

## KRZYSZTOF BINKOWSKI

Uniwersytet im. Adama Mickiewicza w Poznaniu  
<https://orcid.org/0009-0005-8544-2151>  
krzbin@st.amu.edu.pl

# Sztuczna inteligencja a wykładnia prawa – propozycja zastosowania systemów AI do ustalania założenia o racjonalnym prawodawcy

**Artificial intelligence and the interpretation of law – proposal to use AI systems to establish the assumption of a rational legislator**

**Abstract:** Artificial intelligence is developing rapidly and is increasingly present in public discourse, hence questions about its potential application in the legal field are becoming more common. The aim of this thesis is to consider whether AI can improve the process of applying the law, namely establishing a hypothesis for the deep structure of legal science. The research problem is to establish AI's operational capabilities – if AI can adequately reproduce assumption of a rational legislator. In order to consider this, the discussion on the AI definition and methodological issues of legal science were examined. Finally, the conditions of assumption of a rational legislator were reconstructed and it was considered whether a system adapting this assumption could be developed.

**Keywords:** artificial intelligence, interpretation of the law, assumption of a rational legislator, application of the law

## Wprowadzenie

Sztuczna inteligencja (*artificial intelligence*, AI) to prężnie rozwijająca się dziedzina wiedzy, a jej zastosowanie może być bardzo korzystne również w obszarze nauki i praktyki prawa. W niniejszym opracowaniu poddano pod rozważenie, czy systemy sztucznej inteligencji są w stanie pomóc w ustalaniu i precyzowaniu hipotez należących do struktury głębokiej nauki prawa. Wydaje się, że sztuczna inteligencja jest w stanie odtwarzać wzorce ocenne, stąd zaproponowano i rozważono metodologiczne kwestie zastosowania jej do teorii idealizacyjnej dotyczącej racjonalnego prawodawcy. Przy tym za problem badawczy postanowiono uznać kwestię możliwości operacyjnych AI, a mianowicie to, czy sztuczna inteligencja odtwarzałaby model racjonalnego prawodawcy *per se* czy jedynie wyobrażenia o tymże założeniu skutecznie przez organy stosujące prawo. Wydaje się, że ważne byłoby tu jak największe zespolenie rozwiązań zaproponowanych przez teorię prawa z praktyką wykładni prawa. Aby wyprowadzić powyższą hipotezę, należy najpierw przeanalizować dyskusję na temat definicji sztucznej inteligencji oraz rozważyć metodologiczne kwestie nauki prawa. Wreszcie w kwestii założenia o prawodawcy racjonalnym należy odtworzyć warunki związane z tym konstruktem i zastanowić się, czy i w jaki sposób mógłby powstać system adaptujący to założenie.

## 1. Definicja sztucznej inteligencji

Zaczynając rozważania, należy dla porządku przedstawić definicję sztucznej inteligencji. Jest to prężnie rozwijająca się dziedzina informatyki, natomiast – przez nagromadzenie kontekstów z różnych dyscyplin naukowych – trudno podać jedną obowiązującą definicję. Zaczynając od nazwy, sztuczna inteligencja ma na celu odzwierciedlanie działania ludzkiej inteligencji, a w szczególności umiejętność uczenia się. Za Johnem McCarthym można podać, że sama inteligencja jest definiowana jako „obliczeniowa umiejętność służąca do osiągnięcia celów w świecie”<sup>1</sup>. Odwołania do ludzkiej inteligencji są konieczne, gdyż dotychczas nie rozstrzygnięto, jakie konkretnie rodzaje procedur obliczeniowych można uznać za inteligencję (w oderwaniu od inteligencji ludzkiej). Najbardziej precyzyjną definicję AI podano w monografii *Prawo sztucznej inteligencji*: „[AI to] system, który pozwala na wykonywanie zadań wymagających procesu uczenia się i uwzględniania nowych okoliczności w toku rozwiązywania danego problemu i który może w różnym stopniu – w zależności od konfiguracji – działać autonomicznie oraz wchodzić w interakcje z otoczeniem”<sup>2</sup>. „Uczenie się” z kolei to analizowanie i interpretacja danych w celu udoskonalenia działania systemu. Warto przedstawić również podział sztucznej inteligencji funkcjonującej w informatyce. Przyjmuje się, że wyróżniamy: „słabą sztuczną inteligencję” – to program, który ma do dyspozycji bazy danych (najczęściej z ogromną liczbą rekordów) i na ich podstawie uczy się oraz podejmuje decyzje; kolejnym poziomem AI, a może raczej konceptem, bo taki system nie istnieje (a według wielu – nigdy nie powstanie) byłaby „silna AI”, czyli program faktycznie myślący w sposób niestymulowany. Ten rodzaj AI prawdopodobnie byłby obdarzony samoświadomością<sup>3</sup>.

AI jest wszakże zbiorem niejednorodnym, obejmuje wiele różnie działających systemów. W niniejszej pracy zostaną wskazane tylko gałęzie AI mogące ewentualnie znaleźć zastosowanie do teorii i praktyki nauk prawnych. Istotne są tu przede wszystkim systemy AI działające za pomocą rozwiązań logiki matematycznej (*logical AI*). Po wprowadzeniu do nich zdań o faktach, programy mogą wnioskować o prawdziwości innych faktów. Warta uwagi jest również metoda nazywana przez McCarthy’ego mianem *planning*. Tutaj program rozpoczyna pracę z wiedzą ogólną o świecie, z faktami o konkretnej sytuacji oraz z wyznaczonym celem, a następnie z posiadanej wiedzy generuje strategię osiągnięcia celu (sekwencję działań)<sup>4</sup>. Co do aplikacji bazujących na AI, należy niezbędnym wyróżnić systemy ekspertowe.

<sup>1</sup> J. McCarthy, *What Is Artificial Intelligence?* Stanford 2007 <<http://jmc.stanford.edu/articles/whatisai.html>> [dostęp: 23.03.2023].

<sup>2</sup> T. Zalewski, *Rozdział I. Definicja sztucznej inteligencji*, [w:] L. Lai, M. Świerczyński (red.), *Prawo sztucznej inteligencji*, Warszawa 2020, Legalis/el. <<https://sip-1legalis-1pl-1015defdg1cf4.han.amu.edu.pl/document-full.seam?documentId=mjxw62zogi3damrwgaydqnrqoy>> [dostęp: 20.03.2023].

<sup>3</sup> J. McCarthy, op. cit.

<sup>4</sup> Ibidem.

W ich przypadku inżynier wiedzy rozmawia z ekspertem w danej dziedzinie i osadza jego wiedzę w programie, tak aby AI mogła rozwiązywać określone problemy (np. stawianie diagnozy pacjentom). Jest to bardzo obiecujący koncept, lecz trzeba pamiętać, że w podanym przykładzie oprócz samych symptomów choroby istotna jest sytuacja pacjenta, warunki w szpitalu, możliwe leczenie, dlatego systemy ekspertowe muszą być jak najbardziej kompleksowe<sup>5</sup>.

## 2. Metodologia prawa

Po omówieniu reguł działania sztucznej inteligencji należy rozważyć metodologię nauk prawnych i zastanowić się, czy na tym poziomie możliwa jest aplikacja rozwiązań AI do stosowania prawa. Leszek Nowak zwraca uwagę na dwie sfery poznania naukowego. Pierwsza z nich to „struktura głęboka” nauki, czyli zbiór twierdzeń zaakceptowanych przez naukowców jedynie milcząco, druga to „struktura powierzchniowa”, czyli wypowiedzi, które można napotkać w pracach zwyczajowo zaliczanych do nauki, a więc zbiór twierdzeń wypowiedzianych i potwierdzonych przez badaczy. Nowak stwierdza, że zadaniem metodologa jest właśnie wyjaśnienie struktury powierzchniowej nauki na podstawie hipotez dotyczących jej struktury głębokiej. Twierdzi również, że im nauka jest słabiej rozwinięta, tym większy jest obszar struktury głębokiej i tym większa rola metodologii<sup>6</sup>. Idąc dalej, można przyjąć, że jedną z podstawowych metod nauk empirycznych jest metoda idealizacji i konkretyzacji, która polega na wprowadzaniu pewnych założeń upraszczających badany fragment rzeczywistości i ustalaniu hipotez mających obowiązywać przy tych założeniach, które jednocześnie wystarczająco dobrze przybliżają rzeczywistość<sup>7</sup>.

Nauki humanistyczne, w tym nauki prawne, są jednak o tyle różne od nauk ścisłych i przyrodniczych, że ich badacze nie formułują ogólnych prawideł. Zamiast tego wspomniani naukowcy wyjaśniają przedmiot swoich zainteresowań za pomocą metody zwanej „interpretacją humanistyczną”, opracowanej przez Jerzego Kmitę. Jest to szczególna postać metody idealizacji i konkretyzacji, która polega na tym, że w czasie wyjaśniania pewnej czynności przyjmuje się założenia, że podmiot jest racjonalny, czyli uznaje niesprzeczny system wiedzy, a także ma określone preferencje oraz wybiera spośród możliwych do podjęcia czynności te, które prowadzą do najbardziej preferowanego stanu rzeczy<sup>8</sup>. To służy do interpretacji postępowania danej osoby – model racjonalności wyjaśnia, dlaczego dana osoba postąpiła tak a tak. Metody interpretacji humanistycznej, jak zauważa Nowak, mogą być wielce przydatne również na gruncie nauk prawnych do odczytywania

<sup>5</sup> Ibidem.

<sup>6</sup> L. Nowak, *Interpretacja prawnicza. Studium z metodologii prawoznawstwa*, Warszawa 1973, s. 11–12.

<sup>7</sup> Ibidem, s. 16–17.

<sup>8</sup> Ibidem, s. 25–26.

znaczenia tekstów prawnych i struktury głębokiej prawoznawstwa<sup>9</sup>. Zadaniem badacza prawa jest bowiem wyjaśnienie, jaki cel przyświecał prawodawcy tworzącemu przepis prawa i dlaczego posłużył się takimi, a nie innymi metodami, stworzył takie, a nie inne zdania. Wobec tego prawo objaśnia się z perspektywy celów normodawcy. Kmita oraz Nowak największy nacisk kładą właśnie na odczytanie zamiaru użytkownika języka, stwierdzając, że: „znaczenie jest szczególnego rodzaju sensem (celem) użytkowników języka – pytanie o znaczenie wyrażenia w danym języku jest w istocie pytaniem o to, jaki cel komunikacyjny wiąże użytkownik języka (idealny) z użyciem tego wyrażenia. A w rezultacie jest to pytanie o humanistyczną interpretację zachowań językowych”<sup>10</sup>. Można stwierdzić, że ta interpretacja będzie miała dwa aspekty: pierwszy to rozumienie słów (wykładnia językowa), a drugi to zgodność znaczeniowa słów z zamiarem prawodawcy (wykładnia funkcjonalna i systemowa).

Odpowiednie modele AI, jak już wspomniano, posługują się logiką zaadaptowaną do odpowiednich języków programowania. Co do nauk prawnych, to w nich także logika jest sednem języka prawnego, a w połączeniu z metodami wykładni, logika pozwala na odczytanie treści norm prawnych z przepisów prawa. Zarówno w prawie, jak i modelach sztucznej inteligencji istotne zadanie mają reguły wnioskowania. To dzięki nim algorytmy sztucznej inteligencji mogą się uczyć i udoskonalać swój system działania. W prawie natomiast wnioskowanie z norm o obowiązywaniu innych norm pozwala na zachowanie zwięzłości tekstu prawnego<sup>11</sup>. Skoro system prawa wykazuje pewne podobieństwa do systemów sztucznej inteligencji, rządzi się co do zasady podobnym językiem (u podstaw leży logika z nadbudową odpowiednio języka prawnego i języka programowania) oraz również uznaje reguły wnioskowania, to sztuczna inteligencja może pomóc w interpretacji prawa<sup>12</sup>.

Warto wspomnieć jeszcze o kwestii racjonalności, która jest istotna zarówno w teorii prawa (por. ustalenia Kmity i Nowaka), jak i przy tworzeniu sztucznej inteligencji. Stuart Russell i Peter Norvig uznali racjonalność za podejście idealne w projektowaniu AI, w którym systemy AI działają najlepiej. Przeciwstawiono to podejściu ludzkiemu, czyli takiemu, które nie za każdym razem cechuje się racjonalnością. Można stąd wywodzić, że trenując sztuczną inteligencję, niekoniecznie właściwe jest przedstawianie danych płynących z doświadczenia czy historii ludzi, nie zawsze bowiem będą to działania czy wybory racjonalne<sup>13</sup>. Zgodnie ze sformułowanym przez Davida Hume’a błędem naturalistycznym (tj. regułą, że wnioskowanie

<sup>9</sup> Ibidem, s. 32–34.

<sup>10</sup> Ibidem, s. 34.

<sup>11</sup> S. Wronkowska, *Podstawowe pojęcia prawoznawstwa*, Poznań 2002, s. 92–95.

<sup>12</sup> Istotnym problemem będzie jednak adaptacja języka prawniczego (bazującego na języku etnicznym) do języka, którym posługuje się AI, szczególnie ze względu na fakt, że nierzadko pojęcia języka prawniczego są niejednoznaczne.

<sup>13</sup> S. Russell, P. Norvig, *Artificial Intelligence: A modern approach*, London 2020, s. 7–8.

na podstawie istniejącego stanu rzeczy o stanie, jaki istnieć powinien, wnioskowanie z faktów o normach, jest niepoprawne) tworzenie modelu racjonalności nie powinno się odbywać wyłącznie na podstawie istniejącego stanu faktycznego. Aby sformułować model idealny, należy zastosować narzędzia teorii i stworzyć odpowiednie założenia.

### 3. Stronnicza sztuczna inteligencja

Co do tworzenia systemów AI, istotnym problemem jest to, że mogą one stać się stronnicze w swoich wyborach. Zależy to w głównej mierze od danych, które są aplikowane na potrzeby treningu takiego programu, oraz od trenera, który, nawet nieświadomie, może przelać swoje uprzedzenia na działanie AI<sup>14</sup>. Ze względu na to, że sztuczna inteligencja bazuje na modelach statystycznych, może w nieprawidłowy sposób zakładać korelacje między niepowiązаныmi ze sobą cechami. Dla przykładu, jeśli AI prowadzi system rekrutacyjny i jest wyposażona w dane stanowiące, że większość pracowników ma biały kolor skóry, to może automatycznie założyć, iż należy odrzucić pracowników o innym kolorze skóry. Błąd ten wynika m.in. ze zbyt wąskiego doboru danych, wobec czego przy programowaniu systemu AI należy zwracać szczególną uwagę na gromadzenie danych na temat mniejszościowych grup społecznych<sup>15</sup>. Skuteczna może się też okazać metoda zwana „kontrfaktyczną uczciwością” (*counterfactual fairness*). Polega ona na sprawdzeniu, czy rezultat działania algorytmu byłby taki sam w „kontrfaktycznym świecie”, tj. przy zmianie danych dotyczących pochodzenia danej osoby czy jej płci<sup>16</sup>.

Fakt, że AI jest podatna na tworzenie stereotypów, może okazać się ważny dla jej zastosowań w prawie. Wracając do językowego znaczenia stereotypu, jest to „funkcjonujący w świadomości społecznej uproszczony i zabarwiony wartościująco obraz rzeczywistości”<sup>17</sup>. Można wobec tego stwierdzić, że w swojej konstrukcji stereotyp jest pewnym wzorcem ocennym używanym przez ludzi (najczęściej w sposób krzywdzący dla pewnych grup społecznych). Również w prawie występują wartościujące obrazy rzeczywistości, mają one jednak odwrotną funkcję niż powszechnie znane stereotypy, wyznaczają bowiem standard odpowiedniej staranności (np. standard racjonalnego uczestnika obrotu gospodarczego). Wydaje się, że skoro sztuczna inteligencja jest w stanie posługiwać się

<sup>14</sup> J. Feast, *Root Out Bias at Every Stage of Your AI-Development Process*, Harvard Business Review/el., 2020, s. 2 <<https://hbr.org/2020/10/root-out-bias-at-every-stage-of-your-ai-development-process>> [dostęp: 23.03.2023].

<sup>15</sup> Ibidem, s. 2–3.

<sup>16</sup> Ibidem.

<sup>17</sup> Hasło *Stereotyp*, [w:] *Słownik języka polskiego PWN* <<https://sjp.pwn.pl/slowniki/stereotyp.html>> [dostęp: 23.03.2023].

wzorcami ocennymi w ramach swojego działania (wytworzonymi na podstawie danych), to – przy odpowiedniej konstrukcji takiego wzorca – byłaby w stanie rozstrzygać o spełnieniu wzorca ocennego w danym stanie faktycznym. Powstawanie krzywdzących stereotypów w działaniu sztucznej inteligencji opiera się na nieodpowiednim lub zbyt wąskim doborze danych. Wracając do poruszanego już błędu naturalistycznego, być może przy wprowadzeniu odpowiednich założeń modelujących działanie AI (założenie o racjonalnym prawodawcy czy o racjonalnym uczestniku obrotu) można by przekuć wadę systemów AI w narzędzie usprawniające wykładnię przepisów prawnych. Aby to rozważyć, należy omówić koncepcję racjonalności prawodawcy.

#### 4. Założenie o racjonalności prawodawcy

Opierając się na koncepcji interpretacji humanistycznej, zagadnienie zamiaru prawodawcy wydaje się absolutnie kluczowe do ustalania znaczenia tekstów prawnych. Należy wobec tego przeanalizować, w jaki sposób tę wolę można ustalić i kim (lub czym) w istocie jest prawodawca. Wydawać by się mogło, że taki model interpretacji wchodzi na pole psychologii i że najlepszym sposobem zbadania celu byłyby badania empiryczne w postaci ankietowania parlamentarzystów lub przeprowadzania z nimi wywiadów<sup>18</sup>. Prawnicy jednak postępują inaczej i milcząco przyjmują, że prawodawca nie jest osobą psychofizyczną, ale pewnym wyidealizowanym modelem myślowym. Koncepcja interpretacji humanistycznej zakłada racjonalność. Zasada racjonalności, za Nowakiem, polega na tym, że: „jeżeli X uznaje wiedzę W, która jest: (1) niesprzeczna, (2) systemem (tzn. zawiera swe konsekwencje logiczne) oraz X żywi preferencje, które są: (1) asymetryczne i (2) przechodnie, to X podejmuje taką spośród możliwych (w świetle jego wiedzy) czynności, która niezawodnie prowadzi (wedle tej wiedzy) do najwyższej preferowanego stanu rzeczy”<sup>19</sup>.

Wobec tego prawodawca racjonalny w najogólniejszej wersji racjonalności musi posiadać wiedzę, która pozostaje niesprzeczna i tworzy system rządzący się konsekwencjami logicznymi, to znaczy: „jeśli X uznaje zdanie Z oraz zdanie Z' wynika logicznie z Z, to X uznaje zdanie Z'”<sup>20</sup>. Należy przy tym odnotować, że w celu zachowania zwięzłości tekstu prawnego prawodawca nie powinien umieszczać w aktach prawnych norm, które są logicznymi konsekwencjami innych norm<sup>21</sup>. Normodawca musi jednocześnie podejmować decyzje, które w świetle jego wiedzy doprowadzą do preferowanego stanu rzeczy. W odniesieniu do tychże preferencji asymetryczność oznacza, że jeśli prawodawca preferuje stan  $R_1$  przed  $R_2$ , to jednocześnie nie preferuje

<sup>18</sup> L. Nowak, op. cit., s. 35.

<sup>19</sup> Ibidem, s. 38–43.

<sup>20</sup> Ibidem.

<sup>21</sup> S. Wronkowska, op. cit., s. 64–65.

$R_2$  przed  $R_1$ . Przechodniość polega z kolei na tym, że jeśli prawodawca preferuje stan rzeczy  $R_1$  przed  $R_2$ , a  $R_2$  przed  $R_3$ , to jednocześnie preferuje  $R_1$  przed  $R_3$ <sup>22</sup>.

Na tym jednak nie koniec, ponieważ doktryna przyjmuje również inne, dalej idące założenia, zbliżając się do konstrukcji normodawcy doskonałego. Istnieją różne rozwinięcia modelu idealnego, tutaj należy przedstawić ten najbardziej podstawowy, który prezentuje założenia w sposób najbardziej klarowny (i jak się wydaje – możliwy do zaadaptowania na gruncie systemów AI). Oprócz wspomnianych już atrybutów racjonalnego prawodawcy zakłada się, że prawodawca ma pełną znajomość reguł danego języka etnicznego, ma pełną znajomość stanu prawnego, a także dysponuje najbardziej aktualną wiedzą empiryczną. Wydaje się, że ma to służyć jak najpełniejszej realizacji sprawiedliwościowej funkcji prawa. Oczywiście jest, że prawo tworzą ludzie, którzy nie są idealni i popełniają błędy. Nawet ich wyczerpana praca nad aktami prawnymi nie daje gwarancji, że nie pojawią się w nich niedopatrzania tworzące luki prawne lub generujące niesprawiedliwe czy nieracjonalne efekty wykładni językowej. Właśnie w tym celu stosuje się założenie o jak najlepszych kompetencjach prawodawcy, aby w toku wykładni, w wymagających tego sytuacjach, przełamać wykładnię językową, która byłaby niezgodna z intencją prawodawcy. W końcu, prawodawca doskonały ma przed sobą postawiony warunek najważniejszy, a konkretnie warunek ocenny, stanowiący, że prawodawca doskonały powinien dążyć zawsze do rozwiązań „godziwych”, tj. takich, które są aprobowane na gruncie ideologii obowiązującej w danym ustroju<sup>23</sup>.

Reasumując, model idealnego prawodawcy składa się z założeń dotyczących racjonalności oraz z założeń modelujących. W sumie składają się one na poniższe warunki:

- „1) wiedza  $W_L$  prawodawcy L jest niesprzeczna
- 2) wiedza  $W_L$  prawodawcy L jest systemem, to znaczy należą do niej jej własne konsekwencje logiczne
- 3) wiedza  $W_L$  obejmuje wszystkie reguły języka, w jakim formułuje przepisy prawodawca L
- 4) wiedza  $W_L$  jest najlepiej uzasadnioną wiedzą z punktu widzenia aktualnego stanu nauki
- 5) preferencje prawodawcy L są asymetryczne [...]
- 6) preferencje prawodawcy L są przechodnie [...]
- 7) oceny prawodawcy L wyznaczające jego preferencje są pełnym systemem słusznych moralnie ocen<sup>24</sup>.

Wydaje się, że założenia 1–6 mogą być zaadaptowane do tworzenia programu sztucznej inteligencji orzekającego o zamiarze racjonalnego prawodawcy. Miałyby to być program wspomagający decyzję sędziego w czasie dokonywania wykładni

<sup>22</sup> L. Nowak, op. cit., s. 38–43.

<sup>23</sup> Ibidem, s. 43–46.

<sup>24</sup> Ibidem, s. 53–54.

prawa na etapie wykładni funkcjonalnej. Powyższe warunki wyznaczałyby metodologiczną podstawę oceny, czy dana norma prawna odzwierciedla racjonalność prawodawcy. Należy tutaj rozważyć postawiony na wstępie problem badawczy, a mianowicie czy sztuczna inteligencja byłaby w stanie decydować o kształcie racjonalności prawodawcy jako samowystarczального modelu, idealnie odtworzonego ze struktury głębokiej prawoznawstwa, czy jedynie wykorzystywałaby konstrukcję wyinterpretowaną z sądowego orzecznictwa, a więc sędziowskie rozumienie założenia o racjonalnym prawodawcy. Wydaje się, że najbardziej odpowiednia byłaby droga pośrednia, tj. sygnalizowane już zespolenie dorobku teorii prawa z efektami stosowania prawa. Podstawą działania programu AI byłby założenia na temat racjonalności odtworzone przez Nowaka, a uściśleniem modelu byłyby rezultaty faktycznego zoperacjonalizowania tych założeń, czyli informacje pochodzące z orzeczeń sądowych, w których zastosowano omawiany model normodawcy.

Należy jednak pamiętać, że jest to założenie „milcząco zaakceptowane przez prawników”<sup>25</sup>, nie można zatem bazować wyłącznie na orzeczeniach, w których to założenie jest wysłowione. Tu pojawia się zadanie dla AI – program musiałby analizować orzecznictwo i wybierać takie wyroki, w których faktycznie zastosowano założenie o prawodawcy racjonalnym, a co więcej – zastosowano je zgodnie z założeniami wypracowanymi przez teorię prawa (tzn. zastosowano właściwe założenie o prawodawcy racjonalnym, a nie jedynie wyobrażenie tego założenia). Na podstawie takiego zbioru danych program AI mógłby uściślać model normodawcy i orzekać o woli prawodawcy w procesie wykładni prawa. Trzeba też pamiętać, że powyższe warunki zawierają odesłania do systemów pozaprawnych, które musiałyby być ujęte w programie, tj. reguły języka czy aktualny stan nauki. Taki program musiałby również przedstawić uzasadnienie otrzymanego wyniku. Natomiast ostateczna decyzja co do zaakceptowania propozycji programu należałaby do sędziego – tutaj znaczenie ma założenie 7. To sędzia jest osobą wyposażoną w kompetencje moralne oraz kompetencje do rozstrzygania sporów sądowych i to on oceniałby, czy decyzja programu znajduje uzasadnienie w ocenach przypisywanych prawodawcy, a więc rozstrzygał, czy dana propozycja jest „godziwa”.

## 5. Sztuczna inteligencja a prawo do sprawiedliwego sądu

Jak zwraca uwagę Maria Dymitruk, wprowadzenie systemów sztucznej inteligencji do sądownictwa musi być rozważane w kontekście prawa do rzetelnego procesu sądowego wynikającego z art. 6 Europejskiej Konwencji Praw Człowieka<sup>26</sup>. Z jednej

<sup>25</sup> Ibidem, s. 64.

<sup>26</sup> Konwencja o ochronie praw człowieka i podstawowych wolności z dnia 4 listopada 1950 r. (Dz.U. z 1993 r., Nr 61, poz. 284), art. 6.



strony programy AI mogą doskonale realizować postulaty przyspieszenia postępowania sądowego i zredukowania zalegających spraw, gdyż dysponują ogromnym potencjałem obliczeniowym. Z drugiej jednak pojawia się problem uzasadnienia wyroku sądowego, ponieważ możliwość zrozumienia faktów, argumentów i interpretacji prawniczych przemawiających za danym rozstrzygnięciem jest fundamentem realizacji prawa do sprawiedliwego sądu, chroniącym przed arbitralnością<sup>27</sup>. W uzasadnieniu są przedłożone faktyczne i prawne podstawy rozstrzygnięcia sprawy, a także rozumowania prawnicze. Uzasadnienie jest również dowodem przejrzystości systemu sądownictwa, przez co zwiększa zaufanie do władzy publicznej. Jest to także widoczny wynik procesów decyzyjnych sędziego, będący podstawą do kontrargumentacji i ewentualnego zaskarżenia orzeczenia<sup>28</sup>.

Przy działaniu algorytmów sztucznej inteligencji pojawia się znaczący problem: częstokroć programy, szczególnie te oparte na uczeniu maszynowym (*machine learning*), są na tyle skomplikowane, że człowiek nie jest w stanie wyjaśnić, dlaczego ich działanie dało takie, a nie inne rezultaty<sup>29</sup>. Wobec tego klasyczny algorytm AI może być zaprzeczeniem prawa do sprawiedliwego sądu, gdyż jego działanie jest nieprzejrzyste i nieuzasadnialne. Uznanie wyroku AI musiałoby być w takim wypadku oparte jedynie na autorytecie programu i zaufaniu do jego nieomyślności (choć to zaufanie już rzeczywiście istnieje, co zostanie jeszcze omówione). Wobec tego klasyczne programy AI nie mogą być stosowane w sądownictwie. Powstał jednak koncept „wyjaśnialnej sztucznej inteligencji” (*explainable artificial intelligence* – XAI). Taki program musiałby każdorazowo wyjaśniać swoje działania w taki sposób, aby użytkownicy mogli się utwierdzić w racjonalności decyzji algorytmu<sup>30</sup>. Jest to wymagające zadanie, aczkolwiek tylko taki system zapewni sprawiedliwość sądownictwa.

## 6. Zbytnie zaufanie do systemów AI

W kontekście użycia sztucznej inteligencji jako narzędzia wspomagającego podejmowanie decyzji Dymitruk zwraca uwagę na psychologiczne aspekty używania systemów AI. Jak bowiem wynika z badań Komisji Ekspertów ws. Internetowych Pośredników (MSI-NET), rola człowieka może się sprowadzać do wyłącznie formalnego zatwierdzenia rezultatu działania programu, gdyż AI niejednokrotnie działa w sposób niejasny i skomplikowany, a ludziom częstokroć brakuje wiedzy i umiejętności do sprawdzenia poprawności jej wyników, dlatego ludzie są skłonni

<sup>27</sup> M. Dymitruk, *The Right to a Fair Trial in Automated Civil Proceedings*, „Masaryk University Journal of Law and Technology” 2019, vol. 13(1), s. 31–40.

<sup>28</sup> Ibidem.

<sup>29</sup> Ibidem.

<sup>30</sup> Ibidem.

do zatwierdzania działań AI bez ich kontrolowania<sup>31</sup>. Ponadto eksperymenty psychologiczne wskazują, że ludzie są skłonni tak dalece ufać sztucznej inteligencji, iż akceptują nawet oczywiście niepoprawne rezultaty działań algorytmów (np. na polecenie robota podlewają rośliny sokiem pomarańczowym)<sup>32</sup>. To zbytne zaufanie do sztucznej inteligencji może się również przekładać na działanie prawników, mimo że są oni specjalistami w swojej dziedzinie. Jaap Dijkstra w swoim eksperymencie udowodnił, że prawnicy, mając wybór między argumentacją programu a argumentacją człowieka, chętniej wybierają opinię programu, nawet gdy obie są identyczne, uważając tę pierwszą za bardziej obiektywną i racjonalną<sup>33</sup>. Jego badania wskazują, że ludzie, używając sztucznej inteligencji, chcą całkowicie uniknąć procesu podejmowania decyzji, zamiast poprawić jego jakość<sup>34</sup>. Wobec tego programy jedynie wspomagające decyzje stają się faktycznie równe w swej autonomii z algorytmami podejmującymi decyzje samodzielnie, a rola człowieka w sądownictwie może stać się pozorna<sup>35</sup>. Należy zdecydowanie zmienić to ludzkie nastawienie i wprowadzając systemy wspomagające decyzje, zwracać uwagę, że mają one ułatwiać podejmowanie decyzji, a nie odciążać decydentów od ciężaru ich podejmowania<sup>36</sup>.

## Podsumowanie

Przechodząc do konkluzji, z punktu widzenia metodologii zarówno sztuczna inteligencja, jak i nauki prawne opierają się na logice i regułach wnioskowania. Pozwala to na adaptowanie programów AI do prawa. W niniejszej pracy, opierając się na faktach, że AI potrafi wytwarzać na potrzeby swojego działania wzorce ocenne, zaproponowano stworzenie programu odtwarzającego założenie o racjonalnym prawodawcy. Wydaje się, że aby taki program był skuteczny, należałoby najpierw osadzić w nim warunki dotyczące racjonalnego prawodawcy wypracowane przez teorię prawa, a następnie zapewnić materiał do nauki w postaci orzecznictwa, w którym zastosowano założenie o racjonalnym prawodawcy (należy przy tym pamiętać, że oma-

<sup>31</sup> Committee of experts on internet intermediaries (MSI-NET), *Algorithms and Human Rights – Study on the Human Rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*, Council of Europe Study DGI (2017)12, s. 7–10.

<sup>32</sup> M. Salem, G. Lakatos, F. Amirabdollahian, K. Dautenhahn, *Would You Trust a (Faulty) Robots: Effects of error, task type and personality on human-robot cooperation and trust*, 2015, za: M. Dymitruk, op. cit., s. 31–33.

<sup>33</sup> J. Dijkstra, *Legal Knowledge-Based Systems: The Blind Leading the Sheep?*, *International Review of Law, Computers & Technology* 2001, nr 15(2), s. 119–128, za: M. Dymitruk, op. cit., s. 31–33.

<sup>34</sup> M. Dymitruk, op. cit., s. 31–33.

<sup>35</sup> Ibidem.

<sup>36</sup> Jak podkreśla również cytowana wcześniej Komisja Ekspertów: „Indeed, it is not the algorithms themselves but the decision-making processes around algorithms that must be scrutinised in terms of how they affect human rights” [W istocie to nie same algorytmy, ale procesy decyzyjne wokół nich należy przeanalizować w związku z ich wpływem na prawa człowieka – tłum. K.B.] (MSI-NET, s. 8).

wiane założenie należy do struktury głębokiej prawoznawstwa, stąd nie za każdym razem musi być wysłowne przez sąd – sztuczna inteligencja powinna rozpoznawać użycie tego założenia na podstawie zaimplementowanych warunków opracowanych przez teorię prawa). Aby jednak program odtwarzał jak najlepszy model racjonalności prawodawcy, a nie jedynie stereotypowe wyobrażenie o tymże, program każdorazowo musiałby konfrontować obraz wytworzony na podstawie orzeczenia z warunkami wypracowanymi przez teorię prawa. Efekt działania systemu byłby wskazaniem dla sędziego, potwierdzającym lub modyfikującym rezultat wykładni. Zawsze jednak to sędzia musi podejmować ostateczną decyzję, pamiętając, że to do niego należy rozstrzygnięcie, czy rezultat wykładni jest godziwy. Jednocześnie omawiany program musiałby tworzyć uzasadnienie wskazanego rezultatu wykładni, jest to bowiem konieczne dla uczynienia zadość prawu do sprawiedliwego sądu. Sędzia musi przy tym pamiętać, że AI z różnych powodów może popełnić błąd i należy mieć ograniczone zaufanie do wyników działania programu.

## BIBLIOGRAFIA

- Dymitruk, M. (2019). *The Right to a Fair Trial in Automated Civil Proceedings*. „Masaryk University Journal of Law and Technology” 13(1): 27–44.
- Lai, L., Świerczyński, M. (red.). (2020). *Prawo sztucznej inteligencji*. Legalis/el.
- McCarthy, J. (2004). *What is Artificial Intelligence?* Stanford.
- Nowak, L. (1973). *Interpretacja prawnicza. Studium z metodologii prawoznawstwa*. Warszawa.
- Russell, S., Norvig, P. (red.). (2020). *Artificial Intelligence: A modern approach*. London.
- Wronkowska, S. (2002). *Podstawowe pojęcia prawoznawstwa*. Poznań.

