

Sztuczna inteligencja a prawa człowieka, demokracja i rządy prawa. Uwagi na tle Konwencji nr 225 Rady Europy z 2024 r.

1. Wprowadzenie

W dniu 5 września 2024 r. na sesji Komitetu Ministrów Rady Europy w Wilnie, po kilku latach negocjacji, przyjęto Konwencję ramową w sprawie sztucznej inteligencji i praw człowieka, demokracji i rządów prawa (CETS nr 225; dalej: Konwencja)¹. Zgodnie z jej postanowieniami Sekretarz Generalny Rady został zobowiązany do przekazania poświadczonych kopii Konwencji każdemu państwu członkowskiemu, państwom niebędącym członkami Rady, które brały udział w jej opracowaniu, Unii Europejskiej oraz każdemu innemu państwu zaproszonemu do przystąpienia do tego traktatu. Oznacza to, że Konwencja została uznana za traktat półotwarty, czyli taki, któremu od początku nadano możliwie szeroki zakres podmiotowy – od etapu negocjacji aż po okres jego obowiązywania². W praktyce traktatowej Rady Europy rozwiązanie to ma charakter wyjątkowy. Istotne znaczenie dla prawnej oceny traktatu ma również określenie go jako konwencji ramowej, a zatem takiej, która wyznacza jedynie ogólne ramy regulacyjne i będzie wymagać ich uzupełnienia bardziej szczegółową treścią dopiero w procesie stosowania³.

¹ Zob. *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe Treaty Series – No. 225*, Vilnius, 5 September 2024, ust. 1–2, s. 1–33. Wejście Konwencji w życie wymaga jej ratyfikacji przez pięć państw, w tym co najmniej trzy państwa członkowskie. Do dnia 26 lipca 2026 r. nie zaczęła ona jeszcze obowiązywać; podpisało ją dotąd 21 podmiotów, a tylko jeden ratyfikował (EU). Wraz z Konwencją Komitet Ministrów przyjął zwyczajowo obszerne sprawozdanie wyjaśniające, w którego pkt II zastrzeżono, że nie stanowi ono autentycznej wykładni tekstu Konwencji, jakkolwiek może ułatwić zrozumienie jej postanowień. Wnosząc jednak z praktyki skrupulatnego (punkt po punkcie) przyjmowania sprawozdań wyjaśniających, są one czymś więcej niż tylko ułatwieniem rozumienia tekstów konwencji.

² Do negocjacji zaproszono, a następnie wzięły w nich udział: Unia Europejska i jej organy, Argentyna, Australia, Izrael, Japonia, Kanada, Kostaryka, Meksyk, Peru, Stolica Apostolska, Stany Zjednoczone Ameryki i Urugwaj. W charakterze obserwatorów w pracach tych uczestniczyli ponadto przedstawiciele 68 organizacji społeczeństwa obywatelskiego i organizacji przemysłowych, a także organizacji międzynarodowych, takich jak OBWE, OECD i UNESCO.

³ Dosadną ocenę określenia „ramowa” wyraził G. Alfredsson w odniesieniu do innej konwencji ramowej, nazywając ją mianem „ramy nieskończonego malowania” – zob. *idem*, *A Frame an Incomplete*

Genezy Konwencji należy doszukiwać się w rosnącej świadomości znaczenia rozwoju technologii informatycznych i cyfrowych, w szczególności systemów algorytmicznych i sztucznej inteligencji (SI). Na oznaczenie angielskiego terminu *Artificial Intelligence* przyjął się skrót AI, pomimo że jest on powszechnie stosowany również w odniesieniu do znanej organizacji pozarządowej zajmującej się ochroną praw człowieka – Amnesty International. Podobnie nie ma przeszkód, aby w języku polskim termin „sztuczna inteligencja” oznaczać skrótem SI, mimo zbieżności z symbolem odnoszącym się do przyjętego w 1960 r. metrycznego systemu siedmiu miar⁴.

Problematyka sztucznej inteligencji pojawiła się już we wcześniejszych działaniach różnych organów Rady Europy, w szczególności Komitetu *ad hoc* ds. sztucznej inteligencji (Ad hoc Committee on Artificial Intelligence – CAHAI), który następnie został zastąpiony przez Komitet ds. sztucznej inteligencji (Committee on Artificial Intelligence – CAI). Końcowe opracowanie CAHAI dotyczące wpływu technik cyfrowych, w tym sztucznej inteligencji, potwierdziło perspektywy rozwoju narzędzi informatycznych oraz ich ogromne możliwości w sferze gospodarczego dobrobytu indywidualnego i całych grup społecznych. Jednocześnie wskazano zagrożenia ze strony sztucznej inteligencji, obejmujące jej potencjalnie negatywne skutki dla człowieka i społeczeństwa, m.in. takie jak dyskryminacja, nadzorowanie społeczeństwa, osłabienie ludzkiej sprawczości lub władzy publicznej, zniekształcanie informacji, ingerencja wyborcza, cyfrowe wykluczenie czy szkodliwe działania gospodarcze⁵.

2. Cele i motywy

Z wyjątkowo długiego wstępu (preambuli) Konwencji, obejmującego 18 ustępów, można w szczególności odczytać cele i motywy, którymi kierowały się strony tego traktatu. Odwołując się do postanowień Statutu Rady Europy, podkreślono przede wszystkim dążenie do osiągnięcia większej jedności między jej członkami oraz współdziałania na rzecz poszanowania praw człowieka, demokracji i rządów prawa – sztandarowych wartości Rady Europy, które stanowią jednocześnie warunki przystąpienia do niej⁶.

Painting: Comparison of the Framework Convention for the Protection of National Minorities with International Standards and Monitoring Procedures, „International Journal on Minority and Group Rights” 2000, vol. 7, no. 4, s. 291. Zob. też: K. Drzewicki, *European Systems for the Promotion and Protection of Human Rights* [w:] *International Protection of Human Rights: A Textbook*, eds. C. Krause, M. Scheinin, Turku/Åbo 2009.

⁴ Międzynarodowy Układ Jednostek Miar (*International System of Units – SI*) obejmuje siedem podstawowych jednostek: metr (długość), kilogram (masa), sekunda (czas), amper (prąd elektryczny), kelwin (temperatura), mol (ilość substancji), kandela (intensywność światła). Należy jednak zauważyć, że nie wszystkie jednostki układu SI są powszechnie stosowane w Polsce.

⁵ Council of Europe. *Ad Hoc Committee on Artificial Intelligence (CAHAI). Feasibility Study*, Strasbourg, 17 December 2020, CAHAI(2020)23, s. 2–3, www.coe.int/cahai [dostęp: 28.06.2025].

⁶ Szerzej na temat sporów interpretacyjnych wokół wymogów członkostwa zob. K. Drzewicki, *Triada czy sekstet? Kontrowersje wokół warunków członkostwa w Radzie Europy* [w:] *Konstytucjonalizm polski. Refleksje z okazji jubileuszu 70-lecia urodzin i 45-lecia pracy naukowej Profesora Andrzeja Szmyta*, red. A. Gajda, K. Grajewski, A. Ryteł-Warzocho, P. Uziębło, M.M. Wiszowaty, Gdańsk 2020, s. 490–501.

W preambule zwrócono uwagę na potrzebę zapewnienia, aby potencjał tkwiący w sztucznej inteligencji wspierał dobrobyt jednostek i społeczeństw oraz uczynił nasz świat bardziej produktywnym, innowacyjnym i bezpiecznym w sposób szanujący, chroniący i realizujący wspólne wartości Rady Europy. Twórcy Konwencji podkreślili, że sztuczna inteligencja stwarza bezprecedensowe możliwości ochrony tych wartości, wskazując zarazem, że istnieją poważne ryzyka i niebezpieczeństwa wynikające z pewnych oddziaływań SI, które mogą przyczynić się do dyskryminacji i nierówności, podważania demokracji lub służyć celom represyjnym.

Uwarunkowania te skłoniły zainteresowanych aktorów do ustanowienia, na zasadzie priorytetu, globalnie stosowalnych ram prawnych, które miałyby wyznaczać wspólne zasady ogólne i normy regulujące działania w ramach cyklu życiowego (*lifecycle, cycle de la vie*) sztucznej inteligencji, służące skutecznemu zachowaniu wspólnych wartości i wykorzystaniu zalet SI do ich promocji w sposób sprzyjający odpowiedzialnej innowacyjności.

Inspiracji do negocjacji nad Konwencją dostarczyły również liczne i istotne międzynarodowe instrumenty polityczne oraz prawne dotyczące sztucznej inteligencji. Spośród dziesięciu tego rodzaju dokumentów wymienionych w sprawozdaniu wyjaśniającym na szczególną uwagę zasługują:

- Deklaracja Komitetu Ministrów Rady Europy w sprawie możliwości manipulacyjnych procesów algorytmicznych (Decl(13/02/2019)1), przyjęta 13 lutego 2019 r.;
- Zalecenie Rady OECD w sprawie sztucznej inteligencji (OECD/LEGAL/0449), przyjęte 22 maja 2019 r. (ze zmianami z 2023 r.);
- Zalecenie Komitetu Ministrów Rady Europy dla państw członkowskich w sprawie wpływu systemów algorytmicznych na prawa człowieka (CM/Rec(2020)1), przyjęte 8 kwietnia 2020 r.⁷;
- Zalecenie UNESCO w sprawie etyki sztucznej inteligencji, przyjęte 23 listopada 2021 r.;
- Deklaracja z Reykjavíku – Zjednoczeni wokół naszych wartości (Reykjavík Declaration – United around our values), przyjęta 17 maja 2023 r. podczas 4. Szczytu Szefów Państw i Rządów Rady Europy;
- Wytyczne z Hiroszimy i Kodeks postępowania z Hiroszimy dla organizacji rozwijających zaawansowane systemy sztucznej inteligencji (G7 Hiroshima Process International Guiding Principles for Organisations Developing Advanced AI Systems and Hiroshima Process International Code of Conduct for Organisations Developing Advanced AI Systems), przyjęte 30 października 2023 r.;

⁷ Serię zaleceń i rezolucji uchwaliło również Zgromadzenie Parlamentarne Rady Europy w 2020 r. Dotyczyły one m.in. demokratycznego zarządzania sztuczną inteligencją, zapobiegania dyskryminacji związanej z jej wykorzystaniem oraz zastosowania SI w działalności policyjnej i systemie sprawiedliwości karnej, opiece zdrowotnej, na rynkach pracy i w innych podobnych obszarach. Zob. szerzej: *Explanatory Report...*, s. 2.

- Deklaracja z Bletchley państw uczestniczących w Szczycie Bezpieczeństwa Sztucznej Inteligencji (The Bletchley Declaration by Countries Attending the AI Safety Summit), przyjęta 1–2 listopada 2023 r. (zaktualizowana 13 lutego 2025 r.)⁸;
- rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (Dz. Urz. UE L 2024/1689 z 12.07.2024; dalej: rozporządzenie nr 2024/1689)⁹.

Treść preambuły do Konwencji wskazuje na istotne znaczenie, jakie jej strony przyznają problematyce sztucznej inteligencji, zwłaszcza potrzebie wypracowania prawnie obowiązujących reguł konwencyjnych. Ich celem jest zapewnienie, aby działania podejmowane w ramach cyklu życiowego systemów sztucznej inteligencji, które mogą ingerować w poszanowanie praw człowieka, funkcjonowanie demokracji i przestrzeganie rządów prawa w sektorze prywatnym i publicznym, pozostawały w zgodności z postanowieniami analizowanego traktatu.

Niektóre aspekty przedmiotu oraz założeń Konwencji znalazły się w jej art. 1. W przepisie tym wskazano, że celem Konwencji jest zapewnienie, by działania podejmowane w ramach cyklu życiowego SI były w pełni zgodne z prawami człowieka, demokracją i rządami prawa (ust. 1). W związku z tym na strony Konwencji nałożono obowiązek przyjęcia lub utrzymania właściwych środków ustawodawczych, administracyjnych i innych, służących realizacji postanowień tego traktatu. Środki te mają być stopniowane i zróżnicowane stosownie do powagi i prawdopodobieństwa wystąpienia niesprzyjających skutków dla praw człowieka, demokracji i rządów prawa w trakcie cyklu życia systemów SI. Mogą one obejmować środki specjalne lub horyzontalne, niezależnie od rodzaju zastosowanej technologii (ust. 2). Dla zapewnienia skutecznej implementacji Konwencji przez wszystkie strony ustanowiono również mechanizm kontrolny oraz przewidziano współpracę międzynarodową (ust. 3)¹⁰.

Warto zauważyć, że celem Konwencji nie jest kreowanie nowego prawa człowieka ani ustanawianie nowych zobowiązań w sferze ochrony praw człowieka, a także podważanie istniejących mechanizmów ich ochrony. Jej główne założenie sprowadza się do określenia różnorodnych zobowiązań prawnych, o których mowa w rozdziałach II–VI Konwencji, mających służyć wzmocnieniu obowiązujących standardów

⁸ Deklarację z Bletchley zaakceptowało 29 państw, w tym – niezależnie od swoich członków – Unia Europejska. Już po spotkaniu do grona sygnatariuszy dołączyła Nowa Zelandia. Zadziwiający jest brak obecności Polski w tym oraz kolejnych spotkaniach tej grupy państw, a także przykłady negatywnych ingerencji polskich władz ministerialnych w sferę badań naukowych nad sztuczną inteligencją, co ilustrują zmiany personalne w polskim ośrodku badawczym IDEAS NCBR.

⁹ Akt ten stanowi prawdopodobnie najszerszą międzynarodową regulację dotyczącą sztucznej inteligencji – obejmuje on bowiem 180 ustępów preambuły, 113 artykułów wraz z 58 odnośnikami oraz 13 obszernych załączników.

¹⁰ Zob. szerzej: *Explanatory Report...*, s. 4–5.

i mechanizmów ochrony praw człowieka oraz ułatwić ich stosowanie w obliczu wyzwań stawianych przez sztuczną inteligencję.

Interesujący charakter ma używany w Konwencji termin odnoszący się do sztucznej inteligencji w ramach jej cyklu życiowego. W sprawozdaniu wyjaśniającym wskazano, że termin ten zapewnia kompleksowe podejście do zagadnień związanych z zagrożeniami wynikającymi z rozwoju SI i jej niesprzyjającymi oddziaływaniami na prawa człowieka, demokrację i rządy prawa, obejmując wszystkie stadia działań dotyczących sztucznej inteligencji. W konsekwencji zakres ten może objąć cały cykl życiowy, co oznacza, że Konwencja może być stosowana nie tylko wobec obecnych, lecz także przyszłych zagrożeń w obliczu szybkiego i często nieprzewidywalnego rozwoju technologicznego¹¹.

3. Podstawowe definicje

Problematyka sztucznej inteligencji jest szczególnie trudna do pełnego zrozumienia przez prawników, przede wszystkim z uwagi na konieczność posługiwania się wiedzą i terminologią techniczną. Sztuczna inteligencja to pojęcie wieloznaczne, które w zależności od osoby, branży czy dziedziny może być różnie rozumiane. Warto przypomnieć słowa Alberta Einsteina, zgodnie z którymi „inteligencja bez ambicji jest jak ptak bez skrzydeł”, a „prawdziwą oznaką inteligencji jest nie tyle wiedza, co wyobraźnia”. Z kolei Stephen Hawking trafnie stwierdził, iż „inteligencja jest zdolnością do przystosowania się”¹². W pewnej mierze do kwestii definiowania sztucznej inteligencji może odnosić się ostrzeżenie zawarte w rzymskiej dyrektywie wypracowanej przez pretorów, według której *Omnis definitio periculosa est* (definiowanie wszystkiego jest niebezpieczne). Niemniej jednak na potrzeby regulacji prawnych oraz ich stosowania konieczne jest określenie podstawowych pojęć. Należy do nich m.in. termin „algorytm”, rozumiany jako uporządkowany zbiór instrukcji, które wskazują krok po kroku sposób rozwiązania problemu lub wykonania danego zadania. W informatyce algorytmy stanowią podstawę działania programów komputerowych i systemów, umożliwiając automatyzację skomplikowanych procesów. W przypadku algorytmów sztucznej inteligencji najważniejszą rolę odgrywają dane, które są materiałem wejściowym o charakterze zarówno ilościowym, jak i jakościowym. Algorytmy sztucznej inteligencji można zatem określić jako zestawy instrukcji wykorzystujących dane do tworzenia systemów przejawiających inteligentne zachowania i rozwiązujących trudne problemy¹³.

¹¹ *Ibidem*, s. 4 (pkt 14–15).

¹² Szerzej na temat definicji i problemów definicyjnych, historii, modeli oraz zastosowań sztucznej inteligencji zob. R. Hurbans, *Algorytmy sztucznej inteligencji. Ilustrowany przewodnik*, przeł. T. Walczak, Gliwice 2025, s. 1–19. Autor ten twierdzi, że pojęcie „sztuczna inteligencja” zostało wymyślone już w latach 50. XX w., a w kolejnych dekadach dokonano postępów w zakresie robotyzacji, tworzenia języków programowania, badań sieci neuronowych, eksperymentów z samochodami autonomicznymi, a także powstania robotów internetowych i wyszukiwarek.

¹³ *Ibidem*, s. 2.

Sama Konwencja ramowa z 2024 r. poświęca definicji systemów sztucznej inteligencji odrębne postanowienie, które wzorowano na Zaleceniu Rady OECD w celu nadania jej uniwersalnego znaczenia¹⁴. W art. 2 Konwencji wskazano, że: „System sztucznej inteligencji oznacza system oparty na maszynie, który – w celach jawnych lub ukrytych – wnioskuję na podstawie otrzymanych danych wejściowych, jak generować wyniki, takie jak prognozy, treści, rekomendacje lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne. Różne systemy sztucznej inteligencji różnią się poziomem autonomii i adaptacyjności po wdrożeniu”. Tak sformułowanej definicji towarzyszy określenie zakresu stosowania Konwencji (art. 3), obejmującego działania w ramach cyklu życia systemów sztucznej inteligencji, które mogą potencjalnie naruszać prawa człowieka, demokrację i praworządność. W tym celu strony Konwencji będą stosować jej postanowienia do działań związanych z cyklem życia systemów SI podejmowanych przez organy publiczne lub podmioty prywatne działające w ich imieniu, a także uwzględniać zagrożenia i skutki wynikające z szeroko pojętych działań podmiotów prywatnych, ale w zgodzie z przedmiotem i celem Konwencji (art. 3 ust. 1b)¹⁵.

Nieco inne ograniczenia zakresu stosowania Konwencji dotyczą wyłączenia materii związanych z obroną narodową oraz zwolnienia stron z obowiązku jej stosowania w odniesieniu do „ochrony swoich interesów bezpieczeństwa narodowego”¹⁶. Nie ulega wątpliwości, że drugie z tych ograniczeń ma znacznie szerszy charakter, a przy tym pozostaje wysoce kontrowersyjne, choć zostało opatrzone imperatywem prowadzenia działań w sposób zgodny z prawem międzynarodowym, włącznie z zobowiązaniami w zakresie praw człowieka i zasadą poszanowania demokratycznych instytucji i procedur. Wydaje się, że takie ujęcie postanowień w art. 3 Konwencji pozostawia rządowi znaczną swobodę w podejmowaniu decyzji w sferze szeroko rozumianych interesów bezpieczeństwa narodowego.

4. Zobowiązania ogólne oraz zasady dotyczące sztucznej inteligencji

Zobowiązania o charakterze ogólnym zostały określone w postanowieniach art. 4 i 5 rozdziału II Konwencji. Obejmują one wymóg przyjęcia lub utrzymania przez strony

¹⁴ Zalecenie Rady OECD w sprawie sztucznej inteligencji (OECD/LEGAL/0449), przyjęte 22 maja 2019 r. (ze zmianami z 2023 r.) oraz *Explanatory Report...*, s. 6 (pkt 23–25).

¹⁵ Strony określą w oświadczeniu złożonym Sekretarzowi Generalnemu Rady Europy w momencie podpisywania lub składania dokumentu ratyfikacji, przyjęcia, zatwierdzenia lub przystąpienia, w jaki sposób zamierzają wdrożyć to zobowiązanie, stosując zasady i obowiązki określone w rozdziałach II–VI Konwencji do działań podmiotów prywatnych lub podejmując inne odpowiednie środki.

¹⁶ Klauzula ta wynika z nakazu wyrażonego w art. 1 d) Statutu Rady Europy, przyjętego w Londynie dnia 5 maja 1949 r. (Dz. U. z 1994 r. Nr 118, poz. 565), który stanowi: „Sprawy dotyczące obrony narodowej nie wchodzą w zakres kompetencji Rady Europy”. Zob. tekst Statutu: <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU19941180565/O/D19940565.pdf> [dostęp: 28.06.2025] oraz *Wybór Konwencji Rady Europy. Wersja polska*, oprac. K. Drzewicki, Warszawa 1999, s. 13–24.

Konwencji środków zapewniających, by działania podejmowane w ramach cyklu życia sztucznej inteligencji były zgodne ze standardami ochrony praw człowieka przewidzianymi w prawie międzynarodowym i wewnętrzkrajowym (art. 4 Konwencji). Strony zachowują swobodę wyboru środków i sposobów realizacji tych postanowień, pod warunkiem, że pozostają one w zgodzie z przyjętymi zobowiązaniami. Jest to zatem zobowiązanie rezultatu, a nie zobowiązanie środków¹⁷.

Z kolei w odniesieniu do integralności procesów demokratycznych oraz poszanowania rządów prawa na mocy Konwencji ustanowiono wymóg, by środki przyjęte przez strony zmierzały do zapewnienia, że sztuczna inteligencja nie zostanie użyta do podważenia integralności, niezależności i efektywności instytucji i procesów demokratycznych (art. 5 Konwencji). Zagrożenia te nie są zjawiskiem nowym w tych obszarach, jednak w ostatnich latach doszło do zintensyfikowania ich zastosowań, co może stwarzać liczne ryzyka dla demokracji, praw człowieka i rządów prawa, zwłaszcza poprzez podważanie zasadności partycypacji obywatelskiej, złośliwe upowszechnianie dezinformacji, manipulowanie autentycznymi treściami, podejmowanie szkodliwych decyzji wobec jednostek, wpływanie na orzecznictwo sądowe i integralność wymiaru sprawiedliwości, a także nielegalny lub arbitralny nadzór, który może skutkować restrykcjami w sferze wolności zgromadzania się, wolności słowa oraz prawa do prywatności. Powstaje zatem obawa, że technologie SI mogą doprowadzić do eskalacji napięć lub podważania zaufania publicznego. Sztuczna inteligencja zyskała bowiem zdolność do generowania fałszywych informacji oraz wykluczania jednostek z procesów demokratycznych. Twórcy Konwencji wskazali na określone procesy demokratyczne, w których zastosowanie sztucznej inteligencji powinno być poprzedzone wnikliwą oceną potencjalnych zagrożeń oraz ustanowieniem stosownych reguł i gwarancji¹⁸. Wymienione ryzyka są dalekie od zjawisk wyłącznie hipotetycznych, ponieważ część z nich ujawniła się już w praktyce państw demokratycznych, m.in. w postaci złośliwej ingerencji zagranicznej w procesy wyborcze czy rozpowszechniania manipulacji informacyjnych. Stąd też w postanowieniach Konwencji wskazano możliwe działania służące przeciwdziałaniu narastającym zagrożeniom z wykorzystaniem SI.

Zakłada się, że wzmocnieniu powinny ulec przede wszystkim dwa podstawowe założenia determinujące integralność demokracji. Jak ujmują to postanowienia art. 7 Konwencji, do tej rangi urastają godność człowieka i autonomia jednostki, wobec których strony mają przyjąć lub utrzymać środki zapewniające ich poszanowanie w odniesieniu do działań w cyklu życia systemów sztucznej inteligencji. Istotna jest zatem

¹⁷ *Explanatory Report...*, s. 8–10 (pkt 37–41). W sprawozdaniu wyjaśniającym, w celu wskazania międzynarodowych instrumentów z zakresu praw człowieka, które strony Konwencji mogą brać pod uwagę, wymieniono osiem traktatów ONZ, dziesięć instrumentów Rady Europy i Unii Europejskiej oraz siedem innych instrumentów o charakterze regionalnym.

¹⁸ *Ibidem*, s. 10–12 (pkt 42–48). W ramach tej szczególnej ochrony strony Konwencji wskazały osiem ustrojowych rozwiązań demokratycznych: zasadę podziału władz, włącznie z jej wymiarem wertykalnym, efektywny system kontroli i równowagi, pluralizm polityczny, partycypację w procesach demokratycznych, zasadę większości w powoływaniu i poszanowaniu praw mniejszości, poszanowanie rządów prawa oraz niezależności sądowej.

zdolność do kształtowania opinii przez jednostkę i działania podejmowane na tej podstawie oraz zdolność do wywierania wpływu na decyzje podjęte przez jednostkę lub w jej imieniu. W sprawozdaniu wyjaśniającym do Konwencji mocno podkreślono, że działania w ramach cyklu życiowego systemów SI nie powinny prowadzić do dehumanizacji jednostek, podważać ich zdolności lub redukować ich do samych danych, a także prowadzić do antropomorfizacji tych systemów w sposób ingerujący w ludzką godność¹⁹.

W Konwencji przewidziano również wśród zasad ogólnych obowiązek zapewnienia odpowiednich wymogów przejrzystości i nadzoru, dostosowanych do konkretnych kontekstów i zagrożeń, a także utrzymania środków zabezpieczających odpowiedzialność i rozliczalność za negatywny wpływ na prawa człowieka, demokrację i praworządność, wynikający z działań w cyklu życia systemów sztucznej inteligencji (art. 8 i 9). Ponadto działania w cyklu życiowym systemów SI powinny być zgodne z zasadą równości, w tym równości płci, oraz z zakazem dyskryminacji, przy poszanowaniu obowiązującego prawa międzynarodowego i krajowego. W ramach tych zasad strony zobowiązują się do przyjęcia lub utrzymania środków służących przezwyciężaniu nierówności w celu osiągnięcia sprawiedliwych, uczciwych i równych wyników, zgodnie z obowiązującymi krajowymi i międzynarodowymi zobowiązaniami w zakresie praw człowieka w odniesieniu do działań w ramach cyklu życia systemów sztucznej inteligencji (art. 10 Konwencji).

Dalsze postanowienia Konwencji dotyczą ochrony prywatności i danych osobowych oraz proceduralnych gwarancji i zabezpieczeń w krajowych i międzynarodowych procedurach ochronnych (art. 11). Uzupełnieniem tych regulacji są postanowienia dotyczące podejmowania, w odpowiednich przypadkach, środków w celu promowania niezawodności systemów sztucznej inteligencji i zaufania do ich wyników, co może obejmować wymogi w zakresie odpowiedniej jakości i bezpieczeństwa w całym cyklu życia systemów SI (art. 12). Ponadto w Konwencji przewidziano wspieranie bezpiecznych innowacji poprzez tworzenie kontrolowanych środowisk umożliwiających opracowywanie systemów sztucznej inteligencji oraz przeprowadzanie eksperymentów i testów z jej wykorzystaniem, pod nadzorem właściwych organów, przy jednoczesnym unikaniu negatywnego wpływu na prawa człowieka, demokrację i praworządność (art. 13).

W kwestii środków odwoławczych i innych zabezpieczeń proceduralnych w Konwencji przyjęto rozwiązania stanowiące pomost między środkami odwoławczymi z klasycznych traktatów praw człowieka a potencjalnymi naruszeniami wynikającymi z działalności prowadzonej w ramach cyklu życia systemów sztucznej inteligencji. Na mocy art. 14 strony są zobowiązane zapewnić, aby prawo krajowe przewidywało nie tylko odpowiednie środki odwoławcze, lecz także obowiązek dokumentowania wykorzystania systemów SI oraz udostępniania tej dokumentacji osobom, których dotyczyło zastosowanie tych systemów. Zakłada się również, że każda osoba będzie mogła skutecznie skorzystać z wniesienia skargi do właściwego organu władzy publicznej

¹⁹ *Ibidem*, s. 12 (pkt 53–55).

(art. 15). Co więcej, strony mają zapewnić, by osoby wchodzące w relacje interaktywności ze sztuczną inteligencją były informowane, że mają do czynienia z systemem SI, a nie z istotą ludzką (art. 15 ust. 1)²⁰.

5. Ocena i łagodzenie zagrożeń oraz niekorzystnego wpływu

Twórcy Konwencji uznali za stosowne odniesienie się również do kwestii zagrożeń oraz zarządzania ryzykiem i jego skutkami (art. 16). Strony są zobowiązane do przyjęcia lub utrzymania środków służących identyfikacji, ocenie, zapobieganiu i łagodzeniu ryzyka stwarzanego przez systemy sztucznej inteligencji, uwzględniając rzeczywiste i potencjalne skutki ich działania dla praw człowieka, demokracji i praworządności. Środki te powinny być odpowiednio stopniowane i różnicowane. Wśród wskazanych działań wymieniono m.in. uwzględnianie kontekstu oraz zamierzonych zastosowań systemów sztucznej inteligencji, w szczególności w odniesieniu do zagrożeń dla praw człowieka, demokracji i praworządności; ocenę wagi i prawdopodobieństwa potencjalnych skutków, zwłaszcza dla osób, których prawa mogą zostać naruszone; monitorowanie zagrożeń i negatywnych skutków dla praw człowieka, demokracji i praworządności, a także dokumentowanie rzeczywistych i potencjalnych zagrożeń oraz skutków. Ponadto, w stosownych przypadkach, wymagane jest przeprowadzenie testowania systemów SI przed ich udostępnieniem do pierwszego użycia oraz w przypadku ich znaczącej modyfikacji (art. 16 ust. 2).

Zgodnie z Konwencją konieczne jest ponadto odpowiednie przeciwdziałanie negatywnemu wpływowi systemów sztucznej inteligencji na prawa człowieka, demokrację i praworządność. Negatywne skutki i środki zaradcze powinny być dokumentowane i uwzględniane przez odpowiednie środki zarządzania ryzykiem (art. 16 ust. 3). Strony powinny oceniać potrzebę wprowadzenia moratorium, zakazu lub innych odpowiednich środków w odniesieniu do niektórych zastosowań systemów SI, jeżeli uznają, że takie zastosowania są niezgodne z poszanowaniem praw człowieka, funkcjonowaniem demokracji lub praworządnością (art. 16 ust. 4)²¹.

W rozdziale V Konwencji wskazano na potrzebę podejmowania działań o charakterze prewencyjnym, które będą sprzyjać zapobieganiu naruszeniom reguł demokracji, praw człowieka i rządów prawa. Stanowi to potwierdzenie wysokiego poziomu obaw stron Konwencji dotyczących możliwości wystąpienia potencjalnych naruszeń triady wartości, które chcą one nie tylko chronić, lecz także skutecznie im zapobiegać.

²⁰ *Ibidem*, s. 22 (pkt 103–104).

²¹ *Ibidem*, s. 23–24 (pkt 105–112).

6. Zasady i środki implementacji Konwencji

Analizowana Konwencja, w sposób typowy dla konwencji Rady Europy w sferze praw człowieka, przewiduje własny system wdrażania (implementacji). Zapewne ze względu na wagę i złożoność problematyki sztucznej inteligencji w Konwencji przyjęto system implementacyjny oparty na wielości środków implementacji o zróżnicowanym charakterze. W pierwszej kolejności określono grupę wspólnych zasad odnoszących się do wszystkich tych środków, zaczynając od zasady niedyskryminacji. Założono, że ochroną powinny zostać objęte w szczególności osoby narażone na zwiększone ryzyko, w tym osoby niepełnosprawne i dzieci (art. 18). Zasadą powinny stać się również konsultacje społeczne. Dlatego każda strona powinna dążyć do zapewnienia, aby istotne kwestie podnoszone w odniesieniu do systemów sztucznej inteligencji były należycie rozpatrywane w drodze dyskusji publicznej oraz konsultacji z udziałem wielu zainteresowanych stron, z uwzględnieniem kwestii społecznych, ekonomicznych, prawnych, etycznych, środowiskowych i innych (art. 19). Ponadto w Konwencji przewidziano obowiązek wspierania i promocji przez strony odpowiednich umiejętności cyfrowych (*digital literacy*) wśród wszystkich grup społecznych, w tym specjalistycznych umiejętności wśród osób odpowiedzialnych za identyfikację, ocenę, zapobieganie i łagodzenie ryzyka stwarzanego przez systemy sztucznej inteligencji (art. 20).

Wśród organów i mechanizmów wdrażania Konwencji strony przyznały czołowe miejsce Konferencji przedstawicieli stron, która ma stanowić forum służące skutecznej realizacji postanowień tej umowy. Do zadań tego gremium należy m.in. identyfikowanie wszelkich problemów i skutków, zgłoszonych zastrzeżeń i oświadczeń, rozważanie ewentualnego uzupełnienia lub zmiany Konwencji, rozważanie kwestii i formułowanie szczegółowych zaleceń dotyczących interpretacji i stosowania Konwencji, ułatwianie wymiany informacji na temat istotnych zmian prawnych, politycznych lub technologicznych, mających znaczenie dla jej wdrażania, a także ułatwianie polubownego rozstrzygania sporów związanych z jej stosowaniem (art. 23)²².

Kolejnym środkiem implementacji przewidzianym w omawianej Konwencji jest mechanizm sprawozdawczy (art. 24). Każda strona jest zobowiązana do przedłożenia Konferencji, po upływie dwóch lat od związania się Konwencją, a następnie okresowo, sprawozdania szczegółowo opisującego działania podjęte w celu wdrożenia art. 3 ust. 1 lit. a i b. Format i procedurę sporządzania sprawozdań określa zgodnie z regulaminem Konferencja przedstawicieli stron.

Implementacja Konwencji obejmuje również współpracę międzynarodową (art. 25). W tym zakresie strony są zachęcane, w stosownych przypadkach, do wspierania państw, które nie są stronami Konwencji, ale działają zgodnie z jej postanowieniami i dążą do jej ratyfikacji. Ponadto przewidziano, że strony będą wymieniać między sobą istotne i przydatne informacje dotyczące aspektów związanych ze sztuczną

²² Inna grupa postanowień art. 23 Konwencji dotyczy organizacyjnych aspektów zwoływania i funkcjonowania Konferencji, w tym uchwalenia regulaminu i finansowania przez państwa niebędące członkami Rady Europy (art. 23 ust. 3–8).

inteligencją, które mogą mieć znaczący pozytywny lub negatywny wpływ na poszanowanie praw człowieka, funkcjonowanie demokracji i przestrzeganie praworządności, w tym informacji o związanych z nimi ryzykach i skutkach, wynikających z badań naukowych oraz działalności sektora prywatnego. Strony są zachęcane do angażowania w taką wymianę informacji, w stosownych przypadkach, zainteresowanych uczestników i państw niebędących stronami Konwencji.

Poza powyższymi środkami implementacji w Konwencji przewidziano obowiązek ustanowienia lub wyznaczenia przez strony jednego lub więcej skutecznych mechanizmów nadzoru nad przestrzeganiem wynikających z niej zobowiązań (art. 26 ust. 1–4). Każda strona ma zapewnić, aby podmioty działające w ramach tych mechanizmów wykonywały swoje obowiązki w sposób niezależny i bezstronny oraz posiadały niezbędne uprawnienia, specjalistyczną wiedzę i zasoby do skutecznego wypełniania zadań nadzoru nad przestrzeganiem zobowiązań wynikających z postanowień Konwencji, zgodnie z zakresem określonym przez strony. Jeżeli strona przewidziała więcej niż jeden taki mechanizm, powinna podjąć środki w celu ułatwienia skutecznej współpracy między nimi. Jeżeli strona przewidziała mechanizmy inne niż istniejące krajowe struktury praw człowieka, powinna podjąć, w miarę możliwości, działania na rzecz promowania skutecznej współpracy między tymi mechanizmami a wspomnianymi strukturami. Oznacza to, że twórcy Konwencji nie zdecydowali się na powołanie odpowiedniego mechanizmu nadzoru na szczeblu międzynarodowej implementacji tej umowy – jego organizację pozostawiono krajowym systemom państw-stron²³. Rozwiązanie to nie prowadzi jednak do ograniczenia dostępnych środków ochrony prawnej dla osób uznających się za ofiary niekorzystnych działań lub zaniechań. Pozostaje bowiem realna możliwość skorzystania z prawa do skargi indywidualnej na gruncie Europejskiej Konwencji Praw Człowieka²⁴ w sprawach dotyczących m.in. prawa do prywatności, prawa do stowarzyszania, wolności słowa, prawa do udziału w wyborach parlamentarnych i innych.

7. Podsumowanie

Konwencja ramowa w sprawie sztucznej inteligencji i praw człowieka, demokracji i rządów prawa jest instrumentem wyjątkowym. Została sporządzona w odpowiedzi na zaawansowany rozwój i stosowanie SI w praktyce, wykraczające poza warunki wyłącznie

²³ W końcowym rozdziale VII Konwencji zawarto liczne klauzule, które dotyczą zawierania umów między stronami w sprawach objętych jej regulacjami, wpływu prawa Unii Europejskiej na stosowanie Konwencji (art. 27), dokonywania w niej zmian (art. 28), załatwiania sporów między jej stronami (art. 29), podpisania, wejścia w życie i przystępowania do niej (art. 30–31), terytorialnego zakresu jej stosowania (art. 32), klauzuli federalnej (art. 33), zastrzeżeń (art. 34), wypowiedzenia Konwencji (art. 35) oraz wymaganych notyfikacji (art. 36).

²⁴ Konwencja o Ochronie Praw Człowieka i Podstawowych Wolności, sporządzona w Rzymie dnia 4 listopada 1950 r., zmieniona następnie Protokołami nr 3, 5 i 8 oraz uzupełniona Protokołem nr 2 (Dz. U. z 1993 r. Nr 61, poz. 284, ze zm.).

laboratoryjne. Doprowadziło to do objęcia przez sztuczną inteligencję niemal wszystkich dziedzin życia, a nie tylko obszarów techniki, technologii i działalności gospodarczej. Wkroczyła ona również do sfery życia społecznego, medycyny, farmakologii, obronności, ustroju politycznego i prawnego, rządów prawa, demokracji oraz praw człowieka. Ten nowy etap rozwoju sztucznej inteligencji przyniósł jednak nie tylko korzyści dla człowieka i wspólnot, lecz także zagrożenia. Wykorzystanie SI w obiegu informacji oraz podczas wyborów parlamentarnych dostarczyło niepokojących przykładów manipulacji, powodujących poważne ryzyka dla człowieka i organizacji życia zgodnie z regułami etycznymi i prawnymi. Dylemat między korzyściami a zagrożeniami spowodował jednak w miarę szybką reakcję państw oraz innych aktorów krajowych i międzynarodowych zmierzającą do podjęcia odpowiednich kroków zaradczych. Świadczy o tym lista instrumentów wskazana w sprawozdaniu wyjaśniającym, spośród których najbardziej kompleksową regulacją jest rozporządzenie nr 2024/1689²⁵.

Zobowiązania przyjęte przez strony Konwencji mają charakter postanowień ramowych, a zatem wymagają uzupełnienia bardziej szczegółowymi rozwiązaniami w prawie krajowym. Część regulacji ma przy tym charakter nie tylko ochronny, lecz także zapobiegawczy. Szeroko określono również zobowiązania państw w zakresie budowy infrastruktury pozwalającej na regularny nadzór nad procesami stosowania sztucznej inteligencji. Takie podejście wydaje się konieczne, ponieważ zahamowanie lub spowolnienie rozwoju SI jest trudne bądź nawet niemożliwe w ramach istniejących wolności. Konwencja Rady Europy pojawia się zatem już na etapie nieco spóźnionej reakcji na powstające zagrożenia. Czy jednak reakcja państw i organizacji międzynarodowych okaże się skuteczna, pokażą dopiero ich rzeczywiste dokonania. W charakterze przestrogi niech posłuży wskazany poniżej przypadek, ukazujący zarówno osiągnięcia sztucznej inteligencji, jak i zagrożenia, które może ona generować, a także zakres możliwych lub nieosiągalnych środków przeciwdziałania.

8. Zamiast konkluzji

Dalsze konkluzje niniejszego tekstu niech zastąpi przypomnienie historyjki sprzed ponad pół wieku, która pokazuje, jak dalece naturalna inteligencja może reagować, lepiej lub gorzej, na dokonania systemów sztucznej inteligencji.

Nasz wielki pisarz literatury science fiction – Stanisław Lem (aż dziw, że nie został obdarzony Literacką Nagrodą Nobla) w opublikowanej w 1965 r. *Cyberiadzie* opisał następującą historię. Na jednej z odległych planet skonstruowano robota wyposażonego w program układający na życzenie teksty wierszowane. Wywołało to protesty stowarzyszeń literackich, a nawet samobójstwa wśród poetów. W celu podjęcia próby jego „eliminacji” robotowi postawiono ostateczne zadanie – ułożenie rymowanego wiersza. Kazano mu napisać wiersz o cyberotyce, liczący najwyżej sześć wersów, w którym miałyby być o miłości i o zdradzie, o muzyce, o Murzynach, o wyższych sferach,

²⁵ *Explanatory Report...*, s. 2–3 (pkt 7–8).

o nieszczęściu, o kazirodstwie, przy czym wszystkie słowa miały zaczynać się na literę „C”! Robot wykonał polecenie i napisał taki oto wiersz:

Cyprian cyberotoman, cynik, ceniąc czule
Czarnej córy cesarskiej cud ciemnego ciała,
Ciągłe cytrą czarował. Czerwieniła cała,
Cicha, co dzień czekała, cierpiała, czuwała...
...Cyprian ciotkę całuje, cisnąwszy czarnulę!!

Gdy kilka lat później podjęto próbę przetłumaczenia *Cyberiady* na język angielski, okazało się, że znakomity tłumacz dzieł Stanisława Lema, Michael Kandel, nie był w stanie stworzyć w pełni wiernego tłumaczenia tego wiersza. Sens wyznaczonego robotowi zadania oddał w inny, choć zbliżony sposób, pisząc utwór złożony wyłącznie ze słów na literę „S”²⁶.

Seduced, shaggy Samson snored.
She scissored short. Sorely shorn,
Soon shackled slave, Samson sighed,
Silently scheming,
Sightlessly seeking
Some savage, spectacular suicide.

W dosłownym tłumaczeniu na język polski wiersz ten brzmiałby tak:

Uwiedziony, kudłaty Samson chrapał.
Ona ostrzygła krótko. Boleśnie ostrzyżony,
Wkrótce skutą niewolnik, Samson westchnął,
Cicho knując,
Niewidomy poszukujący
Jakiegoś dzikiego, spektakularnego samobójstwa.

Powstaje zatem pytanie, czy dziś, po sześciu dekadach, sztuczna inteligencja byłaby w stanie wiernie przetłumaczyć wiersz robota.

Literatura

- Alfredsson G., *A Frame an Incomplete Painting: Comparison of the Framework Convention for the Protection of National Minorities with International Standards and Monitoring Procedures*, „*International Journal on Minority and Group Rights*” 2000, vol. 7, no. 4, s. 291–304.
- Drzewicki K., *European Systems for the Promotion and Protection of Human Rights* [w:] *International Protection of Human Rights: A Textbook*, eds. C. Krause, M. Scheinin, Turku/Åbo 2009, s. 397–422.
- Drzewicki K., *Triada czy sekstet? Kontrowersje wokół warunków członkostwa w Radzie Europy* [w:] *Konstytucjonalizm polski. Refleksje z okazji jubileuszu 70-lecia urodzin i 45-lecia pracy naukowej*

²⁶ S. Lem, *The Cyberiad: Fables for the Cybernetic Age*, transl. M. Kandel, San Diego 1985, s. 50–51.

- Profesora Andrzeja Szmyta*, red. A. Gajda, K. Grajewski, A. Rytel-Warzocha, P. Uziębło, M.M. Wiśzowaty, Gdańsk 2020, s. 495–506.
- Hurbans R., *Algorytmy sztucznej inteligencji. Ilustrowany przewodnik*, przeł. T. Walczak, Gliwice 2025.
- Lem S., *The Cyberiad, The Cyberiad: Fables for the Cybernetic Age*, transl. M. Kandel, San Diego 1985.
- Wybór Konwencji Rady Europy. Wersja polska*, oprac. K. Drzewicki, Warszawa 1999.

Streszczenie

Krzysztof Drzewicki

Sztuczna inteligencja a prawa człowieka, demokracja i rządy prawa. Uwagi na tle Konwencji nr 225 Rady Europy z 2024 r.

Artykuł poświęcono analizie Ramowej konwencji Rady Europy z 2024 r. dotyczącej sztucznej inteligencji oraz jej relacji z prawami człowieka, demokracją i rządami prawa. Ten półotwarty, wyjątkowy w praktyce Rady Europy traktat wyznacza ramy regulacyjne dla całego cyklu życia systemów SI, nie tworząc nowych obowiązków w zakresie praw człowieka. Przyjmuje definicję SI zgodną z OECD i obejmuje działania władz publicznych oraz podmiotów działających w ich imieniu; wyłącza obronę narodową i dopuszcza szeroką klauzulę bezpieczeństwa. Celem Konwencji jest wykorzystanie korzyści płynących z SI przy jednoczesnym ograniczaniu ryzyk, takich jak dyskryminacja, manipulacja czy nadzór. Strony mają stosować podejście oparte na ryzyku, chronić godność i autonomię jednostki, zapewniać przejrzystość, nadzór i rozliczalność, wspierać równość i niedyskryminację, chronić prywatność, dbać o jakość i bezpieczeństwo SI oraz sprzyjać bezpiecznym innowacjom. Przewidziano też gwarancje środków odwoławczych (m.in. obowiązek informowania o interakcji z SI). Wdrażanie postanowień Konwencji oparto na Konferencji przedstawicieli stron, sprawozdawczości okresowej, współpracy międzynarodowej i krajowych mechanizmach nadzoru.

Słowa kluczowe: konwencja ramowa, ład SI, prawa człowieka.

Summary

Krzysztof Drzewicki

Artificial Intelligence and Human Rights, Democracy, and the Rule of Law: Remarks in the Light of Council of Europe Convention No. 225 of 2024

The article analyzes the Council of Europe's 2024 Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law. This semi-open treaty – exceptional in Council of Europe practice – establishes regulatory parameters for the entire lifecycle of AI-systems without creating new human-rights obligations. It adopts an OECD-aligned definition of AI and applies to actions by public authorities and private entities acting on their behalf; national defence is excluded, and a broad national-security reservation is allowed. The Convention aims

to harness benefits while mitigating risks such as discrimination, manipulation, and surveillance. Parties must adopt a risk-based approach; safeguard human dignity and individual autonomy; ensure transparency, oversight, and accountability; uphold equality and non-discrimination; protect privacy; secure AI quality and safety; and foster safe innovation. It provides guarantees of redress, including the obligation to give notice of AI–human interaction. Implementation is entrusted to a Conference of the Parties, periodic reporting, international cooperation, and domestic supervisory mechanisms.

Keywords: Framework Convention, AI governance, human rights.