

# INTELIGENȚA ARTIFICIALĂ – ÎNTRE TEHNOLOGIE ȘI ETICĂ

Silviu IONIȚĂ<sup>1</sup>  
silviu.ionita@upit.ro

**ABSTRACT:** The present study addressed the issue of artificial intelligence (AI) in connection with both technological and social dimensions. AI is intrinsically an interdisciplinary field covering engineering, natural sciences, mathematics and computer science, psychology, social science, science of communication, etc. Scientific research evolves rapidly and certain progress that foreshadows the emergence of technical systems trends to change people's life style and even provide new models for human society. The impact of advanced research in traditional sub-branches of AI, such as robotics and software engineering research is currently multiplied by deepest insights regarding artificial mind and building the artificial brain. These things sound challenging but also raise new problems and even some fears. AI is no longer just a game of imagination or a software paradigm simulators; it is found already incorporated in physical things created with noble purpose to make our lives better. The super tech clear trend will lead human society in a position to cooperate with increasingly intelligent systems, as intelligent as humans, perhaps superintelligent. Assuming that God will allow us to do that, still remains to answer a lot of questions on one major topic: How human society will live together with others artificial?

**KEYWORDS:** Artificial Intelligence, robotics, artificial brain, autonomous systems, ethics.

## *Metafore și paradigme în inteligența artificială (IA)*

În 1956 John McCarthy<sup>2</sup> a introdus termenul de inteligență artificială oferind și o definiție concisă a acesteia: „*știința și ingineria producerii de*

<sup>1</sup> Prof.univ.dr.inginer la Departamentul de Electronică, Calculatoare și Inginerie Electrică, Universitatea din Pitești; membru titular CRIFST (Filiala CRIFST – Argeș).

<sup>2</sup> 1927–2011, profesor și cercetător la Stanford University, unul dintre fondatorii inteligenței artificiale

*mașini inteligente*”. Definiția lui McCarthy se încadra perfect în spiritul epocii de pionierat a calculatoarelor numerice și a apariției limbajului de programare LISP (LIST Processing) recunoscut ca primul limbaj de programare specific inteligenței artificiale. Astfel, primele concepte despre inteligența artificială au fost dominate de viziunea calculabilității și a limbajelor de programe software. Până la sfârșitul anilor '60 se poate spune că inteligența artificială a parcurs o perioadă a cristalizării conceptuale – epoca definițiilor, a teoremelor și a coniecturilor. Încă în 1931, Kurt Godel<sup>3</sup> afirma prin așa numitele teoreme ale incompletitudinii că nici un sistem formal procedural bazat pe un set consistent de axiome nu poate demonstra toate adevărurile și, mai mult, un astfel de sistem formal nu-și poate demonstra propria consistență. Prin urmare, teoremele lui Godel prefigurau faptul că nici un program de calculator n-ar putea demonstra toate afirmațiile adevărate. Ulterior, în 1950 Alan Turing<sup>4</sup> a propus o convenție referitoare la inteligența artificială consacrată sub forma testului ce-i poartă numele: „dacă o mașină se comporta la fel de inteligent ca o ființă umană, atunci e la fel de inteligentă ca o ființă umană”. Astfel, se acreditează ideea că inteligența unei mașini ar putea fi evaluată, bazându-ne doar pe comportamentul ei. Majoritatea opiniilor, argumentelor și coniecturilor enunțate în epoca respectivă au supraevaluat perspectiva computațională ca viziune dominantă asupra inteligenței artificiale. Chiar dacă acestea reprezintă doar o componentă limitată a ceea ce astăzi înseamnă un sistem inteligent, merită să le menționăm ca puncte de referință pentru evoluția domeniului.

*Concluzia Conferinței din Dartmouth*<sup>5</sup> (1956): „fiecare componentă a procesului de învățare sau oricare alt aspect al inteligenței, poate fi descris atât de precis, încât putem învăța o mașină cum să îl simuleze”. Întrunirea de la Dartmouth College a consacrat practic inteligența artificială ca domeniu de studiu și cercetare.

<sup>3</sup> 1906–1978, logician, matematician și filozof american de origine austriacă

<sup>4</sup> 1912–1954, matematician, logician și criptanalist britanic, pionier, personalitate prominentă în știința calculatoarelor

<sup>5</sup> Dartmouth Summer Research Project on Artificial Intelligence, evenimentul care a consacrat nașterea inteligenței artificiale ca domeniu

*Conjectura simbolurilor fizice* susținută de Allen Newell<sup>6</sup> și Herbert Simon<sup>7</sup>: „un sistem fizic simbolic oferă mijloacele suficiente și necesare pentru a se acționa în mod inteligent”. Aceasta teorie susține ca esența inteligenței constă în manipularea simbolurilor.

*Conjectura inteligenței artificiale „puternică”*: „calculatorul programat adecvat și dotat cu intrări și ieșiri potrivite, ar avea o minte proprie, în aceeași manieră în care oamenii au și ei una”.

*Argumentul creierului artificial* potrivit căruia *creierul poate fi simulat* este susținut de Hans Moravec<sup>8</sup>, Ray Kurzweil<sup>9</sup> și alții. Încurajați de evoluția fabuloasă a electronicii și ingineriei software, aceștia susțin ca e posibilă copierea unui creier corect în software și hardware și, mai mult, o asemenea simulare ar fi identică cu originalul. Acest argument combină ideea că o mașină indeajuns de puternică poate simula orice proces, cu ideea materialistă, conform căreia mintea e rezultatul proceselor fizice din creier.

Pentru a înțelege mai bine modul în care inteligența artificială ridică probleme de natură etică este important să facem o scurtă trecere în revistă a parcursului său conceptual și tehnologic.

*Reperele istorice* în dezvoltarea inteligenței artificiale pot fi găsite încă din antichitate, acestea având o bază mitologică. Talos din Creta este considerat primul robot antic, acesta fiind o mașină complet operațională construită de Hefaistos –zeul metalurgiei și al modelării prin sculptură, unul dintre cei 12 zei ai Olimpului [4]. Mitul lui Pygmalion al înzestrării unei creații materiale sub formă umană –Galatea cu suflet și minte este un alt exemplu în acest sens.

În epoca modernă, tema ființelor artificiale apare mai degrabă în literatura de anticipație, considerată la începutul secolului XX o literatură

<sup>6</sup> 1927–1992, personalitate remarcabilă în domeniul inteligenței artificiale și al psihologiei cognitive umane

<sup>7</sup> 1916–2001, personalitate multilaterală, laureat al premiului Nobel, considerat părintele fondator al mai multor domenii printre care: inteligența artificială, sisteme complexe, procesarea informației și luarea deciziilor

<sup>8</sup> 1948 – robotician, cercetător în domeniul vederii artificiale, scriitor în domeniul impactului tehnologic

<sup>9</sup> 1948 – inventator în domeniul scanării și recunoașterii optice a caracterelor, scriitor futurist

periferică. Frații Karel<sup>10</sup> și Josef Capeck<sup>11</sup> sunt primii care au introdus termenul de robot ca structură antropomorfică în literatură. În opera sa de ficțiune Isac Asimov<sup>12</sup> ridică robotul la rangul de personaj capabil să interacționeze cu oamenii. Roboții lui Asimov sunt implicați în situații critice adesea cu miză crucială pentru existența umanității în care, deznădămintul se derulează în virtutea unor legi ce guvernează comportamentul roboților și mașinilor inteligente. În planul opus ficțiunii – în domeniul tehnologic, evoluția domeniului inteligenței artificiale se poate sintetiza cronologic astfel:

Prima jumătate a sec XX este *epoca fundamentelor* inteligenței artificiale în care s-au pus bazele utilizării logicii formale în raționamente și a metodelor formale de manipulare a simbolurilor. Rezultatele mai multor matematicieni și logicieni cum sunt Kurt Godel, Ian Lucasiewicz<sup>13</sup>, Emil Leon Post<sup>14</sup>, Andrei Kolmogorov<sup>15</sup>, Alonzo Church<sup>16</sup>, Alan Turing, au creat germele apariției inteligenței artificiale ca domeniu de cercetare.

Anii '50 și '60 reprezintă *epoca de început* a inteligenței artificiale, considerată perioada clasică de dezvoltare a domeniului. Este epoca în care calculatoarele electronice și limbajele de programare evoluează rapid, ceea ce permite testarea și validarea modelelor teoretice. Noile științe care apar și se impun în acea perioadă – cibernetică (Norbert Wiener<sup>17</sup>) și teoria informației (Claude Shannon<sup>18</sup>) contribuie deasemenea la dezvoltarea inteligenței artificiale. Calculul neuronal bazat pe modelul propus

<sup>10</sup> 1890–1938, scriitor și dramaturg ceh, autor de proză fantastică și de ficțiune – romanul *War with the Newts* și piesa *R.U.R.* (Rossum's Universal Robots).

<sup>11</sup> 1902–1982, pictor, scriitor și poet ceh, inventatorul de facto al termenului *robot* preluat de Karel Capek în scrierile sale de anticipație.

<sup>12</sup> 1920–1992, autor genial și extrem de prolific al genului science-fiction.

<sup>13</sup> 1878–1954, profesor de logică și filozofie polonez, creator în domeniul logiciilor multivalente.

<sup>14</sup> 1897–1954, matematician și logician american de origine rusa cu contribuții în teoria calculului.

<sup>15</sup> 1903–1987, matematician rus cu contribuții în complexitatea calculului, teoria informației, logica intuiției.

<sup>16</sup> 1903–1995, matematician și logician american cu contribuții în logica matematică și teoria calculatoarelor.

<sup>17</sup> 1894–1964, matematician și logician american recunoscut ca inițiator al domeniului ciberneticii.

<sup>18</sup> 1916–2001, matematician și inginer electronist recunoscut drept părintele teoriei informației.

de Walter Pitts<sup>19</sup> și Warren McCulloch<sup>20</sup> la sfârșitul anilor '40 este conceptualizat sub forma modelelor conexioniste și este implementat în programe de calculator. Impulsionat de aceste cercetări, în 1951 Marvin Minsky<sup>21</sup> proiectează și realizează SNARC (Stochastic Neural Analog Reinforcement Calculator), o rețea neuronală implementată hardware cu tuburi electronice, considerată prima mașină autoinstruibilă. Începe perioada de avânt a rețelelor neuronale marcată de crearea perceptronului de către Frank Rosenblatt<sup>22</sup>, ca model simulativ în 1957 și mai târziu, în 1960, ca sistem hardware cunoscut sub numele de Mark I Perceptron – primul computer care putea să deprindă prin învățare noi abilități folosind tehnica încercare-eroare. Tot în această perioadă apar realizări remarcabile într-o direcție care avea să consacre mai târziu rolul definitoriu al sistemelor expert în dezvoltarea inteligenței artificiale. Astfel, în 1956 Allen Newell, Herbert Simon și John Clifford Shaw<sup>23</sup> prezintă „The Logic Theorist” – primul program de demonstrare a teoremelor, iar în 1959 ei crează „General Problem Solver” – un program care încerca să rezolve orice problemă descrisă prin clauze sub formă de reguli, folosind logica formală. În această perioadă se remarcă programul de jucat șah creat de Arthur Lee Samuel<sup>24</sup> în 1960 – acesta fiind primul program care aplica strategii de evaluare bazate pe calcul simbolic și tehnici de procesare adaptive, care îi permiteau să-și amelioreze performanțele după fiecare partidă jucată.

*O perioadă distinctă* în dezvoltarea inteligenței artificiale se deschide în anul 1965, odată cu începerea construirii la Stanford University a primului sistem expert numit DENDRAL. În această perioadă se lucrează intens pentru dezvoltarea de noi limbaje de programare orientate pe obiect potrivite pentru rezolvarea non-numerică a problemelor de logică. În anul 1965 Lotfi Zadeh<sup>25</sup> publică lucrarea *Fuzzy Sets*, care avea să deschidă

<sup>19</sup> 1923–1969, logician american cu contribuții în calculul neuronal.

<sup>20</sup> 1898–1969, neurofiziolog și cibernetician american cu contribuții în modelarea rețelelor neuronale.

<sup>21</sup> 1927–, personalitate în domeniul științelor cognitive, co-fondator al domeniului inteligenței artificiale.

<sup>22</sup> 1927–1971, psiholog american cu contribuții notabile în inteligența artificială, creatorul perceptronului.

<sup>23</sup> 1922–1991, pionier în programarea calculatoarelor, contributor la dezvoltarea primelor programe aplicative pentru inteligența artificială (alături de A. Newell și H. Simon).

<sup>24</sup> 1901–1990, pionier în domeniul jocurilor pe calculator, creatorul primului program cu inteligență artificială pentru jocul de șah.

<sup>25</sup> 1921–, matematician, inginer electronist, cercetător în inteligența artificială, profesor

drumul către modelarea raționamentului aproximativ cu ajutorul logicii multivalente continue – logica de tip fuzzy. În acest domeniu se remarcă și contribuțiile unor personalități române precum Grigore Moisil<sup>26</sup>, care încă din 1941 a publicat seria de lucrări „Contributions sur l’etude des logiques non-chryspiennes”, ulterior dezvoltând algebrele cunoscute astăzi sub numele de Łukasiewicz–Moisil. Prin cercetări aplicative în domeniul sistemelor fuzzy se disting personalități precum Constantin Virgil Negoită<sup>27</sup> și Horia-Nicolai Teodorescu<sup>28</sup>. Un moment de referință este marcat în 1972 odată cu apariția limbajului PROLOG special dezvoltat pentru programarea logică, utilizând elemente de lingvistică computațională. Un moment de cotitură în istoria inteligenței artificiale este considerat apariția în 1969 a lucrării *Perceptrons* scrisă de Marvin Minsky și Seymour Pappert<sup>29</sup>. În această lucrare, autorii deși recunosc rolul perceptronului ca model esențial de rețea neuronală artificială demonstrează unele limite ale acestuia inducând o perspectivă pesimistă a modelelor conexioniste în general pentru viitorul inteligenței artificiale. Impactul acestei lucrări în comunitatea științifică a avut un efect mai degrabă negativ, declanșând o controversă la acea vreme, în urma căreia conceptul de modelare a cunoștințelor nestructurate într-un cadru numeric, specific rețelelor neuronale a fost oarecum ignorat în favoarea modelelor bazate pe prelucrarea simbolică a cunoștințelor structurate. Până la mijlocul anilor ’80, când acest punct de vedere a fost reconsiderat prin cercetările întreprinse de John Hopfield<sup>30</sup>, David Rumelhart<sup>31</sup> și alții, inteligența artificială a cunoscut așa numita „*iarnă a IA*” o perioadă marcată de reducerea fondurilor de cercetare și în general scăderea interesului comunității științifice pentru acest domeniu. Epoca respectivă se remarcă totuși prin crearea unui număr apreciabil de sisteme expert care au devenit aplicații de referință pentru modelele de prelucrare simbolică a cunoștințelor. Astfel, dacă în anii ’70 se susținea cu tărie că inteligența artificială are limite serioase pentru a fi integrată în

---

emerit la University of California, Berkeley.

<sup>26</sup> 1096–1973, matematician român, considerat părintele informaticii românești.

<sup>27</sup> 1936 –, profesor emerit la City University of New York.

<sup>28</sup> 1951 –, profesor universitar, membru corespondent al Academiei Române.

<sup>29</sup> 1928 –, matematician, informatician și pedagog american de origine sud-africană, unul din pionierii inteligenței artificiale.

<sup>30</sup> 1933 –, personalitate științifică recunoscută pentru inventarea rețelelor neuronale asociative.

<sup>31</sup> 1942–2011, psiholog american cu contribuții în analiza formală a gândirii umane.

aplicații productive, începutul anilor '80 marchează succesul comercial al sistemelor expert.

Se poate spune că anul 1981 a anunțat zorii unei perioade foarte promițătoare pentru evoluția inteligenței artificiale – *perioada contemporană*. Este anul când Ministerul Japonez al Comerțului Internațional și al Industriei demarează un program de cercetare ambițios, cunoscut sub numele de „Calculatorul de generația a 5-a”. Acesta avea drept obiectiv crearea unei structuri hardware și software capabilă să converseze în limbaj natural cu omul, să interpreteze imagini și să raționeze ca o ființă umană. După euforia începutului, rezultatele acestui program devenit de cursă lungă au fost evaluate la fiecare decadă. După 30 de ani, în 2011 ținta majoră a programului era încă neatinsă. Similar, în 1983 SUA declanșază prin DARPA<sup>32</sup> programul de cercetare „Strategic Computing Initiative”, care avea ca obiective finanțarea cercetărilor emergente în domeniul calculatoarelor și al inteligenței artificiale menite să dezvolte inteligența mașinilor. Acest program a avut un impact notabil în dezvoltarea microelectronicii și a supercalculatoarelor, dar în ansamblu obiectivele majore au rămas neîndeplinite până în prezent. Începând cu 1990 inteligența artificială este puternic susținută de dezvoltarea sectorului TIC și se află constant la baza tehnologiilor avansate din diferite industrii oferind beneficii, performanțe spectaculoase și nu în ultimul rând profit financiar. În ciuda unor sincope temporare datorate mai degrabă piețelor de capital, dezvoltarea domeniului inteligenței artificiale a devenit emergentă cu diverse direcții de cercetare, cea ce reflectă practic complexitatea și caracterul său multidisciplinar: sisteme bazate pe cunoștințe, sisteme cognitive, mașini de învățat, inteligența computațională, agenți inteligenți, iar la acestea putem adăuga bioinformatica, web-ul semantic și altele.

*Contextul social și tehnologic – cerințele, realitățile și tendințele actuale pot fi descrise în termeni generici astfel:*

- Societatea informațională este în plină evoluție către societatea bazată pe cunoaștere.
- Cercetarea științifică aplicativă a generat o masă critică de proiecte majore.

<sup>32</sup> Defense Advanced Research Projects Agency, organism responsabil cu dezvoltarea de tehnologii emergente pentru domeniul militar.

- Competitivitatea economică generează o piață concurențială în domenii sensibile: resursele umane, resursele materiale și energetice.
- Dezvoltarea durabilă și sustenabilă a societății impune noi standarde de conviețuire și supraviețuire.

Aceste trăsături determină în prezent evoluția accelerată în domeniul ingineriei sistemelor cu inteligență artificială, pe de o parte și emergența IA cu neuroștiințele și psihologia, pe de altă parte. În prezent se derulează ample cercetări interdisciplinare în cadrul unor proiecte ambițioase cum sunt *creierul artificial*, *autovehiculul autonom*, *robotica humanoidă*, de la care se așteaptă un impact major asupra calității vieții individuale și o mai bună funcționare a societății umane în general. Asistăm astăzi la reconsiderarea inteligenței artificiale din perspectiva filozofică și revigorarea conexiunilor ei cu psihologia, teoria limbajelor naturale și lingvistica computațională. În acest domeniu este locul să menționăm meritul cercetătorilor români Solomon Marcus<sup>33</sup> și Dan Ioan Tufiș<sup>34</sup> pentru contribuții în legătură cu formalizarea și dezvoltarea instrumentelor computaționale de prelucrare a limbajelor naturale. Filozofia minții este un curent de reflecție care s-a remarcat în toate timpurile, începând cu gânditorii antici precum Aristotel<sup>35</sup>, Chrysippus<sup>36</sup>, și continuând cu David Hume<sup>37</sup> considerat inițiatorul științelor cognitive [2], Descartes<sup>38</sup> fondatorul raționalismului, până la reprezentanții contemporani John Searle<sup>39</sup> și Hubert Dreyfus<sup>40</sup>. În lucrarea sa de referință „What Computers can't do. The limits of Artificial Intelligence”, Dreyfus afirmă «*Expertiza umană depinde mai mult de instinctul inconștient, decât de manipularea conștientă a simbolurilor și de faptul că putem „simți” o situație, decât de a avea o înțelegere explicită simbolică*», iar John Searle într-un interviu se pronunță pe cât de simplu, pe atât de clar: «*Doar cu obiecte caracterizate sintactic, cum ar fi cifrele zero și unu și cu un set de reguli formale pentru manipularea lor (programul) nu este suficient pentru*

<sup>33</sup> 1925 –, matematician român, membru titular al Academiei Române.

<sup>34</sup> Cercetător științific, profesor universitar, membru titular al Academiei Române.

<sup>35</sup> 384–322 BC, filozof grec cu idei notabile în logică, teoria silogismului, teoria sufletului.

<sup>36</sup> 279 – c. 206 BC, filozof stoic grec cu reflecții excelente în logică, teoria cunoașterii și etică.

<sup>37</sup> 1711–1776, filozof scoțian remarcabil pentru lucrarea sa de referință *A Treatise of Human Nature* (1739).

<sup>38</sup> 1596–1650, filozof și matematician francez.

<sup>39</sup> 1932–, filozof american cu contribuții în filozofia minții, filozofia limbajului, critic al conjecturii inteligenței artificiale „puternice”.

<sup>40</sup> 1929–, filozof american, cu preocupări în implicațiile filozofiei în inteligența artificială, critic al IA computaționale.



*gândirea în sine, deoarece gândirea implică mai mult decât manipularea de simboluri, ea implică conținut semantic».*

Aceste linii de gândire împreună cu teoriile cognitive generate de progresul constant al cercetărilor în psihologie și neurobiologie trasează cadrul actual de dezvoltare a inteligenței artificiale, focalizat pe modelarea minții umane și construcția creierului artificial. Spectrul actual al cercetărilor legate de inteligența artificială a depășit cu îndrăzneală chiar și limitele pe care le-au enunțat vizionarii corifei în epoca clasică a domeniului, construcția creierului uman fiind de departe cel mai ambițios program de cercetare. Un număr recent al magazinului IEEE Pulse publicație a IEEE Engineering in Medicine and Biology Society, apărut sub genericul „The Brain Connection. Unlocking Secrets to Learning, Disease, and More” [1] este consacrat prezentării stadiului celor mai importante proiecte din domeniul ingineriei creierului: Human Brain Project- cunoscut ca Blue Brain Project, proiectul SpiNNaker și proiectul BrainScaleS. Deși scopul general al acestor proiecte este același, abordările sunt diferite: Blue Brain va fi un creier virtual incorporat într-un supercomputer, în timp ce SpiNNaker se bazează pe o structură neuronală, implementată pe un sistem multiprocesor cu circuite masiv paralele, inspirat din modul de lucru al creierului uman. BrainScaleS se bazează pe o structură neuromorfică hibridă de calcul ierarhizat, similară creierului, care s-a dovedit a fi de 10.000 de ori mai rapidă decât structura biologică originală. Revoluția neurotehnologică este deasemenea o direcție forte în înțelegerea mecanismelor profunde ale creierului uman, progresând în sensul creării dispozitivelor de interfață dintre creier și calculator, respectiv cu alte sisteme externe, cum sunt protezele, diferiți senzori, etc.

Revenind la tehnologie, în societatea noastră construcția de mașini și sisteme inteligente avansează temeinic, valorificând din plin paradigme ale inteligenței computaționale încorporate în produse devenite deja comerciale cum sunt: vederea artificială/analiza imaginilor, prelucrarea limbajului natural, sistemele distribuite de comandă și monitorizare bazate pe agenți. Alte paradigme provocatoare în inteligența artificială cum sunt inteligența socială și „affective computing” își crează drum pentru un transfer tehnologic la scară globală alături de tehnologii emergente cum sunt *smart grid*, *internet of things* și *pervasive computing*. Strâns legat de una dintre problemele noastre cotidiene – transportul – se prefigurează un alt exemplu provocator: *Intelligent driving & Autonomous vehicle*. Grație

evoluției sistemelor bazate pe microsisteme în industria autovehiculelor [3], inteligența încorporată la bordul autovehiculelor și capacitatea acestora de a comunica cu o infrastructură rutieră la fel de inteligentă va atinge o masă critică ce va permite apariția comercială a autovehiculelor autonome.

*Etica privind inteligența artificială* este o parte a eticii tehnologiei specifică roboților și altor ființe/entități cu inteligență artificială. Etica este definită în general ca domeniu de «*studiu teoretic al principiilor și concepțiilor de bază din orice domeniu al gândirii și activității practice*»; un «*ansamblu de norme în raport cu care un grup uman își reglează comportamentul pentru a deosebi ce este legitim și acceptabil în realizarea scopurilor, morală*». Aspectele ce țin de etică sau morală în general constituie adesea un subiect de controversă, atunci când sunt puse la îndoială scopurile cercetării științifice, care de cele mai multe ori, în esență, sunt considerate nobile. Marvin Minsky afirma cu o notă de ironie: „An ethicist is someone who sees something wrong with whatever you have in mind”. Totuși, având în vedere că tehnologia accesibilă ar permite creerea și utilizarea inteligenței artificiale în mod iresponsabil, există motive de îngrijorare în ceea ce privește comportamentul moral al oamenilor în timp ce proiectează, construiesc, utilizează și interacționează cu ființe sau entități inteligente artificiale – acest domeniu de interes fiind numit *robotică*. În aceeași notă de interes, există și temeri cu privire la comportamentul entităților artificiale construite în raport cu creatorii lor umani – domeniu numit *etica mașinilor*. Primele elemente de etică au vizat tocmai acest ultim aspect și au fost magistral enunțate de Issac Asimov. Consacrate sub denumirea de legile roboticii ele reprezintă în fond un set detrei reguli, care se enunță astfel: **Legea 1** – *Un robot nu are voie să pricinuiască vreun rău unei ființe umane, sau, prin neintervenție, să permită ca unei ființe omenesti să i se facă un rău*; **Legea 2** – *Un robot trebuie să se supună ordinelor date de către o ființă umană, atât timp cât ele nu intră în contradicție cu Legea 1*; **Legea 3** – *Un robot trebuie să-și protejeze propria existență, atât timp cât acest lucru nu intră în contradicție cu Legea 1 sau Legea 2*. Ulterior, Asimov a enunțat așa numita **Lege supremă** sau **Legea 0** – *Un robot nu are voie să pricinuiască vreun rău umanității sau să permită prin neintervenție ca umanitatea să fie pusă în pericol*.

*Provocările de ordin etic* se manifestă proporțional cu inevitabila evoluție tehnologică prin apariția unor sisteme tehnice ce tind să schimbe fundamental modul de viață al oamenilor și să ofere chiar noi modele de

funcționare a societății umane. Inteligența artificială nu mai este doar un joc de imaginație sau o paradigmă software simulativă; ea se regăsește deja încorporată în lucruri fizice create cu scopul nobil de a ne face viața mai bună. Tendința clară de supertehnologizare va duce societatea umană în situația de a coopera cu sisteme tot mai inteligente, la fel de inteligente ca și omul, poate chiar superinteligente. Este normal, de pildă, să evaluăm gradul de încredere într-un vehicul autonom care ne transportă, sau în mod particular într-un avion fără pilot uman care efectuează o aterizare pe timp de noapte. În aceeași măsură, trebuie să ne punem problema responsabilității sistemelor inteligente, a răspunderii pentru consecințele acțiunilor lor.

Pornind de la ideea generală unanim acceptată conform căreia creierul uman este mult mai mult decât o mașină de calcul, din perspectiva mecanismelor cognitive creierul artificial ar putea fi realizat într-un interval de timp rezonabil. Potrivit predicției lui *Ray Kurzweil* folosind legea lui Moore<sup>41</sup>, IA ar putea să triumfe conform următoarelor predicții: *până în anul 2029 calculatoarele personale vor avea aceeași putere de procesare ca și creierul uman, și până în 2045, IA va putea să se autoproiecteze și să se autoîmbunătățească.*

În prezent, pe lângă deficitul recunoscut de putere de calcul și prelucrare, există și o insuficiență a modelelor de raționament care trebuie implementate pe hardware-ul existent. Astfel, dintre cele trei tipuri fundamentale de raționament: logic-formal, analogic și intuitiv, pentru ultimul nu există încă un model robust care să poată fi implementat hardware sau software. Ori, tocmai raționamentul intuitiv (Commonsense Reasoning) este cel care contribuie la diferența majoră dintre inteligența umană și cea artificială. Privind holistic, neechivalența dintre cele două tipuri de inteligență este ilustrată în Fig.1.

Efortul de cercetare pe drumul de reproducere prin mijloace artificiale a consistenței intelectului uman și chiar mai adânc, deschide posibilitatea realizării unei structuri de creier artificial care să deruleze procese avansate de prelucrare a informației specifice minții umane: **gândirea, imaginația**

<sup>41</sup> Legea lui Moore descrie o tendință pe termen lung în istoria mașinilor de calcul: numărul de tranzistori care pot fi plasați pe un circuit integrat se dublează aproximativ la fiecare doi ani. Această tendință a continuat de mai bine de o jumătate de secol. Surse din 2005 se așteptau ca această tendință să continue cel puțin până în 2015 sau 2020. (Sursa: <https://ro.wikipedia.org>).

& creativitatea. Demersul de cercetare pentru **implementarea de procese non-cognitive** va conduce în viitor la crearea unor mașini cu statut moral – înzestrate cu conștiință (sau cel puțin, cu un tip particular de conștiință) și chiar la umanizarea mașinilor – acestea devenind capabile să manifeste sentimente, să aibă suflet. Provocările majore țin de apariția mașinilor conștiente de sine, capabile să se reproducă cu performanțe superioare în mod recursiv, ceea ce va conduce la apariția de entități și sisteme superinteligente.



Fig. nr. 1 – Diferența esențială între inteligența artificială și cea a ființei umane

Se justifică două seturi de întrebări care reflectă ambele sensuri ale raportului om-mașină sau creator și creația sa:

A1. În ce măsură sunt acceptabile experimentele pe oameni pentru cercetarea IA?

A2. Este legitim să reproducem creierul uman?

A3. Este legitim să se unească omul cu mașina (*transumanismul* de tip cyborg)?

A4. Avem dreptul să creem lucruri asupra cărora este posibil să nu mai deținem controlul?

B1. Cât de sigure (de încredere) pot fi mașinile inteligente?

B2. Vor fi mașinile inteligente atât de sigure încât să coopereze cu oamenii?

B3. Care va fi nivelul de control între mașini inteligente și oameni?

B4. Care va fi natura raporturilor de subordonare om-mașină inteligentă?

B5. Ce drepturi vor avea mașinile inteligente?

B6. Cum vor răspunde mașinile inteligente pentru deciziile și acțiunile lor?

Nu vom încerca să răspundem aici și acum la aceste întrebări. O vreme, acestea ar putea rămâne simple expresii retorice. Vom remarca

totuși că aspectul tehnologic al inteligenței artificiale tinde în general să fie reglementat, ghidat de standarde specifice ca oricare alt domeniu tehnic. În mod similar vor trebui elaborate și ghiduri etice, directive sau protocoale specifice pentru cercetarea, realizarea și exploatarea sistemelor cu inteligență artificială. Cu siguranță întrebările de mai sus și multe altele vor primi răspunsurile potrivite la momentul potrivit. Dincolo de aceste aspecte formale totuși, pentru dilema *tehnologie versus etică* rămâne o temă de reflecție capitală: Ne va lăsa oare Dumnezeu să facem toate aceste lucruri?

### **Referințe bibliografice:**

- [1] Banks, Jim, The Neurotechnological Revolution, in IEEE Pulse, March/April 2015, Vol.6, No 2, pp. 10–15.
- [2] Fodor, Jerry, Hume Variations, Oxford University Press, 2003.
- [3] Iancu, Ștefan, O istorie recentă: aplicații ale tehnologiei de microsistem în industria auto, Studii și Comunicări/DIS, Vol. VII, 2014.
- [4] \*\*\* Gods And Their Advanced Technologies: Talos A Greek Robot Created By The God Of The Forge – <http://www.messagetoeagle.com/talosrobot.php#.VYwWrvntmko>.