

## REVOLUȚIA INTELIGENȚEI ARTIFICIALE ÎN SISTEMUL JUDICIAR PENAL. EXTRAGEREA ARGUMENTELOR ȘI EVALUAREA RISCULUI

## THE REVOLUTION OF ARTIFICIAL INTELLIGENCE IN THE CRIMINAL JUSTICE SYSTEM. ARGUMENT MINING AND RISK ASSESSMENT

**Cristian-Alin BADEA<sup>1</sup>**

Avocat - Baroul Bihor

### ABSTRACT

In the near future we will rely on algorithms in most of the decisions we make. We already do, but we are not aware enough. The truth is that in recent decades, this is an incredible leap in technology, but it may be the last man will ever do. We will certainly not know, until the last moment when we reach the singularity of AI, so controversial among specialists. Of course, although a cautious approach is needed, I consider it auspicious the use of predictive algorithms to reduce the crime rate, to increase the probability that a sentence will achieve its fundamental purpose for which it was given, to identify and create argumentative structures from extensive amounts of information. It is the moment when humanity intertwines with its reverse, science, and together will form a new era.

**KEYWORDS:** *Artificial Intelligence and Law; criminal risk assessment; argumentative constructions; information extraction; ECHR.*

### REZUMAT

În viitorul apropiat ne vom baza pe algoritmi în cele mai multe decizii pe care le vom lua. Deja o facem, dar nu conștientizăm îndeajuns. Adevărul este că, în ultimele decenii, acest fapt reprezintă un salt incredibil al tehnologiei care poate fi și ultimul pe care omul îl va face vreodată. Cu siguranță nu vom ști până în ultimul moment în care vom atinge singularitatea AI, atât de controversată printre specialiști. Desigur, cu toate că este necesară o abordare prudentă, consider de bun augur utilizarea algoritmilor predictivi pentru a reduce rata infracționalității, pentru a crește probabilitatea ca o sentință să își atingă scopul fundamental pentru care a fost dată și pentru a identifica și crea structuri argumentative din vaste cantități de informații. Este momentul în care umanitatea se împletește cu reversul său, știința, și împreună vor forma o nouă eră.

**CUVINTE-CHEIE:** *Inteligența Artificială și Drept; evaluare risc penal; construcții argumentative; extragere informații; CEDO.*

### CUPRINS

I. Introducere .....	123
II. Clasificări ale inteligenței artificiale .....	124
1. Machine learning .....	124
2. Deep learning .....	126
3. Structura neuronală biologică versus cea artificială .....	126

<sup>1</sup> Email: Badea.cristian.alin@gmail.com.

III. Aplicații ale algoritmilor predictivi în analiza textuală. Extragerea informațiilor și a argumentelor .....	129
1. Categoriile ale procesării lingvistice .....	130
2. Teorema lui Bayes .....	134
3. Clasificatorul Naiv Bayesian .....	136
IV. Implicații ale inteligenței artificiale în evaluarea riscului de recidivă și în condamnarea penală .....	138
1. Instrumente de evaluare a riscului penal .....	140
2. Teoria pedepsei .....	144
V. Algoritmii predictivi și respectarea drepturilor fundamentale ale omului .....	147
1. Prezența biasului în sistemele informatice .....	147
2. Respectarea drepturilor fundamentale ale omului .....	149
VI. Concluzii .....	151

## I. INTRODUCERE

Întreaga istorie a umanității a fost presărată de-a lungul său cu descoperiri științifice fabuloase și opere de artă cutezătoare, fiind reflexia unor suflete înalte înclinate spre progres.

Alături de *Quantum Computing*, *Nanotechnology*, *Neuromorphic Computing* și *Neuralink*, probabil cel mai spectaculos succes al omenirii este descoperirea *Artificial Intelligence (AI)* - Inteligenței Artificiale, disciplină propusă oficial în anul 1956 la conferința de vară susținută la Dartmouth College, în Hanovra, New Hampshire, și, probabil, este cea care va deveni ultima frontieră a umanității când vom atinge singularitatea AI.

Enciclopedia Britannica definește AI astfel, *”este capacitatea unui calculator digital sau a unui robot controlat de un calculator de a efectua sarcini asociate în mod obișnuit cu ființele inteligente.”*<sup>2</sup>

Oxford Dictionary arată că AI este *”teoria și dezvoltarea sistemelor informatice capabile să îndeplinească sarcini care necesită în mod normal inteligență umană, precum percepția vizuală, recunoașterea vorbirii, luarea deciziilor și traducerea dintre limbi.”*<sup>3</sup>

Fundamentul ideilor unor mașinării inteligente, conexiuni neuronale artificiale și programarea unui calculator digital au fost teoretizate de Alan Turing<sup>4</sup>, alături de alți cercetători, printre care și Richard Feynman prin lucrarea sa, *The connection machine*<sup>5</sup>. Algebra Computațională a fost implementată de John Von Neumann prin teoria *Self-replicating automata*<sup>6</sup>, Kurt Gödel a lucrat la teorema de incompletitudine<sup>7</sup> iar Solomon Lefschetz a contribuit prin lucrarea *Algebraic Topology*<sup>8</sup>. Noțiunile de cibernetică și feedback<sup>9</sup> au fost abordate de profesorul MIT, Norbert Wiener.

O personalitate remarcabilă, Alan Turing a demonstrat<sup>10</sup> încă din anul 1936 *universalitatea calculului* de către un calculator și că o mașinărie poate efectua calcule într-un mod arbitrar dar și într-o multitudine

2 Disponibil online: <https://www.britannica.com/technology/artificial-intelligence>, accesat la data 07.08.2020.

3 Disponibil online: [https://www.lexico.com/definition/artificial\\_intelligence](https://www.lexico.com/definition/artificial_intelligence), accesat la data 07.08.2020.

4 A. Turing, *Computing Machinery and Intelligence*, în "Mind", vol. LIX, nr. 236, 1950, pp. 433-460.

5 R. Feynman în W. D. Hillis, "Physics Today", vol. 42, nr. 2, 1989, disponibil [online](#) - accesat la data 07.09.2020.

6 J. V. Neumann, *Theory of Self-Replicating automata*, Ed. University of Illinois Press, 1966, disponibil [online](#) - accesat la data 07.08.2020.

7 R. Panu, *Gödel's Incompleteness Theorems*, The Stanford Encyclopedia of Philosophy, 2020, disponibil [online](#) - accesat la data 07.09.2020.

8 S. Lefschetz, *Algebraic Topology*, American Mathematical Society, 1942, în M. Skilton, F. Hovsepian, *The 4th Industrial Revolution. Responding to the Impact of Artificial Intelligence on Business*, Ed. Macmillan Palgrave, 2018, p. 33.

9 N. Wiener *Invents Cybernetics*, material disponibil online: [https://www.livinginternet.com/i/ii\\_wiener.html](https://www.livinginternet.com/i/ii_wiener.html), accesat la data 08.08.2020.

10 M. Tegmark, *Viața 3.0. Omul în epoca inteligenței artificiale*, Ed. Humanitas, București, 2019, p. 76.

de feluri. De asemenea, a mai arătat că în ipoteza în care două calculatoare dețin aceleași resurse pot face aceleași tipuri de calcule, acest fapt demonstrând *universalitatea* procesului computațional, independența de substrat (materia din care este compus calculatorul). În anul 1950, publică lucrarea *Computing Machinery and Intelligence* în revista *Mind*, unde pune următoarele probleme: "Pot mașinile gândi?", "Pot mașinile să fie nediferențiate lingvistic față de oameni?" și, dacă acestea ar putea câștiga Testul Turing care funcționează ca un joc, *The imitation game*.<sup>11</sup>

Lucrarea de față își propune să abordeze din unghiuri diferite domeniul algoritmilor inteligenți. Pentru început se va face o scurtă introducere în taxonomia tipurilor de inteligență artificială. Ulterior, voi arăta felul în care aceștia impactează extragerea informațiilor și a argumentelor relevante dintr-un text cât și modul de funcționare din punct de vedere matematic a unuia dintre cei mai utilizați algoritmi de clasificare a documentelor. Apoi, se vor avea în vedere alte tipuri de softuri remarcate cu precădere în psihometrică pentru evaluarea riscului de recidivă sau a pericolului social. În final, în scopul contrabalansării cercetării, dezavantajele vor fi puse în perspectivă cu referire la atingerile aduse drepturilor și libertăților fundamentale ale omului.

## II. CLASIFICĂRI ALE INTELIGENȚEI ARTIFICIALE

### 1. Machine Learning

Acesta este studiul algoritmilor care permit programelor unui calculator să se îmbunătățească automat prin experiență.<sup>12</sup> Într-o altă definiție, Kevin P. Murphy, Senior Staff Research Scientist la Google Research, arată că Machine Learning este "învățarea mașinii unui set de metode care pot detecta automat tiparele datelor, și apoi să folosească modelele descoperite pentru a prezice viitorul acestora sau pentru a efectua alte tipuri de luare a deciziilor în condiții de incertitudine (cum ar fi planificarea modului de a colecta mai multe date)."<sup>13</sup>

În acest moment suntem în plină expansiune a erei *big data*. Spre exemplu, sunt 1.7 miliarde de pagini de internet<sup>14</sup>; videoclipuri cumulate ajungând la durata de 500 de ore sunt încărcate la fiecare minut pe Youtube; 5 miliarde de videoclipuri sunt vizionate în fiecare zi în întreaga lume pe platforma Youtube.<sup>15</sup> Genomii a 1000 de oameni, au fost secvențiați de diverse laboratoare din lume pentru a ajuta la predicția bolilor pe care o persoană le poate suferi în timpul vieții, și a lua măsuri pentru a le evita sau reduce impactul acestora prin utilizarea de medicamente și tratamente personalizate<sup>16</sup>. Walmart gestionează mai mult de 1 milion de tranzacții pe oră și are baze de date care conțin mai mult de 2,5 petabytes de informații<sup>17</sup> și așa mai departe.

Aceste imense cantități de date necesită metode automate de analiză, cadru în care intervine conceptul de Machine Learning. Să ne imaginăm că ne dorim de la un calculator să învețe să recunoască singur anumite imagini (de pildă, fețe al unor persoane urmărite internațional pentru infracțiuni deosebit de grave, ori în domeniul medical, să pună un diagnostic de afecțiune malignă pe baza unor radiografii), să conducă singur un autovehicul, să clasifice în grupe, date voluminoase chiar și pentru un calculator obișnuit, să traducă unele texte ori să recunoască vocea unei persoane. Acestor sarcini speciale algoritmi clasici nu le pot face față, de aceea, au fost introduși noii algoritmi, care fie învață singuri, fie sunt antrenați de a recunoaște anumite tipare relevante

11 G. Oppy, D. Dowe, *The Turing Test*, "The Stanford Encyclopedia of Philosophy", 2019, disponibil [online](#) - accesat la data 05.08.2020.

12 T. Mitchell, *Machine Learning*, Ed. McGraw-Hill, New York, USA, 1997, p. 1.

13 K. P. Murphy, *Machine Learning A Probabilistic Perspective*, Ed. The MIT Press Cambridge, Massachusetts London, England, 2012, p. 1.

14 Disponibil online: <https://www.internetlivestats.com/total-number-of-websites/>, accesat la data 10.08.2020.

15 Disponibil online: <https://www.omnicoreagency.com/youtube-statistics/>, accesat la data 10.08.2020.

16 Emc Education Services, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, Ed. Wiley, 2015, p. 4.

17 K. Cukier, *Data, Data Everywhere: A Special Report on Managing Information*, "The Economist", 2010, disponibil [online](#) - accesat la data 10.08.2020.

în cadrul obiectivelor cu care au fost însărcinați.

Machine Learning este o subdiviziune a AI care se împarte în alte subclase în funcție de felul în care dorim să utilizăm diferite softuri, după cum urmează:

**Învățarea supravegheată** – Acesta este cel mai utilizat mod de învățare. Să presupunem că ne dorim de la un algoritm să recunoască o anumită imagine. Pentru început îi vom prezenta ca date intrare (*input*) fotografii cu 2 persoane diferite, fiind interesați doar de descoperirea uneia dintre ele, iar acest rezultat să fie redat sub formă de date de ieșire (*output*). Apoi, îi vom cere să identifice în mod corect acea persoană pe care o căutăm. Prin urmare, în modelul de față, răspunsul corect este oferit în stadiul de pregătire.<sup>18</sup>

Din această clasă de învățare fac parte ca sub-categorii *clasificarea și regresia*, iar ca tipuri de algoritmi întâlnim: arbori de decizie, rețele neuronale artificiale, mașini cu suport vectorial (*SVM*) și învățarea Bayesiană.<sup>19</sup>

**Învățarea nesupravegheată** - este o clasă de învățare automată unde categoriile de informații nu sunt furnizate. În schimb, calculatorul însuși încearcă să realizeze ce distinge anumite date de altele. Așadar, după ce algoritmului i se prezintă mai multe imagini, de mai multe categorii, va determina diferențele dintre ele redând-o pe cea corectă sub formă de output. Un avantaj major este că datele nu trebuie să fie structurate în categorii în stadiul de pregătire, făcând accesibile analizei cantități mari de informații.

Aceasta este o problemă mult mai puțin definită, deoarece nu ni se evidențiază tipurile de modele pe care să le căutăm (spre deosebire de învățarea supravegheată), unde putem compara predicția noastră  $y$  pentru o informație  $x$  la valoarea observată.<sup>20</sup>

Sub-categorii ale acestei forme de învățare sunt *gruparea și asocierea*.

**Învățarea prin întărire** - Acest model încercă să se adapteze în permanență prin încercare și eroare. Modificându-și strategia continuu, primește un feedback din analiza datelor și este ghidat spre rezultatul cel mai bun.<sup>21</sup> Presupunem că obiectivele algoritmului pot fi definite printr-o funcție de recompensă care atribuie o valoare numerică - rambursare imediată - pentru fiecare acțiune distinctă pe care agentul o poate lua la fiecare stare.<sup>22</sup> Inițial, algoritmul încearcă aleatoriu un tipar anume. Dacă în mod întâmplător este cel corect, acesta se va consolida prin recompensă, până în momentul în care învață în mod corect ceea ce ne dorim de la el.

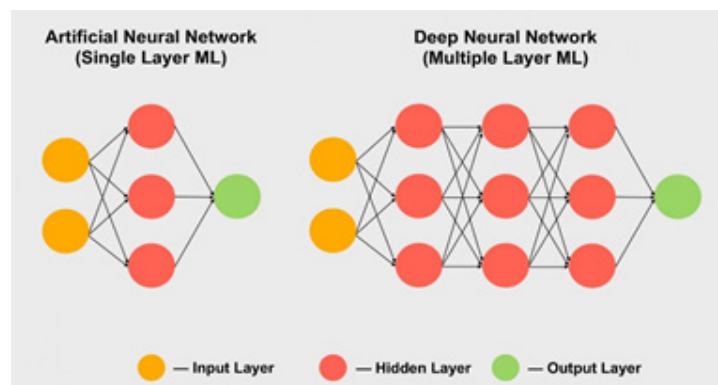


Fig. 1 Comparație ML- DL<sup>23</sup>

18 B. Dupont, Y. Stevens, H. Westermann, M. Joyce, *Artificial Intelligence in the Context of Crime and Criminal Justice*, "Canada Research Chair in Cybersecurity International Centre for Comparative Criminology", Université de Montréal, 2018, p. 18.

19 L. E. Mureșan, *Învățarea Bayesiană*, Lucrare de Diplomă, Universitatea Babeș- Bolyai Cluj-Napoca, Facultatea de Matematică și Informatică, Cluj Napoca, 2011, p. 9.

20 *Ibidem*.

21 J. Hurwitz, D. Kirsch, *Machine Learning For Dummies*, IBM Limited Edition, Ed. John Wiley & Sons, 2018, p. 16.

22 T. Mitchell, *op. cit.*, p. 367.

23 Disponibil online: <https://www.techspot.com/article/2048-machine-learning-explained/>, accesat la data 15.08.2020.

## 2. Deep Learning

Este considerată o subcategorie a Machine Learning, care folosește rețele artificiale de neuroni în straturi succesive - *Artificial Neural Networks*, încercând să imite modul în care funcționează creierul uman în rezolvarea problemelor. Termenul de Deep Learning este utilizat atunci când există multiple straturi ascunse în cadrul unei rețele neuronale.<sup>24</sup>

Așadar, diferența fundamentală dintre Machine Learning și Deep Learning o reprezintă chiar aceste straturi ascunse organizate în structuri ierarhice<sup>25</sup> - *hidden layers* - care reprezintă și în zilele noastre un secret pentru cercetători, numindu-se *black box*,<sup>26</sup> datorită faptului că odată ce au fost introduse datele de input și se rulează secvența către datele de ieșire, dacă intervine o eroare sau algoritmul intenționează să se coreleze pentru a fi mai eficient prin metoda *backpropagation*, inginerii nu înțeleg deocamdată în totalitate cum se ajustează reciproc aceste straturi ascunse.

Rețelele neuronale artificiale conțin alte subcategorii de arhitecturi, cum sunt, de pildă: *deep neural networks*, *deep belief networks*, *recurrent neural networks*,<sup>27</sup> *feedforward neural network* etc.

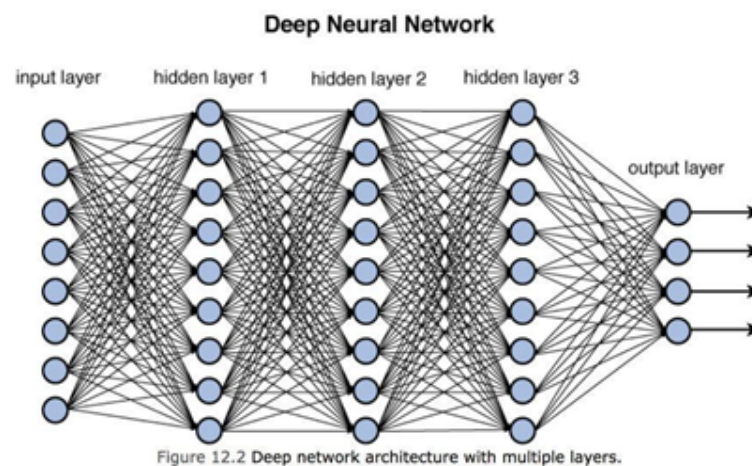


Fig. 2 Rețea Neuronala Artificiala cu mai multe straturi ascunse<sup>28</sup>

## 3. Structura neuronală biologică versus cea artificială

O rețea neuronală artificială standard este formată din multe procesoare simple conectate, numite neuroni, fiecare producând o secvență de activări valorice reale. Neuronii de intrare se activează prin senzori care percep mediul, iar alți neuroni se activează prin conexiuni ponderate de la neuronii anteriori activi. În funcție de problemă și de modul în care sunt conectați neuronii, un astfel de comportament poate necesita lanțuri cauzale lungi ale etapelor de calcul în care fiecare etapă transformă (adesea într-un mod non-liniar) activarea agregată a rețelei.<sup>29</sup>

24 J. Hurwitz, D. Kirsch, *op. cit.*, p. 18.

25 B. Dupont, *op. cit.*, p. 15.

26 B. Robbins, *Machine Learning: How Black is This Beautiful Black Box*, material disponibil [online](#) - accesat la data 15.08.2020.

27 B. M. Gupta, S. M. Dhawan, *Deep Learning Research: Scientometric Assessment of Global*, "Emerging Science Journal", vol. 3, nr. 1, 2019, p. 23.

28 Disponibil online: <https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964>, accesat la data 16.08.2020.

29 J. Schmidhuber, *Deep Learning in Neural Networks: An Overview*, în "Neural Networks Journal", 2014, p. 4.

De obicei, neuronii sunt organizați în straturi, iar acestea vor fi cu atât mai numeroase cu cât problema este mai complexă. Diferite straturi pot efectua transformări în funcție de inputul oferit. Semnalele se deplasează de la primul strat (stratul de intrare), până la ultimul strat (stratul de ieșire), după parcurgerea straturilor de mai multe ori.

Având în vedere capacitățile extraordinare ale creierului uman, cercetătorii din domeniul neuroștiințelor și ale științelor exacte au încercat să reproducă părți componente ale acestuia sub formă artificială și să îl îmbunătățească. Astfel, după unii cercetători,<sup>30</sup> creierul uman este o structură paralelă de procesare și stocare a informației compusă din aproximativ 86.1 miliarde de neuroni interconectați (la o persoană adultă), și în jur de 100 miliarde de neuroni, după alți cercetători.<sup>31</sup> Această structură celulară complexă necesită 20% din energia totală a organismului, deși reprezintă doar 2% din masa totală a acestuia.<sup>32</sup>

Un neuron are rolul fundamental de a primi, conduce, procesa și transmite diferite semnale electrice primite de la organele de simț, diferiți proprioceptori sau de la alți neuroni.<sup>33</sup> Structura unui neuron este formată din:

- **Corpul celular (soma);**

- **Axon**- numit și fibră nervoasă, porțiune a unei celule nervoase (neuron) care transportă impulsurile nervoase departe de corpul celular. Un neuron are de obicei un axon care îl conectează cu alți neuroni, cu celulele musculare sau ale glandelor. Unii axoni pot fi destul de lungi, ajungând, de exemplu, de la măduva spinării până la vârful piciorului. Majoritatea axonilor sunt incluși într-o teacă de mielină, ceea ce crește viteza de transmitere a impulsului; unii axoni mari pot transmite impulsuri cu viteze de până la 90 de metri pe secundă.<sup>34</sup>

- **Arbore dendritic** – Dendritele sunt principala poartă de intrare a informațiilor în neuroni, iar diferitele tipuri de neuroni au modele de ramificare distincte ale dendritei.<sup>35</sup> Progresele în electrofiziologie și modelare de calcul au arătat clar că forma arborelui dendritic este unul dintre factorii cruciali care determină modul în care sunt integrate semnalele provenite din sinapsele individuale.<sup>36</sup>

- **Sinapsa**- Legătura dintre neuroni se realizează prin sinapse. Acestea sunt formațiuni structurale specializate, care se realizează între axonul neuronului presinaptic și dendritele sau corpul celular al neuronului postsinaptic. Legătura interneuronală se face între segmentul presinaptic reprezentat de butonul terminal al axonului și segmentul postsinaptic, reprezentat de o zonă mică din membrana neuronului postsinaptic pe care se aplica butonul terminal. Cele două segmente sinaptice sunt separate printr-un spațiu sinaptic. Deci, legătura dintre neuroni nu se face prin contact direct, ci este mediată chimic, prin eliberarea mediatorului în fanta sinaptică.<sup>37</sup>

---

30 F. A. Azevedo, *Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain*, în "The Journal of Comparative Neurology", vol. 513, nr. 5, 2009, p. 532.

31 S. H. Houzel, *The not extraordinary human brain*, în "Proceedings of the National Academy of Sciences", vol. 109, Supplement 1, 2012, p. 10661.

32 *Ibidem*.

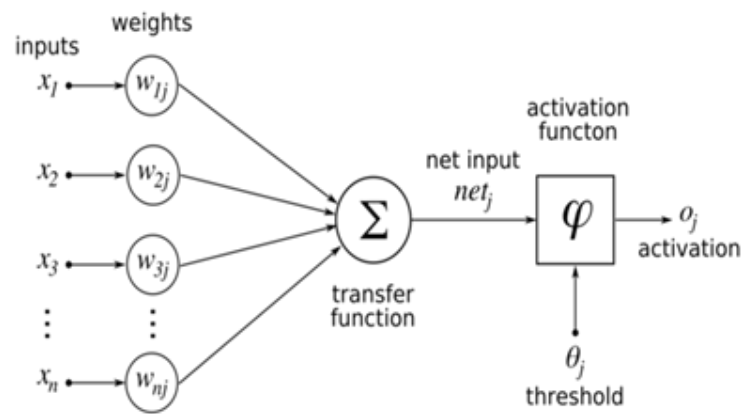
33 M. D. Dobrea, *Tehnici de inteligență computațională. Aplicații în electronică și biomedicină*, Ed. Performantica, Iași, 2017, p. 1, disponibil [online](#) - accesat la data 16.08.2020.

34 Disponibil online: <https://www.britannica.com/science/axon>, accesat la data 16.08.2020.

35 M. Urbanska, M. Blazejczyk, J. Jaworski, *Molecular basis of dendritic arborization*, "Acta neurobiologiae experimentalis", vol. 68, nr. 2, 2008, p. 264.

36 A. T. Gulledge, B. M. Kampa, G. J. Stuart, *Synaptic integration in dendritic trees*, în "Journal of Neurobiology", vol. 64, nr. 1, 2005, pp. 75-90; I. Segev, M. London, *Untangling dendrites with quantitative models*, în "Science", vol. 290, nr. 5492, 2000, pp. 744-750.

37 Introducere în sistemul nervos, material disponibil [online](#) - accesat la data 18.08.2020.

Fig 3. Perceptronul<sup>38</sup>

Similar cu sistemul nervos al omului, **perceptronul** este celula de baza a rețelei neuronale artificiale. În esență, acesta este de fapt o rețea neuronală *unistratificată*, care cuprinde 5 părți fundamentale: valori de intrare (*input*), pondere (*weights*), suma, termenul de deplasament (*bias*), și funcția de activare.

Astfel sinapsele dintre neuroni împreună cu dendritele au devenit intrări, corpul neuronului (soma) a devenit sumator și o funcție de activare, axonul a devenit ieșirea. Biasul (*b*) în limba română, s-ar traduce ca parțialitate, care urmărește un anumit scop. Un exemplu este pareidolia – acea trăsătură a psihicului nostru ce ne face să vedem fețe umane peste tot: să vedem un craniu pe Lună, o față în spuma de cafea etc. Într-un cadru mai larg, e de fapt predispoziția de a găsi tipare în lucrurile pe care le observăm, chiar și acolo unde ele nu există.<sup>39</sup> Ponderea (*w*) este parametrul care are atât rolul de a transforma datele de intrare în straturile ascunse ale rețelei cât și de a stabili cât de mult va fi influențată valoarea de ieșire. Astfel, cu cât variabila *w* este un număr mai mare, cu atât mai mult va "lăsa" ca primul neuron care primește semnalul (input) să influențeze următoarea secvență.

Ideea unui perceptron este de a introduce anumite valori (*x*), de a le înmulți cu ponderea (*w*) pentru a influența mai mult sau mai puțin valoarea finală (*y*), apoi acestea se vor însuma. La această sumă se adaugă valoarea *bias* (*b*) iar rezultatul se trece printr-o funcție de activare. Astfel, perceptronul este un algoritm folosit pentru învățarea supravegheată a clasificării binare<sup>40</sup>, care poate fi exprimat după cum urmează<sup>41</sup>:

Dat fiind un set de intrări  $x_1 \dots x_n$ , dacă o unitate are un set de ponderi corespunzătoare  $w_1 \dots w_n$  și un bias *b*, suma ponderată *z* poate fi reprezentată ca:

$$Z = b + \sum_i (w_i x_i)$$

Însă, este mai convenabil să exprimăm această sumă ponderată printr-un vector. Prin urmare, vom avea (*z*) în ceea ce privește un vector ponderat (*w*), biasul scalar (*b*) și vectorul de intrare (*x*), și vom înlocui suma cu produsul:

$$Z = w \cdot x + b$$

În această ecuație *z* este un număr real.

În cele din urmă, în loc să utilizăm *z*, o funcție liniară a lui *x*, pentru valoarea de ieșire vom utiliza o funcție neliniară *f(z)*. Ne vom referi la ieșirea acestei funcții ca fiind activarea valorii pentru unitatea *a*. După aceea urmează să aplicăm funcția de activare, iar rezultatul va reprezenta ieșirea finală.

$$Y = a = f(z)$$

38 Disponibil online: <https://nl.wikipedia.org/wiki/Perceptron>, accesat la data 18.08.2020.

39 M. Stefanescu, *Introducere în rețele neuronale – Teorie și aplicații*, material disponibil [online](#) - accesat la data 18.08.2020.

40 M. Skilton, F. Hovsepian, *op. cit.*, p. 124-125.

41 D. Jurafsky, J. H. Martin, *Neural Networks and Neural Language Models*, 2019, disponibil [online](#) - accesat la data 20.08.2020.

Așadar, vom aplica una dintre cele 3, cele mai utilizate funcții non-liniare (f. Sigmoidă, f. TanH (tangenta hiperbolică) sau f. Semi-liniară ReLU (Rectified Linear Unit), și anume funcția Sigmoidă:

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

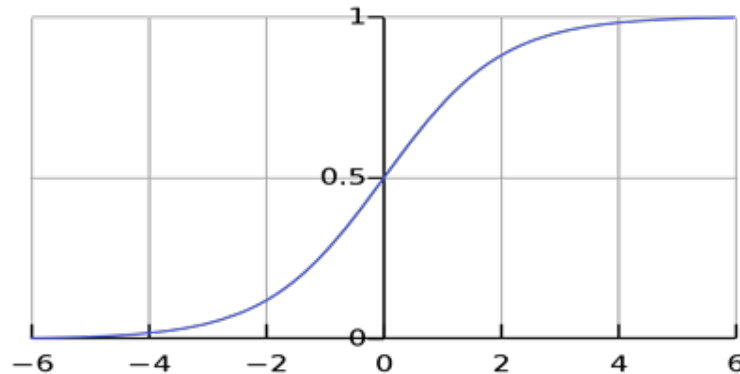


Fig. 4 Funcția sigmoidă<sup>42</sup>

În graficul de mai sus  $y$  ia valori în limita (0, 1) pe axa verticală, iar  $z$  (-6, 6) pe axa orizontală. Dacă substituim ecuația a 2-a în ecuația a 4-a vom avea următorul rezultat:

$$y = \sigma(w \cdot x + b) = \frac{1}{1 + e^{-(w \cdot x + b)}}$$

Desigur, având în vedere că perceptronul este o rețea neuronală unistratificată binară, adică poate returna ca valori de ieșire numai 0 sau 1, folosind aceleași date din ecuația 2, vom avea:

$$z = w_1x_1 + \dots + w_nx_n + b = \sum_{i=1}^n (w_i x_i) + b$$

$$y = \begin{cases} 0, & \text{dacă } w \cdot x + b \leq 0 \\ 1, & \text{dacă } w \cdot x + b > 0 \end{cases}$$

### III. APLICAȚII ALE ALGORITMILOR PREDICTIVI ÎN ANALIZA TEXTUALĂ. EXTRAGEREA INFORMAȚIILOR ȘI A ARGUMENTELOR

După cum se poate anticipa, în ultima decadă, sfera de incidență a inteligenței artificiale este una aflată într-o creștere exponențială, inclusiv în privința științelor juridice. De pildă, una dintre cele mai noi aplicații din domeniul legal, *Ross*<sup>43</sup>, oferă răspunsuri la întrebări de drept bazate pe legislație, jurisprudență și analiză doctrinară. De asemenea, are aptitudinea de a prezice cum va reacționa un judecător în funcție de un anumit argument care urmează să i se prezinte într-o anumită speță. *Lex Machina*<sup>44</sup>, este o altă aplicație de ultimă generație care a fost creată pentru a realiza predicții în domeniul proprietății intelectuale. Algoritmii au mai fost utilizați atât pentru extragerea de argumente din jurisprudența relevantă, literatura de specialitate ori concluziile părților adverse, dar și pentru a concepe argumente ce se pot situa din punct de vedere logic

<sup>42</sup> Disponibil online: [https://en.wikipedia.org/wiki/Sigmoid\\_function](https://en.wikipedia.org/wiki/Sigmoid_function), accesat la data 20.08.2020.

<sup>43</sup> A se vedea platforma companiei Ross, disponibil online: <https://www.rossintelligence.com/>, accesat la data 25.08.2020.

<sup>44</sup> A se vedea platforma companiei LexisNexis, disponibil online: <https://lexmachina.com/legal-analytics/>, accesat la data 25.08.2020.



deopotrivă în favoarea ori în defavoarea unui anumit concept, în funcție de care parte ne situăm ca apărători.

Din punct de vedere structural și funcțional, aceste aplicații sunt denumite *CCLA* (cognitive computing legal app), integrând modele computaționale de raționare juridică *CMLR* (computational models of legal reasoning) sau *CMLA* (computational models of legal argument) și modele de extragere a argumentelor din texte de natură legală.

În dezvoltarea acestor modele, cercetătorii abordează întrebări precum, *Care este modul de a determina înțelesul unei norme juridice ?*, pentru ca un software să decidă dacă aplică acea normă unei situații sau nu, *Cum să diferențiem cazuistica dificilă de cea mai "ușoară" ?* și, *Cum am putea interpreta textele legale în funcție de jurisprudență ?* Până acum, a fost nevoie ca informațiile juridice să fie extrase manual, adică necesita ca acestea să fie citite, iar mai apoi să fie evidențiate părțile relevante din conținutul acestora, ceea ce însemna o pierdere substanțială de timp. Cercetările recente din domeniul *QA* (question-answering), *IE* (information extraction) și extragerea argumentelor, oferă promisiunea de a schimba modul în care vor profesa avocații dar și alți experți din sfera juridică. Mai exact, *CMLR* și *CMLA* vor folosi informația obținută pentru a răspunde la întrebări juridice, pentru a prezice rezultatul unui proces sau a oferi argumente ori explicații în funcție de un anumit scop.<sup>45</sup>

### 1. Categoriile ale procesării lingvistice

Studii care au urmărit cercetarea procesării textului legal pot fi împărțite în următoarele categorii:<sup>46</sup>

**i. Legal Ontology Learning** (oferă o reprezentare structurală a conceptelor legale și interconexiunile dintre acestea<sup>47</sup>). Guiraudé Lame<sup>48</sup> utilizează *NLP* (Natural Language Processing) pentru a recunoaște concepte legale și relațiile semantice dintre acestea. Jose Saias și Paulo Quaresma<sup>49</sup> prezintă o metodologie pentru a crea automat *OWL* (Ontology Web Language) dintr-un set de documente juridice. Stephan Walter și Manfred Pinkal<sup>50</sup> arată un mod de extragere și analiză a unor definiții din 6.000 de decizii ale instanțelor din Germania.

**ii. Legal Information Extraction** (extragerea de date cu caracter juridic). Stephan Walter<sup>51</sup> folosește analiza arborilor decizionali pentru a extrage informații din deciziile instanțelor din Germania. Thorne McCarty<sup>52</sup> prezintă o metodă computațională a interpretărilor semantice a unor texte jurisprudențiale din *outputul* analizei sintactice. Baza analizei textuale constă în cazuistica instanțelor de apel din SUA în dreptul civil.

**iii. Legal Semantic Annotation** (identifică anumite cuvinte sau concepte legale și stabilește relațiile dintre ele). Raffaella Brighi<sup>53</sup> prezintă un mod automat de adnotare a dispozițiilor modificatoare ale legilor din Italia.

45 K. D. Ashley, *Artificial intelligence and legal analytics. New Tools for Law Practice in the Digital Age*, Ed. Cambridge University Press, 2017, pp. 4-5.

46 N. X. Bach, N. L. Minh, T. T. Oanh, A. A. Shimazu, *A Two-Phase Framework for Learning Logical Structures of Paragraphs in Legal Articles*, în "ACM Transactions Asian Language Information Processing", vol. 12, nr. 1, 2013, p. 4.

47 V. Leone, L. D. Caro, S. Villata, *Legal Ontologies and How to Choose Them: the Investigation Tool*, "International Semantic Web Conference", 2018, p. 1.

48 L. Guiraudé, *Using NLP techniques to identify legal ontology components: Concepts and relations*, "Artificial Intelligence and Law", vol. 12, nr. 4, 2004, p. 382.

49 J. Saias, P. Quaresma, *A Methodology to Create Legal Ontologies in a Logic Programming Based Web Information Retrieval System*, în V. R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi, *Law and the Semantic Web. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005, p. 185.

50 S. Walter, M. Pinkal, *Automatic extraction of definitions from German court decisions*, în Proceedings of the Workshop on Information Extraction Beyond The Document, "Association for Computational Linguistics", 2008, p. 21.

51 S. Walter, *Linguistic description and automatic extraction of definitions from German court decisions*, în vol. Proceedings of the 6th International Language Resources and Evaluation, "European Language Resources Association", 2008, p. 2928.

52 T. McCarty, *Deep semantic interpretations of legal texts*, în vol. Proceedings of the 11th International Conference on Artificial Intelligence and Law, Association for Computing Machinery, New York, United States, 2007, p. 2.

53 R. Brighi, L. Lesmo, A. Mazzei, M. Palmirani, D. P. Radicioni, *Towards semantic interpretation of legal modifications through*

Spinosa<sup>54</sup> a creat un sistem de consolidare a legislației italiene bazat pe metadate și algoritmi NLP.

**iv. *Automatic Identification of Legal Terms.*** (identificare a termenilor legali). Scopul proiectului întemeiat de Karel Pala<sup>55</sup> este de a construi un dicționar electronic de termeni juridici și o bază de date de 50.000 de documente de natură legală.

**v. *Legal Knowledge Modeling*** (transformă texte legale în forme logice). Lucrarea lui Nakamura<sup>56</sup> poate fi aplicată în Japonia, deoarece se potrivește pentru limba care nu folosește niciun pronume.

**vi. *Legal Argumentation*** (argumentare juridică). Enrico Francesconi<sup>57</sup> a publicat în numeroase rânduri lucrări în care arată modalitatea de identificare și prelevare a unui argument din text. Moens și echipa sa<sup>58</sup> descriu o investigație privind identificarea argumentelor. Ei au considerat sarcina de detectare ca o problemă de clasificare și au construit un clasificator folosind un set de argumente adnotate. Diferite feluri de trăsături au fost evaluate, inclusiv din punct de vedere lexical, sintactic și semantic. Adam Wyner<sup>59</sup> prezintă o metodă de identificare a argumentelor prin utilizarea formulelor gramaticale, ontologiei, și algoritmi NLP.

**vii. *Legal Automatic Summarisation.*** Claire Grover<sup>60</sup> a conceput un algoritm prin care clasifică propozițiile din documente în funcție de natura lor retorică, apoi acestea sunt selectate pentru a forma un rezumat pe baza rolului lor predefinit.

**viii. *Natural language processing texts.*** Utilizarea *NLP* în analiza morfologică și sintactică în privința textelor legale a fost descrisă de Karel Pala<sup>61</sup> în aplicarea asupra a 50.000 de texte legale, incluzând Constituția, decizii ale instanțelor și alte documente publice. Giulia Venturi<sup>62</sup> descrie o investigație a caracteristicilor sintactice ori lexicale ale construcțiilor legale. În viziunea sa, înțelegerea domeniului specific al limbii are importanțe practice în dezvoltarea de aplicații inteligente.

Reglementările din textele legale conțin adesea noțiuni de o complexitate ridicată. Un labirint intrigant de reguli formulate într-un jargon tehnic descriu ceea ce este legal sau dimpotrivă. Rețelele lor de referințe și excepții sunt de cele mai multe ori inabordabile de către un cetățean obișnuit. Chiar și cei mai experimentați juriști pot întâmpina dificultăți în a reuși să descifreze toate dispozițiile legale pentru analiza unei întrebări sau rezolvarea unei probleme de drept.

*deep syntactic analysis*, în vol. Proceedings of the 21st International Conference on Legal Knowledge and Information Systems, "Frontiers in Artificial Intelligence and Applications", 2008, pp. 202-206.

54 P. L. Spinosa, G. Giardiello, M. Cherubini, S. Marchi, G. Ventur, S. Montemagni, *NLP-based Metadata Extraction for Legal Text Consolidation*, în vol. Proceedings of the 12th International Conference on Artificial Intelligence and Law, 2009, p. 41.

55 K. Pala, P. Rychlý, P. Šmerk, *Automatic identification of legal terms in Czech legal texts*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010, p. 83.

56 M. Nakamura, S. Nobuoka, A. Shimazu, *Towards Translation of Legal Sentences into Logical Forms*, în K. Satoh, A. Inokuchi, K. Nagao, T. Kawamura, *New Frontiers in Artificial Intelligence. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2008, p. 351.

57 E. Francesconi, A. Passerini, *Automatic Classification of Provisions in Legislative Texts*, "Artificial Intelligence and Law", 2007, p. 3; E. Francesconi, G. Peruginelli, *Integrated Access to Legal Literature through Automated Semantic Classification*, "Artificial Intelligence and Law", 2009, p. 34; E. Francesconi, *An Approach to Legal Rules Modelling and Automatic Learning*, The Twenty-Second Annual Conference on Legal Knowledge and Information Systems, Rotterdam, 2009, p. 60.

58 M. F. Moens, E. Boiy, R. Mochales, C. Reed, *Automatic detection of arguments in legal texts*, Proceedings of the 11th International Conference on Artificial Intelligence and Law, Association for Computing Machinery, 2007, p. 225.

59 A. Wyner, R. M. Palau, M. F. Moens, D. Milward, *Approaches to Text Mining Arguments from Legal Cases*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010, p. 60.

60 C. Grover, B. Hachey, I. Hughson, C. Korycinski, *Automatic summarisation of legal documents*, în vol. Proceedings of the 9th International Conference, "Artificial Intelligence and Law", 2003, p. 244.

61 K. Pala, P. Rychlý, P. Šmerk, *Morphological analysis of law texts*, în vol. Proceedings of the 1st Workshop on Recent Advances in Slavonic Natural Language Processing, 2007, p. 21.

62 G. Venturi, *Legal language and legal knowledge management applications*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010, p. 3.

Enrico Francesconi arată în lucrarea sa, dintr-o perspectivă filozofică, analiza unui text legal ca fiind ”un vehicul care conține și transportă reguli/dispoziții (*provisions*) iar ordinea legală este mai degrabă un set de dispoziții decât de legi.”<sup>63</sup> Întregul corp de legi și reglementări, cu articolele și paragrafele lor, poate fi văzut conform unui **profil structural sau formal** și un **profil semantic sau funcțional**.<sup>64</sup> Profilul funcțional se împarte în alte 2 sub-profiluri: *profilul regulativ* și *profilul semantic*. Primul dintre acestea include categoria legislativă iar cel de al doilea reprezintă înțelesul diferitelor reguli. Profilul regulativ poate fi descris în termenii unei dispoziții legale (*provision*). O astfel de dispoziție poate presupune o *definiție, obligație, sancțiune, competență* etc. Acestea pot fi considerate argumentele dispoziției.

Carlo Biagolli a numit *modelul unei dispoziții*<sup>65</sup> ca fiind acea structură din care face parte tipul unei dispoziții și argumentele sale, care este, totodată, un model semantic al unui text legislativ. De asemenea, poate fi considerată o schemă de metadate aptă să descrie din punct de vedere analitic un text de lege. Folosind analiza metadatelor, un text poate fi clasificat semantic în funcție de modelul dispoziției. Francesconi<sup>66</sup> ne oferă un exemplu în acest sens:

”Să presupunem următorul fragment dintr-o lege din Italia:

*Un operator de date care intenționează să prelucreze date cu caracter personal care se încadrează în domeniul de aplicare al acestuia trebuie să îl notifice pe garant.*

Pasajul anterior poate fi interpretat ca fiind o structură a unui text (un paragraf) dar și ca un tip de dispoziție. Prin urmare, fragmentul este o *dispoziție* cu tipul *obligație* care cuprinde *argumentele*:

*Destinatar – Operator*

*Ațiune – Notificare*

*Parte adversă – Garant*”

Utilizarea *modelului de dispoziție* ca fiind elementul component al unui text legal ne ajută să clasificăm informația și să o extragem pentru a fi folosită în structuri superioare cum sunt argumentele juridice.

Voi prezenta pe scurt schema logică și algoritmi implicați în extragerea de informații din text. Nu se va proceda la expunerea pe larg a extragerii argumentelor sau la a arăta cum se pot crea noi construcții argumentative, având în vedere că excede cadrul acestei lucrări din punctul de vedere al complexității demersului și al algoritmilor utilizați, dar se vor face referiri adiacente pentru a explica fenomenul în ansamblul său.

Prin urmare, extragerea informațiilor care stau la baza argumentelor ce vor fi utilizate în pledoariile avocaților, ori în redactarea concluziilor, se realizează prin următoarele etape:

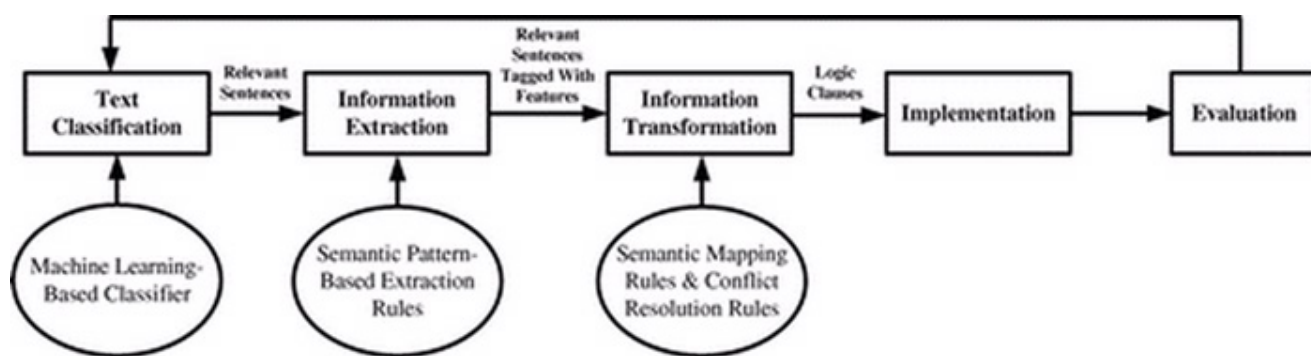
1. Clasificarea textului (*TC*)
2. Extragerea informațiilor (*IE*)
3. Transformarea informației într-un format logic (*Itr*)
4. Implementarea
5. Evaluarea

<sup>63</sup> E. Francesconi, A. Passerini, *op. cit.*, p. 2.

<sup>64</sup> C. Biagolli, *Towards a Legal Rules Functional Micro-Ontology*, ”Journal of Proceedings of workshop LEGONT”, vol. 97, 1997, p. 2.

<sup>65</sup> *Ibidem*.

<sup>66</sup> E. Francesconi, A. Passerini, *op. cit.*, p. 3.

Fig. 5 Extragerea automată a datelor<sup>67</sup>

Astfel, faza clasificării presupune recunoașterea propozițiilor relevante în text, în funcție de obiectivul stabilit. Datele urmărite din acele propoziții menționate sunt extrase și transformate în procesele *IE* și *Itr*. În etapa *TC* se vor filtra propozițiile irelevante, evitând să se prelucreze inutil date necorespunzătoare care ar genera erori și ar conduce la rezultate neconcludente.

Pentru extragerea informației din text sub formă de tipuri funcționale s-au folosit algoritmi *SVM* (Support Vector Machine), arbori de decizie și clasificatori Naive Bayes, iar în privința argumentelor s-au folosit *NLP* (Natural Language processing).<sup>68</sup>

Cercetările privind extragerea informațiilor din dispozițiile legale s-au concentrat pe extragerea, sau clasificarea dispozițiilor în ceea ce privește următoarele tipuri de informații<sup>69</sup>:

- Tipuri funcționale de dispoziții sau norme, cum ar fi definiția, interdicția, obligația și ceea ce este permis.
- Caracteristici legate de funcții, inclusiv elemente sau argumente ale unor tipuri funcționale, cum ar fi subiectul unei obligații sau agentul care acționează, acțiunea (privată individual) și agentul pasiv.
- Domenii sau subiecte de drept, de pildă, dreptul penal, dreptul privat sau dreptul administrativ.
- Profiluri semantice care combină tipuri funcționale, caracteristici sau domenii de referință.
- Concepte de reglementare utile pentru indexarea prevederilor.
- Reguli sau norme juridice, antecedentele și consecințele acestora, inclusiv formulările logice ale regulilor sau ale componentelor acestora.

Modul de funcționare al clasificatorului Bayes este explicat la pct. 3.3. Pentru a extrage informațiile funcționale, Francesconi și Passerini au construit două instrumente: (1) *xmLegesClassifier* și (2) *xmLegesExtractor*. Datele de intrare (*input*) ale **clasificatorului** sunt niște paragrafe de text dintr-o dispoziție legală; datele de ieșire (*output*) reprezintă tipul funcțional prezis sau clasa dispoziției selectate dintr-un set de tipuri și clase candidat.<sup>70</sup> De asemenea, *inputul extractorului* este textul paragrafului și tipul dispoziției prezis anterior de clasificator. *Outputul* este format din fragmente de text care corespund anumitor roluri semantice relevante pentru acel tip de dispoziție.<sup>71</sup>

Odată extrase, aceste informații conceptuale pot fi aplicate dispozițiilor dintr-un fel de marcaj semantic ca metadatae într-o adnotare *XML* și compilate într-o ontologie juridică precum ontologia Dalos.<sup>72</sup> Ca urmare,

67 J. Zhang, N. M. El-Gohary, *Automated Information Transformation for Automated Regulatory Compliance Checking in Construction*, "Journal of Computing in Civil Engineering", vol. 29, nr. 4, 2015, p. 2.

68 K. D. Ashley, *op. cit.*, p. 261.

69 *Ibidem*, p. 260.

70 E. Francesconi în K. D. Ashley, *op. cit.*, p. 263.

71 *Ibidem*.

72 E. Francesconi, S. Montemagni, W. Peters, *Integrating a Bottom-Up and Top-Down Methodology for Building Semantic Resources for the Multilingual Legal Domain*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010, p. 95.

sistemul are câteva informații despre ce înseamnă conceptele din text și le folosește pentru a le ajuta la *regăsirea conceptuală*.<sup>73</sup>

Prin asemănare, Mochales și Moens<sup>74</sup>, în studiul lor asupra extragerii argumentelor, în primă fază, au reprezentat propozițiile ca vectori caracteristici. Pentru învățarea identificării frazelor ca propoziții argumentative, propozițiile au fost marcate în termenii caracteristicilor generale pe baza informațiilor extrase din texte, respectiv:<sup>75</sup>

- fiecare cuvânt, pereche de cuvinte, perechi și tripluri de cuvinte succesive
- etichete POS, inclusiv anumite adverbe, verbe (care indică permisiunea sau obligația precum „poate”, „trebuie” și „ar trebui”)
- timpul și tipul principal al verbului
- categoriile de propoziții anterioare și succesive
- tipul subiectului propoziției (ex. reclamant, pârât, instanța etc.)
- anumite tipare de punctuație
- anumite cuvinte cheie care indică argumentarea, de exemplu, „dar”, „astfel”, „în consecință” și „din această cauză”
- adâncimea ramificațiilor în fraza analizată și numărul de subclauze (ambele sunt un barometru al complexității propoziției)
- anumite statistici de text, inclusiv lungimea propoziției, lungimea medie a cuvântului și numărul semnelor de punctuație

În etapa a doua, au aplicat trei algoritmi de învățare statistică utilizați în mod obișnuit în clasificarea textului: Clasificatorul Bayes, Clasificatorul de entropie maximă și SVM, iar primii doi au fost aplicați pentru a stabili dacă o propoziție a fost argumentativă.<sup>76</sup>

În concluzie, am arătat în termeni generali cum este posibil ca informațiile dintr-un număr mare de documente, decizii, legi etc., pot fi secționare în părți componente pentru a face obiectul etichetării, clasificării iar mai apoi extragerii, pentru ca, în cele din urmă, să conducă la realizarea unor procese computaționale mai complexe.

În continuare este prezentată una dintre modalitățile fundamentale de clasificare ale unui text care este clasificatorul Naiv Bayes.

## 2. Teorema lui Bayes

Cel mai reprezentativ algoritm stohastic este clasificatorul Bayesian. Acesta prezice probabilitatea de apartenență la o clasă anume, cum ar fi de exemplu, probabilitatea ca un text să facă parte dintr-o anumită clasă. Clasificatorul Bayesian simplu, cunoscut sub denumirea de clasificatorul Bayesian Naiv are rezultate bune atunci când este aplicat la seturi de date foarte mari, cum este exemplul literaturii de specialitate și jurisprudenței.<sup>77</sup>

De asemenea, clasificatorul Naiv Bayes este unul dintre cei mai eficienți algoritmi inductivi pentru învățare automată și *data mining*. Performanța sa în clasificare este una surprinzătoare, întrucât se bazează pe

<sup>73</sup> *Ibidem*.

<sup>74</sup> R. Mochales, M. F. Moens, *Argumentation mining*, "Artificial Intelligence and Law", 2011, p. 14.

<sup>75</sup> *Ibidem*.

<sup>76</sup> K. D. Ashley, *op. cit.*, p. 289.

<sup>77</sup> R. G. Crețulescu, *Teză de doctorat (Rezumat). Contribuții la proiectarea sistemelor de clasificare a documentelor*, Universitatea „Lucian Blaga” din Sibiu, Facultatea de Inginerie “Hermann Oberth”, Departamentul de Calculatoare și Inginerie Electrică, 2011, p. 46.

prezumția independenței condiționale,<sup>78</sup> care este rareori adevărată.<sup>79</sup>

Algoritmul la care ne referim pornește de la premisa *naivă*, (care se numește *independență condițională de clasă*)<sup>80</sup> că atributele dintr-o clasă determinată sunt independente. El pleacă de la această prezumție tocmai pentru a ușura calculele. Însă, pentru a înțelege în mod corect cum funcționează algoritmul trebuie să parcurgem teoria care se află la baza sa.

Teorema lui Bayes este fundamentul aplicării clasificatorului Bayesian. Aceasta este o formulă matematică simplă folosită pentru calcularea probabilităților condiționale. Se remarcă în abordări subiectiviste sau Bayesiene ale epistemologiei, statisticilor și logicii inductive,<sup>81</sup> care determină probabilitatea apartenenței evenimentelor sau a obiectelor la o anumită grupă.<sup>82</sup>

Așadar, în paradigma acestei teorii, suntem interesați de determinarea celei mai bune ipoteze pentru un spațiu  $H$ , luând în considerare date de antrenare  $D$ .<sup>83</sup> Dacă ne referim la cea mai bună ipoteză, de fapt, este incidentă cea mai probabilă dintre ipoteze cunoscând datele  $D$  și alte informații inițiale despre probabilitățile diverselor ipoteze din  $H$ .<sup>84</sup> În consecință, metoda Bayes oferă un mijloc de calculare directă a acestor probabilități, mai exact, are posibilitatea de a determina probabilitatea unei ipoteze bazate pe probabilitatea anterioară a acesteia.<sup>85</sup>

Din punct de vedere matematic, în exprimarea acestei teorii vom analiza următoarele notații:

- $h$  este o ipoteză din spațiul ipotezelor  $H$ .
- $P(h)$  reprezintă probabilitatea ipotezei  $h$ ; este deseori numită **probabilitatea anterioară** a lui  $h$  și poate evidenția orice informație inițială pe care o avem despre posibilitatea ca  $h$  să fie o ipoteză corectă.<sup>86</sup>
- $D$  sunt informațiile ce corespund datelor noi care nu au fost utilizate la calcularea probabilității. Sunt date de antrenament.
- $P(D)$  este denumită probabilitate marginală. Este aceeași pentru toate ipotezele posibile luate în considerare. Acest factor nu intră în determinarea probabilităților relative a ipotezelor diferite fiind o probabilitate anterioară informațiilor de antrenament.
- $P(D|h)$  este probabilitatea observării informațiilor  $D$  introduse în conformitate cu ipoteza  $h$ .
- $P(h|D)$  reprezintă **probabilitatea posterioară**  $h$  în funcție de datele de observare  $D$ . Ne dorim să aflăm probabilitatea unei ipoteze date în funcție de informațiile observate.<sup>87</sup>

Prin urmare, teorema lui Bayes calculează probabilitatea posterioară  $P(h|D)$ <sup>88</sup> în funcție de: probabilitatea anterioară  $P(h)$ , probabilitatea anterioară a datelor de antrenament  $P(D)$ , și probabilitatea datelor de antrenament în raport cu ipoteza  $h$  -  $P(D|h)$ <sup>89</sup>:

78 K. P. Murphy, *op. cit.*, p. 30; S. J. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Ed. Prentice Hall, Englewood Cliffs, New Jersey, 1995, p. 429.

79 H. Zhang, *The Optimality of Naive Bayes*, Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, vol. 2, 2004, p. 562.

80 R. G. Crețulescu, *op. cit.*, p. 46.

81 J. James, *Bayes Theorem*, "The Stanford Encyclopedia of Philosophy", 2019 disponibil online: <https://plato.stanford.edu/archives/spr2019/entries/bayes-theorem/>, accesat la data 12.09.2020.

82 L. E. Mureșan, *op. cit.*, p. 12; A. Ciobanu (Aionoe), *Utilizarea Rețelelor Bayesiene și a Diagramelor de Influență în Analiza Disponibilității în Electroenergetică, Rezumat Teză de Doctorat*, Universitatea Tehnică "Gheorghe Asachi" din Iași, 2019, p. 5.

83 L. E. Mureșan, *op. cit.*, p. 13, R. G. Crețulescu, *op. cit.*, p. 47.

84 *Ibidem*, p. 12; K. P. Murphy, *op. cit.*, p. 66-67.

85 *Ibidem*.

86 B. C. Drăgoescu, *Învățarea Bayesiană*, Lucrare de Licență, Universitatea Babeș-Bolyai Cluj-Napoca, Facultatea de Matematică și Informatică, 2015, p. 14-15.

87 K. P. Murphy, *op. cit.*, p. 66.

88 Y. Bengio, A. Courville, P. Vincent, *Representation Learning: A Review and New Perspectives*, "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 35, nr. 8, 2013, p. 7.

89 B. C. Drăgoescu, *op. cit.*, p. 14-15.

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)}$$

În cele mai multe scenarii, urmărim aflarea celei mai probabile ipoteze  $h$  din mulțimea de ipoteze  $H$ , prin raportare la datele de antrenament  $D$ .

Se va numi *MAP* (maximum a posteriori hypothesis)<sup>90</sup> – ipoteza maximă posterioară și reprezintă fiecare dintre ipotezele de probabilitate maximă care se notează cu  $\cdot$ .<sup>91</sup> Pentru determinarea se va utiliza teorema lui Bayes pentru calcularea probabilității ulterioare a fiecărei ipoteze candidat. În concret, se va calcula fiecare ipoteză ulterioară și se va alege cea mai probabilă dintre ele. Astfel, putem afirma că este o ipoteză maximă posterioară<sup>92</sup> dacă:

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}} P(h|D) = \underset{h \in H}{\operatorname{argmax}} \frac{P(D|h) * P(h)}{P(D)} = \underset{h \in H}{\operatorname{argmax}} P(D|h) * P(h)$$

După cum se poate observa  $P(D)$  lipsește din ultima ecuație. Acest aspect se datorează faptului că  $P(D)$  este o constantă care nu influențează sub nici o formă; prin urmare, este rațional să se renunțe la acest termen.

Pot exista în practică situații în care probabilitățile anterioare ale fiecărei ipoteze sunt egale, motiv pentru care vom exclude  $P(h)$ , și vom utiliza doar pentru a identifica cea mai probabilă dintre ipoteze.<sup>93</sup> Orice ipoteză care maximizează este numită *ipoteza cu probabilitatea maximă ML (Maximum Likelihood)*<sup>94</sup> (se va nota cu  $h_{ML}$ )<sup>95</sup>.

$$h_{ML} = \underset{h \in H}{\operatorname{argmax}} P(D|h)$$

Pentru a clarifica relația dintre teorema lui Bayes și problemele învățării automate am făcut referire la datele  $D$  ca la un set de date de antrenare ale unor funcții țintă, iar la  $H$  ca fiind spațiul acestor funcții.<sup>96</sup> De fapt, spectrul de incidență al teoremei este mult mai amplu, cuprinzând domenii mai variate decât sugerează această discuție. Teorema ar putea fi aplicată la fel de bine la orice set  $H$  de propoziții reciproc exclusive a căror sumă a probabilităților este egală cu 1. Aceasta ne oferă principala modalitate de calcul a probabilității ulterioare a fiecărei ipoteze în parte, luând în considerare datele de antrenament.<sup>97</sup>

### 3. Clasificatorul Naiv Bayesian

Algoritmul Naiv Bayesian este un clasificator probabilistic care se bazează pe teorema lui Bayes a probabilităților, presupunând utilizarea independenței caracteristicilor. Într-o formulare simplă, un clasificator Naiv Bayes presupune că prezența unei anumite particularități într-o clasă nu are legătură cu prezența oricărei alte caracteristici. De pildă, un fruct poate fi considerat a fi un măr dacă este roșu, rotund și cu diametrul de aproximativ 5 centimetri. Chiar dacă aceste caracteristici depind unele de altele sau de existența celorlalte caracteristici, toate aceste proprietăți contribuie în mod independent la probabilitatea ca fructul să fie un măr

90 I. Rish, *An empirical study of the naive Bayes classifier*, Workshop on Empirical Methods in Artificial Intelligence, IJCAI 2001, p. 42.

91 S. J. Russell, P. Norvig, *op. cit.*, p. 588; B. C. Drăgoescu, *op. cit.*, p. 15.

92 S. S. Shai, B. D. Shai, *Understanding Machine Learning. From Theory to Algorithms*, Ed. Cambridge University Press, New York, USA, 2014, p. 307.

93 B. C. Drăgoescu, *op. cit.*, p. 15.

94 T. Mitchell, *op. cit.*, p. 157; B. C. Drăgoescu, *op. cit.*, p. 15; K. P. Murphy, *op. cit.*, p. 59-60.

95 S. J. Russell, Peter Norvig, *op. cit.*, p. 589; Y. Bengio, A. Courville, P. Vincent, *op. cit.*, p. 7.

96 T. Mitchell, *op. cit.*, p. 157; B. C. Drăgoescu, *op. cit.*, p. 15.

97 L. E. Mureșan, *op. cit.*, p. 14.

și de aceea este cunoscut sub numele de „naiv”.

Acest model necesită ca fiecare dintre densitățile claselor reprezintă produsul dintre densitățile marginale; adică se presupune că fiecare input este condiționat de fiecare clasă.<sup>98</sup> De asemenea, un clasificator Bayesian este unul optimal, adică obține eroarea minimă posibilă pentru o anumită problemă – evident, numai dacă sunt respectate anumite condiții, iar cea mai restrictivă fiind aceea a mărimii setului de date.<sup>99</sup> Din punct de vedere practic, parametrii pentru modelele naive sunt estimați prin metoda ipotezei cu cea mai mare probabilitate.

Succesul său, în pofida faptului că prezintă elementele unor clase de probabilitate părtinitoare (care ar putea afecta rezultatul final dacă este folosit în alt scop decât clasificarea informațiilor) ori că face estimări sub presupunerea că predicțiile sunt independent condiționate (fără a lua în calcul covarianța dintre predicții), poate fi explicată astfel: optimalitatea în termeni de pierdere zero-unu (eroarea de clasificare) nu este neapărat legată de calitatea de a potrivi o distribuție de probabilitate. Mai degrabă, o clasificare optimă se obține atât timp cât ambele distribuții efective estimate sunt de acord cu privire la clasa cea mai probabilă.<sup>100</sup>

Clasificatorul Bayes se aplică la acele probleme de învățare, unde fiecare instanță este descrisă de succesiunea unor valori corespunzătoare atributelor, iar fiecare funcție target poate lua orice valoare dintr-un spațiu finit.<sup>101</sup> Se va atribui un set de date de antrenament pentru funcția țintă, apoi se prezintă o instanță nouă caracterizată de valorile, iar algoritmului i se cere să realizeze o predicție referitoare la valoarea țintă ori cărei clase îi aparține instanța.<sup>102</sup>

Prin urmare, trebuie să determinăm valoarea funcției țintă cu probabilitatea cea mai mare, luând în considerare vectorul de valori corespunzătoare atributelor ce descriu noua instanță.<sup>103</sup>

$$v_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, \dots, a_n)$$

Pentru a reformula această ecuație putem să ne folosim de teorema lui Bayes și vom obține:

$$\underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, \dots, a_n | v_j) P(v_j)$$

Pentru a estima valoarea ar trebui să se cunoască frecvența incidenței valorii în datele de observare, însă aplicarea aceleiași metode pentru estimarea termenilor nu este fezabilă decât în cazul în care avem un set foarte mare de date.<sup>104</sup> Prin urmare se va aplica de la caz la caz.

Având în vedere că algoritmul Bayes se bazează pe prezumții simplificate referitoare la faptul că valorile atribuite sunt independente condițional, se poate afirma că probabilitatea de a observa succesiunea este determinată de produsul probabilităților pentru fiecare atribut individualizat:

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$$

98 T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Ed. a 2-a, Ed. Springer, 2012, p. 108.

99 D. M. Dobrea, *op. cit.*, pp. 47, 148.

100 I. Rish, *op. cit.*, p. 41; P. Domingos, M. Pazzani, *On the optimality of the simple Bayesian classifier under zero-one loss*, Ed. Kluwer Academic Publishers, 1997, p. 103; L. E. Mureșan, *op. cit.*, p. 21.

101 B. C. Drăgoescu, *op. cit.*, p. 20; T. Mitchell, *op. cit.*, p. 177.

102 T. Mitchell, *op. cit.*, p. 177; L. E. Mureșan, *op. cit.*, p. 21.

103 B. C. Drăgoescu, *op. cit.*, p. 20; T. Mitchell, *op. cit.*, p. 177.

104 *Ibidem*.



Dacă înlocuim în ecuația anterioară cu rezultatul precedent obținem formula pentru determinarea probabilității maxime, unde reprezintă funcția țintă identificată:

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

Avantajul acestui artificiu matematic este că totalitatea termenilor distincți din datele de antrenament care trebuie estimați este foarte mult redus prin comparație cu estimarea tuturor termenilor  $P(a_1, a_2, \dots, a_n | v_j)$  din ecuația precedentă.

Concluzionând, clasificatorul Naiv Bayes implică o etapă de învățare în care sunt estimați diferiți termeni  $P(v_j)$  și  $P(a_i | v_j)$  pe baza frecvenței apariției în datele de observare/antrenare, iar setul de estimări corespunde ipotezelor care au fost învățate în partea de antrenare.<sup>105</sup> Ulterior, această ipoteză este folosită pentru a clasifica fiecare nouă instanță, aplicând formula pentru determinarea probabilității maxime din ultima ecuație. Ori de câte ori este îndeplinită ipoteza de clasificare naivă Bayes referitoare la independența condițională, rezultă că funcția target este identică cu clasificarea dată de ipoteza maximă posterioară.<sup>106</sup>

O diferență interesantă între algoritmul Bayes și alte metode de învățare este că nu există nicio căutare explicită prin spațiul posibilelor ipoteze (în acest caz, spațiul posibile ipoteze este spațiul valorilor posibile care pot fi atribuite diferitelor  $P(v_j)$  și termenii  $P(a_i | v_j)$ ), însă, ipoteza este formată fără a căuta, pur și simplu prin numărarea frecvenței diferitelor combinații de date din exemplele de antrenare.<sup>107</sup>

#### IV. IMPLICAȚII ALE INTELIGENȚEI ARTIFICIALE ÎN EVALUAREA RISCULUI DE RECIDIVĂ ȘI ÎN CONDAMNAREA PENALĂ

Acum că am prezentat în mod lapidar funcționarea unei părți a algoritmilor ce au incidență în analiza textuală, vom aborda într-un sens mai restrâns modul în care inteligența artificială poate impacta cu predilecție domeniul penal prin importanța sentințelor pe care le aplică judecătorii și dacă ar putea fi micșorată recidiva infractorilor.

În ultima decadă, Statele Unite ale Americii asistă la o explozie fără precedent în utilizarea algoritmilor în sfera relațiilor publice.<sup>108</sup> Această creștere exponențială a prezenței softurilor predictive a fost alimentată de numeroși factori, printre care amintim generarea de cantități imense de date - *big data*, de către utilizarea omniprezentă și ultra generalizată a internetului, a dispozitivelor inteligente, precum și punerea unui accent din ce în ce mai apăsător asupra luării deciziilor bazate pe date atât în privința vieții noastre private cât și a politicilor publice.<sup>109</sup> De pildă, în New York, inteligența artificială a fost folosită pentru a distribui copii în școlile publice,<sup>110</sup> evaluarea profesorilor,<sup>111</sup> urmărirea clădirilor pentru inspecții împotriva incendiilor,<sup>112</sup> și

105 T. Mitchell, *op. cit.*, p. 178; L. E. Mureșan, *op. cit.*, p. 22.

106 *Ibidem*.

107 T. Mitchell, *op. cit.*, p. 178.

108 D. Kehl, P. Guo, S. Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, "Harvard Law School", 2017, p. 3, disponibil [online](#) - accesat la data 12.09.2020.

109 N. Duarte, *CDT Statement on Government Use of Algorithmic Decision-Making Tools to NYC Council Committee on Technology*, 2017, disponibil [online](#) - accesat la data 12.09.2020.

110 A. Zimmer, *High Schools Dole Out Misinformation About Admissions Process, Parents Say*, "DNAInfo", 2016, disponibil [online](#) - accesat la data 12.09.2020.

111 C. O'Neil, *Don't Grade Teachers with a Bad Algorithm*, "Bloomberg", 2017, disponibil [online](#) - accesat la data 12.09.2020.

112 *A se vedea Mayor Bloomberg and Fire Commissioner Cassano Announce New Risk-based Fire Inspections Citywide Based on Data Mined from City Records*, 2013, disponibil [online](#) - accesat la data 12.09.2020.

pentru a lua decizii în privința poliției.<sup>113</sup> Acești algoritmi pot prelucra cantități mari de date și pot descoperi tipare sau perspective pentru a lua deciziile cele mai bune; cu toate acestea, se pare că nu sunt factorii de decizie imparțiali.<sup>114</sup>

Filosofia din spatele algoritmilor de evaluare a riscului în privința condamnării penale se bazează pe mișcarea anilor 80, așa-numita revoluție în liniște – *The selective incapacitation movement*<sup>115</sup> după cum a fost descrisă de New York Times. Această teorie se fundamentează pe ideea că ar trebui să existe o anumită ”selecție” a persoanelor care au o predispoziție nativă înspre violență, pericolozitate și risc de recidivă, cum sunt, de pildă, ”criminalii de carieră”,<sup>116</sup> iar apoi sistemul judiciar să îi pună în ”incapacitate” prin excluderea acestora în afara societății prin condamnare la închisoare pentru perioade lungi de timp,<sup>117</sup> prezumându-se că, în acest fel, va scădea rata criminalității generale.<sup>118</sup>

La începutul anilor 1970, cercetările care se desfășurau asupra predictibilității infracțiunilor aveau în prim-plan identificarea elementului de pericolozitate al infractorilor și al capacității acestora de a comite fapte penale cu violență.<sup>119</sup> Cu toate acestea, unele studii<sup>120</sup> au arătat că predicțiile în privința criminalității se dovedesc a fi un fenomen destul de complex care a derivat într-un număr semnificativ de cazuri fals-pozitive.

Conceptul de a pedepsi criminalii nu pentru ce au săvârșit în trecut, ci, mai cu seamă, pentru ce infracțiuni vor putea fi comise în viitor, a reprezentat un punct de cotitură în teoria condamnării penale<sup>121</sup>. Argumentele morale care fundamentează abordarea punerii în incapacitate selectivă a unor indivizi, atrag după sine un conflict între *utilitarieni* (cei care aparțin familiei etice a utilitarismului) și cei care susțin conceptul potrivit căruia infractorii ar trebui să răspundă pentru faptele lor. Viziunea utilitaristă aduce în perspectivă ipoteza condamnării numai dacă este satisfăcută condiția de a reduce rata generală a criminalității și, prin urmare, de a proteja în acest fel interesul public. Pe de altă parte, după cum arată și Thomas Mathiesen: ”este necinstit să solicităm tragerea la răspundere pentru presupusele alegeri care nu au fost încă făcute, adică pentru infracțiunile anticipate care încă nu au fost comise.”<sup>122</sup>

În fine, cu toate că dezbaterea asupra teoriei incapacității selective a apus demult, există totuși două puncte de similaritate între acea perioadă și discuția despre algoritmi de evaluare a riscului în sentințele penale. În acest sens, mă refer la îndoiala legitimă care planează asupra ambelor teze, și anume, că nu reușesc să fie suficient de precise și echidistante față de infractor, prin aceasta afectându-i drepturile și libertățile fundamentale. Tot așa cum în anul 1980 unii își manifestau simpatia și aderența la teoria incapacității selective, azi, cei care pledează pentru algoritmi predictivi o fac din rațiunea de a reduce încarcerarea în masă fără a crește rata criminalității și pentru a îmbunătăți judecata prin analiza tehnică a datelor.<sup>123</sup> A reuși în această privință poate fi un lucru destul de dificil dacă luăm în considerare faptul că instrumentele de evaluare a riscului

---

113 A se vedea *Brennan Center for Justice v. New York City Police Department*, 2017, disponibil online: <https://www.brennancenter.org/legal-work/brennan-center-justice-v-new-york-police-department>, accesat la data 13.09.2020.

114 N. Duarte, *op. cit.*, p. 1.

115 T. Lewin, *Making Punishment Fit Future Crimes*, ”New York Times”, 1982, disponibil [online](#) - accesat la data 12.09.2020.

116 Student-edited journal, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, vol. 96, nr. 2, ”Harvard Law Review”, 1982, p. 511.

117 T. Mathiesen, *Selective Incapacitation Revisited*, ”Law and Human Behavior”, vol. 22, nr. 4, 1998, p. 455.

118 *Ibidem*, p. 456.

119 *Ibidem*, p. 515.

120 J. Monahan, *Predicting Violent Behavior: An Assessment of Clinical Technique*, ”Crime & Delinquency”, vol. 28, nr. 4, 1981, p. 244.

121 D. Kehl, P. Guo, S. Kessler, *op. cit.*, p. 4.

122 T. Mathiesen, *in* D. Kehl, P. Guo, S. Kessler, *op. cit.*, p. 4.

123 CSG Justice Center Staff, *Risk and Needs Assessment and Race in the Criminal Justice System*, The Council of State Governments, 2016, disponibil online: <https://csgjusticecenter.org/reentry/posts/risk-andneeds-assessment-and-race-in-the-criminal-justice-system/>, accesat la data 15.09.2020.

în sentințele penale pot constitui o reiterare a ideilor care au stat la baza incapacității selective.<sup>124</sup>

### 1. Instrumente de evaluare a riscului penal

Din multitudinea de instrumente de evaluare a riscului, Statele Unite consideră că 8 dintre acestea sunt cele mai importante pentru a stabili elementele esențiale cu potențial predictiv, desigur, fiecare având domenii diferite în care își manifestă competența. Într-o enumerare laconică le vom aminti pe următoarele<sup>125</sup>:

i. **COMPAS – Correctional Offender Management Profiling for Alternative Sanctions.** A fost creat de către compania Northpointe și este utilizat de instanțele din Statele Unite în analiza riscului de recidivă a unei persoane condamnate. Acest software evaluează mai multe tipuri de risc, respectiv, recidiva generală, recidiva violentă, nerespectarea măsurilor de supraveghere, a obligațiilor impuse de instanță în cazul

stabilirii unor forme alternative de pedeapsă<sup>126</sup> și evaluarea pericolului social dinainte de proces, pentru a determina dacă se aplică o măsură preventivă sau nu.<sup>127</sup>

Evaluarea riscului de recidivă prin violență se calculează după următoarea formulă<sup>128</sup>:

$$S = a(-w) + a_{\text{prima}}(-w) + h_{\text{violență}}w + v_{\text{edu}} + h_{\text{inc}}$$

Unde  $s$  reprezintă scorul aplicat pentru riscul de recidivă,  $a$  este vârsta actuală,  $a_{\text{prima}}$  este vârsta la care a fost prima dată arestat,  $h_{\text{violență}}$  este istoricul violenței,  $v_{\text{edu}}$  este vocația de a accede la educație,  $h_{\text{inc}}$  este istoria nerespectării legii în general (amenzi contravenționale, alte infracțiuni etc.) iar ponderea  $w$  este o variabilă determinată de intensitatea relației dintre elementul analizat și recidivă.<sup>129</sup>

ii. **IORNS – Inventory of Offender Risk, Needs and Strengths**, este destinat evaluării riscului static, riscului/necesității dinamice, și a factorilor de protecție, după cum se corelează cu recidiva generală, violentă sau cea împotriva libertății sexuale și constă în 4 indici, 8 scale, 14 subclase și 2 scale de validitate.<sup>130</sup> Așadar, cuprinde **ORI**- Overall Risk Index, **SRI**- Static Risk Index și **PSI** - Protective Strength Index, fiecare dintre acești indici oferind date referitoare la probabilitatea de a recădea sub sfera de incidență a infracționalității.

iii. **LSIR (Level of Service Inventory Revised)**, **LS/CMI (Level of Service/Case Management Inventory)** și **LS/RNR (Level of Service/Risk, Need, Responsibility)**.

iv. **LSIR Level of Service Inventory Revised**, instrument de a 3-a generație, măsoară 54 de factori de risc și de nevoi criminogene din 10 domenii care sunt concepute să ofere informații pentru deciziile instanțelor în legătură cu măsurile preventive sau măsurile de supraveghere a inculpatului.<sup>131</sup> Domeniile predictive sunt: istoricul infracțional, educația, situația financiară, familia/starea civilă, acomodarea, modul de relaxare, anturajul, abuzul de alcool sau substanțe, starea emoțională/mentală, atitudinea sau perspectivele individului.<sup>132</sup>

124 D. Kehl, P. Guo, S. Kessler, *op. cit.*, p. 6.

125 L. Stănilă, *Inteligența Artificială. Dreptul Penal și Sistemul de Justiție Penală. Amintiri Despre Viitor*, Ed. Universul Juridic, București, 2020, p. 162.

126 *Ibidem*, p. 163; T. Brennan, W. Dieterich, B. Ehret, *Evaluating The Predictive Validity of the Compas Risk And Needs Assessment System*, "Criminal Justice and Behavior", vol. 36, 2009, pp. 22- 23, disponibil [online](#) - accesat la data 15.09.2020.

127 *A se vedea* ghidul explicat al utilizării Softwareului COMPAS - *Practitioner's Guide to COMPAS Core*, Northpointe, 2005, disponibil [online](#) - p. 29, accesat la data 15.09.2020.

128 *Ibidem*.

129 *Ibidem*.

130 J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, Z. Hamilton, *Handbook of Recidivism Risk/Needs Assessment Tools*, Ed. John Wiley & Sons Ltd, 2018, p. 101.

131 D. Andrews, J. Bonta în A. W. Flores, C. T. Lowenkamp, P. Smith, E. J. Latessa, *Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers*, "Federal Probation", vol. 70, nr. 2, 2006, p. 3;

132 *Ibidem*.

v. *LS/CMI Level of Service/Case Management Inventory*, este un instrument de generația a 4-a care ia în considerare atât componenta riscului/necesităților dar și felul în care este recepționat de către subiectul procesual intervenția justiției.<sup>133</sup> Instrumentul include 11 secțiuni care să permită specialiștilor să evalueze factorii de risc, necesitatea intervenției, receptivitatea față de intervenție, necesitatea și experiența privării de libertate.<sup>134</sup> În psihometrie este folosită adesea *IRT- item response theory*, cunoscută și sub numele de teoria răspunsului latent, care se referă la o familie de modele matematice ce încearcă să explice relația dintre trăsăturile latente (caracteristici sau atribute neobservabile) și manifestările acestora (adică rezultatele observate, răspunsurile sau performanța lor). Acest algoritm matematic stabilește o legătură între proprietățile elementelor ale unui instrument și indivizii care răspund la aceste elemente.<sup>135</sup>

Potrivit autorilor R. Bertrand și J. Blais<sup>136</sup> dacă folosim un *model cu doi parametrii*, parametrul discriminării poate fi interpretat ca indicele de discriminare al elementului din vecinătatea punctului de inflexiune. Formula matematică ce exprimă modelul menționat este:

$$P_j(\theta) = \frac{1}{1 + e^{-1.7 a_j(\theta - \delta_j)}}$$

Unde  $P_j(\theta)$  este **probabilitatea** persoanei examinate de a răspunde corect, este **abilitatea** (capacitatea) de a răspunde corect, reprezintă parametrul discriminării, este **dificultatea** de răspuns la întrebări (sau un prag predefinit),  $e$  este baza logaritmului natural, iar  $1.7$  este o valoare atribuită unei constante  $D$ .<sup>137</sup>

*Dificultatea* unui element (întrebări) se referă la locul în care se poate plasa pe curba abilităților funcționalitatea întrebării (graficul din fig. 6). Mai exact, o întrebare simplă presupune o abilitate scăzută (o poziționare inferioară pe curba abilității), iar o întrebare dificilă un loc superior dispus pe curbă.<sup>138</sup>

*Discriminarea* este proprietatea de a descrie aptitudinea (ridicată sau scăzută) a unei întrebări de a distinge între candidații la examen care se află deasupra sau dedesubtul amplasării pe curbă a întrebării.<sup>139</sup>

În graficul de mai jos,  $b$  reprezintă indicele de dificultate. Prin urmare, prin teoria *IRT*, se poate demonstra frecvența riscului în relație cu răspunsul la unele întrebări. Astfel, cu cât valoarea lui  $b$  este mai mare, cu atât gradul de dificultate al întrebării crește și va fi mai dificil de rezolvat, iar gradul de risc major este mai puțin prevalent în cadrul acestui grup. Dacă valoarea indicelui  $b$  scade, probabilitatea apariției riscului va crește.<sup>140</sup>

**Nota bene!** A nu se confunda probabilitatea de a răspunde corect la o întrebare (marcată pe axa verticală a graficului), cu probabilitatea apariției riscului ce este invers proporțională cu dificultatea (datorită faptului că formula matematică este o fracție cu numărătorul 1).

133 G. Giguère, P. Lussier, *Debunking the psychometric properties of the LS/CMI: An application of item response theory with a risk assessment instrument*, "Journal of Criminal Justice", vol. 46, 2016, p. 209.

134 *Ibidem*.

135 Disponibil online: <https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory>, accesat la data 15.09.2020.

136 R. Bertrand, J.G. Blais, *Modèles de mesure: L'apport de la théorie des réponses aux items*, Ed. Presses de l'Université du Québec, Sainte-Foy, Canada, 2004, p. 131.

137 G. Giguère, P. Lussier, *op. cit.*, p. 211.

138 F. B. Baker, S. H. Kim, *The Basics of Item Response Theory Using R*, Ed. Springer, Pennsylvania, USA, 2017, p. 3.

139 *Ibidem*.

140 G. Giguère, P. Lussier, *op. cit.*, p. 211.

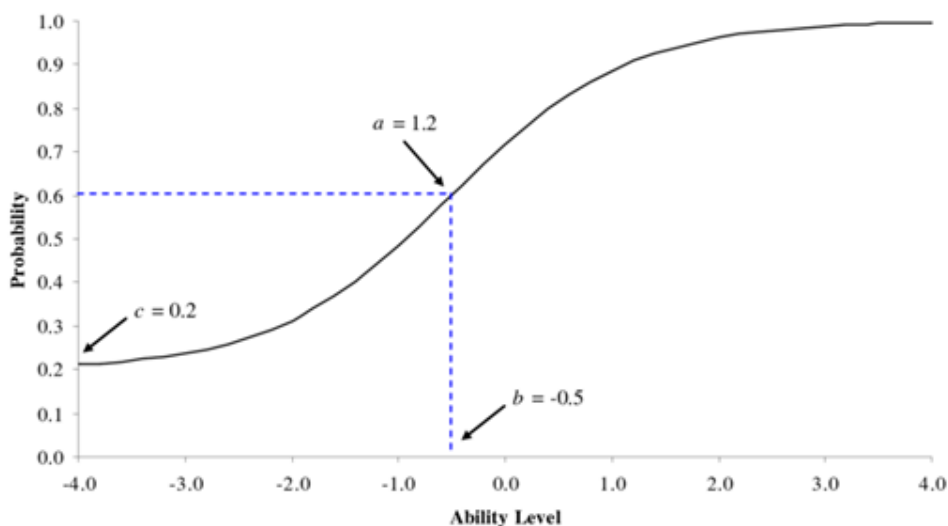


Fig. 6 Reprezintă graficul IRT<sup>141</sup>

vi. *LS/RNR Level of Service/Risk, Need, Responsivity*, este un instrument ce cuprinde 3 caracteristici fundamentale. Principiul *riscului*, se referă la faptul că nivelul de gestionare și intervenție al cazurilor ar trebui să fie proporționale cu riscul de recidivă (adică infracțiuni cu risc ridicat necesită o intervenție intensivă din partea autorităților, în timp ce infractorii cu un risc scăzut necesită intervenții minime).<sup>142</sup> Principiul *necesităților* arată că *intervențiile trebuie îndreptate înspre elementele asociate cu reducerea riscului de recidivă*.<sup>143</sup> Printre trăsăturile care sunt denumite **nevoi criminogene** se află: antecedente penale, personalitate antisocială, atitudine/credințe sau valori care sunt proinfracționale, familie disfuncțională sau probleme în căsnicie, lipsa unor activități sociale și recreative etc.<sup>144</sup> Principiul *responsivității* subliniază utilizarea metodelor de intervenție bazate de dovezi pentru reducerea recidivei, care se întemeiază în primul rând pe modele de învățare cognitiv-socială și care arată importanța adaptării remediilor corelate cu stilul de învățare al infractorului (de exemplu, funcționarea intelectuală).<sup>145</sup>

vii. *ORAS – Ohio Risk Assessment System*. Acest software evaluează indivizii în mai multe etape din sistemul de justiție penală, respectiv, faza dinainte de procesul propriu-zis, supravegherea din comunitate, faza de detenție și reintrarea în închisoare. În al doilea rând, promovează coerența și obiectivitatea în evaluarea riscului și a necesităților infractorilor în sistemul justiției penale.<sup>146</sup> Algoritmul constă în 6 instrumente, respectiv: *Pre-Trial Assessment Tool* (PAT), *Misdemeanor Assessment Tool* (MAT), *Community Supervision Tool* (CST), *Prison Intake Tool* (PIT), *Reentry Tool* (RT), și *Supplemental Reentry Tool* (SRT). În plus, față de acestea există și 3 instrumente de screening, *Misdemeanor Screening Tool* (MST), *Community Supervision Screening Tool* (CSST) și *Prison Screening Tool* (PST) care sunt concepute pentru a permite specialiștilor să determine probabilitatea de a recidiva pentru un număr mare de infractori relativ repede.<sup>147</sup> Toate instrumentele arătate sunt algoritmi

141 L. Keng, *A comparison of the performance of testlet-based computer adaptive tests and multistage tests*, Dissertation Presented to the Faculty of the Graduate School of The University of Texas at Austin in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy, 2008, p. 12.

142 D. Canales, M. A. Campbell, R. Wei, A. Totten, *Prediction of General and Violent Recidivism Among Mentally Disordered Adult Offenders. Test of the Level of Service/Risk-Need-Responsivity (LS/RNR) Instrument*, "Criminal Justice and Behavior", vol. 41, nr. 8, 2014, p. 972.

143 *Ibidem*.

144 D. A. Andrews, J. Bonta, S. J. Wormith, *The recent past and near future of risk and/or need assessment*. "Crime & Delinquency", vol. 52, nr. 1, 2006, p. 7.

145 D. Canales, M. A. Campbell, R. Wei, A. Totten, *op. cit.*, p. 972.

146 J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, Z. Hamilton, *op. cit.*, p. 148.

147 *Ibidem*.

care evaluează atât riscurile statice cât și dinamice.

**viii. OST – Offender Screening Tool** este un instrument de evaluare a riscului inculpatului de a recidiva și a nevoilor criminogene. Conține 9 categorii și 42 de elemente de evaluare. Astfel, fiecare element este analizat pe baza informațiilor disponibile, a unui interviu direct și a unei expertize a personalului calificat care se bazează pe o discuție cu infractorul. Nu ar trebui ca interviul să conștie într-un interogatoriu, care se caracterizează prin utilizarea pe scară largă a întrebărilor de tip închis (da / nu) (de exemplu, ”nu aveți probleme fizice, nu-i așa?”, ”aveți destui bani pentru satisfacerea nevoilor?” etc.) ci să se utilizeze întrebările de tip deschis, care pot oferi o plajă largă de răspunsuri.<sup>148</sup> Referitor la modalitatea de punctare a elementelor, acestea sunt realizate mai degrabă pe tipare de comportament decât pe incidente individuale. Ca regulă generală, utilizatorul OST ar trebui să aibă cel puțin unul sau două motive pentru care fiecare element OST este notat așa cum este. În cele din urmă, ar trebui depuse eforturi pentru a înscrie fiecare element OST pe baza unor dovezi independente, mai degrabă decât să se utilizeze un singur episod comportamental pentru a înscrie mai multe elemente OST.<sup>149</sup>

**ix. STRONG – Static Risk and Offender Needs Guide.** Este clasificat ca un instrument general de predicție a recidivei; cu toate acestea, domeniul său de aplicare a fost extins pentru a oferi o abordare sistemică, previzionând o varietate de rezultate multiple în etapele din sistemul de justiție penală (de exemplu, neprezentarea în cazul citării, infracțiuni, recidivă etc.).<sup>150</sup> Programul începe cu o evaluare cu 26 de întrebări asupra factorilor „statici” și a scorurilor care determină riscurile viitoare în trei domenii: a) infracțiune simplă, b) infracțiune non-violentă și c) infracțiune violentă. Aceste scoruri de risc conduc la clasificarea infractorilor în unul din cele cinci niveluri:<sup>151</sup>

- Risc ridicat de violență
- Risc ridicat cu privire la infracțiunile împotriva patrimoniului
- Risc ridicat cu privire la droguri
- Risc moderat
- Risc scăzut

Ulterior, fiecare infractor cu risc ridicat primește o evaluare a nevoilor criminogene, atinsă printr-un sondaj de 70 de întrebări care acoperă realizările sociale, sistemele de sprijin, trăsăturile cognitive și de personalitate.<sup>152</sup> În cele din urmă, se vor recomanda anumite practici prin care se va urmări ca riscul dinamic de săvârșire a infracțiunilor să fi diminuat.

**x. HART – Harm Assessment Risk Tool.**<sup>153</sup> Instrumentul de evaluare a riscului (HART) prezice riscul unei persoane de a recidiva bazat pe 34 de variabile, dintre care majoritatea se referă la istoricul infracțional. Modelul utilizează algoritmul *random forest forecasting* (metoda arborilor decizionali) pentru a evalua probabilitatea ca o persoană să comită o infracțiune gravă, de pericolozitate medie sau nici o faptă penală în următoarele 24 de luni. Pe baza acestei predicții, indivizii sunt apoi clasificați în grupe de risc ridicat, moderat sau risc scăzut. HART, a fost instruit folosind cinci ani de date (2008- 2012), totalizând în jur de 104.000 evenimente din perioada de probațiune, având capacitatea de a echilibra costurile diferențiale ale erorilor într-un mod în care metodele tradiționale de predicție nu reușesc.

148 Disponibil online: <https://www.azcourts.gov/apsd/Evidence-Based-Practice/Risk-Needs-Assessment/Offender-Screening-Tool-OST>, accesat la data 16.09.2020.

149 *Ibidem*.

150 J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, Z. Hamilton, *op. cit.*, p. 199.

151 <https://www.assessments.com/purchase/detail.asp?SKU=5205#:~:text=Developed%20in%20collaboration%20with%20the,planning%20system%20for%20adult%20offenders>, accesat la data 16.09.2020.

152 *Ibidem*.

153 A. Babuta, *Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making*, ”Rusi”, vol. 38, nr. 2, 2018, p. 1.

**xi. RisCanvi.**<sup>154</sup> RisCanvi, este un protocol de evaluare a riscurilor pe mai multe niveluri pentru prevenirea violenței în închisoare, introdus în sistemul penitenciar din Catalonia. RisCanvi a fost conceput pentru a atinge două obiective principale. Primul dintre acestea a fost îmbunătățirea predicțiilor individualizate ale riscului violențelor viitoare sub forma unor noi infracțiuni, agresiuni *îndreptate* împotriva propriei persoane, comportamente violente în interiorul unităților penitenciare, și pentru a prezice probabilitatea încălcării restricțiilor impuse în perioada de liberare condiționată. În al doilea rând, implementarea algoritmului a fost destinată să generalizeze utilizarea instrumentelor pentru evaluarea riscurilor ca obișnuință pentru profesioniștii din penitenciare și astfel să introducă cele mai bune practici în managementul informațiilor pentru luarea deciziilor. Schimbul de informații, asigură actualizări constante ale datelor și consolidarea instrumentelor tehnice a căror validitate este dovedită empiric.

Acești algoritmi de evaluare a riscului încorporează sisteme de practici bazate pe dovezi (*evidence based practice*) și combinații sofisticate de predicții a riscurilor statice și dinamici. Factorii de risc static sunt deseori asociați cu istoricul penal, vârsta la care a fost prima dată arestată persoana, sexul, problemele anterioare cu abuzul de alcool sau substanțe, tulburări mentale sau încălcarea restricțiilor impuse de autorități.<sup>155</sup> Aceștia fiind corelați de cele mai multe ori cu factorii de risc dinamici, nu sunt vizați strict individual pentru că nu pot fi schimbați.<sup>156</sup> Factorii de risc dinamici contribuie la evaluarea riscului general, având o caracteristică variabilă, adică se pot modifica de-a lungul timpului. De pildă, aceștia pot fi: vârsta, statutul social, locul de muncă, situația efectuării în prezent a unui tratament pentru dezintoxicarea de alcool sau substanțe.<sup>157</sup> Unii dintre factorii dinamici sunt denumiți *nevoi criminogene*<sup>158</sup> care dacă sunt soluționate vor reduce probabilitatea de a avea o activitate infracțională în viitor.<sup>159</sup> Un exemplu elocvent<sup>160</sup> ar putea fi acea situație în care un consumator cronic de substanțe psihoactive ar fi plasat într-un program de dezintoxicare, ceea ce în lanțul de cauzalitate al evenimentelor ar crea, până la urmă, și reducerea riscului de recidivă.

Instrumentele la care am făcut referire anterior vin în sprijinul judecătorului și încearcă să rezolve dilemele cu care acesta se confruntă atunci când trebuie să decidă în cazul unei rețineri sau al unei arestări preventive dacă suspectul/înculpatul reprezintă un *pericol social*. Mai mult, îl poate ajuta să stabilească dacă infractorul va executa pedeapsa în închisoare sau va fi suspendată sub supraveghere, ori, care va fi durata pedepsei pe care urmează să o aplice în funcție de pericolul de recidivă.

## 2. Teoria pedepsei

În funcție de cum se raportează judecătorul la teoria primară a pedepsei, va soluționa conflictul social pe care îl are înaintea sa, însă trebuie să menționez că există puține evidențe în sensul unei corelări directe dintre o pedeapsă ridicată și un risc scăzut de recidivă.<sup>161</sup>

Având în vedere că rezultatele la care se ajunge în urma acestor evaluări ale algoritmilor sunt foarte sensibile și creează un impact puternic în societate, Jim Greiner, profesor la Harvard Law School, și Chris Griffin, directorul de cercetare la același institut, subliniază că ”pentru ca aceste instrumente de evaluare să

154 J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, Z. Hamilton, *op. cit.*, p. 257.

155 J. Austin, *The Proper and Improper Use of Risk Assessment in Corrections*, ”Federal Sentencing Reporter”, vol. 16, nr. 3, 2004, p. 5.

156 N. James, *Risk and Needs Assessment in the Criminal Justice System*, ”Congressional Research Services”, 2018, p. 3.

157 *Ibidem*.

158 D. J. Simourd, *Use of Dynamic Risk/Need Assessment Instruments Among Long-Term Incarcerated Offenders*, ”Criminal Justice and Behavior”, vol. 31, nr. 3, 2004, p. 306.

159 Pew Center on the States, *Risk/Needs Assessment 101: Science Reveals New Tools to Manage Offenders*, Public Safety Performance Project, nr. brief, September 2011, p. 2, disponibil [online](#) - accesat la data 15.09.2020.

160 D. Kehl, P. Guo, S. Kessler, *op. cit.*, p. 9.

161 S. B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, ”Stanford Law Review”, vol. 66, nr. 4, 2014, pp. 855-856.

funcționeze este bine să definim foarte exact ce reprezintă succesul acestora și să îl maximizăm în funcție de obiectiv.”<sup>162</sup>

Astfel, Vintilă Dongoroz, în abordarea sa cu privire la filosofia pedepsei susținea următoarele: ”Când însă reacțiunea represivă a început să fie reglementată, cu drept cuvânt, mintea iscoditoare a omului a trebuit să-și fi pus întrebarea: dar de ce reprimăm?, care este rostul represiunii?, care este justificarea ei?, care este fundamentul său? De răspunsul dat la aceste chestiuni depindea nu numai satisfacerea unei curiozități spirituale, ci însăși chipul în care avea să fie reglementată represiunea, fiindcă este rațional că atunci când cunoaștem rostul unei manifestări, să căutăm a o modela în așa fel, încât acel rost să fie realizat în chipul cel mai desăvârșit.”<sup>163</sup>

În doctrina recentă s-a punctat în același sens: ”Întrebarea ”*de ce pedepsim?*” nu este retorică și nu reprezintă nici un exercițiu de filosofie în sine. Deși la prima vedere este o întrebare de apanajul politicii penale și un răspuns adecvat ne-ar îndreptați la soluții legislative mai eficiente, ea are implicații mult mai profunde, pentru că ne-ar putea conduce la concluzia inutilității nu doar a pedepsei, ci chiar a oricărei forme de organizare a statului.”<sup>164</sup>

De asemenea, Vintilă Dongoroz a făcut distincția absolut necesară între fundament, scop și funcția pedepsei: „Fundamentul pedepsei ne arată cauza pentru care pedepsim, finalitatea pedepsei exprimă efectul pe care îl urmărim pedepsind. De aceea, între fundamentul și finalitatea pedepsei există o legătură indisolubilă, așa încât de conceptul celui dintâi va depinde caracterul celei de a doua. (...) Trebuie, de asemenea, făcută o distincție între scopul (finalitatea) pedepsei și funcțiile acesteia. Funcția nu este decât modalitatea sub care o pedeapsă lucrează în vederea realizării finalității ce se atribuie represiunii. Între funcție și finalitate există un raport ca acela de la mijloc la scop, funcția fiind mijlocul prin care se realizează scopul, deci finalitatea. Aceeași legătură indisolubilă care există între fundamentul pedepsei și finalitatea sa există și între finalitatea pedepsei și funcțiile sale, știut fiind că scopul nu poate fi întreg și corect realizat decât atunci când mijloacele întrebunțate sunt perfect adecvate lui.”<sup>165</sup>

Astfel, individualizarea pedepsei prin determinarea gravității, duratei ori felul executării acesteia se va face prin armonizarea cu una dintre structurile teoriilor pedepsei, respectiv: descurajarea, reabilitarea, retribuția, izolarea de societate (incapacitatea) sau teoria mixtă etc.

În literatura de specialitate,<sup>166</sup> se consideră că există trei orientări majore de justificare a represiunii penale, ca instrument etatizat de control social: sisteme retributive (teorii absolute), sisteme utilitariste (teorii relative) și sisteme mixte (teorii eclecticice).<sup>167</sup>

Teoria privind *descurajarea* sau prevenirea de infracțiuni cuprinde, în esență, ideea că pedeapsa reprezintă o măsură pentru a preveni comiterea unei infracțiuni și descurajarea infractorilor de la a recidiva.<sup>168</sup> Prin aplicarea de pedepse se oferă un exemplu legat de capacitatea generală a societății de a controla comportamentele membrilor ei și de a capacita un răspuns pentru orice acțiune deviantă.<sup>169</sup>

---

162 A se vedea interviul cu J. Greiner și C. Griffin, ”Harvard Law Today”, 2016.

163 V. Dongoroz, *Drept penal*, Institutul de Arte Grafice, București, 1939, p. 42.

164 C. Rotaru, *Fundamentul pedepsei. Teorii moderne*, Ed. C.H. Beck, București, 2006, p. 79.

165 V. Dongoroz, Comentariu în I. Tanoviceanu, *Tratat de drept și procedură penală*, Ed. a II-a revăzută și completată, vol. III, București, 1924-1927, p. 197.

166 I. Tanoviceanu, *op. cit.*, vol. I, p. 31; Tr. Pop, *Curs de Criminologie*, Institutul de Arte Grafice „Ardealul”, Cluj, 1928, p. 189; V. Dongoroz, *op. cit.*, p. 49; C. Rotaru, *op. cit.*, p. 79.

167 M. Dunea, *Analiza receptării în doctrina juridico-penală a principalelor teorii referitoare la fundamentul dreptului de a pedepsi, în relație cu esența noțiunii de impunitate*, Analele Științifice ale Universității „Al.I.Cuza” Iași, Tomul LVII, Științe Juridice, 2011, p. 69.

168 O. R. Rusu, *Etica Pedepsei Privative de Libertate. Teză de Doctorat*, Universitatea „Alexandru Ioan Cuza” Din Iași Facultatea de Filosofie și Științe Social – Politice, Iași, 2018, p. 58.

169 I. Băținaș, *Filosofia Pedepsei. Rezumatul Tezei de Doctorat*, Universitatea Babeș-Bolyai Cluj – Napoca Școala Doctorală de Relații Internaționale și Studii De Securitate, Cluj- Napoca, 2016, p. 15.



Potrivit teoriei *reabilitării*, se are în vedere prevenirea săvârșirii infracțiunilor prin modificarea comportamentului făptuitorului.<sup>170</sup> Se susține că infracțiunea este un efect ce are la baza mai multe cauze puternic viciate, care trebuie să fie eradicate. Scopul acesteia este a modifica comportamentul făptuitorului și a-l reintegra în societate ca un cetățean productiv, afirmându-se că această abordare ar fi mult mai eficientă decât clasicele pedepse.<sup>171</sup> În acest sens se pledează pentru a se recurge la o intervenție de schimbare a comportamentului făptuitorilor făcându-i cetățeni normali ai societății.<sup>172</sup>

Potrivit teoriei *retributive* explicația actului pedepsirii nu necesită recurgerea la elemente exterioare sistemului de referință în care se manifestă noțiunea analizată (elemente extrapunitive), ea fiind identificată în valoarea intrinsecă asociată însăși pedepsei în momentul impunerii sale.<sup>173</sup> Cu alte cuvinte, ceea ce determină măsura represivă și ceea ce se urmărește prin aplicarea ei sunt unul și același lucru: *pedeapsa este un scop în sine*.<sup>174</sup>

Teoria *izolării de societate* a făptuitorului abordează problema protecției sociale urmărite prin aplicarea pedepsei privative de libertate, prin izolarea condamnatului de societate, astfel urmărindu-se reducerea posibilității ca făptuitorul să comită noi infracțiuni.<sup>175</sup>

Teoria *utilitară* statuează că ”scopul pedepsei este utilitatea socială, iar pedeapsa se dă celui care a comis infracțiunea, nu pentru a repara răul făcut, ci pentru a îndrepta pe cei greșiți și pentru a servi, de pildă, altora, pentru binele obștesc.”<sup>176</sup> Platon a fost cel care a dezvoltat substanțial pe marginea acestei idei în Gorgias, în opera lui Aulus Gellius, *Noctes Atticae*: ”Ceea ce trebuie oricărui om supus pedepsei de un judecător care știe să o aplice cu dreptate este de a deveni mai bun și de a scoate un oarecare folos din pedeapsa lui sau de a servi cel puțin de exemplu altora, pentru ca fiind martori la ceea ce suferă, să-i sperie temerea unei astfel de soarte.” Dintre reprezentanții romani ai teoriei, Cicero<sup>177</sup> susținea că orice iertare și pedeapsă este pentru binele public. Seneca<sup>178</sup> spunea că pedeapsa nu se dă din cauza mâniei, ci pentru pază (...) ea nu va privi pe cele trecute ci pe cele viitoare. Ulterior îl citează pe Platon, ”Niciun om înțelept nu pedepsește pe cineva fiindcă a greșit, dar ca să nu mai greșescă.”<sup>179</sup> Juristul italian Giovanni Carmignani, arată în secolul 19 că ”obiectul pedepsei nu este de a reprima răutatea infracțiunii deja comisă, ci de a-i împiedica pe cetățeni de a comite altele. De aceea pedeapsa nu se dă pentru a răzbuna infracțiunea deja comisă, ci, ca să nu se mai întâmple în viitor infracțiuni. În cele din urmă, explicația săvârșirii infracțiunii aparține regulilor moralei teologice și nu celei politice”<sup>180</sup> De asemenea, Jeremy Bentham rezumă în câteva cuvinte această idee: ”Ceea ce justifică pedeapsa, e utilitatea sa majoră, sau mai bine zis necesitatea ei. Infractorii sunt inamici publici. Unde e trebuință ca inamicii să consimtă să fie dezarmați și încarcerați?”<sup>181</sup>

Teoria *justiției absolute* a fost formulată în contradicție cu sistemul utilitarist, și reprezintă teoria retribuției care are ca scop *expitațiunea* (ispășirea pedepsei).<sup>182</sup> Se consideră ca bazele sistemului au fost

170 O.R. Rusu, *op. cit.*, p. 59.

171 *Ibidem*, p. 60.

172 R. Canton, *Why punish? An Introduction to the Philosophy of Punishment*, Ed. Macmillan Publishers Limited, London, 2017, p. 120.

173 C. Rotaru, *op. cit.*, p. 115.

174 O.R. Rusu, *op. cit.*, p. 70.

175 *Ibidem*, p. 59.

176 V. Dongoroz, Comentariu în I. Tanoviceanu, *op. cit.*, vol. I, p. 42.

177 Cicero, *De Officiis*, ”omnis autem et animadversio et castigatio contumelia vacare debet neque ad eius, qui punitur aliquem aut verbis castigat, sed ad rei publicae utilitatem referri.” Disponibil online: <https://www.thelatinlibrary.com/cicero/off1.shtml>, accesat la data 18.09.2020.

178 Seneca, *De Ira*, disponibil online: <https://www.thelatinlibrary.com/sen/sen.ira1.shtml>, accesat la data 18.09.2020.

179 *Ibidem*.

180 G. Carnigani, Comentariu în I. Tanoviceanu, *op. cit.*, vol. I, p. 51.

181 J. Bentham, *Theories des pelnes et de recompenses*, Ed. De L'imprimerie de Vogel Et Schulze, 13, Oxford Street, tom. 1, 1811, p. 7.

182 V. Dongoroz, Comentariu în Ion Tanoviceanu, *op. cit.*, vol. I, p. 54.

teoretizate de Immanuel Kant prin lucrarea *Principii metafizice de drept* publicată în anul 1797. Potrivit lui Kant, justificarea pedepsei vine din dreptatea ei, din legea morală a ispășirii, ca binele să fie răsplătit cu bine, iar răul cu rău.<sup>183</sup> Acest sistem propune într-o anumită măsură revenirea la legea talionului, întrucât utilitatea socială, conservarea binelui public sau intimidarea infractorilor nu mai reprezintă un scop al pedepsei, ci, noul obiectiv va fi format din restaurarea echilibrului prin pedepsirea culpabilului.<sup>184</sup> În cele din urmă, și acest sistem și-a dovedit falimentul, întrucât, în fond, o relație de egalitate între pedeapsă și infracțiune nu este posibilă în niciun fel.<sup>185</sup> Astfel, Émile Garçon se întreba ”Cum ar fi posibilă stabilirea, chiar și prin aproximare, a unei proporții între pedeapsă și greșeala morală?”<sup>186</sup> Enrico Ferri, un distins jurist italian, are un comentariu referitor la aceeași dilemă: ”Nici un judecător din lume nu poate să măsoare greșeala morală a unei persoane, fiindcă nu se pot cunoaște condiționalitățile infinite impuse de ereditatea psiho-fizică, de viața socială și familială, de starea morală, economică și intelectuală, toate acestea în afară de alte infinite influențe datorate contextului social, a căror consecință fatală este infracțiunea”.<sup>187</sup> Prin urmare, ne putem pune întrebarea firească, dacă nu poate fi stabilită o egalitate formală între infracțiune și pedeapsă, cum se poate lua în considerare justiția absolută?<sup>188</sup>

*Teoria mixtă* sau eclectică s-a format din fuziunea sistemului justiției absolute reprezentat de Immanuel Kant cu sistemul utilitarist al lui Bentham. În esență, fundamentul acestei abordări are la bază ideea că dreptul de a pedepsi derivă din justiția absolută și e limitat prin utilitatea socială sau, privind din altă perspectivă, că pedeapsa se întemeiază pe utilitatea socială ținându-se seama de justiția absolută.<sup>189</sup> Cei dintâi susținători ai acestui sistem sunt, în Franța, Victor de Broglie în *Revue française*, anul 1828 și contele Pelegrino Rossi în *Traite de droit penal*, în anul 1829. În același sens, autori români s-au declarat partizani, dintre care îi amintim pe: V. Petroni<sup>190</sup> prin lucrarea *Comentariile dreptului penal*, 1857 și Constantin Eraclide<sup>191</sup> prin *Studii practice asupra comportamentului criminal în 1865*, fiind, totodată, și primii comentatori ai codului penal.

În principiu, a fost arătat ca există o relație direct proporțională între recidivă și obiectivele reabilitării sau incapacitării: de pildă, riscul ridicat de recidivă al unei persoane poate constitui o justificare pentru o detenție de lungă durată sau permanentă, în scopul prevenirii viitoarelor acte infracționale, însă, persoanele a căror probabilitate de a recădea sub incidența legii penale este scăzută, vor putea fi luați în considerare pentru forme de pedeapsă mai puțin severe<sup>192</sup> sau măsuri preventive fără o privare de libertate substanțială.

## V. ALGORITMI PREDICTIVI ȘI RESPECTAREA DREPTURILOR FUNDAMENTALE ALE OMULUI

### 1. Prezența biasului în sistemele informatice

Reprezentarea noastră asupra sistemelor informatice este că acestea au o natură rigidă, total neperturbată de contextele sociale, economice, politice, incapabile de a fi vectorizate de erorile umane de etică, cum ar fi discriminarea, restrângerea libertății sau vieții private etc. În esență, ne închipuim că în afară de restrângerea aptitudinilor strict tehnice ale unui algoritm prin raportare la un scop predefinit, nu pot exista *fracturi de conștiință* în luarea unor decizii, așa cum se întâmplă în cazul ființelor cu discernământ și înzestrate cu spirit

183 *Ibidem*, p. 55.

184 I. Kant, *Metaphysische Anfangsgründe Der Rechtslehre*, Ed. Friedrich Nicolovius, Königsberg, 1797, p. 193.

185 V. Dongoroz, Comentariu în Ion Tanoviceanu, *op. cit.*, vol. I, p. 57.

186 É. A. Garçon, *Le droit pénal: origines--évolution--état actuel*, Ed. Payot & cie, Paris, 1922, p. 144.

187 E. Ferri, *Relazione sul Progetto Preliminare di Codice Penale*, Ed. Vallardi, Milano, 1921, p. 11.

188 V. Dongoroz, în I. Tanoviceanu, *op. cit.*, vol. I, p. 57.

189 *Ibidem*, p. 58.

190 V. Petroni, *Comentariile dreptului penal*, București, 1857, p. 73.

191 C. Eraclide, *Studii practice asupra comportamentului criminal*, București, 1865, p. 58.

192 B. E. Harcourt, *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, Ed. The University of Chicago Press, Chicago, 2007, pp. 32-33; D. Kehl, P. Guo, S. Kessler, *op. cit.*, p. 13.

moral. Vom observa în continuare că acești algoritmi predictivi au aptitudinea să discrimineze, sau, mai degrabă, să aibă o anumită *preferință* în funcție de obiectivul pe care îl stabilim pentru a fi atins.

Termenul *bias* se poate defini ca ”preferința neintenționată a algoritmului pentru o anumită predicție în detrimentul altora, care are ca rezultat implicații improprii din punct de vedere legal sau etic.”<sup>193</sup> Prin urmare, poate fi considerat acel program de calculator care își asumă prejudiciul creat de cel care l-a implementat ori datele viciate introduse.<sup>194</sup>

Cu toate că fenomenul biasului în sistemele informatice a fost abordat într-un sens mai restrâns și, oarecum tangențial de către Deborah G. Johnson și John Mulvey<sup>195</sup> și de James Moor,<sup>196</sup> cea mai complexă lucrare care a apărut pentru prima dată și a teoretizat acest subiect este realizată de Batya Friedman și Helen Nissenbaum în anul 1996.<sup>197</sup>

”În consecință, folosim termenul *prejudecată (BLAS)* pentru a ne referi la sisteme informatice care discriminează sistematic și în mod nedrept împotriva anumitor indivizi sau grupuri de indivizi în favoarea altora. Un sistem discriminează în mod nedrept dacă neagă o oportunitate sau un bun sau dacă atribuie un rezultat nedorit unei persoane sau unui grup de indivizi pe motive nerezonabile sau inadecvate.”<sup>198</sup>

De pildă, poate fi cazul în care instrumentul de evaluare a riscului COMPAS - stabilește că afro-americanii (persoanele de culoare) au un risc de recidivă mai ridicat decât *albi*<sup>199</sup> ori, când se analizează un risc de credit<sup>200</sup> pentru o persoană care dorește să solicite un împrumut și va fi respinsă în urma unui scor slab de returnare al împrumutului din cauza faptului că locuia într-o zonă a orașului cu venituri mici sau o infrafracționalitate mare.

A fost identificat în literatura de specialitate<sup>201</sup> că există mai multe forme ale biasului, și anume: Bias preexistent, Bias tehnic și Bias emergent.

Astfel, substratul *biasului preexistent* rezidă în comportamentele și prejudecățile sociale adânc înrădăcinate în cultura unui popor. Acest tip de bias poate fi *individual*, când atitudinea discriminatoare și independentă a celui care programează sistemul se reflectă în deciziile pe care le va lua algoritmul. Este cazul unui inginer care încorporează prejudecățile sale rasiale în specificațiile pentru softul de acordare a împrumutului sau de stabilire a unui scor de recidivă. Biasul poate fi și *colectiv* când provine din societate, cum ar fi situația unor instituții care încorporează prejudecăți bazate pe gen, sau creează sisteme de învățământ ce favorizează mai degrabă fetele decât băieții.

*Biasul tehnic* rezultă din limitarea tehnologică la care a ajuns în punctul său maxim în momentul utilizării softului. Surse ale erorii mai pot fi și algoritmi decontextualizați în procesul de atribuire al sensului social al unor evenimente, neconformitatea generării unui număr aleatoriu, ori încercarea de a compatibiliza construcțiile umane cum sunt discursul sau intuiția cu calculatoarele.<sup>202</sup>

193 A. Shadowen, *Ethics and Bias in Machine Learning: A Technical Study of What Makes Us 'Good'*, City University of New York, 2017, p. 5, disponibil [online](#) - accesat la data 20.09.2020.

194 G. Megan, *Racist in the Machine: The Disturbing Implications of Algorithmic Bias*. Retrieved, ”World Policy Journal”, vol. 33, nr. 4, 2017, disponibil [online](#) - accesat la data 22.09.2020.

195 D. G. Johnson, J. M. Mulvey, *Computer decisions: Ethical issues of responsibility and bias*. *Statistics and Operations*, Res. Series SOR-93-11, Dept. of Civil Engineering and Operations Research, Princeton Univ., Princeton, 1993.

196 J. Moor, *What is computer ethics?*, ”Metaphilosophy”, vol. 16, nr. 4, 1985, disponibil [online](#) - accesat la data 22.09.2020.

197 B. Friedman, H. Nissenbaum, *Bias in Computer Systems*, ”ACM Transactions on Information Systems”, vol. 14, nr. 3, pp. 330-347, 1996, disponibil [online](#) - accesat la data 22.09.2020.

198 *Ibidem*, p. 332.

199 J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks*, ”Pro Publica”, disponibil [online](#) - accesat la data 23.09.2020.

200 B. Friedman, H. Nissenbaum, *op. cit.*, p. 333.

201 *Ibidem*, p. 330.

202 *Ibidem*, p. 335.

*Biasul emergent* apare de obicei la un anumit interval temporar ulterior momentului realizării proiectării, având drept cauză modificarea unor elemente în structura valorilor culturale, a cunoștințelor înregistrate la nivel general al populației. Astfel, o nouă viziune științifică rezultată în urma unei descoperiri de natură medicală nu poate fi încorporată în vechiul software, discriminând în cele din urmă o categorie de pacienți, care, de altfel, s-ar încadra perfect în noua paradigmă. O altă situație este aceea în care algoritmul a fost conceput având la bază un alt reper de valori decât cel pentru care este utilizat în concret. A mai fost afirmat în doctrină că biasul poate să rezulte din datele de introducere care sunt incomplete sau inadecvate. Prin subreprezentarea sau excluderea anumitor grupuri sau subgrupuri marginalizate social, acest tip de „eroare de eșantionare” duce la rezultate slab calibrate care se intensifică mai degrabă decât să reducă marginalizarea.<sup>203</sup> Chiar dacă s-ar elimina informațiile explicite cu privire la rasă, sex, vârstă, statut socio-economic din aceste modele, o parte din datele viciate remanente continuă să se coreleze cu aceste conținuturi discriminatorii.<sup>204</sup>

Pentru a ajunge la concluzia că o eroare este una de tip bias trebuie ca discriminarea să fie implementată la nivel de sistem, adică să fie una generalizată, nicidecum să fie rezultatul unor apariții singulare. Prin urmare trebuie constatată o practică în acest sens. În plus, prejudecata trebuie să se materializeze într-un rezultat neloyal.<sup>205</sup>

Din punct de vedere statistic, dacă în America, de pildă, rata de recidivă ar fi mai ridicată în privința persoanelor de culoare, introducerea unui astfel de parametru în sistemul de calcul al unui soft ce poate influența pedeapsa aplicată ori stabilirea pericolului social pentru a dispune măsura arestului preventiv, ar pune în discuție discriminarea subiecților procesuali.

De asemenea, în situația asigurării autoturismelor în caz de accidente<sup>206</sup>, sunt situații în care, femeile și bărbații nu se comportă la fel de eficient când ne referim la evitarea unor evenimente nefericite, iar firmele de asigurări ar trebui să le trateze diferit, însă fără a introduce prejudecăți sistematice care ar avea capacitatea de a discrimina.

Într-o lume ideală, sistemele inteligente și algoritmi lor ar fi obiectivi. Din păcate, acestea sunt construite de noi și, ca urmare, ajung să ne reflecte prejudecățile. Înțelegând părtinirea în sine și sursa problemelor, putem proiecta în mod activ sisteme pentru a le evita.<sup>207</sup>

## 2. Respectarea drepturilor fundamentale ale omului

Utilizarea AI, pe lângă beneficiile aduse, care sunt de necontestat, expune societatea la o serie de riscuri generate de o serie de factori. Unii dintre aceștia pot consta în necunoașterea suficient de profundă a sistemului unor algoritmi de inteligență artificială, aspect care reprezintă și cauza pentru care a apărut termenul de *black box*. În același sens, regăsim fenomenul de *bias* sau discriminare explicat anterior. Dar, probabil, cea mai periculoasă formă de generare a unor așa-zise ”erori” care aduc atingere drepturilor și libertăților fundamentale ale cetățeanului este nevoia de control total al populației, de supraveghere prin diverse mijloace, și în ultimă instanță de supunere chiar și a celui mai obișnuit om. Termenul de eroare este folosit într-un sens peiorativ, deoarece rezultatul nu este o culpă la care se ajunge în cazul supravegherii și dirijării maselor, ci este scopul unor interese de mare anvergură.

### A. Dreptul la un proces echitabil și interzicerea discriminării

Acest drept fundamental este consfințit de art. 6 CvEDO. AI are aptitudinea să aducă atingere acestui paragraf prin faptul că, după cum am arătat la pct. 4, anumiți algoritmi specializați determină probabilitatea ca

---

203 I. Gabriel, *The case for fairer algorithms*, 2018, material disponibil [online](#) - accesat la data 23.09.2020.

204 *Ibidem*.

205 B. Friedman, H. Nissenbaum, *op. cit.*, p. 333.

206J. Villasenor, *Artificial intelligence and bias: Four key challenges*, 2019, material disponibil [online](#) - accesat la data 25.09.2020.

207K. Hammond, *5 unexpected sources of bias in artificial intelligence*, 2016, material disponibil [online](#) - accesat la data 25.09.2020.

o anumită persoană să recidiveze, să comită pentru prima dată o infracțiune, ori îi returnează un scor mai mare al violenței și, prin urmare, judecătorul este încurajat să îi aplice o sentință mai dură. În cazul în care subiectul procesual este împiedicat să cunoască modul în care se stabilesc scorurile de risc ori, pur și simplu este nu este suficient de temeinic cunoscută metoda de către specialiști, cred că se ridică probleme în privința respectării art. 6 Conv.EDO.

În cauza *State v. Loomis*,<sup>208</sup> Eric Loomis a fost condamnat la 6 ani de închisoare în urma utilizării unui soft de evaluare a riscului. În acea situație s-a pus problema dacă este încălcat dreptul la nediscriminare și dreptul un proces echitabil în cazul în care acest algoritm folosește genul unei persoane ca factor diferențiator de risc (factor criminogen). Curtea a respins această apărare motivând că ”inclusiunea genului în calcularea riscului promovează mai degrabă acuratețe, servește interesului instituțiilor și al acuzatului, decât ar avea un rol discriminator.”<sup>209</sup>

Consider că, se poate discuta de încălcarea art. 14 CvEDO privind discriminarea numai dacă sentința de condamnare s-a dat în măsură determinantă având la bază elementul de gen. Dacă decizia dată de judecător este întemeiată pe o știință exactă, fundamentată pe matematică, așa cum este statistica, de unde rezultă că genul are o defalcare naturală înspre comiterea de infracțiuni, și se află în combinație cu alți factori criminogeni, nu cred că ar fi incidentă o discriminare.

De asemenea, Protocolul adițional 7 art. 2 CvEDO ar putea fi încălcat dacă împotriva acestor scoruri de risc nu există o cale de atac prin care să poată fi reevaluate, luând în considerare și apărarea acuzatului.

## B. Libertatea de exprimare

În structura dreptului prevăzut de art. 10 din CvEDO se regăsesc libertatea de opinie și libertatea de informare.<sup>210</sup> Dezbaterea<sup>211</sup> despre potențialul efect negativ al motoarelor de căutare în privința respectării dreptului la informare este într-o continuă expansiune.<sup>212</sup> Un astfel de algoritm care este implicat în căutarea informației pe internet utilizează inteligența artificială sub formă de *ML- Machine Learning*. Prin modul său de funcționare, cu cât o informație este mai căutată, cu atât va apărea mai des în *feed-ul* unei persoane, restul rămânând ascunse. Acest fenomen de fragmentare și polarizare a informației a fost denumit *echo chambers*<sup>213</sup> sau *bubble filters*.<sup>214</sup> În cele din urmă modelul de față poate conduce, prin servirea unor date, conform unui program de sortare, la acțiuni politice dirijate, împlinirea unor deziderate economice sau sociale.

Prin urmare, putem considera că atât timp cât suntem împiedicați să accesăm o anumită știre, de pildă, din cauza faptului că softul a emis o predicție bazată pe statistica căutărilor noastre anterioare, acest fapt poate fi interpretat ca o restrângere a dreptului la informare. Însă, așa cum prevede art. 10 alin. (2) CvEDO restrângerea trebuie să fie impusă printr-o lege, să aibă ca scop siguranța publică, apărarea ordinii și prevenirea infracțiunilor etc., și să fie necesare într-o societate democratică. În plus, platformele Facebook și Google, folosesc acești algoritmi automați nu numai pentru filtrarea datelor ci și pentru eliminarea lor definitivă din spațiul de căutare,<sup>215</sup> iar acest lucru pune probleme serioase asupra ingerinței în dreptul la informare.

208 *A se vedea State v. Loomis*, 881 N.W.2d 749, Wisc., 2016.

209 *Ibidem*.

210 C. Bîrsan, *Convenția europeană a drepturilor omului. Comentariu pe articole*, Ed. 2, Ed. All Beck, București, 2005, p. 737.

211 *A se vedea*, David Kaye, *Report on the promotion and protection of the right to freedom of opinion and expression*, pentru a 32-a sesiune a Human Rights Council, (A/HRC/32/ 38), 2016, p. 6.

212 Committee of Experts on Internet Intermediaries (msi-net), *Study on the Human Rights Dimensions of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications*, Council of Europe, 2017, p. 17.

213 R. Saxena, *The social media “echo chamber” is real*, 2017, disponibil [online](#) - accesat la data 23.09.2020.

214 *Ibidem*.

215 Committee of Experts on Internet Intermediaries, *op. cit.*, p. 18.

### C. Dreptul la respectarea vieții private

Vaste cantități de date sunt procesate zilnic de către algoritmiile rețelelor sociale și de motoarele de căutare pe internet. Acestea sunt folosite pentru a determina profilul unei persoane din mai multe punctele de vedere: economic, social, cultural, valoric, politic etc., prin înregistrarea căutărilor unui browser de către *cookie-uri*. Softurile care se folosesc de datele generate de *cookie-uri* pot fi utilizați pentru a crea un profil de advertising, sau chiar comportamental al unui individ., generând, în cele din urmă, prin minarea datelor, un risc major de supraveghere în masă a populației. De pildă, au fost creați algoritmi care pot să estimeze cu acuratețe vârsta, genul, ocupația, și starea civilă a unei persoane numai din analiza locației telefonului.<sup>216</sup> Consiliul pentru Drepturile Omului din cadrul ONU împărtășește aceeași viziune: *”Prelucrarea datelor cu caracter personal pentru profilarea individuală poate duce la discriminare sau decizii care au potențialul de a afecta drepturile omului, inclusiv drepturile economice, sociale și culturale.”*<sup>217</sup>

În ultimul timp, tehnologia pentru recunoașterea facială pune probleme din ce în ce mai mari asupra respectării intimității. Aceasta poate fi utilizată pentru a identifica anumite persoane și a determina tiparele comportamentale uzuale ale lor. De pildă, dacă *LFR* (Live Facial Recognition) este încorporată într-un sistem de camere stradale CCTV, se pot urmări mișcările oamenilor din întregul oraș.<sup>218</sup>

În plus, dreptul la viață privată nu se referă doar la activitățile ce se desfășoară ferite de ochii publicului, ci și la scopul pentru care persoana se hotărăște să întreprindă orice acte consideră de cuviință. Astfel, dacă aceasta apreciază că motivul pentru care acționează în orice fel (de pildă, diferite întâlniri, achiziționarea unor produse, deplasări în anumite locații etc.) dorește să rămână ascuns, și cunoscut doar de ea, trebuie să îi fie respectată intimitatea.

Prin urmare, având în vedere că datele generate de activitatea de pe internet reușesc să conducă la viitoarele opțiuni politice, economice, sociale, sau în cazul softurilor de recunoaștere facială, dezvoltarea unei înțelegeri profunde a activităților zilnice ale noastre, și în urma acestor predicții este posibil să se determine și scopul oricăror acțiuni, este întemeiat să credem că dreptul la viață privată poate fi înfrânt.

## VI. CONCLUZII

Uimitoarea etapă tehnologică în care ne aflăm astăzi trebuie abordată cu prudență. Cu cât beneficiile aduse de sistemele artificiale sunt mai mari cu atât mai mult se impune ca liderii ce iau decizii în implementarea lor să fie lucizi și responsabili. Folosirea instrumentelor de evaluare a riscului penal poate constitui un prilej de optimism în măsura în care, în viitor, prin îmbunătățire, va fi posibil să se reducă rata infracționalității sau precizia unei sentințe penale. Însă, în aceeași măsură, efectele negative au un efect de contrabalansare. Prezența potențialului de întărire sau exacerbare a prejudecăților sociale și de integrare la nivel sistemic este alarmant.

De aceea, este recomandabil ca algoritmiile predictivi să fie utilizați chiar și în viitorul îndepărtat, sub forma unui instrument pe care autoritățile să se bazeze; spre a se folosi în îndrumare. Apreciez că nu trebuie să fie luați în considerare în măsură determinantă, pentru că și aceste sisteme pot fi supuse unor vicii, așa cum am arătat anterior.

Pe de altă parte, nu pot decât să observ o evoluție din ce în ce mai accelerată a softurilor de clasificare a textului pentru extragerea informațiilor și pentru a crea construcții argumentative. Atunci când vor fi utilizate la scară largă, cel mai probabil piața avocaturii și justiția în ansamblul său se vor schimba considerabil, atât din

---

216 S. M. Bellovin, R. M. Hutchins, T. Jebara, S. Zimmeck, *When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning*, "New York University Journal of Law & Liberty", vol. 8, 2014, pp. 557-558.

217 U. N. Human Rights Council Resolution on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/7, 23 Mar. 2017, p. 3.

218 M. Daragh, *Live facial recognition: the impact on human rights and participatory democracy*, disponibil [online](#) - accesat la data 23.09.2020.

punct de vedere pragmatic, dar și prin aducerea la un nou nivel al euristicii.

În mod firesc, în momentul în care în prim-plan este promovată o tehnologie atât de controversată, cercetătorii și întregul mediu academic împărtășesc o optică diferită. Stephen Hawking, scria în ultima sa publicație, ”mă tem de consecințele creării a ceva care poate atinge sau depăși oamenii. Îngrijorarea este că AI ar lua decizii proprii și s-ar reprojeta la un ritm din ce în ce mai mare. Oamenii, care sunt limitați prin evoluția biologică lentă, nu ar putea concura și ar fi înlocuiți.”<sup>219</sup>

Într-un interviu acordat recent pentru The New York Times, Elon Musk, spunea că, având prilejul de a lucra cu AI pentru a dezvolta Tesla, ”ne îndreptăm către o situație în care AI este mult mai inteligentă decât oamenii și cred că acest interval de timp este la mai puțin de cinci ani de acum. Dar asta nu înseamnă că totul se prăbușește în cinci ani. Înseamnă doar că lucrurile devin de neoprit sau ciudate.”<sup>220</sup>

Pentru a nu se ajunge la o dominare totală a AI, cei mai mari cercetători pe care îi preocupă această problemă iminentă, au formulat Principiile ASILOMAR<sup>221</sup> în anul 2017, care constau într-un număr de 23 de reguli în cercetare și etică pentru dezvoltarea unei inteligențe artificiale benefice.

Poate în acest sens ar trebui menționat și Theodore Kaczynski, de partea celor care au fost împotriva unui progres industrial și al ultra computerizării societății prin celebrul său *Manifesto- Industrial Society and Its Future*<sup>222</sup> publicat cu mult timp înaintea punerii problemei Inteligenței Artificiale de o asemenea manieră.

Yuval Noah Harari, pe de altă parte, susține că AI nu ar putea niciodată să surclaseze umanitatea: ”Științifico-fantasticul are tendința de a confunda inteligența cu conștiința și de a porni de la premisa că, pentru a egala sau a întrece inteligența umană, computerele vor trebui să dezvolte conștiință. În realitate însă nu avem nici un motiv să credem că AI va dobândi conștiință, deoarece inteligența și conștiința sunt lucruri cât se poate de diferite.”<sup>223</sup>

Ben Goertzel fondatorul SingularityNET, OpenCog, creatorul robotului Sophia, susține într-un interviu acordat pentru Forbes<sup>224</sup> că, prin compania sa, ”poate ajunge destul de probabil la AGI la nivel uman în 5-7 ani.” Pentru cei care declară că AGI suprauman este înspăimântător sau potențial dăunător, dr. Goertzel afirmă că ”aceste reacții vor părea foarte stupide oamenilor peste câteva decenii de acum înainte, în timp ce își vor desfășura viața care va deveni extrem de ușoară, fericită și fascinantă în comparație cu realitatea din 2020, prin lansarea pe scară largă a sistemelor AGI avansate pentru a gestiona producția, service-ul și toate celelalte slujbe practice cu care oamenii își ocupă acum timpul.”

Dacă temerile cercetătorilor sunt întemeiate, probabil că avântul către implementarea singularității AI reprezintă sacrificiul suprem al umanității către noua paradigmă a răsturnării vârfului lanțului trofic.

## REFERENCES / BIBLIOGRAFIE

**Andrews, D. A., Bonta, J., Wormith, S. J.,** *The recent past and near future of risk and/or need assessment,* ”Crime & Delinquency”, vol. 52, 2006.

**Angwin, J., Larson, J., Mattu S., Kirchner, L.,** *Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks,* ”Pro Publica”, disponibil online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, accesat la 08.09.2020.

219 S. Hawking, *Brief Answers To The Big Questions*, Ed. Bantam Books, New York, 2018, p. 123.

220 *A se vedea interviul acordat de Elon Musk*, disponibil [online](#) - accesat la data 24.09.2020.

221 *A se vedea Principiile Asilomar*, disponibil [online](#) - accesat la data 24.09.2020.

222 T. Kaczynski, *Industrial Society and Its Future*, ”The Washington Post”, 1995, p. 71, parag. 174.

223 Y. N. Harari, *21 de lectii pentru secolul XXI*, Ed. Polirom, 2018, p. 79.

224 *A se vedea interviul acordat de Dr. Ben Goertzel*, disponibil [online](#) - accesat la data 23.09.2020.

**Austin, J.**, *The Proper and Improper Use of Risk Assessment in Corrections*, Federal Sentencing Reporter, vol. 16, nr. 3, 2004.

**Ashley, K.**, *Artificial intelligence and legal analytics New Tools for Law Practice in the Digital Age*, Ed. Cambridge University Press, 2017.

**Azevedo, F. A.**, *Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain*, "The Journal of Comparative Neurology", vol. 513, issue 5, 2009. **Bach, N. X., Minh, N. L., Oanh, T. T., Shimazu, A. A.**, *A Two-Phase Framework for Learning Logical Structures of Paragraphs in Legal Articles*, "ACM Transactions Asian Language Information Processing", vol. 12, nr. 1, 2013.

**Babuta, A.**, *Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making*, "Rusi", vol. 38, nr. 2, 2018.

**Baker, F. B., Kim, S. H.**, *The Basics of Item Response Theory Using R*, Ed. Springer, Pennsylvania, USA, 2017.

**Bătinaș, I.**, *Filosofia Pedepsei. Rezumatul Tezei de doctorat*, Universitatea Babeș-Bolyai Cluj –Napoca Școala doctorală de relații internaționale și studii de securitate, Cluj- Napoca, 2016, p. 15.

**Birsan, C.**, *Conventia europeana a drepturilor omului. Comentariu pe articole. Ediția 2*, ed. All Beck, București, 2005.

**Bellovin, S. M., Hutchins, R. M., Jebara, T., Zimmeck, S.**, *When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning*, "New York University Journal of Law & Liberty", vol. 8, 2014.

**Bengio, Y., Courville, A., Vincent, P.**, *Representation Learning: A Review and New Perspectives*, IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 35, issue 8, 2013.

**Benjamins, V. R., Casanovas, P., Breuker, J., Gangemi, A.**, *Law and the Semantic Web. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005.

**Bentham, J.**, *Theories des peines et de recompenses*, Ed. De L'imprimerie de Vogel Et Schulze, 13, Oxford Street, Tomul 1, 1811.

**Bertrand, R., Blais, J.G.**, *Modèles de mesure: L'apport de la théorie des réponses aux items*, Ed. Presses de l'Université du Québec, Sainte-Foy, Canada, 2004.

**Biagioli, C.**, *Towards a Legal Rules Functional Micro-Ontology*, Journal of Proceedings of workshop LEGONT, vol. 97, 1997.

**Brennan, T., Dieterich, W., Ehret, B.**, *Evaluating The Predictive Validity of the Compas Risk And Needs Assessment System*, "Criminal Justice and Behavior", 2009, vol. 36, pp. 22- 23, diponibil online: [https://www.researchgate.net/publication/242249206\\_Evaluating\\_the\\_predictive\\_validity\\_of\\_the\\_COMPAS\\_Risk\\_and\\_Needs\\_Assessment\\_System](https://www.researchgate.net/publication/242249206_Evaluating_the_predictive_validity_of_the_COMPAS_Risk_and_Needs_Assessment_System), accesat la data 15.09.2020.

**Brighi, R., Lesmo, L., Mazzei, A., Palmirani, M., Radicioni, D. P.**, *Towards semantic interpretation of legal modifications through deep syntactic analysis*, în vol. Proceedings of the 21st International Conference on Legal Knowledge and Information Systems, "Frontiers in Artificial Intelligence and Applications", 2008.

**Canales, D., Campbell, M. A., Wei, R., Totten, A.**, *Prediction of General and Violent Recidivism Among Mentally Disordered Adult Offenders. Test of the Level of Service/Risk-Need-Responsivity (LS/RNR) Instrument*, "Criminal Justice and Behavior", vol. 41, nr. 8, 2014.

**Canton, R.**, *Why punish? An Introduction to the Philosophy of Punishment*, Macmillan Publishers Limited, London, 2017.



**Cicero**, *De Officiis*, disponibil online: <https://www.thelatinlibrary.com/cicero/off1.shtml>, accesat la data 18.09.2020.

**Ciobanu (Aionoae), A.**, *Utilizarea Rețelelor Bayesiene și a Diagramelor de Influență în Analiza Disponibilității în Electroenergetică, Rezumat Teză de Doctorat*, Universitatea Tehnică "Gheorghe Asachi" din Iași, 2019.

**Committee of Experts** on Internet Intermediaries (msi-net), *Study on the Human Rights Dimensions of Automated Data Processing Techniques (In Particular Algorithms) and Possible Regulatory Implications*, Council of Europe, 2017.

**Crețulescu, R. G.**, *Teză de doctorat (Rezumat). Contribuții la proiectarea sistemelor de clasificare a documentelor*, Universitatea „Lucian Blaga” din Sibiu Facultatea de Inginerie “Hermann Oberth”, Departamentul de Calculatoare și Inginerie Electrică, 2011.

**Cukier, K.**, *Data, Data Everywhere: A Special Report on Managing Information.* "The Economist", 2010, disponibil online: <https://www.economist.com/special-report/2010/02/27/data-data-everywhere>, accesat la data 10.08.2020.

**CSG Justice Center Staff**, *Risk and Needs Assessment and Race in the Criminal Justice System*, The Council of State Governments, 2016, disponibil online: <https://csgjusticecenter.org/reentry/posts/risk-andneeds-assessment-and-race-in-the-criminal-justice-system/>, accesat la data 15.09.2020.

**Daragh, M.**, *Live facial recognition: the impact on human rights and participatory democracy*, disponibil online: [https://www.essex.ac.uk/centres-and-institutes/public-engagement/facial-recognition#:~:text=Live%20facial%20recognition%20\(LFR\)%20interferes,person's%20professional%20and%20private%20life](https://www.essex.ac.uk/centres-and-institutes/public-engagement/facial-recognition#:~:text=Live%20facial%20recognition%20(LFR)%20interferes,person's%20professional%20and%20private%20life), accesat la data 23.09.2020.

**Dobrea, D. M.**, *Tehnici de inteligență computațională. Aplicații în electronică și biomedicină*, disponibilonline:[http://www.etc.tuiasi.ro/cin/Courses/Epiom/Course/Curs\\_Rtizrt123XuAsdr/Capitol%20RNA.pdf](http://www.etc.tuiasi.ro/cin/Courses/Epiom/Course/Curs_Rtizrt123XuAsdr/Capitol%20RNA.pdf), accesat la data 16.08.2020.

**Domingos, P., Pazzani, M.**, *On the optimality of the simple Bayesian classifier under zero-one loss*, Ed. Kluwer Academic Publishers, 1997.

**Dongoroz, V.**, *Drept penal*, Institutul de Arte Grafice, București, 1939.

**Drăgoescu, B. C.**, *Învățarea Bayesiană*, Lucrare de Licență, Universitatea Babeș-Bolyai Cluj-Napoca, Facultatea de Matematică și Informatică, 2015.

**Duarte, N.**, *CDT Statement on Government Use of Algorithmic Decision-Making Tools to NYC Council Committee on Technology*, 2017, disponibil online: <https://cdt.org/insights/statement-on-government-use-of-algorithmic-decision-making-tools-to-new-york-city-council-committee-on-technology/>, accesat la data 12.09.2020.

**Dupont, B., Stevens, Y., Westermann, H., Joyce, M.**, *Artificial Intelligence in the Context of Crime and Criminal Justice*, Canada Research Chair in Cybersecurity International Centre for Comparative Criminology, Université de Montréal, 2018.

**Dunea, M.**, *Analiza receptării în doctrina juridico-penală a principalelor teorii referitoare la fundamentul dreptului de a pedepsi, în relație cu esența noțiunii de impunitate*, Analele Științifice ale Universității „Al.I.Cuza” Iași Tomul LVII, Științe Juridice, 2011.

**Eraclide, C.**, *Studii practice asupra comportamentului criminal*, București, 1865.

**Emc Education Services**, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and*

*Presenting Data*, Ed. Wiley, 2015.

**Ferri, E.**, *Relazione sul Progetto Preliminare di Codice Penale*, Ed. Vallardi, Milano, 1921.

**Flores, A. W., Lowenkamp, C. T., Smith, P., Latessa, E. J.**, *Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers*, "Federal Probation", vol. 70, nr. 2, 2006.

**Francesconi, E., Passerini, A.**, Automatic Classification of Provisions in Legislative Texts, "Artificial Intelligence and Law", 2007.

**Francesconi, E., Montemagni, S., Peters, W., Tiscornia D.**, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010.

**Francesconi, E., Peruginelli, G.**, *Integrated Access to Legal Literature through Automated Semantic Classification*, "Artificial Intelligence and Law", 2009.

**Francesconi, E.**, *An Approach to Legal Rules Modelling and Automatic Learning*, The Twenty-Second Annual Conference on Legal Knowledge and Information Systems, Rotterdam, 2009.

**Francesconi, E., Montemagni, S., Peters, W.**, *Integrating a Bottom-Up and Top-Down Methodology for Building Semantic Resources for the Multilingual Legal Domain*, in E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010.

**Friedman, B., Nissenbaum, H.**, *Bias in Computer Systems*, "ACM Transactions on Information Systems", vol. 14, nr. 3, 1996, disponibil online: [https://www.vsdesign.org/publications/pdf/64\\_friedman.pdf](https://www.vsdesign.org/publications/pdf/64_friedman.pdf), accesat la 08.09.2020.

**Gabriel, I.**, *The case for fairer algorithms*, 2018, material disponibil online: [https://medium.com/@Ethics\\_Society/the-case-for-fairer-algorithms-c008a12126f8](https://medium.com/@Ethics_Society/the-case-for-fairer-algorithms-c008a12126f8), accesat la data 10.09.2020.

**Garçon, É. A.**, *Le droit pénal: origines--évolution--état actuel*, Ed. Payot & cie, Paris, 1922.

**Giguère, G., Lussier, P.**, *Debunking the psychometric properties of the LS\CMI: An application of item response theory with a risk assessment instrument*, "Journal of Criminal Justice", vol. 46, 2016.

**Guiraude, L.**, *Using NLP techniques to identify legal ontology components: Concepts and relations*, "Artificial Intelligence and Law", vol. 12, nr. 4, 2004.

**Gupta, B. M., Dhawan, S. M.**, *Deep Learning Research: Scientometric Assessment of Global*, "Emerging Science Journal" vol. 3, nr. 1, 2019.

**Gulledge, A. T., Kampa, B. M., Stuart G. J.**, *Synaptic integration in dendritic trees*, "Journal of Neurobiology", vol. 64, issue 1, 2005.

**Harari, Y. N.**, *21 de lectii pentru secolul XXI*, Ed. Polirom, 2018.

**Harcourt, B. E.**, *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, Ed. The University of Chicago Press, Chicago, 2007.

**Hammond, K.**, *5 unexpected sources of bias in artificial intelligence*, 2016, material disponibil online: [https://flipboard.com/@TechCrunch/5-unexpected-sources-of-bias-in-artificial-intelligence/\\_bHWcY85NT2qn\\_40uag4idQ%3Aa%3A43591897-64d34bacc6%2Ftechcrunch.com](https://flipboard.com/@TechCrunch/5-unexpected-sources-of-bias-in-artificial-intelligence/_bHWcY85NT2qn_40uag4idQ%3Aa%3A43591897-64d34bacc6%2Ftechcrunch.com), accesat la data 10.09.2020.

**Hastie, T., Tibshirani, R., Friedman, J.**, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Ed. a 2-a, Ed. Springer, 2012.

**Hawking, S.**, *Brief Answers To The Big Questions*, Ed. Bantam Books, New York, 2018.

**Houzel, H. S.**, *The not extraordinary human brain*, "Proceedings of the National Academy of Sciences",

vol. 109, Supplement 1, 2012.

**Hillis, W. D.**, "Physics Today", vol. 42, nr. 2, 1989, disponibil online: <http://longnow.org/essays/richard-feynman-connection-machine/>, accesat la data 07.09.2020.

**Hurwitz, J., Kirsch, D.**, *Machine Learning For Dummies*, IBM Limited Edition, Ed. John Wiley & Sons, 2018.

Interviu acordat de E. Musk, disponibil online: <https://www.nytimes.com/2020/07/25/style/elon-musk-maureen-dowd.html>, accesat la data 24.09.2020.

Interviu acordat de Dr. B. Goertzel, disponibil online: <https://www.forbes.com/sites/cognitiveworld/2020/07/14/is-artificial-general-intelligence-agi-on-the-horizon-interview-with-dr-ben-goertzel-ceo--founder-singularitynet-foundation/#5d8c1a4959d0>, accesat la data 23.09.2020.

**James, J.**, *Bayes Theorem*, "The Stanford Encyclopedia of Philosophy", 2019, disponibil online: <https://plato.stanford.edu/archives/spr2019/entries/bayes-theorem/>, accesat la data 12.09.2020.

**James, N.**, *Risk and Needs Assessment in the Criminal Justice System*, Congressional Research Services, 2018.

**Johnson, D. G., Mulvey, J. M.**, *Computer decisions: Ethical issues of responsibility and bias. Statistics and Operations*, Res. Series SOR-93-11, Dept. of Civil Engineering and Operations Research, Princeton Univ., Princeton, 1993.

**Jurafsky D., Martin, H. J.**, *Neural Networks and Neural Language Models*, 2019, disponibil online: <https://web.stanford.edu/~jurafsky/slp3/7.pdf>, accesat la data 20.08.2020.

**Kaczynski, T.**, *Industrial Society and Its Future*, "The Washington Post", 1995.

**Kant, I.**, *Metaphysische Anfangsgründe Der Rechtslehre*, Ed. Friedrich Nicolovius, Königsberg, 1797.

**Kaye, D.**, *Report on the promotion and protection of the right to freedom of opinion and expression*, pentru a 32-a sesiune a Human Rights Council, (A/HRC/32/ 38), 2016.

**Kehl, D., Guo, P., Kessler, S.**, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, "Harvard Law School", 2017.

**Keng, L.**, *A comparison of the performance of testlet-based computer adaptive tests and multistage tests*, Dissertation Presented to the Faculty of the Graduate School of The University of Texas at Austin in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy, 2008.

**Lefschetz, S.**, *Algebraic Topology*, "American Mathematical Society", vol. 27, 1942.

**Leone, V., Caro, L. D., Villata, S.**, *Legal Ontologies and How to Choose Them: the Investigation Tool*, "International Semantic Web Conference", 2018.

**Lewin, T.**, *Making Punishment Fit Future Crimes*, N.Y. Times, 1982, disponibil online: <https://www.nytimes.com/1982/11/14/weekinreview/making-punishment-fit-future-crimes.html>, accesat la data 12.09.2020.

**Mathiesen T.**, *Selective Incapacitation Revisited*, vol. 22, nr. 4, "Law and Human Behavior", 1998.

**McCarty, T.**, *Deep semantic interpretations of legal texts*, în vol. Proceedings of the 11th International Conference on Artificial Intelligence and Law, Association for Computing Machinery, New York, United States, 2007.

**Megan, G.**, *Racist in the Machine: The Disturbing Implications of Algorithmic Bias*. Retrieved, "World

Policy Journal”, vol. 33, nr. 4, 2017, disponibil online: <http://muse.jhu.edu/article/645268/pdf>, accesat la data 08.09.2020.

**Mitchell, T.**, *Machine Learning*, Ed. McGraw-Hill, 1997.

**Mochales, R., Moens, M. F.**, *Argumentation mining*, ”Artificial Intelligence and Law”, 2011.

**Moens, M. F., Boiy, E., Mochales, R., Reed, C.**, *Automatic detection of arguments in legal texts*, în vol. Proceedings of the 11th International Conference on Artificial Intelligence and Law, 2007.

**Moor, J.**, *What is computer ethics?*, ”Metaphilosophy”, vol. 16, nr. 4, 1985, disponibil online: [https://www.researchgate.net/publication/227591551\\_What\\_Is\\_Computer\\_Ethics](https://www.researchgate.net/publication/227591551_What_Is_Computer_Ethics), accesat la data 08.09.2020.

**Monahan, J.**, *Predicting Violent Behavior: An Assessment of Clinical Technique*, ”Crime & Delinquency”, vol. 28, nr. 4, 1981.

**Murphy, K. P.**, *Machine Learning A Probabilistic Perspective*, The MIT Press Cambridge, Massachusetts London, England, 2012.

**Mureșan, L. E.**, *Învățarea Bayesiană*. Lucrare de Diplomă, Universitatea Babeș- Bolyai Cluj-Napoca, Facultatea De Matematică și Informatică, Cluj Napoca, 2011.

**Nakamura, M., Nobuoka, S., Shimazu, A.**, *Towards translation of legal sentences into logical forms*, Annual Conference of the Japanese Society for Artificial Intelligence, 2007.

**Neumann, J. V.**, *Theory of Self-Replicating automata*, editată și completată de A. W. Burks, Ed. University of Illinois Press, 1966, disponibil online: <http://cba.mit.edu/events/03.11.ASE/docs/VonNeumann.pdf>, accesat la data 07.08.2020.

**O’Neil, C.**, *Don’t Grade Teachers with a Bad Algorithm*, Bloomberg, 2017, disponibil online: <https://www.bloomberg.com/view/articles/2017-05-15/don-t-grade-teachers-with-a-bad-algorithm>, accesat la data 12.09.2020.

**Oppy, G., Dowe, D.**, , *The Turing Test*, ”The Stanford Encyclopedia of Philosophy”, 2019, disponibil online: URL =<<https://plato.stanford.edu/archives/spr2019/entries/turing-test/>>, accesat la data 05.08.2020.

**Pala, K., Rychlý, P., Šmerk, P.**, *Automatic identification of legal terms in Czech legal texts*. Semantic Processing Legal Texts, în: E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010.

**Panu, R.**, *Gödel’s Incompleteness Theorems*, ”The Stanford Encyclopedia of Philosophy”, 2020, disponibil online: <https://plato.stanford.edu/archives/fall2020/entries/goedel-incompleteness/>, accesat la data 07.09.2020.

**Pew Center on the States**, *Risk/Needs Assessment 101: Science Reveals New Tools to Manage Offenders*, Public Safety Performance Project, issue brief, 2011, disponibil online: [https://www.pewtrusts.org/~media/legacy/uploadedfiles/pcs\\_assets/2011/pewriskassessmentbriefpdf.pdf](https://www.pewtrusts.org/~media/legacy/uploadedfiles/pcs_assets/2011/pewriskassessmentbriefpdf.pdf), accesat la data 15.09.2020.

**Petroni, V.**, *Comentariile dreptului penal*, București, 1857.

**Pop, Tr.**, *Curs de Criminologie*, Institutul de Arte Grafice „Ardealul”, Cluj, 1928.

**Practitioner’s Guide to COMPAS Core**, Northpointe, 2005, diponibil online: <https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>, accesat la data 15.09.2020.

**Principiile Asilomar**, disponibil online: <https://futureoflife.org/ai-principles/?cn-reloaded=1>, accesat la data 24.09.2020.

**Rish, I.**, *An empirical study of the naive Bayes classifier*, Workshop on Empirical Methods in Artificial Intelligence, 2001.

**Rotaru, C.**, *Fundamentul pedepsei. Teorii moderne*, Editura C.H. Beck, București, 2006.

**Robbins, B.**, Machine Learning: How Black is This Beautiful Black Box, material disponibil online: <https://towardsdatascience.com/machine-learning-how-black-is-this-black-box-f11e4031fdf>, accesat la data 15.08.2020.

**Russell, S. J., Norvig, P.**, *Artificial Intelligence. A Modern Approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

**Rusu, O. R.**, *Etica Pedepsei Privative de Libertate. Teză De Doctorat*, Universitatea „Alexandru Ioan Cuza” Din Iași Facultatea de Filosofie și Științe Social – Politice, Iași, 2018.

**Saias, J., Quaresma, P.**, *A Methodology to Create Legal Ontologies in a Logic Programming Based Web Information Retrieval System*, în V. R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi, *Law and the Semantic Web. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005.

**Satoh, K., Inokuchi, A., Nagao, K., Kawamura, T.**, *New Frontiers in Artificial Intelligence. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2008.

**Saxena, R.**, *The social media “echo chamber” is real*, 2017, disponibil online: <https://arstechnica.com/science/2017/03/the-social-media-echo-chamber-is-real/>, accesat la data 23.09.2020.

**Schmidhuber, J.**, *Deep Learning in Neural Networks: An Overview*, “Neural Networks Journal”, 2014.

**Segev, I., London, M.**, *Untangling dendrites with quantitative models*. “Science”, vol. 290, issue 5492, 2000.

**Seneca**, *De Ira*, disponibil online: <https://www.thelatinlibrary.com/sen/sen.ira1.shtml>, accesat la data 18.09.2020.

**Shai, S. S., Shai, B. D.**, *Understanding Machine Learning. From Theory to Algorithms*, Ed. Cambridge University Press, New York, USA, 2014.

**Shadowen, A.**, *Ethics and Bias in Machine Learning: A Technical Study of What Makes Us ‘Good’*, City University of New York, 2017, disponibil online: [https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1042&context=jj\\_etds](https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1042&context=jj_etds), accesat la data 08.09.2020.

**Simourd, D. J.**, *Use of Dynamic Risk/Need Assessment Instruments Among Long-Term Incarcerated Offenders*, “Criminal Justice and Behavior”, vol. 31, nr. 3, 2004.

**Singh, J. P., Kroner, D. G., Wormith, J. S., Desmarais, S. L., Hamilton, Z.**, *Handbook of Recidivism Risk/Needs Assessment Tools*, Ed. John Wiley & Sons Ltd, 2018.

**Skilton, M., Hovsepian F.**, *The 4th Industrial Revolution. Responding to the Impact of Artificial Intelligence on Business*, Palgrave Macmillan, 2018.

**Spinosa, P., Giardiello, G., Cherubini, M., Marchi, S., Ventur, G., Montemagni, S.**, *NLP-based Metadata Extraction for Legal Text Consolidation*, în vol. Proceedings of the 12th International Conference on Artificial Intelligence and Law, 2009.

**Starr, S. J.**, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, “Stanford Law Review”, vol. 66, nr. 4, 2014.

**Stănilă, L.**, *Inteligența Artificială. Dreptul Penal și Sistemul de Justiție Penală. Amintiri Despre Viitor*, Ed. Universul Juridic, București, 2020.

**Student-edited journal**, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, vol. 96, nr. 2, "Harvard Law Review", 1982.

**Stefanescu, M.**, *Introducere în rețele neuronale – Teorie și aplicații*, material disponibil online: <https://www.code-it.ro/introducere-in-retele-neuronale-teorie-si-aplicatii/>, accesat la data 18.08.2020.

**Tanoviceanu I.**, *Tratat de drept și procedură penală*, Ediția a II-a revăzută și completată, vol. I, București, 1924-1927.

**Tanoviceanu I.**, *Tratat de drept și procedură penală*, Ediția a II-a revăzută și completată, vol. III, București, 1924-1927.

**Tegmark, M.**, *Viața 3.0. Omul în epoca inteligenței artificiale*, Ed. Humanitas, 2019.

**Turing, A.**, *Computing Machinery and Intelligence*, "Mind", Vol. LIX, Issue 236, 1950.

**Urbanska, M., Blazejczyk, M., Jaworski, J.**, *Molecular basis of dendritic arborization*, "Acta neurobiologiae experimentalis", vol. 68, nr. 2, 2008.

**U. N. Human Rights Council Resolution** on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/7, 23 Mar. 2017.

**Venturi, G.**, *Legal language and legal knowledge management applications*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010.

**Villasenor, J.**, *Artificial intelligence and bias: Four key challenges*, 2019, material disponibil online: <https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/>, accesat la data 10.09.2020.

**Walter, S.**, *Linguistic description and automatic extraction of definitions from German court decisions*, în vol. Proceedings of the 6th International Language Resources and Evaluation, "European Language Resources Association", 2008.

**Walter, S., Pinkal, M.**, *Automatic extraction of definitions from German court decisions*, în vol. Proceedings of the Workshop on Information Extraction Beyond The Document, "Association for Computational Linguistics", 2008.

**Wyner, A., Palau, R. M., Moens, M. F., Milward, D.**, *Approaches to Text Mining Arguments from Legal Cases*, în E. Francesconi, S. Montemagni, W. Peters, D. Tiscornia, *Semantic Processing of Legal Texts. Lecture Notes in Computer Science*, Ed. Springer, Berlin, Heidelberg, 2010.

**Zhang H.**, *The Optimality of Naive Bayes*, Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, vol. 2, 2004.

**Zhang, J., El-Gohary, M. N.**, *Automated Information Transformation for Automated Regulatory Compliance Checking in Construction*, "Journal of Computing in Civil Engineering", vol. 29, nr. 4, 2015.

**Zimmer, A.**, *High Schools Dole Out Misinformation About Admissions Process, Parents Say*, "DNAInfo", 2016, disponibil online: <https://www.dnainfo.com/new-york/20161115/kensington/nyc-high-school-admissions-ranking>, accesat la data 12.09.2020