Kodeks etyczny AI – prekursorzy narzucą standard

ROBERT KROPLEWSKI

pełnomocnik Ministra Cyfryzacji ds. Społeczeństwa Informacyjnego

Im bardziej światowy wyścig w rozwoju sztucznej inteligencji przyspiesza, tym większego znaczenia nabiera wypracowanie ram etycznych AI. Ten, kto pierwszy zaproponuje globalny moralny standard, ten z dużym prawdopodobieństwem zdefiniuje reguły dla reszty świata. Czy znajdzie się tam jednak miejsce dla całego spektrum lokalnych specyfik? Jaką rolę może odegrać podejście „godnej zaufania sztucznej inteligencji", rekomendowane przez OECD?

W poszukiwaniu kodeksu etycznego AI

Od 2016 r. obserwujemy znaczne przyspieszenie globalnego wyścigu w badaniach i rozwoju sztucznej inteligencji (AI), zapoczątkowane przez skalowanie pierwszych podejść technicznych do miana wzorca czy standardu światowego. To międzynarodowe konkurowanie dotyczy jednak nie tylko kwestii technicznych czy technologicznych, ale również obszaru etyki AI. Zasadniczo sprowadza się to do pytania o rolę człowieka i jego autonomię wobec automatyki maszyn oraz niewyjaśnialności procesów zachodzących w tzw. „czarnej skrzynce AI".

W świecie wolnego handlu i otwartego internetu, gdzie o przyszłym kształcie rynku w dużej mierze decyduje sukces pierwszego gracza, kwestia który z aktorów o zasięgu globalnym wyznaczy ramy etyczne dla technologii AI, jest szalenie ważna. Istnieje ryzyko, że ten zuniwersalizowany standard szybko stanie się obowiązujący dla pozostałych interesariuszy, często przy pominięciu lokalnych systemów etycznych lub niedoszacowaniu potrzeb związanych z zaufaniem do AI, jej transparentnością, audytowalnością czy rozliczalnością.

> **W świecie wolnego handlu i otwartego internetu, gdzie o przyszłym kształcie rynku w dużej mierze decyduje sukces pierwszego gracza, kwestia który z aktorów o zasięgu globalnym wyznaczy ramy etyczne dla technologii AI, jest szalenie ważna.**

Status quo centrów gospodarczych zostało jednak podważone przez nowych dynamicznych graczy operujących przełomowymi technologiami i podejmujących się ich innowacyjnego zastosowania. To napięcie między starą gospodarką, a nowymi modelami biznesowymi stworzyło zaś szczelinę dla międzynarodowego konsensusu w sprawie wsparcia badań, rozwoju i reguł zastosowań AI, a także ustalenia ram etyki AI, nim wyklaruje się to na drodze konkurencji rynkowej.

Inicjatywa OECD

Pierwszym międzyrządowym standardem dotyczącym sztucznej inteligencji były Rekomendacje OECD w sprawie sztucznej inteligencji (AI), które powstały na wniosek Komitetu w sprawie polityki gospodarki cyfrowej (CDEP) i zostały przyjęte przez Radę OECD 22 maja 2019 r.

Rekomendacje OECD dotyczące AI koncentrują się na tym, w jaki sposób rządy i inne podmioty mogą kształtować zorientowane na człowieka podejście do godnej zaufania sztucznej inteligencji. Pod względem prawnym wyrażają one wspólną aspirację krajów przystępujących do ich wdrożenia.

Rekomendacje mają na celu wspieranie innowacji i zaufania do sztucznej inteligencji poprzez promowanie odpowiedzialnego zarządzania *godną zaufania sztuczną inteligencją* przy jednoczesnym zapewnieniu poszanowania praw człowieka oraz wartości demokratycznych. To jest węzłowa metareguła tych rekomendacji.

W ten sposób po raz pierwszy w treści dokumentu urzędowego uwypuklona została koncepcja *Trustworthy AI*. Warto w tym miejscu zauważyć, że koncepcja „zaufania" jako wartości, mogącej stać się podstawą budowy międzynarodowego konsensusu wyszła także z Polski, i doświadczeń prac polskiej grupy ekspertów, która została zawiązana przy ówczesnym Ministerstwie Cyfryzacji i przyjętego pryncypium ukorzeniania prac etycznych w wartości podstawowej, jaką jest godność ludzka.

Również warta odnotowania jest zmiana podejścia do zarządzania AI, co znalazło wyraz w przyjęciu angielskiego pojęcia *responsible stewardship* czyli zarządzania połączonego z pieczą, co jest szczególne istotne wobec wyzwania zapewnienia nadzoru ludzkiego nad AI.

Rekomendacje uzupełniają już istniejące standardy OECD w obszarach takich jak prywatność, cyberbezpieczeństwo, zarządzanie ryzykiem i odpowiedzialne przywództwo w biznesie, koncentrując się jednak na zagadnieniach związanych ze sztuczną inteligencją. Przełomowość tych rekomendacji polega na tym, że ustanawiają one pierwszy międzynarodowy standard, który jest możliwy do wdrożenia (operacjonalizacji) i jest wystarczająco elastyczny, aby wytrzymać próbę czasu w tej szybko rozwijającej się dziedzinie.

Konsensus definicji AI

W wyniku prac nad rekomendacjami OECD po raz pierwszy udało się osiągnąć międzynarodowy konsens w podejściu do definicji kluczowych pojęć z dziedziny AI takich jak:

- System AI: to system oparty na maszynach, który może, dla danego zestawu zdefiniowanych przez człowieka celów, dokonywać prognoz, rekomendacji lub decyzji mających wpływ na rzeczywistość lub wirtualne środowiska. Systemy AI są zaprojektowane do działania z różnymi poziomami autonomii.
- Cykl życia systemu AI: fazy cyklu życia systemu AI obejmują: i) „projektowanie, dane i modele" – to faza, która jest sekwencją zależną od kontekstu obejmującego planowanie i projektowanie, gromadzenie i przetwarzanie danych, a także budowę modeli; ii) „weryfikacja i walidacja"; iii) „rozmieszczenie"; oraz iv) „działanie i" monitorowanie". Fazy te często odbywają się w sposób iteracyjny i niekoniecznie są sekwencyjne. Decyzja o wycofaniu systemu AI z eksploatacji może nastąpić w dowolnym momencie fazy eksploatacji i monitorowania.
- Wiedza AI: wiedza AI odnosi się do umiejętności i zasobów, takich jak dane, kod, algorytmy, modele, badania, *know-how*, programy szkoleniowe, zarządzanie, procesy i najlepsze praktyki, wymagane do zrozumienia cyklu życia systemu AI i uczestniczenia w nim.
- Aktorzy AI: aktorzy AI to ci, którzy odgrywają aktywną rolę w cyklu życia systemu AI, w tym organizacje i osoby, które wdrażają lub obsługują sztuczną inteligencję.
- Interesariusze: Interesariusze obejmują wszystkie organizacje i osoby zaangażowane lub dotknięte przez systemy sztucznej inteligencji, bezpośrednio lub pośrednio. Podmioty wykorzystujące sztuczną inteligencję stanowią podzbiór interesariuszy.

Zasady odpowiedzialnego zarządzania

Rekomendacje OECD określają również pięć uzupełniających się zasad umożliwiających odpowiedzialne zarządzanie godną zaufania sztuczną inteligencją, i wzywają interesariuszy sztucznej inteligencji do ich promowania i wdrażania. Są nimi:

- wzrost sprzyjający włączeniu społecznemu, zrównoważony rozwój i dobrobyt,
- skoncentrowane na człowieku wartości i uczciwość,
- przejrzystość i wyjaśnialność,
- solidność, odporność i bezpieczeństwo,
- odpowiedzialność.

Co dalej?

Rekomendacje OECD w sprawie sztucznej inteligencji stanowią pierwszy międzyrządowy standard polityki w zakresie sztucznej inteligencji i fundament, na którym można prowadzić dalsze analizy i opracowywać narzędzia wspierające rządy w ich wysiłkach wdrożeniowych.

Co prawda członkami OECD nie są takie kraje jak Chiny czy Rosja, jednak podczas szczytu G20 w Osace, który miał miejsce w czerwcu 2019 r., Rekomendacje OECD zostały zauważone i przyjęte jako referencyjne także przez przywódców państw G20, co pozwoliło nadać dynamiki dla ich promocji także na innych międzynarodowych forach jak UE, UNESCO, Rada Europy i UN. Rekomendacje stały się częścią „Polityki dla rozwoju sztucznej inteligencji w Polsce od roku 2020" oraz stanowiły inspirację m.in. do prac w ramach:

- Komisji Europejskiej, która przygotowała *Przewodnik etyczny dla godnej zaufania sztucznej inteligencji*, R*ekomendacje polityki i inwestycji dla godnej zaufania sztucznej inteligencji* oraz *Listę oceny godnej zaufania sztucznej inteligencji ALTAI*,
- UNESCO, które przyjęło w dniu 24 listopada 2021 r., przy udziale 193 państw, własne *Rekomendacje w zakresie pierwszych globalnych zasady etycznych dla sztucznej inteligencji w zakresie badań, nauki, edukacji i komunikacji*,
- Rady Europy, która na poziomie traktatowym przygotowuje pierwszy wiążący prawnie instrument odpowiadający na wyzwania jakie sztuczna inteligencja tworzy dla praw człowieka, demokracji i praworządności.

Współpraca multilateralna

Również w ramach Unii Europejskiej trwają prace, które mają stworzyć warunki dla wytwarzania i używania na rynku UE godnej zaufania sztucznej inteligencji. Ma się ona stać marką świadczącą o zdolności Unii do wyznaczania trzeciej drogi rozwoju AI obok już istniejących: utylitarnej (USA) i spójności społecznej (Chiny). Trzeba jednak podkreślić, że również te dwa podejścia ulegają modyfikacji, gdyż USA przyjęły nową strategię dotyczącą AI odwołującą się wprost do Rekomendacji OECD i koncepcji godnej zaufania sztucznej inteligencji. Warto też wspomnieć, że przyjęta przez UE i USA wspólna deklaracja z Pittsburga dotycząca rewizji polityki handlowej również odwołuje się do godnej zaufania sztucznej inteligencji i odpowiedzialnego nią zarządzania w ramach łańcuchów wartości. Chiny zaś, wraz z Rosją, zaakceptowały ostateczną treść rekomendacji UNESCO bazujących na Rekomendacjach OECD.

> **Godna zaufania sztuczna inteligencja ma się stać marką świadczącą o zdolności Unii Europejskiej do wyznaczania trzeciej drogi rozwoju AI obok już istniejących: utylitarnej (USA) i spójności społecznej (Chiny).**

Doświadczenie współpracy państw OECD oraz wspomnianych organizacji multilateralnych doprowadziło do wypracowania wspólnego mechanizmu porównawczego aktywności tych organizacji, jakim jest Globalpolicy.AI. Organizacje te są zgodne, że nie konkurują ze sobą w tworzeniu standardów dotyczących AI, ale każda według własnego mandatu nadanego przez państwa członkowskie uzupełnia działania pozostałych organizacji. W ten sposób zapewniona jest spójność podejmowanych przez międzynarodowe gremia wysiłków.

Odpowiedzialne zarządzanie AI wymaga kontekstowego podejścia do rywalizacji międzynarodowej, a jednocześnie skonwertowania potencjału AI dla dobra społeczeństw i ekosystemów gospodarek narodowych przy jednoczesnym wsparciu dla godności człowieka i jego praw podstawowych.

Przy aktualnym stanie techniki AI i stanie podjętych wysiłków politycznych można żywić optymizm, że wykorzystanie nadarzającego się momentu sprawi, że to człowiek zaprogramuje AI i zdobędzie narzędzia, które ochronią go przed prymatem AI.

> **Biorąc pod uwagę aktualny stan techniki AI oraz podjęte wysiłki polityczne, można żywić optymizm, że wykorzystanie nadarzającego się momentu sprawi, że to człowiek zaprogramuje AI i zdobędzie narzędzia, które ochronią go przed prymatem AI.**

Aby ten przełom utrwalić należy nie ustawać w wysiłkach na rzecz przekształcania społeczeństwa informacyjnego w społeczeństwo wiedzy, modelować systemy AI według wypracowanych ram etycznych zapewniających człowiekowi nadzór na systemami AI, (nawet przy pewnych sytuacjach delegowania decyzji przez człowieka maszynie) oraz budować ekosystemy rozwoju AI pozwalające na koordynację zarządzania takimi zasobami, jak dane i biblioteki algorytmów, infrastruktura obliczeniowa czy zwinne finasowanie.

O autorze

Robert Kroplewski – Pełnomocnik Ministra Cyfryzacji ds. Społeczeństwa Informacyjnego, kierownik projektu Platformy Cyfrowej Przemysłu Przyszłości. Radca prawny, ekspert konwergencji nowych technologii i mediów. Właściciel kancelarii kroplewski.com. Animator i współautor „Polityki rozwoju sztucznej inteligencji w Polsce od 2020", Założeń kierunkowych polityki Polski w obszarze „Gospodarka oparta na danych Przemysł+". Inicjator filaru cyfrowego w projekcie Trójmorza. Ekspert OECD, Komisji Europejskiej, UNESCO oraz Rady Europy w zakresie gospodarki cyfrowej, społeczeństwa informacyjnego oraz sztucznej inteligencji. Członek sieci ekspertów sztucznej inteligencji ONE.AI przy OECD oraz ekspert Globalnego Partnerstwa na rzecz Sztucznej Inteligencji. Specjalizuje się w zagadnieniach transferu technologii, społeczeństwa informacyjnego i mediów, ochrony prywatności i prawa do informacji, ochroną konkurencji, własności intelektualnej, zarządzania danymi i etyczną sztuczną inteligencją.

Wydawca

Partnerzy

Pomorski Fundusz Rozwoju sp. z o.o. z siedzibą w Gdańsku
Instytucja Samorządu Województwa Pomorskiego