CZU 004.8
DOI https://doi.org/10.52388/1811-0770.2024.2(252).03

# EVOLUȚIA FUNDAMENTELOR SPRE UN CADRU NORMATIV CONCEPTUAL PRIVIND INTELIGENȚA ARTIFICIALĂ (IA)

**Irina BUZU**
Doctorandă, Universitatea Liberă Internațională din Moldova (ULIM),
Chișinău, Republica Moldova
e-mail: *irina.buzu@gmail.com*
https://orcid.org/0009-0000-1484-5176

**Vitalie GAMURARI**
Doctor în drept, Conferențiar universitar, Universitatea Liberă Internațională din Moldova (ULIM),
Chișinău, Republica Moldova
e-mail: *vgamurari@ulim.md*
https://orcid.org/0000-0003-3634-8554

*Articolul explorează evoluția și fundamentele inteligenței artificiale (IA), de la originile teoretice din secolul XX până la aplicațiile moderne. Autorii discută contribuțiile esențiale ale unor pionieri precum Turing, McCarthy și Minsky, prezentând o cronologie a dezvoltării tehnologice. Se analizează progresele recente în deeplearning, puterea computațională și impactul social și etic al IA. Lucrarea subliniază nevoia redefinirii relației dintre oameni și mașini, propunând o abordare centrată pe obiective umane. Sunt evidențiate atât succesele, cât și limitele actuale ale IA, accentuând rolul crucial al unei inteligențe artificiale compatibile cu valorile umane.*
***Cuvinte-cheie:*** *inteligență artificială, algoritmi, învățare profundă, agenți inteligenți, universalitate computațională.*

## THE EVOLUTION OF THE FOUNDATIONS TOWARD A CONCEPTUAL REGULATORY FRAMEWORK ON ARTIFICIAL INTELLIGENCE (AI)

*The article explores the evolution and foundational principles of artificial intelligence (AI), from its theoretical origins in the 20th century to its modern applications. The authors discuss the essential contributions of pioneers such as Turing, McCarthy, and Minsky, presenting a timeline of technological development. Recent advances in deep learning, computational power, and the social and ethical implications of AI are analyzed. The paper highlights the need to redefine the relationship between humans and machines, proposing a human-centered approach to AI. It out lines both the successes and current limitations of AI, emphasizing the crucial role of artificial intelligence aligned with human values.*
***Keywords****: artificial intelligence, algorithms, deep learning, intelligent agents, computational universality.*

## L'ÉVOLUTION DES FONDEMENTS VERS UN CADRE NORMATIF CONCEPTUEL SUR L'INTELLIGENCE ARTIFICIELLE (IA)

*L'article explore l'évolution et les fondements de l'intelligence artificielle (IA), depuis ses origines théoriques au XXe siècle jusqu'àses applications contemporaines. Les auteurs abordent les contributions essentielles de pionniers tels que Turing, McCarthy et Minsky, en présentant une chronologie du développement technologique. Les progrès récents*

*en apprentissage profond, puissance de calcul et implications sociales et éthiques de l'IAsontanalysés. L'étude souligne la nécessité de redéfinir la relation entre l'humain et la machine, en proposant une approche centrée sur les objectifs humains. Elle met en lumière à la foisles succès et les limites actuelles de l'IA, en insistant sur l'importance d'une intelligence artificielle compatible avec les valeurs humaines.*

***Mots-clés**: intelligence artificielle, algorithmes, apprentissage profond, agents intelligents, universalité computationnelle.*

**ЭВОЛЮЦИЯ ОСНОВ К КОНЦЕПТУАЛЬНОЙ НОРМАТИВНОЙ БАЗЕ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (ИИ)**

*В данной статье рассматривается эволюция и фундаментальные основы искусственного интеллекта (ИИ) – от его теоретических истоков в XX веке до современных технлогий. Авторы анализируют ключевые вклады таких пионеров, как Тьюринг, Маккарти и Мински, представляя хронологию технологического развития. Рассматриваются последние достижения в области глубокого обучения, вычислительных мощностей, а также социальные и этические аспекты ИИ. Работа подчеркивает необходимость переосмысления взаимоотношений между человеком и машиной, предлагая подход, ориентированный на человеческие цели. Освещаются как успехи, так и текущие ограничения ИИ, акцентируя важность соответствия искусственного интеллекта человеческим ценностям.*

***Ключевые слова**: искусственный интеллект, алгоритмы, глубокое обучение, интеллектуальные агенты, вычислительная универсальность.*

## Introducere

Inteligența artificială (IA) reprezintă una dintre cele mai dinamice și controversate ramuri ale științei contemporane, aflându-se la intersecția dintre informatică, matematică, filozofie și etică. De la începuturile sale conceptuale în anii 1950, IA a evoluat spectaculos, reușind să simuleze procese cognitive precum învățarea, raționamentul sau recunoașterea vizuală și auditivă. Această dezvoltare tehnologică a fost posibilă datorită progreselor în algoritmică, creșterii capacității de calcul și accesului la volume masive de date. Cu toate acestea, IA nu este doar un fenomen tehnologic, ci și unul profund social, influențând modul în care interacționăm cu informația, luăm decizii și definim inteligența umană.

Lucrarea de față își propune să analizeze în mod critic evoluția și fundamentele conceptuale ale inteligenței artificiale, punând accent pe etapele-cheie ale dezvoltării sale, contribuțiile teoretice majore și implicațiile actuale asupra societății. Studiul combină o perspectivă istorică cu o reflecție filozofică asupra definiției inteligenței, a scopurilor IA și a modului în care aceasta ar trebui proiectată pentru a fi în armonie cu valorile umane. În acest context, se conturează o direcție de cercetare esențială: trecerea de la o inteligență artificială pur funcțională la una cu adevărat benefică pentru oameni.

## Metodologie

Metodologia utilizată în acest studiu este una teoretico-istorică, bazată pe analiza documentară și sinteza literaturii de specialitate. Autorii compilează și corelează contribuții semnificative din istoria IA, evidențiind momente-cheie, concepte fundamentale și evoluția paradigmelor tehnologice. Sunt utilizate surse academice, rapoarte științifice și lucrări de referință pentru a susține o reflecție critică asupra progresului IA. Demersul se axează pe identificarea etapelor-cheie în dezvoltarea inteligenței artificiale, interpretarea conceptuală a termenului „inteligență" și analiza implicațiilor filozofice și etice ale interacțiunii dintre oameni și mașini.

## Rezultatele cercetării

Rădăcinile inteligenței artificiale (IA) se întind cu mult înapoi, în antichitate, dar începutul său „oficial" a fost în 1956. Doi tineri matematicieni, John McCarthy și MarvinMinsky, îl convinseseră pe Claude Shannon, deja faimos ca inventator al teoriei informației, și pe Nathaniel Rochester, designerul primului computer comercial al IBM, pentru a li se alătura în organizarea unui program de vară la Darthmouth College. Scopul a fost enunțat după cum urmează: „*Studiul urmează să continue pe baza presupunerii că fiecare aspect al învățării sau orice altă trăsătură a inteligenței poate fi, în principiu, descrisă atât de precis încât să poată fi făcută o mașină să o simuleze. Se va încerca să se găsească posibilitatea ca mașinile să folosească limbajul, din abstracții și concepte, să rezolve tipuri de probleme rezervate acum oamenilor și să se perfecționeze. Credem că se poate face un progres semnificativ în una sau mai multe dintre aceste probleme dacă un grup de oameni de știință atent selectat lucrează împreună pentru o vară.*" [1]. Inutil să spunem că a durat mult mai mult decât o vară: încă lucrăm la toate aceste probleme.

În primul deceniu după întâlnirea de la Darthmouth, IA a cunoscut câteva succese majore, inclusiv algoritmul lui Alan Robinson pentru raționament logic de uz general [2, p. 27] un program de joc de dame al lui Arthur Samuel, care a învățat singur să-și bată creatorul [3, p. 216]. Prima bulă IA a izbucnit la sfârșitul anilor 1960, când eforturile timpurii de învățare automată și traducere automată nu au reușit să se ridice la înălțimea așteptărilor. Un raport comandat de guvernul Regatului Unit în 1973 a concluzionat: „În nicio parte a domeniului descoperirile făcute până acum nu au produs impactul major care a fost promis" [4, p. 12]. Cu alte cuvinte, mașinile pur și simplu nu erau suficient de inteligente.

La mijlocul anilor 1980, IA a cunoscut o revigorare uriașă datorită potențialului comercial al așa-numitelor sisteme experte. A doua bulă IA s-a spart când aceste sisteme s-au dovedit a fi nepotrivite pentru multe dintre sarcinile care le-au fost delegate. Din nou, mașinile pur și simplu nu erau suficient de inteligente. A urmat o iarnă IA. Comunitatea AI și-a învățat lecția iar domeniul a devenit mult mai matematic. S-au făcut conexiuni cu disciplinele de lungă durată ale probabilității, statisticii și teoria controlului. Semințele progresului de astăzi au fost semănate în timpul acelei ierni IA, inclusiv lucrările timpurii asupra sistemelor de raționament probabilistic la scară largă și ceea ce mai târziu a devenit cunoscut sub numele de învățare profundă (*din en. deeplearning*). Începând cu 2011, tehnicile de învățare profundă au început să producă progrese dramatice în recunoașterea vorbirii, recunoașterea vizuală a obiectelor și traducerea automată - trei dintre cele mai importante probleme deschise în acest domeniu. Potrivit unor estimări, mașinile acum egalează sau depășesc capacitățile umane în aceste domenii. În 2016 și 2017, AlphaGo de la Deep Mind i-a învins pe Lee Sedol, fostul campion mondial la Go, și pe KeJie, actualul campion - evenimente despre care unii experți au prezis că nu se vor întâmpla până în 2097 [5]. Acum IA generează acoperire media de la o pagină aproape în fiecare zi. Mii de companii start-up au fost create, alimentate de un val de fonduri de risc. Investițiile care provin de la fonduri de risc, guverne naționale și corporații importante sunt de zeci de miliarde de dolari anual - mai mulți bani în ultimii cinci ani decât în întreaga istorie anterioară a domeniului. Pentru a înțelege mai bine această evoluție, prezentăm câteva momente-cheie din istoria IA - tabelul 1.

**Tabelul 1. Reprezentarea cronologică a evoluției Inteligenței Artificiale**

| An | Eveniment |
|---|---|
| **1943** | McCulloch&Pitts propun baza rețelelor neuronale într-un articol fundamental. |
| **1950** | Alan Turing propune Testul Turing pentru măsurarea inteligenței mașinilor. |
| **1951** | Minsky&Edmonds construiesc SNAR, prima rețea neuronală computerizată. |
| **1956** | Conferința Dartmouth marchează nașterea IA ca domeniu de studiu. |
| **1957** | Rosenblatt dezvoltă Perceptron-ul, prima rețea neuronală artificială capabilă de învățare. |
| **1965** | Weizenbaum dezvoltă ELIZA, un program de procesare a limbajului natural. |
| **1967** | Newell& Simon dezvoltă General Problem Solver (GPS), un sistem IA de rezolvare a problemelor. |
| **1974** | Prima „iarnă IA" - scădere a finanțării și interesului față de IA. |
| **1980** | Sistemele expert devin populare în afaceri și medicină. |
| **1986** | Hinton, Rumelhart& Williams dezvoltă backpropagation pentru antrenarea rețelelor neuronale profunde. |
| **1997** | Deep Blue (IBM) învinge campionul mondial la șah, GarryKasparov. |
| **2002** | iRobot lansează Roomba, primul aspirator robotizat bazat pe IA. |
| **2011** | Watson (IBM) învinge campioni umani la Jeopardy! |
| **2012** | DeepMind dezvoltă o rețea neuronală care recunoaște pisici în videoclipuri YouTube. |
| **2014** | Facebook dezvoltă DeepFace, un sistem avansat de recunoaștere facială. |
| **2015** | AlphaGo (DeepMind) învinge un campion uman de Go. |
| **2016** | AlphaGo învinge cel mai bun jucător mondial de Go într-o serie de meciuri. |
| **2020** | OpenAI lansează GPT-3, un progres major în procesarea limbajului natural. |
| **2021** | AlphaFold2 (DeepMind) rezolvă problema plierii proteinelor. |
| **2022** | Google concediază un inginer pentru afirmații privind conștiința modelului LaMDA. |
| **2023** | Artiști dau în judecată companii AI pentru utilizarea artei protejate prin drepturi de autor. |
| **2024** | Guvernele implementează reglementări globale privind utilizarea etică a IA. |

*Sursa:* Elaborat de către autor.

Istoria IA a fost condusă de o singură mantră: cu cât mai inteligent, cu atât mai bine. În cartea sa *"Humancompatible. AI andthe problem of control",* Stuart Russell susține că aceasta este o greșeală - nu din cauza unei frici vagi de a fi depășiți de mașini, ci din cauza modului în care am înțeles inteligența însăși [6, p. 368]. Conceptul de inteligență este esențial pentru cine suntem – de aceea ne numim HomoSapiens, sau om înțelept. După mai bine de două mii de ani de auto-reflecție, am ajuns la o caracterizare a inteligenței care se poate rezuma la următoarele: *Oamenii sunt **inteligenți** în măsura în care ne putem aștepta ca acțiunile **noastre** să **ne** atingă obiectivele.* Toate celelalte caracteristici ale inteligenței - perceperea, gândirea, învățarea, inventarea și așa mai departe - pot fi înțelese prin contribuțiile lor la capacitatea

noastră de a acționa cu succes. Încă de la începuturile IA, inteligența mașinilor a fost definită într-un mod similar: *Mașinile sunt **inteligente** în măsura în care ne putem ca acțiunile **lor** să-și atingă obiectivele.*

Deoarece mașinile, spre deosebire de oameni, nu au obiective proprii, noi le oferim obiective de realizat. Cu alte cuvinte, construim mașini de optimizare, introducem obiective în ele și pornesc. Cu toate acestea, nu vrem mașini inteligente în acest sens. Dezavantajul acestui model a fost evidențiat în 1960 de Norbert Wiener, un profesor legendar la MIT și unul dintre cei mai importanți matematicieni de la mijlocul secolului al XX-lea. Wiener tocmai văzuse că programul de joc de dame al lui Arthur Samuel învăța să joace dame mult mai bine decât creatorul său, iar această experiență l-a determinat să scrie o lucrare prevăzătoare, dar puțin cunoscută, „*Some Moral andTechnicalConsequences of Automation"*. Iată cum afirmă el punctul principal: *Dacă folosim, pentru a ne atinge scopurile, o agenție mecanică cu a cărei funcționare nu putem interveni eficient... ar fi bine să fim siguri că scopul introdus în mașină este scopul pe care ni-l dorim cu adevărat"* [7, p. 1356].

Problema este chiar acolo în definiția de bază a IA. Spunem că mașinile sunt inteligente în măsura în care se poate aștepta ca acțiunile lor să își atingă obiectivele, dar nu avem o modalitate fiabilă de a ne asigura că obiectivele lor sunt aceleași cu obiectivele noastre. Ce se întâmplă dacă, în loc de a permite mașinilor să-și urmărească obiectivele, insistăm ca acestea să urmărească obiectivele noastre? O astfel de mașină, dacă ar putea fi proiectată, ar fi nu doar inteligentă, ci și benefică pentru oameni și, prin urmare, am obține asta: *Mașinile sunt **benefice** în măsura în care ne putem aștepta ca acțiunile **lor** să **ne** atingă obiectivele.* Eliminarea presupunerii că mașinile ar trebui să aibă un obiectiv definit înseamnă că ar trebui să reconfiguram o parte din fundația și suprastructura IA. Rezultatele ar fi o nouă relație simbiotică sau asistată între oameni și mașini. Dar pot acele mașini să fie considerate inteligente?

A avea o definiție rezonabilă a inteligenței este primul ingredient în crearea mașinilor inteligente. Al doilea ingredient este o mașină (sau un computer) în care poate fi realizată definiția. Suntem atât de obișnuiți cu computerele încât abia le observăm puterile incredibile. Dacă ai un laptop sau un desktop sau un telefon inteligent, uită-te la el: o cutie mică, cu o modalitate de a introduce caractere. Doar tastând, putem crea programe care transformă cutia în ceva care sintetizează imagini în mișcare, traduce engleza în chineză, ascultă și vorbește, învinge campionii mondiali la șah. Această capacitate a unei singure mașini de a efectua orice proces pe care ți-l poți imagina se numește *universalitate* [8, p. 231], un concept introdus pentru prima dată de Alan Turing în 1936. Universalitatea înseamnă că nu avem nevoie de mașini separate pentru aritmetică, traducere automată, șah, înțelegere a vorbirii sau animație: o singură mașină face totul. Lucrarea lui Turing care introduce universalitatea a fost probabil cea mai importantă lucrare tematică scrisă vreodată. În ea, autorul descrie un simplu dispozitiv de calcul care ar putea accepta ca intrare (*din en. input*) descrierea oricărui alt dispozitiv de calcul, împreună cu intrarea acelui al doilea dispozitiv și, prin simularea funcționării celui de-al doilea dispozitiv la intrarea sa, să producă aceeași ieșire (*din en. output*) ca și al doilea. dispozitivul ar fi produs. Numim acum acest prim dispozitiv o mașină Turing universală. Pentru a-și demonstra universalitatea, Turing a introdus definiții precise pentru două noi tipuri de obiecte matematice: mașini și programe. Împreună, mașina și programul definesc o secvență de evenimente. Domeniul care a apărut prin urmare, știința computerizată, a provocat o explozie în următorii șaptezeci de ani, generând o gamă largă de concepte, design-uri, metode și aplicații noi, precum și șapte dintre cele mai valoroase opt companii din lume.

Conceptul central al informaticii este cel de algoritm, care este o metodă precis specificată pentru a calcula ceva. Algoritmii sunt, până acum, părți familiare ale vieții de zi cu zi: un algoritm cu rădăcină pătrată dintr-un calculator de buzunar primește un număr ca intrare și returnează rădăcina pătrată a acelui număr ca rezultat; un algoritm de joc de șah ia o poziție de șah și returnează o mișcare; un algoritm de găsire a rutei ia o locație de pornire, o locație a obiectivului și o hartă a străzilor și returnează traseul rapid de la început la obiectiv. Algoritmii pot fi descriși în limbaj sau în notație matematică, dar pentru a fi implementați trebuie să fie codați ca programe într-un limbaj de programare.

Hardware-ul computerelor contează deoarece computerele mai rapide, cu mai multă memorie și putere de calcul, permit algoritmilor să ruleze mai rapid și să gestioneze mai multe informații. Progresele în acest domeniu sunt bine-cunoscute. Primul computer programabil electronic comercial, Ferranti Mark I [9], putea executa aproximativ o mie ($10^3$) de instrucțiuni pe secundă și avea aproximativ o mie de octeți de memorie principală. Cel mai rapid computer de la începutul anului 2019, aparatul Summit de la Laboratorul Național OakRidge din Tennessee [10], execută aproximativ $10^{18}$ instrucțiuni pe secundă (de o mie de trilioane de ori mai rapid). Acest progres a rezultat din progresele în dispozitive electronice și fizică, permițând un grad incredibil de miniaturizare. Deși comparația între calculatoare și creier nu este deosebit de semnificativă, cifrele pentru Summit depășesc ușor capacitatea brută a creierului uman, care are aproximativ $10^{15}$ sinapse și un timp de ciclu de aproximativ o sutime de secundă, pentru un maxim teoretic de aproximativ $10^{17}$ operațiuni pe secundă. Cea mai mare diferență este consumul de energie: Summit folosește de aproximativ un milion de ori mai mult consum [11, p. 368].

În anii 1950, computerele erau descrise în presa populară drept super-creiere care erau „mai rapide decât Einstein” [12, p. 668]. Cu toate acestea, nu putem spune că computerele sunt la fel de puternice ca creierul uman, pentru că concentrarea asupra puterii brute de calcul ratează complet ideea. Numai viteza nu ne va oferi IA. Rularea unui algoritm prost proiectat pe un computer mai rapid nu face algoritmul mai bun; înseamnă doar că primești răspunsul greșit mai repede. Principalul efect al mașinilor mai rapide a fost acela de a scurta timpul pentru experimentare, astfel încât cercetarea să poată progresa mai repede. Marele matematician francez din secolul al XVII-lea Blaise Pascal a fost primul care a dezvoltat un calculator mecanic real și practic. Deși nu putea decât să adună și să scadă și a fost folosit în principal în biroul de colectare a impozitelor al tatălui său, l-a determinat pe Pascal să scrie: „Mașina aritmetică produce efecte care sunt mai aproape de gândire decât de toate acțiunile animalelor” [13, p. 368].

Tehnologia a făcut un salt dramatic înainte în secolul al XIX-lea, când matematicianul și inventatorul britanic Charles Babbage a proiectat *AnalyticalEngine* (Motorul Analitic), o mașină universală programabilă în sensul definit mai târziu de Turing. A fost ajutat în munca sa de Ada, Contesa de Lovelace, fiica lordului Byron. În timp ce Babbage spera să folosească *AnalyticalEngine* pentru a calcula cu precizie tabele matematice și astronomice, Lovelace și-a înțeles adevăratul potențial [14, p. 669], descriindu-l în 1842 drept o „mașină care gândește sau... o mașină rezonabilă” care ar putea raționa despre „toți subiectele din univers”. Multă vreme, *AnalyticalEngine* nu a fost niciodată construit, iar ideile lui Lovelace au fost în mare măsură uitate. Odată cu munca teoretică a lui Turing din 1936 și impulsul ulterior al celui de-al Doilea Război Mondial, mașinile de calcul universale au fost în sfârșit realizate în anii 1940. Gândurile despre crearea inteligenței au urmat imediat. Lucrarea lui Turing din 1950, „*ComputingMachineryand Intelligence*” este cea mai cunoscută dintre multe lucrări timpurii despre posibilitatea mașinilor inteligente. El

a propus un test operațional pentru inteligență, numit *jocul imitației*, care ulterior a devenit cunoscut sub numele de *testul Turing* [15, p. 438]. Testul măsoară comportamentul mașinii - în special, capacitatea sa de a păcăli un interogator uman, făcându-l să creadă că este uman.

Conceptul central al IA moderne este agentul inteligent, ceva care percepe și acționează. Modul în care construim agenți inteligenți depinde de natura problemei cu care ne confruntăm. Aceasta, la rândul său, depinde de trei lucruri:

1. natura mediului în care agentul va opera;
2. observațiile și acțiunile care leagă agentul de mediu;
3. obiectivul agentului.

IA a făcut multe progrese în probleme precum jocurile de societate și puzzle-urile care sunt observabile, dezordonate, deterministe și au reguli cunoscute. Pentru tipurile de probleme mai ușoare, cercetătorii IA au dezvoltat algoritmi destul de generali și eficienți și o înțelegere teoretică solidă; adesea, mașinile depășesc performanța umană în acest tip de probleme. Probleme precum conducerea unui guvern sau predarea biologiei moleculare sunt mult mai dificile. Au medii complexe, în mare parte neobservabile, mult mai multe obiecte și tipuri de obiecte, nu există o definiție clară a acțiunilor, în mare parte reguli necunoscute, multă incertitudine și scale de timp foarte lungi. Când construim sisteme IA pentru aceste tipuri de sarcini, acestea trebuie să necesite o mare cantitate de inginerie specifică problemei și sunt adesea foarte fragile.

Progresul către generalitate are loc atunci când elaborăm metode care sunt eficiente pentru problemele mai dificile dintr-un anumit tip de metodă care necesită mai puține și mai slabe presupuneri, astfel încât să fie aplicabile la mai multe probleme. IA generală (*din en. General purpose AI sau GPAI*) ar fi o metodă care este aplicabilă pentru toate tipurile de probleme și funcționează eficient pentru cazuri mari și dificile, cu foarte puține presupuneri. O astfel de metodă de uz general nu există încă, dar ne apropiem. Doar pentru a exemplifica, o versiune de AlphaGo numită AlphaZero a învățat recent să depășească AlphaGo la Go și, de asemenea, să depășească Stockfish (cel mai bun program de șah din lume, mult mai eficient decât orice om) și Elmo (cel mai bun program shogi din lume, de asemenea, mai bun decât orice om). AlphaZero a făcut toate acestea într-o singură zi [16].

Când AlphaGo l-a învins pe Lee Sedol și mai târziu pe toți ceilalți jucători de top Go, mulți oameni au presupus că, deoarece o mașină a învățat de la zero să învingă rasa umană la o sarcină despre care se știa că este foarte dificilă chiar și pentru oamenii foarte inteligenți, a fost doar o chestiune de timp înainte ca IA să preia controlul. Chiar și unii sceptici s-ar putea să fi fost convinși când AlphaZero a câștigat la șah și shogi, precum și la Go. Dar AlphaZero a avut limitări: funcționa doar în clasa de jocuri discrete, observabile, pentru doi jucători, cu reguli cunoscute. Abordarea pur și simplu nu va funcționa deloc pentru conducere, predare, conducerea unui guvern sau preluarea lumii.

## Concluzii

Evoluția inteligenței artificiale (IA) evidențiază un parcurs marcat de alternanța dintre entuziasm tehnologic și perioade de stagnare, cunoscute sub numele de „ierni IA”. Fiecare avans major – de la algoritmii timpurii până la rețelele neuronale profunde – a contribuit la extinderea capacităților IA, însă a adus în același timp noi provocări conceptuale și etice. Progresele recente în domeniul învățării automate și al procesării limbajului natural au permis obținerea unor performanțe comparabile sau superioare celor umane în sarcini specifice, demonstrând potențialul extraordinar al acestor tehnologii.

Totuși, dezvoltarea IA ridică întrebări esențiale legate de controlul și alinierea obiectivelor mașinilor cu valorile și interesele umane. Inteligența nu ar tre-

bui definită doar prin capacitatea de atingere a unui scop, ci și prin contextul etic al acțiunilor întreprinse. Astfel, autorii subliniază necesitatea unei schimbări de paradigmă: în locul unor mașini pur inteligente, este esențial să construim mașini benefice, capabile să susțină și să îmbunătățească viața umană. Acest lucru presupune regândirea relației dintre om și tehnologie, dar și dezvoltarea unor mecanisme fiabile de control și reglementare care să asigure utilizarea responsabilă a inteligenței artificiale în societate.

## Referințe bibliografice


1. DARTMOUTH COLLEGE. *Artificial Intelligence (AI) coined at Dartmouth, Dartmouth College*. [citat 03.11.2023] Disponibil: https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth.

2. ROBINSON, John Alan. *A machine-oriented logic based on theresolutionprinciple.*[online] În: Journal of the ACM, 1965, vol. 12, p. 23-41. [citat 03.11.2023] Disponibil: https://doi.org/10.1145/321250.321253

3. SAMUEL, Arthur. *Somestudies in machinelearningusingthe game of checkers*. [online] În: IBM Journal of ResearchandDevelopment, 1959, vol. 3, p. 210-229. [citat 03.11.2023] Disponibil: https://ieeexplore.ieee.org/document/5392560lore

4. LIGHTHILL, James. *Artificial intelligence: A general survey*. [online] În: Artificial Intelligence: A Paper-Symposium, ScienceResearch Council of Great Britain, 1973, p. 1-16. [citat 03.11.2023] Disponibil: https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf

5. JOHNSON, George. *To test a powerful computer, play an ancient game*. [online] The New York Times, 1997. [citat 03.11.2023] Disponibil: https://www.nytimes.com/1997/07/29/science/to-test-a-powerful-computer-play-an-ancient-game.html

6. RUSSELL, Stuart. *HumanCompatible. AI andthe Problem of Control*. Londra: PenguinBooks, 2020. ISBN 978-0-141-98750-7, 368 p.

7. WIENER, Norbert. *Some moral andtechnicalconsequences of automation*. [online] În: Science, 1960, vol. 131, p. 1355-1358. [citat 12.06.2024] Disponibil: https://www.science.org/doi/10.1126/science.131.3410.1355

8. TURING, Alan. *On computablenumbers, with an applicationtotheEntscheindungsproblem*. [online] În: Proceedings of the London Mathematical Society, 2nd ser., 1936, vol. 42, p. 230-265. [citat:12.06.2024] Disponibil: https://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf

9. University of Manchester. Ferranti Mark 1 – The firstcommerciallyavailable computer. Manchester: [citat 12.06.2024] Disponibil: https://curation.cs.manchester.ac.uk/digital60/www.digital60.org/birth/manchestercomputers/mark1/ferranti.html.

10. OakRidge National Laboratory. [citat 12.06.2024] Disponibil: https://www.olcf.ornl.gov/summit/.

11. RUSSELL, Stuart. *HumanCompatible. AI andthe Problem of Control*. Londra: PenguinBooks, 2020. ISBN 978-0-141-98750-7, 368 p.

12. MENABREA, Luigi Federico. *Sketch of theAnalyticalEngineinventedby Charles Babbage*. [online] În: ScientificMemoirs, vol. III, ed. R. Taylor, 1843, p. 666-673. [citat 12.06.2024] Disponibil: https://johnrhudson.me.uk/computing/Menabrea_Sketch.pdf

13. RUSSELL, Stuart. *HumanCompatible. AI andthe Problem of Control*. Londra: PenguinBooks, 2020. ISBN 978-0-141-98750-7, 368 p.

14. MENABREA, Luigi Federico. *Sketch of theAnalyticalEngineinventedby Charles Babbage*. [online] În: ScientificMemoirs, vol. III, ed. R. Taylor, 1843, p. 666-673. [citat 12.06.2024] Disponibil:https://johnrhudson.me.uk/computing/Menabrea_Sketch.pdf

15. TURING, Alan. *Computingmachineryand intelligence*. [online] În: Mind, 1950, vol. 59, nr. 236, p. 433-460. [citat 12.06.2024] Disponibil: https://academic.oup.com/mind/article-abstract/LIX/236/433/986238

16. SILVER, David. et al. *Masteringchessandshogiby self-play with a general reinforcementlearningalgorithm.* [online] În: arXiv. 2017, 1712.018015. [citat 12.06.2024] Disponibil: https://arxiv.org/abs/1712.018015.