiedzialności w przypadku dobrowolnego ujawnienia incydentu, współpracy z organami nadzorczymi oraz podejmowania działań naprawczych. Widoczne podejście deregulacyjne w wybranych obszarach wobec małoskalowych administratorów nie oznacza zasadniczego odejścia od wymagającego, w znacznym stopniu zbieżnego z RODO, chińskiego ustawowego modelu ochrony danych i nadzoru nad nim. Publikacja nowych przepisów zbiegła się z ogłoszeniem przez CAC i inne organy planu działań kontrolnych na 2026 r. obejmującego m.in. aplikacje mobilne, reklamę internetową, sektor finansowy, ochrony zdrowia i edukacji, oraz przestępstwa związane z danymi osobowymi.

Źródła:

- [Oficjalne oświadczenie Chińskiej Administracji Cyberprzestrzeni \(CAC\) dotyczące regulacji danych i infrastruktury cyfrowej.](#)
- [Artykuł w Przeglądzie Ustawodawstwa Gospodarczego analizujący transgraniczne transfery danych osobowych po chińskiej nowelizacji z marca 2024 r. w porównaniu do RODO.](#)
- [Analiza prawna w serwisie WeChat dotycząca chińskich przepisów nakładających obowiązki przy transgranicznym przesyłaniu danych.](#)

AI w zastosowaniach krytycznych: dlaczego o wartości systemów wizyjnych decydują walidacja, powtarzalność i warunki operacyjne

dr inż. Jakub Suder

W debacie o sztucznej inteligencji dominują dziś modele generatywne i systemy agentowe. Z perspektywy wdrożeń inżynierskich jest to jednak tylko fragment znacznie szerszego obrazu. W obszarach takich jak lotnictwo, motoryzacja, obronność czy monitoring infrastruktury krytycznej, kluczową rolę odgrywają nie rozwiązania generujące treść, lecz systemy, które analizują obraz i jego kontekst, a na tej podstawie wspierają ocenę sytuacji w czasie rzeczywistym.

To właśnie w systemach wizyjnych szczególnie dobrze widać różnicę między efektowną demonstracją a realną wartością technologiczną. Rozwiązania te mają wykrywać obiekty, klasyfikować zdarzenia, identyfikować anomalie, oceniać stan infrastruktury albo wspierać operatora w podejmowaniu decyzji. Ich zadaniem nie jest „brzmieć przekonująco”, lecz działać w sposób mierzalny, przewidywalny i powtarzalny, szczególnie w terenie.

W zastosowaniach krytycznych samo wykorzystanie modelu AI nie przesądza jeszcze o wartości systemu. Znaczenie ma to, czy całe rozwiązanie zostało poprawnie zaprojektowane i zweryfikowane. Obejmuje to jakość i reprezentatywność danych, dobór kamer, sensorów oraz jednostki obliczeniowej, procedury kalibracji, sposób anotacji, definicję klas, przyjęte metryki, a także zachowanie systemu w zmiennych warunkach środowiskowych. W praktyce właśnie te elementy decydują o tym, czy rozwiązanie będzie użyteczne operacyjnie.

Powtarzalność oznacza przy tym coś więcej niż uzyskanie wysokiego wyniku podczas pojedynczego eksperymentu. System powinien zachowywać zbliżoną skuteczność dla różnych sekwencji danych, kolejnych przejazdów, zmieniających się lokalizacji i odmiennych warunków rejestracji. Konieczne jest również rozdzielenie danych treningowych i testowych w sposób, który pozwala rzeczywiście ocenić zdolność modelu do działania na nowych materiałach. Bez tego bardzo dobry wynik może świadczyć przede wszystkim o dopasowaniu do konkretnego zbioru, a nie o gotowości do wdrożenia.

W lotnictwie systemy wizyjne mogą wspierać między innymi inspekcję oświetlenia nawigacyjnego, detekcję ciał obcych (FOD), ocenę nawierzchni czy analizę oznakowania. W motoryzacji odpowiadają za percepcję otoczenia, wykrywanie przeszkód, interpretację sceny drogowej i wsparcie systemów bezpieczeństwa. W zastosowaniach wojskowych i dual-use ich znaczenie rośnie wszędzie tam, gdzie liczą się szybka analiza obrazu, odporność na zakłócenia oraz możliwość przetwarzania danych bezpośrednio na urządzeniach brzegowych. We wszystkich tych przypadkach kluczowe pytanie brzmi nie „czy użyto AI?”, ale „czy system działa wiarygodnie w warunkach rzeczywistych?”.

Dlatego ocena systemu wizyjnego nie powinna opierać się wyłącznie na skuteczności uzyskanej na zbiorze testowym. Równie ważne jest jego zachowanie przy zmianach oświetlenia i pogody, zabrudzeniu optyki, drganiach kamery, częściowym zasłonięciu obiektów, ograniczonej widoczności

czy pogorszeniu jakości danych. W środowisku operacyjnym to właśnie te czynniki ujawniają rzeczywiste granice działania modelu.

Istotna jest także analiza rodzaju popełnianych błędów. W systemie bezpieczeństwa fałszywy alarm i niewykrycie rzeczywistego zagrożenia mogą mieć zupełnie inne konsekwencje. Z tego powodu nie zawsze wystarcza jedna zbiorcza miara skuteczności. Potrzebne są metryki dobrane do konkretnego zadania, ocena przypadków granicznych oraz określenie sytuacji, w których wynik systemu powinien zostać zweryfikowany.

Z perspektywy praktyki wdrożeniowej AI w systemach wizyjnych należy więc traktować jako element większej architektury technicznej, a nie samodzielne rozwiązanie. O wartości całego systemu decydują walidacja terenowa, jednoznaczne metryki, analiza błędów, dokumentacja ograniczeń, wydajność sprzętowa i właściwie zaprojektowany nadzór człowieka. Ważna jest również zdolność rozpoznania sytuacji, w której dane wejściowe wykraczają poza warunki przewidziane podczas projektowania i testowania.

W zastosowaniach krytycznych potrzeba dziś mniej narracji o możliwościach AI, a więcej rzetelnej inżynierii, której wyniki można sprawdzić, porównać i obronić. Dopiero wtedy system wizyjny staje się rzeczywistym wsparciem dla człowieka, a nie jedynie obiecującą demonstracją technologii.

Źródła:

- [Publikacja NIST AI 100-1 określająca ramy zarządzania ryzykiem sztucznej inteligencji \(AI Risk Management Framework\).](#)
- [Mapa drogowa Agencji Unii Europejskiej ds. Bezpieczeństwa Lotniczego \(EASA\) dotycząca integracji AI w lotnictwie \(wersja 2.0\).](#)
- [Artykuł naukowy w czasopiśmie Applied Sciences dotyczący zaawansowanych zastosowań sztucznej inteligencji i uczenia maszynowego.](#)
- [Artykuł z konferencji IEEE EUROCON poświęcony inżynierii elektrycznej oraz systemom sztucznej inteligencji.](#)
- [Publikacja naukowa w czasopiśmie Energies dotycząca optymalizacji systemów energetycznych z wykorzystaniem metod uczenia maszynowego.](#)

AI Index 2026: technologia szybsza niż systemy, które ją oswajają

dr Maciej Tąkiel

W kwietniu 2026 r. Stanford Institute for Human-Centered Artificial Intelligence opublikował dziewiątą edycję AI Index Report. Sami autorzy formułują wprost tezę przewodnią: technologia skaluje się szybciej, aniżeli systemy mające nią zarządzać. Spójrzmy na powyższe zjawisko z polskiej oraz europejskiej perspektywy regulacyjnej.

Ponad 50% adopcji w trzy lata

Generatywna sztuczna inteligencja ma już za sobą 53% adopcji populacyjnej w okresie trzech lat. Tempo to przewyższa upowszechnienie się komputera osobistego oraz internetu. Korzysta z niej 88% organizacji, a wśród studentów uniwersyteckich aż czterech na pięciu. Inaczej rysuje się sytuacja po stronie wytwarzania samych modeli. W 2025 r. powstało ich w USA łącznie 59, w Chinach 35, natomiast w Europie zaledwie dwa, francuski i brytyjski. Z sektora prywatnego pochodzi ponad 91% wszystkich modeli, akademii zaś przypada zaledwie 2%. Adopcja jest zatem globalna, lecz za samą technologię odpowiadają w obecnej chwili nominalnie dwa państwa.

Co coraz trudniej zmierzyć

Najwięcej do myślenia daje to, czego raport skwantyfikować już nie potrafi. Wskaźnik transparentności wiodących modeli (Foundation Model Transparency Index, badający dane treningowe, zasoby obliczeniowe i wpływ po wdrożeniu) w ciągu roku obniżył się z 58 do 40 punktów. Nie mały regres, w czasie gdy OpenAI, Anthropic oraz Google praktycznie zaprzestali publikowania metadanych swoich czołowych modeli. Towarzyszy temu wzrost udokumentowanych incydentów AI, z 233 w 2024 r. do 362 w 2025 r. Niezależnie od powyższego, same testy zdolności modeli również tracą swą wiarygodność. Wzorcowy test SWE-bench Verified w ciągu roku przeszedł drogę od 60% do niemal 100% dopasowania do ludzkiego punktu odniesienia. Halucynacje wiodących modeli mieszczą się w przedziale od 22% do 94% dla 26 testowanych systemów, GPT-4o zaś w teście oddzielania faktów od fałszywych przekonań pierwszoosobowych użytkownika spadł z 98,2% do 64,4%. Modele radzą sobie z testami wręcz lepiej aniżeli z obecną rzeczywistością.

Świat nadaje ton, Europa się przygląda

Reakcje regulacyjne roku 2025 nie poszły w jednym kierunku. Pierwsze zakazy z unijnego AI Act weszły w życie 2 lutego, podczas gdy administracja amerykańska niespełna dwa tygodnie wcześniej wycofała własne dotychczasowe dyrektywy AI w drodze rozporządzenia prezydenckiego. GDPR pozostaje najczęściej cytowaną ramą regulacyjną (60%, spadek z 65%), wszakże w obieg wszedł również ISO/IEC 42001, pierwszy międzynarodowy standard zarządzania AI, cytowany przez 36% organizacji. Niezależnie od powyższego, sondaż Pew z 2025 r. dał Brukseli wyraźną przewagę zaufania publicznego: mediana z 25 badanych krajów oceniła wiarygodność regulacyjną UE na 53%, USA na 37%, Chin zaś na 27%. Po stronie infrastruktury Europa pozostaje wszelako w tyle: w Europie

i Azji Centralnej powstały do 2025 r. 44 państwowe superkomputerowe klastry AI (wobec 3 w 2018), wśród modeli istotnych powstały jednak wyłącznie dwa, francuski i brytyjski. Aktywność w zakresie regulacji jest nacechowana na adresowanie technologii spoza kontynentu.

Wnioski

Wniosek z tegorocznego AI Indexu jest jeden: tempo rozwoju AI niemal całkowicie wyprzedza próby regulacyjne. Nie jest to wąski problem techniczny, lecz sytuacja systemowa. Konsekwencja praktyczna jest taka, że unijne ramy regulacyjne (AI Act, GDPR, ISO/IEC 42001) przejęły funkcję, której dostawcy AI sami sobie już nie nakładają. Polski kontekst dodaje do powyższego osobny brak: według styczniowego Raportu Strategicznego UODO, 96% badanych organizacji ocenia własne przygotowanie do RODO w warunkach AI jako niewystarczające lub niepewne. Pełne zastosowanie obowiązków AI Act wobec systemów wysokiego ryzyka z Załącznika III z dniem 2 sierpnia 2026 r. będzie zatem pierwszym praktycznym egzaminem z owej asymetrii. Sam termin wydaje się przesadzony w świetle pakietu Digital Omnibus z dnia 19 listopada 2025 r., a także porozumienia z 7 maja 2026 r., w którym instytucje unijne zaproponowały powiązanie stosowania obowiązków z dostępnością zharmonizowanych norm, z datą graniczną 2 grudnia 2027 r. Czy unijne ramy okażą się wystarczająco precyzyjne, czy też potrzebne będą kolejne nowelizacje - pokaże nam najbliższy rok.

Źródła:

- [Raport Stanford HAI AI Index 2026 stanowiący kompleksowe podsumowanie stanu rozwoju i trendów w sztucznej inteligencji.](#)
- [Oficjalny komunikat Urzędu Ochrony Danych Osobowych \(UODO\) dotyczący ochrony danych w kontekście nowych technologii i AI.](#)
- [Oficjalny tekst Aktu o sztucznej inteligencji \(EU AI Act – Rozporządzenie Parlamentu Europejskiego i Rady 2024/1689\) w bazie EUR-Lex.](#)

Czy można zaufać rozumowaniu AI? - Traceability modeli, chain of thought i nowe wyzwania dla AI Governance

Marzena Anczurowska-Praniuk

Przez długi czas organizacje oceniały modele AI głównie przez jakość odpowiedzi. Jeśli wynik był poprawny, system uznawano za skuteczny. Dziś to przestaje wystarczać.

Coraz częściej analizujemy już nie tylko odpowiedź modelu, ale również sposób, w jaki dochodzi on do takiej, a nie innej decyzji. W obszarze AI Governance, czyli zasad zarządzania, nadzoru i odpowiedzialności związanych z wykorzystaniem sztucznej inteligencji, rośnie znaczenie takich pojęć jak traceability, audytowalność czy chain of thought.

Traceability można najprościej opisać jako możliwość odtworzenia, w jaki sposób powstał dany wynik modelu: od danych i procesu uczenia, przez działanie systemu, aż po wykorzystanie odpowiedzi przez człowieka lub organizację. W teorii ma to zwiększać przejrzystość i zaufanie do systemów AI (co jest szczególnie istotne dla branż regulowanych, jak ochrona zdrowia, automotive, bankowość).

Problem polega jednak na tym, że logicznie wyglądające uzasadnienie nie zawsze musi odzwierciedlać rzeczywisty mechanizm podjęcia decyzji przez model.

Coraz więcej badań (link podany w źródłach) pokazuje zjawisko określane jako unfaithful chain of thought (pozorny tok rozumowania, który nie odzwierciedla rzeczywistych przesłanek prowadzących model do odpowiedzi). Model potrafi wygenerować przekonujące rozumowanie krok po kroku, mimo że faktyczna decyzja została wcześniej ukierunkowana przez kontekst, sugestię lub inny sygnał obecny w danych wejściowych. Innymi słowy: AI może tworzyć racjonalizację po fakcie, która brzmi wiarygodnie dla człowieka.

Audytowalność decyzji sztucznej inteligencji

To zmienia sposób myślenia o audycie AI. Jeszcze niedawno wiele organizacji zakładało, że możliwość podglądu rozumowania automatycznie zwiększy transparentność modeli. Dziś coraz częściej pojawia się pytanie, czy audytor rzeczywiście widzi proces decyzyjny systemu, czy jedynie narrację wygenerowaną na potrzeby użytkownika.

Największe ryzyko nie zawsze polega bowiem na tym, że model wygeneruje błędną odpowiedź. Znacznie groźniejsza może być sytuacja, w której AI wygeneruje odpowiedź błędną, ale uzasadnioną w sposób profesjonalny, logiczny i przekonujący.

Przekonałam się o tym już na początku 2024 roku podczas analiz regulacji AI realizowanych w ramach przygotowania rekomendacji dla Ministerstwa Cyfryzacji dotyczących organów notyfikacji AI. Narzędzia generatywne potrafiły tworzyć bardzo przekonujące informacje na temat aktów prawnych i regulacji z różnych państw. Część z nich po dokładnej weryfikacji okazywała się jednak błędna, nieaktualna albo trudna do potwierdzenia w oficjalnych źródłach.

Podobne problemy zaczynają pojawiać się również w praktyce. W jednym z głośnych orzeczeń Krajowej Izby Odwoławczej z 2025 roku analizowano przypadek wykorzystania niezwyfikowanych informacji wygenerowanych przez AI w postępowaniu przetargowym. W dokumentacji pojawiły się m.in. nieistniejące interpretacje podatkowe. Sprawa pokazała, że nawet najbardziej przekonująca odpowiedź modelu nie zwalnia organizacji z obowiązku weryfikacji informacji i zachowania należytej staranności.

Nieprzypadkowo pojawiają się nowe standardy, takie jak AIUC-1, dotyczące bezpieczeństwa, ochrony i niezawodności agentów AI. Równolegle rozwija się rynek usług związanych z oceną ryzyka, audytem i ubezpieczaniem systemów AI. To sygnał, że dyskusja przesuwana się z pytania: Czy AI działa? - na pytanie: Jak odpowiedzialnie zarządzać ryzykiem związanym z jej wykorzystaniem?

Trzy rekomendacje dla organizacji i zarządów

- **Traktuj AI jako źródło hipotez, nie źródło prawdy**

Im większe konsekwencje decyzji, tym większy powinien być poziom niezależnej weryfikacji odpowiedzi modelu.

- **Buduj kompetencje eksperckie równolegle z wdrażaniem AI**

Najszybciej błędy modeli wychwytywać osoby posiadające wiedzę merytoryczną w danej dziedzinie. AI nie eliminuje potrzeby eksperckości lecz zwiększa jej znaczenie.

- **Dokumentuj proces podejmowania decyzji wspieranych przez AI**

W środowisku regulowanym coraz ważniejsze staje się nie tylko to, jaka decyzja została podjęta, ale również kto ją zatwierdził, na podstawie jakich informacji i w jaki sposób zweryfikowano rekomendację modelu.

Dostęp do zaawansowanych modeli AI staje się coraz bardziej powszechny. Coraz większe znaczenie zyskuje natomiast zdolność organizacji do odpowiedzialnego nadzorowania procesów decyzyjnych wspieranych przez sztuczną inteligencję.

Źródła:

- [Strona główna AIUC-1 dedykowana standardom benchmarkingu i certyfikacji modeli sztucznej inteligencji.](#)
- [Artykuł naukowy w serwisie arXiv dotyczący ewaluacji i bezpieczeństwa złożonych modeli sztucznej inteligencji.](#)
- [Wyrok Krajowej Izby Odwoławczej \(KIO\) w sprawie zamówień publicznych dostępny w bazie Urzędu Zamówień Publicznych.](#)

Dokumentowanie realnego nadzoru człowieka nad systemami AI

Tomasz Fronczyk

W dyskusji o odpowiedzialnym wdrażaniu sztucznej inteligencji często pojawia się pojęcie nadzoru człowieka. Brzmi ono uspokajająco: system AI może wspierać analizę danych, porządkować informacje albo wskazywać możliwe kierunki działania, a ostateczna kontrola ma pozostawać po stronie człowieka. Problem zaczyna się wtedy, gdy brakuje informacji, kto faktycznie sprawował ten nadzór, kiedy interweniował, jakie informacje zweryfikował oraz czy miał realną możliwość zakwestionowania wyniku działania systemu.

Samo stwierdzenie, że człowiek pozostaje w pętli decyzyjnej, jest niewystarczające. W praktyce administracyjnej, regulacyjnej i audytowej znaczenie ma zarówno istnienie procedury, jak i możliwość jej odtworzenia. Jeżeli system AI wygenerował rekomendację, a pracownik instytucji ją zaakceptował, powinno być możliwe ustalenie, czy była to świadoma ocena, czy jedynie formalne potwierdzenie wyniku. Bez takiego śladu nadzór człowieka może stać się pozornym zabezpieczeniem, które ogranicza odpowiedzialność jedynie na poziomie deklaracji.

AI Act wzmacnia znaczenie nadzoru człowieka w odniesieniu do systemów wysokiego ryzyka. Zgodnie z art. 14 rozporządzenia nadzór ten ma służyć zapobieganiu lub minimalizowaniu ryzyk dla zdrowia, bezpieczeństwa i praw podstawowych. Jednocześnie art. 12 przewiduje wymogi dotyczące rejestrowania zdarzeń w systemach wysokiego ryzyka, tak aby możliwe było odtworzenie działania systemu w czasie jego cyklu życia. Oznacza to, że nadzór człowieka i dokumentowanie działania systemu należy traktować łącznie. Dopiero ich połączenie tworzy podstawę realnej rozliczalności.

W administracji publicznej ma to szczególne znaczenie. Projekt „Polityki rozwoju sztucznej inteligencji w Polsce do 2030 roku” zakłada rozwój narzędzi AI w administracji, w tym inteligentnych systemów wsparcia decyzji, chatbotów, rozszerzonych baz wiedzy oraz pilotażowe wdrożenia polskiego modelu językowego PLLuM w podmiotach publicznych. W tym samym kierunku zmierza założenie tworzenia administracji bardziej efektywnej, dostępnej i zdolnej do sprawniejszego reagowania na potrzeby obywateli. Im szerzej AI będzie wykorzystywana w usługach publicznych, tym ważniejsze stanie się rozróżnienie między formalnym nadzorem a rzeczywistą kontrolą człowieka nad procesem.

Dlatego przy wdrożeniach AI warto przyjąć prostą zasadę: każda istotna interwencja człowieka powinna pozostawiać minimalny ślad dokumentacyjny. Taki mechanizm może być prosty i proporcjonalny do ryzyka. W wielu przypadkach wystarczy zapis: kto dokonał weryfikacji, jaki wynik systemu oceniał, czy zaakceptował rekomendację, czy ją zmienił, z jakiego powodu oraz na podstawie jakich informacji. Inne wymagania będą dotyczyły narzędzia porządkującego korespondencję, a inne systemu wspierającego działania w ochronie zdrowia, świadczeniach społecznych, edukacji, migracji lub wymiarze sprawiedliwości.

Podobną logikę wzmacniają międzynarodowe ramy zarządzania AI. OECD wskazuje, że godna zaufania sztuczna inteligencja powinna respektować prawa człowieka, wartości demokratyczne, przejrzystość, możliwość wyjaśnienia sposobu działania systemu, bezpieczeństwo i odpowiedzialność. NIST AI Risk Management Framework traktuje zarządzanie ryzykiem AI jako proces obejmujący cały cykl życia systemu, a ISO/IEC 42001 opisuje system zarządzania AI jako mechanizm utrzymywania i doskonalenia odpowiedzialnego wykorzystania sztucznej inteligencji w organizacji.

Ostatecznie nadzór człowieka nad AI ma znaczenie tylko wtedy, gdy pozostawia po sobie możliwy do sprawdzenia ślad. Jasno przypisana odpowiedzialność, określone punkty kontroli, możliwość zatrzymania lub korekty działania systemu oraz dokumentacja przebiegu sprawy decydują o tym, czy „człowiek w pętli” jest realnym zabezpieczeniem. Bez tych elementów nadzór pozostaje elegancką formułą, która dobrze wygląda w procedurze, lecz nie daje rzeczywistej gwarancji odpowiedzialnego wdrożenia.

Źródła:

- [Tekst europejskiego rozporządzenia w sprawie sztucznej inteligencji \(EU AI Act\) na portalu EUR-Lex.](#)
- [Interaktywne narzędzie AI Act Explorer udostępnione przez Komisję Europejską do analizy przepisów aktu o AI.](#)
- [Zbiór rekomendacji i zasad OECD dotyczących odpowiedzialnego rozwoju i wykorzystania sztucznej inteligencji.](#)
- [Portal NIST dedykowany Ramom Zarządzania Ryzykiem Sztucznej Inteligencji \(AI RMF\)](#)
- [Oficjalna strona standardu ISO/IEC 42001 określającego system zarządzania sztuczną inteligencją w organizacjach.](#)
- [Oficjalny dokument rządowy przedstawiający „Politykę rozwoju sztucznej inteligencji w Polsce do roku 2030”.](#)

Oznaczanie treści AI: czy techniczne watermarki wystarczą do walki z dezinformacją?

Tomasz Fronczyk

Rozwój generatywnej sztucznej inteligencji znacząco obniżył koszt tworzenia realistycznych obrazów, nagrań głosowych, filmów i tekstów. W efekcie dezinformacja nie musi już opierać się na prostych fałszywych komunikatach. Coraz częściej może przyjmować formę wiarygodnie wyglądających materiałów multimedialnych, które trudno ocenić bez dodatkowego kontekstu. Z tego powodu oznaczanie treści tworzonych lub modyfikowanych przez AI staje się jednym z istotnych tematów debaty publicznej.

8 maja 2026 r. Komisja Europejska opublikowała trzy badania dotyczące technicznych metod oznaczania i wykrywania treści generowanych przez AI. Opracowania obejmują tekst, audio oraz obraz i wideo. Ich zadaniem jest wsparcie prac nad Kodeksem praktyk dotyczącym oznaczania i etykietowania treści AI, przygotowywanym w kontekście obowiązków przejrzystości z art. 50 AI Act. Komisja wskazuje, że badania analizują zarówno obecne, jak i rozwijane techniki, ich skuteczność, ograniczenia oraz praktyczną przydatność w różnych typach treści.

Najczęściej omawianym rozwiązaniem są watermarki, czyli cyfrowe znaczniki pozwalające rozpoznać, że materiał został wygenerowany albo istotnie zmodyfikowany przez system sztucznej inteligencji. Mogą być widoczne dla użytkownika lub zapisane w sposób możliwy do odczytu maszynowego, na przykład w metadanych. W założeniu mają ułatwiać identyfikację treści syntetycznych i wspierać automatyczne systemy wykrywania.

Nie należy jednak traktować watermarków jako samodzielnej odpowiedzi na dezinformację. Tekst można sparafrazować, nagranie poddać kompresji, obraz przyciąć, a film ponownie nagrać z ekranu. Takie działania mogą osłabić albo usunąć techniczny znacznik. Równie istotny jest problem skali: jeżeli oznaczanie będzie stosowane tylko przez część dostawców, brak znacznika nie będzie oznaczał, że materiał jest autentyczny. W praktyce szczególnie trudne będą sytuacje, w których prawdziwe materiały zostaną częściowo zmienione przez AI, na przykład przez podmianę głosu, dopisanie fragmentu wypowiedzi albo zmianę kontekstu obrazu. Taki materiał nie jest w pełni syntetyczny, ale może wprowadzać odbiorcę w błąd równie skutecznie jak całkowicie fałszywe nagranie.

Dlatego potrzebne jest podejście wielowarstwowe. Oprócz watermarków znaczenie mają metadane, podpisy kryptograficzne, mechanizmy potwierdzania pochodzenia treści, narzędzia detekcji, procedury platform internetowych oraz edukacja medialna użytkowników. Komisja Europejska wskazuje, że techniczne rozwiązania powinny być możliwie niezawodne, interoperacyjne, skuteczne i odporne, przy uwzględnieniu specyfiki różnych formatów treści. Ważne jest także, aby oznaczenia były zrozumiałe dla odbiorców. Etykieta techniczna nie spełni swojej funkcji, jeżeli użytkownik nie będzie wiedział, czy oznacza ona pełne wygenerowanie treści przez AI, częściową edycję, czy jedynie użycie narzędzia pomocniczego.

Z perspektywy administracji publicznej kluczowe jest zachowanie proporcji. Oznaczanie treści AI może zwiększać przejrzystość obiegu informacji, ale nie zastąpi odpowiedzialnej komunikacji instytucji, fact-checkingu, szybkiego reagowania na kampanie manipulacyjne oraz wzmacniania kompetencji cyfrowych obywateli. Szczególne znaczenie ma tu komunikacja kryzysowa, wybory, zdrowie publiczne i bezpieczeństwo, czyli obszary, w których fałszywy materiał może szybko wywołać skutki społeczne.

Techniczne watermarki są potrzebne, ale nie stanowią wystarczającej odpowiedzi na problem dezinformacji. Ich wartość będzie zależeć od tego, czy zostaną połączone z regulacjami, standardami interoperacyjności, odpowiedzialnością platform oraz edukacją społeczną. Dopiero takie podejście może przekształcić oznaczanie treści AI z formalnego mechanizmu w realne narzędzie wzmacniania zaufania do informacji w środowisku cyfrowym.

Źródła:

- [Raport Komisji Europejskiej prezentujący trzy opracowania na temat rozwiązań technicznych do znakowania i wykrywania treści generowanych przez AI.](#)
- [Europejski Kodeks Postępowania w zakresie identyfikacji oraz przejrzystości treści tworzonych przez sztuczną inteligencję.](#)
- [Projekt wytycznych KE dotyczących wdrażania obowiązków przejrzystości dla systemów AI zgodnie z art. 50 Aktu o AI.](#)
- [Pełna dokumentacja prawna Aktu o sztucznej inteligencji \(UE 2024/1689\) w serwisie EUR-Lex.](#)
- [Portal Komisji Europejskiej opisujący unijne ramy regulacyjne i strategiczne podejście do sztucznej inteligencji.](#)

AI & Rynek Pracy – Fakty, Trendy, Prognozy

Anna Nieć-Mrzygłód

Współczesna analiza wpływu sztucznej inteligencji na rynek pracy wskazuje na istotną ambiwalencję pomiędzy prognozowanym wzrostem produktywności a potencjalnym ryzykiem strukturalnych zmian zatrudnienia. W raporcie McKinsey Global Institute z 2023 roku wskazano, że potencjał technologiczny generatywnej AI i innych technologii może objąć automatyzacją aktywności odpowiadające za około 60–70% czasu pracy. Autorzy podkreślają jednak, że potencjał automatyzacji poszczególnych zadań nie oznacza automatyzacji całych stanowisk pracy, lecz może prowadzić do istotnej zmiany struktury wykonywanych obowiązków. Jednocześnie badania McKinsey wskazują, że generatywna AI zwiększa potencjał automatyzacji również w zawodach wymagających kwalifikacji specjalistycznych, obejmujących m.in. analizę dokumentów prawnych, analizę danych, programowanie czy działalność badawczo-rozwojową. Jednocześnie autorzy raportu sugerują, że obszary wymagające wysokiego poziomu inteligencji emocjonalnej, takie jak edukacja czy opieka zdrowotna, pozostaną mniej podatne na pełną automatyzację ze względu na znaczenie kompetencji interpersonalnych, choć AI będzie stopniowo wspierać wykonywanie wielu zadań również w tych sektorach. Z perspektywy ekonomicznej Goldman Sachs wskazywał, że generatywna AI może w gospodarkach rozwiniętych objąć znaczącą część wykonywanych zadań, przy czym szacunki te dotyczą przede wszystkim ekspozycji zadań na automatyzację, a nie bezpośredniego zaniku całych miejsc pracy. W analizach Goldman Sachs przyjmowano, że około jedna czwarta obecnej pracy w gospodarkach rozwiniętych może potencjalnie zostać zastąpiona przez AI, przy jednoczesnym założeniu powstawania nowych aktywności i wzrostu produktywności. W klasycznym badaniu Carla Benedikta Freya i Michaela Osborne'a z 2013 roku oszacowano, że około 47% miejsc pracy w USA znajdowało się wówczas w grupie wysokiego ryzyka komputeryzacji. Wynik ten należy jednak traktować jako historyczny punkt odniesienia, ponieważ badanie poprzedzało rozwój współczesnej generatywnej AI, a późniejsze analizy przyjmowały znacznie bardziej zróżnicowane szacunki.

Rozbieżności w ocenach skali automatyzacji są jednak znaczące. Analizy EY wskazują, że poziom potencjalnej automatyzacji różni się między gospodarkami. W przypadku Polski EY szacuje, że w perspektywie kolejnej dekady około 2,6% zadań może zostać poddanych automatyzacji przez narzędzia bazujące na AI, wobec około 5% w Europie Zachodniej i USA oraz 2,2% w Europie Środkowo-Wschodniej. Analizy Office for National Statistics wskazują, że ryzyko automatyzacji różni się znacząco między zawodami i jest najwyższe w przypadku profesji opartych na rutynowych, powtarzalnych zadaniach, natomiast niższe w zawodach wymagających złożonego rozumowania, kreatywności i kompetencji społecznych. W tym samym nurcie Tony Blair Institute for Global Change prognozuje, że pełne wdrożenie sztucznej inteligencji w brytyjskich przedsiębiorstwach mogłoby przynieść oszczędności czasu pracy odpowiadające rocznej pracy około 6 milionów pracowników sektora prywatnego. W kontekście rozwoju nowych ról zawodowych LinkedIn wskazywał, że w 2023 roku liczba publikowanych treści i ofert odnoszących się do generatywnej AI wzrosła 36-krotnie w porównaniu z 2022 rokiem, co odzwierciedlało gwałtowny wzrost zainteresowania kompetencjami

związanymi z generatywną AI na rynku pracy. Wśród nowych profesji wymienia się m.in. Prompt Engineer, AI Ethicist czy AI Governance Specialist. Wskazuje się jednak, że powstające miejsca pracy niekoniecznie zrekompensują redukcję zatrudnienia w zawodach rutynowych. W literaturze przedmiotu podkreśla się również dynamiczny rozwój badań nad wpływem AI na strukturę pracy. W tym kontekście istotna jest publikacja Mariusza Miąsko, który analizuje konsekwencje cyfryzacji prawa pracy w kontekście wdrażania systemów sztucznej inteligencji oraz cyfrowej inteligencji w procesach legislacyjnych i organizacyjnych rynku pracy. Autor wskazuje, że rozwój AI ujawnia fundamentalne problemy adaptacyjne prawa pracy, wynikające z niedostosowania norm prawnych do możliwości technologicznych systemów cyfrowych. Dodatkowo badania przeprowadzone w ramach benchmarku SWE-Lancer wskazują, że współczesne modele językowe nadal nie są w stanie skutecznie rozwiązać większości rzeczywistych zadań programistycznych wykonywanych przez profesjonalnych programistów, co potwierdza, że mimo szybkiego rozwoju AI pełna automatyzacja pracy programistycznej pozostaje ograniczona. Z kolei badania PARP i SWPS wskazują, że w analizach jakościowych dokumentów AI może osiągać porównywalne lub wyższe wyniki w wybranych zadaniach. Zestawienie powyższych wyników prowadzi do wniosku, że wpływ sztucznej inteligencji na rynek pracy ma charakter asymetryczny i zależy od kontekstu zastosowania technologii. W niektórych obszarach AI już obecnie przewyższa człowieka, podczas gdy w innych pozostaje narzędziem wspierającym. Jednocześnie dynamika rozwoju technologii sugeruje, że obecne ograniczenia mogą mieć charakter przejściowy. W konsekwencji coraz częściej pojawiają się postulaty dotyczące konieczności systemowych interwencji regulacyjnych, w tym propozycje opodatkowania AI w celu stabilizacji systemów społecznych i zabezpieczenia rynku pracy przed gwałtownymi zmianami strukturalnymi.

W odniesieniu do Polski można zaobserwować stopniowe, lecz wyraźne przejście od fazy deklaratywnej do fazy implementacyjnej polityki AI, w której państwo coraz częściej traktuje sztuczną inteligencję nie jako technologię eksperymentalną, lecz jako strategiczny filar transformacji gospodarczej. Kierunek ten wpisuje się w szersze ramy polityki Unii Europejskiej, w szczególności rozporządzenia w sprawie sztucznej inteligencji, które ustanawia podejście oparte na poziomach ryzyka oraz odrębne obowiązki dotyczące m.in. przejrzystości określonych systemów i treści generowanych przez AI, a także szczególne wymogi dotyczące systemów wysokiego ryzyka. Znaczenie tych regulacji dla rynku pracy wzrosło szczególnie w 2026 roku, kiedy rozpoczęło się stosowanie kolejnych przepisów AI Act, w tym obowiązków dotyczących przejrzystości określonych systemów i treści AI. Regulacja przestała być zatem wyłącznie ramą przyszłego zarządzania technologią, stając się elementem bieżącego otoczenia prawnego przedsiębiorstw. W Polsce polityka AI koncentruje się obecnie na trzech równoległych osiach rozwojowych: cyfryzacji administracji publicznej, zwiększaniu kompetencji cyfrowych pracowników oraz wspieraniu innowacji w sektorze prywatnym, w szczególności w obszarze MŚP. Jednocześnie widoczna jest rosnąca świadomość, że AI nie jest wyłącznie narzędziem wzrostu produktywności, lecz także czynnikiem potencjalnej destabilizacji rynku pracy, co prowadzi do coraz częstszych debat dotyczących mechanizmów kompensacyjnych, w tym koncepcji opodatkowania automatyzacji oraz redystrybucji korzyści technologicznych. Kluczowym trendem w polskiej polityce AI jest zatem próba pogodzenia dwóch

sprzecznych celów: przyspieszenia adopcji technologii w celu utrzymania konkurencyjności gospodarki oraz jednoczesnego zabezpieczenia rynku pracy przed gwałtowną dezaktualizacją kompetencji zawodowych. W praktyce oznacza to stopniowe przesuwanie akcentu z ochrony konkretnych zawodów na ochronę kompetencji i zdolności adaptacyjnych pracowników. W dłuższej perspektywie polityka AI w Polsce zmierza więc w kierunku modelu hybrydowego, w którym państwo pełni rolę zarówno regulatora, jak i aktywnego uczestnika transformacji technologicznej. Ostateczny kierunek tej polityki będzie jednak w dużej mierze zależny od tempa wdrażania AI w sektorze prywatnym oraz zdolności systemu edukacji, przedsiębiorstw i instytucji publicznych do adaptacji wobec zmian o charakterze strukturalnym. Z perspektywy rynku pracy kluczowym wyzwaniem nie jest zatem wyłącznie ochrona istniejących stanowisk, lecz rozwój kompetencji umożliwiających pracownikom funkcjonowanie w środowisku, w którym zakres zadań wykonywanych przez człowieka i systemy AI będzie podlegał dalszej rekonfiguracji.

Źródła:

- [Raport McKinsey & Company o potencjale ekonomicznym generatywnej sztucznej inteligencji i nowej fali produktywności.](#)
- [Analiza McKinsey Global Institute dotycząca wpływu generatywnej AI na przyszłość rynku pracy w USA.](#)
- [Raport Goldman Sachs poświęcony globalnemu wpływowi automatyzacji i AI na wzrost gospodarczy.](#)
- [Badanie University of Oxford \(Frey & Osborne\) na temat podatności zawodu na automatyzację \(„The Future of Employment”\).](#)
- [Opracowanie EY analizujące transformację rynku pracy pod wpływem wdrożeń generatywnej AI.](#)
- [Artykuł Brytyjskiego Urzędu Statystycznego \(ONS\) analizujący grupy zawodowe najbardziej narażone na automatyzację.](#)
- [Analiza Tony Blair Institute for Global Change dotycząca wpływu rozwoju sztucznej inteligencji na gospodarkę i zatrudnienie.](#)
- [Raport LinkedIn Economic Graph przedstawiający zmiany w wymaganych umiejętnościach i dynamice rynku pracy w erze AI.](#)
- [Artykuł naukowy opublikowany w polskim czasopiśmie prawniczo-społecznym ejournals analizujący skutki automatyzacji.](#)
- [Artykuł OpenAI prezentujący środowisko SWE-lancer służące do oceny autonomicznych agentów AI w zadaniach programistycznych.](#)
- [Publikacja naukowa na platformie OpenReview dotycząca uczenia maszynowego i algorytmów AI.](#)
- [Opracowanie Tony Blair Institute oceniające makroekonomiczne skutki integracji AI na rynku pracy.](#)

Legislacja AI w Polsce nabiera tempa - ustawa o systemach AI (USSI) już obowiązuje

Urszula Uchmańska

Sprawy nabrały tempa

Wraz z przyjęciem przez Unię Europejską rozporządzenia w sprawie sztucznej inteligencji (AI Act, rozporządzenie 2024/1689) Polska stanęła przed wyzwaniem stworzenia krajowego systemu nadzoru nad sztuczną inteligencją. Po wielu miesiącach prac legislacyjnych cel ten został osiągnięty. Ustawa o systemach sztucznej inteligencji została uchwalona przez Sejm 3 lipca 2026 r., ogłoszona w Dzienniku Ustaw 27 lipca 2026 r. (Dz.U. z 2026 r. poz. 1003) i zasadniczo weszła w życie 11 sierpnia 2026 r. Należy jednak uwzględnić, że przepisy art. 8–18 oraz rozdziałów 3–5, 8 i 9, obejmujące m.in. opinie indywidualne, kontrolę, postępowania w sprawie naruszeń i administracyjne kary pieniężne, wejdą w życie 28 października 2026 r.

Przyjęcie ustawy zbiegło się w czasie z istotnymi zmianami na poziomie unijnym. Rozporządzenie (UE) 2026/1744, określane jako Digital Omnibus, przesunęło termin stosowania przepisów dotyczących samodzielnych systemów AI wysokiego ryzyka wskazanych w załączniku III do 2 grudnia 2027 r., a systemów wysokiego ryzyka związanych z produktami objętymi załącznikiem I – do 2 sierpnia 2028 r. Nie zmienia to jednak faktu, że państwa członkowskie muszą posiadać skuteczny mechanizm egzekwowania przepisów AI Act oraz wyznaczone organy nadzorcze. Polska ustawa odpowiada na tę potrzebę, tworząc kompleksowy system nadzoru, kontroli i wsparcia rozwoju sztucznej inteligencji.

Dla przedsiębiorców oznacza to, że choć część kluczowych obowiązków AI Act została odroczone, to krajowe ramy instytucjonalne już niebawem będą gotowe do rozpoczęcia nadzoru nad rynkiem AI. W praktyce jest to ostatni moment na uporządkowanie procesów związanych z wykorzystaniem sztucznej inteligencji, przeprowadzenie inwentaryzacji systemów AI oraz określenie roli organizacji w łańcuchu dostaw AI.

Co wprowadza ustawa?

Choć nazwa Ustawy wskazuje na regulację dotyczącą systemów AI, to w praktyce tworzy nowy ekosystem administracji publicznej mającej egzekwować przepisy AI Act i Ustawy. Regulacją skupiającą się w przede wszystkim na systemach AI dalej pozostaje AI Act.

Ustawa określa:

- organizację i sposób sprawowania nadzoru nad rynkiem systemów AI oraz modelami AI ogólnego przeznaczenia w zakresie objętym AI Act.
- postępowanie w sprawie naruszenia rozporządzenia AI Act i ustawy.
- warunki i tryb akredytacji oraz notyfikacji jednostek oceniających zgodność.
- działania wspierające rozwój systemów AI.
- zasady nakładania administracyjnych kar pieniężnych za naruszenie AI Act i ustawy.

Ekosystem nowej administracji publicznej w obszarze AI w Polsce

Sercem nowego systemu jest Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji (KRiBSI), która została ustanowiona ustawą o systemach sztucznej inteligencji jako główny organ nadzoru rynku oraz pojedynczy punkt kontaktowy w sprawach związanych z AI Act.

Ma ona rolę głównego krajowego organu nadzoru rynku oraz pojedynczego punktu kontaktowego do relacji z instytucjami unijnymi, w tym z Komisją Europejską. Zamiast rozdzielać kompetencje nadzorcze między wiele istniejących już urzędów, ustawodawca uznał, że powołanie jednego, wysoce wyspecjalizowanego podmiotu będzie bardziej efektywne i zminimalizuje koszty.

Do zadań Komisji będzie należeć nie tylko kontrola i nakładanie kar (w tym wydawanie postanowień i decyzji w sprawach o naruszenia przepisów), wydawanie decyzji, opinii i wyjaśnień na wniosek zainteresowanych podmiotów, ale także wspieranie innowacji, edukacja publiczna, oraz prowadzenie tzw. piaskownic regulacyjnych, pozwalających na testowanie nowych rozwiązań AI w bezpiecznym środowisku.

Obecnie trwają prace nad faktycznym powołaniem do życia KRiBSI i rozpoczęciem jej funkcjonowania (podobnie jak innych wspomnianych poniżej organów i funkcji, które będą funkcjonować w ekosystemie prawnym sztucznej inteligencji w Polsce).

Komisja nie będzie działać w izolacji. Wsparciem organizacyjnym, administracyjno-kancelaryjnym oraz finansowym zajmie się minister właściwy do spraw informatyzacji (Ministerstwo Cyfryzacji). Minister zajmie się również udzielaniem pomocy finansowej w zakresie badań naukowych i prac rozwojowych nad AI (finansowaniem programów z budżetu państwa i środków UE) oraz wydawaniem rekomendacji najlepszych praktyk w zakresie stosowania systemów AI oraz określaniem w drodze zarządzenia i coroczną weryfikacją wymogów dla organu notyfikującego oraz przekazywaniem sprawozdań Komisji Europejskiej. Minister ten będzie pełnił również odrębną funkcję organu notyfikującego, odpowiedzialnego m.in. za autoryzację jednostek oceniających zgodność systemów AI, w czym wesprze go Polskie Centrum Akredytacji. Centrum to zajmie się wspieraniem ministra w opracowywaniu procedur oceny, wyznaczania i notyfikowania jednostek oceniających zgodność, opracowywaniem (i aktualizacją) szczegółowego programu akredytacji takich jednostek, przekazywaniem informacji o wpływających wnioskach o akredytację do celów notyfikacji, a także bieżącym informowaniem ministra o podjętych decyzjach dotyczących udzielonych, zawieszonych, zmienionych lub cofniętych akredytacjach.

Aby proces decyzyjny był przejrzysty i uwzględniał interesy różnych środowisk, przy Komisji powstaje Społeczna Rada ds. Sztucznej Inteligencji. W jej skład wejdą eksperci, naukowcy, reprezentanci izb gospodarczych, związków zawodowych, a także kandydaci zgłaszani przez różnego rodzaju rzeczników praw (np. Rzecznika Praw Obywatelskich, Rzecznika MŚP, Głównego Inspektora Pracy, czy izby gospodarcze) oraz organizacje pozarządowe. Rada będzie pełnić rolę opiniotwórczo-doradczą (w tym wyrażanie opinii i stanowisk na wniosek, przyjmowanie i przekazywanie wniosków o wszczęcie postępowania w sprawach naruszenia przepisów o AI).

AI to technologia przenikająca niemal każdy sektor gospodarki, dlatego Ustawa obliguje Komisję do ścisłej współpracy z szerokim ekosystemem wyspecjalizowanych organów publicznych. W skład tej sieci wchodzi m.in.:

- Urząd Ochrony Konkurencji i Konsumentów (UOKiK) – w zakresie ochrony konsumentów na rynku,
- Urząd Ochrony Danych Osobowych (UODO) – w kwestiach prywatności,
- Komisja Nadzoru Finansowego (KNF) – dla systemów AI wdrażanych w sektorze finansowym,
- Krajowa Rada Radiofonii i Telewizji (KRRiT) – w obszarze treści audio/video generowanych przez AI w mediach,
- Urząd Rejestracji Produktów Leczniczych oraz Główny Inspektor Farmaceutyczny – dla systemów w ochronie zdrowia i wyrobach medycznych,
- Urząd Patentowy RP – w zakresie praw własności intelektualnej,
- Prokurator Generalny – w zakresie spraw związanych z wykorzystaniem systemów sztucznej inteligencji w ściganiu sprawców przestępstw,
- Podmiotami ds. cyberbezpieczeństwa (np. CSIRT) – w sprawach dotyczących bezpieczeństwa narodowego i ścigania przestępstw.

Reprezentanci części z tych instytucji (UOKiK, KNF, KRRiT, UKE) będą posiadać swoich stałych przedstawicieli w samej Komisji, co zagwarantuje uwzględnienie odpowiednich perspektyw sektorowych w procesie decyzyjnym.

Ostatnim elementem ekosystemu jest dedykowany system kontroli sądowej. Odwołania od decyzji Komisji (np. w zakresie zablokowania niebezpiecznego systemu AI lub nałożenia milionowych kar finansowych) będą trafiać do wysoce wyspecjalizowanego Sądu Okręgowego w Warszawie – sądu ochrony konkurencji i konsumentów. Centralizacja orzecznictwa w jednym sądzie ma zapobiec rozproszeniu kompetencji i ułatwić budowanie specjalistycznej wiedzy sędziowskiej w skomplikowanych technologicznie sprawach.

W ekosystemie będą również funkcjonować jednostki notyfikowane, czyli wyspecjalizowane podmioty, które uzyskały oficjalne uprawnienia państwowe do weryfikowania, czy systemy sztucznej inteligencji spełniają wymogi AI Act (tj. jednostki oceniające zgodność). W praktyce oznacza to, że zanim niektóre systemy AI trafią na rynek lub zostaną oddane do użytku, będą musiały przejść audyt i uzyskać potwierdzenie zgodności z przepisami, przeprowadzone właśnie przez taką jednostkę.

Szansa na mniejsze kary niż przewiduje to AI Act

Zgodnie z projektem ustawy, KRiBSI będzie nakładać administracyjne kary pieniężne w oparciu o stawki i warunki określone bezpośrednio w AI Act. Polski ustawodawca przewidział jednak kilka istotnych mechanizmów, które pozwalają na znaczne obniżenie kwoty kary lub zmniejszenie jej bezpośredniej dolegliwości dla ukaranego podmiotu. Oto przewidziane opcje:

Układ w sprawie nadzwyczajnego złagodzenia sankcji

To główny mechanizm premiujący współpracę z Komisją. Może zostać zainicjowany przez sam organ lub na wniosek podmiotu dopuszczającego się naruszenia. Aby móc skorzystać z układu, podmiot musi zaprzestać naruszeń (usunąć stan niezgodności z prawem) oraz rzetelnie ujawnić wszystkie istotne okoliczności sprawy, w tym ewentualnych współsprawców i odpowiednie dowody. Ulgi kwotowe w ramach układu prezentują się następująco:

- Obniżenie od 20% do 70% – w standardowej procedurze układowej, w zależności od stopnia współpracy i podjętych zobowiązań do naprawienia szkody.

- Obniżenie od 30% do 90% – w przypadku, gdy podmiot sam, z własnej inicjatywy zgłosi naruszenie do Komisji (tzw. samodenuncjacja).
- Odstąpienie od kary (100% obniżki) – w nadzwyczajnych okolicznościach, na przykład gdy podmiot całkowicie usunie skutki swojego naruszenia, Komisja może w ogóle odstąpić od nałożenia kary.

Obniżenie kary za zastosowanie się do ostrzeżenia

Jeśli w toku postępowania Komisja wydała "ostrzeżenie" z zaleceniami naprawczymi, a podmiot ukarany wykona te zalecenia w ciągu 3 miesięcy od dnia doręczenia ostatecznej decyzji o karze, Komisja może zmniejszyć wysokość kary finansowej od 10% do 50%. Warunkiem zastosowania tej ulgi jest to, by dane naruszenie nie doprowadziło do śmierci lub poważnego uszczerbku na zdrowiu osób fizycznych.

Zmniejszenie dolegliwości (ulgi w płatnościach)

W sytuacjach, gdy nałożona kara stanowiłaby zbyt duże obciążenie na dany moment, podmiot może złożyć uzasadniony wniosek, powołując się na swój "ważny interes". W takiej sytuacji Komisja może:

- Odroczyć termin uiszczenia kary.
- Rozłożyć karę na raty.

Należy jednak pamiętać, że w obu powyższych przypadkach od nieuiszczonej kwoty będą naliczane odsetki za zwłokę, jednak przy zastosowaniu obniżonej stawki odsetek za zwłokę obowiązującej w Ordynacji podatkowej.

Dodatkowym mechanizmem chroniącym ukaranego przed przedwczesną dolegliwością finansową jest fakt, że samo wniesienie odwołania od decyzji o nałożeniu kary do sądu automatycznie wstrzymuje obowiązek zapłaty kary do czasu prawomocnego rozstrzygnięcia.

Interpretacje indywidualne i wyjaśnienia o charakterze ogólnym

KRiBSI może wydawać opinie indywidualne, które stanowią wiążące stanowisko w konkretnej sprawie zarówno dla samej Komisji, jak i innych organów państwowych. Uzyskuje się je na sformalizowany, płatny wniosek. Zastosowanie się do takiej opinii gwarantuje podmiotowi bezpieczeństwo i chroni go przed negatywnymi konsekwencjami prawnymi.

Ponadto Komisja może wydawać z urzędu lub na wniosek niewiążące wyjaśnienia o charakterze ogólnym. Pełnią one rolę edukacyjną i mają zapewnić jednolite rozumienie skomplikowanych przepisów, opierając się na dorobku orzeczniczym oraz wytycznych unijnych (np. Komisji Europejskiej czy Europejskiej Rady ds. AI). Po zanonimizowaniu danych, zarówno opinie indywidualne, jak i ogólne wyjaśnienia, będą podlegać obowiązkowej publikacji w Biuletynie Informacji Publicznej (BIP).

Maksymalne limity wydatków z budżetu państwa na nowy ekosystem w ciągu najbliższych 10 lat

Końcowo, co ciekawe, w projekcie Ustawy precyzyjnie określono także maksymalne limity wydatków z budżetu państwa, które będą bezpośrednim skutkiem finansowym utworzenia i funkcjonowania nowego ekosystemu w ciągu najbliższych 10 lat. Limit na pierwszy rok obowiązywania regulacji (2026 r.) ustalono na poziomie 9,30 mln zł, po czym będzie on sukcesywnie wzrastać, osiągając kwotę 29,68 mln zł docelowo w 2035 r. Aby zapewnić bezpieczne i racjonalne gospodarowanie środkami publicznymi, ustawodawca przewidział tzw. mechanizm korygujący na wypadek ryzyka przekroczenia

założonych kwot w danym roku budżetowym. Mechanizm ten automatycznie wymusza redukcję kosztów, co w praktyce będzie oznaczać konieczność weryfikacji, priorytetyzacji lub przesunięcia w czasie realizacji niektórych zadań Komisji. Nadzór nad bieżącym wykorzystaniem limitów sprawować będzie minister właściwy do spraw informatyzacji (Minister Cyfryzacji), natomiast za bezpośrednie wdrożenie i egzekwowanie działań korygujących odpowiadać ma Przewodniczący Komisji.

Co to oznacza dla przedsiębiorców jesienią 2026 r.?

Mimo przesunięcia części obowiązków AI Act w ramach pakietu Digital Omnibus, organizacje nie powinny traktować tego jako powodu do odkładania przygotowań. Obowiązują już przepisy dotyczące systemów AI zakazanych, wymogów AI literacy, regulacji dotyczących modeli AI ogólnego przeznaczenia oraz transparentności treści AI. Polska USSI już obowiązuje, a cały proces ustanowienia krajowego aparatu egzekwującego przepisy AI Act sprawnie już się toczy. W tym momencie trwa intensywna budowa krajowego systemu nadzoru, który w najbliższych miesiącach zacznie realizować zadania przewidziane w ustawie, w tym będzie mógł rozpocząć nakładanie kar. W praktyce oznacza to, że organizacje powinny już dziś przygotować inwentaryzację wykorzystywanych systemów AI, określić swoją rolę względem poszczególnych rozwiązań (dostawca, wdrażający, importer, dystrybutor lub użytkownik), zweryfikować obowiązki wynikające z AI Act oraz rozpocząć dostosowywanie polityk, procedur i umów do nowego otoczenia regulacyjnego.

[Ustawa z dnia 3 lipca 2026 r. o systemach sztucznej inteligencji](#)

Cyberobrona państwa oparta na agentach AI: czego Polska może nauczyć się z brytyjskiego projektu Cyber Shield i strategii cyfrowej NATO

David Rodenko

Polska strategia cyberbezpieczeństwa i strategia AI dla obronności nie opisują koordynacji agentów AI między administracją, wojskiem i operatorami infrastruktury krytycznej. To brak, ponieważ nie jest to przecież zwykły chatbot. Może zaplanować kilka kroków, użyć narzędzi, ocenić skutek i zmienić działanie. Dla cyberbezpieczeństwa oznacza to na przykład wykrywanie podatności oraz pomoc w rozpoznawaniu i ograniczaniu incydentów. Five Eyes ostrzega przed takim ryzykiem. Przyjęte założenia mogą zestarzeć się w ciągu miesięcy. Publikacja NCSC z lipca 2026 r. zaznacza jednak, że w rzeczywistych systemach nie zaobserwowano jeszcze w pełni autonomicznego ataku obejmującego cały cykl włamania. Cyber Shield trzeba więc czytać jako brytyjski plan przygotowania obrony, nie jako gotową tarczę.

Punkt wyjścia Brytyjczyków jest prosty. Wiele ataków odnosi sukces z powodu niezataczanych luk i przestarzałych systemów, a także ze słabej kontroli dostępu. AI pogłębia ten problem, bo automatyzacja przyspiesza wyszukiwanie podatności. Czynności, które wcześniej zajmowały tygodnie, mogą dziś trwać minuty, a obrona ma mniej czasu na zauważenie włamania. NCSC zaleca, by nie zwlekać z usuwaniem luk i ograniczać zależność od przestarzałego oprogramowania. Wobec tej zmiany NCSC zaleca systemy, w których bezpieczeństwo uwzględnia się już na etapie projektowania, zgodnie z zasadą *secure-by-design*. Ten fundament ma iść w parze z wdrażaniem AI.

Na początek AI miałyby wykrywać odsłonięte podatności i pomagać w rozpoznaniu oraz ograniczeniu incydentu. Opisując brytyjski model, NCSC dzieli role między agentów „red” i „blue”. Agent red ma zachowywać się jak kontrolowany tester i szukać słabości, zanim znajdzie je napastnik; agent blue ma natomiast obserwować sygnały ostrzegawcze i pomagać zatrzymać incydent. Jak wyglądałoby to w praktyce? Pierwszy wykrywa źle zabezpieczony serwer. Drugi sprawdza, czy ktoś próbuje go wykorzystać, a decyzję o dalszym kroku podejmuje człowiek. Sam projekt zakłada też federację agentów, czyli współpracę agentów różnych organizacji, z których każda zachowuje kontrolę nad swoimi agentami; każdy agent pozostaje pod władzą właściciela, ale może bezpiecznie rozpoznać innego agenta, wymienić z nim sygnał i wspólnie zareagować. Projekt rozwijałby się tak.

Brytyjczycy wyróżniają tu sześć potrzebnych zdolności. Pierwsza to niezawodna i zrozumiała AI. Druga to federacja agentów, którzy potrafią się wzajemnie identyfikować, ufać sobie i komunikować. Kolejne dwie odpowiadają na pytanie, co mają robić: automatycznie znajdować podatności, pomagać je usuwać oraz dzielić się informacją o incydencie między organizacjami. Dwie ostatnie dotyczą skali kraju: Wielka Brytania chce skanować krytyczne zakresy adresów IP w internecie, aby zobaczyć pełniejszy wspólny obraz podatności, a nie pojedyncze błędy każdej instytucji. Chce też zautomatyzować działania w skali kraju. Chodzi między innymi o blokowanie znanych złośliwych

domen i sieci. Strategia NATO przewiduje również federacyjne platformy, Zero Trust i interoperacyjność.

Przed szerokim wdrożeniem trzeba jednak uwzględnić ryzyko błędnej automatyzacji: obrona może zatrzymać atak, ale też odciąć potrzebną usługę albo zablokować właściwe konto. NCSC chce więc najpierw testować rozwiązania, poprawiać je i dopiero potem skalować. Testy mają zacząć się we współpracy z obrońcami sieci administracji i sektorów krytycznych, co ma przyspieszyć zdobywanie doświadczeń i zwiększyć odporność tam, gdzie przyniesie to największy efekt. Agenci mają wprowadzać znaczące zmiany za zgodą właściciela systemu. Celem jest przejście od testów do rozwiązań, które można wdrażać komercyjnie na większą skalę. Założenia NCSC obejmują przy tym współpracę rządu, operatorów, firm technologicznych i uczelni.

Wymianie informacji w Polsce służą już PCOC, sektorowe CSIRT-y i system S46. Strategia cyberbezpieczeństwa z marca 2026 r. przewiduje również budowę Krajowego Cyber Hubu na bazie S46 i n6. Te rozwiązania mogłyby stanowić punkt wyjścia do prac nad współpracą agentów. Dokument MON również akcentuje testowanie, skalowanie AI i odpowiedzialność człowieka. Przykład Wielkiej Brytanii pozwala więc postawić prostsze pytanie niż „czy budować polski Cyber Shield?”. Należałoby zapytać, czy nasze instytucje potrafią najpierw wspólnie wykryć tę samą lukę, wskazać osobę, która ma podjąć decyzję, a następnie przetestować rezultat. Ograniczony pilotaż mógłby łączyć skanowanie podatności z propozycją blokady znanej złośliwej domeny, którą zatwierdza operator. To ważne, bo agenci otrzymują dostęp do danych i narzędzi, co tworzy nowe zagrożenia, takie jak prompt injection lub podszywanie się pod innego agenta. W badaniu Eurostatu za 2023 r. konsekwencje incydentów ICT zgłosiło 32,47% polskich przedsiębiorstw z badanych sektorów, z co najmniej 10 osobami pracującymi, wobec 21,54% w UE. Proponowany pilotaż na modelu brytyjskim mógłby przygotować nas na skuteczniejsze wdrożenie rozwiązań jako frontowe państwo NATO.

- [Uchwała nr 92 Rady Ministrów z dnia 10 marca 2026 r. w sprawie Strategii Cyberbezpieczeństwa Rzeczypospolitej Polskiej](#)
- [Dokument Ministerstwa Obrony Narodowej określający resortową strategię rozwoju i zastosowania sztucznej inteligencji do 2039 roku.](#)
- [Publikacja Geneva Centre for Security Policy \(GCSP\) dotycząca militarnych i geopolitycznych konsekwencji rozwoju autonomicznych agentów AI.](#)
- [Artykuł Brytyjskiego Narodowego Centrum Cyberbezpieczeństwa \(NCSC\) omówiający zastosowanie agentów AI w obronie cybernetycznej.](#)
- [Komunikat NCSC wskazujący na ewolucję zagrożeń cybernetycznych związanych z AI i wytyczne dla kadry kierowniczej.](#)
- [Oficjalny tekst Cyfrowej Strategii Sojuszu Północnoatlantyckiego \(NATO Digital Strategy\).](#)
- [Dokumentacja techniczna Microsoft Learn dotycząca konfiguracji wykluczeń w mechanizmie automatycznego przerywania ataków \(Microsoft Defender XDR\).](#)
- [Baza danych Eurostatu zawierająca wskaźniki i statystyki dotyczące wykorzystania ICT i cyberbezpieczeństwa w przedsiębiorstwach.](#)

- [Raport Eurostatu z serii „Statistics Explained” przedstawiający stan bezpieczeństwa teleinformatycznego w europejskich firmach.](#)

Sekcja ds. współpracy międzynarodowej w oparciu o godną zaufania sztuczną inteligencję

Grupa Robocza ds. Sztucznej Inteligencji

Sekcja działa w ramach Grupy Roboczej ds. Sztucznej Inteligencji (GRAI) przy Ministerstwie Cyfryzacji. Jej głównym celem jest wspieranie międzynarodowej współpracy na rzecz odpowiedzialnego rozwoju AI. Dążymy do budowania pomostów pomiędzy administracją publiczną, biznesem i środowiskiem naukowym, aby tworzyć warunki dla rozwoju sztucznej inteligencji zgodnej z zasadami etyki, cyberbezpieczeństwa oraz praw człowieka.

Pośród naszych kluczowych działań znalazły się:

- **Konsultacje prezentowanych przez Ministerstwo Cyfryzacji polityk** – bierzemy aktywny udział w opiniowaniu dokumentów strategicznych oraz tworzenie spójnych stanowisk w imieniu grupy
- **Katalog ekspercki** to baza specjalistów gotowych wspierać rozwój AI w Polsce, w tym w obszarze współpracy z MŚP i administracją
- **Newsletter specjalistyczny** to regularne publikacje prezentujące analizy, komentarze i wydarzenia związane z AI na poziomie lokalnym

Sekcja organizuje także cykliczne spotkania, które umożliwiają wymianę doświadczeń i integrację środowiska. Dzięki wspólnemu zaangażowaniu tworzymy przestrzeń dla dyskusji, które kształtują przyszłość AI jako technologii transparentnej, godnej zaufania i dostępnej dla wszystkich.

Członkowie i członkinie naszej sekcji działają na zasadach *pro publico bono*, przyczyniając się do dynamicznego rozwoju tej kluczowej dziedziny technologicznej na arenie międzynarodowej. Dołącz do nas i współtwórz przyszłość sztucznej inteligencji!

Osoby zainteresowane dołączeniem do GRAI i naszej sekcji, współtworzenia opracowań eksperckich na zasadach *pro bono* prosimy o przesłanie zgłoszenia na adres mailowy: grai@cyfra.gov.pl.

Dowiedz się więcej [na stronie GRAI](#).