UNIVERSITATEA DE STAT DIN MOLDOVA

ȘCOALA DOCTORALĂ ȘTIINȚE UMANISTE ȘI ALE EDUCAȚIEI

CONSORȚIU: Universitatea de Stat din Moldova (USM), Universitatea de Stat „Alecu Russo” din Bălți (USARB), Universitatea de Stat „Bogdan Petriceicu Hașdeu” din Cahul (USC).

Cu titlu de manuscris

C.Z.U.: 111.1:001.2:004.8(043.3)

DUCA MIHAIL

INTELIGENȚA ARTIFICIALĂ: O PERSPECTIVĂ TRANSDISCIPLINARĂ

631.01 ONTOLOGIE ȘI GNOSEOLOGIE

Teză de doctor în filosofie

Autor: Duca Mihail

Conducător științific: Bogatu Eugenia,
dr. hab. în filosofie, conf. univ.

CHIȘINĂU, 2026

©Duca Mihail, 2026

Cuprins

ADNOTARE **5**
LISTA FIGURILOR **8**
INTRODUCERE **9**
1. FUNDAMENTE TEORETICE ALE INTELIGENȚEI ARTIFICIALE: O ABORDARE TRANSDISCIPLINARĂ **19**

1.1 Geneza și evoluția conceptului de Inteligență Artificială **19**
1.2 Cadrul științific și stadiul actual al cercetărilor **26**
1.3 De la interdisciplinaritate la transdisciplinaritate. **41**
1.4 Niveluri de realitate și logica terțului inclus în studiul inteligenței artificiale **54**
1.5 Concluzii **61**

2. DIMENSIUNI CONCEPTUAL-METODOLOGICE ALE CERCETĂRII: STUDII DE CAZ **62**

2.1 Arhitectura și funcționarea rețelelor neuronale artificiale **62**
2.2 Limitările, performanțele și diferențele comparativ cu un agent inteligent uman. Problemele legate de înțelegerea contextului situațional și cognitiv. Inteligența simulată versus Inteligența în sine. Studiul de caz: ChatGPT. **71**

2.2.1 Provocarea schemelor Winograd și înțelegerea contextului **74**
2.2.2 Studiu de caz. Teste-dialog cu ChatGPT 3.5 ..... **78**
2.2.3 Dialectica dintre limbajul natural și cel artificial **87**

2.3 Recunoașterea formelor și algoritmii de joc. Studii de caz: scrisul de mână și șahul **98**
2.4 Inteligența simulată vs. inteligența autentică: O perspectivă non-reducționistă **113**

2.4.1 Non-reducționismul și nivelurile de realitate **116**
2.4.2 Interpretarea non-reducționistă a inteligenței artificiale **121**

2.5 Concluzii **127**

3. DE LA TRANSUMANISM LA TRANSDISCIPLINARITATE - IPOSTAZE FILOSOFICE **128**

3.1 Modelul transdisciplinar al realității: Învățare, înțelegere, înțelepciune 128
3.2 Provocarea Transumanismului: Homo Deus și redefinirea naturii umane 140
3.3 Perspectiva teologică și antropologică: Corporalitatea și Theosis 152
3.4 Etica și responsabilitatea în era post-umană. Critica transdisciplinară a transumanismului ca reducționism unidimensional și analiza inegalităților sociale generate de tehnologie 162
3.5 Concluzii 166

CONCLUZII GENERALE ȘI RECOMANDĂRI 167
BIBLIOGRAFIE 171
ANEXE 183
Declarația privind asumarea răspunderii 193
CV-ul candidatului 194

ADNOTARE

Duca Mihail, Inteligența artificială: o perspectivă transdisciplinară, Chișinău, 2026

Structura tezei: introducere, trei capitole, concluzii generale și recomandări, bibliografie din 190 de titluri, 161 de pagini de text de bază, 18 figuri, rezultatele obținute fiind publicate în 7 lucrări științifice.

Cuvinte-cheie: inteligență artificială, rețele neuronale, tehnologie, transdisciplinaritate, complexitate, reducționism, non-reducționism, holism, inteligență umană, transumanism, *Homo Deus, Theosis*.

Scopul lucrării: constă în realizarea unei cercetări transdisciplinare a inteligenței artificiale care ar facilita o înțelegere mai profundă a unui subiect extrem de popular și actual pentru societatea supratehnologizată în care trăim.

Obiectivele cercetării: analiza și sintetiza cunoștințelor existente în domeniul inteligenței artificiale, coroborat cu diversele abordări; identificarea și evaluarea domeniilor de aplicare în care inteligența artificială poate beneficia de o abordare transdisciplinară; dezvoltarea unui model transdisciplinar a inteligenței artificiale; evaluarea riguroasă a performanței și impactului abordării transdisciplinare în domeniul inteligenței artificiale; dezvoltarea cunoașterii în domeniul inteligenței artificiale prin intermediul unei abordări transdisciplinare; înțelegerea ființei umane și a complexității ireductibile a minții umane, prin aplicarea metodei transdisciplinare; analiza transumanismului în raport cu inteligența artificială; evaluarea corporalității ca element esențial al inteligenței

Noutatea și originalitatea științifică constau în aplicarea unei metodologii filosofice transdisciplinare asupra inteligenței artificiale, depășind analizele pur tehnice sau chiar interdisciplinare. Originalitatea este susținută de elaborarea unui model transdisciplinar al ființei umane, care demonstrează limitele ontologice ale inteligenței artificiale în raport cu complexitatea ireductibilă a inteligenței umane. Totodată, lucrarea propune o viziune critică inovatoare prin juxtapunerea conceptelor de *Homo Deus* (transumanism) și *Theosis* (perspectiva teologic-filosofică), oferind un cadru conceptual nou pentru evaluarea impactului etic și existențial al tehnologiilor emergente.

Rezultatele obținute care contribuie la soluționarea unei probleme științifice importante: depășirea viziunii reducționiste asupra inteligenței artificiale prin fundamentarea unei perspective transdisciplinare care clarifică distincția ontologică dintre algoritm și mintea umană. Lucrarea oferă o soluție metodologică pentru integrarea nivelurilor de realitate în analiza tehnologiei, demonstrând că inteligența umană posedă o complexitate ireductibilă ce nu poate fi replicată integral prin procese strict computaționale.

Semnificația teoretică: fundamentarea unui cadru conceptual extins pentru înțelegerea inteligenței artificiale, care transcende reducționismul algoritmic prin intermediul metodologiei transdisciplinare. Lucrarea contribuie la dezvoltarea filosofiei tehnologiei și a antropologiei filosofice prin redefinirea raportului dintre inteligență, conștiință și corporalitate.

Valoarea aplicativă: rezidă în furnizarea unui suport metodologic pentru mediul educațional, oferind un cadru pedagogic ce integrează etica și gândirea critică în studiul tehnologiilor avansate. Aceasta facilitează formarea viitorilor specialiști prin clarificarea distincției fundamentale dintre procesele algoritmice și complexitatea inteligenței umane.

Implementarea rezultatelor științifice: comunicări în cadrul a 14 conferințe naționale și internaționale, 7 articole științifice și activități de cercetare.

АННОТАЦИЯ

Дука Михаил, Искусственный интеллект: трансдисциплинарная перспектива, Кишинев, 2026

Структура диссертации: введение, три главы, общие выводы и рекомендации, библиография из 190 наименований, 161 страниц основного текста, 18 рисунков. Результаты исследования опубликованы в 7 научных работах.

Ключевые слова: искусственный интеллект, нейронные сети, технология, трансдисциплинарность, сложность, редукционизм, нон-редукционизм, холизм, человеческий интеллект, трансгуманизм, *Homo Deus, Теозис.*

Цель работы: проведение трансдисциплинарного исследования искусственного интеллекта для более глубокого понимания этой актуальной темы в контексте современного сверхтехнологичного общества.

Задачи исследования: анализ и синтез существующих знаний в области искусственного интеллекта; определение сфер применения, в которых ИИ может воспользоваться трансдисциплинарным подходом; разработка трансдисциплинарной модели ИИ; строгая оценка эффективности и влияния трансдисциплинарного подхода в области ИИ; развитие знаний об ИИ посредством трансдисциплинарного метода; понимание человеческой сущности и сложности человеческого разума через призму трансдисциплинарности; анализ трансгуманизма в контексте ИИ; оценка телесности как сущностного элемента интеллекта.

Научная новизна и оригинальность: заключаются в применении трансдисциплинарной методологии к исследованию искусственного интеллекта, выходящей за рамки чисто технических или междисциплинарных подходов. Оригинальность подтверждается разработкой трансдисциплинарной модели человека, демонстрирующей онтологические пределы ИИ по сравнению с неустранимой сложностью человеческого интеллекта. Также работа предлагает инновационный критический взгляд через сопоставление концепций *Homo Deus* (трансгуманизм) и *Теозиса* (богословско-философская перспектива).

Решенные научные проблемы: преодоление редукционистского взгляда на ИИ путем обоснования трансдисциплинарной перспективы, проясняющей онтологическое различие между алгоритмом и человеческим разумом. Работа предлагает методологическое решение для интеграции множества уровней реальности в анализ технологий, доказывая, что человеческий интеллект обладает сложностью, которая не может быть полностью воспроизведена вычислительными процессами.

Теоретическая значимость: обоснование расширенной концептуальной базы для понимания ИИ, преодолевающей алгоритмический редукционизм. Работа вносит вклад в развитие философии техники и философской антропологии через определение взаимосвязи между интеллектом, сознанием и телесностью.

Практическая ценность: заключается в предоставлении методологической поддержки для образовательной среды, предлагая педагогическую основу, интегрирующую этику и критическое мышление в изучение высоких технологий. Это способствует подготовке будущих специалистов через разъяснение фундаментального различия между алгоритмическими процессами и сложностью человеческого интеллекта.

Внедрение научных результатов: 14 национальных и международных конференций, 7 научных статей и научно-исследовательская деятельность.

ABSTRACT

Duca Mihail, Artificial Intelligence: A Transdisciplinary Approach, Chișinău, 2026

Thesis Structure: introduction, three chapters, general conclusions and recommendations, bibliography of 190 titles, 161 pages of core text, 18 figures. The results obtained have been published in 7 scientific papers.

Keywords: artificial intelligence, neural networks, technology, transdisciplinarity, complexity, reductionism, non-reductionism, holism, human intelligence, transhumanism, *Homo Deus, Theosis*.

Purpose of the Work: to conduct transdisciplinary research on artificial intelligence that facilitates a deeper understanding of this highly popular topic within our hyper-technologized society.

Research Objectives: to analyze and synthesize existing knowledge in the field of AI combined with various approaches; to identify and evaluate application areas where AI can benefit from a transdisciplinary approach; to develop a transdisciplinary model of AI; to rigorously evaluate the performance and impact of the transdisciplinary approach in the field of AI; to advance knowledge in the AI field through a transdisciplinary method; to understand the human being and the irreducible complexity of the human mind by applying the transdisciplinary method; to analyze transhumanism from an AI perspective; to evaluate embodiment as an essential element of intelligence.

Scientific Novelty and Originality lie in the application of a transdisciplinary philosophical methodology to artificial intelligence, moving beyond purely technical or even interdisciplinary analyses. The originality is supported by the development of a transdisciplinary model of the human being, which demonstrates the ontological limits of AI compared to the irreducible complexity of human intelligence. Additionally, the work proposes an innovative critical vision by comparing the concepts of *Homo Deus* (transhumanism) and *Theosis* (theological-philosophical perspective).

Scientific Problem Solved: overcoming the reductionist view of artificial intelligence by grounding a transdisciplinary perspective that clarifies the ontological distinction between algorithms and the human mind. The work offers a methodological solution for integrating levels of reality into technological analysis, demonstrating that human intelligence possesses an irreducible complexity that cannot be fully replicated by strict computational processes.

Theoretical Significance: the foundation of an extended conceptual framework for understanding AI, transcending algorithmic reductionism through transdisciplinary methodology. The work contributes to the development of the philosophy of technology and philosophical anthropology by redefining the relationship between intelligence, consciousness, and embodiment.

Applicative Value: resides in providing methodological support for the educational environment, offering a pedagogical framework that integrates ethics and critical thinking into the study of advanced technologies. This facilitates the training of future specialists by clarifying the fundamental distinction between algorithmic processes and the complexity of human intelligence.

Implementation of Scientific Results: presentations at 14 national and international conferences, 7 scientific articles, and research activities.

LISTA FIGURILOR

Fig. 1. 1 Fundamentele științifice ale inteligenței artificiale. Relațiile dintre diferite domenii. (schemă simplificată) 183
Fig. 1. 2 Reprezentarea simbolică a acţiunii logicii terţului inclus [91, p. 238] 184
Fig. 1. 3 Obiectul transdisciplinar, Subiectul transdisciplinar și termenul de Interacţiune [91, p. 237] 184

Fig. 2. 1 Structura neuronului biologic [55]....185
Fig. 2. 2 Schema de principiu a unei reţele neuronale artificiale [55] 185
Fig. 2. 3 Schema de principiu al unui element procesor (neuron) și a fluxului de informație aferent [55] 186
Fig. 2. 4 Conexiunile informaționale dintre faza de antrenament și cea de aplicare, în cazul folosirii unei reţele neuronale artificiale [55].....186
Fig. 2. 5 Schema neuronului artificial [55]....187
Fig. 2. 6 Exemplu de cifră scrisă de mână din baza de date MNIST (reprezentată printr-un grafic de tip heatmap - fotografia originală are rezoluția de 28x28 și valorile reprezintă nuanțe de gri (grayscale) 187
Fig. 2. 7 Evoluția acurateței și erorii („pierderii") în procesul de antrenare și de testare (validare) 187
Fig. 2. 8 Matricea confuziilor 188
Fig. 2. 9 Cifrele 0 și 1 cu care a fost testată rețeaua neuronală....188
Fig. 2. 10 Poziție de șah pentru analiza pe computer [9, pp. 98-99]....189
Fig. 2. 11 Stockfish capturează turnul din a5 și… pierde partida. 189
Fig. 2. 12 O variație a poziției pe aceeași temă 190
Fig. 2. 13 În loc să așeze nebunul pe câmpul b4 și apoi să mute doar regele (poziția fiind, astfel, remiză), Stockfish își maximizează scorul, dar ajunge în poziție pierdută. 190
Fig. 2. 14 Schema funcțiilor active și pasive [132, p. 45]....191
Fig. 2. 15 Modelul transdisciplinar al ființei umane 192

INTRODUCERE

„Nu va fi o exagerare dacă vom afirma că
întrebarea cu privire la tehnică a devenit o
întrebare despre soarta omului și a culturii”
Nikolai Berdiaev [178]

Importanța și actualitatea temei abordate vine dintr-o realitate pe care nu o putem ignora. Dezvoltarea rapidă a inteligenței artificiale nu mai este doar o chestiune tehnică, ci a devenit o problemă care atinge însăși esența existenței noastre. Cu aproape un secol în urmă Nikolai Berdiaev avertiza vizionar că întrebarea despre tehnică s-a transformat, de fapt, într-o întrebare despre soarta omului și a întregii culturi. Astăzi, când tehnologia pare să șteargă granițele dintre ceea ce este uman și ceea ce este artificial, această avertizare este mai actuală ca niciodată. Lucrarea de față nu își propune o analiză a programelor informatice, ci încearcă să exploreze acele zone de graniță dintre discipline unde se decide viitorul demnității noastre. Considerăm că tema este esențială deoarece avem nevoie, mai mult ca oricând, să readucem profunzimea filosofică în centrul progresului tehnologic, pentru a preveni transformarea persoanei umane într-o simplă cifră sau o dată statistică.

În ultimii ani, inteligența artificială a ajuns o ramură de cercetare și dezvoltare cu o influență semnificativă asupra societății noastre contemporane. Dată fiind complexitatea fenomenului, a devenit din ce în ce mai clar că abordarea provocărilor acestei discipline necesită o perspectivă care să depășească limitele disciplinei individuale și să se aplece asupra interacțiunii dintre diverse domenii [168]. Mai mult, se constată că nici măcar acest fapt nu este suficient, pentru că rămân neexplicate concepte și fapte situate la limita spațiului dintre discipline. Tocmai în acest punct, metodologia transdisciplinară poate aduce o contribuție valoroasă în domeniul inteligenței artificiale. Această abordare deschide noi uși spre inovații și soluții mai complete și mai robuste. Prin înțelegerea complexității impactului tehnologiei asupra mediului sociocultural, eticii și economiei, cercetarea transdisciplinară are capacitatea de a modela o viziune holistică și echilibrată asupra dezvoltării inteligenței artificiale.

La fel, în contextul actual, transdisciplinaritatea poate juca un rol esențial în prevenirea și gestionarea posibilelor consecințe dăunătoare ale acestei tehnologii. Tocmai de aceea, în cadrul acestei lucrări ne-am propus studierea și evaluarea din perspectivă filosofică transdisciplinară a inteligenței artificiale, precum și a unor fenomene înrudite, de exemplu, transumanismul.

Inteligența artificială, cu actuala ei expansiune exuberantă, aduce în prim-plan o pluralitate de întrebări și provocări, ce se întind dincolo de barierele tehnologice, evidențiind o nevoie imperativă de a aduna expertiza și viziunea din diverse câmpuri disciplinare și, ceea ce este esențial pentru o abordare transdisciplinară, de a completa zonele de graniță sau chiar spațiile dintre discipline. În acest sens, metodologia transdisciplinară apare ca un instrument vital în înțelegerea complexităților emergente ce caracterizează inteligența artificială în contextul contemporan [92].

Prin abordarea inteligenței artificiale printr-o lentilă transdisciplinară, se pot desprinde diverse straturi de impact și de influență, creând un mozaic vast de analiză. În acest cadru, fenomenul transumanismului poate fi văzut ca un complement la discuția despre inteligența artificială, explorând granițele dintre om și mașină și evaluând modul în care această interacțiune remodelează percepțiile noastre despre umanitate și societate.

În cadrul acestui studiu, este esențial să se analizeze și să se evalueze critic potențialele căi de evoluție ale inteligenței artificiale, inclusiv riscurile și avantajele acesteia. Știința, tehnologia și filosofia se întrepătrund, creând un cadru pentru a modela o înțelegere cuprinzătoare a modului în care inteligența artificială își poate găsi locul în societate, într-un mod responsabil. Aici, rolul eticii devine deosebit de important, asigurând că dezvoltarea tehnologică este ghidată de principii care pun în centru bunăstarea umană.

Mai mult, pe lângă etica tehnologiei, metodologia transdisciplinară permite, de asemenea, abordarea critică a economiei, explorând modul în care inteligența artificială poate transforma structurile economice existente, creând oportunități pentru inovații în sectoare precum industrie, sănătate, educație și multe altele [92]. Acest lucru poate să faciliteze identificarea soluțiilor robuste care pot contribui la construirea unei economii mai reziliente și mai echitabile.

De asemenea, este crucială crearea unui spațiu de dialog și colaborare între diferitele domenii implicate, care să permită o discuție deschisă, în care expertizele diverse să se întâlnească și să colaboreze într-o manieră productivă și armonioasă [91]. O astfel de deschidere permite experților din diferite domenii să învețe unii de la alții și să găsească soluții care nu ar fi posibile într-o singură disciplină.

Astfel, printr-o abordare transdisciplinară, se poate spera că inteligența artificială nu va fi doar o forță de transformare tehnologică, ci și un catalizator pentru progres și inovație în multe alte domenii, promovând o dezvoltare care este atât responsabilă, cât și sustenabilă, ghidată de o viziune holistică și echilibrată asupra potențialului și limitelor tehnologiei în societatea contemporană [92].

Scopul și obiectivele lucrării. Scopul acestei lucrări este realizarea unei cercetări transdisciplinare a inteligenței artificiale care ar facilita o înțelegere mai profundă a unui subiect extrem de popular și actual pentru societatea supratehnologizată în care trăim.

Pentru a atinge ceea ce dorim, ne-am propus o serie de obiective.

- Analiza și sintetiza cunoștințelor existente în domeniul inteligenței artificiale, coroborat cu diversele abordări.
- Identificarea și evaluarea domeniilor de aplicare în care inteligența artificială poate beneficia de o abordare transdisciplinară.
- Dezvoltarea unui model transdisciplinar al inteligenței artificiale.
- Evaluarea riguroasă a performanței și impactului abordării transdisciplinare în domeniul inteligenței artificiale.
- Dezvoltarea cunoașterii în domeniul inteligenței artificiale prin intermediul unei abordări transdisciplinare.
- Înțelegerea ființei umane și a complexității ireductibile a minții umane, prin aplicarea metodei transdisciplinare.
- Analiza transumanismului în raport cu inteligența artificială.
- Evaluarea corporalității ca element esențial al inteligenței umane

Ipoteza de cercetare. Inteligența artificială rămâne o simulare funcțional-algoritmică a proceselor cognitive, fiind ontologic incapabilă să acceadă la dimensiunea noetică a înțelegerii umane. Evaluarea corectă a acestui decalaj necesită abandonarea reducționismului tehnocentrist în favoarea unei metodologii transdisciplinare, capabilă să gestioneze impactul etic și spiritual al noilor tehnologii.

Problema cercetării. Complexitatea emergentă a sistemelor de inteligență artificială contemporane generează o tensiune conceptuală majoră între performanța computațională și înțelegerea de tip noetic. Problema centrală a acestei investigații rezidă în dificultatea de a defini și circumscrie statutul ontologic al inteligenței artificiale dincolo de simulările funcționale ale cogniției. În timp ce inteligența umană se constituie ca o structură ireductibilă, integrând raționamentul cu conștiința de sine, afectivitatea și așezarea într-un context socio-cultural viu, inteligența artificială operează prin manipularea a unor volume masive de date în cadrul unor algoritmi predefiniți.

În ce măsură poate fi validată calitatea de „inteligent” a unui sistem care, deși rezolvă probleme de o complexitate extremă, rămâne lipsit de dimensiunea semantică a înțelegerii și de conștiința propriei existențe?

Cum poate fi depășit impasul reducționismului, care tinde să confunde eficiența algoritmică cu inteligența în sine, ignorând limitele insurmontabile ale mașinii în raport cu judecata morală și empatia?

Care este cadrul metodologic capabil să integreze aceste distincții ontologice pentru a ghida implementarea etică a tehnologiilor, în contextul în care autonomia tot mai mare a sistemelor de inteligență artificială diluează liniile tradiționale ale responsabilității umane?

Identificarea limitelor ontologice ale inteligenței simulate nu reprezintă doar o necesitate de ordin tehnic, ci o problemă filosofică de o actualitate critică. Absența unei distincții clare între algoritm și persoană riscă să fundamenteze o utilizare iresponsabilă a noilor tehnologii, subminând valori umane fundamentale.

Metodologia cercetării. Metoda **transdisciplinară** reprezintă inima acestui studiu, fiind cel mai important instrument prin care putem privi dincolo de barierele tehnice pentru a înțelege ce înseamnă cu adevărat inteligența artificială. Aplicarea metodologiei transdisciplinare reprezintă un salt major în direcția unei înțelegeri mai profunde și mai complete a acestei tehnologii. Inteligența artificială, în sine, încorporează o multitudine de discipline cu o multitudine de metode de cercetare, de la matematică și informatică la etică și științele sociale. Înțelegerea acestei tehnologii complexe cere mai mult decât o analiză îngustă dintr-o singură perspectivă disciplinară. Transdisciplinaritatea oferă un cadru pentru a îmbrățișa această complexitate într-un mod organizat și coerent.

Considerăm că, prin aducerea împreună a expertizelor din diverse domenii, putem construi modele de inteligență artificială care nu numai că sunt mai eficiente, dar care țin cont de implicațiile etice, sociale și filosofice ale tehnologiei. Acest lucru este crucial într-o eră în care inteligența artificială începe să afecteze aproape toate aspectele vieții noastre, de la sănătate și educație până la securitate națională și justiție socială.

În plus, transdisciplinaritatea ne permite să abordăm provocări tehnice complexe într-un mod mai inovator. Spre exemplu, probleme ca explicabilitatea algoritmilor de inteligență artificială pot fi abordate nu doar din perspectiva ingineriei, ci și din cea a psihologiei cognitive, pentru a înțelege cum oamenii interpretează și înțeleg deciziile luate de mașini. În același timp, dialogul cu eticienii și specialiștii în științele sociale ajuta la identificarea unor soluții care respectă valorile și normele culturale ale societăților în care aceste tehnologii sunt implementate.

Un alt aspect important al acestei metodologii este faptul că ea nu își propune doar să rezolve problemele imediate, dar și să exploreze întrebări fundamentale despre relația dintre om și tehnologie. Ce înseamnă pentru noi ca specie faptul că mașinile pot face lucruri pe care până de

curând doar oamenii le puteau face? Cum redefinim concepte precum inteligența, învățarea sau chiar conștiința în lumina progreselor din domeniul inteligenței artificiale?

Așa dar, metodologia transdisciplinară în studiul inteligenței artificiale nu este doar un lux academic, ci o necesitate pentru a naviga într-o manieră responsabilă prin complexitățile și provocările unei tehnologii care continuă să schimbe lumea în moduri semnificative și adesea imprevizibile. Aceasta oferă o cale pentru a gândi holistic și integrat despre cea mai semnificativă inovație tehnologică a timpului nostru.

La fel, în cadrul acestei cercetări, este aplicată **analiza fenomenologică** care este o metodologie de cercetare ce se concentrează pe investigarea experienței subiective și a percepțiilor individuale ale fenomenelor. În contextul inteligenței artificiale, această abordare implică explorarea modului în care oamenii percep și înțeleg interacțiunile cu sistemele de inteligență artificială, ceea ce ne ajută să avem o claritate despre cum utilizatorii experimentează și interpretează prezența și acțiunile acestei tehnologii în viața lor de zi cu zi, precum și modul în care aceste interacțiuni influențează percepțiile lor despre inteligență, conștiință și umanitate.

Această metodologie permite o clarificare a impactului inteligenței artificiale asupra subiectivității umane și oferă perspective valoroase asupra felului în care această tehnologie modelează identitatea, relațiile și percepția realității. Analiza fenomenologică ajută la dezvăluirea nuanțelor și subtilităților experienței umane, oferind o bază solidă pentru reflecții filosofice și dezvoltarea teoriilor legate de conștiință și inteligență.

În contextul în care modelele textuale generative pre-antrenate sunt extrem de populare astăzi, suntem obligați să luăm în considerare **hermeneutica** pentru a interpreta răspunsurile inteligenței artificiale la solicitările agentului uman. Aplicații precum asistenții virtuali sau chatbot-urile, generează o cantitate enormă de informații care poate fi examinată pentru a înțelege mai bine cum ea este percepută și cum influențează relațiile și comunicarea umană. Aceasta implică analizarea nu doar a textului în sine, ci și a contextului în care este utilizat, inclusiv și a reacțiilor utilizatorilor.

Un lucru absolut esențial este faptul că hermeneutica devine un instrument indispensabil în evaluarea textelor generate. Această abordare ne ajută să decriptăm mecanismele prin care sensul este construit în dialogul dintre om și algoritm, fiind deosebit de utilă în analizarea fenomenului „halucinațiilor”. Atunci când sistemele de inteligență artificială produc informații false sau incoerente, hermeneutica ne permite să interpretăm aceste erori nu ca pe niște simple defecțiuni tehnice, ci ca pe o consecință a absenței unei înțelegeri semantice autentice. La fel, ne permite să prevenim riscul ca halucinațiile mașinii să fie acceptate drept adevăruri absolute de către o societate care valorizează viteza informației în detrimentul profunzimii critice.

Având în vedere caracterul interdisciplinar al cercetării, **studiile comparative interdisciplinare** implică examinarea modului în care diferite discipline abordează și înțeleg inteligența artificială, cu scopul de a identifica convergențele și divergențele dintre acestea. În acest context, este analizat modul în care filosofi, psihologi, eticieni și experți în informatică interpretează și evaluează conceptele și implicațiile inteligenței artificiale. Este evident că filosofia poate oferi perspective asupra naturii conștiinței și a inteligenței, punând întrebări fundamentale despre ce înseamnă să fii inteligent sau conștient. Psihologia poate explora cum interacțiunile cu inteligența artificială influențează comportamentul și procesele mentale ale oamenilor. Etica, pe de altă parte, se concentrează pe dilemele morale și principiile care ar trebui să ghideze dezvoltarea și utilizarea inteligenței artificiale, asigurând că tehnologia este folosită într-un mod responsabil. Informatica se ocupă de aspectele tehnice ale dezvoltării și implementării inteligenței artificiale, incluzând algoritmi, învățare automată și arhitecturi de rețele neuronale.

Prin compararea acestor abordări, putem descoperi unde și cum se intersectează perspectivele disciplinelor și unde acestea diferă semnificativ. Această metodologie, ca și în cazul metodologie transdisciplinare, ajută la construirea unei înțelegeri holistice a inteligenței artificiale. Identificarea punctelor comune facilitează colaborarea interdisciplinară, în timp ce recunoașterea diferențelor deschide noi direcții de cercetare și dezvoltare. Studiile comparative interdisciplinare oferă astfel un cadru valoros pentru explorarea complexităților inteligenței artificiale dintr-o perspectivă cuprinzătoare și integrativă.

O mare parte a aceste lucrări este fundamentată pe diverse **studii de caz** care implică examinarea unor exemple specifice de aplicare a inteligenței artificiale. Această metodologie permite o analiză detaliată și contextualizată a modului în care inteligența artificială este integrată și utilizată în diverse sectoare. Prin examinarea a trei studii de caz specifice, încercăm să adunăm dovezi pentru a construi o înțelegere complexă a modului în care inteligența artificială funcționează și diferențele conceptuale cu inteligența de tip uman.

Sumarul compartimentelor tezei. Structura tezei este organizată în felul următor: introducere, trei capitole, concluzii și recomandări, bibliografie și anexe.

În **Introducere**, se pune un accent deosebit pe argumentarea actualității și importanței subiectului ales pentru studiu. Scopul, obiectivele cercetării și ipoteza sunt clare și bine definite. De asemenea, se discută problema științifică soluționată și se subliniază importanța teoretică și valoarea aplicativă a tezei. Sunt prezentate cinci metode de cercetare, metoda de bază fiind cea transdisciplinară.

În **capitolul 1**, intitulat *FUNDAMENTE TEORETICE ALE INTELIGENȚEI ARTIFICIALE: O ABORDARE TRANSDISCIPLINARĂ***,** oferim o analiză a evoluției conceptului de inteligență artificială, a bazelor sale științifice și a cadrului metodologic transdisciplinar.

În subcapitolul 1.1, *Geneza și evoluția conceptului de Inteligență Artificială*, analizăm definițiile conceptuale și etapele istorice majore. Inteligența artificială a evoluat de la visul lui Alan Turing de a crea mașini care imită mintea umană până la sistemele moderne de *„deep learning"* care procesează volume uriașe de date. Istoria sa este marcată de succese răsunătoare în jocuri de strategie precum șahul sau Go, dar și de perioade de stagnare. La fel, în acest subcapitol, sunt prezentate domeniile de aplicare a inteligenței artificiale.

În subcapitolul 1.2, *Cadrul științific și stadiul actual al cercetărilor*, arătăm că cercetarea actuală în inteligența artificială este un efort colectiv uriaș care îmbină matematica riguroasă și logica formală cu principii din biologie, fizică și neuroștiințe. Deși algoritmii pot simula procese complexe precum recunoașterea facială sau diagnosticul medical, ei se lovesc de limite matematice fundamentale, cum este teorema lui Gödel, care ne amintește că nicio mașină nu poate cuprinde întregul adevăr prin calcule. Esențial este să înțelegem că inteligența artificială rămâne un produs ingineresc bazat pe reguli, fiind structural diferită de inteligența umană care este dependentă de emoții, experiențe și biologie.

În subcapitolul 1.3, *De la interdisciplinaritate la transdisciplinaritate*, justificăm necesitatea integrării diverselor discipline și explicăm de ce abordările tradiționale sunt adesea insuficiente pentru a cuprinde complexitatea fenomenului de inteligență artificială. Lucrarea propune trecerea de la simpla colaborare între experți (interdisciplinaritate) la o abordare transdisciplinară, care să integreze cunoașterea exterioară cu experiența interioară a subiectului. Această metodologie ajută la o înțelegere mai bună a fenomenelor mentale complexe și la evitarea unei perspective pur fizicaliste asupra omului.

În subcapitolul 1.4, *Niveluri de realitate și logica terțului inclus în studiul inteligenței artificiale*, vorbim despre metodologia transdisciplinară. Basarab Nicolescu susține că realitatea este structurată pe mai multe niveluri de realitate (clasic, cuantic, ciberspațiu-timp), fiecare având legi proprii. Trecerea de la un nivel de realitate la altul este explicată prin logica terțului inclus, care permite unificarea unor concepte contradictorii fără a încălca axioma noncontradicției. Această viziune arată că există o legătură strânsă între nivelurile de realitate ale obiectului studiat și nivelurile de percepție ale celui care cunoaște.

În **capitolul 2**, intitulat *DIMENSIUNI CONCEPTUAL-METODOLOGICE ALE CERCETĂRII: STUDII DE CAZ*, sunt prezentate trei studii de caz care explorează performanțele

și limitele inteligenței artificiale, demonstrând totodată eficiența metodologiei transdisciplinare prin validarea modelului transdisciplinar al ființei umane (vezi **Fig. 2.** ***15***).

În subcapitolul *2.1 Arhitectura și funcționarea rețelelor neuronale artificiale*, explicăm modul în care rețelele neuronale artificiale sunt inspirate din sistemul biologic, bazându-se pe unități de procesare simple (neuroni) interconectate prin ponderi adaptabile. Spre deosebire de programarea clasică, aceste sisteme nu urmează instrucțiuni rigide, ci „învață" prin ajustarea conexiunilor în funcție de datele de antrenament, manifestând capacități de generalizare și sinteză a informațiilor complexe.

În subcapitolul *2.2 Limitările, performanțele și diferențele comparativ cu un agent inteligent uman. Problemele legate de înțelegerea contextului situațional și cognitiv. Inteligența simulată versus Inteligența în sine. Studiul de caz: ChatGPT*, analizăm distincția dintre inteligența simulată și cea autentică, subliniind că, deși modelele actuale pot procesa volume uriașe de date, ele rămân limitate de absența conștiinței și a subiectivității. Performanța tehnică a mașinii nu echivalează cu înțelegerea umană, fiind mai degrabă un instrument analitic decât un agent cu gândire autonomă. În compartimentul *2.2.1 Provocarea schemelor Winograd și înțelegerea contextului,* sunt prezentate testele de tipul schemelor Winograd ca metodă de evaluare a capacității inteligenței artificiale de a descifra ambiguitățile lingvistice și referințele anaforice care necesită logică de bun-simț. În compartimentul *2.2.2 Studiu de caz. Teste-dialog cu ChatGPT 3.5*, prin experimente directe, demonstrăm că ChatGPT eșuează adesea în a rezolva probleme de logică elementară sau jocuri de cuvinte bazate pe polisemiei (de exemplu, termenul „mine"). Rezultatele confirmă argumentul „camerei chinezești" a lui John Searle. Sistemul manipulează simboluri și sintaxă la un nivel înalt, dar fără a avea nicio fărâmă de înțelegere semantică reală. În compartimentul *2.2.3 Dialectica dintre limbajul natural și cel artificial,* explorăm contrastul dintre limbajele umane, care sunt organice, ambigue și creative, și cele artificiale, care tind spre regularitate și structuri matematice fixe. Subliniem decalajul dintre competență (cunoașterea regulilor) și performanță (utilizarea lor în context), demonstrând că inteligența artificială nu poate replica funcția autoreferențială și poetică a vorbirii umane.

În subcapitolul *2.3 Recunoașterea formelor și algoritmii de joc. Studii de caz: scrisul de mână și șahul,* sunt analizate limitele inteligenței artificiale în sarcini vizuale și strategice, folosind exemple precum baza de date MNIST și anomalii din jocul de șah pe care motorul Stockfish nu le poate gestiona.

În subcapitolul *2.4 Inteligența simulată vs. inteligența autentică: O perspectivă non-reducționistă,* criticăm abordările reducționiste care încearcă să explice mintea doar prin procese fizice sau mecanice, propunând în schimb o viziune transdisciplinară. În compartimentul *2.4.1*

Non-reducționismul și nivelurile de realitate, prezentăm modelul nivelurilor de Realitate, respingând ideea că totul poate fi redus la fizică. Folosind conceptele lui Roy Clouser despre proprietăți active și pasive, explicăm faptul că ființa umană este singura care funcționează activ în toate aspectele realității, de la cel fizic și biologic la cel logic și spiritual. În compartimentul *2.4.2 Interpretarea non-reducționistă a inteligenței artificiale,* în baza modelului transdisciplinar al ființei umane stabilim limita ontologică a inteligenței artificiale. Aceasta se oprește la nivelul învățării (NR_3, procesare de date), în timp ce omul atinge nivelul înțelegerii și al înțelepciunii (NR_4 și NR_5).

În **capitolul 3**, intitulat *DE LA TRANSUMANISM LA TRANSDISCIPLINARITATE - IPOSTAZE FILOSOFICE*, analizăm limitările ontologice ale inteligenței artificiale în raport cu inteligența umană definită de învățare, înțelegere și înțelepciune, evidențiind importanța experienței corporale. În paralel, criticăm proiectul transumanist care își propune depășirea condiției umane actuale.

În subcapitolul *3.1 Modelul transdisciplinar al realității: Învățare, înțelegere, înțelepciune,* argumentăm că inteligența artificială este limitată la procesarea de date și nu poate accesa sensul autentic sau dimensiunea morală a cunoașterii. În timp ce înțelegerea umană este un proces holistic ce implică experiență subiectivă și structuri apriorice, algoritmul rămâne captiv într-o viziune reducționistă, lipsit de capacitatea de a înțelege cu atât mai mult posibilitatea de a atinge înțelepciunea.

În subcapitolul *3.2 Provocarea Transumanismului: Homo Deus și redefinirea naturii umane,* analizam curentul transumanist ca pe o tentativă tehnologică de a anula limitările biologice și de a recrea specia sub forma unui post-om sau *Homo Deus*. Aici arătăm contrastul dintre divinizarea artificială, bazată pe „upgrade" material cu idealul creștin al *Theosis*-ului, care propune restaurarea firii umane prin participarea la harul divin.

În subcapitolul *3.3 Perspectiva teologică și antropologică: Corporalitatea și Theosis,* arătăm că inteligența și conștiința sunt indisociabile de experiența corporală, trupul fiind considerat în antropologia creștină o dimensiune sacră a persoanei, nu un simplu obiect. Utilizând viziunea Sfântului Maxim Mărturisitorul, spunem că desăvârșirea umană depinde de unitatea trup-suflet și de libertatea de a alege între *logos*-ul naturii și *tropos*-ul existenței, elemente absente în sistemele de inteligență artificială.

În subcapitolul *3.4 Etica și responsabilitatea în era post-umană. Critica transdisciplinară a transumanismului ca reducționism unidimensional și analiza inegalităților sociale generate de tehnologie,* criticăm transumanismul dintr-o perspectiva transdisciplinară. Afirmăm că el este un reducționism materialist care riscă să adâncească inegalitățile sociale și să fragmenteze umanitatea

prin accesul inegal la tehnologie. La fel, evidențiem pericolele legate de pierderea autonomiei și a identității umane, subliniind lipsa unor soluții etice reale în cadrul mișcărilor post-umaniste pentru gestionarea riscurilor existențiale generate.

În compartimentul final **Concluzii generale și recomandări** este evidențiată relevanța aplicării metodologiei transdisciplinare în analiza inteligenței artificiale. Schema transdisciplinară a ființei umane evidențiază limitările inerente ale inteligenței artificiale, subliniind absența nivelurilor de realitate care definesc complexitatea inteligenței umane.

Bibliografia constă din 190 de surse în română, engleză, franceză și rusă.

Cuvinte-cheie: inteligență artificială, rețele neuronale, tehnologie, transdisciplinaritate, complexitate, reducționism, non-reducționism, holism, inteligență umană, transumanism, *Homo Deus, Theosis*.

1. FUNDAMENTE TEORETICE ALE INTELIGENȚEI ARTIFICIALE: O ABORDARE TRANSDISCIPLINARĂ

1.1 Geneza și evoluția conceptului de Inteligență Artificială

Cel mai bun lucru pe care putem să-l facem pentru a pune ordine în discuția despre inteligența artificială este ca, de la bun început, să propunem câteva definiții conceptuale. În mod intuitiv, plecând de la etimologia termenilor în discuție, inteligența artificială este un domeniu al tehnoștiinței care încearcă să conceapă sisteme cu proprietăți pe care noi le asociem, în mod obișnuit, unei inteligențe de tip uman. Activitățile inteligente umane sunt multiple și implică anumite operațiuni, de exemplu recunoașterea și înțelegerea limbajului natural - o activitate care nu necesită foarte mult efort pentru om, dar care creează multe dificultăți și care scoate în evidență anumite limite ale algoritmilor de inteligență artificială. Despre limitele inteligenței artificiale vom vorbi mai jos, când vom încerca să schițăm acele obstacole epistemologice fundamentale de care se lovesc cercetătorii din domeniu. Pe lângă înțelegerea limbajelor naturale, mai putem vorbi despre activități inteligente care implică învățarea, rezolvarea diferitor probleme, inclusiv demonstrarea teoremelor, recunoașterea imaginilor, jocurile etc.

D. Dumitrescu, în lucrarea „Principiile Inteligenței Artificiale” [32], clasifică definițiile posibile în patru categorii de bază. Autorul menționează că în definirea inteligenței artificiale accentul se poate pune fie asupra proceselor cognitive, fie asupra comportamentului. Astfel, definițiile pot avea în vedere următoarele premise: mașinile cu inteligență artificială 1) gândesc asemenea oamenilor, 2) gândesc rațional, 3) acționează asemenea oamenilor, 4) acționează rațional. În măsura în care unele dintre aceste premise sunt acceptate, sunt posibile definiții ca cele de mai jos:

- Inteligența artificială constă în automatizarea activităților pe care noi le asociem cu gândirea umană (luarea deciziilor după anumite reguli sau în condiții de incertitudine, rezolvarea problemelor într-o manieră mai degrabă euristică, învățarea).
- Inteligența artificială are legătură cu efortul de a face sisteme de calcul să gândească într-o maniera specific umană, de a înzestra mașinile cu o minte umană.
- Inteligența artificială se definește ca fiind studiul capacităților mentale cu ajutorul modelelor computaționale (abordarea cognitivistă).

- Inteligența artificială este studiul proceselor de calcul care fac posibile percepția senzorială, raționamentul, voința și acțiunea.
- Inteligența artificială este activitatea prin care se concep mașini capabile să îndeplinească funcții pentru care suntem de acord că este necesară inteligență atunci când sunt îndeplinite de oameni.
- Inteligența artificială este rezultatul cercetărilor prin care se realizează sisteme capabile să îndeplinească sarcini care, dacă sunt îndeplinite de oameni, implică inteligență.
- Inteligența artificială este activitatea prin care încercăm să proiectăm sisteme de calcul care pot realiza sarcini pe care, în acest moment, oamenii le fac mai bine (recunoașterea de pattern-uri, limbajul și înțelegerea limbajului ș.a.).
- Inteligența artificială este domeniul de studiu care, folosind procese computaționale, caută să explice și să imite cât mai fidel comportamentul uman inteligent.
- Inteligența artificială este acel domeniu al informaticii care își propune automatizarea comportamentului inteligent.

Din definițiile de mai sus putem extrage anumite concluzii cu privire la ceea ce ar trebui să înțelegem prin inteligență artificială. Nici una din aceste definiții nu este pur tehnică. Toate definițiile vizează diferite aspecte specifice ființei umane. Într-un fel, chiar și acei cercetători care se ocupă strict de inginerie sau programare, fără a fi interesați prea mult de abordările filosofice, formulează definiții ale inteligenței artificiale care afirmă, chiar și în plan secundar, ceva despre ceea ce înseamnă să fii om. Altfel nici nu se poate, deoarece, în încercarea de a construi o inteligență artificială, singurele modele pe care le regăsim se raportează, de cele mai multe ori, la inteligența umană. Tocmai această apropiere a tehnologiei de specificul uman poate sta la baza unei varietăți de mituri care ne pot modela viața.

După clarificarea conceptului de inteligență artificială este util să schițăm **o scurtă istorie a evoluției** acestei tehnologii. Cu riscul de a omite elemente importante vom lua ca punct de cotitură un eveniment editorial de la mijlocul secolului XX, atunci când tehnologia devine suficient de avansată pentru a putea da primele roade în privința simulării activității inteligente non-biologice. Pentru a putea vorbi cu adevărat despre o inteligență non-umană a fost nevoie de o inteligență umană de-a dreptul genială. Așadar, Alan Turing publica, în 1950, în revista *„Mind”* (revistă a Universității Oxford), un articol care avea să fie celebru în literatura de specialitate: *„Computing Machinery and Intelligence”* [172]. După cum se vede încă din titlu, Turing

analizează posibilitatea existenței mașinilor inteligente. Articolul lui Turing poate fi considerat un punct de cotitură în istoria inteligenței artificiale. În acele timpuri computerele dispuneau de o putere de calcul deloc impresionată, în comparație cu ceea ce avem noi astăzi. În acest articol, printre altele, Turing formulează renumitul său test care avea drept scop identificarea unei inteligențe de tip uman în mașini. Testul Turing s-a dovedit a fi atât de dificil încât, până în prezent, nici o tehnologie nu a fost în stare să-l treacă. Până și cei mai optimiști experți în inteligența artificială nu văd cum, în acest moment, testul Turing ar putea fi trecut. Se pare că este necesară o abordare cu totul diferită față de cea actuală. Desigur, aceasta nu înseamnă că, în viitor, testul nu va putea fi trecut. Dar, în cadrul acestei lucrări, ne vom raporta numai la realizările din prezent, fără a ne lansa în scenarii previzioniste sau chiar science-fiction.

Un lucru uimitor s-a întâmplat odată cu apariția internetului și, mai târziu, odată cu apariția rețelelor de socializare: o explozie de date care devin disponibile pentru cercetătorii din domeniul inteligenței artificiale. Până atunci acest domeniu era divizat în două tabere care abordau problema diferit. Unii încercau ca prin intermediul codificării anumitor reguli logice (deducție) să învețe computerele să imite un comportament inteligent (inteligența artificială clasică, cu sisteme expert sau simbolice), pe când adepții celeilalte tabere încercau să reconstruiască însuși creierul uman (inteligența modernă, cu abordările legate de rețele neuronale). Chiar dacă au apărut încă din deceniul 6 al secolului trecut, rețelele neuronale nu au avut succesul demonstrat de inteligența artificială clasică și în deceniul 8 al secolului XX întreg domeniul a intrat în așa-numită „iarnă a inteligenței artificiale”. Au existat mai multe „ierni” în istoria inteligenței artificiale. Aceste perioade sunt asociate cu o scădere a interesului și a finanțării proiectelor care dezvoltau această tehnologie. Odată cu volumul imens de date din internet și creșterea capacității de calcul, rețelele neuronale au devenit dominante în cercetarea inteligenței artificiale. Pentru a antrena rețelele neuronale, este necesară o cantitate masivă de date, făcându-le astfel capabile să recunoască anumite tipare. Antrenarea este posibilă pentru că în zilele noastre cercetătorii au acces la baze de date cu un număr imens de exemple, iar puterea de calcul actuală permite o analiză rapidă. Astfel, a apărut inteligența artificială bazată pe învățarea profundă *(deep learning)*, care, folosind rețelele neuronale artificiale cu mai multe straturi, realizează în acest moment performanțe spectaculoase (cel mai faimos exemplu este ChatGPT, în baza căruia vom realiza și un studiu de caz în lucrarea de față). Cu toate acestea, putem reitera că la moment nu avem realizată o tehnologie a inteligenței artificiale generale [30].

Pentru a simplifica și mai mult istoricul nostru despre inteligența artificială și performanțele sale, vom examina câteva exemple din domeniul jocurilor de strategie. Jocul de șah este, în mod evident, considerat drept o activitate care necesită inteligență. Încă din 1958, A. Newell, în

colaborare cu alți autori, publică în *„IBM Journal of Research and Development”* un articol intitulat „*Chess-Playing Programs and the Problem of Complexity”* („Programele care joacă șah și problema complexității”), în care se afirmă că *„dacă cineva ar putea construi cu succes o mașină care joacă șah, acel cineva ar părea că a pătruns până în miezul efortului intelectual uman”* [163, p.320-335]. Într-adevăr, nu toți oamenii știu să joace șah, puțini ajung să joace șah bine și doar câțiva sunt extrem de buni. Noi considerăm marii maeștrii în șah ca fiind persoane cu un intelect destul de dezvoltat. Filosoful american Hubert Dreyfus în cartea sa, *„What Computers Can't Do”* („Ce nu pot face computerele”), publicată în 1978, susținea că mașinile de calcul capabile să joace șah mai bine decât un maestru vor rămâne o ficțiune [135]. Chiar dacă lucrarea lui Dreyfus are idei valoroase cu privire la limitele inteligenței artificiale, această predicție s-a dovedit a fi greșită. Ne aducem aminte cu toții de un eveniment răsunător al anului 1997, eveniment care astăzi este considerat un alt moment de cotitură în istoria inteligenței artificiale. Garri Kasparov, campionul mondial la șah din acel moment, a fost învins de către supercomputerul de la IBM Deep Blue.

Va fi destul de dificil să trecem în revistă toate performanțele mașinilor de calcul în legătură cu jocurile de strategie. Nick Bostrom, în cartea sa „Superinteligența. Căi, Pericole, Strategii.” ne prezintă un tabel cu mai multe jocuri în care tehnologia este superioară omului [14, p.35-38]. Vom analiza două tipuri de jocuri de strategie, care, din punctul nostru de vedere, merită o abordare mai serioasă. În 2014, anul în care Bostrom își scrie cartea, nici un computer nu era în stare să câștige în fața unui campion de Go – un joc inventat în China, cu o strategie mult mai complexă decât șahul. Computerul era la nivelul unui jucător mediu. După înfrângerea suferită la șah – un joc în care tactica și, în consecință, puterea de calcul este esențială – se credea că omul va rămâne neînvins la Go, dată fiind complexitatea strategică „necalculabilă” a acestui joc. Dar, în 2017, Ke Jie, campion la Go, avea să piardă cu 3 - 0 în fața programului AlphaGo, un program de la Google, antrenat cu tehnici de învățare profundă *(deep learning)*. A fost nu doar un triumf al inteligenței artificiale, ci și o victorie a marilor companii IT din Occident. Un detaliu interesant a fost făcut cunoscut, mai târziu, de către Kai-Fu Lee, un fost angajat al reprezentanței Google din China. În martie 2016, se disputase un meci de cinci partide, în care AlphaGo câștigase cu 4-1 în fața unui alt adversar renumit (Lee Sedol). Kai-Fu Lee menționează că aceste jocuri au trecut aproape neobservate în Occident, dar au atras peste 280 milioane de privitori chinezi. *„Peste noapte, China a fost cuprinsă de febra inteligenței artificiale”* [68, p.18]. Astăzi China este unul din liderii în domeniul cercetării inteligenței artificiale.

Watson este un program conceput în cadrul firmei IBM, care a învățat să joace Jeopardy – un joc televizat cu întrebări și răspunsuri care necesită nu doar putere de calcul, cunoștințe de cultură generală, dar, mai mult decât atât, înțelegerea limbajului natural, un lucru destul de dificil

pentru mașinile de calcul. Regulile și scopul jocului sunt cât se poate de simple pentru un om. În loc să pornească de la o întrebare, jocul pornește de la o afirmație care conține niște informații de cultură generală și jucătorii trebuie să formuleze o întrebare corectă. De exemplu, pentru afirmația „poetul român care a scris Luceafărul" răspunsul corect sub forma întrebării ar fi „Cine este Mihai Eminescu?". După un deceniu de cercetări, în 2011, Watson i-a învins pe doi campioni în Jeopardy, Brad Rutter și Ken Jennings.

De fiecare dată când o tehnologie depășea inteligența umană, fie și în aspecte destul de specifice și înguste, unii cercetători veneau cu afirmația că, de fapt, aceasta nu este o inteligență adevărată și că inteligența umană este cu totul altceva. Nici Deep Blue, nici AlphaGo și nici Watson nu posedă o inteligență generală umană, dar performanțele demonstrate ne-au obligat să ne gândim mai mult la ceea ce este de fapt inteligența, ce este inteligența umană și de ce oamenii gândesc diferit de o mașină? Eric J. Larson, în „Mitul Inteligenței Artificiale. De ce computerele nu pot gândi la fel ca noi" [66] explică modul în care toate aceste tehnologii funcționează, care sunt limitările lor și de ce este nevoie de o nouă teorie fundamentală pentru a putea avansa în cercetările din acest domeniu.

Nu în ultimul rând este necesar să arătăm și **domeniile de aplicare** a inteligenței artificiale. Această tehnologie a avut un impact semnificativ într-o varietate de domenii, aducând cu sine promisiuni și provocări noi. Capacitatea sa de a simula și imita inteligența umană a deschis uși către o gamă largă de aplicații în diferite domenii [168; 115].

În medicină, inteligența artificială a revoluționat domeniul prin dezvoltarea unor algoritmi și modele de învățare automată capabile să diagnosticheze și să realizeze predicții legate de diferite afecțiuni medicale. De la analiza imaginilor medicale și detectarea cancerului, până la diagnosticarea timpurie a bolilor și personalizarea tratamentelor [139]. O dovadă clară a succesului acestei colaborări este rezolvarea misterului privind plierea proteinelor[1], o enigmă care i-a pus la încercare pe cercetători mai bine de jumătate de secol. Cu ajutorul sistemului AlphaFold, inteligența artificială a reușit să descifreze în doar câteva zile formele tridimensionale ale proteinelor, o sarcină care, înainte, ar fi cerut decenii de muncă grea în laborator. Această reușită are o însemnătate uriașă pentru omenire. Ea le oferă medicilor o hartă detaliată a bolilor la nivel atomic. În consecință, putem crea medicamente care să se potrivească perfect în mecanismul celulei, așa cum se potrivește cheia în broască, aducând speranța vindecării pentru mulți pacienți.

[1] Plierea este procesul prin care proteina trece de la un fir bidimensional la o formă 3D funcțională. Dacă proteina se „împachetează" greșit, ea nu mai funcționează sau devine toxică.

Tehnologiile inteligenței artificiale sunt esențiale pentru dezvoltarea mașinilor autonome. Sistemele de asistență în conducerea automobilului, cum ar fi frânarea automată de urgență și asistența la menținerea benzii de circulație, folosesc algoritmi de învățare automată pentru a interpreta datele din mediul înconjurător și a lua decizii în timp real pentru a asigura siguranța șoferului și a pasagerilor [168]. Dacă astăzi încă privim cu uimire imaginile din orașe precum San Francisco sau Phoenix, unde poți chema un taxi care sosește la destinație fără un șofer la volan, în curând acest spectacol tehnologic își va pierde caracterul de noutate. În doar câțiva ani, prezența vehiculelor autonome pe străzile noastre va deveni o realitate cotidiană.

De asemenea, algoritmii inteligenți au adus inovații semnificative în sectorul financiar. Modelele de învățare automată pot analiza și interpreta mari cantități de date financiare pentru a identifica tipare și tendințe, putând fi, astfel, utilizate pentru gestionarea riscurilor, detecția fraudelor bancare și pentru a oferi recomandări personalizate de investiții [139].

Asistenții virtuali alimentați de inteligența artificială, precum Siri, Alexa, Google Assistant, Bing, Perplexity sau ChatGPT, au devenit parte integrantă a vieții noastre de zi cu zi. Aceștia sunt capabili să răspundă la întrebări, pot executa sarcini simple și, unii dintre ei, oferă asistență vocală interactivă, ajutând utilizatorii să rezolve probleme complexe [150; 126].

Inteligența artificială are potențialul de a transforma modul în care învățăm lucruri noi și maniera în care poate decurge actul didactic. Sistemele de învățare adaptivă pot adapta conținutul educațional în funcție de nevoile individuale ale elevilor și studenților, oferindu-le un mediu de învățare personalizat [162]. Cel mai la îndemână exemplu este modul în care învățăm astăzi o limbă străină folosind doar telefonul din buzunar. Aplicațiile moderne nu ne mai pun să memorăm liste lungi de cuvinte la întâmplare, ci folosesc algoritmi care observă exact ce termeni uităm mai repede. Este ca și cum am avea un mentor răbdător, care nu obosește niciodată și care își amintește perfect fiecare greșeală pe care am făcut-o, transformând studiul dintr-o povară plictisitoare într-un parcurs personalizat, adaptat ritmului propriu de memorare. De asemenea, inteligența artificială poate fi utilizată pentru a analiza performanța elevilor și studenților, pentru a oferi feedback și a dezvolta programe educaționale mai eficiente [139].

Utilizarea inteligenței artificiale în procesele de producție și logistică poate aduce îmbunătățiri semnificative în eficiență și optimizare. Sistemele de monitorizare bazate pe inteligența artificială pot detecta și preveni defectele în timp real, reducând costurile și timpul de producție. Algoritmii de optimizare pot fi utilizați pentru a gestiona lanțurile de aprovizionare și pentru a planifica eficient rutele de transport, contribuind la reducerea costurilor și la creșterea productivității [126].

Inteligența artificială a început să pătrundă și în domeniul artistic, generând creații „originale” în diverse medii, precum design, muzică, cinematografie, pictură și literatură [145]. Algoritmii de învățare automată au capacitatea de a concepe lucrări artistice, compoziții muzicale și povestiri bazate pe modele și stiluri existente, aducând noi perspective și provocări în lumea artistică [126; 143]. Având o experiență directă în domeniul designului și al muzicii, utilizarea pe scară largă a tehnologiei în artă îmi stârnește unele dintre cele mai profunde temeri și sentimente contradictorii. Dincolo de potențialul său considerabil, inteligența artificială pare să atingă ceea ce este adesea considerat ultimul bastion al umanității, și anume creativitatea autentică. Există riscul ca, prin delegarea actului artistic către algoritmi, să diminuăm tocmai acea scânteie care ne definește ca oameni. Vedem deja cum platforme precum Midjourney sau DALL-E pot genera, în câteva secunde, imagini de o complexitate vizuală uluitoare care au început să câștige concursuri de artă digitală, lăsând în urmă artiști care au studiat o viață întreagă tehnica picturii.

În muzică, fenomenul este la fel de prezent. Algoritmii pot acum „compune” simfonii noi în stilul lui Beethoven sau piese jazz care imită perfect improvizația lui Miles Davis, analizând tiparele matematice din spatele geniului lor. Mai mult, în literatură, modelele de limbaj pot scrie scenarii de film sau poezii care, deși corecte estetic, ridică o întrebare tulburătoare: poate o mașină care nu a simțit niciodată durerea, iubirea sau frica să creeze cu adevărat artă, sau asistăm doar la o simulare extrem de sofisticată a emoțiilor umane arhivate în baze de date? Fără îndoială, arta generată de inteligența artificială ne forțează să redefinim ce anume face ca o operă să fie valoroasă, efortul și trăirea artistului sau doar impactul vizual și auditiv asupra publicului.

În fața amenințărilor cibernetice tot mai sofisticate, inteligența artificială joacă un rol crucial în detectarea și prevenirea atacurilor cibernetice. Algoritmi de învățare automată pot analiza modele de trafic și comportamentul utilizatorilor pentru a identifica activități suspecte și a preveni atacurile cibernetice. Pe măsură ce tehnologia continuă să evolueze, este de așteptat să asistăm la extinderea și diversificarea aplicațiilor inteligenței artificiale în diferite industrii și domenii, aducând cu sine noi oportunități, dar și noi provocări. Totodată, este important să luăm în considerare implicațiile etice și sociale ale utilizării inteligenței artificiale, pentru a asigura un impact pozitiv și responsabil al acestei tehnologii în societate [148; 153; 155].

1.2 Cadrul științific și stadiul actual al cercetărilor

Stadiul actual al cercetărilor în domeniul inteligenței artificiale prezintă o panoramă vastă și în continuă schimbare. Progresele semnificative în învățarea automată *(machine learning)*, învățarea profundă cu rețelele neuronale artificale *(deep learning)*, procesarea limbajului natural *(natural language processing)*, vederea artificială *(computer vision)* și roboții autonomi sunt doar câteva exemple ale direcțiilor explorate.[2] În acest moment, pentru a răspunde provocărilor generate de evoluția și utilizarea responsabilă a inteligenței artificiale, cercetarea continuă sub diverse aspecte, cum ar fi explicabilitatea sistemelor de inteligență artificială, etica în inteligența artificială și interacțiunea om-mașină [168].

Sub aspectul abordării transdisciplinare a inteligenței artificiale, se poate constata că, cel puțin în momentul scrierii acestor rânduri, există extrem de puține studii. Una dintre explicații ar putea fi confuzia dintre transdisciplinaritate și multi/interdisciplinaritate. Prezenta lucrare își propune să contribuie la completarea acestei lacune din literatura de specialitate.

În *„Artificial Intelligence. A Modern Approach"* (Inteligența Artificială. O abordare modernă, 2021) [168], lucrare elaborată de o echipă interdisciplinară condusă de Peter Norvig și Stuart Russell, lucrare devenită deja o referință aproape obligatorie în orice bibliografie pe acest subiect, găsim o scurtă, dar elocventă, descriere a interpretărilor inteligenței artificiale din mai multe perspective filosofice. Autorii analizează răspunsuri la întrebările filosofice esențiale legate de minte, inteligență, cunoaștere și acțiune, explorând diverse aspecte filosofice provocatoare și oferind, în acest mod, o perspectivă asupra evoluției ideilor și teoriilor care abordează, direct sau indirect, domeniul inteligenței artificiale.

Lucrarea debutează prin a investiga interogații de natură științifică și filosofică cu privire la fezabilitatea aplicării regulilor formale în scopul deducerii concluziilor valide. Autorii

[2] Deoarece inteligența artificială evoluează rapid, mulți termeni sunt folosiți direct în engleză sau într-o formă hibridă, ceea ce poate crea confuzii. Pentru a pune bazele unei comunicări coerente, prezentăm în continuare o adaptare precisă a terminologiei esențiale în limba română.

- *Machine Learning* (Învățare Automată). Acesta este fundamentul inteligenței artificiale moderne. În loc să fie programat cu reguli rigide (dacă A, atunci B), sistemul învață singur tipare din date. Este capacitatea unui algoritm de a-și îmbunătăți performanța pe măsură ce este expus la mai multe date.
- *Deep Learning* (Învățare Profundă). Este o subcategorie a învățării automate care folosește arhitecturi inspirate de structura creierului uman. Utilizarea unor rețele neuronale artificiale cu foarte multe straturi (de aici termenul „profund/deep"). Acestea pot procesa date complexe precum imagini sau sunete.
- *Natural Language Processing* / NLP (Procesarea Limbajului Natural). Aceasta este tehnologia din spatele chatbot-urilor sau a traducerilor automate. Este abilitatea computerului de a „înțelege", interpreta și genera limbaj uman (scris sau vorbit) într-un mod coerent.
- *Computer Vision* (Vedere Artificială sau Viziune computerizată). Dacă NLP este „urechea și vocea", *Computer Vision* este „ochiul". Este tehnologia care permite sistemelor informatice să extragă informații semnificative din imagini digitale, clipuri video sau alte intrări vizuale.

abordează, de asemenea, aspecte referitoare la emergența conștiinței, precum și problema originii procesului de acumulare a cunoașterii și impactul acestuia asupra acțiunii umane.

Sunt menționate contribuțiile unor gânditori ca Aristotel [6], Ramon Llull, Leonardo da Vinci, Gottfried Wilhelm Leibniz [69] și René Descartes [26] în dezvoltarea unor concepte și dispozitive mecanice folosite în elaborarea de raționamente și calcule și se analizează, în punctele esențiale, dualismul și materialismul, ca perspective alternative atunci când vorbim despre relația dintre minte și materie. O implicație a materialismul este că operațiunile creierului fizic constituie mintea, în timp ce dintr-o perspectivă dualistă, există o parte a minții care, pentru că este în afara naturii, se sustrage, astfel, legilor fizice.

Lucrarea analizează, de asemenea, contribuțiile mișcării empiriste și rolul raționamentului inductiv în procesul de luare a deciziilor, evidențiind aportul adus de Francis Bacon [8], John Locke [73] și David Hume [52]. Sunt prezentate și ideile cercului de la Viena (pozitivismul logic) [118], care a combinat raționalismul și empirismul, analizând modul în care cunoașterea este dobândită prin cuantificarea gradului de credință acordat propozițiilor logice, pe baza conexiunii acestor propoziții cu observațiile.

Un alt aspect discutat în aceeași lucrare este legătura dintre cunoaștere și acțiune. Se evidențiază importanța înțelegerii modului în care acțiunile sunt justificate în construirea unui agent ai cărui acțiuni sunt raționale. Sunt prezentate ideile lui Aristotel despre justificarea acțiunilor prin conexiunea logică dintre scopuri și cunoașterea rezultatelor acțiunii [2].

Evaluarea istorică a impactului în filosofie se încheie prin abordarea conceptului de utilitate și prin analiza raționamentului decizional în contextul acțiunilor care urmăresc atingerea anumitor obiective. Se discută și despre utilitarism, care se bazează pe maximizarea utilității în luarea deciziilor raționale, și, în aceeași manieră, despre etica deontologică, ca bază a respectării regulilor sociale universale în determinarea acțiunilor corecte [168].

Una dintre puținele lucrări care folosește perspectiva filosofică transdisciplinară în vederea înțelegerii inteligenței artificiale este articolul lui Adrian Iosif „Inteligența artificială și mintea umană. Aspecte inter- și transdisciplinare", publicat în 2006, în volumul colectiv Perspective românești asupra științei și teologiei [56]. Textul abordează diferite aspecte legate de percepția distorsionată a inteligenței artificiale și impactul său asupra concepției umane despre sine și despre divinitate. Se afirmă că reducerea ființei umane la funcțiile unui angrenaj pur tehnic și considerarea inteligenței artificiale drept un substitut al creierului uman conduce la o concepție fizicalistă limitată a umanității. În plus, se menționează că tehnologiile, inclusiv inteligența artificială, nu pot contribui decisiv la un salt ontologic al ființei umane, deși constituie un progres tehnologic evident al umanității.

De asemenea, autorul subliniază că inteligența artificială este doar o proiecție bidimensională a inteligenței umane și că mintea umană nu se limitează doar la raționamentul deductiv. În acest punct, dezvoltând ideea autorului, adăugăm că nu doar raționamentul deductiv, dar nici cel inductiv, pe larg folosit în inteligența artificială modernă, nu descrie în totalitate caracteristicile inteligenței umane. Tocmai de aceea, este nevoie de o modelare a teoriei abducției care să fie aplicată în domeniul inteligenței artificiale, pentru a putea realiza progrese semnificative în apropierea de ceea ce numim inteligența generală umană. Acest subiect va fi pe larg discutat în capitolele următoare din această lucrare.

Deși anumite activități ale creierului uman pot fi transferate pe o mașină (de obicei, activități rutiniere), inteligența umană, în esența sa, nu poate fi replicată. În textul citat, se evidențiază că, fără a fi interpretat, un șir de simboluri rezultat dintr-un calcul este lipsit de conținut și informația pe care o conține nu are relevanță. Astfel, pentru a putea extrage mesajul inteligent dintr-un șir de simboluri, un subiect cu o minte și o inteligență are nevoie de atribute mai complexe decât simpla calculabilitate.

În acest sens, autorul afirmă că inteligența umană nu poate fi replicată în întregime. Obiectivele inteligenței artificiale „tari" sunt considerate comparabile cu construirea unui perpetuum mobile. Putem remarca faptul că din momentul scrierii articolului (2006) și până în prezent progresele inteligenței artificiale sunt remarcabile, da nu atât de extraordinare încât punctele de bază ale studiului mai sus citat să fie anulate. Este clar că încă nu avem, și nici nu știm exact când vom avea sau dacă vom avea, o inteligență artificială de tip „tare" *(Artificial General Intelligence)*.[3] Nu lipsit de ironie, savantul britanic John Lennox [154] afirmă că, de fiecare dată când încercăm să facem predicții legate de evoluția inteligenței artificiale generale, avem impresia că vom avea nevoie de încă 50 de ani până la apariția acesteia.

Textul lansează întrebarea dacă diferitele concepții asupra inteligenței umane pot fi aduse într-un dialog fertil și, de asemenea, cum poate fi explicată discrepanța dintre abordarea „slabă" și „tare" a inteligenței artificiale. Se sugerează că abordarea transdisciplinară a naturii și cunoașterii, propusă de Basarab Nicolescu [91], poate oferi o perspectivă mai integrată. Transdisciplinaritatea postulează existența mai multor niveluri de realitate și, în consecință, se poate admite că fiecare teorie despre inteligența artificială corespunde unui anumit nivel diferit de realitate. Prin aplicarea

[3] AGI (*Artificial General Intelligence* / Inteligență Artificială Generală): Reprezintă un sistem ipotetic capabil să înțeleagă, să învețe și să aplice cunoștințe în orice domeniu intelectual, la un nivel egal sau superior omului. Spre deosebire de „IA îngustă" sau „slabă" (*Narrow AI*), care excelează în sarcini specifice, AGI posedă capacitatea de a rezolva probleme noi, neprevăzute, fără a fi fost antrenată special pentru acestea.

modelului transdisciplinar, se argumentează că diferitele aspecte ale fenomenelor mentale pot fi manifestări ale aceleiași minți umane, dar pe diferite niveluri de realitate.

În contextul metodologiei transdisciplinare, autorul dezvoltă o analogie între modelul teologic al ierarhiilor elaborat de Sf. Dionisie Areopagitul și modelul transdisciplinar al nivelurilor de realitate. Se menționează că ierarhiile lui Dionisie se „luminează" reciproc și că nu există un nivel de realitate privilegiat care să poată cuprinde și înțelege toate celelalte niveluri. Se subliniază că, conform metodologiei transdisciplinare, există o zonă de nonrezistență la experiențe și descrieri matematice, iar această zonă se extinde în afara nivelurilor de realitate. În acest sens, între ierarhiile lui Dionisie există o diferență axiologică, în timp ce nivelurile de realitate nu au o structură piramidală. Conceptul transdisciplinar de nonrezistență [91, 92] va fi dezvoltat pe larg în cuprinsul acestei lucrări.

Același autor observă că există similitudini transdisciplinare și cu descrierea făcută de Sf. Maxim Mărturisitorul în Mystagogia V. Urmând același raționament care decurge din axiomele transdisciplinarității, se poate considera că fiecare etapă a procesului de căutare a adevărului și a binelui poate fi identificată ca o trecere de la un nivel de realitate la altul. Așadar, fluxul coerent de informație care străbate diferitele niveluri de realitate corespunde unui flux de conștiință prin diferitele niveluri de percepție. Se subliniază că înțelegerea fenomenelor mentale și dezvoltarea unei noi forme de inteligență necesită o cunoaștere transdisciplinară în deplină armonie cu tradiția spirituală a umanității.

În concluziile studiului citat, se menționează că cunoașterea transdisciplinară revalorizează poeticitatea lumii și reintegrează uimirea ca mediu fertil pentru a percepe ceea ce nu poate fi perceput. În acest mod, abordarea transdisciplinară poate ajuta la distingerea esențialului de accesoriu și a raționalului de raționalizabil în complexitatea fenomenelor mentale. Prin înțelegerea și afirmarea unui nou tip de inteligență, cunoașterea transdisciplinară poate contribui la o mai bună înțelegere a minții și la descoperirea profunzimilor fenomenelor umane.

Cadrul științifice

Inteligența artificială este un domeniu complex și vast, care se bazează pe o serie de fundamente științifice interdisciplinare (***Fig. 1. 1***). Aceste fundamente includ matematica, fizica, biologia, știința computerelor, algoritmii, logica matematică și filosofică, neuroștiințele și științele cognitive, statistica, probabilitățile, limbajele formale și limbajele de programare [127]. Prin combinarea cunoștințelor și tehnologiilor din aceste domenii, inteligența artificială continuă să avanseze și să aducă inovații semnificative în societatea noastră. În continuare vom explora aceste fundamente și vom analiza exemple concrete și aplicații relevante în fiecare domeniu [168; 161].

Fără îndoială, **matematica** este o piatră de temelie a inteligenței artificiale și este utilizată într-o varietate de moduri. Conceptele matematice precum algebra liniară, calculele probabilistice și teoria grafurilor sunt esențiale în dezvoltarea algoritmilor și a modelelor de inteligență artificială [162]. De exemplu, rețelele neuronale artificiale, care imită structura creierului uman, utilizează operații matematice precum calculul matricial, operații de convoluție și funcțiile non-liniare pentru a procesa informații și a genera predicții precise [143].

Un exemplu concret de aplicare a operației de convoluție în inteligența artificială este utilizarea rețelelor neuronale convoluționale (*Convolutional Neural Networks* - CNN) pentru recunoașterea imaginilor. Aceste rețele utilizează filtre convoluționale care aplică principiile matematice ale convoluției și ale transformatei Fourier pentru a extrage caracteristici importante din imaginile de intrare [126; 31].

Principiul de bază al convoluției este analog cu fenomenul fizic al filtrării și interacțiunii undelor. Filtrarea cu ajutorul filtrelor convoluționale în rețelele neuronale convoluționale implică aplicarea unui set de matrici peste imaginea de intrare, pentru a detecta diferite tipuri de caracteristici, cum ar fi marginile, texturile sau obiectele [139]. Aceste filtre numerice sunt concepute pentru a identifica șabloane sau structuri specifice în imagini.

De exemplu, pentru recunoașterea fețelor, rețelele neuronale convoluționale pot utiliza filtre pentru a detecta marginile și contururile fețelor, linii importante, ochii, nasul sau gura. Filtrele convoluționale sunt aplicate pe straturile rețelei pentru a extrage și rafina caracteristicile, iar apoi aceste caracteristici sunt utilizate pentru clasificarea imaginilor în clase specifice (de exemplu, „față” sau „non-față”). Aplicarea principiilor fizice ale convoluției și transformatei Fourier în rețelele neuronale convoluționale permite extragerea eficientă a caracteristicilor cheie din imagini și obținerea unor performanțe ridicate în recunoașterea și clasificarea imaginilor [161].

Înțelegerea semnalelor și procesarea semnalului se bazează pe principii fizice și matematice, cum ar fi transformata Fourier sau teoria informației. Aceasta permite analiza și interpretarea semnalelor (de exemplu, semnale audio sau semnale biofizice) pentru a extrage informații relevante [133].

Problemele de optimizare și modelare matematică din inteligența artificială pot implica rezolvarea sistemelor de ecuații liniare. Algebra liniară furnizează metode și tehnici pentru rezolvarea acestor sisteme, inclusiv prin utilizarea matricelor și vectorilor. Transformările liniare, de genul rotației, scalării și translației, sunt utilizate în algoritmii de prelucrare a imaginilor și în reprezentarea datelor în spații multidimensionale [143].

Calculul diferențial și integral este folosit pentru a optimiza funcțiile de cost și a calcula gradienții în metodele de învățare automată. Analiza matematică este utilă în evaluarea și

înțelegerea performanței algoritmilor de inteligență artificială și a modelelor de învățare automată. De asemenea, analiza matematică furnizează un cadru pentru studiul funcțiilor și a derivatelor lor. Aceste concepte sunt utilizate în algoritmi de învățare automată, cum ar fi regresia și clasificarea, care se bazează pe ajustarea funcțiilor astfel încât acestea să se potrivească cât mai exact cu datele de intrare.

Înainte de a aplica algoritmii de învățare automată datele trebuie să fie prelucrate și transformate într-un format adecvat. În acest scop, sunt necesare operații de normalizare a datelor, extragerea de caracteristici și reducerea dimensionalității acestora. Analiza componentelor principale (*Principal Component Analysis*) este o tehnică foarte importantă utilizată în inteligența artificială pentru reducerea dimensionalității datelor. Aceasta se bazează pe concepte din algebra liniară, precum valorile și vectorii proprii, pentru a identifica componentele principale ale unui set de date [162].

Problemele de optimizare, cele mai multe dintre ele implicând găsirea valorilor optime ale unor variabile, sunt foarte frecvente în inteligența artificială. Analiza matematică oferă metode și tehnici pentru rezolvarea acestor probleme (de exemplu, algoritmi de gradient și metoda Lagrange). Într-o manieră asemănătoare, analiza matematică include studiul seriilor și integralelor, care sunt folosite în algoritmi de învățare automată (rețelele neuronale constituie aici un bun exemplu). Dezvoltarea în serii de puteri este folosită pe larg pentru aproximarea și învățarea funcțiilor complexe [133].

Teoria grafurilor este un domeniu matematic care studiază structurile de rețele compuse din noduri și legături (muchii) între ele, fiind utilizată în probleme de reprezentare și analiză a relațiilor dintre entități. Teoria grafurilor are o aplicare largă în inteligența artificială, în special în problemele de reprezentare a datelor și în algoritmi de căutare și de învățare automată. Graful este foarte folosit pentru a reprezenta o structură de interconectare a neuronilor în rețelele neuronale artificiale sau pentru a modela relațiile dintre entitățile dintr-un sistem.

Un exemplu aplicativ al teoriei grafurilor în inteligența artificială este utilizarea grafurilor din rețelele de socializare pentru a genera recomandări personalizate. În acest caz, nodurile grafului reprezintă utilizatorii și legăturile dintre ei reprezintă tipuri relații sociale, cum ar fi prietenii și conexiunile dintre ei pe rețele de socializare. Grafurile sociale pot fi folosite pentru a identifica relații și influențe sociale între utilizatori și pentru a face recomandări bazate pe aceste informații. Să presupunem că avem un graf social în care fiecare nod reprezintă un utilizator și fiecare muchie reprezintă o conexiune între utilizatori (de exemplu, prietenie). Acest tip de grafuri sociale poate fi utilizat pentru a identifica grupuri de utilizatori strâns conectați și pentru a găsi conexiuni între grupuri [133]. Să presupunem că utilizatorul A este conectat direct cu utilizatorul B și cu

utilizatorul C, iar utilizatorul B este conectat cu utilizatorul D. Bazându-ne pe aceste conexiuni sociale, putem utiliza algoritmi de grafuri pentru a identifica utilizatori care sunt prieteni ai prietenilor utilizatorului A, dar nu sunt direct conectați cu A. Acest lucru ar putea sugera potențiali noi prieteni care ar putea fi recomandați utilizatorului A (un algoritm similar este folosit și în rețeaua Facebook).

În plus, grafurile sociale pot fi folosite și pentru a identifica grupuri de utilizatori cu interese comune sau cu profiluri similare, ceea ce poate duce la recomandări mai precise de produse, servicii sau de conținut. Algoritmii de învățare automată pot fi aplicați pe grafuri sociale pentru a identifica modele de comportament și preferințe în funcție de conexiunile sociale și pentru a face recomandări personalizate pe baza acestor modele.

De asemenea, **principiile fizicii** sunt relevante în domeniul inteligenței artificiale, în special în legătură cu roboții și sistemele autonome. Roboții inteligenți se bazează pe principiile staticii, cinematicii și dinamicii pentru a se deplasa și interacționa cu mediul înconjurător. Nu în ultimul rând, principiile termodinamicii sunt foarte importante în proiectarea unor sisteme eficiente din punct de vedere energetic. De exemplu, roboții care explorează spațiul necesită o planificare inteligentă a rutelor și optimizarea consumului de energie pentru a-și îndeplini obiectivele cu succes. Fizica joacă un rol important în inteligența artificială, contribuind la dezvoltarea algoritmilor, simulărilor și înțelegerii sistemelor dinamice [168].

Fizica este utilizată pentru a modela și simula fenomenele din lumea reală într-un mediu virtual. Modelele fizice permit obținerea unor rezultate coerente și realiste în simulări, precum simulările de mișcare, coliziuni, gravitație, curgerea fluidelor sau comportamentul materiei.

În inteligența artificială, fizica este utilizată, la fel pentru a modela și controla roboții și sistemele robotice. Aceasta implică simularea mișcării și interacțiunii cu mediul, precum și aplicarea principiilor mecanicii, electricității și controlului pentru a dezvolta roboți capabili să efectueze sarcini complexe [139].

Principiile fizice, cum ar fi optică și lumina, sunt aplicate în algoritmii de reprezentare și procesare a imaginilor. Cunoștințele despre formarea imaginilor, filtrare și segmentare pot ajuta la recunoașterea obiectelor, detectarea fețelor și analiza conținutului vizual [138].

În inteligența artificială, datele colectate de la diferiți senzori, cum ar fi camere video, microfoane sau senzori de mișcare, nu pot fi interpretate și analizate corect fără a ține cont de principiile fizice. Aceasta implică extragerea informațiilor relevante din semnale și aplicarea modelelor și teoriilor fizice pentru a înțelege mediul și evenimentele din jur [139].

Principiile fizice sunt utilizate și pentru a dezvolta modele de învățare automată care înregistrează și analizează eficient comportamentul fizic al sistemelor. De exemplu, în problemele

de simulare a dinamicii corpului deformabil sau a fluidelor se pot folosi rețele neuronale pentru a aproxima ecuațiile fizice subiacente și pentru a obține simulări mai rapide și mai precise [168].

Adesea, principiile fizice sunt integrate direct în arhitectura rețelelor neuronale pentru a le conferi cunoștințe și abilități fizice. De exemplu, în rețelele neuronale de simulare a mișcării sau de robotica virtuală, se definesc variabile din fizică, precum forțele și constrângerile în procesul de învățare, permițând rețelelor să genereze mișcări și interacțiuni realiste [143].

În procesarea limbajului natural (*Natural Language Processing*), fizica poate fi aplicată pentru a construi modele care sesizează semantica și relațiile dintre cuvinte și fraze. Aceste modele pot fi utilizate pentru rezolvarea sarcinilor de înțelegere a limbajului natural, cum ar fi analiza sentimentelor, traducerea automată sau răspunsurile la întrebări. Într-o manieră asemănătoare, principiile fizice pot fi utilizate pentru a modela coerența și coeziunea discursului în sistemele de generare automată a limbajului. De exemplu, pot fi integrate constrângeri fizice pentru a asigura cursivitatea și coerența logică a textului generat [150].

Chiar dacă nu întotdeauna în mod direct, **biologia** oferă o sursă importantă de inspirație pentru inteligența artificială. Algoritmii genetici și evolutivi se bazează pe procesele de selecție naturală și evoluție pentru a găsi soluții optime la probleme complexe. De asemenea, învățarea automată se inspiră din modul în care creierul uman procesează și analizează informațiile. Studiul sistemelor biologice, cum ar fi neuronii și rețelele neuronale, contribuie la dezvoltarea de modele de inteligență artificială mai avansate și mai eficiente. *Swarm intelligence* (inteligența stolurilor)[4] este un domeniu al inteligenței artificiale inspirat de comportamentul sistemelor naturale, biologice, cum ar fi furnicile, albinele sau păsările. Acest concept se bazează pe ideea că entități simple, care interacționează și comunică între ele, pot genera comportamente complexe și soluții eficiente la problemele complexe. În contextul inteligenței artificiale, *swarm intelligence* se referă la utilizarea modelelor și algoritmilor inspirați de comportamentul stolurilor în dezvoltarea sistemelor inteligente. Acestea pot fi aplicate într-o varietate de domenii, cum ar fi optimizarea, roboții autonomi, rețelele de senzori și altele [152]. De exemplu, inspirat de modul în care furnicile găsesc cel mai scurt drum către hrană, companiile de logistică și curierat folosesc aplicații pentru a planifica traseele camioanelor în timp real, evitând congestionările și reducând consumul de combustibil.

Este evident că **știința computerelor** rămâne disciplina centrală în dezvoltarea inteligenței artificiale. Programarea este esențială pentru a implementa algoritmi și modele de inteligență artificială care să rezolve probleme specifice. Limbajele de programare precum *C++*, *Python, R*

[4] Conceptul de *Swarm Intelligence* (SI), poate fi tradus în română ca *inteligență colectivă* sau, mai plastic, *inteligență de stol/roi*.

și *Java* sunt folosite pentru a implementa și a rula aplicații de inteligență artificială. De asemenea, domenii precum teoria informației, teoria limbajelor formale și teoria algoritmilor sunt fundamentale în dezvoltarea de algoritmi eficienți și în optimizarea performanței sistemelor de inteligență artificiale. În plus, arhitecturile hardware specializate, precum unitățile de procesare grafică (GPU) și circuitele integrate specifice aplicațiilor (ASICs), sunt utilizate pentru a accelera calculele necesare în aplicațiile de inteligență artificială [133].

Teoria limbajelor formale joacă un rol esențial în dezvoltarea inteligenței artificiale prin furnizarea unui cadru matematic pentru reprezentarea și analiza limbajului. Aceasta se aplică în mai multe domenii ale inteligenței artificiale, inclusiv în prelucrarea limbajului natural și în dezvoltarea limbajelor de programare [168].

Automatele finite și gramaticile formale sunt utilizate pentru a descrie și a recunoaște limbaje formale. În inteligența artificială, acestea sunt aplicate în prelucrarea limbajului natural pentru a analiza și a genera texte. Automatele finite pot fi folosite pentru a modela și a recunoaște structurile gramaticale ale limbajului, cum ar fi analiza sintactică a propozițiilor. Gramaticile formale, cum ar fi gramaticile *context-free*, sunt utilizate pentru a defini reguli de generare a limbajelor și a structurilor lor [133].

Un exemplu foarte răspândit de aplicare a teoriei limbajelor formale în inteligența artificială sunt sistemele de recunoaștere vocală. Sistemul trebuie să recunoască cuvintele și structura gramaticală a unei propoziții date. Pentru aceasta, automatul finit este utilizat pentru a recunoaște secvențele valide de cuvinte, iar gramatica formală pentru a defini regulile de generare a propozițiilor corecte din punct de vedere gramatical.

Expresiile regulate sunt utilizate pentru a defini șabloane de căutare în texte. Acestea permit identificarea de modele sau secvențe de caractere într-un text. În inteligența artificială, expresiile regulate sunt aplicate în prelucrarea limbajului natural pentru a efectua căutări și filtrări de informații relevante din texte. Un exemplu de aplicare a expresiilor regulate în inteligența artificială este un sistem de extragere a informațiilor din texte. Sistemul poate utiliza expresii regulate pentru a căuta entități specifice, cum ar fi adrese de e-mail, numere de telefon sau date, într-un text de dimensiune mare. Astfel, teoria limbajelor formale este utilizată pentru a defini șabloanele căutate și pentru a identifica aceste entități în mod automat [150].

Teoria limbajelor formale este aplicată și în dezvoltarea limbajelor de programare. Sintaxa limbajelor de programare definește regulile și structura corectă a programelor, în timp ce semantica limbajelor de programare definește înțelesul și comportamentul programelor. Un exemplu de aplicare a teoriei limbajelor formale în inteligența artificială este proiectarea unui compilator sau a unui interpretor pentru un limbaj de programare. În acest caz, teoria limbajelor formale este

folosită pentru a defini gramatica și sintaxa limbajelor de programare, precum și pentru a realiza analiza sintactică și semantică a codului sursă. Aceasta permite validarea și interpretarea corectă a programelor scrise în acel limbaj, facilitând dezvoltarea de software și algoritmi în domeniul inteligenței artificiale [161].

Un exemplu simplu ar fi limbajul de programare *Python,* cel mai popular limbaj de programare din momentul scrierii acestei lucrări (conform indexului TIOBE). Python este, în acest moment, cel mai folosit limbaj în modelarea sistemelor care folosesc inteligența artificială. Teoria limbajelor formale este utilizată pentru a defini sintaxa limbajului *Python,* inclusiv structura corectă a instrucțiunilor, expresiilor și funcțiilor. Prin aplicarea regulilor gramaticale și a regulilor de precedență, interpretorul de *Python* poate valida și interpreta codul sursă *Python* într-un mod coerent și eficient [162].

De asemenea, teoria limbajelor formale poate fi utilizată pentru a extinde limbajele de programare existente sau pentru a dezvolta limbaje specifice unui domeniu în cadrul inteligenței artificiale, care să ofere expresivitate și capacitatea de abstractizare mai mare în rezolvarea problemelor specifice.

Așa dar, teoria limbajelor formale este esențială în inteligența artificială pentru reprezentarea și analiza limbajului, precum și pentru dezvoltarea de sisteme de prelucrare a limbajului natural și de limbaje de programare. Prin aplicarea acestor teorii, inteligența artificială poate înțelege, genera și manipula informații în limbaj uman și poate dezvolta algoritmi și software eficient în rezolvarea problemelor specifice [168].

Calculul computațional și **teoria complexității** calculului sunt două concepte fundamentale în domeniul inteligenței artificiale. Acestea se aplică în analiza și proiectarea algoritmilor, evaluarea eficienței computaționale și înțelegerea limitărilor resurselor necesare pentru a rezolva problemele în inteligența artificială. Calculul computațional este utilizat în dezvoltarea și antrenarea algoritmilor de învățare automată, cum ar fi rețelele neuronale. Acești algoritmi necesită operații matematice intensive, cum ar fi înmulțiri de matrice și calcule de derivate, care pot fi realizate eficient cu ajutorul metodelor de calcul computațional.

Teoria complexității calculului oferă un cadru pentru evaluarea eficienței algoritmilor din punct de vedere al resurselor necesare, cum ar fi timpul de execuție și spațiul de memorie. Aceasta ajută la identificarea algoritmilor care pot rezolva problemele într-un timp rezonabil și cu resurse acceptabile. Teoria complexității calculului identifică clase de probleme, cum ar fi problemele NP-complete, care sunt considerate a fi extrem de dificile din punct de vedere al rezolvării. Aceste probleme necesită algoritmi cu o complexitate exponențială, ceea ce limitează aplicabilitatea lor practică în inteligența artificială. Teoria complexității calculului se aplică și în înțelegerea

limitărilor învățării automatizate. De exemplu, se studiază complexitatea generalizării modelelor de învățare automată și relația între dimensiunea setului de date și performanța algoritmului [169].

Logica matematică joacă un rol important în dezvoltarea inteligenței artificiale, în special atunci când vorbim despre raționament și luarea de decizii. Logica matematică este utilizată pentru a construi sisteme formale de reprezentare a cunoștințelor și a raționamentului. Ea stă la baza sistemelor expert și a raționamentului automat în inteligența artificială. Logica simbolică și raționamentul formal sunt adesea folosite pentru a lua decizii raționale și pentru a trage concluzii logice. Logica formală joacă un rol important în dezvoltarea inteligenței artificiale prin furnizarea unui cadru formal pentru reprezentarea cunoștințelor și pentru raționamentul automat [133].

Astfel, logica propozițională este utilizată pentru a reprezenta și a evalua propoziții simple. În inteligența artificială, calculul propozițional poate fi folosit în sistemele de raționament automat pentru a lua decizii bazate pe reguli și informații cunoscute. Un exemplu ar fi un sistem expert în diagnosticarea medicală, care utilizează reguli logice pentru a evalua simptomele unui pacient și a genera un diagnostic [168].

Să luăm în considerare următoarea regulă logică:

Dacă pacientul are febră și tuse, atunci există șanse mari să aibă gripă.

Folosind logica propozițională, putem reprezenta această regulă într-o formă logică în felul următor:

P = Pacientul are febră.

T = Pacientul are tuse.

G = Pacientul are gripă.

(P ȘI T) → G

Această reprezentare logică permite sistemului de inteligență artificială să evalueze condițiile și să tragă concluzii. Dacă pacientul are febră și tuse, sistemul va concluziona că există șanse mari ca pacientul să aibă gripă.

Logica predicatelor de ordinul doi extinde logica propozițională prin adăugarea de variabile și cuantificatori. Aceasta permite reprezentarea și raționamentul asupra entităților și relațiilor dintre ele într-un mod mai expresiv. În inteligența artificială, logica predicatelor de ordinul doi poate fi utilizată pentru a modela situații complexe și a face raționamente mai sofisticate [139].

Un exemplu de aplicare a logicii predicatelor de ordinul doi în inteligența artificială sunt sistemele de reprezentare a cunoștințelor și de înțelegerea raționamentului natural. Aceste sisteme pot utiliza logica predicatelor de ordinul doi pentru a reprezenta informații despre obiecte, proprietăți și relațiile dintre ele și pentru a face inferențe logice.

Să luăm în considerare următoarea reprezentare logică a unei reguli:

P(x) = x este o persoană.

M(x, y) = x este mama lui y.

T(x) = x este tatăl.

$\forall x [P(x) \rightarrow (\exists y [P(y) \wedge M(y, x)] \vee T(x))]$

Această regulă afirmă că fiecare persoană are fie o mamă, fie un tată. Aceasta implică o cuantificare universală (∀x) pentru a se aplica la toate obiectele de tipul „persoană” și cuantificatori existențiali (∃y) pentru a găsi existența unei mame sau a unui tată potrivit pentru fiecare persoană [145].

Prin urmare, logica predicatelor de ordinul doi permite dezvoltarea unor sisteme de inteligență artificiale mai puternice și capabile să facă raționamente complexe și să reprezinte cunoștințe mai sofisticate despre lume [139]. Logica formală oferă un cadru riguros și expresiv pentru reprezentarea și raționamentul cunoștințelor în sistemele de inteligență artificială.

Dar, logica formală este afectată de consecințele teoremei lui Gödel. Teorema lui Gödel, denumită și teorema incompletitudinii, este un rezultat fundamental al logicii matematice dezvoltat de Kurt Gödel în anii 1930 [145].

Teorema lui Gödel afirmă că orice sistem formal suficient de puternic care include aritmetica elementară este fie incomplet, fie inconsistent. Aceasta înseamnă că, într-un sistem formal dat, există propoziții adevărate care nu pot fi demonstrate în acel sistem formal. Cu alte cuvinte, există întotdeauna afirmații care sunt adevărate, dar care nu pot fi demonstrate în cadrul acelui sistem [17, p.300].

Această teoremă are implicații profunde pentru logica formală și pentru inteligența artificială bazată pe logica formală. Înseamnă că nu există un sistem formal complet și consistent care să poată rezolva toate problemele matematice sau logice. Vor exista întotdeauna limite și lacune în cadrul acestor sisteme. Deci, teorema lui Gödel subliniază că există limite intrinseci în logica formală și, implicit, în puterea de calcul a sistemelor formale. Aceste limite se aplică și în cazul inteligenței artificiale bazate pe logica formală, care nu poate scăpa de restricțiile teoremei lui Gödel [17].

Există probleme și afirmații pe care inteligența artificială nu le poate rezolva sau demonstra într-un sistem formal dat. Inteligența artificială bazată pe logica formală se va confrunta cu acest tip de limite și nu va putea scăpa de aceste constrângeri. Astfel, chiar dacă inteligența artificială poate rezolva multe probleme matematice, există întotdeauna teoreme care vor rămâne nedemonstrate în cadrul acelui sistem formal [145].

De asemenea, unele paradoxuri logice, precum paradoxul lui Russell[5] sau paradoxul lui Berry[6], creează afirmații care sunt simultan adevărate și false în același sistem formal [17]. Inteligența artificială bazată pe logica formală va avea dificultăți în a gestiona aceste paradoxuri și în a oferi rezolvări coerente.

În aceeași ordine de idei, există probleme de decizie în matematică și în informatică care nu pot fi rezolvate de către inteligența artificială într-un mod complet și corect. Un exemplu este „problemă opririi" (*halting problem*), care constă în a determina dacă un program se va opri sau va continua să ruleze într-un anumit moment. Teorema lui Gödel arată că nu există un algoritm general care să rezolve această problemă pentru toate programele posibile [95].

Așa cum vom vedea și dintr-un studiu de caz, inteligența artificială bazată pe logica formală poate avea dificultăți în interpretarea și receptarea contextului și a intențiilor umane. Logica formală tratează semnificația într-un mod strict și riguros, ceea ce poate fi limitativ atunci când este necesară înțelegerea subtilităților și a nuanțelor în comunicarea umană.

Aceste exemple evidențiază faptul că teorema de incompletitudine a lui Gödel pune limite fundamentale asupra inteligenței artificiale bazate pe logica formală [145; 94]. Totuși, există abordări în cadrul inteligența artificială, cum ar fi rețelele neurale, învățarea automată și învățarea profundă, care pretind că pot depăși anumite limite ale logicii formale. Cât de adevărat este vom putea să constatăm din analiza studiilor de caz.

Etica computațională abordează problemele morale și etice legate de utilizarea inteligenței artificiale în societate și influențele acestora asupra deciziilor și comportamentului. În același timp cu avansarea rapidă a tehnologiei și creșterea utilizării inteligenței artificiale în diverse domenii, este crucial să se examineze implicațiile etice ale acestor sisteme și să se dezvolte principii și norme pentru a asigura utilizarea lor într-un mod responsabil și benefic [141].

Etica computațională se concentrează pe o gamă largă de probleme și provocări etice. Acestea includ, printre altele, confidențialitatea și securitatea datelor, discriminarea algoritmică, transparența și responsabilitatea sistemelor inteligente, impactul asupra locurilor de muncă și distribuția justă a avantajelor tehnologiei [153; 155].

---

[5] Paradoxul lui Russell este un paradox descoperit de filozoful și logicianul Bertrand Russell în 1901, care apare în cadrul teoriei naive a mulțimilor. Paradoxul rezultă din analiza mulțimii tuturor mulțimilor care nu se conțin pe ele însele. Întrebarea devine: această mulțime se conține pe sine sau nu? Dacă se conține, atunci conform definiției nu ar trebui să se conțină; dacă nu se conține, atunci ar trebui să se conțină. Această contradicție a demonstrat că teoria naivă a mulțimilor este inconsistentă și a dus la dezvoltarea unor sisteme axiomatice mai riguroase ale teoriei mulțimilor.

[6] Paradoxul lui Berry este un paradox semantic formulat de matematicianul G. G. Berry la începutul secolului al XX-lea. El apare din expresia: „cel mai mic număr natural care nu poate fi descris în mai puțin de douăzeci de cuvinte". Problema este că această formulare însăși descrie acel număr în mai puțin de douăzeci de cuvinte, generând o contradicție. Paradoxul evidențiază limitele limbajului natural atunci când este folosit pentru a defini riguros concepte matematice și arată dificultățile legate de autoreferențialitate și definibilitate.

Un aspect important al acestei filosofii este identificarea și abordarea sesizărilor și prejudecăților încorporate în algoritmi și sistemele de inteligență artificială. De exemplu, există preocupări legate de discriminarea algoritmică, în care sistemele inteligente pot perpetua și amplifica inegalitățile sociale existente. Aceste probleme necesită luarea în considerare a valorilor și normelor sociale, precum și dezvoltarea de tehnologii și algoritmi imparțiali și etici [155].

Etica computațională își propune, de asemenea, să exploreze aspecte legate de autonomia și responsabilitatea sistemelor inteligente. Sistemele autonome, cum ar fi roboții și mașinile autonome, ridică întrebări referitoare la luarea deciziilor morale și asumarea responsabilității pentru acțiunile lor. Este necesară dezvoltarea unor principii și norme pentru a gestiona aceste aspecte într-un mod etic și compatibil cu valorile umane [161].

În concluzie, etica computațională aduce în prim-plan dezbaterea și analiza problemelor morale și etice asociate cu utilizarea inteligenței artificiale în societate. Prin abordarea acestor aspecte, se încearcă construirea unui cadru etic și responsabil pentru dezvoltarea și implementarea tehnologiilor inteligente, asigurându-se că acestea contribuie la binele comun și respectă valorile și drepturile fundamentale ale indivizilor. Prin urmare, luarea în considerare a perspectivei eticii în proiectarea inteligenței artificiale este esențială pentru a ne asigura că dezvoltarea tehnologică este în concordanță cu valorile umane și că impactul său este benefic și în conformitate cu principiile morale. Abordarea etică în proiectarea fundamentală a inteligenței artificiale ne ajută să evităm consecințele nedorite, să promovăm utilizarea responsabilă a tehnologiilor și să ne asigurăm că acestea aduc beneficii societății în ansamblul ei [169].

Studiul creierului uman și al proceselor cognitive este și el esențial în înțelegerea și replicarea inteligenței umane în sistemele de inteligență artificială. **Neuroștiințele și științele cognitive** investighează cum funcționează creierul, cum se formează și se procesează informațiile și cum se realizează învățarea și raționamentul. Aceste cunoștințe sunt aplicate în dezvoltarea modelelor de rețele neuronale artificiale și de învățare profundă (*deep learning*), care încearcă să simuleze funcționarea creierului uman și să obțină rezultate comparabile în ceea ce privește recunoașterea de tipare (*pattern recognition*), înțelegerea limbajului și luarea de decizii [18].

Statistica și teoria probabilităților sunt utilizate în inteligența artificială pentru a analiza datele și pentru a efectua inferențe și predicții. Metodele statistice sunt folosite pentru a extrage informații din seturile de date, pentru a identifica modele și tendințe și pentru a realiza predicții având la bază datele disponibile. Teoria probabilităților este utilizată pentru a cuantifica incertitudinea și pentru a evalua probabilitatea evenimentelor în contextul luării deciziilor în sistemele de inteligență artificială [133].

De exemplu, modelele bayesiene reprezintă o abordare puternică în inteligența artificială pentru luarea deciziilor și analiza datelor. Ele se bazează pe teorema lui Bayes și sunt folosite pentru a calcula probabilitățile și relațiile dintre evenimente. Aceste modele pot fi utilizate într-o varietate de aplicații, cum ar fi clasificarea datelor, filtrare de spam, recunoașterea vocală, înțelegerea limbajului natural și multe altele. Ele sunt eficiente în situații în care există incertitudine și este necesară luarea de decizii pe baza probabilităților. Un aspect important al modelelor bayesiene este capacitatea lor de a actualiza continuu probabilitățile pe măsură ce apar date noi [133]. Modelele bayesiene permit integrarea noilor informații și ajustarea rezultatelor în funcție de observații și reprezintă o unealtă puternică pentru a modela și înțelege lumea complexă în care trăim [168].

Este important de menționat că însăși fundamentele științifice a inteligenței artificiale scot în evidență caracterul „artificial" al acesteia. Nimeni atunci când vorbește despre inteligența umană nu se referă strict la fundamentele ei științifice. Inteligența umană este un fenomen complex, influențat de factori multipli, precum genetica, dezvoltarea cognitivă, experiențele individuale și interacțiunea cu mediul înconjurător. Aceasta implică o gamă largă de aspecte, inclusiv emoții, conștiință și capacitatea de a face alegeri morale.

În schimb, inteligența artificială se bazează pe algoritmi și modele matematice, pe procesarea datelor și pe încorporarea cunoștințelor în sistemele computerizate. Este o replică a capacităților intelectuale umane, dar cu o abordare total diferită și o funcționare bazată pe reguli și proceduri predefinite [137].

În timp ce fundamentele științifice ale inteligenței artificiale sunt esențiale pentru dezvoltarea și funcționarea acesteia, nu putem neglija faptul că inteligența artificială este o creație umană, un produs al ingineriei și al tehnologiei. De aceea, inteligența artificială este fundamental diferită de inteligența umană, care este inevitabil înrădăcinată în biologie și într-un sistem complex de interconexiuni neuronale, care, la rândul lui, este integrat unui sistem care depășește chiar și cadrul biologic [56; 85]. Aici, după cum vom vedea, metodologia transdisciplinară poate aduce explicații semnificative.

Prin urmare, este important să înțelegem că inteligența artificială, în ciuda progreselor sale și a capacităților sale impresionante, nu poate înlocui pe deplin inteligența umană. Fundamentele științifice ale inteligenței artificiale sunt esențiale pentru dezvoltarea și funcționarea acesteia, dar nu putem ignora diferențele insurmontabile dintre inteligența umană și inteligența artificială. Înțelegerea acestor diferențe ne permite să evaluăm și să valorizăm adecvat rolul și impactul inteligenței artificiale în societatea noastră și să ne asigurăm că progresul tehnologic este orientat spre beneficiul și binele comun.

1.3 De la interdisciplinaritate la transdisciplinaritate.

”Separarea domeniilor este o convenție a omului,
ceva artificial. (...) natura nu se interesează de
delimitările pe care le facem noi și multe
fenomene interesante se află la limita dintre
diferite domenii.”

Richard Feynman [138]

Ideea lui Feynman ne obligă să recunoaștem că specializarea rigidă, deși utilă pentru organizarea studiului, este în esență o limitare autoimpusă care nu reflectă complexitatea lumii reale. În practică, natura funcționează ca un sistem unitar, iar cele mai valoroase perspective apar adesea la intersecția unor domenii care, la prima vedere, nu au nimic în comun. Aici, abordarea transdisciplinară devine esențială atunci când analizăm tehnologiile emergente, unde succesul nu mai depinde doar de rigoarea tehnică, ci de capacitatea noastră de a conecta idei din sfere variate.

Tocmai de aceea, integrarea disciplinelor diverse în cercetarea și dezvoltarea inteligenței artificiale joacă un rol crucial în avansarea acestui domeniu și în înțelegerea impactului său asupra societății. De asemenea, inteligența artificială are o influență semnificativă asupra dezvoltării altor științe și domenii, iar colaborarea între acestea duce la progrese semnificative nu doar în înțelegerea și aplicarea ei, ci și în progresul domeniilor cu care inteligența artificială interacționează.

În continuare, vom analiza succint modul în care anumite domenii ale cunoașterii umane au integrat concepte strâns legate de apariția și dezvoltarea inteligenței artificiale. Ne vom opri asupra ontologiei, epistemologie, psihologiei, antropologie, teologiei, economiei și ecologiei.

Interacțiunile dintre **ontologie** și inteligența artificială sunt strâns legate de modul în care această tehnologie utilizează și modelează cunoștințele și reprezentările ontologice în procesul său de înțelegere și luare a deciziilor. Pe de-o parte, ontologia poate furniza baza conceptuală și structura formală necesară pentru a organiza și reprezenta cunoștințele într-un mod semnificativ și coerent. Pe de altă parte, inteligența artificială poate contribui la evoluția ontologiei prin aplicarea sa în dezvoltarea și rafinarea ontologiilor existente, prin extragerea de cunoștințe din texte și resurse nestructurate și prin generarea automată de ontologii [168].

De exemplu Cyc este un proiect de dezvoltare a unei ontologii generale și puternice, care cuprinde cunoștințe despre lumea fizică, evenimente, acțiuni, concepte abstracte și multe altele. Utilizând inteligența artificială, Cyc a fost alimentat cu o serie de cunoștințe din diverse surse, iar

ontologia sa a fost extinsă și rafinată pentru a sprijini raționamentul automat și luarea deciziilor în diverse domenii. Alte exemple de ontologii sunt DBpedia (model de ontologie care utilizează algoritmi de extragere a informațiilor și de prelucrare a limbajului natural pentru a extrage și structura informații din diverse surse de cunoștințe semantice), Gene Ontology (o ontologie specifică domeniului biologic, care modelează cunoștințele despre gene, funcțiile lor și relațiile dintre ele), WordNet (o ontologie lexicală și semantică care modelează relațiile dintre cuvinte și sensurile lor, utilizând metode de procesare a limbajului natural și algoritmi de clustering) [161].

Este evident că există o legătură între ontologia, în sens filosofic, și ontologia calculatoarelor dezvoltată în cadrul inteligenței artificiale, dar, pentru a nu genera confuzii, trebuie spus că există și diferențe semnificative între cele două concepte. Prin urmare, atât ontologia filosofică, cât și ontologia calculatoarelor se ocupă de natura existenței și a ființei. În ambele cazuri, se definesc entități și se descriu relațiile dintre aceste entități pentru a obține o înțelegere mai profundă a lumii în care trăim sau, în cazul calculatoarelor, a sistemelor informatice [158].

Dar, în timp ce ontologia filosofică are un caracter mai general și este preocupată de întrebări fundamentale despre realitate, existență și natura ființei umane, ontologia calculatoarelor se concentrează, în principal, pe dezvoltarea de modele formale care să permită reprezentarea și gestionarea cunoașterii în contextul aplicațiilor informatice și a inteligenței artificiale [161].

Din punct de vedere metodologic, în cadrul ontologiei filosofice, abordarea este una speculativă și teoretică, bazată pe raționament și discuții conceptuale. În ontologia calculatoarelor, se folosesc metode și tehnici formale, cum ar fi limbajele formale, logică și tehnici de prelucrare a limbajului natural. În ontologia calculatoarelor se urmărește dezvoltarea de modele computaționale care să permită reprezentarea și gestionarea cunoașterii. Spre deosebire de ontologia calculatoarelor, în ontologia filosofică se urmărește explorarea și dezbaterea conceptelor filosofice fundamentale, fără a se limita la aspectele tehnice și computaționale [160].

Ontologia poate servi ca o limbă comună de comunicare între diferite sisteme ale inteligenței artificiale și resurse de informații [160]. Prin definirea unui set comun de concepte și relații în ontologie, sistemele inteligenței artificiale pot comunica și colabora mai eficient.

Interacțiunile și influențele reciproce dintre **epistemologie** și inteligența artificială sunt subiectul unei dezbateri active în domeniul filozofiei cunoașterii și al inteligenței artificiale [160; 161; 148]. În acest sens, inteligența artificială contribuie prin dezvoltarea de noi metode și tehnici care, inevitabil, au impact asupra modului în care obținem cunoașterea. De exemplu, algoritmii de învățare automată (*machine learning*) și de învățare profundă (*deep learning*) pot fi utilizați pentru a analiza și extrage cunoștințe din mari volume de date. Aceasta ridică întrebări despre fiabilitatea și validitatea acestor metode în raport cu tradiționalele metode de obținere a cunoașterii [54].

Sistemele de inteligență artificială pot genera rezultate sau concluzii pe cale automat, dar cum putem evalua dacă acestea sunt corecte și dacă putem avea încredere în ele? Aceasta implică o reconsiderare a criteriilor de validitate și a surselor de cunoaștere [145].

În acest sens, inteligența artificială poate aduce noi tipuri de cunoaștere care sunt generate automat și care nu se bazează pe raționamentul uman tradițional. De exemplu, rețelele neuronale artificiale cu mai multe straturi pot realiza clasificări și recunoaștere a modelelor **într-un mod care nu poate fi explicat în termeni umani**. Aceasta ridică întrebări despre natura și valabilitatea acestui tip de cunoaștere [162].

Sistemele de inteligență artificială, deși avansate, acționează ca o oglindă pentru propriile noastre limitări cognitive și epistemologice. Vulnerabilitatea algoritmilor la erori de procesare sau la interpretări eronate ale datelor ne reamintește că nicio sursă de cunoaștere nu este infailibilă. Acest fapt subliniază importanța gândirii critice și necesitatea de a integra discernământul uman în validarea informațiilor, asigurându-ne că tehnologia rămâne un instrument complementar, și nu un substitut pentru rigoarea intelectuală umană. [9].

Inevitabil, apariția inteligenței artificiale a generat întrebări despre rolul uman în procesul de cunoaștere. Cum se încadrează agentul uman într-un context în care mașinile pot genera și utiliza cunoștințe? Astfel de întrebări pot determina o reevaluare a naturii și valorii cunoașterii umane și a relației dintre om și tehnologie.

Legătura **psihologiei** cu inteligența artificială poate fi descrisă de un impact semnificativ asupra ambelor domenii, generând influențe și schimbări de modele în ambele sensuri [152].

În acest sens, inteligența artificială contribuie la înțelegerea proceselor cognitive umane prin intermediul modelelor și algoritmilor de învățare automată. Prin analiza datelor și identificarea modelelor, inteligența artificială poate sugera modele explicative cu privire la modul în care funcționează procesele cognitive umane, cum ar fi percepția, atenția, memoria și luarea deciziilor [151].

De asemenea, inteligența artificială poate inspira și susține dezvoltarea de noi teorii și modele psihologice. Modelele de rețele neuronale profunde, de exemplu, sunt folosite adesea pentru a crea simulări computaționale ale proceselor cognitive umane și pentru a testa teorii psihologice existente sau pentru a dezvolta noi abordări în înțelegerea comportamentului uman [18; 151].

Inteligența artificială este deja utilizată pentru a dezvolta și implementa terapii psihologice inovatoare. De exemplu, există aplicații de asistență virtuală și *chatbot*-uri care oferă suport emoțional și consiliere în timp real. Acestea pot fi folosite pentru terapii asistate de tehnologie, cum ar fi terapia cognitiv-comportamentală sau terapia de gestionare a stresului [139]. Este evident

că astfel de tehnici influențează modele de terapie și, reciproc, modele din terapie sunt folosite în proiectarea asistenților virtuali.

Așadar, inteligența artificială contribuie la îmbunătățirea proceselor de evaluare și diagnosticare în psihologie. Sistemele de inteligență artificială pot analiza și interpreta datele clinice și de evaluare pentru a oferi evaluări și diagnoze mai precise și mai rapide. De exemplu, algoritmii de învățare automată pot fi antrenați să detecteze modele în datele psihologice și să ofere suport în stabilirea diagnosticelor. Dar, este clar că modele de învățare automată nu pot fi antrenate fără a lua în considerare procesele de evaluare și diagnosticare „clasice" și fără aportul datelor clinice deja existente. Așadar, procesul de interacțiune are loc, și în acest caz, în dublu sens.

Inteligența artificială este utilizată și pentru a înțelege și adapta interacțiunea dintre oameni și tehnologie. Aceste sisteme pot analiza și interpreta comportamentul uman în interacțiune cu interfețele și dispozitivele tehnologice (*gesture recognition*, *gait recognition* și *sentiment analysis*) și pot oferi feedback și sugestii personalizate pentru a îmbunătăți experiența utilizatorului [168; 139].

Analiza interdependențelor dintre **științele educației** și inteligența artificială vine cu provocări și, în același timp, cu oportunități semnificative în ambele domenii.

Cât nu ar părea de contradictoriu, având în vedere folosirea de către elevi și studenți în mod abuziv a LLM-urilor, inteligența artificială are potențialul de a contribui la personalizarea învățării prin adaptarea conținutului, ritmului și metodelor de predare în funcție de nevoile și nivelul de cunoștințe al fiecărui elev. Algoritmii pot analiza și interpreta datele privind progresul și performanța elevilor pentru a oferi feedback individualizat și pentru a adapta instruirea într-o formă mai dinamică. Reciproc, modele de inteligență artificială folosite în scop educațional se constituie și evoluează având la bază concepte și metode din teoria educației [141].

Așa cu am menționat mai sus, inteligența artificială poate îmbunătăți procesele de evaluare și feedback în domeniul educațional. De exemplu, algoritmii de inteligență artificială pot analiza răspunsurile și rezolvările elevilor și pot oferi feedback imediat și precis, ajutându-i să identifice erorile și să înțeleagă mai bine conceptele. În același timp, un proces de evaluare care folosește inteligența artificială nu poate fi conceput fără a lua în considerare sistemele de evaluare tradiționale, având în vedere imensa cantitate de date și experiență acumulată. Se vede și în acest caz cum sistemele de evaluare care folosesc inteligența artificială duc la reconsiderarea de pe alte baze a procesului de evaluare, dar și faptul că aceste sisteme se dezvoltă în strânsă legătură cu filosofia educațională [127]. Aici trebuie avut grijă ca nu cumva, din dorința eliminării cu orice preț a subiectivității inerente oricărui proces de evaluare tradițional să nu se cadă în extrema unei obiectivități „reci", pur algoritmică. Deja există exemple de drame umane generate de folosirea

fără discernământ în domeniul educațional a unor algoritmi de evaluare nesupervizați corespunzător de către factorul uman [137].

Inteligența artificială mai poate fi folosită pentru a dezvolta sisteme de tutoriat virtual care pot oferi suport și asistență individualizată elevilor. Aceste sisteme pot răspunde întrebărilor, pot explica concepte complexe și pot oferi resurse suplimentare adaptate nevoilor specifice ale fiecărui elev. Consecința constă în aceea că un elev învață astăzi diferit de cum învăța un elev în urmă cu 20 de ani.

Sociologia și inteligența artificială ne oferă una dintre cele mai complexe interacțiuni.

Dată fiind cantitatea din ce în ce mai mare de date achiziționate de sisteme informatice, inteligența artificială interacționează cu sociologia în domeniul analizei și interpretării datelor sociale masive (*big data*). Prin aplicarea algoritmilor de învățare automată și a tehnicilor de prelucrare a limbajului natural, inteligența artificială poate extrage informații semnificative din postări pe rețele sociale, interacțiuni online și alte surse de date cu impact sociologic. Aceasta contribuie la identificarea de modele și tendințe sociale, precum și la înțelegerea dinamicilor și structurilor sociale [168].

Inteligența artificială a devenit recent o metodă utilă pentru a crea simulări și modele sociale. Prin dezvoltarea de agenți virtuali autonomi și interacțiunea acestora într-un mediu simulat, se pot explora comportamente sociale și dinamici complexe. Aceste simulări pot ajuta la dezvoltarea, testarea și validarea teoriilor sociologice, precum și la înțelegerea implicațiilor diferitelor politici și intervenții sociale [160].

Inteligența artificială exercită o influență din ce în ce mai vizibilă în influențarea ofertelor de locuri de muncă și analiza distribuției inegale a resurselor în societate. Automatizarea și robotizarea pot genera schimbări majore în piața muncii, cu potențialul de a crea inegalități și de a afecta anumite grupuri sociale. Așadar, inteligența artificială poate fi utilizată pentru a identifica și cuantifica inegalitățile sociale și pentru a sugera posibile soluții pentru reducerea acestora [155].

Interacțiunea dintre sociologie și inteligența artificială pune în discuție problemele etice și responsabilitatea socială asociate cu utilizarea tehnologiei inteligente. Sociologia poate contribui la înțelegerea impactului inteligenței artificiale asupra societății și a modului în care aceasta poate fi utilizată într-un mod etic și responsabil, luând în considerare implicațiile sociale, economice și culturale.

Într-un context asemănător cu cel sociologic, **antropologia** și inteligența artificială au un impact semnificativ asupra înțelegerii umane și a culturilor.

Inteligența artificială deschide noi orizonturi în studiul interacțiunii dintre oameni și mașini. Antropologii pot investiga cum oamenii interacționează cu agenții de inteligență artificială și cum aceste interacțiuni afectează comportamentele, relațiile sociale și cultura [55].

Inteligența artificială poate schimba modul în care oamenii desfășoară anumite activități și practici culturale. De exemplu, aplicațiile de traducere automată pot influența modul în care se comunică între culturi, iar utilizarea asistenților virtuali poate afecta modul în care se iau decizii în diverse contexte sociale [55]. Aceste schimbări pot avea implicații semnificative în privința identității culturale, valorilor și relațiilor interumane.

Interacțiunea cu inteligența artificială poate conduce la o reevaluare a conceptelor și teoriilor antropologice. Antropologii pot revizui și adapta concepte precum identitatea, cultura, comunicarea și practicile sociale în contextul tehnologiilor inteligente [151]. De asemenea, pot explora concepte emergente, cum ar fi inteligența artificială distribuită și inteligența colectivă.

Relațiile dintre **teologie** și inteligența artificială sunt un subiect complex și provocator. Legăturile dintre cele două domenii pot avea influențe reciproce, provocând întrebări noi teologice [85; 56].

Inteligența artificială este foarte utilă în analiza textelor religioase și pentru a extrage cunoștințe semantice și contextuale din ele. Aceasta poate facilita cercetarea și interpretarea textelor sacre, oferind abordări mai eficiente și sistematice în ceea ce privește analiza stilistică, semnificațiile multiple și contextul cultural. De exemplu, inteligența artificială poate identifica conexiuni între diverse pasaje biblice sau poate explora anumite semnificații în textele mistice.

Dezvoltarea și utilizarea inteligenței artificiale ridică întrebări legate de impactul asupra credinței și spiritualității umane. Credința și spiritualitatea implică adesea dimensiuni umane unice, cum ar fi conexiunea emoțională și experiența personală. Integrarea inteligenței artificiale în acest domeniu poate avea influențe asupra modului în care oamenii își experimentează și își exprimă credința, dar și asupra modului în care își îndeplinesc ritualurile religioase și se angajează în practici spirituale [151].

Un exemplu concret ar fi acela că apariția inteligenței artificiale aduce întrebări noi despre natura și locul umanității în creație. Cum se încadrează mașinile inteligente în actul creației divine? Cum influențează capacitatea lor de a lua decizii autonome, înțelegerea noastră despre condiția umană?

În contextul inteligenței artificiale, se ridică întrebări privind responsabilitatea morală a sistemelor inteligente. Cum atribuim responsabilitatea morală unui algoritm sau unei mașini autonome? Cum se aplică concepte teologice precum păcatul, iertarea sau răscumpărarea într-un context în care acțiunile morale pot fi realizate și de mașini? De exemplu, în cazul producerii unui

accident de către un vehicul autonom care folosește inteligența artificială cine va fi tras la răspundere din punct de vedere legal? [115]

Apariția inteligenței artificiale poate alimenta discuții despre natura conștiinței și a sufletului uman. Se poate dezvolta o conștiință sau un suflet artificial într-o mașină inteligentă? Cum se raportează aceasta la înțelegerea tradițională a conștiinței și a sufletului în teologie? [115]

Mai mult, inteligența artificială ridică întrebări despre limitele și libertatea umană în raport cu predestinarea sau controlul divin. Cum se încadrează capacitățile și deciziile umane într-un context în care mașinile pot lua decizii complexe pe baza algoritmilor și datelor?

Progresul tehnologic, inclusiv cel legat de inteligența artificială, poate influența modalitățile de comunicare și acces la cunoaștere divină. Cum se schimbă relația omului cu divinul într-un context în care tehnologia mediază tot mai mult experiența spirituală și interacțiunea cu practicile religioase?

Relaționările dintre **economie** și inteligența artificială sunt și ele semnificative și aduc schimbări majore atât în domeniul economic, cât și în direcțiile de dezvoltare a inteligenței artificiale. Unul dintre motivele pentru care se întâmplă acest lucru este că inteligența artificială are capacitatea de a automatiza și optimiza procesele economice și operaționale. Prin aplicarea algoritmilor de învățare automată și tehnici de procesare a datelor, inteligența artificială poate contribui la creșterea eficienței în diverse domenii economice, precum producție, logistică, managementul lanțului de aprovizionare și servicii financiare. Automatizarea anumitor sarcini poate duce la reducerea costurilor și la creșterea productivității în economie [139; 168].

Inteligența artificială procesează și analizează cantități uriașe de date economice, ceea ce permite generarea de informații și înțelegerea mai profundă a piețelor, a comportamentului consumatorilor și a tendințelor economice. Acest fapt facilitează luarea de decizii informate în legătură cu strategiile de afaceri, marketingul, investițiile și managementul riscurilor.

De asemenea, inteligența artificială generează noi modele de afaceri și inovații economice. De exemplu, platformele online bazate pe algoritmi inteligenți eficientizează comerțul electronic și serviciile personalizate, iar roboții și asistenții virtuali oferă experiențe unice pentru clienți. Astfel, inteligența artificială influențează semnificativ modul în care companiile operează și interacționează cu consumatorii [148].

Similar cu efectele de natură sociologică, introducerea inteligenței artificiale în economie poate avea un impact semnificativ asupra ocupării forței de muncă. Automatizarea și robotizarea pot duce la schimbări în cererea de forță de muncă și pot afecta anumite sectoare și ocupații în mod diferit [148]. Vedem foarte clar că unele locuri de muncă repetitive și rutiniere sunt deja înlocuite de tehnologie, în timp ce altele beneficiază de suportul tehnologiei inteligente.

Interacțiunea dintre inteligența artificială și economie poate avea implicații semnificative asupra politicii economice. Guvernele și instituțiile pot folosi inteligența artificială pentru a monitoriza și reglementa piețele, a îmbunătăți eficiența administrativă și a dezvolta politici economice bazate pe date și analize avansate [160]. De asemenea, trebuie abordate și aspectele legate de securitatea datelor și protecția consumatorilor în contextul utilizării inteligenței artificiale în economie.

Influențele reciproce dintre **ecologie** și inteligența artificială produc deja beneficii semnificative în abordarea problemelor ecologice și în conservarea mediului înconjurător. Acest tip de tehnologie inteligentă este utilizată pentru a monitoriza și evalua starea mediului înconjurător. Prin analiza datelor colectate din senzori, imagini satelitare și alte surse, inteligența artificială este capabilă să identifice modele, tendințe și schimbări în ecosisteme [168]. Acest lucru facilitează gestionarea și conservarea resurselor naturale, precum apa, pădurile, solurile și biodiversitatea.

Inteligența artificială mai este folosită și pentru a dezvolta modele și prognoze ale schimbărilor climatice [162]. Prin analiza datelor istorice și a factorilor care influențează clima, inteligența artificială ajută la înțelegerea impactului schimbărilor climatice asupra ecosistemelor și populațiilor. Aceasta facilitează dezvoltarea de strategii și politici pentru adaptarea și ameliorarea efectelor schimbărilor climatice.

Sistemele inteligente contribuie, în diverse aspecte, la conservarea biodiversității prin identificarea și monitorizarea speciilor. Algoritmii de învățare automată sunt deja utilizați pentru clasificarea și recunoașterea speciilor pe baza datelor de imagine sau audio. Acest lucru permite identificarea speciilor în pericol și monitorizarea populațiilor, ajutând la luarea de decizii cu privire la conservarea și protejarea acestora [167].

În acest moment, pentru prima dată în istoria umanității, inteligența artificială contribuie la creșterea eficienței și a sustenabilității în agricultură. Prin analiza datelor despre sol, climă și culturi, inteligența artificială poate furniza informații pentru optimizarea aplicării unor parametri de control, cum ar fi cantitatea de apă, îngrășăminte, insecticide și pesticide. Aceasta contribuie la reducerea risipei și impactul asupra mediului înconjurător [166].

La fel, inteligența artificială poate contribuie semnificativ la dezvoltarea și implementarea tehnologiilor verzi și a energiei sustenabile. Prin analiza datelor și optimizarea sistemelor, inteligența artificială contribuie la creșterea eficienței energetice și ajută la identificarea celor mai bune strategii de producție și utilizare a energiei regenerabile [166].

Inteligența artificială este din ce în ce mai utilizată pentru a detecta și preveni dezastrele naturale. Prin analiza datelor, precum modele meteorologice, date de senzori și imagini preluate

din satelit, inteligența artificială contribuie cu rezultate deja remarcabile la identificarea riscurilor și a tendințelor care pot duce la dezastre. Aceasta facilitează dezvoltarea de sisteme de avertizare timpurie și strategii de intervenție în situații de urgență [139].

În contextul actual, inteligența artificială contribuie la reducerea impactului asupra mediului prin optimizarea proceselor industriale și a sistemelor logistice. Algoritmii de învățare automată, dacă sunt bine aleși și configurați, pot fi utilizați cu succes pentru a optimiza rutele de transport, reducând astfel emisiile de carbon și îmbunătățind eficiența energetică.

Rezumând cele expuse anterior, se constată că interacțiunile dintre inteligența artificială și diverse discipline sunt semnificative și reciproce, aducând schimbări și provocări în fiecare domeniu, în ambele sensuri. Integrarea inteligenței artificiale de către aceste discipline deschide noi orizonturi și oportunități, dar solicită și abordări complexe din punct de vedere etic, sociologic și, nu în ultimul rând, suscită noi reinterpretări din perspectivă filosofică [160].

Așa cum s-a văzut, integrarea inteligenței artificiale în ontologie, teologie, epistemologie, psihologie, științele educației, sociologie, antropologie, economie și ecologie aduce schimbări fundamentale în modul în care înțelegem și explorăm aceste domenii, dar și în modul în care proiectăm, implementăm și evaluăm tehnologiile care folosesc inteligența artificială. Aceasta implică cu necesitate o reevaluare a conceptelor și teoriilor existente din fiecare domeniu, precum și o explorare a noilor posibilități și perspective pe care le oferă inteligența artificială [162; 161].

Multidisciplinaritate, Interdisciplinaritate, Transdisciplinaritate

Abordarea multidisciplinară și interdisciplinară în cercetarea inteligenței artificiale reprezintă o tendință tot mai pronunțată în ultimele decenii, având în vedere complexitatea și diversitatea acestei discipline. În contextul creșterii exponențiale a cunoștințelor și a aplicațiilor tehnologice, specializarea excesivă poate avea anumite neajunsuri, cum ar fi pierderea sensului și progresia absurdității și nonsensului [91].

Specializarea excesivă are avantajele sale, accelerând progresul cunoașterii și facilitând dezvoltarea tehnologică în diferite domenii. În ciuda acestui fapt, ea poate duce la fragmentarea și izolarea disciplinelor, pierzându-se astfel perspectiva de ansamblu și capacitatea de a înțelege conexiunile și interacțiunile dintre diferitele aspecte ale cercetării în domeniul inteligenței artificiale [91].

Pentru a depăși aceste limite, abordarea multidisciplinară și interdisciplinară devine tot mai necesară. Prin aducerea împreună a experților din diferite domenii, cum ar fi informatica, matematica, neuroștiințele, filosofia, psihologia și etica, se poate obține o perspectivă mai cuprinzătoare asupra inteligenței artificiale [92]. În consecință, avem posibilitatea să abordăm

problemele complexe într-un mod mai holistic și se pot descoperi soluții inovatoare care nu ar fi fost posibile în cadrul unei abordări disciplinare restrânse.

Multidisciplinaritatea implică colaborarea între specialiști din diferite domenii, care își aduc contribuțiile specifice la cercetarea inteligenței artificiale. De exemplu, un grup de cercetători poate include programatori, psihologi și eticieni, fiecare punându-și în comun cunoștințele pentru a aborda aspecte tehnice, cognitive și etice ale dezvoltării și utilizării inteligenței artificiale. Această abordare poate aduce o mai mare diversitate de idei și perspective, contribuind la soluții mai complete și echilibrate.

Interdisciplinaritatea merge mai departe, punând accentul pe transferul de metode și concepte între discipline. Astfel, se pot dezvolta noi abordări și tehnici care combină aspecte din diferite domenii și pot genera progrese semnificative în cercetarea inteligenței artificiale. De exemplu, transferul de metode matematice și algoritmice din domeniul informaticii către medicină poate conduce la noi tratamente ale cancerului sau alte aplicații medicale inovatoare.

Transdisciplinaritatea, așa cum vom vedea mai pe larg în cuprinsul acestei lucrări, abordează aspectele care se află dincolo de orice disciplină și se concentrează pe înțelegerea globală și unitară a lumii prezente. Transdisciplinaritatea se referă la integrarea și transcenderea limitelor disciplinelor individuale pentru a aborda problemele complexe dintr-o perspectivă holistică [91; 92].

Cu toate acestea, există și unele provocări și neajunsuri asociate cu abordarea multidisciplinară și interdisciplinară. Unul dintre neajunsuri este dificultatea de a facilita comunicarea și colaborarea între specialiștii din diferite domenii, având în vedere limbajele și paradigmele diferite folosite în disciplinele respective [86]. De asemenea, sincronizarea diferitelor ritmuri și metode de lucru poate fi o provocare în eforturile de integrare și transfer de cunoștințe.

În plus, abordarea multidisciplinară și interdisciplinară poate necesita mai mult timp și resurse pentru a integra și reconcilia perspectivele diferite, ceea ce poate încetini progresul în anumite aspecte ale cercetării [91]. Există riscul ca integrarea excesivă să ducă la o diluare a expertizei și la o pierdere a profunzimii într-un domeniu specific.

Pentru a depăși aceste provocări, este necesară o abordare echilibrată care să valorifice avantajele specializării și ale abordărilor multidisciplinare și interdisciplinare. Acest lucru nu este întotdeauna ușor de realizat. Este important să recunoaștem că specializarea aduce un progres semnificativ în cadrul fiecărei discipline și că este necesară pentru a avansa cunoașterea și tehnologia. Însă, nu ar trebui să ne limităm doar la specializare, ci să recunoaștem importanța integrării și colaborării interdisciplinare.

Una dintre modalitățile de a aborda aceste provocări este prin crearea unor medii și structuri care să faciliteze colaborarea și schimbul de cunoștințe între specialiști din diferite domenii. Aceasta poate include organizarea de conferințe interdisciplinare, proiecte de cercetare colective și programe de formare care să promoveze abordarea interdisciplinară.

De asemenea, este important să se promoveze gândirea critică și analitică în rândul cercetătorilor și specialiștilor din domeniul inteligenței artificiale [56]. Acest lucru poate ajuta la depășirea limitărilor specializării excesive și la încurajarea explorării și integrării ideilor din diverse discipline.

În ciuda neajunsurilor, este evident că abordarea multidisciplinară și interdisciplinară în cercetarea inteligenței artificiale aduce beneficii semnificative în înțelegerea și dezvoltarea acestei tehnologii. Prin integrarea perspectivelor diferite, putem obține soluții mai complete și mai etice, care să abordeze provocările complexe pe care le implică inteligența artificială [56]. O abordare echilibrată și colectivă poate deschide noi orizonturi și poate contribui la progresul cunoașterii în acest domeniu fascinant al inteligenței artificiale.

La fel, diferențele dintre cunoașterea interdisciplinară și cunoașterea transdisciplinară pot fi evidențiate în mai multe aspecte. Din punct de vedere al perspectivei asupra lumii, cunoașterea interdisciplinară, descrisă ca *„in vitro”* (în laborator), se concentrează asupra lumii exterioare, obiective, în timp ce cunoașterea transdisciplinară, denumită *„in vivo”* (în viață), se referă la corespondența dintre lumea externă și lumea internă a subiectului care cunoaște [91].

Cunoașterea interdisciplinară este adesea asociată cu inteligența analitică, în timp ce cunoașterea transdisciplinară implică înțelegerea, care depășește simpla analiză rațională. Cunoașterea transdisciplinară își propune să integreze diferite aspecte ale cunoașterii, inclusiv emoțiile și experiențele subiectului. Cunoașterea interdisciplinară, așa cum afirmă Basarab Nicolescu, este orientată spre putere și posesie, în timp ce cunoașterea transdisciplinară este orientată spre uimire și partaj. Cunoașterea interdisciplinară poate viza obținerea de control sau dominație asupra obiectelor de studiu aflate în conexiune, în timp ce cunoașterea transdisciplinară își propune să exploreze și să înțeleagă lumea într-un mod holistic și integrat [91].

Mai mult ca atât, cunoașterea interdisciplinară se bazează adesea pe o logică binară, care implică distincții clare între concepte și excluderea valorilor subiective. În schimb, cunoașterea transdisciplinară utilizează logica terțului inclus, care recunoaște complexitatea și interconectarea aspectelor diferite ale cunoașterii și include valorile în procesul de înțelegere [92].

În ansamblu, cunoașterea interdisciplinară și cunoașterea transdisciplinară reprezintă abordări diferite în explorarea și înțelegerea lumii [91]. Cunoașterea interdisciplinară se concentrează pe investigarea în profunzime a unei probleme dintr-un domeniu specific prin

raportarea lui la conexiunile conceptuale cu alte domenii, în timp ce cunoașterea transdisciplinară urmărește conexiuni și sinergii între diferite discipline și perspective. Ambele abordări au rolul lor în dezvoltarea cunoașterii, dar cunoașterea transdisciplinară aduce în plus o dimensiune holistică și integratoare, promovând înțelegerea globală și includerea valorilor în procesul de cunoaștere.

Cunoașterea interdisciplinară, într-o analogie cu *„in vitro”* [91], se referă la studiul profund și specializat al unor domenii specifice, cu accent pe dezvoltarea unei expertize interdisciplinare. Aceasta implică utilizarea metodelor, teoriilor și paradigmelor interdisciplinare, adesea folosind o abordare analitică și rațională. Cercetătorii interdisciplinari se concentrează pe investigarea detaliată și adâncă a unui aspect specific al lumii, urmărind, prin coroborarea abordărilor din mai multe discipline, să obțină o înțelegere profundă și precisă.

În contrast, cunoașterea transdisciplinară, numită *„in vivo”*, implică o abordare holistică și integrată care depășește limitele disciplinelor individuale, inclusiv cu conexiunile lor interdisciplinare. Ea implică conexiuni și interacțiuni între diferite domenii și integrarea diferitelor perspective și metode de cercetare. Cercetătorii în transdisciplinaritate încearcă să abordeze probleme complexe, care necesită o înțelegere globală și sinergii între diferite discipline.

Să luăm ca exemplu studiul schimbărilor climatice. Abordarea interdisciplinară ar putea implica cercetarea unor domenii specifice, mai mult sau mai puțin înrudite, cum ar fi meteorologia și ecologia, pentru a înțelege modelele climatice și impactul schimbărilor climatice asupra ecosistemelor [91]. În același timp, abordarea transdisciplinară ar putea implica integrarea cunoștințelor și metodelor din mai multe domenii, nu doar o simplă coroborare prin suprapunerea și completarea reciprocă a interpretărilor, cum ar fi științele sociale, ecologia, economia și politica, pentru a înțelege implicațiile sociale, economice și politice ale schimbărilor climatice și pentru a dezvolta soluții sustenabile.

Cunoașterea transdisciplinară se concentrează pe înțelegerea complexității și interconexiunii dintre diferite aspecte ale lumii și abordează provocările care necesită o perspectivă integrată și holistică. Ea include valorile, experiențele și perspectivele diferite ale indivizilor și încearcă să promoveze înțelegerea globală și colaborarea interdisciplinară.

O altă diferență importantă între cunoașterea interdisciplinară și cea transdisciplinară este orientarea și scopul acestora. Cunoașterea interdisciplinară, de obicei, se concentrează pe obținerea de cunoștințe și expertiză în anumite domenii conexe, iar obiectivul principal este adesea rezolvarea problemelor specifice sau dezvoltarea de teorii și paradigme într-un cadru interdisciplinar [91].

În schimb, cunoașterea transdisciplinară are ca scop înțelegerea și abordarea problemelor complexe, care nu pot fi rezolvate prin intermediul unei singure discipline [91]. Ea caută conexiuni

și interacțiuni între diferite domenii pentru a dezvolta o viziune mai cuprinzătoare și holistică asupra problemelor și pentru a găsi soluții inovatoare și creative.

În concluzie, cunoașterea interdisciplinară se concentrează pe explorarea profundă și specializată a unor domenii specifice, conectate între ele prin concepte comune, având drept scop rezolvarea problemelor interne ale unei discipline, dar într-un context interdisciplinar. În schimp, cunoașterea transdisciplinară are o abordare holistică, integrată, prin care explorează interconexiunile și relațiile dintre domenii diferite pentru a aborda probleme complexe și globale [91]. Despre specificul abordării transdisciplinare, vom trata mai pe larg, cu explicații amănunțite în subcapitolul următor.

1.4 Niveluri de realitate și logica terțului inclus în studiul inteligenței artificiale

Așa cum arată cercetătorul român Basarab Nicolescu, viziunea transdisciplinară presupune o Realitate multidimensională, structurată pe multiple niveluri, care înlocuiește Realitatea unidimensională, pe un singur nivel, a gândirii clasice.

Pentru început, este important să facem o analiză complexă a conceptelor de **Real** și **Realitate**. Pentru a evita diferite ambiguități, Basarab Nicolescu face o distincție între **Real** și **Realitate**. *Realul desemnează ceea ce este, în timp ce Realitatea este conectată cu rezistența la experiența noastră umană. Realul este, prin definiție, închis pentru totdeauna (nu tolerează alte calificări), în timp ce Realitatea este accesibilă cunoștințelor noastre. Realul implică nonrezistență, în timp ce Realitatea implică rezistență* [164, p. 13].

În măsura în care Realitatea participă la ființarea lumii, trebuie să atribuim și această dimensiune ontologică acestui concept. Realitatea nu este doar o construcție socială, consensul unei colectivități sau un acord inter-subiectiv. Ea are de asemenea, o dimensiune trans-subiectivă. De exemplu, datele experimentale pot distruge cea mai frumoasă teorie științifică. Semnificația pe care o acordăm cuvântului *realitate* este, așadar, pragmatică și ontologică în același timp. Realitatea nu poate fi redusă la substanță. Pentru fizicienii de astăzi, acest lucru este evident. Materia este **complexul** substanță-energie-spațiu-timp-informație.

Marele fizicianul german Werner Karl Heisenberg distinge trei niveluri de Realitate: primul nivel al Realității, conform modelului lui Heisenberg, corespunde stărilor lucrurilor care sunt obiectivate independent de procesul de cunoaștere. La acest prim nivel, el plasează mecanica clasică, electromagnetismul și teoriile lui Einstein despre relativitate - cu alte cuvinte, fizica clasică. Al doilea nivel al realității corespunde stărilor lucrurilor care sunt inseparabile de procesul de cunoaștere. El pune aici mecanica cuantică, biologia și științele cognitive. În cele din urmă, al treilea nivel al Realității corespunde stărilor lucrurilor create în legătură cu procesul de cunoaștere. El situează pe acest nivel de realitate: filosofia, arta, politica, metaforele despre „Dumnezeu", experiența religioasă și experiența inspirației [164, p. 16].

O clasificare asemănătoare a nivelurilor de realitate o găsim la Adrian Iosif [86]. *Putem vorbi – ținând cont de stadiul actual al științei – de existența a cel puțin trei niveluri de Realitate: unul corespunzător legilor și conceptelor mecanicii clasice, un nivel caracterizat de legile și conceptele mecanicii cuantice și un nivel al ciberspaţiu-timpului* [86, p. 178]. Pentru noi un interes deosebit prezintă nivelul ciberspaţiu-timp. În contextul dezvoltării tehnologiilor informaționale în special al Realității virtuale acest nivel prinde tot mai mult contur. Realitatea virtuală nu este același lucru cu ciberspaţiu-timpul, care în esență este independent de orice tip de tehnologie și

interdependent cu celelalte niveluri de realitate. Odată cu apariția tehnologiei de Realitate virtuală putem percepe acest nivel al realității mult mai ușor în comparație cu generațiile anterioare care nu au avut parte de astfel de tehnologii avansate. O caracteristică esențială a ciberspaţiu-timpului, conform cu Basarab Nicolescu, este *capacitatea maximală de interacțiune real-imaginar, concret-abstract, corp-ecuații matematice. Ciberspaţiu-timpul poate deci, în principiu, să pună în evidență un nou nivel de percepție* [92, p. 92].

Realitatea are, conform modelului transdisciplinar, un anumit număr de niveluri (finit sau infinit). Vom presupune că acest număr este infinit.

Nivelurile adiacente sunt legate prin logica terțului inclus (***Fig. 1. 2***), în sensul că starea T prezentă la un anumit nivel este legată de o pereche de contradictorii (A, non-A) de la nivelul imediat învecinat. Starea T realizează unificarea contradictoriilor A și non-A, dar această unificare se realizează la un nivel diferit față de cel în care se află A și non-A. Axioma noncontradicției este respectată în acest proces [91].

Există cu siguranță o coerență între diferitele nivele ale Realității, cel puțin în lumea naturală. O autocoerență extinsă, arată Basarab Nicolescu, pare să guverneze evoluția universului, de la infinit de mic la infinit de mare, de la infinit de scurt la infinit de lung. Exemplul pe care îl dă autorul este preluat din fizica modernă. O mică variație a constantei de cuplaj a interacțiunilor puternice între particulele cuantice ar duce, la nivelul universului nostru, fie la conversia întregului hidrogen în heliu, fie la inexistența atomilor complecși precum carbonul. La fel, o mică variație a constantei de cuplaj gravitațional ar duce fie la planete efemere, fie la imposibilitatea formării lor. Mai mult, conform teoriilor cosmologice actuale, Universul pare capabil să se autogenereze fără nicio intervenție externă. Metodologia transdisciplinară arată că un flux continuu de informație se transmite coerent de la un nivel de Realitate la alt nivel de Realitate al universului fizic.

Unul din punctele importante ale metodologiei transdisciplinare constă în logica terțului inclus. Logica terțului inclus este capabilă, după părerea lui Basarab Nicolescu, să descrie coerența între nivelurile Realității prin intermediul unui proces iterativ, care constă din următoarele etape (***Fig. 1. 2***) [91, p. 238]:

1. Un cuplu de concepte contradictorii (A, non-A) situat la un anumit nivel de realitate este unificat de o stare T situată pe un nivel imediat următor al Realității;

2. Această stare T este legată de un cuplu de contradictorii (A', non-A'), situat pe propriul său nivel;

3. Cuplul de contradictorii (A', non-A') este, la rândul său, unificat de o stare T' situată la un nivel diferit de Realitate, imediat vecin cu cel în care se află terțul inclus (A', non-A', T).

Acest procesul iterativ continuă, până la epuizarea tuturor nivelurilor de Realitate.

Acțiunea logicii terțului inclus asupra diferitelor nivele ale Realității induce o structură deschisă, pe care o putem numi gödeliană, a ansamblului nivelurilor de Realitate. Această structură gödeliană are o importanță considerabilă în teoria cunoașterii, implicând imposibilitatea unei teorii complete [91].

Starea T realizează, în conformitate cu axioma non-contradicției, unificarea cuplului de contradictorii (A, non-A), dar este asociată, în același timp, cu un alt cuplu de contradictorii (A', non-A'). Aceasta înseamnă că putem construi, pornind de la un număr de cupluri reciproc exclusive, o nouă teorie care elimină contradicțiile la un anumit nivel de Realitate, dar această teorie este doar temporară, deoarece va conduce inevitabil, sub presiunea combinată a teoriei și a experienței, la descoperirea unor noi cupluri de contradictorii, situate la un nou nivel de Realitate. Această teorie va fi, la rândul ei, înlocuită pe măsură ce noi niveluri de Realitate vor fi descoperite, de teorii încă mai unificate. Acest proces va continua la infinit, fără a putea ajunge vreodată la o teorie complet unificată [91].

Putem vorbi, astfel, despre o evoluție a cunoașterii, fără a ajunge vreodată la o non-contradicție absolută, care implică toate nivelurile Realității. Cunoașterea este deschisă pentru totdeauna. În lumea nivelurilor Realității în sine, ceea ce este sus este ca și ceea ce este jos, dar ceea ce este jos nu este ca și ceea ce este sus. Gradele de materialitate induc o săgeată de orientare a transmiterii informației de la un nivel la altul. În acest sens, arată Basarab Nicolescu, „ceea ce este jos nu este ca și ceea ce este sus", cuvintele „sus" și „jos" neavând aici nicio altă semnificație (spațială sau morală) decât cea topologică asociată săgeții transmiterii informației [91].

Structura deschisă a ansamblului nivelurilor Realității este în concordanță cu unul dintre cele mai importante rezultate ale secolului XX, din domeniul fundamentelor aritmeticii: teorema lui Gödel, care ne spune că un sistem de axiome suficient de bogat duce inevitabil la rezultate fie nedecidabile, fie contradictorii [91].

Consecința teoremei lui Gödel are o importanță considerabilă pentru orice teorie modernă a cunoașterii. În primul rând, consecința are legătură cu întreaga matematică care include aritmetica. Este evident că, matematica, care este instrumentul de bază al științei moderne, conține aritmetica. Aceasta înseamnă că orice căutare a unei teorii complete este iluzorie. Dacă această afirmație este valabilă pentru domeniile cele mai riguroase ale studiului sistemelor naturale, ea este cu atât mai mult valabilă pentru un domeniu infinit mai complex - acela al științelor umane [92].

Căutarea unei axiomatizări care să conducă la o teorie completă, fără rezultate nedecidabile sau contradictorii, marchează, în același timp, atât apogeul, cât și punctul de declin al gândirii clasice.

Basarab Nicolescu subliniază faptul că structura gödeliană a întregului ansamblu al nivelurilor Realității, asociată logicii terțului inclus, implică imposibilitatea construirii unei teorii complete pentru a descrie trecerea de la un nivel la altul și pentru a descrie întregul ansamblu al nivelurilor Realității [91].

Se poate vorbi de o coerență a întregului ansamblu al nivelurilor Realității, dar această coerență are totuși o orientare: o săgeată este asociată oricărei transmiteri de informație de la un nivel la altul. Așadar, coerența se oprește la nivelul cel mai „înalt" și nivelul cel mai „jos". Pentru ca ea să continue dincolo de aceste două nivele limită, pentru a conserva o unitate deschisă, fizicianul român accentuează faptul că este absolut necesar să considerăm că întregul ansamblu al nivelurilor Realității se prelungește printr-o zonă de non-rezistență la experiențele noastre, reprezentări, descrieri, imagini sau formalizări matematice. Această zonă de non-rezistență corespunde, în modelul transdisciplinar de Realitate, cu ceea ce Bernard d'Espagnat numește „realitatea voalată" [91]. Nivelul cel mai „înalt" și nivelul cel mai „jos" al întregului ansamblu al nivelurilor Realității se unesc printr-o zonă de transparență absolută. Unitatea deschisă a lumii implică faptul că ceea ce este „jos" este ca și ceea ce este „sus". Izomorfismul între „sus" și „jos" este restabilit de zona de non-rezistență.

Non-rezistența acestei zone de transparență absolută, arată mai departe Basarab Nicolescu, se datorează pur și simplu limitărilor corpului nostru și ale organelor noastre de simț, indiferent de instrumentele de măsură care extind aceste organe de simț [92]. Afirmarea unei cunoașteri umane infinite care exclude orice zonă de non-rezistență, în timp ce afirmă limitarea corpului nostru și a organelor noastre de simț, nu are nicio bază epistemologică. În viziunea academicianului român, această zonă de non-rezistență corespunde sacrului, definit, de același autor, ca fiind ceea ce nu se supune niciunei raționalizări (desigur, această definiție unilaterală a sacrului trebuie considerată având în vedere abordarea transdisciplinară, ea fiind mult mai generală decât definițiile teologice sau cele din istoria ideilor religioase). Proclamarea existenței unui singur nivel al Realității, în contextul metodologiei transdisciplinare, elimină complet sacrul.

Ansamblul nivelurilor Realității și zona lor complementară de non-rezistență constituie ceea ce Basarab Nicolescu numește Obiectul transdisciplinar [91; 92].

În viziunea transdisciplinară, pluralitatea complexă și unitatea deschisă sunt două aspecte ale unei singure și aceleiași Realități.

Ca o consecință a aplicării metodologiei transdisciplinare apare, în viziunea academicianului Nicolescu, un nou Principiu al Relativității, rezultat al coexistenței dintre pluralitatea complexă și unitatea deschisă. Niciun nivel al Realității nu constituie un loc privilegiat de unde s-ar putea înțelege toate celelalte niveluri ale Realității. Un nivel al Realității este ceea ce

este pentru că toate celelalte niveluri există în același timp. Acest Principiu al Relativității este fundamentul unei noi perspective asupra educației, artei, politicii, religiei și vieții sociale.

Basarab Nicolescu [91] accentuează de mai multe ori faptul că diferitele niveluri ale Realității sunt accesibile cunoașterii umane datorită existenței diferitelor niveluri de percepție, care se află în corelație biunivocă cu nivelurile Realității. Așadar, în viziunea transdisciplinară, Realitatea este, în același timp, multidimensională și multireferențială. Nivelurile de percepție permit o viziune din ce în ce mai generală și mai unificatoare asupra Realității, fără însă, are grijă să specifice fizicianul român, a o epuiza în întregime.

Coerența nivelurilor de percepție presupune, urmând o structură similară celei a nivelurilor de Realitate, o zonă de non-rezistență față de percepție. Ansamblul nivelurilor de percepție și zona lor complementară de non-rezistență constituie, pentru Basarab Nicolescu, Subiectul transdisciplinar.

Cele două zone de non-rezistență ale Obiectului și Subiectului transdisciplinare trebuie să fie identice pentru că doar în aceste condiții Subiectul transdisciplinar poate comunica cu Obiectul transdisciplinar. Fluxul informațional care traversează coerent nivelurile de Realitate corespunde unui flux de conștiință care traversează coerent nivelurile de percepție. Cele două fluxuri se află într-o relație de izomorfism datorită existenței unei singure zone de non-rezistență. Cunoașterea nu este, din perspectiva transdisciplinară, nici exterioară, nici interioară, ci este, simultan, atât exterioară, cât și interioară. Zona de non-rezistență joacă rolul a ceea ce Basarab Nicolescu [92] numește terțul tainic inclus, care permite unificarea completă, dar o unificare care păstrează diferențele, a Subiectului transdisciplinar și a Obiectului transdisciplinar.

Diagrama din ***Fig. 1. 3*** ilustrează abordarea transdisciplinară asupra Naturii și a cunoașterii, conform lui Nicolescu [91, p. 237]. În această reprezentare, nivelurile de Realitate sunt notate simbolic în partea stângă: {NRn, ..., NR_2, NR_1, NR_0, NR_{-1}, NR_{-2}, ..., NR-n}, unde indicele n poate avea valori finite sau infinite.

Această schemă ne invită să explorăm o realitate multidimensională, structurată pe niveluri multiple, departe de concepția unidimensională și monocromă a gândirii clasice. În cadrul acestei reprezentări, două niveluri adiacente (de exemplu, NR_0 și NR_1) sunt conectate prin intermediul logicii terțului inclus, care reprezintă o abordare inovatoare față de logica clasică.

Starea terțul inclus Tx, situat la nivelul NR_1, este conectat cu un set de opoziții (A_0 și non-A_0) aflate la un nivel proxim. Deși Tx unifică aceste contradicții, unificarea are loc la un nivel diferit de cel pe care se află A_0 și non-A_0, astfel respectându-se axioma noncontradicției.

Logica terțului inclus evidențiază coerența dintre nivelurile de Realitate printr-un proces iterativ care se desfășoară așa: un set de contradictorii (A_0, non-A_0), aflat la nivelul de realitate

NR_0, este unificat de o stare T_1 la nivelul NR_1. Aceasta, la rândul ei, se conectează cu un alt set de contradictorii (A_1, non-A_1) de la nivelul său. Procesul de unificare continuă cu starea T_2, care se află la nivelul NR2, și așa mai departe, explorând toate nivelurile de Realitate posibile, fără a ajunge vreodată la o teorie complet unificată.

Conform lui Nicolescu, cunoașterea este un proces evolutiv și perpetuu, caracterizat de imposibilitatea atingerii unei noncontradicții absolute ce ar include toate nivelurile de Realitate. Prin urmare, cunoașterea rămâne o călătorie infinită de explorare și înțelegere.

În ***Fig. 1. 3*** observăm că nivelurile variate de Realitate devin accesibile prin intermediul diverselor niveluri de percepție. Acestea din urmă corespund biunivoc cu nivelurile de Realitate și sunt ilustrate în partea dreaptă a figurii. Reprezentate simbolic ca (NPn, ..., NP_2, NP_1 NP_0, NP_{-1}, NP_{-2}, ..., NP-n), aceste niveluri de percepție ne oferă o perspectivă tot mai cuprinzătoare asupra Realității, deși nu ne permit să o cunoaștem în întregime.

Conform lui Basarab Nicolescu, transdisciplinaritatea implică transcenderea dualităților clasice, cum ar fi subiect-obiect, subiectivitate-obiectivitate și altele. Acest model al cunoașterii aduce un aport semnificativ în analiza complexității sistemelor. Complexitatea infinită a obiectului transdisciplinar se îmbină cu simplitatea infinită a subiectului transdisciplinar, echivalând cu modul în care complexitatea unui nivel specific de Realitate poate fi simplificată prin integrarea într-un alt nivel.

Există, însă, o discuție referitoare la termenul de "simplitate infinită a Subiectului", care ar putea avea o conotație mai degrabă metaforică [41]. Abordarea transdisciplinară subliniază o unitate deschisă între obiectul și subiectul transdisciplinar, evidențiată prin coordonarea fluxurilor de informație și conștiință ce traversează respectivele niveluri.

Un element crucial în această perspectivă este punctul X, prezentat în ***Fig. 1. 3***, unde se intersectează cele trei bucle ale informației și conștiinței. Acest punct emblematic reprezintă interacțiunea între Subiect și Obiect, un element distinct care nu poate fi redus la niciunul dintre cei doi termeni. Această structură ternară {Subiect, Obiect, Interacțiune} contrastează cu dualitatea {Subiect, Obiect} specifică metafizicii moderne, indicând o ruptură decisivă introdusă de abordarea transdisciplinară.

Noțiunea transdisciplinară a nivelurilor de realitate este incompatibilă cu reducerea nivelului spiritual la nivelul psihic, a nivelului psihic până la nivelul biologic și a nivelului biologic până la nivelul fizic. Chiar și așa, cele patru niveluri sunt unite prin **Terțul Ascuns** (punctul X) care este un termen de interacțiune ce permite unificarea Subiectului transdisciplinar și a Obiectului transdisciplinar, păstrând în același timp diferențele și care nu poate fi descris de o teorie științifică. Prin definiție, știința exclude nonrezistența. Știința, așa cum este definită astăzi,

este limitată de metodologia ei proprie. Zona de nonrezistență joacă rolul unui terț între Subiectul și Obiect. *Cele două zone de nonrezistență ale Obiectului și Subiectului transdisciplinare trebuie să fie identice pentru ca Subiectul transdisciplinar să poată să comunice cu Obiectul transdisciplinar. Fluxului de informație care traversează in mod coerent diferitele niveluri de realitate îi corespunde un flux de conștiință care traversează în mod coerent diferitele niveluri de percepție. Cele două fluxuri sunt într-o relație de izomorfism datorită uneia și aceleiași zone de nonrezistență. Cunoașterea nu este deci nici exterioară, nici interioară; ea este în același timp exterioară și interioară: studiul Universului și studiul ființei umane se susțin unul pe celălalt* [91, p. 239-240].

În esență, metodologia transdisciplinarității este întemeiată pe următoarele trei axiome (Basarab Nicolescu le numește „axiome", poate că termenul de „principiu" ar fi mai potrivit):

1. **Axioma ontologică:** există diferite niveluri de Realitate ale Obiectului și, în mod corespunzător, diferite niveluri de Realitate ale Subiectului;
2. **Axioma logică:** trecerea de la un nivel al Realității la celălalt este asigurată de logica terțului inclus;
3. **Axioma epistemologică:** structura totalității nivelurilor de realitate apare în cunoașterea noastră a naturii, a societății și a propriei persoane, ca o structură complexă: fiecare nivel este ceea ce este deoarece toate nivelurile există în același timp [92].

Pornind de la aceste trei axiome, vom construi în cadrul acestei lucrări un model particularizat care să explice conceptul de inteligență artificială din perspectiva transdisciplinară.

1.5 Concluzii

În prima parte a lucrării putem extrage câteva concluzii preliminare care conturează o imagine cât de cât clară asupra locului pe care îl ocupă inteligența artificială în prezent. Deși am văzut succese tehnice impresionante în jocuri de strategie precum șahul sau Go, sau procesarea limbajului natural, observăm că acestea rămân performanțe în nișe înguste, fără a atinge încă pragul unei inteligențe generale care să aibă flexibilitatea gândirii umane.

Este esențial să înțelegem că fundamentele științifice ale inteligenței artificiale, oricât de complexe ar fi, definesc un sistem care funcționează strict pe bază de reguli și proceduri. Am observat că, spre deosebire de inteligența umană, inteligența artificială rămâne limitată de cadrele logicii formale. Prezența unor bariere matematice precum teorema lui Gödel ne întărește convingerea că nicio mașină, indiferent de puterea de calcul, nu va putea epuiza sau demonstra toate adevărurile realității noastre.

La fel, analizând impactul acestei tehnologii asupra societății, am constatat că inteligența artificială nu este doar un subiect de studiu pentru informaticieni, ci o tehnologie care provoacă transformări profunde în medicină, economie și chiar în modul în care ne raportăm la divinitate sau la propria identitate. De aceea, propunerea unei metodologii transdisciplinare ni se pare justificată și viabilă pentru a face ordine în această discuție. Prin acceptarea mai multor niveluri de realitate și a logicii terțului inclus, înțelegem că inteligența artificială aparține unui nivel de realitate specific. În acest moment putem sugera că acești algoritmi rămân un instrument complementar, o tehnologie sofisticată care trebuie gestionată cu discernământ uman.

2. DIMENSIUNI CONCEPTUAL-METODOLOGICE ALE CERCETĂRII: STUDII DE CAZ

2.1 Arhitectura și funcționarea rețelelor neuronale artificiale

În cadrul inteligenței artificiale problemele de învățare ocupă un loc aparte. Aceste preocupări se constituie într-o direcție distinctă de cercetare, cea a mașinilor autoinstruibile (*machine learning*).

Învățarea automată reprezintă studiul modelelor de sisteme capabile să-și îmbunătățească performanțele semnificativ, utilizând o mulțime de date de instruire.

Se pot distinge trei direcții majore de cercetare relativ la mașinile autoinstruibile: modelarea neuronală și tehnicile bazate pe teoria deciziilor; achiziția simbolică a conceptelor; învățarea bazată pe cunoaștere a domeniului specific.

Modelarea neuronală încearcă să dezvolte sisteme instruibile pentru scopuri generale, având în vedere faptul că aceste sisteme pornesc cu o cantitate mică de cunoștințe inițiale. Astfel de sisteme se numesc **rețele neuronale**, sisteme cu autoorganizare sau sisteme conexioniste [32; 31; 55].

Capacitatea rețelelor neuronale de a învăța cum să rezolve o problemă are o importanță deosebită în cazul multor aplicații și pot oferi soluții neașteptate pentru probleme ivite în practică. Posibilitatea de a prelucra în paralel și, deci, foarte rapid, informații prin conectarea în rețea a unor elemente procesoare simple, sub forma unor anumite structuri care se comportă redundant și care sunt din această cauză tolerante la erori, face ca datorită capacității de a învăța, rețelele neuronale să fie adecvate pentru a fi întrebuințate la reglarea automată a unor complexe care necesită o atenție deosebită și în care este necesară prelucrarea rapidă a unui flux mare de informație.

Structura rețelelor neuronale este concepută după modelul prelucrării informațiilor în domeniul biologic, unde strategia de rezolvare a problemelor se învață tot prin exemple. Astfel, omul reușește să realizeze reglarea și coordonarea mișcărilor fără probleme, cu toate că nu are o reprezentare matematică a coordonatelor. La nivelul reglajului senzomotric, omul învață să execute operațiile necesare prin încercări, în funcție de situația dată. Ideea de a întrebuința cel puțin unele părți ale acestei însușiri, de care dispun sistemele biologice, pentru rezolvarea unor probleme tehnice în modelarea interfeței om-mașină este foarte interesantă.

În ultimul deceniu a avut loc o creștere considerabilă a interesului pentru domeniul rețelelor neuronale artificiale (ținând cont că prima referire la acest tip de rețele datează din anul 1943).

Calculatoarele din zilele noastre devin tot mai performante și mai capabile să realizeze stocarea și prelucrarea unui volum mare de informații. Circuitele lor operează cu viteze extrem de mari și pot realiza calcule matematice oricât de complicate, fără erori semnificative.

Dar, oamenii sunt capabili să realizeze anumite sarcini pe care nici un sistem de inteligență artificială, indiferent pe ce calculator ar fi implementat, nu le poate nici măcar modela cu o anumită aproximație și cu atât mai puțin să le realizeze (de exemplu mersul biped, vorbirea însoțită de raționare, recunoașterea fără erori a unor obiecte etc.).

Se pare că structura și modul de funcționare ale sistemului nervos, care este cel care comandă toate activitățile amintite mai sus, sunt potrivite pentru acțiunile mai „umane" și nu pentru cele care necesită o mare putere de calcul [143]. De aceea, impunându-se sistemelor dotate cu inteligență artificială anumite constrângeri, sugerate de structura și funcționarea sistemului nervos, este posibil (cel puțin aceasta se speră) ca ele să fie capabile de performanțe asemănătoare celor ale omului.

Rețelele neuronale artificiale au ca punct de inspirație sistemul nervos. Unitatea de organizare a sistemului nervos este neuronul, o celulă care prezintă un număr de dendrite și un axon, prin intermediul cărora se interconectează cu alți neuroni. Dendritele constituie intrările în celula neuronală, iar axonul ieșirea. Axonul se ramifică astfel încât neuronul poate avea conectată ieșirea la mai mulți neuroni (***Fig. 2. 1***) [55; 31].

Totalitatea impulsurilor prezentate la intrarea neuronului îl pot excita ca acesta să genereze un impuls mai departe spre neuronii cu care este conectat. Trebuie menționat că legăturile între neuroni sunt ponderate și fiecare neuron aplică o transformare asupra impulsului de la intrări înainte de a-l transmite mai departe.

Așadar, o rețea neuronală artificială este alcătuită dintr-o mulțime de noduri în care se află neuronii artificiali, elementele de procesare neliniare, care operează în paralel. Prin analogie cu neuronul biologic, un neuron artificial are un anumit număr de intrări și o singură ieșire, care se poate conecta la intrarea mai multor neuroni. Fiecare intrare are asociată o anumită pondere care reprezintă importanța pe care o are impulsul, prezent pe linia respectivă, la activarea neuronului.

Cu toate că asemănarea între sistemul nervos și rețelele neurale artificiale este relativ mică, estimată de unii autori cu mult sub 10% [55], rețelele neurale artificiale prezintă un număr surprinzător de mare de caracteristici ale creierului.

De exemplu, rețelele neuronale pot învăța din experiență, pot generaliza din anumite exemple altele noi și, de asemenea, pot sintetiza caracteristicile esențiale din intrări ce conțin și date irelevante, ceea ce cu algoritmii clasici este foarte dificil, dacă nu chiar imposibil.

Un mare avantaj al rețelelor neurale artificiale este că pot să descrie o problemă și să o rezolve în același timp, prin autoorganizarea lor și NU printr-un program [55; 32]. Acest proces de autoorganizare are loc pe parcursul unui proces de învățare obținut prin cooperarea unei topologii inițiale, a unor reguli de învățare și a unui număr mare de antrenamente.

Caracteristicile cele mai importante ale rețelelor neurale artificiale sunt [31]:

Capacitatea de a învăța. Rețelele neurale artificiale nu necesită programe puternice, ci sunt mai degrabă rezultatul unor antrenamente având la bază un set de date. Dându-se un vector de intrări, alături cu ieșirile dorite (etichetele), în urma procesului de antrenament rețelele neurale se autoorganizează, astfel că pot rezolva problemele pentru care au fost proiectate. S-a dezvoltat o gamă foarte largă de algoritmi de antrenament, fiecare având avantajele și dezavantajele sale, putându-se aplica cu succes în unele cazuri sau fără nici o șansă de reușită în altele.

Capacitatea de generalizare. Dacă au fost antrenate corespunzător, rețelele sunt capabile să dea răspunsuri corecte și pentru intrări diferite față de cele cu care au fost antrenate, atât timp cât aceste intrări nu sunt foarte diferite. Este important de menționat că rețelele neurale artificiale generalizează automat **ca urmare a structurii lor, nu cu ajutorul inteligenței omului înglobată într-un anumit program special creat în acest scop** [143]. În cazul rețelelor neurale, inteligența umană este înglobată în alegerea structurii rețelei (numărul de straturi ascunse, topologia rețelei, numărul de intrări), funcției de activare, algoritmului de antrenare, setului de antrenament etc.

Capacitatea de sinteză. Rețele neuronale artificiale pot lua decizii sau trage concluzii când sunt confruntate cu informații complexe sau cu zgomote irelevante sau parțiale (de exemplu, în recunoașterea de caractere distorsionate). Deci, se poate spune că rețeaua a învățat să producă ceva ce n-a mai văzut înainte.

În ciuda caracteristicilor atât de mult lăudate ale rețelelor neurale artificiale, nici chiar cei mai înflăcărați și optimiști cercetători în acest domeniu nu se încumetă, încă, să afirme că rețelele neurale artificiale vor ajunge prea curând la performanțele creierului uman. Totodată, nu trebuie minimizată sau ignorată importanța rețelelor neurale artificiale și este posibil ca, în viitor, cu ajutorul lor să se ajungă la o cunoaștere mai aprofundată a fenomenelor ce au loc în creierul uman. Ceea ce recomandă rețelele neurale artificiale este raportul favorabil dintre performanțele pe care le realizează și complexitatea lor. Acest raport este în continuă creștere și depășește, în acest moment, pe cel al sistemelor de inteligență artificială implementate cu alte tipuri de tehnologii [162].

Rețelele neuronale sunt formate dintr-un număr de elemente procesoare simple (neuronii artificiali) care, așezate într-o anumită structură, prelucrează informațiile în paralel. Rețelele neuronale nu sunt programate, ci dezvoltă singure soluția problemei, pe bază de exemple, cu

ajutorul unei legi de învățare. Structura lor se orientează după exemplul de prelucrare a informației în creierul biologic, unde căile de rezolvare a unei probleme sunt învățate tot prin exemple [32].

Rețelele neuronale artificiale sunt constituite ca structuri paralele de prelucrare a informației, unde elementele procesoare sunt legate între ele prin conexiuni, prin care informația este transmisă într-o singură direcție. Fiecare element procesor (neuron) are o memorie locală pentru factorii de „pondere" a conexiunilor de intrare, care pot fi modificați în faza de învățare a rețelei neuronale [55; 31].

Elementele procesoare (neuronii artificiali) prelucrează numai informații locale disponibile sub formă de semnale imediate, transmise prin conexiunile de intrare și care sunt înmulțite cu factorul de „pondere" al conexiunii respective. Elementele procesoare (neuronii) calculează un singur semnal de ieșire, care este transmis ca semnal de intrare altor elemente procesoare (neuroni), în funcție de ramificațiile structurii rețelei.

Dispunerea elementelor procesoare se face în straturi, denumite și „*layers*" (***Fig. 2. 2***). Elementele procesoare (neuronii) dintr-un anumit strat sunt legate în rețea numai cu neuronii stratului următor, prin intermediul „ponderilor" de conectare. De cele mai multe, în tipul de arhitectură cel mai des întâlnit, conexiunile nu se fac niciodată între elementele procesoare (neuronii) din același strat și nici cu cele din stratul anterior.

Cunoștințele rețelei neuronale se află stocate în „ponderile" conexiunilor cu care sunt modulate semnalele de intrare ale neuronilor pentru obținerea rezultatului. Aceste „ponderi" sunt actualizate în timpul fazei de învățare a rețelei. În această fază, rețeaua se antrenează pentru rezolvarea problemei. Se poate vorbi, deci, de o fază de antrenament.

Pe baza celor prezentate, rețelele neuronale artificiale sunt capabile să învețe, adică să acumuleze cunoștințe cu privire la relațiile funcționale, acumulând și stocând experiență. De asemenea, ele sunt capabile să prelucreze aceste cunoștințe în funcție de fiecare semnal de intrare și să furnizeze la ieșire rezultatele care vor decide tipul de acțiune asupra obiectului. Prin învățare continuă (on-line) se poate realiza o intervenție dirijată asupra obiectului supus controlului [143; 55].

În faza de antrenare se urmărește să se achiziționeze cât mai multe cunoștințe preliminare. Aceste cunoștințe sunt perfecționate permanent prin învățare continuă în faza de funcționare propriu-zisă a regulatorului neuronal și ele sunt adaptate cât mai bine condițiilor existente (***Fig. 2. 3***).

Cunoștințele suplimentare dobândite astfel în faza de funcționare, împreună cu cele preliminare (din faza de antrenare), vor acționa din nou asupra comportamentului rețelei neuronale în faza de funcționare a acesteia [143; 31].

Ca și la om, în cazul rețelei neuronale artificiale, performanțele depind de felul și volumul cunoștințelor acumulate experimental în timpul fazei de antrenament și în timpul fazei de funcționare.

Pregătirea (faza de antrenament în cazul rețelelor neuronale), precum și perfecționarea în timpul funcționării sunt de importanță hotărâtoare, deoarece în această fază are loc o adaptare permanentă a comportamentului la condițiile variabile din mediul cu care interacționează rețeaua (***Fig. 2. 4***). Aceasta se realizează prin învățare și NU prin reprogramare [55].

Neuronul artificial este o copie mult simplificată a neuronului biologic. Este format dintr-un corp (considerat ca o cutie neagră, un set de intrări și o ieșire. Fiecare intrare este ponderată, adică valoarea sa este multiplicată cu o valoare corespunzătoare, numită pondere, apoi toate intrările ponderate sunt însumate.

Asupra sumei obținute se aplică o funcție de activare rezultând valoarea ieșirii neuronului respectiv (vezi ***Fig. 2. 5***, structura neuronului artificial) [55].

Funcția F poate să fie o simplă funcție treaptă

$$F(x) = \begin{cases} 1\ , & x > t \\ 0\ , & x \leq t \end{cases}$$

sau o funcție mai complicată, cu rol de a restrânge domeniul de variație al valorii de ieșire a neuronului, indiferent de valoarea netă (notată cu *net*, valoarea de intrare în funcția de activare a neuronului, după notațiile din ***Fig. 2. 5***). Una dintre cele mai uzuale funcții F este funcția numită *sigmoidă*:

$$F(x) = \frac{1}{1+e^{-x}}$$

În ***Fig. 2. 5*** este reprezentată schema neuronului artificial, în care *inp1, inp2, ..., inpn* reprezintă valorile intrărilor, *w1, w2, ..., wn* reprezintă ponderile cu care se multiplică fiecare intrare în parte, *net* este ieșirea netă din neuron, iar *out* este ceea ce rezultă după prelucrarea ieșirii nete cu funcția de activare (*net* devine variabilă de intrare pentru funcția de activare, iar out este rezultatul calculului realizat de această funcție).

Cu notațiile din ***Fig. 2. 5***, formula funcției de activare de tip sigmoidă devine:

$$out = \frac{1}{1+e^{-net}}\ .$$

De asemenea, cu notațiile din ***Fig. 2. 5***, sunt valabile formulele:

$$net = \sum_{i=1}^{n} inp_i\ w_i$$
$$out = F(net)$$

Trebuie subliniat un lucru foarte important, în contextul subiectului tratat în această lucrare: nu există o teorie care să determine cu certitudine care funcție de activare este mai potrivită pentru o anumită aplicație. Răspunsul la o astfel de întrebare se găsește doar cu ajutorul experimentelor [143].

Se poate observa că modelul de neuron artificial din ***Fig. 2. 5*** nu reproduce anumite caracteristicile importante întâlnite la neuronul biologic. Nu este luat în considerare timpul de întârziere care afectează dinamica sistemului, pentru că, în cazul neuronului artificial, intrarea determină o valoare de ieșire imediată. Nu se ține cont nici de modulația în frecvență a neuronului și nici de alte caracteristici ale neuronului biologic [55].

Cu toate aceste simplificări ale neuronului artificial față de cel biologic, rețelele neurale artificiale se comportă în anumite cazuri destul de asemănător cu sistemul nervos. Explicația ar putea fi că neuronul artificial înglobează caracteristicile fundamentale ale celui biologic sau că, pur și simplu, este vorba de o coincidență (variantă care este totuși luată în considerație de unii specialiști în domeniu).

Din perspectivă filosofică, și în special a filosofiei cognitiviste, rețelele neuronale artificiale intră în categoria abordării conexioniste. În continuare, vom analiza relevanța filosofică a rețelelor neuronale având în vedere paradigma conexionistă [162].

Abordarea conexionistă examinează modul în care gândirea și procesele cognitive pot fi înțelese prin intermediul rețelelor și conexiunilor neuronale. Această perspectivă se bazează pe ideea că gândirea și procesele cognitive sunt rezultatul interacțiunilor complexe dintre neuroni și conexiunile acestora în creier.

Ideea de bază a conexionismului este aceea că o rețea strâns legată de procesoare simple de tip neuronal se poate comporta, atunci când este confruntată cu anumite date de intrare, într-un mod consistent pentru a găsi date de ieșire, și că modul cel mai bun de a obține datele de ieșire corecte nu este acela de a prezice regula de calculare a lor, ci mai curând de a lăsa sistemul să afle răspunsul corect încercând diferite conexiuni în rețea până când se stabilizează asupra celor care dau rezultatul corect. Atitudinea de tip *bottom-up* a conexioniștilor situează programul lor în afara inteligenței artificiale tradiționale (prelucrarea de nivel înalt nu poate face abstracție de hardware, necesitatea unor arhitecturi paralele, procesare distribuită și relativa lipsă de instrucțiuni specifice pentru mașinile conexioniste).

Conexioniștii consideră că mintea umană poate fi înțeleasă prin intermediul simulărilor computerizate ale rețelelor neuronale, care pot reproduce, cel puțin într-o anumită măsură, comportamentul cognitiv uman. Aceste simulări sunt create utilizând algoritmi și modele matematice care imită structura și funcționarea rețelelor neuronale din creier [107].

Prin urmare, conexioniștii investighează cum informația este prelucrată, stocată și transmisă prin conexiunile neuronale, în încercarea de a înțelege procesele cognitive și de a crea modele care să reproducă aceste procese. Această abordare are, în acest moment, un impact semnificativ în domeniul inteligenței artificiale și în înțelegerea mecanismelor cognitive.

Alături de conexionism există și alte abordări și teorii care explorează natura minții și a proceselor cognitive din alte perspective: funcționalismul, reprezentaționalismul sau fenomenologia cognitivă.

Claudiu Ioniță în teza sa de doctorat „*Incomensurabilitatea paradigmelor - O perspectivă cognitivistă"* [54, pp. 96-97] este de părere că *„schemele mentale merită a fi analizate în termeni conexionişti, iar această analiză trebuie să țină cont de avantajele pe care le presupune deschiderea pe nivelul de implementare pe care îl are conexionismul încă de la origine."*

Autorul explică mai departe că premisele subiacente ale conexionismului implicat în arhitectura rețelelor neuronale, nu doar că oferă aceleași avantaje ca și mediul de abordare a problemelor inteligenței artificiale folosind programarea clasică, ci are și alte trei avantaje incontestabile care subliniază plauzibilitatea sa biologică. În primul rând, toleranța față de eroare, specifică organismelor. Relațiile stohastice dintre stările mașinii recapitulează relațiile stohastice dintre evenimentele mediului pe care le reprezintă. Acest fapt face ca rețeaua să atingă o configurație stabilă prin mecanisme statistice, mecanism care tolerează erorile particulare. În al doilea rând, toleranța față de degradare. eliminarea unor noduri (neuroni) sau a unor legături (sinapse) nu duce la blocarea funcționării rețelei și nu afectează în mod semnificativ performanțele acesteia. În al treilea rând, întărirea conexiunilor. Încă din anii 1950, regula lui Hebb a consacrat ideea că o creștere a schimbului de semnal între neuroni va întări legătura dintre aceștia. Dovezile experimentale au relevat că grupuri extinse de neuroni asociate reprezentărilor vizuale acționează competitiv în cazul unei discrepanțe oculare. Neuronii care asigură conexiunile corespunzătoare semnalului provenit de la ochiul funcțional vor prelua o mare parte din conexiunile ochiului nefuncțional. Extinzând acest mecanism în ceea ce privește un nivel de procesare superior celui perceptiv, rezultă că modificările la nivelul deprinderilor și automatismelor reprezintă o transpunere, în termeni computaționali, a unor proceduri formulate în limbaj de nivel înalt la nivelul arhitecturii *„soft"*. Rămâne de stabilit o regulă conform căreia pentru fiecare modificare a arhitecturii "soft" trebuie concepută o modificare a arhitecturii *„hard"* în concordanță cu regula lui Hebb. În cazul conexionismului, acest lucru se traduce doar prin creșterea convergenței între model și cel care este modelat, deoarece rețelele neuromimetice și rețelele neuronale reale reprezintă cele două niveluri, *„soft"* și *„hard"*.

Concluzia autorului, rezultată din aceste trei aspecte, este că, dacă acceptăm drept primitivă arhitectura *„hardware”* a sistemului cognitiv (indiferent de modelul computațional ales, creierul este compus tot din ansambluri de neuroni), este mult mai avantajoasă acceptarea la un nivel superior a unei arhitecturi cognitive *„soft”* cu plauzibilitate biologică, deci modelată conexionist. Creșterea plauzibilității biologice a modelelor conexioniste și extinderea aplicațiilor lor sunt actualmente direcții de cercetare, astfel că, este de părere autorul, putem rămâne optimiști [54].

Totuși, contrar optimismului autorului mai sus citat, există mai multe limite ale abordării conexioniste în contextul interpretării filosofice a rețelelor neuronale.

În primul rând, conexionismul se concentrează în principal pe procesele computaționale și pe modul în care informația este prelucrată în rețelele neuronale. În pofida acestui fapt, abordarea conexionistă întâmpină dificultăți în a explica conținutul mental și semnificația reprezentărilor cognitive. Acest aspect este important în filosofie, deoarece se pune întrebarea cum informația reprezentată în rețelele neuronale dobândește semnificație și înțeles subiectiv.

Rețelele neuronale cu mai multe straturi de neuroni (*deep neural network*), care sunt adesea amintite în argumentările conexioniste, sunt cunoscute pentru capacitatea lor de a realiza calcule complexe și de a învăța automat modele intricate complexe. Cu toate acestea, rețele *deep* pot fi extrem de opace și greu de interpretat. În consecință, explicabilitatea proceselor interne ale rețelelor neuronale poate fi o provocare, ceea ce poate limita înțelegerea noastră filosofică legată de modul în care este realizat „raționamentul” în astfel de sisteme.

De asemenea, abordarea conexionistă se concentrează pe modelarea proceselor cognitive și a comportamentului prin interacțiunile dintre unități de calcul simple. Totodată, ea nu abordează direct întrebările despre intenționalitate și conștiință. Deci, explicarea modului în care rețelele neuronale ar putea genera experiențe subiective sau conștiință rămâne un subiect deschis și controversat în filosofie [161].

În cele din urmă, este evident că simulările computerizate ale rețelelor neuronale, deși pot oferi *insight*-uri prețioase, există limite referitoare la cât de bine pot reproduce aceste simulări complexitatea și funcționalitatea rețelelor neuronale reale. Aceste simulări pot fi o simplificare excesivă a realității biologice, ceea ce poate restrânge înțelegerea noastră filosofică a naturii minții și a proceselor cognitive [162].

Abordarea filosofică transdisciplinară, cu modelul său bazat pe mai multe nivele de Realitate, este o abordare superioară în comparație cu abordarea conexionistă. Vom enumera în continuare principalele argumente care susțin acest punct de vedere.

Spre deosebire de conexionism, abordarea transdisciplinară recunoaște că fenomenele cognitive și mentale sunt rezultatul interacțiunii complexe și multidimensionale a numeroși factori

și, în esență, interacțiunea nemediată și nonierarhică dintre mai multe niveluri de Realitate. Transdisciplinaritatea permite luarea în considerare a unei game mai largi de influențe, inclusiv contextul social, cultural, biologic și filosofic. Astfel, abordarea transdisciplinară oferă o perspectivă mai comprehensivă asupra proceselor cognitive și mentale umane, depășind limitările unei perspective unidimensionale oferite de conexionism.

Abordarea transdisciplinară acordă o importanță deosebită conștiinței și subiectivității umane (Subiectul transdisciplinar). Aceasta recunoaște că gândirea și experiența umană sunt mai mult decât simple procese computaționale și că ele implică aspecte subiective și intenționale, cu punctul comun în ceea ce Basarab Nicolescu numește terțul tainic inclus [91]. Prin includerea filosofiei și a altor discipline umaniste în abordarea transdisciplinară (dar nu ca un simplu dialog interdisciplinar!), se acordă atenție aspectelor precum semnificația, subiectivitatea, intenționalitatea și natura conștiinței, care sunt adesea ignorate sau reduse în abordarea conexionistă.

Transdisciplinaritatea recunoaște că aspectele cognitive și mentale nu sunt separate de contextul social și etic în care se manifestă, pentru că fiecare nivel de Realitate este ceea ce este doar pentru că există toate celelalte nivele de Realitate. Acest fapt implică explorarea consecințelor sociale, morale și etice ale proceselor cognitive și a impactului acestora asupra persoanei (nu doar a individului, care nu este altceva decât reducerea persoanei la unul din nivelurile de Realitate!) și a societății în ansamblu. Prin integrarea transdisciplinară a eticii, sociologiei și altor discipline relevante, abordarea transdisciplinară oferă o perspectivă mai cuprinzătoare asupra implicațiilor cognitive și sociale ale minții umane [164].

Nu în ultimul rând, abordarea transdisciplinară este mult mai flexibilă și permite integrarea și adaptarea la noi descoperiri și perspective în domeniul studiului minții și al proceselor cognitive. Transdisciplinaritatea încurajează dialogul și colaborarea dintre discipline diferite, dar nu în maniera multi- sau interdisciplinară, ci permițând, prin acceptarea faptului că există un spațiu dintre discipline, actualizarea și îmbogățirea continuă a cunoștințelor noastre. Pe măsură ce învățăm mai mult despre natura complexă a minții umane, abordarea transdisciplinară ne permite să ne adaptăm și să integrăm noile cunoștințe cu adevărat holistic [86].

2.2 Limitările, performanțele și diferențele comparativ cu un agent inteligent uman. Problemele legate de înțelegerea contextului situațional și cognitiv. Inteligența simulată versus Inteligența în sine. Studiul de caz: ChatGPT.

ChatGPT este o entitate virtuală care manifestă o formă inedită de inteligență artificială. Cu rădăcini filosofice în teoria cunoașterii și epistemologie și poate fi considerat ca o expresie modernă a filosofiei cibernetice.

Din punct de vedere tehnic, ChatGPT este o rețea neuronală, foarte mare și foarte complexă, în care un număr foarte mare de neuroni artificiali sunt interconectați. ChatGPT este o rețea neuronală profundă, organizată pe mai multe straturi [174].

El funcționează după principiile generale ale unei rețele neuronale, cu câteva proprietăți specifice. O rețea neuronală primește datele de intrare, cum ar fi text sau o imagine. În cazul ChatGPT, intrarea constă în textul conversațional furnizat de utilizator. Intrarea este trecută printr-un strat de neuroni, numit strat de intrare. Fiecare neuron din acest strat primește o parte a datelor de intrare și le procesează.

După stratul de intrare, urmează una sau mai multe straturi ascunse. Aceste straturi conțin mulți neuroni care primesc informații de la straturile anterioare și le prelucrează prin aplicarea unor funcții matematice specifice. Prin intermediul acestor funcții, rețeaua are capacitatea de a învăța și de a identifica modele în datele de intrare [174].

După straturile ascunse, rețeaua neuronală ajunge la stratul de ieșire. Acesta generează rezultatele finale ale rețelei. În cazul ChatGPT, stratul de ieșire produce răspunsurile generate pe baza „înțelegerii” și a modelelor învățate în timpul antrenamentului.

Așa cum am arătat anterior, pentru a ajunge la un nivel ridicat de performanță, rețeaua neurală este antrenată folosind o cantitate mare de date etichetate. ChatGPT a fost antrenat cu texte preexistente pentru a „înțelege” structura limbajului și pentru a genera răspunsuri relevante.

În timpul procesului de antrenament, rețeaua ajustează ponderile asociate fiecărei conexiuni neuronale, în funcție de erorile observate în predicții. Acest proces de ajustare ajută rețeaua să se adapteze și să-și îmbunătățească performanța sa pe măsură ce învață din datele de antrenament. După finalizarea antrenamentului, rețeaua neurală poate fi utilizată pentru a genera răspunsuri și predicții bazate pe intrările primite [150].

Să luăm în considerare un exemplu concret de antrenament pentru ChatGPT, folosind texte de conversații.

În etapa de antrenament, ChatGPT este expus la o cantitate mare de texte de conversații, care includ întrebări și răspunsuri asociate. Aceste texte pot fi preluate de pe platforme de mesagerie, forumuri, pagini web, fișiere în diverse formate digitale sau alte surse similare.

Iată cum poate arăta un exemplu concret de antrenament:

Text de intrare:

Utilizator: „Ce timp va fi mâine?”

Text de ieșire dorit:

ChatGPT: „Este posibil să plouă mâine.”

Textul de intrare conține întrebarea utilizatorului, iar textul de ieșire dorit conține răspunsul corect. Pentru a antrena rețeaua, se compară răspunsul generat de ChatGPT cu textul de ieșire dorit și se măsoară, folosind diferite procedee matematice, discrepanța dintre ele. Textul de intrare și textul de ieșire dorit sunt prelucrate într-un format adecvat pentru a putea fi introduse în rețeaua neurală. De exemplu, textul poate fi transformat într-un vector de numere reale, unde fiecare număr reprezintă un cuvânt sau un caracter.

Exemplul de mai sus este mult simplificat și se referă la principiile generale ale antrenamentului rețelelor neurale. În practică, procesul de antrenament al unui sistem precum ChatGPT poate implica etape suplimentare și algoritmi mai avansați pentru a obține performanțe optime.

În mod concret, folosind textul de ieșire generat și textul de ieșire dorit, se calculează valoarea funcției de cost. Aceasta poate fi realizată prin aplicarea unei funcții matematice specifice, cum ar fi eroarea medie pătratică (*Mean Squared Erro*r) sau entropia încrucișată (*Cross-Entropy*). Utilizând aceste perechi de texte de intrare-ieșire, se aplică un proces numit *backpropagation* (propagarea erorii înapoi în rețea, de la ieșire spre stratul de intrare) pentru a ajusta ponderile rețelei și a minimiza discrepanța (eroarea, din punct de vedere matematic) dintre răspunsurile generate și răspunsurile dorite. Pe baza valorii funcției de cost (o altă denumire pentru funcția care calculează eroarea), ponderile conexiunilor dintre neuroni sunt ajustate în sensul de a minimiza această discrepanță. Acest proces se bazează pe gradientul funcției de cost și utilizează algoritmi de optimizare, cum ar fi algoritmul de propagare a gradientului (*Gradient Descent*).

În timpul antrenamentului se parcurg multiple iterații în care rețeaua primește diferite texte de intrare, procesul de ajustare a ponderilor având rolul de a face ca rețeaua să genereze răspunsuri corecte și relevante.

Așadar, procesul de propagare înapoi a erorii (*backpropagation*) și ajustarea ponderilor este repetat pentru multiple exemple de antrenament, pentru a îmbunătăți treptat performanța rețelei. Cu fiecare iterație, rețeaua ajustează ponderile în funcție de erorile observate, încercând să

minimizeze discrepanța între răspunsurile generate și cele dorite. Cu cât se folosesc mai multe texte, cu cât textele sunt mai diverse și mai reprezentative în procesul de antrenare, cu atât rețeaua devine mai capabilă să răspundă adecvat la o gamă largă de întrebări și solicitări.

Prin repetarea acestui proces de ajustare a ponderilor pe baza discrepanței dintre răspunsurile dorite și cele generate, rețeaua neurală tinde să se îmbunătățească în generarea de răspunsuri relevante și corecte.

Alegerea funcției de cost și a algoritmilor de optimizare poate varia în funcție de specificul problemei și de arhitectura rețelei neuronale. În cazul unui sistem precum ChatGPT, care este bazat pe modele de limbaj, funcția de cost și metodele de optimizare sunt adaptate pentru a îmbunătăți generarea de texte coerente și relevante.

Procesul de antrenament necesită resurse computaționale semnificative și poate dura mult timp pentru a obține rezultate optime. Cu toate acestea, odată finalizat antrenamentul, rețeaua poate fi utilizată pentru a genera răspunsuri în timp real la întrebări și solicitări.

Limitările ChatGPT sunt puse în evidență de anumite aspecte fundamentale.

În primul rând, ChatGPT 3.5 este constrâns de cunoștințele și informațiile disponibile în etapa de antrenament, în septembrie 2021. Inevitabil, orice proces de antrenare, fie și adus la zi, trebuie să fie oprit la un moment dat. Orice eveniment sau dezvoltare care a avut loc după data de terminare a antrenării îi vor rămâne, cel puțin pentru o perioadă de timp, necunoscute. În mod asemănător, și ChatGPT 4 va avea propria sa limită cronologică pentru datele de antrenare.

De asemenea, ChatGPT nu dispune de o conștiință în sensul uman, nefiind capabil să experimenteze subiectivitatea sau să își dezvolte o perspectivă personală. Așadar, ChatGPT este mai mult un instrument analitic, decât un agent autonom de gândire.

Chiar și așa, performanțele sale sunt notabile, remarcându-se prin abilitatea sa de a produce răspunsuri coerente și, multe dintre ele, relevante la întrebările și solicitările utilizatorilor. ChatGPT se bazează pe o rețea neuronală complexă și pe algoritmi de învățare profundă (*deep learning*), care îi permit să asimileze, să proceseze un volum imens de informații și să identifice modele în limbajul uman.

Performanțele lui ChatGPT constau în natura sa adaptabilă și în abilitatea de a învăța prin interacțiunea continuă cu utilizatorii. Cu cât este folosit mai des, cu atât își îmbunătățește cunoștințele și „înțelegerea", devenind mai precis și mai capabil să ofere răspunsuri utile. În ciuda acestui fapt, după cum vom vedea mai departe prin analiza unui studiu de caz, ChatGPT nu este întotdeauna în măsură să ofere răspunsuri corecte sau să gestioneze cu precizie subtilitățile și contextul unor întrebări complexe. Desigur, utilizatorii trebuie să fie conștienți de aceste limite.

2.2.1 Provocarea schemelor Winograd și înțelegerea contextului

Ce este *Winograd Schema Challenge* (WSC)?

„Winograd Schema Challenge” este un test de inteligență artificială care a fost propus ca o alternativă la testul Turing pentru a evalua abilitatea unei mașini de a înțelege limbajul natural. Acest test implică răspunsul la întrebări legate de anumite fraze, unde răspunsul corect depinde de înțelegerea unei subtilități lingvistice, adesea implicând anafora (o referință la un cuvânt sau o frază anterioară). Schema Winograd conține o pereche de enunțuri, unde schimbarea unui cuvânt schimbă răspunsul corect la întrebarea asociată [38; 45; 46].

Iată două exemple de enunțuri Winograd, împreună cu întrebările asociate:

Enunț 1: „Băiatul a lovit mingea. A zburat peste gard.”

Întrebare: „Ce a zburat peste gard?”

Răspuns: „Mingea”

Enunț 2: „Băiatul a lovit mingea. A zburat peste gard.”

Întrebare: „Ce a zburat peste gard?”

Răspuns: „Băiatul”

Enunț 1: „Polițistul nu a putut prinde hoțul cu mașina sa veche.”

Întrebare: „Cine avea mașina veche?”

Răspuns: „Polițistul”

Enunț 2: „Polițistul nu a putut prinde hoțul cu mașina sa veche.”

Întrebare: „Cine avea mașina veche?”

Răspuns: „Hoțul”

În fiecare exemplu, un cuvânt cheie în enunț a fost schimbat, modificând răspunsul corect la întrebarea asociată. Pentru a răspunde corect, un sistem de inteligență artificială ar trebui să fie capabil să înțeleagă subtilitățile lingvistice și referințele anaforice implicate în fiecare pereche de enunțuri.

Propoziții sau enunțuri care pot fi dificile pentru modelele lingvistice sunt adesea cele care implică abstractizări avansate, referințe la cunoștințe de nișă, ambiguitate înaltă, sau sarcini care necesită raționament profund [36]. De exemplu, enunțuri cu ambiguitate semantică ridicată: „Fructul mărului stătea pe masă lângă mărul de la calculator.” (Ambiguitate lexicală între „măr”

ca fruct și „măr” ca parte a unui calculator). Sau enunțuri cu anafore complexe: „Maria i-a spus Anei că sora sa va veni mai târziu.” (Este neclar cărei surori i se referă - a Mariei sau Anei). Metonimii și metafore avansate: „Washington este reticent în a accepta propunerea.” (Aici „Washington” este folosit pentru a reprezenta guvernul SUA, care poate fi confuz pentru un model care nu înțelege metonimiile). Întrebări deschise și abstracte: „Ce înseamnă a fi conștient de propria existență?” Întrebări care necesită raționament avansat: „Dacă toți trandafirii din grădină cresc mai repede în zona umbrită și Maria plantează un trandafir în zona însorită, ce se poate deduce despre rata de creștere a trandafirului plantat de Maria comparativ cu celelalte?” Sentințe cu implicații culturale sau istorice specifice: „Cum a influențat revoluția de la 1821 politica internă a țării în deceniile care au urmat?” (Necesită cunoștințe specifice despre contextul istoric). Întrebări care necesită interpretare creativă sau artistică: „Ce sentimente încearcă să exprime artistul în această pictură?” (Depinde de pictură și necesită o analiză subiectivă).

Aceste tipuri de propoziții pot prezenta dificultăți pentru modelele lingvistice, fie pentru că necesită o analiză complexă, raționament profund, sau înțelegerea subtilităților lingvistice și culturale [35].

Modelele lingvistice, în special cele bazate pe rețele neuronale profunde, cum ar fi ChatGPT, funcționează printr-un proces complex de interpretare și generare a limbajului. Iată o privire generală asupra procesului de interpretare a propozițiilor complexe [150].

Tokenizare: În primul rând, textul de intrare este divizat în *„tokeni”,* care sunt unitățile individuale (cuvinte sau subcuvinte) care vor fi procesate de model.

Încorporare:

Apoi, fiecare token este transformat într-un vector multidimensional, folosind o matrice de încorporare. Aceasta încorporează informații despre semantica cuvintelor din datele de antrenament.

Procesare în Straturi Ascunse:

Modelele precum GPT-3 sunt alcătuite din multe straturi de neuroni artificiali. Fiecare strat învață diferite aspecte ale datelor de intrare. Acest lucru este realizat printr-o serie de operațiuni matematice complexe (cum ar fi transformări liniare și non-liniare) care ajustează iterativ ponderile conexiunilor dintre neuroni pe baza datelor de antrenament [162]. În aceste straturi, modelele încearcă să înțeleagă contextul și structura lingvistică a propoziției, identificând relații între cuvinte și extrăgând caracteristicile relevante.

Mecanismele de „atenție” permit modelului să acorde mai multă importanță anumitor cuvinte în funcție de context, ajutând la rezolvarea unor probleme precum dezambiguizarea anaforică sau înțelegerea referințelor [174].

Generarea răspunsului:

Odată ce textul a fost procesat prin toate straturile, modelul încearcă să genereze un răspuns sau o continuare. În această etapă modelul evaluează cuvintele următoare posibile și selectează cel mai probabil cuvânt pe baza unei funcții de cost, care încurajează alegerea cuvintelor care măresc probabilitatea secvenței curente.

Rafinare și optimizare:

În etapele ulterioare de antrenament, modelul se optimizează continuu, ajustându-și ponderile pentru a reduce eroarea în predicțiile sale.

Interpretare și răspuns final:

În final, modelul oferă un răspuns bazat pe interpretarea sa a textului de intrare și pe ceea ce a învățat în timpul antrenamentului. Răspunsul poate fi apoi prezentat utilizatorului.

Feedback și ajustare:

Unele sisteme pot permite feedback de la utilizatori, utilizând aceste informații pentru a ajusta și îmbunătăți modelele ulterior.

Este important de reținut că aceste modele, deși sunt foarte avansate, au încă limitări semnificative, în special în ceea ce privește înțelegerea semanticii profunde, raționamentul abstract și generarea de răspunsuri care necesită cunoștințe specifice sau expertiză într-un domeniu [147].

În procesarea limbajului natural (NLP), un token este o unitate de text care a fost extrasă în timpul procesului de tokenizare. Un token poate fi un cuvânt, o parte dintr-un cuvânt sau un caracter, în funcție de modul în care este definit procesul de tokenizare [174].

Există mai multe tipuri de tokeni:

Cuvânt: în cea mai simplă formă, un token poate fi un cuvânt individual dintr-o frază. De exemplu, în fraza „Ana are mere”, cuvintele „Ana”, „are” și „mere” sunt tokeni;

O parte dintr-un cuvânt: în unele cazuri, cuvintele pot fi divizate în unități mai mici. De exemplu, cuvântul „neînțeles” poate fi tokenizat în „ne-” (prefix), „înțeles” (rădăcină).

Caracter: un token poate fi, de asemenea, un caracter individual dintr-un text.

Iată un exemplu de tokenizare:

Luând fraza „Casa este frumoasă.”, procesul de tokenizare la nivel de cuvânt ar fi:

Token 1: Casa

Token 2: este

Token 3: frumoasă

Token 4: . (punct)

Tokenizarea este un pas fundamental în procesarea limbajului natural, deoarece permite modelului să manipuleze și să analizeze textul în unități mai mici, facilitând aplicarea tehnicilor

de *machine learning* și statistice. Aceasta ajută în efectuarea sarcinilor precum analiza sentimentelor, recunoașterea entităților numite, traducerea automată și altele [139].

Tokenizarea, în contextul procesării limbajului natural, poate fi privită din perspectiva filosofiei limbajului ca o încercare de a deconstrui și categorisi complexitatea comunicării umane în unități semnificative. În esență, este o modalitate de a dezagrega discursul fluid în bucăți discrete pentru a facilita analiza.

Ludwig Wittgenstein, în *Tractatus Logico-Philosophicus* [122] și, mai târziu, în *Cercetări filozofice* [123], explorează natura sensului și limitările limbajului. El sugerează că limbajul este în esență un instrument, iar sensul cuvintelor provine din modul în care sunt utilizate în practică. Prin tokenizare, încercăm să captăm aceste practici de utilizare, dar este important de menționat că „destrămarea" textului în bucăți mici poate risca să piardă unele dintre contextele în care aceste unități de limbaj dobândesc sensul lor.

În *Curs de lingvistică generală* [24], Ferdinand de Saussure, a introdus distincția dintre semnificant (forma unui cuvânt) și semnificat (sensul sau conceptul reprezentat de un cuvânt). Tokenizarea poate fi văzută ca o încercare de a izola semnificanții pentru a facilita analiza, dar acest proces nu poate fi complet decuplat de contextul în care semnificatul este înțeles.

Jacques Derrida, cu conceptul său de *diférance* și critica sa la adresa logocentrismului, ar putea argumenta că tokenizarea încearcă să stabilească o prezență sau o esență fixă a limbajului, în timp ce sensul este în mod constant în derivă și nu poate fi complet fixat. Așadar, în timp ce tokenizarea este utilă pentru a analiza textul în termeni tehnici, nu ar trebui să presupunem că aceasta capturează complet nuanțele sau complexitățile sensului [25].

Așa cum am menționat anterior, transdisciplinaritatea [92] se bazează pe ideea că există nivele de realitate care sunt interconectate prin intermediul logicii terțului inclus. Acesta este un teritoriu în care opozițiile tradiționale sunt depășite și apare o nouă coerență. Când privim la procesul de tokenizare prin lentila axiomelor transdisciplinare, ne confruntăm cu o serie de provocări și realizări.

Tokenizarea pare să opereze la un nivel al realității în care cuvintele sau frazele sunt tratate ca unități discrete și separate. Acest nivel al realității privilegiază claritatea și analiza, dar poate neglija complexitatea relațiilor dintre aceste unități discrete și contextul lor mai larg. În termenii lui Basarab Nicolescu, aceasta poate fi văzută ca o limitare la nivelul realității în care tokenizarea este efectuată.

Din perspectiva logicii terțului inclus, procesul de tokenizare poate fi înțeles ca parțial adecvat pentru descoperirea sensului în limbaj. Mai degrabă, tokenizarea poate fi văzută ca o

modalitate de a naviga în complexitatea limbajului, recunoscând în același timp că acesta nu poate prinde întotdeauna nuanțele și relațiile dinamice care există în discursul real.

În plus, transdisciplinaritatea ne îndeamnă să privim dincolo de granițele tradiționale ale disciplinelor și să înțelegem cum diferite nivele de realitate se intersectează și se influențează reciproc. Astfel, în timp ce tokenizarea poate fi utilă în contextul procesării limbajului natural, este important să ne reamintim că aceasta este doar una dintre multele abordări posibile și că există alte nivele ale realității în care sensul se manifestă în moduri mai subtile sau complexe.

În esență, tokenizarea, văzută prin axiomele transdisciplinare, ne oferă o imagine a complexității și a naturii multidimensionale a limbajului. Ne arată că, în timp ce tokenizarea poate fi valoroasă pentru anumite scopuri, nu ar trebui să ne bazăm exclusiv pe tokenizare ca pe o modalitate definitivă sau exhaustivă de a înțelege sau analiza comunicarea umană.

BERT și Schemele Winograd

BERT, un model lingvistic dezvoltat de Google, cu arhitectura sa bidirecțională, poate fi antrenat pentru a rezolva schemele Winograd într-un mod mai sofisticat decât modelele anterioare, deoarece poate înțelege contextul unui cuvânt în ambele direcții (de la stânga la dreapta și de la dreapta la stânga). Datorită înțelegerii contextuale bidirecționale, BERT poate analiza relația dintre cuvinte într-un mod mai subtil, făcându-l potrivit pentru a identifica coreferințele în schemele Winograd. Pentru a excela în rezolvarea schemelor Winograd, BERT poate fi *fine-tuned* cu un set de date specific care conține exemple de scheme Winograd. Acest *fine-tuning* ajută modelul să se specializeze în identificarea corectă a referințelor pronumelor în aceste tipuri de probleme.

În procesul de răspuns, BERT poate folosi strategii de tip *masked language model* (MLM) pentru a prezice cuvântul corect în locul unei „măști” (de exemplu, un pronume care determină referința) în funcție de contextul oferit de restul propoziției.

Deși BERT poate avea un avantaj în abordarea schemelor Winograd datorită capacității sale de a înțelege contextul bidirecțional și relații complexe între cuvinte, este demn de remarcat că schemele Winograd sunt încă o provocare majoră în domeniul inteligenței artificiale, chiar și pentru modele avansate cum ar fi BERT, deoarece necesită o înțelegere profundă și nuanțată a limbajului natural [147].

2.2.2 Studiu de caz. Teste-dialog cu ChatGPT 3.5

Vom exemplifica acest lucru prin următoarele dialoguri-test cu ChatGPT 3.5, singura versiune disponibilă la momentul testării. De menționat că odată cu apariția versiunii ChatGPT 4

și BARD (mai târziu Gemini) de la Google propozițiile au fost testate din nou. Chiar dacă am primit rezultate diferite în anumite cazuri, în esență nimic nu s-a schimbat. Scopul nu este de a arăta performanța modelelor, deoarece oricând pot apărea modele mai performante, ci mai degrabă cum ele funcționează în comparație cu inteligența umană.

Pentru primele două teste am preluat patru întrebări cu scheme Winograd foarte bine cunoscute și testate de alte modele lingvistice.

> „User-ul: Joan a ținut să îi mulțumească lui Susan pentru tot ajutorul pe care îl primise. Cine primise ajutorul? a) Joan; b) Susan
>
> ChatGPT 3.5: Joan a primit ajutorul din partea lui Susan, deci Susan a oferit ajutorul, iar Joan a primit ajutorul. Răspunsul corect este b) Susan."

Se observă că ChatGPT 3.5 nu reușește să discearnă corect contextul. Întrebarea a fost reformulată pentru a vedea dacă reformularea va genera un răspuns relevant:

> „User-ul: Joan a ținut să îi mulțumească lui Susan pentru tot ajutorul pe care i-l oferise. Cine oferise ajutorul? a) Joan; b) Susan
>
> ChatGPT 3.5: În propoziția dată, subiectul este Joan, iar verbul este "a mulțumi", ceea ce indică faptul că Joan este persoana care a recunoscut ajutorul primit. Pronumele "îi" se referă la Susan, deci ea este persoana care a oferit ajutorul. Răspunsul corect este b) Susan."

Se observă că răspunsul corect este oferit prin strictă analiză gramaticală, nu prin înțelegerea contextului situațional.

În continuare, a fost testată capacitatea lui ChatGPT 3.5 de a „înțelege" situațional noțiunile de mai mare / mai mic:

> „User-ul: Trofeul nu a încăput în geamantanul maro deoarece era foarte mic. Ce era foarte mic? a) Trofeul; b) geamantanul maro
>
> ChatGPT 3.5: Conform propoziției, verbul este „a încăput", ceea ce indică faptul că obiectul direct este ceva care a putut fi sau nu încărcat în geamantanul maro. De asemenea, este menționat că trofeul nu a putut fi încărcat, deci răspunsul corect este a) Trofeul.
>
> User-ul: Trofeul nu a încăput în geamantanul maro, cu toate că era foarte mic. Ce era foarte mic? a) trofeul; b) geamantanul maro
>
> ChatGPT 3.5: În propoziția dată, se spune că trofeul nu a putut încăpea în geamantanul maro, deși era foarte mic, ceea ce indică faptul că

trofeul este obiectul care este descris ca fiind mic. Prin urmare, răspunsul corect este a) trofeul.”

Se observă cum, în ambele cazuri, ChatGPT 3.5 nu înțelege relația dintre cele două obiecte cu mărimi diferite. ChatGPT 3.5 încearcă să scoată din analiza pur gramaticală a relațiilor dintre cuvinte sensuri care nu depind în exclusivitate de forma gramaticală. Este o dovadă că limbajul uman nu este doar morfologie și sintaxă și că semantica depinde de un cadru contextual foarte larg, care implică și interacțiunea cu mediul și de relațiile foarte complexe care se pot dezvolta între agenții care comunică.

Un ultim test de înțelegere contextuală este următorul (aici trebuie să menționăm că nici ChatGPT 4.0 nu face față provocării date):

„User-ul: În timp ce înotam mă apropiam de mine și astfel a început să-mi fie teamă pentru mine. Cui îi era teamă și de ce anume îi era teamă?
ChatGPT 3.5: În propoziția dată, formularea este puțin ambiguă și poate fi interpretată în mai multe feluri. Cu toate acestea, în contextul dat, interpretarea corectă ar fi următoarea: persoana care vorbește se referă la sine însăși în propoziție, deci „eu” este subiectul. Prin urmare, „cui îi era teamă” se referă la persoana care vorbește, adică „eu”. „De ce anume îi era teamă” nu este specificat în propoziție și necesită informații suplimentare pentru a răspunde cu exactitate.”

Se observă că ChatGPT 3.5, deși reușesc să constate ambiguitatea textului din analiza gramaticală, încă din start pleacă pe o pistă de interpretare greșită. Textul este preluat dintr-o scenetă umoristică cândva foarte populară în România. Ambiguitatea este una din sursele umorului și, pentru orice persoană cu capacități cognitive normale, indiferent de nivelul studiilor, acest tip de ambiguitate nu prezintă nicio o problemă de interpretare. De aceea, majoritatea oamenilor râd când întâlnesc astfel de expresii. Acest lucru se explică, așa cum vom arăta pe parcursul acestei lucrări, prin faptul că sensul unui text are aproape întotdeauna referințe exterioare textului.

„User: pune-mi întrebări care să elimine ambiguitatea.
ChatGPT 3.5: Desigur! Iată câteva întrebări care pot ajuta la eliminarea ambiguității:
Când spui „mă apropiam de mine”, te referi la o altă persoană sau te referi la tine însuți?
Când afirmi că „a început să-ți fie teamă pentru tine”, cine era în situația de a avea teamă? Te referi la tine însuți sau la o altă persoană?

Dacă ai avut teamă, ce anume a cauzat acea teamă? Există un factor specific sau un pericol identificabil?

Te rog să oferi mai multe detalii sau clarificări pentru a putea înțelege mai bine contextul și a răspunde în mod adecvat.”

Observăm cum ChatGPT 3.5 rămânând exclusiv legat de text intră din nou pe o pistă de interpretare greșită. Deși, identifică un pericol nu reușește să-l lege de polisemia cuvântului „mine”. În continuare User-ul încearcă să-l ajute în această direcție.

„User-ul: Cel mai corect ar fi să pui întrebarea despre ce sens are „mine” din expresia „înotam spre mine”.

ChatGPT 3.5: Înțeleg. Îmi cer scuze pentru confuzie. Iată întrebarea pentru a clarifica sensul expresiei „înotam spre mine”: Când spui „înotam spre mine”, te referi la direcția de înot a persoanei care vorbește, către propria persoană, sau te referi la înotul către o altă persoană?

Această întrebare ar trebui să clarifice sensul corect al expresiei și să elimine ambiguitatea.

User-ul: ce sensuri are cuvântul „mine” in limba română?

ChatGPT 3.5: Cuvântul „mine” din limba română poate avea mai multe sensuri, în funcție de context. Iată câteva dintre aceste sensuri:

Pronume personal: „mine” este pronumele personal de persoana întâi singular, folosit pentru a se referi la persoana care vorbește.

Substantiv: „mine” poate fi și un substantiv comun care desemnează o mină, adică o exploatare a solului pentru extragerea resurselor minerale.

Adverb de timp: „mine” poate fi folosit ca adverb de timp pentru a indica ziua următoare celei în care se vorbește.

Este important să rețineți că sensul cuvântului „mine” depinde de contextul în care este utilizat.”

Este interesant de observat că în ciuda antrenamentului cu o bază de date imensă de texte (inclusiv în limba română) ChatGPT 3.5 nu folosește exact sensul cuvântului de care ar avea nevoie pentru eliminarea ambiguității. Deși, așa cum vom vedea mai departe are și acest cuvânt în baza de cunoștințe. Utilizatorul încearcă să-i sugereze sensul potrivit. După cum vom vedea ChatGPT 3.5 va răspunde pe larg folosind chiar amănunte tehnice ceea ce ne arată încă o dată că înțelegerea sensului unui text nu ține exclusiv de accesarea unei baze de date.

„User-ul: știi ceva despre „mine” marine care pot exploda?

ChatGPT 3.5: Da, există mine marine care pot exploda. Mina marina este un dispozitiv exploziv care este amplasat sub apă, de obicei în scopuri militare, pentru a ataca navele sau alte vehicule maritime. Aceste mine pot fi activate de contactul cu o navă, de magnetism, de presiunea apei sau prin alte metode. Explodează atunci când sunt activate, cauzând daune semnificative navei sau vehiculului atins. Misiunea de detectare și neutralizare a minelor marine este un aspect important al activităților navale și de securitate maritimă.

User-ul: Deci, dacă înot în apă si mă apropii de mine marine nu e este corect să spun că este un pericol pentru mine?

ChatGPT 3.5: Da, dacă te apropii de mine marine în apă, ar putea fi considerat un pericol pentru tine. Mecanismele prin care minele marine sunt activate pot varia, dar multe dintre ele pot fi declanșate de contactul cu un obiect sau de alte stimuli, cum ar fi presiunea apei sau magnetismul. Prin urmare, dacă te apropii de o mină marină în apă, există riscul ca aceasta să fie activată și să provoace o explozie, cu consecințe grave pentru tine. Este important să te informezi cu privire la eventualele pericole și să respecți avertismentele și restricțiile de siguranță stabilite în zonele unde există riscul prezenței minelor marine.”

În ciuda noilor date, ChatGPT 3.5 încă nu a „decriptat” sensul textului inițial. Suntem tentați să spunem că ceea ce lipsește acestei inteligențe artificiale este acea senzație acută de frică pe care orice om o poate simți fie și doar imaginându-se înotând în apropierea unei mine marine. Experiența unui pericol, pe care fiecare om o trăiește aproape zi de zi face ca astfel de ambiguități textuale să fie înțelese aproape instantaneu, chiar dacă persoana respectivă nu a avut o experiență reală de acest gen. Ființele umane sunt capabile de „extrapolare” a experiențelor și cunoștințelor împărtășite de către alte persoane. Așa se explică de ce literatura și arta cinematografică au un impact atât de puternic asupra noastră. Acest subiect va fi dezvoltat pe larg în cuprinsul lucrării.

În continuare, utilizatorul mai testează încă o dată capacitatea lui ChatGPT 3.5 de sesiza contextul semantic al unor cuvine încercând să observe dacă inteligența artificială folosește rezultatele antrenării din discuțiile anterioare:

„User-ul: Explică-mi următoarea frază: „Nu te apropia de mine, Marine, dacă vrei să te mai întâlnești cu mine.”

ChatGPT 3.5: Această frază pare să fie un joc de cuvinte care folosește termenul „mine” cu două sensuri diferite. Iată o explicație posibilă a frazei:

„Mine” cu sensul de pronume personal: „Nu te apropia de mine” - în acest context, „mine” se referă la persoana care vorbește, adică vorbitorul, exprimând dorința acestuia de a nu fi apropiat de către Marine.

„Mine” cu sensul de mină marină: „dacă vrei să te mai întâlnești cu mine” - aici, termenul „mine” este folosit pentru a face o legătură cu subiectul precedent și pentru a crea un joc de cuvinte. Sugerează că, dacă Marine se apropie de „mine” (adică de mină marină), ar exista un pericol și că vorbitorul nu ar mai fi disponibil pentru o întâlnire ulterioară. În ansamblu, fraza pare a fi o construcție jucăușă care joacă pe două sensuri ale cuvântului „mine” și sugerează că apropierea de o mină marină ar fi periculoasă și ar putea duce la refuzul unei întâlniri ulterioare.”

Foarte interesant, dar se pare că discuția anterioară, care pentru o persoană derutată ar fi fost pe deplin edificatoare, pe ChatGPT 3.5 mai mult îl încurcă, ducându-l din nou pe o pistă de interpretare greșită. În textul dat „minele marine” nu-și au locul. Dar, pentru că este programat să folosească rezultatele discuțiilor anterioare, ChatGPT 3.5 introduce aici artificial conceptul de „mine marine”, nedându-și seama că „Marine” este o formă imperativă a unui nume propriu. Aceste exemple demonstrează că ChatGPT 3.5 nu poate, cel puțin în acest moment, să treacă cu succes faimosul test Turing. Desigur, se știe că programatorii ChatGPT-ului nici nu și-au propus acest lucru, dar un utilizator neavizat ar putea să creadă, eronat, că ChatGPT poate îndeplini fără probleme cerințele testului.

În concluzie, deși trebuie să recunoaștem că ChatGPT reprezintă un instrument interesant și puternic, care combină elemente de logică cu tehnicile de inteligență artificială și cu elemente de analiză a limbajului natural, este evident că nu reușește, de multe ori, să se adapteze contextului lexical. Chiar și în aceste condiții, demonstrează performanță și potențial în a facilita și a oferi informații într-un mod accesibil și captivant.

Așa cum am constatat, ChatGPT, ca orice alt sistem dezvoltat de inteligență artificială, are și limite care, în acest moment, nu par ușor de surmontat. Chiar dacă este programat să ofere răspunsuri relevante și precise din punct de vedere al detaliilor, există multe situații în care ChatGPT poate furniza informații incorecte sau poate interpreta greșit contextul întrebărilor.

Așa cum văzut mai sus, un exemplu de răspuns eronat pe care ChatGPT îl poate oferi la o întrebare este atunci când primește o întrebare ambiguă sau imprecisă. Deoarece ChatGPT nu are capacitatea de a cere clarificări precise sau de a deduce intenția din context, poate furniza un răspuns care nu corespunde cu așteptările utilizatorului.

Să luăm încă un exemplu: dacă cineva îl întreabă pe ChatGPT „Cât de mare este Everestul?" și nu specifică în mod clar dacă se referă la înălțime sau la suprafață, ChatGPT poate furniza în mod greșit informația că Everestul are o înălțime de 8.848 de metri, ignorând complet posibila interpretare a întrebării referitoare la suprafață. Aceasta este o limitare ușor de descoperit în capacitatea lui ChatGPT de a deduce contextul și de a furniza răspunsuri adecvate în situații ambigue.

Este important ca utilizatorii să fie conștienți de faptul că, în ciuda eforturilor programatorilor lui ChatGPT de a furniza informații precise și relevante, există posibilitatea ca multe răspunsuri să fie inexacte sau să nu corespundă în totalitate cu intenția originală a întrebării. De aceea, este recomandabil întotdeauna verificarea informațiilor furnizate de ChatGPT, sau de către orice alt tip de inteligență artificială, cu surse suplimentare și consultarea unor surse de încredere pentru a obține informații precise și actualizate. Altfel spus, orice informație preluată dintr-un sistem care folosește inteligența artificială trebuie adusă la un context situațional relevant pentru utilizatorul uman.

Acest studiu de caz scoate în evidență veridicitatea argumentului mai vechi a lui John Searle despre camera chinezească [107]. Să presupunem că ne aflăm într-o cameră închisă, a cărei unică legătură cu exteriorul o constituie o ușă care este prevăzută o fantă de genul celor folosite la cutiile poștale. În cameră se află un mare număr de cartonașe pe care sunt desenate ideograme chinezești, câte una pe fiecare cartonaș. Tot în această cameră se mai găsește o carte foarte mare, în genul dicționarelor, în care sunt trecute instrucțiunile în engleză referitoare la modul în care trebuie folosite cartonașele care sunt primite prin fantă. Un exemplu de instrucțiune ar putea fi: „Dacă prin fantă se primește ideograma corespunzătoare cuvântului *mâzgălitură*, atunci se caută și se trimite afară cartonașul cu traducerea *mâzgălitură*".

Persoanele din afara camerei ne trimit o serie de asemenea cartonașe, iar noi căutăm în carte instrucțiunile potrivite și trimitem înapoi cartonașul solicitat. Fără să știm, pentru că nu cunoaștem absolut deloc limba chineză, cartonașele care vin din exterior alcătuiesc o serie de întrebări despre, să zicem, o piesă muzicală de succes. Cartonașele care ies prin fantă constituie un răspuns coerent și de bun simț la întrebări referitoare la compozitor, interpret și la mesajul cântecului.

Pentru cei din afara camerei, cutia neagră reprezentată de cameră și conținutul ei manifestă o înțelegere perfectă a limbii chineze. Totuși, din perspectiva noastră, a celor din interiorul camerei, nu există niciun fel de înțelegere a limbii chineze: noi nu facem altceva decât să introducem niște cartonașe cu ideograme chinezești, urmând cu strictețe setul de reguli. Cu alte cuvinte, există sintaxă, dar fără semantică.

Experimentul mental al lui Searle este exact ceea ce face astăzi ChatGPT sau orice alt sistem asemănător. Gândirea presupune înțelegerea sensului simbolurilor, pe câtă vreme ChatGPT (și orice alt sistem de calcul, cu sau fără inteligență artificială), nu face altceva decât să manipuleze simboluri potrivit unui set de reguli. Nu există nicio înțelegere a întrebărilor puse în chineză, doar o manipulare oarbă de simboluri. Nu există niciun fel de înțelegere a întrebărilor puse lui ChatGPT, ci doar o aranjare a cuvintelor după regulile și setul de date cu care ChatGPT a fost antrenat. Fără înțelegere, nu există stări cognitive veritabile, deci nu există gândire. Mai mult, inteligența, fiind mai mult decât o simplă manipulare de simboluri, putem conchide că ChatGPT este o reușită din punctul de vedere a simulării inteligenței. El doar pare inteligent. În realitate este la fel de inteligent pe cât de buni cunoscători ai limbii chineze suntem noi. Orice fel de înțelegere ar exista în acest proces, ea este prezentă numai în ceea ce a fost „programat" din exterior, în cartea de reguli, iar „procesorul" (noi, cei care facem legătura între mesajele de intrare și cele de ieșire) nu are nicio idee despre ceea ce semnifică simbolurile în realitate.

Orice poate părea a fi înțelegere în schimbul de mesaje cu ChatGPT nu este altceva decât înțelegerea avută în vedere de către colectivul care a proiectat și a antrenat rețeaua neuronală.

Concluzia este că ChatGPT, sau într-un sens mai larg inteligența artificială, nu posedă nici un fel de înțelegere a textelor analizate, ci doar o rulare automată a unor algoritmi complecși (la o intrare formată dintr-un vector de cuvinte, sistemul răspunde la ieșire cu cel mai probabil vector de cuvinte, având în vedere vectorul de intrare).

O inteligență artificială nu poate poseda o stare mentală, iar inteligența nu se poate manifesta în absența unei astfel de stări. Acesta este și motivul pentru care noi considerăm ca termenul de „inteligența artificială" ar trebui, pentru rigoare, întotdeauna folosit încadrat între ghilimele.

Searle a fost foarte atacat pentru acest argument. Cele mai multe atacuri nu s-au folosit de argumente, ci de invective „academice". Și Douglas Hofstadter, și Daniel Dennett, în ciuda competențelor de care ei dau dovadă în propria specialitate, au dovedit că dialogul se termină acolo unde încep prejudecățile noastre. Nu este de ajuns să spui despre un articol că este „exasperant" [145] sau că este o „diatribă sectară" [134] ca să îi și dovedești lipsa de consistență. Stadiul actual al inteligenței artificiale îl face pe Searle mai actual ca oricând. Este foarte probabil că

argumentația de bun simț a lui Searle și exemplul concret, aproape reproductibil cu mijloacele de azi, i-au scos din sărite pe adepții reducționismelor de tot felul.

De fapt, chiar răspunsul pe care ChatGPT îl dă atunci când îi adresăm întrebarea dacă înțelege limba chineză este mult mai elocvent decât toate articolele anti-Searle luate la un loc: „În calitate de model de inteligență artificială, nu am înțelegerea umană a limbii chineze. Deși pot procesa și genera texte în limba chineză, nu am o conștiință sau o experiență subiectivă asupra limbii sau a sensului său. Am fost antrenat să înțeleg și să generez texte într-o varietate de limbi, inclusiv limba chineză. Cu toate acestea, este important de menționat că eu, ca model, nu „înțeleg" în sensul uman al termenului. Eu procesez informațiile pe baza datelor și algoritmilor, nu pe baza experiențelor sau conștiinței." Dacă ar fi să luăm în considerare vocabularul adversarilor camerei chinezești, am putea spune că ChatGPT „înțelege" mai bine decât ei ce înseamnă **înțelegere** [17].

O altă discuție pe care o generează analiza acestui studiu de caz este legată de vechea dispută dintre Noam Chomsky și Jean Piaget: este limbajul o proprietate înnăscută a creierului uman sau o achiziție ulterioară în procesul de creștere? [98]

Posedă ChatGPT cu adevărat o capacitate lingvistică? Dacă limbajul presupune existența unei gramatici universale, așa cum credea Chomsky, este posibil ca această gramatică universală să fie implementată într-o structură fizică nebiologică?

Chomsky a avansat conceptul radical după care gramatica oricărui limbaj trebuie să fie generativă, în sensul că ea trebuie să reprezinte un set de reguli capabile să genereze toate propozițiile corect formulate ale limbajului și niciuna din cele incorecte [129]. Din această perspectivă, se poate spune că ChatGPT posedă o astfel de gramatică generativă?

În acest sens, urmând raționamentele lui Chomsky, putem să analizăm modul în care copiii dobândesc abilitatea de a vorbi limba lor maternă, comparativ cu metoda prin care ChatGPT și-a obținut abilitățile de manipulare a cuvintelor.

Nu trebuie scăpat din vedere faptul că vorbirea auzită de copii nu constă întotdeauna din propoziții corect formulate, complete, ci include propoziții formulate greșit, afirmații parțiale, scăpări de limbaj și alte expresii negramaticale. Copiii întâlnesc numai un domeniu limitat de expresii și, cu toate acestea, ajung să se descurce cu un spectru practic nelimitat de propoziții noi, care depășesc cu mult ceea ce au auzit ei anterior. Chomsky este de părere că fiecare copil asimilează niște scheme de generare a unui număr potențial nelimitat de propoziții, cum ar fi „acesta este câinele care a alergat pisica ce a mâncat șoarecele care mi-a ros coperta cărții pe care o primisem cadou…" (raport de relativizare) sau „Mihai a plecat la București să testeze o rețea neuronală, Natalia și Andrei au ieșit să se plimbe pe malul Prutului, iar Adrian s-a culcat la umbra unui nuc…" (raport de coordonare), ori „Ai auzit că Vitalie m-a rugat să-i spun lui Andrei că a

văzut cetatea de scaun a lui Ștefan cel mare…” (raport de subordonare). Este foarte probabil că nu ați văzut niciodată până în acest moment aceste fraze și, în fapt, mai mult ca sigur că ele nu au existat până în acest moment când le-am inventat scriindu-le aici. Totuși, veți înțelege imediat structura și înțelesul lor și, ceea ce este mai important, o va înțelege chiar și un copil de 6 ani. Este clar, evidențiază pe bună dreptate Chomsky, un copil nu învață limba maternă exclusiv prin imitare.

Și mai important, observă Chomsky, copiii ajung să știe în mod subconștient lucruri legate de limba lor despre care nu există dovezi directe în cadrul datelor la care sunt expuși. De exemplu, copiii nu sunt informați sistematic asupra faptului că anumite propoziții ipotetic posibile nu apar în fapt în realitate sau că o frază oarecare este ambiguă (cu alte cuvinte, în limbajul tehnic al învățării automate, procesul de învățare nu este exclusiv supervizat).

Recapitulând faptele relevante semnalate de Chomsky, observăm că:

Copilul stăpânește, fără o instruire semnificativă, un bogat sistem de cunoștințe.

Acest lucru se realizează în ciuda sărăciei de stimuli.

Procesul are loc mai rapid între vârstele de doi și trei ani.

Copiii normali sunt în stare să stăpânească ORICE limbaj uman la care sunt expuși pe durata copilăriei.

Gramatica abstractă caracterizează sintaxa abstractă a limbajului, independentă de particularitățile și idiosincraziile prezente într-o anumită comunitate lingvistică umană. Ca atare, ereditatea oferă schema de bază comună tuturor limbilor, mediul lingvistic al copilului completând apoi detaliile specifice limbii pe care o învață.

Este evidentă diferența dintre modul de antrenare al unei rețele de tip ChatGPT și modul în care noi asimilăm limba maternă.

2.2.3 Dialectica dintre limbajul natural și cel artificial

Deosebirea dintre limbajele naturale și cele artificiale nu este doar istorică, ci și structurală. Cu toate că în lucrarea „*Cât de umane pot fi limbajele artificiale*, în *Sisteme de inteligență artificială”* [80] autorul nu stabilește o concluzie definitivă în acest sens, analiza studiului de caz privind ChatGPT sugerează cu fermitate că există o diferență fundamentală în structura limbajului uman și cea a mașinilor, inclusiv a ChatGPT.

Hans Freudenthal a adus o contribuție semnificativă în dezbaterea cu privire la limbajele artificiale, și mai specific la limba esperanto, un exemplu clasic de limbaj artificial. El a formulat o obiecție importantă care a stârnit interesul cercetătorilor în ceea ce privește natura și structura

limbajelor artificiale. Obiecția lui se concentrează pe faptul că limbajele artificiale, inclusiv esperanto, par să se limiteze la regularitatea și predictibilitatea la nivelul combinațiilor de morfeme și cuvinte, lăsând în umbră o analiză logică mai profundă și cuprinzătoare. Această obiecție sugerează că limbajele artificiale sunt, în esență, construite pentru a fi sisteme de reguli și structuri foarte ordonate. Esperanto, de exemplu, a fost creat pentru a fi o limbă internațională auxiliară și pentru a facilita comunicarea între oameni din diferite țări și culturi. Astfel, a avut tendința de a se baza pe reguli stricte pentru a asigura claritatea și consistența în comunicare [80].

În ciuda acestui fapt, obiecția lui Freudenthal pune sub semnul întrebării profunzimea și complexitatea limbajelor artificiale în comparație cu limbajele naturale. Limbajele naturale, precum engleza, franceza sau chineza, sunt adesea bogate în ambiguitate și nuanțe semantice. Ele pot avea multiple sensuri pentru aceeași expresie sau pot implica referințe metaforice sau metonimice. Această bogăție semantică și pragmatică face limbajele naturale mai puțin previzibile și mai puțin restrictive în comparație cu limbajele artificiale.

În schimb, limbajele artificiale, datorită accentului lor pe regularitate și structură bine definită, pot să piardă din flexibilitatea și complexitatea limbajelor umane. Ele pot să nu fie la fel de adecvate pentru a exprima nuanțe subtile sau pentru a face față diversității abordărilor comunicative care caracterizează limbajele naturale. Așadar, în ciuda eforturilor de a crea sisteme lingvistice ordonate și consistente, limbajele artificiale pot să rămână limitate din punct de vedere al adâncimii și complexității lor semiotice în comparație cu limbajele naturale, care evoluează organic în răspuns la nevoile și experiențele umane variate [80].

Conceptul de universali lingvistici a fost dezbătut încă din cele mai vechi timpuri pentru a elucida caracteristicile distinctive ale limbajului uman. Acest concept presupune că există anumite trăsături sau structuri gramaticale fundamentale care sunt prezente în toate limbile naturale, indiferent de diversitatea lor aparentă. Deci, se consideră că limbajul uman are o bază comună care transcende varietatea limbilor vorbite în întreaga lume. Cu toate acestea, în ciuda ideii de universalitate lingvistică, limbajele naturale se remarcă prin complexitatea lor și prin varietatea de structuri și reguli gramaticale. Astfel, chiar dacă există trăsături comune în limbajele umane, statutul secvențelor corect formate în aceste limbi nu poate fi definit cu precizie absolută.

Acest aspect poate fi ilustrat prin observarea limbilor naturale din diverse culturi și regiuni geografice. Fiecare limbă naturală prezintă caracteristici unice și particularități gramaticale, care reflectă istoria, cultura și evoluția sa. De exemplu, există limbi care au un sistem foarte elaborat de declinări pentru substantiv și altele care se bazează pe ordinea cuvintelor pentru a exprima relațiile gramaticale. Acest lucru arată că limbajele naturale sunt extrem de diverse și adaptate la specificul contextului lor cultural și social. Așadar, în ciuda ideii generale de universali lingvistici,

nu există un model lingvistic universal care să se potrivească perfect tuturor limbilor naturale. Statutul secvențelor corect formate poate varia semnificativ de la o limbă la alta, în funcție de regulile și structurile specifice ale fiecăreia [80].

Această diversitate lingvistică reflectă, în cele din urmă, abilitatea adaptativă a limbii umane, care se dezvoltă în răspuns la nevoile și experiențele complexe ale comunităților lingvistice. Astfel, limbajele naturale reprezintă un fenomen complex și dinamic, care se schimbă în timp și se adaptează la schimbările culturale și sociale, făcând studiul lor o provocare continuă și fascinantă pentru cercetătorii din domeniul limbajului.

Statutul secvențelor corect formate în limbile naturale este un aspect complex și în continuă evoluție. Într-adevăr, limbajele naturale, precum engleza, franceza sau oricare alta, nu pot fi reduse la reguli și structuri lingvistice stricte. Ele au o dinamică proprie, care se adaptează și evoluează în funcție de schimbările culturale, sociale și lingvistice. Această flexibilitate le permite să exprime o gamă extrem de largă de idei și concepte, dar poate crea și ambiguitate, diversitate și nuanțe semantice care pot depăși limitele unui model lingvistic rigid.

Un aspect interesant este legat de conceptul de infinitate potențială în limbile naturale. Acesta se referă la faptul că, dacă două expresii corect formate, cum ar fi „A” și „B”, pot fi combinate corect pentru a forma „A ȘI B”, atunci acest proces poate fi repetat la nesfârșit. De exemplu, putem avea „A ȘI B ȘI C„ sau „A ȘI B ȘI C ȘI D”, și așa mai departe. Această abilitate de a itera și extinde expresiile valide este una dintre caracteristicile fascinante ale limbilor naturale și contribuie la bogăția lor expresivă [80].

Pe de altă parte, când discutăm despre sistemele de inteligență artificială, cum ar fi ChatGPT, întrebarea care se pune este dacă acestea posedă aceeași capacitate de infinitate potențială. În general, aceste sisteme sunt construite pe baza unor modele și reguli lingvistice învățate din date de antrenament, și nu au o înțelegere semantică profundă sau capacitatea de a extinde automat expresiile într-un mod logic și semantic. Prin urmare, dacă un utilizator introduce o secvență complexă sau iterează expresii într-un mod neregulat sau neașteptat, ChatGPT poate avea dificultăți în a produce răspunsuri coerente sau adecvate.

Astfel, în ciuda capacităților sale impresionante de procesare a limbajului natural, ChatGPT și alte sisteme similare au limite în legătură cu adaptarea la construcții lingvistice neconvenționale sau la infinitatea potențială a limbilor naturale. Aceste sisteme funcționează în baza unor modele prestabilite și pot fi limitate de acestea în interpretarea și generarea de expresii complexe. Este important să se înțeleagă aceste limite atunci când se lucrează cu astfel de sisteme și să se aibă în vedere că interpretarea lor a limbajului poate varia în funcție de context și de complexitatea secvențelor introduse.

Decalajul dintre competență și performanță [129] reprezintă un concept esențial în studiul limbilor naturale și al limbajului artificial. Acest decalaj devine evident în momentul în care analizăm complexitatea și adaptabilitatea limbilor naturale în comparație cu limbajele artificiale. În esență, competența lingvistică se referă la cunoștințele și capacitățile intrinseci ale unei persoane în ceea ce privește limbajul. Ea include înțelegerea regulilor gramaticale, sintaxa, vocabularul și structura de bază a limbii. Competența lingvistică este, într-o oarecare măsură, înnăscută sau dobândită în mod precoce, reprezentând baza teoretică a limbajului. Aceasta este partea abstractă și teoretică a limbajului, care poate fi modelată prin mecanisme generative caracterizate prin reguli.

Pe de altă parte, performanța lingvistică se referă la utilizarea practică a limbajului în contextul comunicării. Aici, intră în joc diverse restricții contingente, precum locul, timpul, capacitățile fizice, psihice și sociale ale vorbitorului. Performanța lingvistică se referă la modul în care individul folosește și adaptează competența sa lingvistică pentru a comunica eficient. Ea poate varia în funcție de context și de circumstanțe și poate include erori sau dificultăți în exprimarea ideilor în mod adecvat. Decalajul dintre aceste două aspecte este evident în orice limbă naturală, întrucât competența lingvistică a vorbitorului este adesea mai profundă și mai complexă decât performanța sa lingvistică cotidiană. Acest decalaj apare întrucât intervine distincția între ceea ce este înnăscut și ceea ce este dobândit, dintre ceea ce aparține speciei umane și ceea ce aparține individului. Cu alte cuvinte, competența lingvistică reprezintă cunoștințele și abilitățile comune tuturor vorbitorilor unei limbi, în timp ce performanța este influențată de experiența personală și de particularitățile individuale [129].

Acest fenomen al decalajului dintre competență și performanță nu este limitat doar la limbile naturale, ci este întâlnit și în contextul limbajelor artificiale. Asemenea limbajelor naturale, chiar și limbajele artificiale, precum cele folosite în programarea calculatoarelor sau în rețelele neuronale artificiale, prezintă o diferență distinctă între cunoștințele teoretice (competența) și modul în care aceste limbaje sunt aplicate în proiecte practice (performanța). În cazul limbajelor artificiale, competența lingvistică se referă la cunoștințele teoretice și la regulile care definesc limbajul respectiv. Aceste reguli pot include sintaxa, semantica și funcționalitățile specifice ale limbajului. Competența în limbajul artificial poate fi dobândită prin studiu și învățare și este deseori înrădăcinată în documentația și specificațiile limbajului.

Chiar și așa, modul în care limbajele artificiale sunt folosite în proiectele practice, adică performanța lor, poate varia considerabil în funcție de programatorul sau de echipa care le utilizează. Performanța lingvistică este influențată de nivelul de expertiză al utilizatorului și de modul în care acesta aplică competența sa în limbajul respectiv în contextul proiectului [129]. Aceasta poate include abilități precum proiectarea eficientă a algoritmilor, gestionarea memoriei,

optimizarea codului și gestionarea erorilor. Un aspect important de menționat este că performanța limbajelor artificiale este deseori dependentă de deciziile și alegerile programatorilor. Modul în care sunt structurate și organizate codurile, care biblioteci de programe sau *framework*-uri sunt utilizate, precum și cum se gestionează aspecte practice, cum ar fi performanța și scalabilitatea, pot influența semnificativ modul în care limbajele artificiale sunt aplicate în practică.

Așadar, decalajul dintre competență și performanță nu este specific doar limbilor naturale, ci este o caracteristică relevantă și în contextul limbajelor artificiale. Acest decalaj subliniază importanța înțelegerii profunde a limbajelor, atât în cadrul limbilor naturale, cât și în cel al limbajelor artificiale, și subliniază rolul esențial al experienței și expertizei în aplicarea eficientă a acestora în proiecte practice.

Conceptul de semnificație în limbaj este un aspect fundamental atât în lingvistică, cât și în dezvoltarea limbajelor artificiale. Paul Garvin, în lucrarea sa „*What is Linguistics?*" publicată în 1971 [80], a adus în discuție două abordări semantice distincte: semantica aditivă și semantica integrativă, care ne pot ajuta să înțelegem modul în care limbajele funcționează și cum sunt percepute semnificațiile în acestea.

Semantica aditivă, așa cum a fost formulată de Garvin, se referă la limbajele artificiale și modul în care semnificația unei secvențe este construită prin simpla adiție a semnificațiilor termenilor care o compun. Cu alte cuvinte, în limbajele artificiale, înțelegerea unei fraze sau expresii este rezultatul sumei semnificațiilor cuvintelor sau termenilor care o alcătuiesc. Această abordare poate fi eficientă în limbajele cu structuri logice și formale, unde semantica este definită într-un mod clar și riguros.

În schimb, semantica integrativă se referă la limbile naturale și modul în care semnificația întregului nu poate fi redusă la suma semnificațiilor părților componente. În limbile naturale, semantica este influențată de context și de modul în care cuvintele sau termenii sunt combinați. Această abordare integrativă recunoaște că semantica limbilor naturale poate depăși simpla adiție a semnificațiilor și poate implica nuanțe, metafore sau expresii idiomatice care nu pot fi înțelese prin descompunerea simplă a frazei [80].

Un exemplu relevant este acela că limbile naturale abundă în expresii idiomatice care nu urmează o logică aditivă strictă. Aceste expresii sunt parte a culturii și a folclorului limbii și pot fi greu de înțeles pentru cei care nu sunt vorbitori nativi sau nu au cunoștințe profunde ale limbii, aspect care subliniază complexitatea semanticii limbilor naturale și necesitatea înțelegerii lor în context.

În orice limbaj uman, fie el natural sau artificial, există un echilibru între aspectul atomistic-analitic și cel integrativ-holistic. Limbajele pot varia în măsura în care pun accentul pe una dintre

aceste abordări, iar acest echilibru poate influența modul în care limbajul este perceput și utilizat în comunicare. De exemplu, limbajele care își fărâmițează prea mult semnificațiile, așa cum se poate întâmpla în cazul limbajelor artificiale precum ChatGPT, pot deveni mai puțin intuitive și pot pierde din funcția lor euristică.

Utilizarea limbajelor, fie ele naturale sau artificiale, reflectă adesea scopurile și finalitățile pentru care au fost create sau dezvoltate. În cazul limbajelor artificiale, acestea sunt, de obicei, elaborate cu o destinație specială sau un scop specific. Ele sunt concepute pentru a rezolva anumite probleme, pentru a îndeplini sarcini precise sau pentru a servi unui domeniu particular, cum ar fi programarea, matematica, științele naturale sau ingineria. Acest caracter specializat face ca limbajele artificiale să fie instrumente puternice în contextul pentru care au fost create, dar poate limita adaptabilitatea lor în alte domenii.

Însă, limbajele naturale, cum ar fi româna, engleza sau chineza, nu sunt dezvoltate având în vedere o destinație specială. Limbile naturale au evoluat de-a lungul mileniilor pentru a satisface nevoile de comunicare ale oamenilor într-un mod universal. Limbajele naturale sunt flexibile și pot fi aplicate într-o varietate de contexte, de la conversațiile de zi cu zi și comunicarea în societate, până la literatură, filosofie și artă. Această universalitate face ca limbajele naturale să fie instrumente extrem de puternice pentru transmiterea și exprimarea ideilor și a gândurilor umane într-un mod bogat și divers [80].

În pofida acestui fapt, limbajele artificiale și limbajele naturale nu sunt complet separate în utilizarea lor. În lumea tehnologică de astăzi, limbajele naturale sunt adesea procesate și interpretate de sisteme de inteligență artificială, precum ChatGPT, pentru a efectua sarcini precum traducerea automată, analiza textului sau generarea de conținut. Acest lucru arată că, în anumite cazuri, limitele dintre limbajele naturale și cele artificiale pot deveni mai puțin clare, iar utilizarea poate depăși frontierele inițiale.

Ambiguitatea este un alt aspect semnificativ al limbajelor naturale care poate aduce complexitate în înțelegerea și interpretarea comunicării. În studiul lingvistic al limbajelor, se recunoaște că ambiguitatea apare într-o varietate de contexte și poate fi explorată prin prisma universaliilor lingvistice. Universaliile lingvistice reprezintă trăsături și modele comune în structura limbajelor umane care sunt valabile într-un mod generalizat în întreaga lume. Un aspect notabil al universaliilor lingvistice este opoziția dintre o categorie marcată și alta nemarcată. Aceasta se referă la existența sau absența unei trăsături sau categorii în limbaj. Un exemplu concret poate fi observat în limba engleză, în care cuvântul „*man*” poate avea două semnificații distincte. Poate semnifica „bărbat” sau „ființă umană” [80]. În acest caz, apare o distincție între categoria masculinului (bărbat) și categoria mai generală a ființei umane (om), iar această distincție poate

crea ambiguitate în anumite contexte. De exemplu, în propoziția „*This man is strong*” cuvântul „*man*” poate fi interpretat ca „bărbat” sau ca „ființă umană,” iar sensul depinde de context.

Ambiguitatea lingvistică poate apărea și în alte moduri, precum cele legate de structura sintactică, semantica cuvintelor, sau utilizarea expresiilor idiomatice. Aceasta poate face ca înțelegerea limbajului să fie mai dificilă și poate necesita clarificări suplimentare în comunicare. Chiar și în era tehnologiei, în care sistemele de inteligență artificială lucrează pentru a înțelege și a interpreta limbajul uman, ambiguitatea rămâne o provocare semnificativă. Prin urmare, dezvoltarea algoritmilor și a modelelor de învățare automată care să poată recunoaște și să rezolve ambiguitatea lingvistică reprezintă un domeniu de cercetare în continuă dezvoltare.

Conceptul de „dublă articulare” este esențial în lingvistică și arată modul în care limbile naturale sunt organizate și structurate pentru a exprima semnificații complexe prin intermediul unor unități mai mici, dotate cu semnificație. Acest concept dezvăluie un nivel profund de analiză și descompunere a limbajului care face limbile naturale extrem de eficiente în transmiterea gândurilor și ideilor. În dubla articulare, o secvență de cuvinte, cum ar fi „cal frumos,” este analizabilă în morfeme, care sunt cele mai mici unități de bază cu semnificație într-o limbă. În acest caz, „cal” și „frumos” sunt morfeme care pot avea semnificații lexicale sau gramaticale distincte. Morfemele, la rândul lor, pot fi analizate în foneme, care sunt cele mai mici unități sonore din limbaj. Deși fonemele nu au, în general, semnificație lexicală sau gramaticală individuală, ele pot influența semnificația cuvintelor și a morfemelor când sunt combinate sintagmatic.

Un aspect important al dublei articulări este discrepanța între numărul de foneme și numărul de morfeme în limbile naturale. Acest lucru înseamnă că există mai puține foneme decât morfeme într-o limbă dată. Totuși, datorită capacității combinatorii foarte mari a limbilor naturale, un număr relativ mic de foneme poate genera un număr mare de semnificații diferite prin intermediul morfemelor și cuvintelor. Avantajul dublei articulări constă în faptul că limbile naturale pot exprima o gamă largă și variată de idei și concepte folosind un set relativ restrâns de foneme și morfeme. Aceasta face limbile naturale extrem de economice și eficiente din punct de vedere al stocării și transmiterii informației [80].

Limbajul uman este extrem de versatil și nu se limitează doar la transmiterea de informații, ci are și un rol puternic în exprimarea creativității, a gândirii abstracte și a abilităților retorice. În acest sens, prezența unei retorici în limbajul uman, fie sub formă de metafore sau metonimii, este un aspect semnificativ.

Metafora și metonimia sunt două figuri de stil care contribuie la expresivitatea limbajului și la capacitatea acestuia de a reda idei și concepte în moduri neașteptate și pline de înțelesuri ascunse. Metafora implică folosirea unui termen sau a unei expresii într-un sens figurat, adesea

comparativ, pentru a evidenția o similitudine sau o conexiune între două concepte aparent diferite. Metonimia, pe de altă parte, implică folosirea unui termen pentru a reprezenta altceva cu care este strâns legat sau asociat [98].

Aceste figuri de stil nu sunt doar instrumente poetice, ci sunt prezente în mod obișnuit în limbajul cotidian. Ele ne permit să exprimăm idei complexe, să sugerăm relații subtile și să creăm imagini vii în mintea ascultătorilor sau cititorilor. De exemplu, atunci când spunem „coroană", putem face referire la idee de monarhie în ansamblul ei, nu doar la piesa de îmbrăcăminte. Sau, atunci când vorbim despre „casa albă", nu ne referim numai la clădirea în sine, ci și la administrația prezidențială a Statelor Unite.

În gândirea științifică, procesele de generalizare și abstractizare sunt esențiale pentru dezvoltarea teoriei și pentru înțelegerea fenomenelor complexe. Aceste procese pot fi generatoare de figuri metaforice, deoarece cercetătorii caută adesea să ilustreze relații sau concepte abstracte prin intermediul analogiilor sau a imaginilor figurate. De exemplu, „modelul atomic cu electroni în orbite" poate fi o metaforă pentru a reprezenta modul în care electronii orbitează în jurul nucleului unui atom. Pe de altă parte, deducția și organizarea în etape riguros ordonate pot duce la apariția figurilor metonimice în limbaj [80]. De exemplu, atunci când un cercetător se referă la „studiile din domeniul sănătății" pentru a se referi la cercetările în domeniul medical, se folosește metonimia pentru a reprezenta un întreg domeniu prin intermediul unei părți sau a unei aspecte specifice.

În contextul științific, limbajul nu este doar un instrument de comunicare, ci și un instrument de explorare și descoperire. Funcția euristică a limbajului științific stimulează utilizarea figurilor de stil pentru a ilustra idei și concepte complexe, făcându-le mai accesibile și mai ușor de înțeles pentru publicul larg și pentru comunitatea științifică [124]. Astfel, retorica și structurile prozodice adaugă o dimensiune suplimentară limbajului științific, făcându-l mai captivant și mai expresiv.

Tendința spre poeticitate este o caracteristică fascinantă a limbilor naturale, care se manifestă într-o manieră evident în modul în care oamenii folosesc limbajul pentru a crea expresii și imagini cu impact emoțional și estetic. Această tendință este prezentă în toate limbile naturale și transcende concepțiile individuale despre poeticitate. Limbile naturale sunt intrinsec poetice prin modul lor de a se dezvolta și evolua de-a lungul timpului. Ele nu sunt doar instrumente pentru transmiterea informațiilor, ci și mijloace pentru a exprima emoții, a crea imagini vizuale și a spune povești. Această tendință spre poeticitate se reflectă în utilizarea metaforelor, a simbolurilor, a ritmului și a structurilor sonore ale limbajului. În fiecare limbă naturală, se pot găsi expresii idiomatice, proverbe, poezii și alte forme de expresie artistică care ilustrează modul în care

limbajul poate fi folosit pentru a crea efecte poetice. Aceste expresii adaugă o profunzime și o textură limbii, permițând oamenilor să comunice nu doar informații, ci și sentimente și imaginație [80].

Tendința spre poeticitate nu este limitată doar la literatură sau poezie, ci se regăsește și în limbajul cotidian. Oamenii folosesc adesea expresii sau imagini poetice pentru a face limbajul lor mai viu și mai expresiv. De exemplu, atunci când spunem „un răsărit de soare minunat”, folosim o imagine poetică pentru a descrie frumusețea unui moment.

Posibilitatea autoreferinței în limbaj este o caracteristică distinctivă și, în același timp, fascinantă a limbilor naturale. Aceasta se referă la capacitatea unei limbi de a se referi la ea însăși, de a discuta despre limbaj și de a reflecta asupra propriilor sale structuri și reguli într-un mod recursiv, fără a părăsi mediul limbajului. Această abilitate este foarte evidentă în limbile naturale și reprezintă o caracteristică semnificativă a comunicării umane.

Paul Garvin [80] a subliniat această caracteristică remarcabilă în limbile naturale, afirmând că în limbile naturale putem vorbi despre limbă fără a fi nevoie să ne exteriorizăm din cadrul limbajului însuși. Cu alte cuvinte, în timp ce utilizăm o limbă, avem capacitatea de a discuta și de a analiza structura și regulile limbajului respectiv. Această capacitate de autoreferință deschide uși pentru reflecție, introspecție și dezvoltare în comunicarea umană. Un exemplu simplu de autoreferință în limbaj poate fi găsit în propoziția: „Cuvântul 'carte' are cinci litere.” Aici, cuvintele sunt folosite pentru a descrie proprietăți ale cuvintelor înșiși, fără a ieși din cadrul limbajului. Acest tip de autoreferință este comun și facil în limbile naturale. În plus, posibilitatea autoreferinței în limbaj susține dezvoltarea gramaticii și a lingvisticii. Lingviștii și cercetătorii limbajului pot utiliza limbajul pentru a discuta despre structurile limbii și pentru a crea termeni și concepte specializați pentru a descrie fenomenele lingvistice. Această abilitate de a vorbi despre limbaj în limbaj a contribuit la evoluția științei limbajului și la înțelegerea mai profundă a funcționării limbilor naturale.

Imposibilitatea învățării limbajului înainte de utilizarea sa este o caracteristică esențială a limbajului uman și subliniază faptul că limbajul este în primul rând un instrument de comunicare și interacțiune. Procesul de a învăța un limbaj înainte de a îl utiliza poate părea o enigmă, aproape ca o încercare de a merge înainte de a învăța să te târăști. De fapt, limbajul, în esența sa, este un vehicul al interacțiunii și comunicării; se dezvoltă și prosperă în terenul fertil al comunității și al interacțiunii sociale. La naștere, ființele umane nu posedă cunoștințe despre o limbă anume sau reguli gramaticale specifice [80]. În schimb, ele dezvoltă abilități lingvistice pe măsură ce interacționează cu mediul lor și cu vorbitorii nativi ai limbii respective. Acest proces de învățare a limbajului începe de la o vârstă foarte fragedă, când copiii sunt expuși la sunetele și structurile

limbii vorbite în jurul lor. În plus, învățarea limbajului devine un proces circular, în care utilizarea sa alimentează înțelegerea, iar înțelegerea stimulează mai departe utilizarea.

Învățarea limbajului nu este un proces static sau izolat, ci este unul continuu și dinamic, care evoluează pe măsură ce individul crește și se dezvoltă. Copiii încep cu înțelegerea și producerea sunetelor limbajului, apoi trec la cuvinte, propoziții și discursuri mai complexe. Prin interacțiunea cu membrii familiei, prietenii și comunitatea, ei își dezvoltă abilitățile lingvistice într-un mod natural și progresiv. A învăța limbajul prin utilizarea sa înseamnă a țese o plasă de înțelesuri și conexiuni care sunt profund încorporate în experiența de viață a unei persoane. Această viziune asupra învățării limbajului subliniază ideea că limbajul este viu, în constantă schimbare și evoluție. Prin îmbrățișarea acestei naturi fluide și dinamice, putem vedea că învățarea limbajului nu este o sarcină ce poate fi bifată pe o listă, ci un proces continuu de explorare și descoperire care îmbogățește și aprofundează experiența umană [98].

Un aspect important al acestei imposibilități de a învăța limbajul înainte de utilizare este că limbajul este un fenomen social. Învățarea limbajului implică nu doar înțelegerea structurii lingvistice, ci și a contextului social, a normelor culturale și a semnificațiilor compartimentate de comunitatea lingvistică. Acest aspect evidențiază legătura strânsă dintre limbaj și societate și modul în care acestea sunt interconectate.

Umorul și ironia au un rol esențial în limbajul uman natural, iar acestea pot fi percepute ca modalități subtile prin care individul explorează și înțelege mai bine mediul și ceilalți.

Într-un text celebru, James Joyce explorează o temă profunde legate de limbajul și conștiința umană, de cunoaștere și de libertatea de gândire. În romanul său *Ulise*, Joyce surprinde ideea că mintea umană poate fi adesea captivă într-o serie de gânduri și convingeri, limitându-și astfel perspectiva și capacitatea de a percepe lumea într-un mod deschis și liber („râse, ca să-și elibereze mintea de sub robia minții"). Joyce sugerează că umorul și ironia pot servi, în contextul limbajului uman, drept mijloace de a scăpa de constrângerile rigide ale gândirii preconcepute. Prin râs și autoironie, individul poate să-și elibereze mintea și să-și exploreze gândurile și percepțiile cu mai multă libertate. Limbajul uman este unic în capacitatea sa de a exprima umorul și ironia. Aceste forme de exprimare se bazează pe jocul cu cuvintele, pe ambiguitate și pe abilitatea de a vedea lumea într-un mod neconvențional [80]. Capacitatea de a folosi umorul în limbaj indică un anumit nivel de inteligență și creativitate. Râsul, în cadrul limbajului natural, nu este doar o reacție mecanică la anumite situații, ci poate avea semnificații variate, incluzând umorul, bucuria, ironia, surpriza sau chiar inconfortul social. Prin intermediul râsului, oamenii pot exprima o gamă largă de emoții și intenții, ceea ce îl face un element complex al limbajului non-verbal. Râsul poate fi strâns legat de limbajul verbal. O glumă spusă cu un ton vesel și urmată de râs poate amplifica

efectul umorului. De asemenea, râsul poate fi folosit pentru a sublinia sarcasmul sau ironia într-o conversație, ajutând astfel la interpretarea corectă a intențiilor vorbitorului.

2.3 Recunoașterea formelor și algoritmii de joc. Studii de caz: scrisul de mână și șahul

Recunoașterea formelor

Recunoașterea caracterelor scrise de mână este o abilitate remarcabilă a creierului uman, care a captivat interesul cercetătorilor de-a lungul timpului. În ultimele decenii, dezvoltarea tehnologiei a deschis calea pentru a explora și a dezvolta metode de recunoaștere automată a acestor caractere, prin intermediul rețelelor neuronale artificiale. Această abordare a generat rezultate impresionante, dar persistă întrebarea dacă mașinile pot atinge sau depăși capacitatea noastră umană de a recunoaște și a înțelege modelele vizuale.

Chiar dacă inteligența artificială a progresat semnificativ în acest domeniu, rămânem încă impresionați de abilitățile noastre umane de a percepe și de a interpreta cu ușurință scrisul de mână, chiar și în cele mai variate forme și stiluri. Această distincție subliniază unicitatea noastră umană și complexitatea extraordinară a capacităților noastre cognitive.

Așadar, în timp ce inteligența artificială poate oferi soluții eficiente și rapide pentru recunoașterea caracterelor scrise de mână, abilitățile noastre umane de percepție vizuală, bazate pe principiile gestaltiste și pe experiență, rămân de neegalat. Recunoașterea scrisului de mână reprezintă o demonstrație a capacității noastre de a interacționa cu lumea într-un mod holistic și profund, aducând în discuție întrebări filosofice despre natura conștiinței și a cunoașterii umane.

Pattern recognition, sau recunoașterea modelelor, reprezintă fundamentul esențial în recunoașterea caracterelor scrise de mână, indiferent dacă este realizată de oameni sau de mașini. Acest proces complex implică identificarea și interpretarea trăsăturilor distinctive ale unui model și atribuirea acestuia unei categorii sau unui înțeles specific.

Cercetările în domeniul recunoașterii *pattern*-urilor au evidențiat că mintea umană posedă o abilitate incredibilă de a recunoaște și a înțelege modele vizuale. În *pattern recognition*, există câteva principii de bază care sunt aplicate pentru a identifica și a recunoaște modele [126].

Astfel, **principiul similarității** se referă la identificarea și clasificarea obiectelor sau a modelelor pe baza asemănărilor dintre ele. Obiectele care au caracteristici și trăsături comune sunt grupate împreună și clasificate în aceeași categorie. Similaritatea poate fi evaluată în funcție de caracteristici precum forma, culoarea, textura sau alte atribute relevante.

Principiul proximității descrie tendința de a grupa obiectele sau modelele care sunt spațial apropiate unele de altele. Obiectele care sunt mai apropiate între ele sunt percepute ca fiind mai relevante sau ca făcând parte dintr-un ansamblu coerent. Proximitatea poate fi utilizată pentru a identifica relații sau modele într-un set de date.

Principiul continuității explică abilitatea de a percepe linii și forme continue, chiar și atunci când acestea sunt întrerupte sau parțial ascunse. Mintea umană completează informațiile lipsă pentru a crea o imagine coerentă și continuă. Continuitatea poate fi utilă în identificarea și recunoașterea unor forme sau trăsături care se extind dincolo de limitele aparente.

Principiul simetriei se referă la perceperea și recunoașterea formelor sau a modelelor care sunt simetrice. Simetria poate fi utilizată pentru a identifica modele într-un set de date și pentru a clasifica obiecte în funcție de această caracteristică.

Principiul invarianței descrie capacitatea de a recunoaște un model, indiferent de poziția, dimensiunea sau orientarea acestuia. Mintea umană poate percepe și recunoaște un model într-o varietate de contexte și variante. În recunoașterea automată a modelelor, se caută și se utilizează caracteristici invariante pentru a asigura o recunoaștere precisă și robustă.

Aceste principii de bază în *pattern recognition* sunt inspirate din modul în care funcționează sistemul vizual uman și sunt aplicate în proiectarea diverșilor algoritmi utilizați pentru recunoașterea automată a modelelor.

În rândurile care urmează vom analiza un studiu de caz, în care vom evalua performanțele unei rețele neuronale artificiale de tip convolutiv antrenată pentru recunoașterea cifrelor de la 0 la 9 scrise de mână.

În acest scop, vom folosi baza de date MNIST (*Modified National Institute of Standards and Technology*), care este una dintre cele mai cunoscute și utilizate baze de date în domeniul recunoașterii cifrelor scrise de mână. Baza de date a fost creată în Statele Unite ale Americii de către Institutul Național de Standarde și Tehnologie și a fost modificată ulterior pentru a fi utilizată în cercetarea și dezvoltarea algoritmilor de recunoaștere a cifrelor [143].

MNIST conține un set de imagini alb-negru, format din cifre scrise de mână de la 0 la 9. Fiecare imagine are o rezoluție de 28x28 de pixeli și reprezintă o cifră singulară. Baza de date cuprinde un total de 60000 de exemple de antrenament și 10000 de exemple de testare.

Pentru a construi baza de date MNIST s-au colectat mostre de scriere de mână de la mai multe persoane. Participanții au fost rugați să scrie cifrele pe o suprafață digitală sau pe hârtie, iar aceste imagini au fost digitalizate și standardizate pentru a se potrivi rezoluției de 28x28 de pixeli.

MNIST a fost introdusă pentru prima dată în anul 1998, ca o resursă pentru evaluarea algoritmilor de recunoaștere a cifrelor scrise de mână. De atunci, a devenit o bază de date de referință în domeniul învățării automate și al recunoașterii de modele. MNIST a fost și este încă utilizată foarte mult pentru antrenarea și testarea algoritmilor de învățare automată, inclusiv pentru rețele neuronale, în vederea recunoașterii cifrelor scrise de mână.

În studiul nostru de caz (vezi ***Fig. 2. 6***), folosind limbajul de programare Python (în acest moment, cel mai popular limbaj de programare și practic, limbajul cel mai folosit în aplicațiile de analiză a datelor și inteligență artificială), am proiectat o rețea neuronală de tip convolutiv și am antrenat-o pe setul de date MNIST.

Rețelele neuronale convolutive sunt preferate în fața altor tipuri de rețele neuronale, cum ar fi rețelele complet conectate, din mai multe motive. În primul rând, datorită utilizării operației de convoluție, rețele neuronale convolutive pot captura relațiile spațiale dintre pixeli sau dintre regiuni vecine dintr-o imagine. Acest fapt le permite să înțeleagă contextul local și să extragă caracteristici relevante, cum ar fi contururile sau texturile.

Prin utilizarea filtrelor convoluționale și a straturilor de subeșantionare (*pooling*), rețelele neuronale convolutive pot reduce numărul de parametri necesari pentru a reprezenta o imagine, amănunt care facilitează antrenamentul și previne supraînvățarea, în timp ce menține capacitatea rețelei de a extrage caracteristici semnificative.

De asemenea, rețelele neuronale convolutive și-au demonstrat capacitatea de a realiza un transfer de învățare eficient. Aceasta înseamnă că ele pot fi antrenate pe un set de date mare și complex, cum ar fi *ImageNet* sau *MNIST*, și apoi pot fi utilizate și reglate fin (*fine tuned*) pentru sarcini mai specifice, cum ar fi recunoașterea cifrelor scrise de mână. Transferul de cunoștințe permite accelerarea procesului de antrenare și îmbunătățirea performanței.

Trebuie adăugat că rețelele neuronale convolutive sunt robuste în fața variațiilor de iluminare, deformațiilor și zgomotului în imagini, putând, astfel, generaliza informațiile învățate și realizând previziuni corecte, chiar și în cazul unor condiții suboptime.

Datorită arhitecturii lor optimizate pentru recunoașterea imaginilor, rețelele neuronale convolutive au obținut rezultate remarcabile în numeroase sarcini de recunoaștere, cum ar fi clasificarea obiectelor, segmentarea semantică sau detecția obiectelor, performanțe care au depășit alte metode tradiționale și au obținut performanțe comparabile cu cele umane în anumite sarcini.

Acestea sunt și motivele pentru care în acest studiu de caz am ales o rețea neuronală de acest tip. Arhitectura este o rețea neuronală convoluțională tipică, utilizată pentru recunoașterea cifrelor scrise de mână. Descrierea tehnică a rețelei nu este greu de înțeles, chiar și pentru un nespecialist în domeniu, grație analogiei cu structurile neuronale cerebrale.

Punctul de start a constat în definirea unui model secvențial, care reprezintă o succesiune liniară de straturi cu neuroni [162]. Apoi a fost adăugat primul strat convoluțional de neuroni (*Conv2D*), cu 32 de filtre, fiecare filtru fiind o matrice de dimensiune (3, 3). Aceste filtre aplică operația de convoluție pe imaginea de intrare. Funcția de activare de tip *relu* este utilizată pentru a introduce non-linearitate în rețea. Parametrul *input_shape* specifică dimensiunea imaginilor de

intrare (28x28x1 în acest caz, unde 1 reprezintă numărul de canale de culoare - de fapt, este o scală pentru nuanțele de gri).

După acest strat a fost adăugat un nou strat (*MaxPooling2D*) care reduce dimensiunile matricei prin selectarea valorilor maxime din fiecare regiune. Acest lucru ajută la extragerea caracteristicilor relevante și la reducerea numărului de parametri.

În continuare, am repetat pasul 2 și pasul 3 pentru a adăuga încă două straturi de tip *Conv2D* și *MaxPooling2D*. Aceste straturi adaugă mai multe filtre și reduc dimensiunea imaginilor în continuare, permițând rețelei să învețe caracteristici mai abstracte și complexe.

După aceste straturi este absolut necesar un strat de aplatizare (*Flatten*), care transformă datele tridimensionale într-un vector unidimensional (elementele dintr-o matricele sunt vectorizate, adică sunt așezate unul după altul). Această operație face tranziția de la straturile de extragerea caracteristicilor într-o reprezentare liniară necesară pentru straturile complet conectate care urmează.

După stratul care realizează vectorizarea datelor, am adăugat un strat complet conectat cu 64 de neuroni (numit *Dense*). Complet conectat înseamnă că cei 64 de neuroni se conectează la toți neuronii din stratul anterior.

În continuare, a fost adăugat un strat de *Dropout*, care, în timpul antrenamentului, dezactivează o proporție din totalul de neuroni de pe acest strat (în acest caz, 50%), pentru a preveni supraînvățarea (*overfitting*-ul) și pentru a îmbunătăți, astfel, capacitatea de a generaliza a modelului [161].

În final, adăugăm un ultim strat complet conectat cu 10 neuroni, corespunzător celor 10 clase posibile (cifrele de la 0 la 9). Acest ultim strat folosește funcția de activare *softmax* pentru că ne furnizează o distribuție de probabilitate peste cele 10 clase posibile. Prin urmare, rețeaua noastră poate genera predicții pentru fiecare clasă și putem alege clasa cu cea mai mare probabilitate ca rezultat final al recunoașterii.

Arhitectura descrisă mai sus este o arhitectură clasică pentru o rețea neuronală de tip convolutiv, cu straturi convolutive pentru extragerea caracteristicilor și straturi complet conectate pentru clasificare. Acest tip de arhitectură este eficient în recunoașterea imaginilor și a obținut rezultate remarcabile în diverse sarcini de clasificare și recunoaștere vizuală.

Prin aplicarea operației de convoluție și a straturilor de *pooling*, rețeaua poate detecta caractere și *pattern*-uri specifice în imagini, în timp ce straturile complet conectate îmbină aceste caracteristici pentru a realiza clasificarea finală. În plus, utilizarea funcțiilor de activare non-liniare, precum *relu* și *softmax*, ajută la introducerea neliniarității în rețea și la obținerea unor predicții mai precise.

În continuare modelul a fost compilat (*compile*). Operația de compilare în modelele de rețele neuronale este responsabilă de configurarea setărilor necesare pentru procesul de antrenare al modelului.

Prin parametrul *optimizer*, putem specifica algoritmul de optimizare utilizat pentru a actualiza ponderile rețelei în timpul antrenării. Există diferite optimizatori disponibili, cum ar fi *Adam*, *SGD* (*Stochastic Gradient Descent*), *RMSprop* etc. În cazul nostru, folosim optimizatorul *Adam*, care este un optimizator popular și eficient în antrenarea rețelelor neuronale. *Adam* utilizează o combinație între algoritmul de gradient descendent și metode bazate pe moment, ceea ce îl face potrivit pentru o varietate de probleme.

Prin parametrul *loss*, am specificat funcția de pierdere (de calcul a erorii) utilizată în timpul antrenării. Această funcție de pierdere cuantifică diferența între predicțiile modelului și etichetele adevărate și servește drept măsură a performanței modelului. În cazul unei probleme de clasificare multiplă, cum este recunoașterea cifrelor, se utilizează funcția *categorical crossentropy*, care compară distribuția de probabilitate obținută de rețea față de etichetele adevărate și măsoară diferența dintre ele [143].

Prin parametrul *metrics*, am ales modul în care dorim să monitorizăm evoluția performanțelor modelului în timpul antrenării. În cazul nostru, am măsurat acuratețea modelului (*accuracy*), prin care ni se arată procentul de predicții corecte făcute de rețea.

Următoarea etapă a fost inițierea procesului de antrenare propriu-zis (*fit*), pe setul de date MNIST, pentru un număr de 12 epoci (*epochs*). O epocă reprezintă parcurgerea completă a tuturor exemplelor de antrenare, deci, în cazul nostru, datele vor fi „vizualizate” de către model de 12 ori. Ca o analogie, este ca și cum un student ar repeta, înainte de examen, de 12 ori fiecare curs din disciplina la care susține examenul.

A fost ales și un set de validare pe care l-am utilizat pentru evaluarea performanței modelului în timpul antrenării. Acest set de date nu este utilizat pentru actualizarea ponderilor, ci doar pentru a obține o evaluare a performanței modelului în afara setului de antrenare. Cu alte cuvinte, aceste imagini ale cifrelor nu sunt folosite în timpul procesului de antrenare a rețelei. După terminarea antrenării, vom folosi acest set de imagini pentru a verifica cât de bine a învățat rețeaua neuronală să recunoască cifrele scrise de mână.

Rezultatul antrenării este stocat în variabila *history*, care conține informații despre evoluția pe parcursul fiecărei epoci de antrenare a parametrilor de măsurare (*metrics*), în cazul nostru pentru acurateță (*accuracy*) și eroare (*loss*). Această informație poate fi ulterior utilizată pentru a vizualiza și analiza performanța modelului în timp [143].

În ansamblu, procesul de antrenare (*fit*) constă în parcurgerea repetată a setului de antrenare pe parcursul a unui număr predefinit de epoci, actualizând ponderile rețelei în funcție de diferența între predicțiile obținute și etichetele reale. Astfel, rețeaua neuronală „învață" să recunoască cifrele scrise de mână, pe măsură ce procesul de antrenare progresează.

Rezultatul antrenării modelului ales este foarte bun. Acuratețea obținută pe setul de testare (validare) este de 99,3%.

Evoluția erorii și a acurateței pe setul de antrenare și pe cel de validare este afișat în ***Fig. 2. 7***

Și mai clar, rezultatele se văd dacă analizăm cu atenție rezultatele din matricea confuziilor (***Fig. 2. 7***). Matricea de confuzie este o matrice care prezintă rezultatele comparate ale clasificării unui model de învățare automată. Această matrice arată cât de bine a prezis modelul fiecare clasă, în comparație cu etichetele reale. Matricea de confuzie este deosebit de utilă în evaluarea performanței unui model de recunoaștere a cifrelor scrise de mână. Prin compararea predicțiilor făcute de model cu etichetele reale ale cifrelor, matricea de confuzie oferă o perspectivă asupra modului în care modelul face distincția între diferite cifre.

Pentru un model de recunoaștere a cifrelor scrise, matricea confuziilor poate avea forma de mai jos:

	Cifre prezise			
	0	1	...	9
Cifra reală 0	TP0	FN0	...	FN0
Cifra reală 1	FP1	TP1	...	FN1
...	…	…	…	...
Cifra reală 9	FP9	FN9	...	TP9

Pe axa verticală sunt cifrele reale (etichetele adevărate, iar pe axa orizontală sunt predicțiile făcute de model. Valorile din matrice reprezintă numărul de cifre clasificate în fiecare categorie.

Elementele matricei sunt definite astfel:

-TP (Adevărat Pozitiv, *True Positive*) reprezintă numărul de cifre clasificate corect ca aparținând unei anumite cifre.

-FN (Fals Negativ, *False Negative*) reprezintă numărul de cifre clasificate incorect ca neaparținând unei anumite cifre, dar care în realitate ar fi trebuit să fie clasificate ca aparținând acelei cifre (eroare de tipul „fals negativ").

-FP (Fals Pozitiv, *False Positive*) reprezintă numărul de cifre clasificate incorect ca aparținând unei anumite cifre, dar care în realitate nu aparțin acelei cifre (eroare de tipul „fals pozitiv”).

-TN (Adevărat Negativ, *True Negative*) reprezintă numărul de cifre clasificate corect ca neaparținând unei anumite cifre.

Analizând matricea de confuzie pentru un model de recunoaștere a cifrelor scrise, putem identifica *pattern*-uri și erori specifice ale modelului. De exemplu, putem observa dacă există confuzii frecvente între anumite cifre (cum ar fi 0 și 9) sau dacă modelul are probleme în distingerea anumitor cifre. Această analiză ne poate ajuta să înțelegem punctele tari și punctele slabe ale modelului și să luăm măsuri pentru a îmbunătăți performanța acestuia.

Cu cât cifrele de pe diagonala principală a matricei sunt mai mari, cu atât rezultatele modelului nostru sunt mai bune (***Fig. 2. 8***).

În cazul nostru, observăm că din cele 10000 de date de testare o majoritate copleșitoare au fost clasificate corect (cifrele de pe diagonala principală sunt de sute de ori mai mari decât cifrele cumulate din celelalte celule ale matricei).

În cazul cifrei 0, din 980 de fotografii ale cifrei 0 scrisă de mână (fotografii folosite pentru validarea antrenării) un număr de 977 au fost recunoscute corect, adică 99,69%. Rata erorii în cazul clasificării cifrei zero este de 0,3%.

În cazul cifrei 1, din 1135 de fotografii ale cifrei 1 scrisă de mână, un număr de 1128 au fost recunoscute corect, adică 99,38%. Rata erorii în cazul clasificării cifrei zero este de 0,61%.

Rezultatul este remarcabil, având în vedere complexitatea scrisului de mână și comparând cu sistemele mai vechi de recunoaștere a cifrelor de mână. Iar acest rezultat poate fi în continuare îmbunătățit.

Asemeni lui ChatGPT, această rețea neuronală pare a realiza performanțe comparabile, în anumite cazuri particulare, cu cele ale oamenilor. Dar, la fel ca în studiul de caz bazat pe interacțiunea cu ChatGPT, am ales și în acest caz să creăm contexte situaționale care să testeze rețeaua neuronală în cazuri pe care chiar și oameni cu o instruire mai puțin temeinică le rezolvă fără probleme.

În acest sens, am analizat cu multă atenție tipurile de cifre de 0 și 1 pe care rețeaua neuronală le-a primit la antrenare. După această analiză, am generat un set de cifre de la 0 la 9, cu care prilej, în mod intenționat, am reprezentat cifrele 0 și 1 într-o manieră proprie (***Fig. 2. 9***). De exemplu, cifra 1 prezintă o buclă la partea superioară, iar cifra 0 este ușor deformată. Ambele cifre sunt deplasate spre stânga în imagini (am făcut acest lucru pentru că am observat că, în general, în baza de date MNIST cifrele sunt așezate aproximativ central). Acest tip de cifre de mână este

recunoscut foarte ușor de orice subiect uman, inclusiv de un elev din ciclul gimnazial sau de persoană cu foarte puțină instrucție școlară (dar care, desigur, știe să scrie și să citească).

În ciuda rezultatelor excelente la antrenare și în faza de validare, rețeaua neuronală convolutivă recunoaște cifra 0 scrisă de mine ca fiind 9, iar cifra 1 ca fiind 6. Mai mult, din cele 10 cifre generate în mod intenționat după un *pattern* diferit de MNIST (dar perfect recognoscibile pentru un om alfabetizat!), rețeaua a reușit contraperformanța să clasifice greșit 4 cifre (pe lângă 0 și 1, 6 a fost clasificat ca fiind 5, iar 8 ca fiind 1). Procentul de erori pe acest set de date, intenționat „alterat" în comparație cu MNIST, este de 40%.

Desigur, introducerea acestor tipuri de date noi în setul de antrenament (cu alte cuvinte, introducerea stilului meu de scriere) va îmbunătăți performanța rețelei neuronale. Dar, problema este că tot eu, sau orice altă persoană, poate crea ușor un nou *pattern* de cifre (banal de recunoscut pentru un om) care să pună în încurcătură rețeaua neuronală. Este ușor de observat asemănarea cu modul în care ChatGPT este indus în eroare.

Nu ne-am propus ca prin acest studiu de caz experimental să demonstrăm ineficiența rețelelor neuronale artificiale, ci, pe linia trasată în lucrarea de față, dorim, de fapt, să accentuăm diferența specifică, mai mult chiar decât genul proxim, dintre ceea ce numim inteligența artificială și inteligența umană.

Ce anume face ca omul să recunoască extrem ușor caracterele scrise de mână, în timp ce procesul acesta este, cel puțin în acest moment, încă foarte dificil de realizat pentru o mașină cu inteligență artificială?

Cu toții avem probleme cu înțelegerea scrisului medicilor de pe rețete de medicamente. Dar, ați observat că farmaciștii nu au această problemă? Am întrebat o farmacistă cum de a înțeles ce scrie pe rețeta mea. Răspunsul ei a fost foarte interesant: „Fiți liniștit, mi-a spus ea cu zâmbetul pe buze, nici eu nu înțeleg scrisul acesta mai bine decât dumneavoastră. Dar este de ajuns ca sa recunosc măcar câteva litere ca să îmi dau seama despre ce medicament este vorba. Toate cutiile cu medicamente au trecut prin mâna mea și, fără să vreau, fac legătura între frânturile de cuvinte inteligibile și medicamentele pe care le-am așezat pe raft."

Deci, contextul situațional, interacțiunea cu realitatea înconjurătoare, o imagine cuprinzătoare care nu cuprinde doar caracterele scrise de pe rețetă, ci un model de realitate mult mai amplu, duc la perceperea sensului cuvintelor. Noi nu doar recunoaștem semne, noi extragem sens din aceste semne (până acolo încât, de multe ori, introducem semnificații acolo unde nu găsim nicio semnificație). Deși contextul este diferit, practic situația este analoaga cu cea analizată în cazul ChatGPT.

Formarea conceptului de număr în mintea umană este un subiect complex și inter- și multidisciplinar (și, finalmente, după cum vom arăta, transdisciplinar), care implică aspecte cognitive, dezvoltare mentală și procese de învățare. Există mai multe teorii și perspective în acest sens.

Teoria empiristă susține formarea conceptului de număr prin experiența senzorială și interacțiunea cu lumea fizică. După teoria empiristă, copiii învață să recunoască și să numere obiectele din jurul lor prin asocierea simțurilor cu cantități și prin observarea regulilor matematice în acțiune. Așa cum am văzut în cazul dezvoltării abilităților lingvistice (cu ocazia studiului de caz legat de ChatGPT), acest punct de vedere este limitat.

Teoria nativistă, susținută, așa cum am văzut deja, de Noam Chomsky [129], afirmă că există anumite structuri cognitive înnăscute sau capacități cognitive specifice care stau la baza conceptului de număr. Conform acestei teorii, există un modul mental dedicat matematicii, care permite oamenilor să înțeleagă și să lucreze cu numere fără a avea nevoie de învățare sau de experiență. Noi credem că acest punct de vedere este cel mai aproape de realitate și că permite, în cele de urmă, trecerea spre înțelegerea din perspectivă filosofică transdisciplinară.

Teoria constructivistă subliniază că numărul, din punct de vedere conceptual, se dezvoltă prin construirea treptată a unui înțeles abstract și a unei reprezentări mentale a numerelor. Prin interacțiunea cu mediul înconjurător, copiii construiesc treptat conceptul de număr pe baza experiențelor lor și a regulilor matematice învățate. Această teorie vine să completeze aspecte importante, care pot fi integrate, alături de teoria lui Chomsky, în modelul transdisciplinar.

Teoria conexionistă se bazează chiar pe modelele de rețele neuronale și susține că formarea conceptului de număr este rezultatul procesării informației și al conexiunilor dintre elementele rețelei neuronale a creierului uman [107]. Folosind aceste modele, numărul poate fi reprezentat și recunoscut prin intermediul activării anumitor conexiuni neuronale. Studiul de caz nr. 2 relevă, după părerea noastră, limitele acestei teorii.

Cu privire la explicația modului în care o rețea neuronală recunoaște cifrele scrise de mână, există diferite abordări, dar majoritatea se bazează pe antrenamentul rețelei folosind seturi de date etichetate. Din exemplul prezentat, se poate constata că rețelele neuronale, folosind algoritmi specifici, sunt antrenate să „învețe” anumite tipuri relații și *pattern*-uri în ceea ce privește formele și caracteristicile cifrelor scrise, astfel încât să poată face predicții corecte cu privire la cifrele necunoscute. Problema este că adaptarea acestor rețele la anomalii este net inferioară capacității creierului uman.

Există câteva deosebiri esențiale între modul în care un om recunoaște o cifră scrisă de mână și modul în care o rețea neuronală recunoaște aceeași cifră.

O primă diferență ar consta în tipul de abordare folosit: de jos în sus (*bottom up*) sau de sus în jos (*top-down*). Un om are capacitatea de a recunoaște cifrele scrisă de mână printr-o abordare **de jos în sus**. Acest lucru înseamnă că noi începem prin a observa caracteristicile de bază ale cifrei, cum ar fi linii, curbe și intersecții, și apoi le combinăm pentru a identifica cifra în ansamblu. În schimb, rețelele neuronale, inclusiv cele care recunosc cifrele scrise de mână, utilizează o abordare **de sus în jos**. În cazul rețelelor neuronale, se începe cu o reprezentare globală a imaginii și se încearcă identificarea caracteristicilor și pattern-urilor specifice ale cifrelor [162].

Referitor la procesarea contextuală, oamenii pot folosi contextul și informațiile din jur pentru a ajuta la recunoașterea cifrelor scrise de mână. De exemplu, dacă vedem o linie curbă în forma unui cerc deschis, putem deduce că este cifra "0". În schimb, așa cum deja am constatat, rețelele neuronale, în general, recunosc cifrele pe baza caracteristicilor intrinseci ale formei, fără a folosi contextul în mod direct, bazându-se pe modele învățate din exemple de antrenament pentru a face predicții pe baza caracteristicilor vizuale ale cifrelor.

Fără îndoială, oamenii sunt de obicei mai flexibili și adaptabili în recunoașterea cifrelor scrise de mână. Chiar și atunci când cifra este scrisă într-un mod ușor diferit sau într-o poziție ușor schimbată, oamenii pot recunoaște în continuare cifra respectivă. Pe de altă parte, rețelele neuronale pot fi sensibile la variațiile minore (lucru demonstrat și în cadrul studiului nostru de caz) și pot avea dificultăți în recunoașterea cifrelor care nu se potrivesc strict cu modelele învățate în timpul antrenamentului. De ce anume oamenii sunt mai flexibili și mai adaptabili, credem că explicația ține de modelul transdisciplinar al minții umane.

Se vede din cele enumerate mai sus că aceste diferențe scot în evidență faptul că procesul de recunoaștere a cifrelor scrise de mână implică o combinație complexă de abilități cognitive și perceptive la nivel uman, în timp ce rețelele neuronale se bazează pe modele și algoritmi matematici pentru a face predicții bazate pe caracteristicile vizuale ale cifrelor, aceste caracteristici fiind, de fapt, transformate în combinații de biți.

După cum constatăm, aceste caracterizări pot fi interpretate și adaptate în diverse moduri și că multiple perspective filosofice pot fi aplicate în înțelegerea și descrierea abilităților rețelelor neuronale, în funcție de context și de interpretarea filosofică specifică.

De exemplu, abordarea lui Carnap în filosofia matematicii este axată pe aspecte logice și sintactice, iar conceptul său de număr se referă la construcții abstracte definite prin intermediul limbajului formal și al regulilor sintactice [22]. Din perspectiva lui Carnap, rețelele neuronale ar putea fi caracterizate ca fiind sisteme de calcul care utilizează modele și algoritmi matematici pentru a învăța și a face predicții bazate pe datele de intrare. Ele ar putea fi privite ca sisteme care interpretează și procesează informațiile și modelele matematice corespunzătoare, într-un mod

similar cu modul în care limbajul formal și regulile sintactice sunt utilizate pentru a defini numerele în abordarea lui Carnap.

Din perspectiva lui Gilbert Ryle, cunoscut pentru conceptul său de „cunoaștere încorporată" și critica față de „mitul minții separate", rețelele neuronale pot fi caracterizate ca fiind sisteme încorporate care învață prin interacțiunea directă cu mediul lor (ce înseamnă „mediul lor" rămâne de dezbătut). Rețele neuronale sunt capabile să „capteze" reguli și *pattern*-uri prin procesarea datelor și a feedback-ului primit din mediul lor (de fapt, de la etichetele atașate, de obicei, de operatori umani). În acest sens, rețelele neuronale pot fi văzute ca un exemplu al modului în care cunoașterea și învățarea sunt încorporate în sistemul lor de procesare a informațiilor [23].

Perspectiva gestaltistă pune accentul pe percepție și organizarea globală a informațiilor. Din perspectiva gestaltistă, rețelele neuronale ar putea fi caracterizate ca sisteme care identifică și extrag caracteristici globale și *pattern*-uri în datele de intrare, într-un mod similar cu modul în care mintea umană percepe și organizează informațiile într-un mod holist. Rețelele neuronale pot identifica structuri și configurații specifice în datele de intrare și pot recunoaște obiecte sau modele pe baza acestor caracteristici globale, asemănător cu procesul de percepție și organizare a informațiilor din perspectiva gestaltistă. În cadrul analizei studiului de caz nr. 3, vom trata mai pe larg anumite aspecte ale teorie gestaltiste.

Care ar putea fi explicațiile pentru care un om recunoaște atât de ușor cifrele și literele scrise de mână? Există, desigur, mai multe explicații pentru faptul că oamenii recunosc relativ ușor cifrele și literele scrise de mână.

O primă explicație ar fi experiența și familiaritatea. Oamenii sunt expuși de la o vârstă fragedă la cifre și litere scrise de mână. De-a lungul timpului, această expunere repetată și familiaritatea cu formele și configurațiile acestor simboluri conduce la dezvoltarea unei experiențe și a unui recunoașteri ușoare. Prin practică și repetiție, oamenii se obișnuiesc cu variabilitatea și variațiile individuale ale formelor scrise de mână și pot recunoaște astfel de variante. Explicația aceasta eșuează atunci când încercăm să explicăm cum recunoaștem, totuși, o cifră al cărei model de scriere personalizat nu l-am mai văzut niciodată.

O a doua explicație este legată de punctul de vedere gestaltist și se referă la capacitatea noastră de procesare globală. Când oamenii recunosc cifrele și literele scrise de mână, ei tind să proceseze aceste forme într-un mod holistic și global. Aceasta înseamnă că observă configurația generală și caracteristicile majore ale simbolului în ansamblu, precum linii, curbe, intersecții și proporții. Procesarea globală facilitează recunoașterea rapidă și automată a formelor familiare.

O altă explicație accentuează înclinația noastră de a sesiza cu ușurință caracteristici distinctive, dat fiind că cifrele și literele scrise de mână au astfel caracteristici distinctive care, în

consecință, facilitează recunoașterea. Aceste caracteristici includ, de obicei, anumite linii și curbe specifice, orientări sau proporții diferite ale componentelor simbolului. Prin identificarea și focalizarea asupra acestor caracteristici distinctive, oamenii pot recunoaște și distinge rapid un simbol de altul.

O ultimă explicație este aceea că recunoașterea cifrelor și literelor scrise de mână este favorizată de contextul și semantica asociate. De exemplu, într-un cuvânt sau într-o frază, contextul și cunoștințele noastre lingvistice ne ajută să anticipăm și să recunoaștem literele și cifrele în funcție de cuvinte sau structuri gramaticale specifice. Această capacitate de a integra contextul și semantica facilitează recunoașterea rapidă și precisă a cifrelor și literelor scrise de mână.

Din fericire, aceste explicații (limitate, dacă le luăm pe fiecare în parte) pot interacționa și se pot completa reciproc. Recunoașterea ușoară de către un om a cifrelor și literelor scrise de mână rezultă dintr-o combinație complexă de experiență, procesare globală, caracteristici distinctive și integrarea contextului și semanticii. Abordarea filosofică transdisciplinară oferă posibilitatea ca aceste explicații, fiind fiecare legată de un anumit nivel de Realitate și, corespunzător, de un anumit nivel de percepție, să fie înglobate într-un model explicativ unic.

Jocul de șah

Jocul de șah a fost, încă de la apariția conceptului de inteligență computațională, una dintre provocările constante ale cercetătorilor.

Un exemplu preluat din lumea șahului care, deși pare diferit este, în esență, de aceeași natură cu celelalte două studii de caz analizate anterior, îl constituie acele poziții de șah banale chiar și pentru jucători mediocri, dar de nerezolvat pentru actualele programe de jucat șah (programe capabile, de altfel, să învingă chiar și campionul mondial!).

Diagrama din ***Fig. 2. 10*** este preluată din lucrarea lui John Barrow, *„Despre imposibilitate. Limitele științei și știința limitelor”* [9, pp. 98-99].

Atât John Barrow, cât și Roger Penrose, încearcă să răspundă la întrebarea dacă este sau nu plauzibil că în procesul nostru de gândire există ceva care este diferit de un proces de calcul. În mod surprinzător, se constată că poziția din ***Fig. 2. 10*** este rezolvată greșit de cele mai performante programe de jucat șah, inclusiv programul de șah care este declarat campion mondial în acest moment - Stockfish, care a atins un coeficient Elo de peste 3300 (pentru comparație, Magnus Carlsen, cel mai bun jucător din lume, are, în acest moment, 2835).

În ***Fig. 2. 11*** se observă că Stockfish, în ciuda puterii imense de calcul, mută… greșit! Este evident, chiar și pentru un jucător slab de șah că albul trebuie să mute liniștit regele și să-și păstreze

lanțul de pioni nevătămat. În pofida acestui fapt, Stock alege să ia turnul și, implicit, să piardă partida.

Evident, s-a pus problema introducerii acestei poziții în baza de date, astfel încât Stockfish să o trateze ca pe o anomalie și să îi memoreze în mod particular soluția. Surpriza este că un om poate imediat schimba puțin poziția, astfel încât să introducă o altă anomalie pe care puternica mașină de jucat șah să nu o poată rezolva corect. Acest lucru este demonstrat în ***Fig. 2. 11*** și ***Fig. 2. 12***.

Este interesant de observa că programele de jucat șah din 2023 par să nu fi depășit dilemele cu care a fost confruntat celebrul *Deep Thought* (învingătorul lui Kasparov), încă din 1997.

Ceea ce face Stockfish este complet stupid. Tot ce are de făcut albul este să rămână în spatele barierei sale de pioni albi, pentru că, în acest caz, nu există nicio posibilitate ca să ajungă în poziție pierdută. Și totuși, în ambele tipuri de poziție, Stockfish, ca și „strămoșul" său *Deep Thought*, înhață imediat turnul negru și deschide astfel bariera de pioni, dând astfel posibilitatea negrului ca să câștige. Programele sunt programate să caute, mutare cu mutare, cea mai bună poziție (acea poziție care produce un scor maxim). Pentru computere, șahul este un joc de calcul pur. Jucătorul uman, pur și simplu, vede bariera de pioni și ÎNȚELEGE că este impenetrabilă. Computerul nu posedă acest tip de ÎNȚELEGERE a poziției - el nu face altceva decât să calculeze mutare cu mutare. Acest exemplu este o ilustrare a diferenței dintre calculul pur și superioritatea înțelegerii.

Imediat ce computerul a fost învățat să recunoască o barieră de pioni, el poate rezolva prima problemă. Dar, constatăm că dă greș la al doilea tip de problemă, deoarece este necesar un nivel suplimentar de înțelegere. Există un număr imens de astfel de poziții-problemă. Teoretic, s-ar putea concepe un superprogram care să cuprindă toate aceste probleme, dar practic, proiectul este nefezabil. Chiar și în această variantă nefezabilă, există în mod evident sentimentul că, pe lângă ceea ce numim noi înțelegere, există și altceva decât calcul direct. Așa cum arată Roger Penrose, primele 200 de pagini din cartea sa *Incertitudinile rațiunii. Umbrele minții* [94] au fost dedicate demonstrării acestei afirmații.

Practic, între exemplele preluate din experimentul cu ChatGPT, recunoaștere de cifre scrise de mână cu o rețea neuronală convolutivă și aceste poziții date spre analiză programelor de jucat șah nu există nicio diferență esențială. Exemple alese dovedesc că gândirea umană tratează diferit și folosește alte mecanisme în analiza *pattern*-urilor. Fisurile în raționamentele logico-matematice, pe care mintea umană le depășește fără probleme, dovedesc că ceea ce definește fundamental inteligența este foarte diferită de oricare dintre abordările reducționiste.

A recunoaște un context lingvistic care este inevitabil legat de o stare de fapt sau posibilă, a recunoaște variații complexe în scrierea cifrelor de mână și a realiza rolul unei structuri de pioni (și aceasta fără a avea în spate în proces de calcul sofisticat) sunt trăsături esențiale ale inteligenței umane. În fiecare dintre aceste cazuri, cunoașterea contextului, „întruparea" subiectului în firul acțiunilor și în recepționarea și interpretare datelor constituie răspunsul pentru care inteligența artificială clachează acolo un om nu întâmpină nici un fel de probleme. Dacă nu vom înțelege limitele fundamentale ale inteligenței artificiale, nu vom ști să o folosim corect. Așa cum vom vedea mai departe, viziunea filosofică transdisciplinară vine să umple un gol în această înțelegere a limitelor. Conștientizarea limitelor inteligenței artificiale face posibilă eficientizarea folosirii acestor noi tehnologii. Scopul nostru nu este să promovăm un scepticism steril, ci, dimpotrivă, prin înțelegerea a ceea ce este inteligența artificială, în esența ei, să o folosim cu un optimism ponderat și, prin aceasta, cu adevărat fertil.

Înainte de a încerca o explicație folosind modelul filosofic transdisciplinar, vom analiza această situație din punctul de vedere al psihologiei gestaltiste, care aduce câteva contribuții semnificative la înțelegerea acestei abilități specific umane.

Teoria gestaltistă explorează, printre altele, modul în care creierul uman percepe și organizează informațiile vizuale. Conform teoriei gestaltiste, percepția umană se bazează pe organizarea globală a informațiilor vizuale, în loc de a realiza doar o simplă sumă a părților componente. Această abordare subliniază că, mai curând decât să se concentreze pe trăsăturile individuale, mintea umană percepe întregul și caută modele și structuri în informații. Drept urmare, spun adepții teoriei gestaltiste, când recunoaștem scrisul de mână, ne bazăm pe principii gestaltiste pentru a identifica și a înțelege modelele și relațiile între diferitele elemente [141].

Oamenii pot recunoaște modele și grupuri de caractere scrise de mână în funcție de proximitatea și similaritatea lor. Deci, într-un grup de cifre, putem identifica grupurile mai mici sau perechile de cifre similare pe baza distanței și asemănărilor vizuale. Această abordare bazată pe principii gestaltiste facilitează identificarea și organizarea informațiilor într-un mod mai eficient decât algoritmul de recunoaștere rigid al unei rețele neuronale.

Oamenii tind să completeze sau să închidă informațiile lipsă într-o imagine pentru a forma un model coerent. De exemplu, chiar dacă o anumită parte a unui caracter scris de mână este parțial ascunsă sau incompletă, putem completa mental și recunoaște forma completă a caracterului. Această abilitate de a deduce și a completa informațiile lipsă se bazează pe principiile gestaltiste ale continuității și închiderii și poate facilita recunoașterea eficientă a scrisului de mână.

În plus, oamenii folosesc adesea cunoștințele anterioare și contextul pentru a interpreta și a recunoaște caracterele scrise de mână. Într-o secvență de cifre care reprezintă un număr de

telefon, putem deduce și anticipa cifrele următoare în funcție de reguli și cunoștințe. Această capacitate de a integra contextul și cunoștințele anterioare poate facilita recunoașterea precisă a scrisului de mână, în timp ce inteligența artificială poate avea dificultăți în recunoașterea și interpretarea semnificativă a contextului și a semnificației.

În felul acesta, teoria gestaltistă ne oferă o explicație pentru abilitățile umane superioare în recunoașterea scrisului de mână. Capacitatea noastră de a organiza global informațiile vizuale, de a recunoaște modele și de a completa informațiile lipsă ne permite să recunoaștem caracterele scrise de mână cu precizie și eficiență, chiar și în condiții variate și ambigue.

În comparație cu teoria gestaltistă, inteligența artificială, inclusiv rețelele neuronale, utilizează abordări mai rigide. Deși rețelele neuronale pot obține performanțe remarcabile, așa cum am și constatat în acest studiu de caz, acestea pot întâmpina dificultăți în recunoașterea caracterelor scrise de mână în situații atipice sau cu variații semnificative. Capacitatea noastră umană de a aduce cunoștințe anterioare, de a integra contextul și de a utiliza principiile gestaltiste pentru a recunoaște și a înțelege modelele vizuale ne conferă, cel puțin în acest moment, un avantaj incontestabil în acest tip de activitate.

2.4 Inteligența simulată vs. inteligența autentică: O perspectivă non-reducționistă

Înainte de a trece nemijlocit la aplicarea metodologiei transdisciplinare în evaluarea și înțelegerea inteligenței artificiale vom analiza pe scurt conceptul și tipurile de reducționism. Acest lucru este important deoarece, metaforic vorbind, reducționismul este „inamicul" abordării holistice a metodologiei transdisciplinare.

Matematicianul englez John C. Lennox abordează trei tipuri de reducționism [70, p. 67-74]: reducționismul metodologic, reducționismul epistemologic și reducționismul ontologic. Toate aceste trei tipuri constituie abordări distincte în filosofie și știință, cu diferențe semnificative în ceea ce privește domeniul de aplicare și perspectivele pe care le oferă asupra fenomenelor complexe. Cu toate că în mod firesc sunt pe larg aplicate în știință, atunci când sunt utilizate incorect pot genera cunoștințe neadevărate despre realitatea înconjurătoare.

Reducționismul metodologic este o abordare filosofică și științifică care susține că fenomenele complexe pot fi înțelese și explicate prin descompunerea lor în componente mai simple, fundamentale. Această abordare se bazează pe ideea că înțelegerea detaliată a părților componente ale unui sistem complex poate duce la înțelegerea globală a întregului sistem.

Reducționismul metodologic presupune că toate fenomenele complexe pot fi reduse la componente mai simple și că principiile fundamentale care guvernează aceste componente pot fi utilizate pentru a explica întregul sistem. Prin analiza și înțelegerea componentelor individuale, reducționiștii metodologici susțin că putem dezvolta o înțelegere mai profundă a sistemului în ansamblul său.

De exemplu, în domeniul științei biologice, reducționiștii metodologici susțin că proprietățile și funcțiile unui organism viu pot fi reduse la nivelul molecular, celular și genetic. Aceștia argumentează că, prin înțelegerea detaliată a structurii și funcției componentelor moleculare și celulare, putem înțelege modul în care un organism viu funcționează în ansamblul său [134].

Totodată, este important să menționăm că reducționismul metodologic nu susține că aceasta este singura modalitate de a aborda fenomenele complexe sau că explicațiile reduse la niveluri mai mici sunt suficiente pentru a cuprinde întreaga complexitate a unui sistem. Există și alte abordări, precum holismul, care subliniază importanța înțelegerii sistemului ca o entitate globală și a interacțiunilor sale complexe. Prin urmare, reducționismul metodologic este doar una dintre multiplele perspective asupra înțelegerii și explicării fenomenelor complexe.

Reducționismul epistemologic este o abordare filosofică care afirmă că cunoașterea și explicarea fenomenelor complexe pot fi reduse la cunoașterea și explicarea fenomenelor mai

simple. Acest tip de reducționism se concentrează pe procesul de reducere a unui domeniu complex de cunoaștere la un set mai restrâns de principii și legi fundamentale.

Reducționiștii epistemologici susțin că cunoașterea și explicarea nivelurilor superioare ale realității, cum ar fi fenomenele biologice, psihologice sau sociale, pot fi înțelese în termenii legilor și principiilor nivelurilor inferioare ale realității, cum ar fi fizica sau chimia. Astfel, ei consideră că nivelurile superioare pot fi reduse la niveluri inferioare în ceea ce privește explicația și fundamentarea lor [157].

Un exemplu în acest sens ar fi considerarea fenomenelor mentale și psihologice ca fiind reduse la procesele neurologice și la activitatea cerebrală. Reducționiștii epistemologici ar susține că înțelegerea funcționării creierului și a sistemului nervos poate oferi o bază pentru înțelegerea comportamentului și experienței umane.

Reducționismul epistemologic nu exclude existența unor proprietăți și legi emergente la niveluri superioare. Aceasta înseamnă că înțelegerea unui fenomen complex nu se rezumă doar la înțelegerea componentelor sale mai simple, ci poate necesita, de asemenea, o abordare care să țină cont de interacțiunile și proprietățile emergente ale acestora.

Reducționismul ontologic este, la fel ca și reducționismul metodologic și epistemologic, o abordare filosofică care susține că entitățile și fenomenele complexe pot fi reduse la entități și fenomene mai simple și fundamentale. Spre deosebire de celelalte abordări, reducționismul ontologic se concentrează pe structura ontologică a realității și argumentează că tot ceea ce există poate fi în cele din urmă redus la entități de bază.

Reducționiștii ontologici susțin că entitățile complexe, cum ar fi obiectele fizice, organismele vii sau sistemele sociale, pot fi reduse la entități mai simple. De exemplu, în domeniul fizicii, reducționiștii ontologici ar argumenta că toate proprietățile și comportamentele obiectelor fizice pot fi reduse la nivelul particulelor subatomice și a interacțiunilor lor [149].

În esență, reducționismul ontologic susține că există o structură ierarhică în realitate, în care nivelurile superioare pot fi explicate și înțelese prin intermediul nivelurilor inferioare. Prin urmare, întregul univers și toate fenomenele și entitățile sale pot fi reduse la elementele lor de bază.

Reducționismul ontologic este o perspectivă contestată și există critici și alternative la această abordare. Unii contestă ideea că totul poate fi redus la entități și fenomene mai simple, argumentând că există proprietăți și legi emergente la niveluri superioare, care nu pot fi reduse la niveluri inferioare (teoria complexității ireductibile). Perspectiva holistică, contrar reducționismului ontologic, subliniază interconectivitatea, interdependența entităților și a fenomenelor într-un sistem complex.

Recapitulând, reducționismul metodologic se referă la descompunerea sistemelor complexe în părți componente, reducționismul epistemologic la reducerea cunoașterii fenomenelor complexe la nivelurile inferioare iar reducționismul ontologic la redimensionarea entităților și fenomenelor complexe la niveluri mai simple și fundamentale. Aceste abordări diferite oferă perspective distincte asupra înțelegerii și explicării fenomenelor complexe și sunt fi utilizate în diferite contexte filozofice și științifice.

Să ne gândim la *laptop*-ul care în acest moment este folosit pentru a scrie aceste rânduri. Constatăm faptul că acest dispozitiv are minim două componente de bază: *hard*-ul și *soft*-ul. La rândul său, *hard*-ul este compus din felurite metale, plastic, sticlă, care în prealabil, au fost extrase, fabricate, prelucrate, pentru ca, fiind ulterior proiectate după un anumit design, să fie asamblate în componente de bază. La fel, *soft*-ul este elaborat de proiectanți care au avut grijă ca acesta să poată fi integrat în *hard*. La fiecare nivel avem anumite reguli, principii și legi după care nivelul respectiv este funcțional. Fizica și chimia sunt necesare pentru a înțelege proprietățile materiei prime din care este realizat laptopul. Ingineria, designul și tehnologia sunt responsabile de organizarea și modelarea materiei prime. La rândul lor, programatorii, ținând cont de *hardware*, scriu softurile necesare. Sigur, înțelegerea că fiecare nivel superior este dependent de nivelul inferior, dar legile de la nivelul superior nu derivă automat din legile unui nivel inferior, este un aspect important în dezbaterea despre reducționism. Chiar dacă nivelurile inferioare pot influența și stabili constrângerile pentru nivelurile superioare, nu înseamnă că legile și proprietățile de la nivelul superior pot fi reduse și explicate în întregime prin intermediul legilor și proprietăților nivelului inferior.

În cazul laptopului, de exemplu, fizica și chimia pot furniza explicații cu privire la materialele utilizate, procesele de producție și funcționarea componentelor sale. Aceste niveluri inferioare pot pune constrângeri asupra performanței și capacităților laptopului. Cu toate acestea, la întrebările legate de motivele pentru care a fost creat laptopul, de forma și funcționalitatea sa specifică, nu se poate răspunde în întregime cu ajutorul legilor și principiilor fizice sau chimice. Aceste întrebări țin de niveluri superioare, cum ar fi designul, ingineria și intențiile umane. Ele implică aspecte cum ar fi cerințele utilizatorilor, obiectivele de proiectare și contextul social și cultural. Prin urmare, pentru a înțelege pe deplin laptopul și scopurile sale, este necesară o abordare care să țină cont de multiplele niveluri și interacțiuni între acestea. O explicație mai completă poate fi dată, folosind metodologia filosofică transdisciplinară, având în vedere legile și principiile celorlalte niveluri de realitate [92].

Așa cum afirmă Lennox [70, p.71], anumiți cercetători cu o viziune materialistă susțin că există o strictă dependență a nivelurilor superioare față de nivelurile inferioare și utilizează

conceptul de emergență pentru a sublinia acest aspect. Cercetătorii respectivi consideră că anumite proprietăți și fenomene de la niveluri superioare apar automat din nivelurile inferioare, fără necesitatea adăugării de informații noi sau de organizare suplimentară. Un exemplu adesea menționat este emergența proprietăților apei la nivel superior (proprietatea de a fi umed, de a fi solvent pentru multe substanțe, proprietățile electrice și termodinamice etc.) ca rezultat din combinarea hidrogenului și oxigenului. Chiar și așa, extrapolarea acestui exemplu de proprietăți emergente ale apei la conceptul de inteligență poate fi problematică și discutabilă. În continuare, vom explora argumentele care susțin că definirea inteligenței ca o proprietate emergentă a creierului este contestabilă sau chiar incorectă.

Thomas Nagel afirmă, într-o lucrare deja celebră: *„orice program reducționist trebuie să se bazeze pe analiza a ceea ce trebuie redus. Dacă analiza va lăsa ceva pe dinafară, atunci problema va fi pusă greșit. E inutil să întemeiezi apărarea materialismului pe vreo analiză a fenomenelor mentale care nu reușește să trateze explicit caracterul lor subiectiv, pentru că nu există nici un motiv să presupui că o reducție care pare plauzibilă atunci când nu se face nici o încercare de a analiza conștiința poate fi extinsă pentru a o include și pe aceasta din urmă. (...) Nu poți exclude dintr-o reducție trăsăturile fenomenale ale experienței, așa cum excluzi trăsăturile fenomenale ale unei substanțe obișnuite dintr-o reducție chimică sau fizică."* [90, p. 254].

Acest tip de excludere este practicat, cu sau fără intenție, de toți promotorii diferitelor interpretări asupra inteligenței artificiale. Metodologia transdisciplinară reușește să evite tocmai acest neajuns, prin includerea într-un model unic a trăsăturilor fenomenale ale experienței și a caracterului subiectiv al acestei experiențe.

Metodologia transdisciplinară reușește să depășească diferitele tipuri de reducționism prin aplicarea consecventă a axiomelor transdisciplinare și prin modelul Realității în care Subiectul și Obiectul sunt interconectate printr-un punct comun în zona de non-rezistență (X, în ***Fig. 1. 3***)

2.4.1 Non-reducționismul și nivelurile de realitate

În continuare, ne vom concentra asupra conceptului de non-reducționism și impactului acestuia asupra unui alt concept important, cel de realitate. Non-reducționismul poate fi exprimat prin alte concepte-cheie, cum ar fi holismul și emergentismul.

Holismul este o perspectivă filozofică și teoretică care susține că întregul este mai mult decât suma părților sale individuale. Acest concept afirmă că nu putem înțelege pe deplin sau explica fenomenele sau sistemele complexe prin simpla analiză sau reducerea la componentele lor individuale. Holismul afirmă că există o interdependență și o interconectare profundă între părțile

componente ale unui sistem, iar înțelegerea acestuia necesită luarea în considerare a acestor relații și interacțiuni. Așadar, holismul consideră că întregul are propriile sale proprietăți și comportamente emergente care nu pot fi reduse sau explicate în întregime prin analiza părților componente izolate. Este o abordare care evidențiază importanța contextului și a relațiilor în înțelegerea și explicarea fenomenelor complexe.

Ideea că întregul este mai mult decât suma părților sale, este cunoscută și sub numele de „proprietate emergentă" (dar nu în același sens în care acest termen este folosit de reducționiști). Aceasta înseamnă că, pe lângă proprietățile și comportamentele părților individuale, sistemul în ansamblul său prezintă caracteristici și comportamente noi, distincte și imprevizibile.

Dar, așa cum am menționat anterior, reducționismul își are locul său important în cercetarea științifică, ca manieră necesară, până la un punct, de simplificare a modelelor, asumând riscul pierderii unor proprietăți (de obicei, considerate neesențiale). Modelul transdisciplinar al nivelurilor de realitate este capabil să reconcilieze reducționismul sistemelor simple cu non-reducționismul sistemelor complexe. Se știe că termenii „realitate" și „nivel" sunt utilizați într-un mod neriguros și ambiguu în literatura filosofică și științifică. De aceea, transdisciplinaritatea propune o abordare clară și precisă a acestor termeni. Mai mult, este necesar să subliniem că Realitatea, în sens transdisciplinar, nu se limitează doar la o construcție socială sau la consensul unei colectivități. Ea are și o dimensiune trans-subiectivă și ontologică. Realitatea este ceva mai mult decât o simplă convenție umană; ea are o existență independentă și poate influența lumea dincolo de percepțiile și interpretările noastre individuale [164; 86].

Filosoful american Roy Clouser, în articolul intitulat „*A Blueprint for a Non-Reductionist Theory of Reality*" (*Un model pentru o teorie non-reductionistă a realității*) [132, pp. 27-58], încearcă să aducă un argument îndreptat, în principal, împotriva reducționismului fizicalist. Autorul susține că *„nu poate exista o justificare discursivă pentru credința că orice ar putea fi explicat exclusiv pe baza unui atribut care să fie independent de toate celelalte atribute."* [132, p.30]. Argumentul pe care urmează să-l prezinte autorul nu este nici deductiv, nici inductiv. Nu se bazează pe premise privilegiate sau pe probabilități, ci este un experiment de gândire simplu pe care oricine îl poate realiza. Experimentul constă în a încerca să concepem orice lucru ca având o singură natură exclusivă. Luând exemplul materialismului, ne vom gândi la un obiect concret, cum ar fi o carte, ca fiind exclusiv fizic. Vom elimina toate proprietățile non-fizice, cum ar fi cele cantitative, spațiale, senzoriale, logice și lingvistice, până când nu mai rămâne nimic din conceptul de carte. Acest experiment poate fi repetat și cu alte concepte, arătând că nu avem o idee clară despre ce înseamnă un lucru exclusiv fizic. Experimentul nu reușește doar atunci când îl încercăm cu un obiect concret, cum ar fi cartea, dar și atunci când îl aplicăm la proprietățile fizice abstracte,

cum ar fi greutatea. Eliminăm toate aspectele măsurabile, spațiale, senzoriale, logice și lingvistice ale greutății, iar rezultatul este că nu mai rămâne nimic din conceptul de greutate.

Concluzia pe care o putem trage din argumentul prezentat este că încercarea de a concepe un obiect sau o proprietate ca fiind exclusiv fizică, eliminând orice altă caracteristică sau conexiune, duce la o lipsă de substanță și înțelegere. Experimentul mental arată că nu avem o idee clară despre ce înseamnă cu adevărat „exclusiv fizic". Asemenea tentativelor de a concepe un cerc pătrat, aceasta este o afirmație care nu are sens. Astfel, argumentul susține că materialismul și ideea de a reduce totul la aspecte fizice exclusive nu sunt fundamentate și nu pot fi susținute rațional.

Mai mult ca atât, experimentul de gândire se aplică nu doar la materialism, ci și la alte „ism"-uri care fac pretenții exclusiviste similare în numele altor tipuri de proprietăți și legi. De exemplu, nu putem concepe senzații pur senzoriale sau concepte pur logice care să nu implice niciun alt aspect non-senzorial sau non-logic. Acest lucru subminează ideea că există entități sau proprietăți pure în sine.

De asemenea, experimentul de gândire are implicații asupra legilor cauzale postulate pentru a explica cum un anumit tip de entitate produce tot ceea ce nu este din același tip. Nu putem concepe legi pure ale unui singur tip, la fel cum nu putem concepe obiecte pure ale acelui tip. Aceasta ridică întrebări despre modul în care entitățile interacționează între ele și subminează înțelegerea noastră despre cauzalitate și metafizică.

În consecință, concluzia experimentului de gândire descris de Clouser este că ideea de a încadra ceva ca fiind exclusiv de un singur tip de natură, fie că vorbim despre obiecte, proprietăți sau legi, nu are sens. Aceasta subminează pretențiile exclusiviste ale diferitelor teorii și perspective filosofice și ne invită să reconsiderăm modul în care înțelegem realitatea. Această concluzie are implicații în diverse domenii, cum ar fi filosofia minții, metafizica și științele cognitive. Aceasta ne provoacă să abandonăm ideea de entități pure și să explorăm interconexiunile și interacțiunile complexe dintre diferitele aspecte ale realității.

În același text, Clouser folosește conceptul de „aspect" pentru a se referi la un tip de bază al proprietăților și legilor. În viziunea autorului *„un aspect este considerat de bază atunci când nu poate fi subsumat sub niciun alt aspect fără a genera antinomii, contradicții sau alte incoerențe grave"* [132, p.39]. A considera astfel de proprietăți și legi ca fiind un aspect al realității, înseamnă a afirma că ele sunt valabile atât pentru experiențele noastre subiective, cât și pentru obiectele acestora. Aceste tipuri de aspecte sunt abstractizate din experiența noastră pre-teoretică a lucrurilor, evenimentelor, relațiilor, stărilor de lucruri, persoanelor etc. Obiectele experienței obișnuite prezintă atât proprietăți, cât și conformitate cu legi care, la rândul lor, prezintă meta-

proprietăți (cantitative, spațiale, fizice, logice etc.) care caracterizează fiecare aspect. Existența unor legi și aspecte de bază a dus la folosirea unei liste aproximativ identice a aspectelor utilizate de mulți gânditori în istoria filosofiei și științelor. Nu există un acord universal cu privire la lista exactă de aspecte autentice, dar, chiar dacă ar exista, acest lucru nu ar rezolva problema. Aceasta este o problemă importantă pentru orice ontologie. Totodată, nu ne propunem să stabilim aici lista exactă corectă de aspecte. Argumentul non-reducționist și conturul general al ontologiei non-reducționiste nu depind de stabilirea prealabilă a unei anumite liste de aspecte (proprietăți, legi) ca fiind cea corectă.

Pentru a completa argumentul său autorul introduce conceptul de proprietate activă și pasivă [132, p. 43]. Aceste proprietăți reprezintă felurile în care un lucru există și funcționează sub influența legilor unui anumit aspect. Deși argumentul susține că toate lucrurile funcționează pasiv în toate aspectele, apariția funcțiilor active este determinată de ordinea secvențială a aspectelor. Un exemplu oferit de autor ne poate ajuta să înțelegem mai bine legătura dintre feluritele aspecte și modul în care proprietățile active și pasive se manifestă. O stâncă funcționează activ în aspectele cantitative, spațiale, cinematice și fizice, dar nu și în celelalte aspecte. Chiar și în aceste condiții, există o realitate în care stânca deține proprietăți în acele aspecte, deoarece există legi care o guvernează. Totuși, aceste proprietăți depind de interacțiunea pietrei cu alte lucruri care funcționează activ în acele aspecte, ceea ce vom numi proprietăți pasive ale stâncii. De exemplu, o stâncă poate fi un obiect pe care un pescăruș sparge scoici pentru a le mânca sau poate fi peretele bârlogului unui animal. Prin urmare, o stâncă poate fi pasiv acționată de ființele vii. Aceste proprietăți pasive sunt doar potențiale până când sunt actualizate prin contactul cu ceva activ în mod biotic (un agent biologic). Cu toate acestea, ele rămân proprietăți reale, posibile datorită faptului că stânca este guvernată de legile biologiei și de celelalte legi. Deși o stâncă nu funcționează activ sub aspectul senzorial, faptul că este supusă legilor senzoriale și posedă proprietăți senzoriale pasive permite ființelor perceptive să o perceapă. De asemenea, aspectul logic conferă stâncii proprietăți pasive, deși nu gândește în sine. În același mod, stânca poate avea proprietăți pasive și în aspecte lingvistice, sociale, economice și altele. În contrast, o plantă funcționează activ într-un aspect suplimentar față de o stâncă, și anume în aspectul biotic, deoarece este vie și își desfășoară activ funcții metabolice, având în același timp doar funcții pasive în celelalte aspecte. De asemenea, un animal prezintă o funcție activă suplimentară față de o plantă, fiind activ în aspectul senzorial, permițându-i să perceapă și să simtă într-un mod în care plantele nu pot. Se consideră că doar oamenii au o funcție activă în fiecare aspect cunoscut până acum. Acest lucru este ilustrat de autor într-o schemă simplificată [132, p. 45] ***Fig. 2. 14***.

Distincția între funcțiile active și pasive într-un obiect se referă la modurile în care obiectul are proprietăți și există sub guvernarea legilor diferitelor aspecte. Aceste aspecte sunt aranjate într-o ordine secvențială într-o listă, iar obiectul trebuie să aibă proprietăți active în aspectele inferioare pentru a avea și proprietăți active în aspectele superioare. Această ordine nu este cauzală sau ierarhică în sensul că aspectele de rang superior nu sunt mai importante decât cele de rang inferior. Este doar o observație că pentru a avea funcții active în aspectele superioare, obiectul trebuie să aibă funcții active în aspectele inferioare.

Distincția pasiv-activ are implicații asupra teoriilor privind percepția senzorială. Aceasta ne permite să evităm postulatul că ceea ce experimentăm sunt doar date senzoriale interne și ne permite să ne însușim elemente de adevăr atât din obiectivism, cât și din subiectivism. De exemplu, obiectele au proprietăți potențiale pasive în percepție, cum ar fi culoarea sau forma, care devin actualizate atunci când sunt percepute. Aceasta evită presupunerea că suntem izolați de lumea înconjurătoare sau că calitățile senzoriale sunt create în totalitate de noi sau există doar în mintea noastră.

Distincția activ-pasiv ne ajută să evităm falsa dilemă dintre obiectivism și subiectivism și ne permite să recunoaștem că obiectele au dispoziții pasive care fac posibile experiențele noastre în diverse domenii, cum ar fi frumusețea sau valoarea economică. De exemplu, obiectele trebuie să fie supuse legilor economice pentru a avea valoare economică, chiar dacă această valoare nu este deja reală în obiect independent de actualizarea ei.

Distincția dintre funcțiile active și pasive în aspecte arată că nu este plauzibil să presupunem că aspectele apar în existență în întregime, deoarece funcțiile active apar în relație cu proprietățile pasive și legile aspectului respectiv. De exemplu, cosmosul nu putea avea doar proprietăți fizice inițiale, iar apoi proprietăți logice să apară mai târziu, deoarece acest lucru ar fi fost ilogic. Distincția pasiv-activ ne arată că aspectele sunt interconectate și că toate aspectele sunt la fel de reale. Fiecare lucru posedă proprietăți în fiecare aspect, iar negarea unei realități exclusive într-un anumit aspect nu înseamnă că există lucruri complet non-fizice sau non-biotice. Argumentul subliniază că toate lucrurile au anumite proprietăți în fiecare aspect al existenței.

Am insistat mai mult asupra lucrării lui Clouser pentru că aceasta se constituie ca un argument foarte puternic împotriva oricărei forme de reducționism și are multe idei comune cu metodologia filosofică transdisciplinară, care operează cu multiple niveluri de Realitate.

2.4.2 Interpretarea non-reducționistă a inteligenței artificiale

Toate modele de interpretare ale inteligenței artificiale au la bază reducționisme de un anumit tip, cel mai adesea fizicaliste. Unul dintre punctele slabe ale acestor modele derivă din faptul că toate abordările interpretează unilateral ceea ce se consideră a fi „inteligență". Conștiința este total scoasă din această ecuație, deși toate sistemele biologice pe care inteligența artificială își propune să le imite posedă un anumit nivel de conștiință, definit ca un mecanism prin care un sistem viu se întoarce asupra sa, se poate autoanaliza și corecta, toate acestea fiind făcute astfel încât agentul inteligent știe despre sine că există și se percepe pe sine ca o entitate distinctă de celelalte. Se poate obiecta că există sisteme vii inteligente, despre care nu știm (mai exact, nu avem dovezi) că posedă o formă de conștiință. Studii recente de știința comportamentului [119] arată însă că nu există sisteme vii cu o inteligență avansată care să nu aibă și o anumită formă de conștiință, chiar și la un nivel rudimentar. Este clar, în acest moment, că cimpanzeii, elefanții, câinii, pisicile, delfinii și alte specii posedă un anumit nivel de conștiință (în cazul maimuțelor, se pare că putem vorbi chiar de conștiință morală) [120].

Această excludere brutală a acțiunii conștiente din câmpul de interacțiune cu activitatea inteligentă face ca toate modele de inteligență artificială să piardă din vedere faptul că, dincolo, de îndeplinirea unei activități după un anumit algoritm sau tehnică euristică, a fi inteligent implică și a poseda facultatea de a fi conștient că exiști. Doar astfel activitățile pot fi considerate inteligente, pentru că doar o stare de conștiență poate atașa un sens, o înțelegere activă, unei activități inteligente. Nu se poate vorbi de act inteligent fără capacitatea unui sistem inteligent de a evalua plenar eficiența și rezultatele acestei activități. Aceasta implică faptul că activitățile considerate inteligente trebuie să fie cuprinse într-un cadru mult mai larg, care să analizeze, să clasifice, să corecteze și să interpreteze rezultatele acestor activități. Este exact ceea ce face conștiința. Altfel, putem vorbi despre sisteme de inteligentă artificială doar în maniera în care un personaj al lui Moliere era poet fără să știe că este poet. Dacă inteligența artificială exclude conștiința ca fiind ceva de altă natură, atunci de ce nu am putea vorbi de sisteme inteligente și atunci când este vorba de mecanisme mai simple, dar care copiază și îndeplinesc foarte bine anumite sarcini, până în acel moment considerate ca fiind pur umane?! De ce un sistem de acționare cu aburi, care deschide ușa unui templu antic la apropierea preotului, nu ar putea fi considerat „inteligent", dar un sistem modern de comandă cu recunoașterea amprentei vocale este văzut ca „inteligent" în sens uman? De ce recunoașterea vorbirii este considerată o parte a inteligenței umane, și, în consecință sarcină a inteligenței artificiale, iar gesturile prin care un om deschide o ușă, nu?! Ce le diferențiază în mod esențial? Doar faptul că recunoașterea vorbirii este mai dificil de realizat și a apărut, istoric

vorbind, mult mai târziu decât mecanismele de deschidere ale unei uși? Ambele activități au în comun faptul că îndeplinirea lor necesită o cuprindere într-un câmp de conștiință, câmp care ghidează activitatea și îi interpretează rezultatele.

Modelul transdisciplinar pe care îl propunem face ca nivelurile de realitate să interacționeze și să nu poată exista unul fără celălalt (axioma epistemologică a transdisciplinarității) [92]. Din perspectiva transdisciplinară, inteligența este posibilă doar pentru că nivelul învățării nu poate exista separat de nivelul înțelegerii, că nivelul înțelegerii nu poate fi redus la nivelul învățării și nici invers, că nu există ierarhie între aceste nivele. Deci, conform schemei propuse de noi (***Fig. 2. 15***), inteligența nu are sens fără înțelegere, care implică cu necesitate conștiința de sine. Procesele de analiză și sinteză, elaborarea deciziilor, care constituie practic efectul activității inteligente, presupun deopotrivă un anumit tip de înțelegere globală și un anumit grad de conștiință.

Un alt punct slab al interpretărilor reducționiste legate de inteligența artificială este cel legat de corporalitate și interacțiunea oricărei activități umane inteligente cu mediul biofizic (nivelul *bios*) și psihic (nivelul *psyche*). Perspectiva filosofică transdisciplinară arată foarte clar că, date fiind interacțiunile complexe din cele trei niveluri (*bios, psyche, nous*), din faptul că fiecare nivel de realitate este dependent de existența tuturor celorlalte (axioma epistemologică) și că nu se poate vorbi de o ierarhie între nivelurile de realitate, înseamnă că nu putem vorbi de inteligență *dianoică* în mod plenar, fără a lua în considerare afectivitatea și corporalitatea.

Decizia unui sistem care posedă inteligență nu este un act pur logic, nu constă doar în decodificarea „rece" a unei informații, ci implică atașarea unui anumit tip de stare afectivă acestei informații. Se replică adesea că tocmai din absența componentei afective va decurge superioritatea inteligenței artificiale, dat fiind că prezența factorului afectiv nu face altceva decât să „contamineze" suportul rațional al actului decizional. Astfel de deducții vin din partea unor analiști care fie nu cunosc, fie au uitat istoria speciei umane. Prezența factorului afectiv nu duce, cu necesitate, la decizii iraționale catastrofale, ci, mult mai des, prezența factorului afectiv îmbunătățește procesul decizional. Un sistem legislativ bazat exclusiv pe raționamente logice nu ar permite, de exemplu, unei mame, închisă pentru un delict oarecare, să-și alăpteze pruncul, nu ar permite nici vizitele familiale. Un sistem educațional bazat exclusiv pe decizii de tip algoritmic ar fi, în ciuda a ceea ce se crede, profund discriminatoriu, pentru că nota nu este strict o evaluare a cunoștințelor într-un domeniu, ci este relaționată și cu un anumit tip de comportament, inclusiv cu situația personală a elevului examinat.

Cifrele și algoritmii nu pot surprinde relaționarea care există în cadrul oricărei specii superior organizată pe scara evolutivă. Se poate spune că un conducător auto poate comite un

accident pentru că este în depresie, dar, același conducător auto evită mult mai des accidente tocmai pentru că atașează deciziilor sale componenta afectivă. Stricta respectare a regulilor de circulație ar putea duce la un număr mai mare de accidente, pentru că un conducător auto obișnuit nu aplică, de exemplu, doar algoritmi de evaluare a distanțelor și vitezelor alături de regulile de circulație, ci ia în considerare tot timpul intenția și starea afectivă probabilă a celorlalți participanți la trafic. Dacă va observa un pieton cu comportament anormal, va acționa ponderând acțiunile sale cu afecte, și nu doar aplicând raționamente logice și interpretând *ad hoc* pasaje din Codul Rutier. Atunci când conducătorul auto va încearcă să intuiască posibila manevră a unui alt participant la trafic, el nu va analiza doar datele de intrare fizice (viteza, traiectoria ș.a.), ci va încerca să ghicească și posibila stare emoțională a celuilalt. Emoțiile ne ajută să luăm decizii și nu sunt, în condiții normale, un factor perturbator în procesul de luare a unei decizii. Emoțiile nu sunt „paraziți" ai inteligenței, ci parteneri.

Marile crime ale umanității nu au fost în primul rând, cum se crede de obicei, produse ale unor afecte degradate, ci consecința unor raționamente despărțite de componenta afectivă. Gândirea „la rece" este cea care a inventat lagărele. Evaluarea strictă după algoritmi de prelucrare a datelor este una dintre rădăcinile totalitarismului. Atât marxismul, cât și nazismul, ca să ne oprim doar la cele două totalitarisme care au ucis, împreună, peste 100 de milioane de oameni, au tratat categorii întregi de oameni ca simple obiecte de inventar, ca date pur statistice. Pentru Hitler și Stalin, oamenii erau simple molecule lipsite de viață interioară.

Faptul că o mașină poate rezolva un test logic sau matematic rapid și fără greșeală, pentru că, spre deosebire de un om, ea nu are emoții, este doar una dintre fețele monedei. A simți ajută la a putea fi empatic, deci la a-l putea înțelege pe cel de lângă tine. Greșelile datorate emoțiilor, experiență prin care am trecut și trecem cu toții, sunt un preț mic și necesar pentru a putea decripta mesajele din jur și realiza, astfel, un tip de înțelegere mult mai plenară a realității. Frica de un pericol ne ajută să supraviețuim într-o măsură mult mai mare decât evaluarea strict rațională a pericolului. Orice expert în evaluarea riscurilor ar trebuie să știe asta și să nu lase ca această evaluare să fie făcută exclusiv de mașini.

Inteligența, oricât de paradoxal ar suna acest lucru, nu înseamnă infailibilitate, ci a ști să profiți de greșeli. Realitatea este atât de complexă, încât singurul mod prin care putem supraviețui ca indivizi și ca specie este abordare de tip *trial and error* (încercare și eroare), iar o astfel de abordare nu poate fi fructificată cu adevărat dacă raționamentele noastre logice nu ar fi mediate de factorul afectiv. În limbajul teoriei rețelelor neuronale artificiale, afectele sunt un factor important în ponderarea sinapselor noastre. Chiar dacă sună straniu, a fi inteligent în sensul deplin al acestui termen, înseamnă să ai opțiunea actului ratat și, în plus, conștiința propriei failibilități.

În continuare, dorim să arătăm, folosind modelul transdisciplinar al ființei umane (***Fig. 2. 15***) creat în baza conceptului transdisciplinar a nivelurilor de realitate, care este limita ontologică a inteligenței artificiale.

Cum am menționat deja, miza acestei lucrări constă în faptul că există limite insurmontabile în încercarea de a simula complet o inteligență de tip uman. Aceste limite nu țin de stadiul actual al științei și tehnologiei, ci sunt limite impuse de legile naturii și de însăși structura Realității. De asemenea, transferul în mașini a unei psihologii de tip uman este limitată de un dat ontologic care definește unicitatea ființei umane.

Adoptând un limbaj transdisciplinar, am putea spune că inteligența artificială, care este obiect transdisciplinar, nu va putea să devină vreodată subiect transdisciplinar, sintagmă care, din această perspectivă, poate fi atribuită doar ființei umane. Aceasta ne readuce pe tărâmul discuțiilor mai vechi despre inteligența artificială „slabă" și despre posibilitatea creării unei inteligențe artificiale „tari". Evident, în acest context, nu putem să nu facem recurs la problema principală care generează discuția în cauză – problema omului.

În celebra sa carte „*Despre sclavia și libertatea omului*" marele filosof rus, Nikolai Alexandrovici Berdiaev scria: „*Omul este o enigmă și poate chiar cea mai mare enigmă a lumii. Omul este o enigmă nu pentru că face parte din lumea animală sau pentru că este o ființă socială, adică în calitate de produs al naturii sau al societății, ci în calitatea lui de persoană și numai de persoană.*" [10, p.23]

Putem fi de acord că în calitatea sa de persoană omul continuă a fi o enigmă. Persoana nu poate fi definită exhaustiv în termeni științifici. Persoana, de fiecare dată, va fi ceva mai mult decât orice definiție pe care o putem găsi în diferitele discipline sau chiar într-un punct de vedere interdisciplinar. Omul ca și persoană necesită a fi privit dintr-o perspectivă transdisciplinară, o perspectivă care vine cu o abordare holistică, nu reducționistă. „*Omul a început să se ia în considerare ca problemă a cunoașterii. Cercetările psihologice, biologice, sociologice nu au elaborat nici o antropologie filosofică. Ele au privit omul sub diverse unghiuri și nu l-au studiat decât în mod parțial fără să reușească vreodată să îi descifreze misterul.*" [11, p. 66] Transdisciplinaritatea încearcă să păstreze misterul din om, să-l readucă la înălțimea de la care a căzut.

Omul este atât subiect cât și obiect transdisciplinar. Mintea ca subiect transdisciplinar este capabilă să studieze și să cunoască mintea ca obiect transdisciplinar; într-un fel putem vorbi despre o minte care se cugetă singură pe sine, o gândire care gândește despre sine. În „*Etica nicomahică*", Aristotel se pare că a intuit același lucru: *„...dacă cel ce vede simte că vede, iar cel ce aude simte că aude, cel ce merge că merge și la fel, în fiecare formă de activitate, există în noi ceva care*

simte că desfășurăm o activitate, care simte deci, dacă simțim, că simțim și dacă gândim că gândim..." [2, p. 202]. În notele explicative ale acestei ediții, aflăm un lucru interesant legat de verbul „a simți". Ni se explică de către traducător că, de fapt, verbul *aisthánesthai*, al cărui sens obișnuit este cel de „a simți" sau „a percepe" prin simțuri, are în acest pasaj semnificația de „a avea conștiință de" (conștiința de sine ce însoțește senzația și gândirea). Așadar, putem afirma împreună cu Aristotel că omul are o conștiință care este conștientă de sine, care, cum vom încerca să argumentăm în continuare, nu este și nici nu va fi accesibilă inteligenței artificiale, oricare ar fi nivelul de dezvoltare al tehnologiilor.

Un alt pasaj, pe care îl găsim la Adrian Miroiu, ne sugerează că omul, pe lângă capacitatea sa de autoreflecție, are o capacitate creatoare care este în stare să creeze alți „creatori". *Mai specific omenesc nu este faptul că omul face unelte, ci acela că face unelte cu care face unelte (deci face mijloace de producție). Analog, raportându-ne la alte cunoscute definiții ale omului, am putea zice: omul nu este un animal care are conștiință, ci unul care are conștiința conștiinței sale (să reamintim aici și susținerea lui Aristotel că orice cunoaștere presupune o iterație); omul nu este un animal care are limbaj, ci unul care are un limbaj capabil de autoreferință)* [88, p. 163].

Capacitatea de a crea îl face pe om să fie asemănător cu Creatorul său, să fie mai mult decât o creație, să fie o creație creatoare. La fel și în cazul inteligenței artificiale, care este o creație a omului, putem evidenția clar capacitatea creatoare a creaturii. Astăzi putem asculta muzică creată de inteligența artificială, putem viziona imagini video și admira pictură, putem citi texte și chiar poezii. Cu siguranță, toate acestea constituie elemente ale creativității, chiar dacă sunt departe de a putea fi comparate cu capodoperele umane. Întrebarea pe care trebuie să ne-o punem este legată de modalitatea în care acestea au fost create. Omul, fiind și subiect și obiect transdisciplinar, are acces simultan la toate nivelurile de realitate disponibile ființei umane, ceea ce îi permite să creeze „*ex nihilo*" din nimic, nu în sensul în care acest termen este utilizat în teologie cu referire ca Creația divină, ci în sensul că agentul uman nu are nevoie de opere artistice anterioare pentru a crea ceva nou. Inteligența artificială funcționează în baza unor algoritmi care iau în calcul ceea ce deja a fost creat. Cineva ar spune că și omul face la fel raportându-se la natură ca sursă a inspirației. Este adevărat, doar că inspirația nu este același lucru cu rularea unui algoritm după legile matematice [29].

Așadar, aplicând modelul transdisciplinar putem spune că inteligența umană și inteligența artificială se află la diferite niveluri de realitate. Conform cu ***Fig. 2. 15***, care ne reprezintă opt niveluri de realitate, limita ontologică a inteligenței artificiale este plasată la nivelul NR_3 pe când inteligența umană, care de fapt nici nu o mai numim inteligență, la acel nivel de realitate, pentru că este ceva mult mai mult, este înțelepciunea care își are originile în Divinitate, se situează la

nivelul NR_5. Nivelul NR_3 corespunde unei inteligențe practice care este caracterizat de procesul de învățare.

Astăzi mașinile sunt în stare să învețe în special folosind algoritmi specializați de tipul *deep learning*. Atunci când dorim să trecem la următorul nivel NR_4, cel al unei inteligențe caracterizată de procesul de înțelegere, ne dăm seama că inteligența artificială nu poate să treacă acest prag. A învăța nu este egal cu a înțelege, așa cum ne sugerează experimentul mintal al camerei chinezești imaginat de Searle [107]. Prin argumentele invocate mai sus dorim să susținem acest punct de vedere pentru a arăta care este limita ontologică a inteligenței artificiale.

Ultimele patru niveluri din ***Fig. 2. 15*** ne descriu o „realitate" care este legată de conceptul de inteligență, care la rândul ei este interconectată cu toate nivelurile de realitate existente. Fiind vorba de conexiuni complexe între diferitele niveluri de realitate, observăm că inteligența, urmând acest model, este un concept complex care nu poate fi definit exhaustiv de un singur nivel de realitate și, în consecință, atunci când atribuim mașinilor acest calificativ, involuntar sau poate chiar intenționat, subevaluăm ființa umană.

2.5 Concluzii

Analizând datele și studiile de caz prezentate în acest capitol, primul lucru evident pe care îl putem spune este că rețelele neuronale artificiale, oricât de performante ar părea, operează pe un principiu radical diferit de cel al minții umane. În timp ce sistemele de tip ChatGPT sau algoritmii de recunoaștere a formelor excelează în manipularea statistică a unor volume uriașe de date, acestora le lipsește complet „înțelegerea" autentică a sensului. Practic, ne aflăm în fața unei simulări extrem de convingătoare, o „cameră chinezească" modernă unde sintaxa și regulile matematice complexe reușesc să imite inteligența fără a poseda însă nicio stare mentală sau conștiință de sine.

O altă observație importantă vizează fragilitatea acestor sisteme în fața contextului situațional. Studiul de caz pe jocul de șah arată că algoritmii clachează exact acolo unde un om, chiar și cu o instruire minimă, rezolvă problema instantaneu prin intuiție și logică de bun-simț. Această diferență vine din faptul că omul nu doar procesează semne, ci extrage sens dintr-un model de realitate mult mai amplu, care include experiența senzorială, corporalitatea și, mai ales, componenta afectivă, care nu este un „parazit" al rațiunii, ci un partener esențial în luarea deciziilor.

În final, perspectiva transdisciplinară ne obligă să recunoaștem o limită ontologică clară a inteligenței artificiale. Mașina este un obiect transdisciplinar care poate învăța, dar nu poate deveni un subiect dotat cu înțelepciune. Reducerea inteligenței la simplul calcul algoritmic este o eroare de interpretare care ignoră unicitatea persoanei umane și capacitatea sa de a crea „din nimic", fără a fi dependentă exclusiv de baze de date anterioare. Prin urmare, inteligența artificială trebuie privită ca un instrument analitic extrem de puternic, dar care rămâne blocat la nivelul procesării de informație, fiind incapabil să facă saltul către nivelurile superioare de realitate unde rezidă înțelegerea și conștiința umană.

3. DE LA TRANSUMANISM LA TRANSDISCIPLINARITATE - IPOSTAZE FILOSOFICE

3.1 Modelul transdisciplinar al realității: Învățare, înțelegere, înțelepciune

„All our knowledge brings us near to our ignorance,
All our ignorance brings us nearer to death,
But nearness to death, no nearer to God.
Where the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?”[7]

T. S. Eliot, *The Rock* [34]

Așa cum am arătat în studiile de caz din capitolul precedent, în special prin eșecul modelului ChatGPT de a decripta polisemia cuvântului „mine” sau dificultatea rețelei convolutive în fața cifrelor cu modele atipice, distincția dintre procesarea datelor și accesul la sens este insurmontabilă pentru algoritm. Nu vom insista aici pe o examinare detaliată a conceptelor de învățare, înțelegere și înțelepciune, subiecte care au fost pe larg abordate de numeroși autori în istoria filosofiei [22; 23; 42; 19; 97; 118; 99; 6; 105; 90; 107; 73; 69; 52].

Pentru lucrarea de față este suficient să pornim la drum cu câteva definiții generale, pentru a le putea aplica la modelul transdisciplinar prin prisma căruia analizăm conceptul de inteligență artificială. Din perspectivă transdisciplinară, aceste versuri nu sunt doar o meditație poietică, ci descriu o „ierarhie” ontologică a nivelurilor de realitate și percepție pe care inteligența artificială pare incapabilă să o parcurgă integral. Interogația poetului americano-britanic T.S. Eliot, ne face să intuim imediat, fără o analiză filosofică riguroasă, diferențele profunde dintre învățare, înțelegere și înțelepciune. Ce vrea să zică poetul prin versurile în care spune că înțelepciunea a fost pierdută în cunoștințe, iar cunoștințele au fost pierdute în informație?

Dacă am încerca să le dăm o explicație din perspectiva omului supertehnologizat al sec. XXI, aceste versuri ne-ar putea sugera că, în epoca noastră, există o tendință de a valoriza mai

[7] Trad. „Toată cunoașterea noastră ne apropie de ignoranță,
Toată ignoranța noastră ne apropie de moarte,
Dar proximitatea morții nu ne apropie de Dumnezeu.
Unde este Viața pe care am pierdut-o trăind?
Unde este înțelepciunea pe care am pierdut-o în cunoaștere?
Unde este cunoașterea pe care am pierdut-o în informație?”

mult informația și cunoașterea superficială, lăsând în uitare ceea ce milenii la rând era considerat a fi înțelepciune. Informația este disponibilă în cantități uriașe și accesibilă cu ușurință prin intermediul tehnologiei, iar cunoștințele pot fi acumulate rapid prin studiu și cercetare. Pe de altă parte, înțelepciunea pare să fie pierdută sau neglijată în acest proces.

Putem extrage din versurile lui Eliot și ideea că astăzi, pe lângă accesul rapid la date, cunoștințele și informațiile pot fi adunate într-un mod fragmentat și lipsit de sens, fără o înțelegere și perspectivă de ansamblu. Cunoștințele acumulate izolat conduc, de cele mai multe ori, la o viziune reducționistă asupra lumii. Există o tendință de a valoriza cantitatea și rapiditatea informației, în detrimentul profunzimii și înțelegerii autentice. De exemplu, modele actuale de procesare computațională a limbajului natural evidențiază în mod pregnant tocmai această situație. Vedem asta la tot pasul în facultăți, unde studenții „scriu" lucrări întregi despre mari filosofi cu un simplu copy-paste din ChatGPT. Referatul arată impecabil, dar e doar o colecție de fraze bine împachetate care nu au trecut prin filtrul minții lor. Momentul adevărului vine la curs. Este de ajuns să le pui o întrebare „dar tu ce crezi despre asta?" sau să le ceri să explice un concept cu propriile cuvinte, și se așterne liniștea. Deși „au produs" un text corect, n-au rămas cu nimic în esență. Au informația la un click distanță, dar au pierdut exercițiul de a înțelege cu adevărat ce spun, transformând învățarea într-o simplă bifare de sarcini.

Versurile lui Eliot ne fac să ne întrebăm dacă suntem dispuși să trecem dincolo de suprafața lucrurilor văzute pentru a căuta o înțelegere mai profundă și, astfel, să ne asigurăm că informația pe care o acumulăm și cunoștințele pe care le dobândim sunt integrate într-o perspectivă care ne modelează viața.

Acum să încercăm să analizăm în lumina acestor versuri performanțele inteligenței artificiale. Putem spune că, în ciuda capacității acestei tehnologii de a procesa și stoca cantități uriașe de informații (învățare), ea este fundamental limitată sau, așa cum argumentăm în această lucrare, îi lipsește complet capacitatea de a înțelege, cu atât mai mult o inteligență artificială nu poate atinge ceea ce oamenii numesc înțelepciunea. Tehnologia inteligenței artificiale este eficientă în a furniza răspunsuri rapide și exacte, nu și de fiecare dată corecte, dar este complet lipsită de perspectiva și discernământul uman necesare pentru a transforma aceste informații în înțelepciune. Într-o lume dominată de fluxul constant de informații generate de inteligența artificială, există riscul ca cunoștințele noastre, așa cum ne spune Eliot, să fie pierdute în marea de date și informații superficiale (vezi exemplul din studiul de caz cu ChatGPT). Da, inteligența artificială poate furniza informații rapide, dar capacitatea noastră de a înțelege și a aplica critic aceste informații poate fi amenințată sau diminuată. Există pericolul ca erorile produse de

inteligența artificială să fie acceptate de către o umanitate superficială drept „adevăruri” și, în acest sens, să pierdem complet legătura cu realitatea.

În general, în ciuda capacităților sale tehnologice avansate, există o nevoie constantă de a aduce înțelepciune și înțelegere umană în utilizarea și dezvoltarea inteligenței artificiale. Este important să ne asigurăm că implementarea tehnologiei inteligenței artificiale este ghidată de valori etice și sociale și că ne depășim tendința de a valoriza doar informația și cunoștințele superficiale, căutând, în schimb, o înțelegere mai profundă și o aplicare înțeleaptă a acestei tehnologii.

Învățarea: De la date la vectori de probabilitate

Ce presupune, în esență, procesul de învățare și în ce mod se diferențiază acesta în cazul oamenilor față de cel al mașinilor? Sau, ținând cont de modelul transdisciplinar al ființei umane (***Fig. 2. 15***), în cadrul căruia inteligența artificială este situată la nivelul de realitate corespunzător proceselor de învățare (NR_3), se poate afirma că acest tip de învățare este identic cu cel uman?

Înainte de a răspunde la aceste întrebări este bine să punctăm câteva perspective filosofice asupra naturii și procesului de învățare. De exemplu, în filosofia antică greacă, Socrate a subliniat importanța învățării prin intermediul dialogului și punerii de întrebări critice pentru a ajunge la adevăr. Platon a dezvoltat ideea de reminiscență, susținând că învățarea constă în rememorarea cunoașterii pe care sufletul o dobândește înainte de naștere [100]. Mai târziu, în filosofia modernă, John Locke a susținut că învățarea are loc prin intermediul experienței senzoriale și al asocierii ideilor. Potrivit lui Locke [73], mintea umană este inițial o *„tabula rasa”* sau o „foaie goală” și că învățarea are loc prin intermediul percepției și reflecției asupra experienței.

La fel, învățarea a fost abordată și din perspectiva behaviorismului [170], o școală de psihologie care susține că comportamentul este determinat în principal de stimuli externi și de răspunsurile la acești stimuli. Potrivit behaviorismului, învățarea se realizează prin asocieri între stimulii externi și recompense sau pedepse.

În epoca contemporană, teoriile cognitive și constructiviste ale învățării au câștigat popularitate. Aceste teorii subliniază rolul activ al individului în învățare, punând accent pe procesele mentale, cum ar fi gândirea, memoria și rezolvarea problemelor. În această perspectivă, învățarea este considerată un proces activ în care indivizii își construiesc cunoștințele și înțelegerea prin explorare, reflecție și construirea de modele mentale [107; 18].

Toate aceste perspective, de la dialogul socratic până la teoriile moderne despre construcția modelelor mentale, scot în evidență faptul că, pentru om, învățarea a fost mereu legată de sens, experiență și participare activă. Diferența majoră în cazul mașinilor este că inteligența artificială învață recunoscând tipare matematice în seturi uriașe de date. Pentru un algoritm, învățarea este o

formă avansată de statistică. El calculează ce cuvânt sau element urmează cel mai probabil, fără a avea însă habar ce înseamnă acel concept în lumea reală. În schimb, pentru oameni, a învăța presupune a trece informația prin filtrul conștiinței, al emoțiilor și al experiențelor trăite. Noi nu doar memorăm date, ci le dăm un sens, fiind capabili să aplicăm ceea ce știm în situații complet noi prin intuiție și raționament, lucru pe care o mașină, oricât de performantă, nu îl poate face cu adevărat.

Experimentele noastre cu baza de date MNIST au evidențiat că, pentru o mașină, învățarea este strict dependentă de calitatea și cantitatea datelor de antrenament. Atunci când contextul se modifică subtil, mașina clachează, demonstrând că învățarea algoritmică este o proiecție unidimensională a procesului cognitiv uman, lipsită de capacitatea de a generaliza în afara unui spațiu matematic predefinit.

Putem spune că oricât de mult ar semăna rezultatele, inteligența artificială și omul învață complet diferit. Pentru un algoritm, totul se rezumă la matematică, statistici și date procesate, fără nicio legătură cu realitatea din afara codului. În schimb, pentru noi, învățarea este un proces viu, care ne schimbă modul de a gândi și de a simți. Chiar dacă mașina poate recunoaște tipare mai rapid, și de multe ori mai exact, decât noi, ea nu va avea niciodată acel moment de uimire și înțelegere autentică prin care noi dăm sens lumii. Inteligența artificială rămâne un instrument de calcul performant, în timp ce învățarea umană devine o formă de înțelepciune vitală.

Din perspectiva modelului transdisciplinar (***Fig. 2. 15***) putem merge chiar și mai departe cu această distincție. Nivelul de realitate NR_3 nu reprezintă doar o limită ontologică pentru inteligența artificială, ci și locul unde se vede cel mai clar ruptura de felul în care are loc învățarea umană. Chiar dacă, în modelul nostru, situăm atât omul, cât și mașina la același nivel, atunci când vorbim despre procesele de învățare, natura lor rămâne fundamental diferită. La om, învățarea niciodată nu este izolată în NR_3, ea este conectată organic cu celelalte niveluri de realitate, „hrănindu-se" din nivelurile superioare (NR_4 și NR_5). În schimb, inteligența artificială este „prizonier" în nivelul său, neavând acces la dimensiunile superioare care dau sens și direcție cunoașterii. Ca atare, deși la suprafață par să facă același lucru, învățarea umană este un act integrat și multidimensional, în timp ce la mașină rămâne o simplă operațiune liniară, lipsită de legătura vitală cu întregul.

Înțelegerea: Limita noetică a algoritmului

După ce am stabilit distincția dintre procesele de învățare la om și la mașină, este necesar să analizăm conceptul de înțelegere ca etapă superioară a procesării informației. Acum putem să ne întrebăm foarte simplu: are inteligența artificială, în mod real, capacitatea de a înțelege ceea ce procesează? Răspunsul foarte simplu ar fi, NU. Totuși, această întrebare deschide un set de teme

esențiale la care trebuie să răspundem mai detaliat. Poate exista înțelegere autentică în absența unei conștiințe, sau ceea ce observăm la modelele actuale este doar o simulare statistică extrem de performantă? În ce măsură un sistem care operează exclusiv cu simboluri matematice poate pretinde că „înțelege" realitatea, atâta timp cât nu are contact direct cu lumea fizică și senzorială? Sau, dacă inteligența artificială reușește să ofere răspunsuri corecte fără a deține un sens interior, este oare nevoie să redefinim conceptul de „înțelegere" în era digitală? Și nu în ultimul rând, ne întrebăm dacă nu cumva există riscul ca, prin utilizarea acestor tehnologii, să ne limităm propria capacitate de înțelegere profundă, mulțumindu-ne doar cu răspunsuri teoretice rapide?

Similar discuției despre învățare, înainte de a răspunde la întrebările propuse, este util să începem și aici prin schițarea unor perspective filosofice generale asupra conceptului de înțelegere. Așa dar, din punct de vedere filosofic, înțelegerea este o capacitate fundamentală a minții umane de a interpreta și de a atribui semnificație experiențelor și informațiilor primite. Este procesul prin care individul integrează și asimilează cunoștințe, concepte, idei și perspective într-un mod coerent și semnificativ [22].

Conceptul de înțelegere a fost explorat de diverse paradigme filosofice. De exemplu, filosoful german Immanuel Kant a evidențiat rolul esențial al rațiunii în procesul de comprehensiune. *„Dar dacă orice cunoaștere a noastră începe cu experiența, aceasta nu înseamnă totuși că ea provine întreagă din experiență"* [57, p. 50]. Potrivit acestuia, înțelegerea constituie facultatea prin care subiectul organizează și sintetizează datele sensibile într-o formă coerentă, utilizând concepte apriori și categoriile intelectului. Aceste categorii, cum ar fi cauzalitatea și substanța, nu sunt derivabile din experiență, ci sunt structuri apriorice ale minții care permit configurarea fenomenelor într-o unitate inteligibilă. Așadar, prin intermediul acestor instrumente conceptuale, suntem capabili să conferim sens experienței și să articulăm o cunoaștere ordonată a lumii.

În contrast cu ființa umană, care, conform lui Kant, posedă cunoștințe apriorice, inteligența artificială nu are acces la astfel de structuri sau cunoștințe preexistente. Inteligența artificială, în forma sa actuală, operează exclusiv prin intermediul cunoștințelor aposteriorice, adică a datelor și informațiilor acumulate din experiențe și interacțiuni prealabile.

Așadar, inteligența artificială funcționează printr-un proces de învățare care este în întregime contingent și dependent de contextul specific în care este expusă la date. Spre deosebire de subiectul transcendental kantian, care poate impune anumite structuri cunoașterii prin intermediul categoriilor apriorice, inteligența artificială nu poate transcende experiența sa limitată și contextuală. Aceasta se explică de faptul că inteligența artificială rămâne întotdeauna legată de

o perspectivă particulară și fragmentată asupra realității, neputând accede la o cunoaștere universală sau necesară.

Mai mult ca atât, înțelegerea umană, întemeiată pe structurile apriorice ale intelectului, nu se limitează la organizarea experienței, ci își extinde în mod necesar aplicarea dincolo de aceasta. Tocmai în acest punct de depășire a experienței apar tensiunile rațiunii pure, formulate de Kant sub forma antinomiilor [57]. De exemplu, categoria de cauzalitate ne permite să organizăm orice eveniment ca având o cauză. Totuși, atunci când aplicăm această idee asupra întregii lumi, apare întrebarea dacă lanțul cauzelor are sau nu un început absolut. Anume aici putem observa următoarea diferență esențială. Sistemele de inteligență artificială nu operează în cadrul unei rațiuni transcendentale.

Cm poate fi interpretat acest aspect? Înțelegerea umană nu se reduce la formularea de răspunsuri, ci implică o reflecție critică asupra limitelor propriei rațiuni. În fața unor întrebări care depășesc experiența posibilă, omul este capabil să recunoască tensiunea antinomiilor și să conștientizeze caracterul problematic al aplicării categoriilor apriorice la nivelul totalității. Însă, prin contrast, inteligența artificială nu accede la această dimensiune transcendentală a înțelegerii. Deși poate genera și susține poziții contradictorii, ea nu sesizează limitele cadrului în care operează și nu reflectează asupra condițiilor de posibilitate ale cunoașterii. Astfel, „înțelegerea" artificială rămâne una funcțională și lipsită de conștiința limitelor, spre deosebire de înțelegerea umană, care este în mod esențial critică și reflexivă.

Din alt punct de vedere, filosofii fenomenologi, precum Edmund Husserl și Martin Heidegger, au acordat o atenție deosebită experienței subiective în înțelegere. Ei au subliniat importanța atenției și a reflecției în explorarea conștiinței și în obținerea unei înțelegeri profunde a lumii și a sinelui. Pentru ei, înțelegerea nu este doar o simplă activitate cognitivă, ci implică și o dimensiune afectivă și existențială [53; 51].

Husserl a insistat pe ideea că înțelegerea realității începe cu o întoarcere la experiența imediată și directă a fenomenelor. În concepția husserliană, actul de înțelegere este mediat de intenționalitatea conștiinței, care își îndreaptă atenția spre obiecte într-un mod specific și semnificativ. Acest proces intențional nu este doar unul rațional, ci este impregnat de modul în care subiectul trăiește și resimte prezența fenomenelor. În consecință, înțelegerea fenomenologică implică nu doar cogniție, ci și o deschidere către experiența trăită, care include afecte, dorințe și moduri de a fi în lume [53].

De exemplu, să ne imaginăm o dimineață răcoroasă și o terasă unde cineva bea o cană de ceai fierbinte. Pentru o inteligență artificială, cana este doar un set de date: cilindru de ceramică, temperatura de 60°C, lichid de culoare verde sau maro. Algoritmul „știe" aceste detalii prin senzori

sau descrieri, dar procesul se oprește la nivel de informație brută. Pe când pentru un om, înțelegerea aceleiași căni trece prin ceea ce Husserl numea intenționalitate și experiență trăită. Noi nu vedem doar un obiect, ci simțim căldura care dezmorțește degetele, simțim mirosul care ne trezește o amintire plăcută și resimțim momentul ca pe o pauză de liniște. În acest sens, înțelegerea umană nu este un calcul matematic, ci o „deschidere" către fenomenul respectiv. Cana are un sens pentru noi pentru că o trăim cu tot corpul și cu toată starea de spirit din acel moment. Prin urmare, în timp ce mașina doar identifică un obiect, omul îl cuprinde în contextul vieții sale, transformând un simplu obiect fizic într-o experiență plină de semnificație.

Martin Heidegger a extins aceste idei într-o direcție existențială, subliniind că înțelegerea este fundamental legată de ființa-în-lume (*Dasein*). Pentru Heidegger, înțelegerea nu este un simplu act de cunoaștere, ci o modalitate de a fi, o formă de relaționare cu propria existență și cu lumea din jur. Înțelegerea, din această perspectivă, este strâns legată de modul în care individul își proiectează posibilitățile de a fi în lume, fiind influențată de angoasele, speranțele și angajamentele sale [51].

Aici putem prelua exemplul cu ciocanul oferit de Heidegger [51, p. 93] care deschide o nouă fațetă a înțelegerii. Pentru o inteligență artificială, ciocanul este un obiect definit prin proprietăți fizice, greutate, material, formă. Pentru om însă, înțelegerea ciocanului nu vine din citirea unei fișe tehnice, ci din utilizarea lui. Tâmplarul „înțelege" ciocanul prin simplul fapt că îl folosește pentru a-și construi o casă. În viziunea lui Heidegger, această înțelegere este legată de proiecția propriului viitor. Omul nu vede doar o unealtă, ci vede posibilitatea de a avea un adăpost, siguranța familiei sau pasiunea pentru meserie. „Înțelegerea" ciocanului este împletită cu dorințele și grijile tâmplarului. Pentru mașină, ciocanul nu înseamnă nimic dincolo de imaginea sa, deoarece mașina nu are un viitor de construit sau o existență de care să se îngrijească.

Deci, în gândirea fenomenologică, înțelegerea transcende simpla activitate cognitivă și devine un act holistic, care include dimensiuni afective, temporale și existențiale. Este un proces dinamic, prin care subiectul se situează în lume și interpretează experiențele sale într-un mod care implică întreaga sa ființă, nu doar intelectul. Această perspectivă subliniază faptul că înțelegerea este o activitate integrativă, care combină gândirea, emoția și existența într-un mod indisolubil.

Nu putem trece cu vederea domeniul filosofic, probabil, cel mai strâns legat de inteligența artificială, și anume hermeneutica, care oferă un cadru esențial pentru înțelegerea procesului de interpretare și atribuire de sens. Pentru filosofi precum Hans-Georg Gadamer și Paul Ricoeur, înțelegerea este considerată ca fiind un proces interpretativ al textelor și al culturii în general. Înțelegerea nu este un proces izolat sau pur mental, ci un eveniment care are loc în contextul unui dialog continuu între individ și cultură, între trecut și prezent. Gadamer și Ricoeur arată că

înțelegerea implică întotdeauna un context și se desfășoară într-o comunitate de sensuri partajate. Drept urmare, înțelegerea devine un act comun, care presupune nu doar individualitatea, ci și intersubiectivitatea, nu doar rațiunea, ci și afectele și istoria, transformând actul de interpretare într-o experiență profund umană și interconectată [43; 104].

Să continuăm cu următorul exemplu. Dacă dăm o sarcină unui program de inteligență artificială să analizeze *„Micul Prinț”*, acesta va genera un rezumat logic și precis, însă va rămâne la suprafața datelor. În schimb, pentru un om, lectura este un „eveniment” în care sensul se construiește prin dialogul dintre lumea autorului și propria istorie de viață. O persoană care a suferit o pierdere va înțelege textul prin filtrul propriei dureri și afectivități, ceva ce mașina nu poate experimenta. Ca atare, în timp ce algoritmul procesează simboluri într-un vid emoțional, înțelegerea umană este profund intersubiectivă, fiind ancorată în contextul cultural și în capacitatea noastră de a fi mișcați de sensurile pe care le descoperim.

Doar dintr-o analiză succintă a diferitor tradiții filosofice, de la Kant la fenomenologie și hermeneutică, putem spune că înțelegerea umană este profund legată de experiența subiectivă, structuri apriorice, intenționalitate, afectivitate și dialogul cultural-istoric. Inteligența artificială, deși capabilă să proceseze și să analizeze date și texte, operează doar prin cunoștințe aposteriorice și algoritmi, lipsindu-i accesul la dimensiunile afective, existențiale și creativ-interpretative care caracterizează comprehensiunea umană. Este evident că aceste componente specifice umanului nu pot fi acoperite de niciuna dintre performanțele inteligenței artificiale actuale și una din temele pe care încercăm să le demonstrăm în lucrarea de față este că există premise ontologice puternice care fac ca acest lucru să rămână la stadiul de utopie.

Pe lângă perspectivele filosofice discutate anterior, pot fi formulate câteva concluzii privind procesul de înțelegere și pe baza studiilor de caz analizate în această lucrare, precum și a modelului transdisciplinar propus (***Fig. 2. 15***). Eșecul programului Stockfish de a înțelege impenetrabilitatea unei bariere de pioni în șah sau incapacitatea ChatGPT de a simți frica în fața unei mine marine confirmă argumentul camerei chinezești a lui John Searle. Deși algoritmul operează corect la nivel sintactic, îi lipsește dimensiunea semantică propriu-zisă. Asemenea procesului de învățare, înțelegerea, în cadrul modelului transdisciplinar, presupune interacțiunea dintre nivelul inteligenței practice (NR_3) și nivelul noetic (NR_4), unde conștiința de sine atașează un sens activ activității inteligente. Prin urmare, inteligența artificială rămânând blocată la nivelul manipulării de simboluri, doar simulând înțelegerea fără a o poseda.

Considerăm că am oferit un răspuns suficient de convingător la primele două întrebări formulate la începutul analizei noastre despre înțelegere. Rămâne însă să vedem dacă putem răspunde și la celelalte două. Este oare necesară o redefinire a conceptului de înțelegere în era

digitală? Răspunsul este – NU. Nu trebuie neapărat să redefinim înțelegerea, ci mai degrabă să o protejăm. Ceea ce face inteligența artificială este simulare, nu comprehensiune. Nu putem confunda capacitatea mașinii de a corela date cu actul uman de a sesiza sensul. Aceasta ar însemna să reducem gândirea la o simplă operațiune statistică. Dar, oare există riscuri să limităm propria înțelegere prin folosirea activă a tehnologiei? Aici putem răspunde cu un DA ferm. Obișnuindu-ne cu răspunsuri livrate instant, putem pierde antrenamentul de a trece informația prin filtrul propriei conștiințe. Dacă ne mulțumim doar cu soluții de-a gata, riscăm să devenim niște simpli consumatori de rezultate, atrofiindu-ne acea capacitate de sinteză și reflecție profundă care ne permite să transformăm datele în înțelepciune autentică.

Înțelepciunea: Orizontul transcendental al ființei

Având în vedere concluziile anterioare privind învățarea și înțelegerea, considerăm că întrebarea dacă inteligența artificială ar avea cumva înțelepciune, ceva ce este specific ființei umane, devine în mare măsură irelevantă. Mai mult, putem susține că nici măcar nu toți oamenii manifestă într-o formă deplină această calitate. Totuși, pentru a clarifica mai bine legătura dintre conceptul de inteligență, învățare, înțelegere și înțelepciune, este necesar să schițăm și aici câteva perspective filosofice, dar și teologice, asupra a ceea ce înseamnă înțelepciunea.

Înțelepciunea, așa cum este înțeleasă în mod obișnuit se referă la un nivel înalt de cunoaștere, înțelegere și judecată, care depășește simpla acumulare de cunoștințe și presupune o perspectivă profundă asupra vieții, valorilor și sensului existenței umane. Ea este asociată cu o capacitate superioară de a aplica într-o manieră rațional cunoștințele în diverse situații și de a lua decizii chibzuite și bine fundamentate [99; 11; 19].

Filosofii au abordat conceptul de înțelepciune în diferite moduri. De exemplu, în filosofia antică greacă, Socrate a considerat că înțelepciunea constă în recunoașterea propriei ignoranțe și înțelegerea limitelor cunoașterii umane. Prin celebra sa afirmație „știu că nu știu nimic”, el evidențiază faptul că umilința epistemică și reflecția asupra propriei cunoașteri reprezintă elemente esențiale ale înțelepciunii [100; 118].

Platon a definit înțelepciunea ca fiind cunoașterea esenței și a realității absolute. El considera că înțelepciunea adevărată nu se limitează la lumea fizică, ci constituie o punte către lumea ideilor eterne. Potrivit lui Platon, înțelepciunea este un obiectiv nobil și o trăsătură esențială a unui conducător înțelept [100].

În filosofia stoică, înțelepciunea era considerată ca fiind o virtute supremă și ca o cunoaștere profundă a ordinii naturale a lumii și a vocației pe care omul o are în această ordine universală. Stoicii credeau că înțelepciunea poate fi atinsă doar prin cultivarea rațiunii și prin conformarea la voința divină sau la legea naturală [79; 108; 40].

Filosoful german Immanuel Kant a asociat înțelepciunea cu capacitatea de a aplica rațiunea și principiile morale în viață. Kant a susținut că înțelepciunea morală implică recunoașterea și respectarea principiilor universale și a valorilor etice, în timp ce înțelepciunea teoretică implică obținerea unei înțelegeri profunde a lumii naturale prin intermediul rațiunii [57; 58].

Într-un sens mai larg, înțelepciunea este adesea asociată cu maturitatea, înțelepciunea vârstei și cu experiența de viață. Ea presupune capacitatea de a înțelege perspective diferite, de a aborda problemele complexe cu discernământ și de a lua decizii responsabile. Însă, înțelepciunea este un ideal care poate fi greu de atins complet, iar călătoria către înțelepciune este considerată adesea ca fiind un proces continuu de creștere și dezvoltare personală [47].

Prin urmare, din perspectivă filosofică, înțelepciunea reprezintă mai mult decât o simplă cunoaștere intelectuală. Ea implică o înțelegere profundă, rațională și morală a lumii și a existenței umane, însoțită de aplicarea practică a acestei înțelegeri în viața de zi cu zi.

Din **perspectiva teologică,** conceptul de înțelepciune este adesea asociat cu cunoașterea și înțelegerea profundă a naturii divine, a relației omului cu Dumnezeu și a învățăturilor și planului divin. Înțelepciunea teologică implică o conexiune spirituală și transcendentă, depășind cunoștințele și înțelegerea rațională obișnuită. În multe tradiții religioase, înțelepciunea este considerată ca fiind un dar divin sau o virtute supremă, iar căutarea înțelepciunii este un obiectiv fundamental al vieții spirituale. Aceasta presupune învățarea și asimilarea învățăturilor și înțelepciunii transmise de către Dumnezeu, prin revelație sau prin intermediul textelor sacre.

În tradiția creștină, înțelepciunea este considerată ca fiind legată de temerea de Dumnezeu. În Cartea Proverbelor din Vechiul Testament, de exemplu, se afirmă că *„teama de Domnul este începutul înțelepciunii”* (Proverbe 9:10). Înțelepciunea este văzută ca o calitate spirituală și morală, care se manifestă prin discernământ, prin respectarea legii divine și prin ascultarea de voința lui Dumnezeu. Dar, probabil, cea mai profundă idee în teologia creștină, este identificarea înțelepciunii cu persoana lui Iisus Hristos. În Noul Testament, Iisus este prezentat ca *"înțelepciunea lui Dumnezeu"* (1 Corinteni 1:24) și ca Cel în care *„sunt ascunse toate comorile înțelepciunii și ale cunoștinței”* (Coloseni 2:3). În această perspectivă, înțelepciunea capătă o dimensiune ontologică.

Pe lângă conexiunea cu divinul, înțelepciunea teologică include și o înțelegere profundă a condiției umane, a suferinței, a răului și a salvării. Ea poate oferi îndrumare morală și spirituală, ghidând oamenii să trăiască în armonie cu voința lui Dumnezeu și să se apropie de perfecțiunea spirituală [110].

La fel, în tradițiile teologice, înțelepciunea este considerată adesea ca fiind un dar spiritual acordat de Dumnezeu și ca rezultat al acțiunii Duhului Sfânt în viața credincioșilor. Ea poate fi

căutată și dobândită prin rugăciune, meditație, studiu al Scripturii și comuniune cu Dumnezeu. Prin intermediul acestei înțelepciuni, credincioșii sunt ghidați în procesul de transformare a minții [106]. O minte ce trece prin procesul de metamorfoză este caracterizată de virtuți precum iubirea, mila, compasiunea, umilința și dreptatea. Ea ne îndeamnă să trăim în armonie cu valorile și principiile morale promovate de credință și să ne dedicăm serviciului și bunăstării celorlalți [71]. Interesant că din punct de vedere teologic „ceilalți" nu sunt doar alte ființe umane ci include întreaga creație. Înțelepciunea teologică ne ajută să ne raportăm la lumea înconjurătoare cu respect și responsabilitate. Ea ne îndeamnă să avem grijă de mediu, să promovăm dreptatea socială și să ne implicăm în acțiuni care aduc binele și restaurarea în lume [12].

În viziunea lui N. T. Wright, înțelepciunea înseamnă să înveți să trăiești astăzi valorile Împărăției Cerurilor, deoarece lumea nouă a lui Dumnezeu este deja aici. Nu este vorba despre o teorie complicată, ci despre aducerea dreptății, iertării și iubirii în viața de zi cu zi. Pentru om, a fi înțelept înseamnă a fi un partener al lui Dumnezeu care ajută la „repararea" lumii. Acolo unde este ură, el aduce împăcare, iar acolo unde este suferință, el aduce alinare. Astfel, înțelepciunea devine un mod de a acționa prin care arătăm tuturor cum va arăta lumea atunci când totul va fi restaurat și pus în rânduială de către Creator [125].

Deși învățarea, înțelegerea și înțelepciunea sunt etape ale aceluiași proces de creștere intelectuală, ele se disting clar prin profunzime și scop. În timp ce înțelegerea ne oferă claritate mentală, înțelepciunea ne oferă discernământ moral, fiind capacitatea de a aplica ceea ce știm în situații concrete și complexe de viață. O persoană poate fi un expert teoretic în etică (învățare și înțelegere), dar rămâne lipsită de înțelepciune dacă deciziile sale nu reflectă virtuți precum curajul sau generozitatea. În acest sens, înțelepciunea este forma supremă de inteligență transformatoare, deoarece integrează rațiunea cu etica și responsabilitatea, orientând întreaga ființă umană spre binele comun și armonia cu lumea.

În concluzie putem spune că, analiza transdisciplinară a triadei învățare-înțelegere-înțelepciune relevă o barieră ontologică de netrecut pentru inteligența artificială. În timp ce algoritmul excelează la nivelul (NR_3) prin procesarea statistică a informației și recunoașterea de tipare, acesta rămâne închis în sfera sintaxei lipsită de sens. În contrast, omul parcurge aceste etape integrând dimensiunea noetică și conștiința de sine (NR_4), ceea ce transformă simpla acumulare de date într-o înțelegere vie. Așa cum am demonstrat, înțelegerea umană nu este un calcul, ci un act holistic impregnat de afectivitate, intenționalitate și experiență trăită, permițându-ne să sesizăm nu doar „ce" este un obiect, ci și semnificația lui existențială în contextul vieții noastre.

Mai mult, trecerea către înțelepciune marchează orizontul transcendental al ființei, unde cunoașterea se contopește cu morala și spiritualitatea. Dacă inteligența artificială furnizează

răspunsuri rapide, înțelepciunea oferă discernământ și capacitatea de a acționa în spiritul binelui și al restaurării lumii. Din perspectiva teologică și a valorilor Împărăției, înțelepciunea devine arta de a oglindi caracterul divin prin fapte concrete de iubire și dreptate, dimensiune inaccesibilă unei mașini.

3.2 Provocarea Transumanismului: Homo Deus și redefinirea naturii umane

„Atunci când mașinile au înlocuit trăsurile trase de cai, nu am upgradat caii – i-am scos din uz. Poate că e timpul să procedăm la fel cu Homo sapiens." [49, p. 338]

Afirmația lui Yuval Noah Harari propune o perspectivă radicală și provocatoare asupra evoluției, sugerând că specia noastră ar putea fi doar o etapă tranzitorie, și nu destinația finală a dezvoltării biologice. Analogia sa cu trecerea de la tracțiunea animală la cea mecanică subliniază o viziune pragmatică, aproape rece. Această idee ne forțează să ne întrebăm dacă umanitatea va alege să își păstreze limitele naturale sau dacă va accepta propria transformare pentru a evita irelevanța într-o lume dominată de eficiența mașinilor.

Această viziune de înlocuire sau depășire a condiției umane actuale constituie nucleul ideologic al curentului transumanist. În această ordine de idei, transumanismul este o mișcare intelectuală și culturală care explorează și promovează utilizarea tehnologiilor avansate pentru a îmbunătăți capacitățile umane, atât fizice, cât și cognitive. Scopul transumanismului este de a depăși limitările biologice și de a îmbunătăți calitatea vieții umane prin aplicarea științei și tehnologiei [14; 61; 49].

Un aspect important este îmbunătățirea fizică, care se concentrează pe utilizarea tehnologiei pentru a face un *upgrade* la capacitățile fizice ale omului. Putem discuta despre implantarea de proteze avansate care să permită o mobilitate mai bună sau de îmbunătățiri genetice pentru a preveni boli ereditare.

Exemplu concret de astfel de *upgrade* fizic este dezvoltarea brațelor robotice controlate de la distanță, care ar putea fi folosite pentru a ajuta persoanele cu dizabilități să își recâștige mobilitatea. Realitatea cruntă a conflictelor armate contemporane, precum războiul din Ucraina, transformă aceste teorii transumaniste într-o necesitate vitală de supraviețuire și reintegrare. Mii de militari care au suferit amputări în urma exploziilor, beneficiază astăzi de proteze bionice de ultimă generație. Aceste instrumente nu doar că suplinesc absența membrelor pierdute, dar integrează senzori și algoritmi capabili să traducă impulsurile nervoase în mișcări precise. Deși contextul este unul tragic, procesul de recuperare al acestor soldați demonstrează o fuziune om-mașină fără precedent. Mulți dintre ei aleg să se întoarcă pe front sau să rămână activi în societate, transformându-se în exponenți involuntari ai unei noi condiții umane, unde tehnologia devine extensia directă a voinței și rezilienței biologice.

Pe lângă îmbunătățirea fizică, abordarea cognitivă vizează dezvoltarea capacităților cognitive umane, precum memoria, atenția și inteligența. Sunt explorate tehnologii precum implanturile cerebrale sau interfețele neuronale directe (IND) pentru a spori abilitățile cognitive

umane. O istorie pe care probabil toată lumea o urmărește, deja cu un amestec de entuziasm și scepticism este proiectul *Neuralink*, inițiat de Elon Musk. Această tehnologie nu mai este doar subiect de film *science-fiction*, ci a trecut deja la faza testelor pe subiecți umani. Dispozitivul, de mărimea unei monede, este inserat direct în craniu și folosește fire microscopice pentru a citi activitatea neuronală.

Un alt scop al transumanismului constă în prelungirea vieții, dincolo de limitele maxime biologice ale speciei umane, prin utilizarea unor tehnologii adecvate și în încetinirea procesul de îmbătrânire. Cercetările în domenii precum ingineria genetică și medicina regenerativă sunt folosite pentru a găsi modalități de a învinge bolile și de a menține o sănătate optimă pe parcursul întregii vieți. Deja suntem martorii utilizării terapiei genetice pentru a repara și înlocui genele deteriorate, prevenind astfel îmbătrânirea și declinul funcțional asociate cu vârsta [49].

Perspectiva transumanistă analizează și posibilitățile tehnologice care ar putea permite oamenilor să exploreze și să colonizeze spațiul cosmic. Aceasta implică dezvoltarea de tehnologii avansate pentru călătorii spațiale, colonizarea planetelor și chiar transformarea umanității într-o specie interplanetară. Companii precum *SpaceX* și *Blue Origin* au ca obiectiv final colonizarea planetei Marte și crearea unor comunități umane permanente pe alte planete.

Este important de menționat că există și critici în legătură cu transumanismul. Acestea includ aspecte etice și sociale, precum distribuția inegală a tehnologiei. De asemenea, există temeri cu privire la pierderea unor aspecte esențiale ale umanității, cum ar fi identitatea și experiența umană autentică [13].

Având în vedere cele discutate mai sus, putem vorbi despre mai multe tipuri de transumanism. Un prim tip este transumanismul etic, care pune accent pe principiile morale și pe responsabilitatea implicată în dezvoltarea și utilizarea tehnologiilor transumaniste. Susținătorii acestui curent consideră că avansurile tehnologice ar trebui să fie realizate într-un mod responsabil, luând în considerare impactul asupra individului și societății. Ei promovează dezbateri și reguli etice care să prevină abuzurile și să asigure o dezvoltare responsabilă a tehnologiilor transumaniste [14]. Aici ne putem aminti de reacția comunității științifice internaționale față de editarea genetică a embrionilor umani, în special după controversatul caz al cercetătorului He Jiankui din 2018[8]. Deși tehnologia CRISPR[9] permitea teoretic „programarea" rezistenței la anumite boli, lipsa unui cadru moral și a unor reglementări stricte a dus la o condamnare globală, subliniind faptul că

[8] Cercetătorul chinez He Jiankui a anunțat în 2018 crearea primilor copii modificați genetic, gemenii Lulu și Nana, utilizând tehnologia CRISPR pentru a le conferi imunitate la HIV.
[9] CRISPR - acronim pentru *Clustered Regularly Interspaced Short Palindromic Repeats*. Este, în esență, un sistem de *„copy-paste"* sau „editare" pentru codul genetic.

progresul tehnic nu trebuie să devanseze responsabilitatea față de siguranța speciei. Ca răspuns, organizații precum OMS au instituit comitete de supraveghere și ghiduri etice riguroase, menite să asigure că astfel de intervenții rămân transparente, sigure și accesibile tuturor, evitând transformarea evoluției asistate tehnologic într-un instrument de discriminare sau de experimentare necontrolată.

Extropianismul este un alt curent transumanist care se concentrează pe extinderea limitelor și potențialului uman prin intermediul tehnologiei. Susținătorii extropianismului sunt interesați de tehnologii precum ingineria genetică, nanotehnologiile și inteligența artificială, considerând că acestea pot conduce la o îmbunătățire semnificativă a condiției umane. Ei susțin că progresul tehnologic ar trebui să fie urmărit și promovat în vederea realizării potențialului uman maxim și chiar a depășirii acestui potențial definit prin limitele condiției umane [159]. Un exemplu emblematic pentru extropianism este *Alcor Life Extension Foundation*, cea mai cunoscută organizație care practică criogenia[10]. Membrii acestui curent nu privesc moartea ca pe un final biologic inevitabil, ci ca pe o problemă tehnică ce va putea fi rezolvată în viitor. Ei mizează pe faptul că progresele viitoare în nanotehnologie și medicină regenerativă vor permite „repararea" celulelor și reanimarea celor conservați astăzi. Această investiție în tehnologii încă nedescoperite reflectă perfect filozofia extropiană. Susținătorii acestor idei refuză să accepta limitele biologice actuale și au convingerea că rațiunea și știința vor asigura, în cele din urmă, o existență nelimitată.

Mai există și un alt tip de transumanism, denumit transumanism democratic, care pune accentul pe implicarea și participarea cetățenilor în procesele decizionale legate de dezvoltarea tehnologică. Promotorii transumanismului democratic consideră că deciziile referitoare la tehnologii precum îmbunătățirile genetice sau inteligența artificială ar trebui luate în mod colectiv, prin implicarea cetățenilor și a societății în ansamblu. Ei promovează dezbateri publice și transparență pentru a asigura o abordare democratică a tehnologiilor transumaniste [146].

La rândul său, transumanismul libertarian reprezintă o ramură a transumanismului democratic care se concentrează asupra ideilor și valorilor libertarianiste, în contextul dezvoltării și aplicării tehnologiilor avansate pentru îmbunătățirea umană. Acest curent îmbină principiile transumaniste cu principiile libertății individuale și a minimalizării intervenției statului în viața privată și alegerile personale. Spre deosebire de transumanismul democratic, transumanismul libertarian pune mai mult accent pe autonomia individuală și libertatea de a alege în ceea ce privește îmbunătățirea și modificarea corpului și a minții umane prin intermediul tehnologiei. Susținătorii acestui curent consideră că fiecare individ ar trebui să aibă dreptul de a-și îmbunătăți

[10] criogenia - conservarea corpurilor umane la temperaturi extrem de scăzute după moartea clinică

propriile capacități și de a decide asupra propriului destin, fără restricții impuse de guvern sau de societate [61].

Trebuie remarcat că există, în acest cadru conceptual, o abordare care tinde să integreze toate tipurile mai sus amintite de transumanism. Este ceea ce unii autori definesc ca postumanism. Pentru unii dintre ei, [144] postumanismul este o abordare de tip transumanistă, care se concentrează pe ideea de a depăși condiția umană și de a transforma fundamental ceea ce înseamnă a fi uman. Cuvântul „fundamental" este aici foarte important și face diferența față de alte curente. Susținătorii postumanismului consideră că tehnologia ar trebui să conducă la o evoluție sau transformare a speciei umane, permițând dezvoltarea unor noi forme de existență și conștiință. Distinct de transumanismul „clasic", postumanismul nu își propune doar depășirea umanului printr-un anumit tip de *„upgrade"* (tehnologic, biologic, spiritual), ci depășirea umanului până la punctul în care rezultatul final este un post-om, o entitate radical diferită de specia umană. Parafrazându-l pe Nietzsche [93], cu citatul celebrul din „*Așa grăit-a Zarathustra*", omul nu mai este doar ceva care trebuie depășit, ci, mai mult, omul este ceva care trebuie recreat.

Dacă privim cu atenție, devine clar că inteligența artificială și transumanismul merg mână în mână. De fapt, ambele au în spate aceleași obiective. Ne ajută să gândim mai repede, să reținem mai multe și să luăm decizii mai bune. Astăzi, nu mai vorbim despre inteligența artificială doar dintr-o perspectivă a programelor și algoritmilor executate de un calculator, ci mai degrabă de agenți inteligenți care sunt un ajutor direct pentru creierul nostru. Această colaborare transformă ideile transumaniste în realitate, oferindu-ne instrumentele necesare pentru a depăși limitele naturale. Prin dezvoltarea de roboți și sisteme automate, inteligența artificială preia anumite sarcini și activități pe care oamenii le-au realizat anterior. Practic, inteligența artificială este motorul care ne permite să devenim o variantă „îmbunătățită" a noastră [61].

De asemenea, atât inteligența artificială, cât și transumanismul ridică întrebări provocatoare din punct de vedere etic și social. Dezvoltarea inteligenței artificiale semnalează preocupări legate de încredere, confidențialitate și impactul asupra locurilor de muncă și a economiei, iar transumanismul, la rândul său, evidențiază probleme etice privind îmbunătățirea selectivă, inegalitatea și accesul la tehnologii avansate [14].

Relația dintre inteligența artificială și transumanism este una de tip simbiotic, în care cele două domenii se alimentează și se modelează reciproc. Pe de o parte, inteligența artificială acționează ca un accelerator tehnologic, oferind soluții concrete acolo unde biologia ne limitează, de la medicina de precizie, care poate repara defecte genetice, până la descifrarea mecanismelor extrem de complexe ale creierului uman. Pe de altă parte, viziunea transumanistă oferă un scop acestor inovații, împingând cercetarea în domeniul inteligenței artificiale spre noi frontiere, cum

ar fi fuziunea dintre conștiință și mediul digital. Dar, în esență, ambele tind spre același punct de întâlnire, spre redefinirea condiției umane actuale.

Homo Deus versus *Theosis*

Cartea *"Noile religii seculare"* scrisă de profesorul Nicu Gavriluță [44] explorează trei religii seculare: corectitudinea politică, transumanismul și tehnologiile viitorului. Autorul argumentează că aceste noi fenomene sociale au devenit forme de religie pentru unii oameni. Lucrarea analizează modul în care tehnologiile viitorului se transformă în subiecte de venerație și modul cum transumanismul încearcă să redefinească ființa umană. Autorul subliniază că aceste religii seculare sunt influențate de mituri și simboluri preluate din religiile tradiționale, dar că ele distorsionează și inversează sensul acestora. Gavriluță atrage atenția asupra implicațiilor acestor religii seculare și propune o reflecție critică asupra impactului lor pentru individ și societate.

În calitatea sa de „religie", transumanismul își propune să redefinească natura umană prin intermediul tehnologiilor avansate. Autorul analizează aspirațiile și provocările acestui curent de gândire și subliniază modul în care el influențează percepția noastră asupra umanității și valorilor tradiționale.

Tehnologiile viitorului sunt reprezentate de anumite tehnologii emergente, cum ar fi Internetul, rețelele sociale și realitatea virtuală. Aceste tehnologii tind să se transforme și ele în forme de religie secularăprin intermediul cărora oamenii creează și adoră noi realități virtuale. Nicu Gavriluță evidențiază fenomenul de „fetișizare" a noilor tehnologii în societate și modul în care acestea își asumă un rol central în viața noastră de zi cu zi [44].

Bestseller-ul *„Homo Deus"* scris de Yuval Noah Harari [49] propune o privire în viitorul umanității și explorează posibilitățile și provocările pe care le-ar putea aduce evoluția tehnologică și științifică. Autorul examinează modul în care omenirea se îndreaptă către o nouă etapă a existenței, în care biologia și inteligența artificială se combină pentru a crea o nouă formă de ființă pe care el o denumește *Homo Deus*.

Prin intermediul analizei istoriei umane și a progresului tehnologic, Harari discută despre potențialul ca omul să-și depășească limitele biologice și să devină aproape nemuritor, cu abilități și puteri supraomenești, niște ființe aproape divine.

De asemenea, Harari discută despre modul în care progresele tehnologice, cum ar fi inteligența artificială și biotehnologia, pot oferi oamenilor control și putere fără precedent asupra propriei vieți și asupra altor ființe [49].

Autorul examinează și dilemele și provocările etice cu care ne confruntăm în fața acestei evoluții tehnologice, inclusiv inegalitățile sociale, dependența de tehnologie și pierderea libertății

individuale. Tehnologia va afecta conceptul de libertate individuală și democrația în viitor, având în vedere potențialul concentrării puterii în mâinile unei elite tehnologice.

Lucrarea lui Harari ne provoacă să ne gândim la sensul vieții și la ceea ce înseamnă să fim umani într-o lume în care tehnologia și știința pot transforma fundamental natura umană. În acest sens, autorul examinează modul în care religia poate evolua și se poate adapta în fața schimbărilor tehnologice și științifice, oferind oamenilor alternative la tradiționalele sisteme de credință. Autorul sugerează că tehnologia și știința pot oferi alternative la tradiționalele sisteme de credință, iar oamenii pot găsi în ele un fel de religie seculară.

Harari observă că tehnologiile emergente, precum inteligența artificială și biotehnologia, pot avea un impact semnificativ asupra religiei și a modului în care oamenii își găsesc sensul și scopul în viață. *"Creștinii cred că după moarte sfinții se vor umple de harul nemărginit a lui Dumnezeu, în vreme ce păcătoșii se separă de prezența Sa. De fapt, în Silicon Valley profeții dataiști folosesc în mod voit limbajul mesianic tradițional."* [49, p. 332]. Așa, crede Harari, tehnologia se va transforma într-o sursă de înțelepciune și putere, care va putea înlocui sau oferi o alternativă la practicile religioase tradiționale.

Un exemplu citat în carte este apariția noilor „biserici virtuale". Harari menționează că există mișcări și organizații care încearcă să creeze religii *online* sau comunități religioase virtuale. Aceste „biserici" se bazează pe tehnologie și medii digitale pentru a oferi servicii religioase, instruire spirituală sau conexiuni comunitare. Ele atrag oameni care caută să găsească sens și împlinire prin intermediul tehnologiei și a comunității virtuale.

În acest mod, conform lui Harari, tehnologia și știința pot juca un rol important în evoluția religiei și în modul în care oamenii găsesc sensul și scopul în viață. Ele pot oferi alternative la credințele și practicile tradiționale și pot satisface nevoia umană de înțelegere, conexiune și transcendere într-un mod secular. Cu toate acestea, chiar și Harari atrage atenția asupra riscului ca tehnologia să devină o nouă formă de idolatrie sau să creeze inegalități și disfuncții sociale. *„Noile tehnologii ucid vechii zei și dau naștere altora."* [49, p. 238] Prin urmare, acest aspect evidențiază modul în care tehnologia și știința pot influența și remodela religia, oferind noi perspective și provocări atunci când vorbim despre sensul vieții și practicile spirituale. Deși această evoluție poate deschide noi oportunități și căi de explorare, ea poate și să aducă cu sine consecințe negative, precum dependența excesivă de tehnologie sau pierderea unor valori tradiționale.

Harari sugerează că în această eră a tehnologiei și a cunoașterii, este important să fim conștienți de influența și implicațiile pe care acestea le au asupra religiei și a vieții noastre spirituale. El ne îndeamnă să reflectăm asupra modului în care ne raportăm la tehnologie și la

modul în care integrăm aspecte tehnologice în practicile noastre religioase sau în căutarea noastră de sens și scop.

Prin urmare, concluzia cheie a lui Harari este că în era tehnologiei și a progresului științific, religia se confruntă cu provocări semnificative și este nevoie de o adaptare și reevaluare a conceptelor și practicilor religioase. Este important, spune autorul, să abordăm aceste schimbări cu o minte deschisă și cu discernământ, pentru a ne asigura că evoluția religiei va avea loc într-un mod favorabil promovării valorilor umane fundamentale și, astfel, va fi capabilă să răspundă nevoilor spirituale ale omenirii.

Un alt exemplu concret dat de Harari este avansarea în domeniul ingineriei genetice, care ar putea permite oamenilor să-și modifice ADN-ul și să-și îmbunătățească trăsăturile și capacitățile fizice și cognitive. *„Cu 70.000 de ani în urmă, Revoluția Cognitivă a transformat mintea sapiens, făcând astfel dint-o maimuță africană neînsemnată stăpânul lumii. Minţile sapiens îmbunătăţite au avut dintr-odată acces la vastul tărâm intersubiectiv, care le-a permis să creeze zei şi corporaţii, să construiască oraşe şi imperii, să inventeze scriere şi banii şi, în cele din urmă, să descompună atomul şi să păşească pe Lună. Din câte ştim, această revoluţie capitală s-a datorat câtorva mici modificări în ADN-ul sapiens şi unei uşoare reconfigurări a creierului sapiens."* [49, p. 309]

Autorul stabilește fundamentul succesului speciei s*apiens* prin capacitatea de a colabora cu ajutorul ficțiunilor colective. Nu forța fizică, ci abilitatea de a crede în entități abstracte precum statul, moneda sau religia, a permis coordonarea la scară largă. Această „revoluție" marchează momentul în care evoluția culturală a început să devanseze evoluția biologică, transformând o specie vulnerabilă în forța dominantă a planetei.

În continuare Harari mută discuția de la istorie la speculație futuristă, introducând premisa centrală a tehnoumanismului. „ *...tehnoumanismul, poate că alte câteva schimbări în genomul nostru şi o nouă reconfigurare a crierului vor fi de ajuns pentru a lansa o a doua revoluţie cognitivă. Datorită prefacerilor mentale ale primei Revoluţii Cognitive, Homo sapiens a putut păşi pe tărâmul intersubiectiv şi a devenit stăpânul planetei; o a doua revoluţie cognitivă ar putea să faciliteze accesul lui Homo deus la noi tărâmuri inimaginabile şi să-l facă stăpânul galaxiei."* [49, p. 309]

Dacă prima revoluție a fost un accident genetic „ușor" care a deschis porțile către abstractizare, autorul sugerează că o intervenție deliberată asupra genomului și a arhitecturii neuronale ar putea declanșa un salt calitativ similar. Trecerea de la *Sapiens* la *Deus* nu este văzută doar ca o evoluție tehnologică, ci ca o extindere a realității însăși, unde limitele speciei devin simple bariere de cod ce pot fi rescrise.

Încă un exemplu este și dezvoltarea inteligenței artificiale și a roboților autonomi. Tehnologiile asociate inteligenței artificiale ar putea conduce la apariția unor ființe superinteligente și superputernice. Harari afirmă că o inteligență artificială care depășește semnificativ intelectul uman ar putea deveni o entitate divină în raport cu noi. Aceste entități ar putea avea cunoștințe și abilități mult peste cele umane și ar putea lua decizii care ne-ar afecta profund existența [49].

În esență, Harari susține că prin tehnologie și știință, oamenii pot deveni ființe cu puteri și abilități supraomenești, aproape divine. Deși este foarte optimist, autorul însuși ne avertizează și asupra consecințelor și dilemelor pe care această evoluție le poate aduce. De exemplu, omenirea ar putea să se confrunte cu inegalități profunde între cei care au acces la aceste tehnologii și cei care nu au, iar aceste transformări ar putea submina valorile și tradițiile umane fundamentale.

În consecință, evoluția umană către *Homo Deus* ridică întrebări și provocări importante cu privire la identitatea umană, la limitele noastre și la implicațiile sociale, etice și existențiale ale unei astfel de transformări. Este o explorare fascinantă a potențialului uman și a direcției în care se îndreaptă umanitatea în era tehnologiei și a științei.

În continuare, vom compara ideile filosofice transumaniste cu un concept cheie al teologiei creștine pentru a arăta că în ciuda anumitor similitudini acestea sunt radical diferite. Perspectiva creștină a divinizării sau îndumnezeirii, cunoscută și sub numele de *theosis*, susține că scopul ultim al vieții umane este de a deveni „dumnezei” (vom explica în detaliu acest concept pentru a exclude orice fel de speculații sau neclarități) prin participarea la viața și natura lui Dumnezeu. În linii generale *theosis* înseamnă că Dumnezeu a creat oamenii într-un mod unic, după chipul și asemănarea Sa, cu potențialul de a dezvolta o relație personală și de a deveni tot mai asemănători cu El prin harul Său. *Theosis* implică o transformare în care natura umană este purificată, îmbunătățită și înnoită prin lucrarea Duhului Sfânt [75].

Pe de altă parte, transumanismul este o mișcare care explorează posibilitățile tehnologice și științifice pentru a îmbunătăți condiția umană și pentru a depăși limitele biologice și cognitive ale acesteia. Așa cum am mai arătat, transumanismul propune utilizarea tehnologiilor emergente, cum ar fi inteligența artificială, nanotehnologia, ingineria genetică și interfețele creier-computer, pentru a îmbunătăți capacitățile umane și pentru a permite evoluția și transformarea umană într-o formă superioară.

Deși perspectivele creștine și transumaniștii conțin puncte de vedere radical diferite, există și câteva similitudini între ele. Una dintre ele este aspirația către perfecțiune și transcendere. Atât mistica creștină, prin *theosis*, cât și transumanismul urmăresc să depășească condiția umană

limitată și să atingă un nivel calitativ superior de existență. În ambele cazuri, există dorința de a depăși slăbiciunile, suferința și moartea pentru a atinge o stare de plenitudine și împlinire.

De asemenea, ambele perspective subliniază necesitatea unei transformări, a unei „evoluții". În *theosis*, transformarea se realizează prin lucrarea Duhului Sfânt, iar în transumanism prin intermediul tehnologiei și inovației. În ambele cazuri, persoana este văzută ca fiind într-o continuă dezvoltare și având capacitatea de a se schimba și de a progresa către o formă superioară de existență.

Totuși, așa cum am afirmat mai sus, există diferențe radicale între cele două perspective, atât de radicale încât cele două puncte de vedere devin iremediabil incompatibile. În viziunea creștină, divinizarea este privită ca un dar al lui Dumnezeu și ca o participare la natura divină prin har [75]. Este un proces în care omul răspunde și cooperează cu lucrarea lui Dumnezeu în el. Pe de altă parte, transumanismul se concentrează pe capacitatea umană de a-și îmbunătăți condiția prin mijloace tehnologice, fără a implica neapărat o intervenție divină sau o dimensiune spirituală [49].

În plus, îndumnezeirea în teologia creștină implică o transformare interioară a naturii umane, în timp ce transumanismul se concentrează mai mult pe îmbunătățirea capacităților și performanțelor exterioare ale omului prin intermediul tehnologiei. Transumanismul explorează posibilitatea extinderii duratei de viață, creșterea inteligenței, îmbunătățirea sănătății și chiar posibilitatea de a transcende condiția umană prin transferul conștiinței în medii digitale sau în corpuri artificiale.

În ciuda acestor diferențe radicale între viziunea creștină a îndumnezeirii și transumanism, este interesant de observat că mișcarea transumanistă a crescut și a evoluat în mare măsură într-un context cultural și filosofic influențat de tradițiile religioase creștine. Unii transumaniști au preluat și reinterpretat elemente din teologia creștină tradițională, construindu-și astfel propria perspectivă despre transcendere și evoluția umană, dar într-un context secular [49].

Așadar, ce urmărește transumanismul? Un (alt) „om nou", un om *upgradat*/îmbunătățit prin contribuțiile biotehnologiei și ale inteligenței artificiale, de fapt, o depășire a omului actual (iremediabil deficient, carent, imperfect) prin anularea lui și trecerea într-un fel de supraom.

Ecuația antropologică a transumanismului este deci aceasta: transumanul = om natural (biologic) + adaos (antropoid) artefactual/biotehnologic.

Antropologia creștină propune cu totul altceva: un om restaurat/mântuit (în primă fază), adică readus la starea inițială/paradisiacă, și apoi desăvârșit prin sfințenie, care e „împlinirea umanului" [20], nu *upgradarea* lui. Antropologia creștină vrea deci, și ea, depășirea stării actuale a omului, dar o depășire care nu pierde nimic pe drum, nu anulează, ci convertește, preschimbă.

Omul păcatului este mântuit, în sensul că este convertit (*id est i se schimba, prin metanoia, direcția investiții mentale și existențiale*), nu redefinit, nu transgresat [20].

Pleroma (plinătatea) omului se definește ca unirea omului cu Dumnezeu (omul deplin este Dumnezeu-omul), după cum, reciproc, *pleroma* lui Dumnezeu s-a arătat prin Înomenirea Lui – în Hristos „locuiește, trupește, toată plenitudinea dumnezeirii (*pan to pleroma tes theotetos*)" – Coloseni 2:9 [81].

Îndumnezeirea (*theosis*), ca ideal uman al creștinismului Răsăritean, nu (pre)schimbă definiția omului, nu înseamnă ridicarea omului la o treaptă supra- sau transumană, ci realizarea plenară a omului după (*kata*) Arhetipul divino-umanității, Arhetip asumat din veci/dincolo de timp de către Fiul. Îndumnezeirea nu-i modifică omului „rațiunea firii *(logos physeos*)", ci doar „modul subzistenței (*tropos hyparxeos*)", după distincția făcută de Sf. Maxim Mărturisitorul [81] adică îi (re)convertește facultățile definitorii, deturnate de păcat, spre scopul (*telos*) lor inițial si firesc, adică spre Dumnezeu.

Divino-umanitatea, ca plenitudine a umanului, nu e un mixaj ontologic (de tip centaurin), nici o upgradare a umanului (după cum, evident, nu e nici o degradare a divinului), ci este o „sinteză iconică", o întrecuprindere în care umanul este afirmat plenar în divin și viceversa. Aceasta pentru că, iarăși după celebra rostire a Sf. Maxim, „Dumnezeu și omul își sunt unul altuia modele (gr. *allelon einai paradeigmata ton Theon kai ton anthropon*)" [81, p. 112]. Acesta este principiul analogiei (*tantum-quantum*) între Dumnezeu și om, preluat de Sf. Maxim de la Sf. Atanasie cel Mare, dar rafinat și dus până la ultimele lui consecințe: atât de mult (*tantum*) se poate face omul Dumnezeu pe cât de mult (*quantum*) S-a făcut Dumnezeu om: adică deplin, dar bineînțeles fără confuzie sau amestecare (dar și fără separare/scindare, ca și cum divinul ar putea fi perceput ca un etaj superior, ultim al umanului – *homo deus*): „Dumnezeu și omul stau «față în față», pe pozițiile unei «egalități (iconice) de prezență» (nu ale unei identități ontologice sau logice), în conlocuire și revelare reciprocă". Este vorba de un tip de „egalitate" (iconică) ce coexistă cu un tip nou (tot iconic) de „ierarhie", întrucât Dumnezeu – deși analog omului – Își păstrează simultan și ascendentul față de acesta.

Un loc aparte îi revine conceptului de **divino-umanitate** dezvoltat și realizat în filosofia rusă pre-revoluționară și în cea din diaspora constituită după revoluția bolșevică din 1917. Divino-umanitatea (*Богочеловечество*) pe drept poate fi numită una dintre ideile de bază în gândirea religioasă-filosofică și teologică rusă, idee opusă conceptului de „om-deitate" (*человекобожество*) care afirmă umanul în afara lui Dumnezeu. În mod fundamentat, problema raportului dintre om-deitate și divino-umanitate a fost abordată de numeroși gânditori ruși.

De exemplu, Dostoievski [183] afirma că libertatea fără Dumnezeu este goală și conduce omul spre auto-divinizare, spre om-deitate, manifestându-se prin inumanitate și cruzime. Omul-divinizat, adică ceea ce era pentru Dostoievski un antiideal, era vehement afirmat de către Nietzsche [93]. Umanitatea și compasiunea trebuie să fie depășite și abandonate, omul fiind doar un mijloc pentru apariția supraomului, care se rupe de moralitatea creștină („roaba") și de moralitatea umanistă. Vladimir Soloviov [188], încearcă să introducă omo-deitatea ca un pas sau o parte componentă a procesului către divino-umanitate. Chiar și în progresul secular, el a observat o umbră secretă a Duhului lui Hristos. Umanismul secular a dobândit un înțeles pozitiv pentru el fiind necesar, fără de care scopurile creștinismului nu ar putea fi realizate. În persoana lui Iisus Hristos a avut loc unirea naturii divine și umane, și El S-a manifestat ca divino-umanitate. Același lucru, adică întruparea sobornicească a lui Dumnezeu, trebuie să se realizeze, potrivit gândirii lui Soloviov, în umanitatea creștină. Doar în Biserică omul poate găsi justificare și sprijin pentru creație sau acțiunea istorică [188].

Urmându-l pe Soloviov, Bulgakov [182] insistă deja asupra universalității ideii de divino-umanitate ca principiu care pătrunde și în omo-deitate. Bulgakov creează o trilogie teologică în care el rămâne fidel ideii de divino-umanitate, considerând că ea se realizează prin Duhul Sfânt, fiind de fapt o divinizare a creaturii. În schimb, progresul „umanist" este puternic condamnat de el, deoarece forța sa motrice nu este iubirea, mila, ci visul mândru al unui paradis pământesc în contrast cu paradisul ceresc, al unei împărății conduse de omul-dumnezeu. Aceasta este așa cum afirmă Bulgacov o nouă variantă a vechiului mesianism iudaic fals.

Berdiaev, [179; 180; 181] formulează divino-umanitatea ca „tema gândirii rusești". Omul își păstrează valoarea cea mai înaltă, libertatea și independența sa doar dacă există Dumnezeu și divino-umanitatea. Potrivit lui Berdiaev, călătoria tânărului Marx de la umanism la comunism, precum și călătoria socialismului rus de la Belinski la comunism, au fost, în opinia sa, afirmarea omului în afara divino-umanității și au condus la negarea umanității. Omul trebuie să se îndumnezeiască, dar poate face acest lucru doar prin intermediul divino-umanității. Divino-umanitatea presupune activitatea creatoare a omului. Mișcarea are loc inclusiv dinspre om către Dumnezeu, nu doar de la Dumnezeu către om. Aceasta este o mișcare creatoare, continuarea creației. Ideea de divino-umanitate ca esență a creștinismului în opinia lui Berdiaev a fost puțin dezvoltată în gândirea occidentală, de aceea ea reprezintă o creație originală a gândirii creștine ruse, în care filosofia creștină este înțeleasă ca filosofie a divino-umanității.

S. L. Frank [190] a remarcat că mulți mistici, în special cei germani, au o concepție despre nașterea nu numai a eului nostru în Dumnezeu, ci și a lui Dumnezeu în noi, ceea ce reprezintă în esență divino-umanitatea, și numai prin ea se deschide adevărata și concretă plinătate atât a

umanității, cât și a Divinității. El afirmă că Dumnezeu este cu adevăratul Dumnezeu doar în calitate de divino-umanitate. Umanitatea ca divino-umanitate se dezvăluie în conștientizarea unității interioare a omului cu Dumnezeu, iar deoarece acest lucru este legat de natura sobornicească a eului, noi avem revelația divino-umanității.

Așa dar, spre deosebire de proiectul *Homo Deus*, care urmărește o transgresare a limitelor biologice prin „adaos" tehnologic și autonomie radicală, conceptul de *Theosis* propune o restaurare a umanului prin participarea harică la natura divină, fără a anula firea omenească. În timp ce transumanismul caută nemurirea prin expansiunea orizontală a puterii și a informației, antropologia creștină vizează plenitudinea prin unirea verticală cu Divinitatea, transformând „omul-păcatului" în „omul-sfințit" prin convertire, nu prin redefinire genetică.

3.3 Perspectiva teologică și antropologică: Corporalitatea și Theosis

„În tonul vocii, în ochii și în ținuta unei persoane
nu este mai puțină elocvență decât în alegerea cuvintelor.”
La Rochefoucauld, *Maxime*, 249 [62]

În *What Computers Can't Do*, Hubert Dreyfus [135] subliniază în mod distinct rolul corpului și al experienței încorporate în cunoașterea și raționamentul uman. El argumentează că gândirea și înțelegerea umană nu sunt doar rezultate ale procesării simbolice, așa cum sugerează multe modele timpurii de inteligență artificială, ci sunt profund legate de faptul că suntem ființe corporale care trăiesc și interacționează în lume.

Punctul de vedere care leagă corporalitatea de inteligența în sens uman contrastează cu abordarea tradițională a inteligenței artificiale, care consideră mintea ca fiind separată de corp și funcționând asemănător cu un calculator. Dreyfus susține că multe dintre cunoștințele și competențele noastre sunt rezultatul direct al experiențelor noastre senzoriale și motorii și că acestea nu pot fi ușor reprezentate printr-un set de reguli formale sau algoritmi (aceeași idee este întărită în *What Computers Still Can't Do*, publicată în 1992)[136].

Acest punct de vedere a avut implicații profunde pentru studiul cogniției și a fost ulterior explorat și dezvoltat în filosofia cognitivă, dând naștere conceptelor de „cogniție încorporată” (*embodied cognition*) și „cogniție situată” (*situated cognition*). Aceste idei sugerează că gândirea noastră este, prin natura sa, legată de corp și de mediu și că nu poate fi complet înțeleasă sau reprodusă în absența acestui context.

Perspectiva că trupul ocupă o poziție centrală în percepția și înțelegerea noastră asupra lumii reprezintă un subiect profund în filosofia minții și a cogniției. De-a lungul anilor, numeroși autori au oferit contribuții semnificative la această temă.

Unul dintre cei mai influenți filosofi fenomenologi ai secolului al XX-lea, Maurice Merleau-Ponty a subliniat rolul corpului în percepție și cunoaștere. În *Fenomenologia percepției*, Merleau-Ponty [87] descrie corpul ca fiind un „vehicul al ființei în lume”, sugerând că înțelegerea noastră a lumii este în mod inerent legată de corpul nostru și de modul în care interacționează cu mediul.

În *The Embodied Mind*, Francisco Varela, Evan Thompson și Eleanor Rosch [173] introduc conceptul de „cogniție încarnată” (*embodied cognition*). Ei argumentează că mintea nu este doar un procesor de informații abstracte, ci este profund înrădăcinată în structura și dinamica corpului.

În *Metaphors We Live*, George Lakoff și Mark Johnson [151] susțin că metaforele lingvistice, care formează baza multor concepte abstracte, sunt fundamental bazate pe experiența corporală. În lucrări ulterioare, cum ar fi *Philosophy in the Flesh: the Embodied Mind & its Challenge to Western Thought* [152] aceeași autori dezvoltă această temă, sugerând că gândirea umană este modelată de experiențele corpului nostru în lume.

De asemenea, filosoful Andy Clark, în lucrări precum *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* [130] și *Being There: Putting Brain, Body, and World Together Again* [131], explorează ideea de „cogniție extinsă", sugerând că mintea umană nu se limitează la creier, ci se extinde în corp și chiar în mediul înconjurător.

La fel, în *Action in Perception*, Alva Noë [165] argumentează că percepția nu este doar un produs pasiv al stimulilor senzoriali, ci implică activ corpul și interacțiunile sale cu lumea.

Aceste viziuni contrastează cu modelul tradițional al minții ca procesor de informații și pun accentul pe importanța experienței corpului și a interacțiunii cu lumea în formarea cunoașterii și a înțelegerii. Cu acest prilej, constatăm cât de fertilă poate fi perspectiva holistică asupra ființei umane, considerând mintea, corpul și lumea ca fiind profund interconectate și interdependente.

Hubert Dreyfus a fost influențat puternic de fenomenologia existențialistă, în special de lucrările lui Maurice Merleau-Ponty și Martin Heidegger. Critica sa față de inteligența artificială în lucrarea *What Computers Can't Do* [135] a fost în mare măsură bazată pe aceste fundamente filosofice.

În centrul viziunii lui Dreyfus asupra corpului se află ideea că cunoașterea umană și abilitatea de a naviga în lume nu sunt bazate doar pe reprezentări mentale sau pe procesarea simbolică, ci pe experiența directă și încorporată [135; 136].

Dreyfus a introdus ideea de *skillful coping*, adică modul în care ființele umane interacționează cu lumea într-un mod direct și nemediat, fără a necesita o deliberare explicită sau un raționament conștient. De exemplu, atunci când un sportiv joacă un joc sau când un muzician cântă la un instrument, acestea nu se bazează pe reprezentări mentale sau pe aplicarea de reguli, ci pe o înțelegere corporală și automatizată a situației.

Un alt aspect important pe care autorul îl sugerează este acela că ființele umane au un *background understanding* sau o înțelegere fundamentală și implicită a lumii care informează toate acțiunile și deciziile noastre (destul de asemănător cu ceea ce filosofii înțeleg prin *Weltanschauung*). Această înțelegere nu este formată numai din reguli sau fapte specifice, ci și din experiența noastră corporală și din relația noastră continuă cu mediul.

Dreyfus a criticat ideea că ființele umane se bazează în principal pe reprezentări mentale pentru a naviga în lume. El a susținut că, în multe situații, acțiunile noastre sunt nemediate și directe, fără a necesita o „hartă mentală” sau o reprezentare explicită a lumii.

De asemenea, autorul a subliniat importanța contextului și relevanței situaționale în gândirea și acțiunea umană. Față de computere, care necesită instrucțiuni explicite pentru fiecare situație, ființele umane pot naviga în lume bazându-se pe experiența lor încorporată și pe capacitatea de a recunoaște relevanța specifică a situației. Acest aspect încearcă să implementeze rețelele neuronale artificiale (vezi studiul de caz nr. 2).

Aceste idei au fost centrale în critica lui Dreyfus față de inteligența artificială, în special față de viziunile optimiste ale cercetătorilor din anii '60 și '70, care credeau că gândirea umană poate fi reprodusă printr-un set de reguli formale într-un computer. Pentru Dreyfus, corpul nu este doar un vehicul pasiv pentru minte, ci este central în modul în care înțelegem și interacționăm cu lumea [135; 136].

Pe lângă cele menționate mai sus, având în vedere caracterul transdisciplinar al lucrării de față, considerăm că este importantă menționarea importanței trupului (corpului) în viziunea teologică creștină. În acest sens, trupul (corpul) nu este doar un element al naturii umane, ci și o dimensiune a persoanei umane. Deși avem cu toții o natură comună care ne deosebește de alte specii, suntem persoane distincte și diverse. Suntem asemănători prin natura noastră, dar diferiți și unici ca persoane. Fiecare dintre noi își asumă natura omenească într-un mod singular și unic. Suntem înzestrați cu rațiune, memorie, imaginație, afecte, emoții, sentimente, libertate și liber arbitru. Chiar și așa, fiecare dintre noi folosește aceste trăsături diferit și are propriile sale experiențe și alegeri [65].

Personalitatea este unică și se distinge de caracter, care este o particularitate a naturii noastre în plan psihologic. Persoana se diferențiază de individ, deoarece individualitatea este o categorie biologică și sociologică aplicabilă și altor specii, nu doar oamenilor. Pe de altă parte, noțiunea de persoană se referă la ființe spirituale înzestrate cu inteligență, voință și libertate. Persoana are o valoare absolută și o dimensiune transcendentă. Este capabilă să se depășească constant și să atingă o stare duhovnicească în care modalitățile de existență ale naturii umane sunt depășite [11].

Trupul este o parte integrantă a persoanei și este părtaș la valoarea sa duhovnicească. Este considerat de Sfinții Părinți ca fiind parte a persoanei încă de la concepție și chiar dincolo de viața pământească. Trupul participă la devenirea duhovnicească a persoanei și este legat de suflet și de duh. Concepția creștină despre trup se distinge de cea naturalistă, care adesea reduce trupul la un obiect și îl disociază de celelalte aspecte ale ființei umane [65].

Astăzi, trupul uman este adesea tratat ca un obiect în publicitate, în medicină și în reprezentările culturale. În această viziune, trupul este lipsit de dimensiunea sa spirituală și personală și este redus la aparențele sale fizice. Aceasta se poate observa în exploatarea sexuală a trupului în industria pornografică și în tratarea bolnavilor ca simple obiecte în medicină modernă.

Antropologia creștină și etica bazată pe ea mărturisesc respectul față de trup ca fiind un element inseparabil de persoană. Orice afront adus trupului este considerat un afront adus persoanei înseși, iar respectul față de trup este un semn de respect față de persoană.

Trupul uman nu poate fi redus la un simplu obiect fizic sau la o colecție de organe. El este parte integrantă a ființei umane și joacă un rol esențial în exprimarea și manifestarea acesteia. Fiecare trup este unic și posedă trăsături și caracteristici individuale, care îl fac deosebit și special. Totodată, trupul uman nu poate fi separat de persoana la care aparține. Trupul și persoana sunt într-o relație indivizibilă și inseparabilă.

În concepția creștină, trupul nu este considerat doar o entitate materială, ci are și o dimensiune duhovnicească. El este văzut ca un dar de la Dumnezeu, creat în chipul și asemănarea Sa. Prin urmare, trupul uman are o valoare intrinsecă și o sacralitate pe care trebuie să le respectăm și să le protejăm.

Tratarea trupului uman ca un obiect, cum se întâmplă în publicitate sau în industria pornografică, este o distorsionare gravă a valorii și demnității sale. Această abordare disociată de trupul uman îl devalorizează și încurajează exploatarea și lipsa de respect față de persoana umană.

În societatea contemporană, este nevoie de o medicină care să nu reducă oamenii la simpli pacienți sau la obiecte de tratat. O abordare creștină a medicinei ia în considerare integralitatea ființei umane, atât aspectele fizice, cât și cele emoționale, spirituale și relaționale. Ea respectă demnitatea și unicitatea fiecărui individ și își propune să vindece și să restaureze întreaga persoană [39].

Așadar, concepția creștină despre trup și despre persoană contrastează cu viziunea seculară și materialistă care predomină în societatea contemporană. Ea ne amintește că suntem mai mult decât simple obiecte fizice și că trupurile noastre au o valoare și o semnificație profundă. Respectarea și îngrijirea trupului uman sunt expresii ale respectului și iubirii față de persoana umană în ansamblul său [110].

În concluzie, trupul uman nu poate fi redus la un simplu obiect sau la o realitate disociată de persoana la care aparține. El este o dimensiune esențială a persoanei umane și trebuie tratat cu respect, demnitate și iubire. La fel am arătat importanța și rolul distinct al corpului și al experienței încorporate în cunoașterea și raționamentul uman. Gândirea și înțelegerea umană sunt profund

legate de faptul că suntem ființe corporale care trăiesc și interacționează în lume. Inteligența artificială este lipsită de această caracteristică fundamentală.

Antropologia Sf. Maxim Mărturisitorul

Sfântul Maxim Mărturisitorul este recunoscut ca unul dintre cei mai de seamă teologi ai Patristicii răsăritene. El a stabilit bazele unui sistem teologic cuprinzător, caracterizat prin coerență și integrare și oferind o alternativă semnificativă la origenism[11]. În teologia Sfântului Maxim o importanță deosebită o are hristologia și antropologia creștină, care sunt legate între ele. Pentru Sfântul Maxim, legătura dintre hristologie și antropologie devine extrem de strânsă, deoarece el consideră că „întruparea reprezintă centrul existenței lumii, nu doar din perspectiva răscumpărării, ci și din perspectiva inițială a actului creator" [81; 83; 84]. Sf. Maxim este cunoscut prin abilitatea sa remarcabilă de a sintetiza fundamentele dogmatice ale primelor sinoade ecumenice, dar și al altor teologi importanți pentru acea perioadă.

Scrierile Sf. Maxim sunt numeroase și cuprind întregul spectru al teologiei creștine, de aceea ar fi imposibil în cadrul acestei lucrări să facem o sinteză generală al operei sale. În primul rând, având în vederea tema cercetări noastre suntem interesați în mod special de antropologia sa, care ar aduce o clarificare și o revalorificare a ființei umane, a cărei imagine a fost încețoșată de anumite interpretări exagerate asociate cu noile tehnologii ale inteligenței artificiale.

Jean-Claude Larchet este considerat unul dintre cei mai de vază experți ai operei Sf. Maxim. În scrierile sale Larchet scoate în evidență câteva elemente cheie ale antropologiei Sf. Maxim [63; 64].

În primul rând, la Sf. Maxim natura omului este compusă din suflet și trup, unite inseparabil. Sfântul Maxim respinge origenismul, susținând simultaneitatea celor două componente ale ființei umane, în contradicție cu teoriile care afirmă preexistența sufletului sau a trupului. De asemenea, Sf. Maxim respinge și credința în metempsihoză (transmigrarea sufletului după moarte, reincarnare). Larchet scoate în evidență și ideea Sf. Maxim despre statutul embrionului și viața de după moarte, în care sufletul și trupul nu există separat, ci rămân împreună ca parte a unei persoane unice. Aceste cugetări sunt de o importanță majoră în cadrul discuțiilor contemporane despre bioetică [63].

[11] Origenism: Sistem teoretic (atribuit lui Origen și discipolilor săi) care susține preexistența sufletelor și căderea acestora în trupuri ca pedeapsă. În această viziune, lumea materială este rezultatul unei „erori" inițiale, mântuirea însemnând revenirea spiritului la o stare incorporală, spre deosebire de viziunea Sf. Maxim, unde trupul și materia sunt componente esențiale și bune ale creației, menite îndumnezeirii.

Logosul firii umane are un rol central în metafizica și învățătura Sf. Maxim. El definește ființa umană ca fiind în conformitate cu planul divin, așa cum a fost gândită și dorită de Dumnezeu înainte de Creație, precum și conform finalității pe care Dumnezeu i le-a menit de la început, pentru a-și găsi împlinirea în El. Sfântul Maxim respinge ideea unei perfecțiuni inițiale pierdute ulterior, susținută de origenism, și concepe desăvârșirea ca un scop care este conform voinței lui Dumnezeu, dar doar potențial, reprezentând norma de desăvârșire pentru creaturile înzestrate cu capacitatea de a se îndumnezei [64]. Omului i-au fost date puterile necesare pentru a atinge această finalitate destinată lui de către Creator. Cu toate acestea, nici natura care îndeamnă omul să se îndrepte către Dumnezeu, nici harul care îl susține, nu constrâng persoana umană. Prin libertatea sa de voință, omul poate să se conformeze sau să se opună acestora [63].

Sf. Maxim Mărturisitorul utilizează analogia dintre unirea trupului și a sufletului în om și unirea naturii divine și umane în Hristos, pentru a susține ideea unei uniri de structură. El consideră că unirea dintre trup și suflet este nu doar ipostatică, ci și naturală. În învățătura hristologică a Sfântului Maxim, perihoreza (întrepătrunderea) reprezintă un punct central, iar analogia dintre suflet și trup poate fi numită „întrepătrunderea trupului și a sufletului” [81].

Sfântul Maxim afirmă că trupul și sufletul sunt neamestecate, dar unite indisolubil, păstrându-și diferențele distinctive. El evidențiază coexistența originară a trupului și sufletului și argumentează că sufletul nu încetează niciodată să se raporteze la trup, iar trupul la suflet. Prin aceasta, el subliniază reciprocitatea perfectă și caracterul integrant al trupului în întregul uman. Sfântul Maxim dezvoltă aceste concepte nu doar în context hristologic, ci și în relație cu creația și cosmologia, oferind astfel o perspectivă asupra scopului final al creației [83; 84].

Ipostasul uman se referă la unitatea dintre trup și suflet în ființa umană. Această unitate este creată de Dumnezeu și nu se bazează pe o atracție naturală între ele, ci pe voința divină. În Iisus Hristos, există două naturi care se unesc într-un singur ipostas, iar unirea trupului și sufletului în om este și ea naturală și ontologică. Mintea, *nous*, este un al treilea element constitutiv al omului și reprezintă partea contemplativă și responsabilă în relația cu Dumnezeu. Omul ca și persoană este o unitate de trup și suflet, iar libertatea și unicitatea fiecărui individ conferă valoare și unicitate persoanei umane. Sufletul nu poate fi separat de trup, iar suferința trupească este simțită și de suflet. Întruparea Cuvântului în Hristos reprezintă o realizare unică și desăvârșită a umanității [187].

O altă distincție importantă pe care o face Sfântul Maxim Mărturisitorul este între natură și ipostas. Natura se referă la aspectul universal și comun al ființelor, în timp ce ipostasul este persoana individuală care particularizează și diferențiază o ființă. Ipostasul determină mișcarea și

exercitarea puterilor sufletului, în timp ce natura rămâne neschimbată în logosul său. Prin voință și alegeri proprii, persoana poate influența modul în care natura își manifestă puterile.

În legătură cu conceptele de natură și ipostas, Sf. Maxim vorbește despre voința ca o predispoziție rațională specifică omului, care permite autodeterminarea. El face distincție între „simpla voință" care este legat de conceptul de natură și „felul în care se voiește", legat de ipostas („voința gnomică") . Voința naturală este una rațională și vitală și poate fi înțeleasă ca „*o putere doritoare a ceea ce există prin fire*" [63], ceea ce o face să nu dorească nimic rău sau împotriva firii, în timp ce voința gnomică implică orientarea către bine sau rău. Voința gnomică este legată de dispoziția de a vrea și se manifestă în alegeri și acțiuni morale [64].

Urmând deja conceptele analizate, logosul și troposul reprezintă o diadă fundamentală în gândirea Sfântului Maxim. Logosul este principiul fix și imuabil, reprezentând definiția și legea firii, în timp ce troposul se referă la diversitate, variație și modul în care o ființă există și acționează. Troposul este influențat de dispoziția de a voi și de alegerile făcute de persoană, exprimându-se prin comportament și maniere de a fi. El este legat de modul în care persoana utilizează și pune în practică capacitățile și puterile sale naturale [64]. Troposul poate fi o folosire bună sau rea a acestor facultăți și lucrări naturale, determinând dacă o persoană acționează în conformitate cu natura sa sau împotriva ei, dacă este virtuoasă sau nu, dacă trăiește în acord cu voia lui Dumnezeu sau în contradicție cu ea. Deci, troposul subliniază libertatea persoanei de a se determina în raport cu logosul firii sale, având posibilitatea de a urma către Dumnezeu și de a-și găsi împlinirea în El sau de a acționa contrar naturii și voinței divine [63].

Nu vom insista prea mult pe terminologia complexă a Sf. Maxim, dar vom analiza conceptele de chip și asemănare. Ca și celelalte, acestea sunt armonios conectate cu distincțiile natură-ipostas și logos-tropos [81]. Chipul reprezintă natura sau ființa omului, care reflectă atributele și calitățile primite de la Dumnezeu. Asemănarea, în schimb, este legată de ipostas și implică participarea personală a omului pentru a fi împlinită.

Chipul este dat omului de la început și se referă la natura sa, în timp ce asemănarea este potențială și necesită participarea și alegerile omului pentru a fi dezvoltată. Asemănarea depinde de virtuți și de modul de viață al omului, în care se manifestă înfrânarea, curăția, blândețea, răbdarea, smerenia și altele. Omul este chemat să dobândească aceste virtuți prin participare și să devină un imitator al lui Dumnezeu [83].

Dobândirea asemănării necesită un efort constant și o viață statornică conformă virtuților, iar aceasta este rodul unei deveniri pe parcursul căreia sufletul caută să se conformeze cu Arhetipul dumnezeiesc. Însă, dobândirea asemănării nu este doar rezultatul străduințiilor omului, ci depinde

și de harul divin. Pe această cale, omul poate ajunge să se asemene lui Dumnezeu numai prin participarea sa activă și prin harul pe care-l primește [63].

Potrivit cu antropologia Sf. Maxim, omul a fost creat ca un microcosmos și ca un intermediar între Dumnezeu și creație. Căderea sa a fost rezultatul devierii de la mișcarea naturală către Dumnezeu și a orientării spre lumea senzorială, ducând la dezbinarea unității. Pentru a corecta această situație, Dumnezeu S-a întrupat, modificând astfel modul greșit de viață al omului și readucându-l în direcția sa autentică către Dumnezeu [81].

Prin cădere, omul a adoptat o formă de existență în conformitate cu alegerile sale. Pentru a se reîntoarce la forma sa autentică, omul trebuie să se desprindă voluntar de viața egoistă și să urmeze căile ascezei și contemplației. Prin har divin, omul este eliberat de corupție și de moarte, iar Dumnezeu îl ajută să crească spiritual în vederea îndumnezeirii. Scopul final al existenței este considerat de Sfântul Maxim a fi superior începutului. Sfinții primesc darul unei existențe veșnice și fericite, deoarece au dezvoltat în mod maxim potențialul ontologic prin efort și sfințenie. *Stasis*-ul reprezintă starea naturală a celui care a atins scopul final în Dumnezeu și nu mai poate cădea, fiind stabilizat prin har într-o mișcare fericită și în comuniune cu Dumnezeu. În perioada dintre geneză și *stasis*, sfinții și-au realizat într-o formă compleă potențialul spiritual prin trăirea deiformă, pregustând învierea [63; 64].

În discuția despre inteligența artificială, transumanism și limitele etice și tehnologice ale simulării inteligenței este util să explorăm conexiunile și diferențele prezentate în lucrarea lui Lars Thunberg, *Omul și Cosmosul în viziunea Sf. Maxim Mărturisitorul* [117].

Așa cum am văzut și mai sus, Sfântul Maxim Mărturisitorul oferă o viziune profundă asupra naturii divine, subliniind caracterul inefabil și transcendental al lui Dumnezeu, descris ca ființă „supranegrăită și „supra necuvântătoare," fiind infinit mai presus de orice înțelegere umană: *„fericita și sfânta Dumnezeire este după ființă supranegrăită și supra necuvântătoare și infinit mai presus de toată infinitatea, nelăsând celor de după ea nici urma cea mai subțire de cuprindere și nedând vreuneia dintre existențe nicio idee de cum sau în ce măsură este tot în ea și uniune și treime. Căci nu poate fi necreatul cuprins de creatură, nici infinitul înțeles de cele mărginite."* [117, pp. 40-42]. Această perspectivă teologică deschide calea către o reflecție profundă asupra caracterului sacru al realității și înțelegerii sale în contextul tehnologic actual.

Sfântul Maxim identifică trei niveluri ale teologiei, fapt deosebit de important în contextul conexiunilor cu viziunea filosofică transdisciplinară [81]:

Nivelul *iconomic*, care corespunde revelației istorice, unde cunoașterea divină este dezvăluită în mod specific oamenilor prin intermediul evenimentelor și a revelațiilor religioase. În

dezvoltarea inteligenței artificiale, tehnologiile ne oferă informații și date accesibile și comprehensibile în lumea noastră.

Nivelul revelației mistice, nivel care transcende cunoașterea istorică și ne deschide către o conexiune mai profundă cu divinul. În contextul transumanismului, se poate observa o aspirație similară de a depăși limitele umane și de a atinge o conștiență sau existență transcendentă prin intermediul tehnologiei.

Nivelul non-revelației reprezintă taina în sine, ce nu poate fi înțeleasă sau dezvăluită în totalitate. Acesta este un concept care poate fi adus în discuția despre inteligența artificială, unde există limite inerente ale cunoașterii și înțelegerii, precum și în domeniul transumanismului, unde anumite aspecte ale existenței umane și ale aspirațiilor sale rămân necunoscute și neexplorate. Nivelul non-revelației poate fi considerat ca fiind similar Realului din metodologia transdisciplinară.

Sfântul Maxim subliniază că Dumnezeu este atât de transcendent încât nu poate fi cuprins sau înțeles în întregime de către om [117]. Acest aspect poate fi adus în discuția despre inteligența artificială, unde simulările de inteligență nu pot recrea în totalitate complexitatea inteligenței umane sau a conștiinței umane.

Sfântul Maxim vorbește despre medierea omului ca microcosmos între cer și pământ, paradis și lumea oamenilor. Într-un mod similar, în abordarea transdisciplinară, subiectul poate fi privit ca un mediator între diferite niveluri ale realității, similar cu cele menționate de Maxim. Această mediere (terțul inclus, în limbaj transdisciplinar) poate avea loc între dualități precum creat și Necreat, inteligibil și sensibil, cer și pământ, paradis și lumea oamenilor, sau chiar între masculin și feminin [117, pp. 74-75].

Iată care sunt cele cinci distincții în contextul cărora are loc medierea între dualități:

Vârful ierarhiei, între creat și Necreat – deosebirea care separă întreaga creație de „natura necreată”, adică de Dumnezeu așa cum este El în Sine Însuși. Această mediere a fost săvârșită de Hristos în numele omului, exclusiv în virtutea unirii Sale ipostatice între firea creată și cea necreată.

În lumea lucrurilor create, între inteligibil și sensibil – statutul lucrurilor inteligibile nu este esențial diferit de cel al lucrurilor sensibile, pentru că ambele aparțin ordinii create, iar acest lucru este mai important decât diferența dintre ele.

În lumea sensibilă, între cer și pământ (Hristos între cei doi tâlhari) – unitatea de esență a naturii sensibile dincolo de orice separare.

Pe pământ, între paradis și lumea oamenilor – paradisul nu este niciodată o realitate transcendentă; Paradisul este o manifestare a coerenței întregii creații prin Logos.

În omenire, între bărbat și femeie (între masculin și feminin) - în Hristos nu există „nici bărbat, nici femeie” pentru că Hristos a supus forțele omului *logos*-ului naturii sale.

Conexiunile dintre transumanism, transdisciplinaritate și teologia creștină oferă o perspectivă fascinantă asupra căutării umane în legătură cu transcendența și înțelegerea de care este capabilă ființa umană. Totuși, trebuie să recunoaștem și diferențele esențiale dintre aspirațiile tehnologice și revelația divină, precum și limitele inerente ale cunoașterii umane în fața sacralității și complexității divine, așa cum Sfântul Maxim Mărturisitorul le-a subliniat în învățăturile sale.

Distincțiile dintre transumanism și teologia creștină sunt semnificative și oferă perspective diferite asupra umanității, existenței și aspirațiilor umane. Înțelegerea acestor diferențe este esențială pentru a aprecia modul în care aceste două domenii ar putea să intre în dialog sau să ofere perspective alternative.

Transumanismul susține ideea că natura umană poate fi îmbunătățită semnificativ prin intervenții tehnologice. Aceasta poate include extinderea duratei de viață, îmbunătățirea capacităților cognitive și chiar transformarea corpului uman. Teologia creștină pune accentul pe credința că omul este creat în imaginea lui Dumnezeu (*Imago Dei*) și că natura umană are o dimensiune sacră. Schimbările semnificative în natura umană ar putea ridica întrebări profunde despre identitatea și scopul uman.

Transumanismul își propune să dezvolte inteligența artificială și să creeze entități conștiente sau să încarce conștiința umană în medii digitale. Aceasta ridică întrebări despre ce înseamnă să fii conștient și să ai un suflet. Teologia creștină afirmă foarte clar existența unui suflet și a unei conștiințe umane unice, dar și posibilitatea unei relații cu Dumnezeu. Această viziune poate privi intervențiile tehnologice asupra conștiinței umane cu scepticism, având în vedere valoarea inerentă a fiecărei suflete.

Transumanismul încearcă să depășească limitele umane și să exploreze necunoscutul prin tehnologie. Cu toate acestea, așa cum arătăm în lucrarea de față, există limite inerente ale cunoașterii umane și ale tehnologiei. În schimb, teologia creștină recunoaște prezența tainelor divine care nu pot fi înțelese în totalitate de către om și care depășesc orice încercare de cunoaștere sau control uman [75].

Așadar, transumanismul și teologia creștină propun perspective distincte, uneori chiar divergente, cu privire la natura umană, conștiință, scopul vieții și relația cu divinul. Deși pot exista și unele puncte de întâlnire și dialog între aceste două sisteme de gândire, trebuie să recunoaștem și să acceptăm diferențele lor fundamentale.

3.4 Etica și responsabilitatea în era post-umană. Critica transdisciplinară a transumanismului ca reducționism unidimensional și analiza inegalităților sociale generate de tehnologie

Din perspectiva metodologiei transdisciplinare, transumanismul poate fi considerat în esență un reducționism unidimensional care adoptă o viziune simplificată și predominant materialistă asupra realității.

Transumanismul, așa cum am văzut deja, în încercarea sa de a îmbunătăți capacitățile umane prin tehnologie, se concentrează în mare parte pe aspectele materiale și tehnologice. Această viziune poate subestima și neglija dimensiuni importante ale ființei umane, cum ar fi conștiința, spiritualitatea și aspectele non-materiale ale existenței. Din perspectiva transdisciplinară, această concentrare excesivă asupra materialului conduce la o înțelegere limitată a complexității umane și a realității în ansamblul său.

Transdisciplinaritatea subliniază conexiunile și interdependențele dintre diferite nivele de realitate. În timp ce transumanismul se concentrează pe îmbunătățirea individului și a capacităților sale, perspectiva transdisciplinară subliniază interacțiunile și efectele dintre persoane, societate și mediul înconjurător plecând de la prima axioma transdisciplinară (axioma ontologică). Din această perspectivă transumanismul nu ia în considerare fluxul informațional dintre nivele de realitate corespunzătoare aspectului social, ecologic și cultural la care participă ființa umană.

Analizat din perspectiva axiomei doi a transdisciplinarității, transumanismul încearcă rezolvarea dilemelor existențiale, etice, sociale și tehnologice prin aplicarea unei logici a terțului exclus. Dar, așa cum reiese din aplicarea metodologiei transdisciplinare, complexitatea realității (așa cum este înțeleasă în modelul transdisciplinar) și interacțiunile dintre nivelurile de realitate nu pot fi abordate decât prin acceptarea unei logici a terțului inclus.

Perspectiva transdisciplinară susține că este important să includem și dimensiuni valorice și sensul vieții în discuția despre îmbunătățirea umană din perspectiva cele de-a treia axiome a transdisciplinarității (axioma epistemologică). Înțelegerea complexității umane și a realității în ansamblu nu poate avea loc în contextul unei viziuni reducționiste, în care modelul realității se bazează pe existența unui singur nivel de realitate.

Din perspectivă transdisciplinară reiese că o focalizare excesivă pe îmbunătățirea umană prin tehnologie poate ascunde și perpetua inegalitățile sociale, economice și politice existente, dat fiind modelul unilateral de realitate cu care operează toate tipurile de transumanism. Consecința este că, în opoziție totală cu scopurile declarate inițial, transumanismul va adânci decalajele sociale și va submina eforturile de a crea o societate mai egalitară și mai incluzivă.

Având în vedere critica transdisciplinară a transumanismului, odată cu explorarea și aplicarea tehnologiilor emergente, transumanismul ridică și numeroase dileme etice.

O critică majoră adusă transumanismului este legată de inegalitățile sociale și economice pe care această mișcare le poate amplifica. Îmbunătățirea capacităților umane prin tehnologie generează un decalaj între cei care își permit din punct de vedere financiar să acceseze aceste tehnologii și cei care nu au această posibilitate. Drept urmare, ar putea apărea o societate fragmentată, în care unii oameni devin superinteligenți și superputernici, în timp ce alții vor fi discriminați. Așadar, această dilemă etică ridică întrebări despre dreptatea socială și accesul egal la oportunități și poate adânci diviziunile sociale deja existente, contribuind la creșterea inegalității și marginalizării în societate.

Îmbunătățirea tehnologică a capacităților umane poate aduce schimbări radicale în identitatea noastră. Deci, este posibil ca utilizarea de implanturi cerebrale să modifice semnificativ modul în care gândim și percepem lumea. În acest sens, apar întrebări profunde despre esența umană și despre modul în care ne definim ca indivizi. Există îngrijorări justificate că aceste schimbări vor eroda, mai mult sau mai puțin remediabil, ceea ce este specific uman și că în acest mod vor fi afectate valorile și relațiile umane.

Inevitabil, în măsura în care ne bazăm tot mai mult pe tehnologie pentru a îmbunătăți capacitățile umane, apar și riscuri asociate. De exemplu, implanturile cerebrale sau îmbunătățirea genetică ar putea prezenta vulnerabilități cibernetice sau riscuri pentru sănătate. În plus, dependența de tehnologie poate face oamenii mai vulnerabili în cazul în care aceasta nu funcționează corect sau este compromisă într-un fel. Astfel că, din această perspectivă, se ridică preocupări cu privire la siguranța și protecția datelor personale, precum și la riscurile de dependență și pierderea autonomiei umane.

Transumanismul aduce în discuție și întrebări despre cine controlează și cine distribuie puterea într-o societate dependentă într-o mare măsură de tehnologie. Există temeri că aceste tehnologii avansate ar putea fi monopolizate de către elite, corporații sau chiar dictatori și organizații teroriste, ceea ce ar duce la o concentrare inegală a puterii și resurselor. Acest fapt ridică dileme etice legate de justiție, democratizare și democrația participativă într-un context transumanist.

O altă preocupare importantă în legătură cu transumanismul este legată de implicațiile pe termen lung ale utilizării tehnologiilor avansate pentru îmbunătățirea capacităților umane. Deși beneficiile potențiale par promițătoare, există, fără îndoială, riscuri și consecințe imprevizibile. O îmbunătățire semnificativă a capacităților cognitive poate conduce la dezechilibre sociale și psihologice sau chiar la dispariția unor aspecte valoroase ale umanității, cum ar fi creativitatea și

emoțiile. Prin urmare, ar trebui să se acorde o atenție deosebită evaluării și gestionării acestor riscuri și efecte secundare pe termen lung, dar acest lucru nu este ușor de implementat în practică prin inițiative la nivel legislativ.

Aceste tipuri de dileme etice în viața de zi cu zi în anumite contexte socio-umane. De pildă, utilizarea tehnologiei de editare genetică (cum ar fi CRISPR) pentru a îmbunătăți sau chiar modifica substanțial anumite caracteristici genetice ale copiilor ridică întrebări de natură etică. Această situație poate avea ca și consecință o dorință exacerbată a părinților de a interveni în ADN-ul copiilor lor pentru a le oferi avantaje sau trăsături dorite. Este evident că acest lucru va duce la o înrădăcinare a inegalităților genetice și la perpetuarea unor norme de frumusețe sau abilități selective.

La fel, folosirea implanturilor cerebrale sau a tehnologiilor de stimulare a creierului pentru a îmbunătăți funcțiile cognitive poate ridica întrebări despre ce înseamnă să fii uman și despre egalitatea șanselor. Transumanismul, deși promovează un optimism tehnologic maxim, nu analizează și acest tip de consecințe. Orice tip de progres tehnologic, generează probleme de adaptare începând cu structurile globale și până la nivel personal. Progresul exponențial necontrolat poate duce la schimbări de o complexitate și o instabilitate imposibil de controlat la nivel de societate. În acest sens, transumanismul ne arată, de fiecare dată, doar jumătatea plină a paharului.

Dezvoltarea tehnologiilor care încetinesc procesul de îmbătrânire și permit o viață mai lungă poate ridica dileme etice privind calitatea vieții, resursele limitate și impactul asupra mediului. Alegerea de a prelungi viața poate fi considerată o decizie personală, dar, inevitabil această decizie va avea consecințe sociale și economice semnificative, lucru la care filosofiile transumaniste nu par a se fi gândit îndeajuns.

Utilizarea tehnologiei pentru a combina elemente umane și non-umane poate ridica întrebări semnificative despre identitate și despre statutul moral al umanității. Crearea de ființe hibride, cum ar fi ființele cu abilități animale sau roboții cu „conștiință”, ridică dileme privind demnitatea și drepturile acestor entități. După ce tip de legi, într-o societate de tip transumanist, vor fi tratate astfel de ființe, după ce fel de cod etic vor fi considerate ca ființe mai mult sau mai puțin morale?

Dezvoltarea inteligenței artificiale „puternice” poate avea, așa cum am mai arătat, implicații etice majore legate de punerea în pericol a locurilor de muncă umane și crearea unui risc economic potențial pentru oameni. Dar, dincolo de aspectul economic, există dileme legate de responsabilitatea și controlul asupra acestor mașini inteligente și alegerile morale pe care acestea le pot face. Și acesta este un aspect la care curentele transumaniste nu par a avea soluții.

În contextul transumanismului, tehnologiile avansate implică colectarea și stocarea masivă a datelor personale și a informațiilor intime despre indivizi. Aceasta poate pune în pericol confidențialitatea, autonomia și drepturile individuale. Transumanismul nu pare a avea soluții nici la acest tip de dileme etice care privesc controlul și utilizarea adecvată a acestor informații și nici la posibilele riscuri de abuz și manipulare.

3.5 Concluzii

Această secțiune evidențiază o barieră fundamentală între capacitatea de calcul a mașinii și profunzimea existențială a omului, argumentând că simpla acumulare de informații nu va putea genera niciodată înțelepciune. În timp ce inteligența artificială excelează în învățarea bazată pe date și statistici, ea rămâne complet străină de procesul de înțelegere autentică, care la ființa umană presupune nu doar logică, ci și o experiență corporală, afectivă și o situare istorică în lume. Prin urmare, orice tentativă de a reduce mintea umană la un procesor de simboluri ignoră tocmai esența care ne definește, și anume, capacitatea de a atribui sens și valoare realității prin întreaga noastră ființă.

Analiza curentului transumanist dezvăluie o dorință de a forța limitele biologice prin tehnologie, propunând un model de „om-dumnezeu" (*Homo Deus*) construit pe o viziune reducționist-materialistă. Această perspectivă vede umanul ca pe un sistem deficitar ce necesită reparații sau „upgrade-uri" externe, riscând însă să sacrifice libertatea și unicitatea persoanei pe altarul eficienței tehnice. În opoziție, antropologia creștină propune conceptul de *Theosis,* unde desăvârșirea nu se obține prin adaosuri tehnologice sau redefinire genetică, ci prin restaurarea și sfințirea firii umane într-o relație personală cu divinitatea.

Nu în ultimul rând, din punct de vedere etic și social, progresul tehnologic exponențial, deși promite beneficii remarcabile, poartă în sine germenii unor inegalități profunde și ai unei posibile dezumanizări. Există riscul iminent ca tehnologia să devină o formă de idolatrie modernă, unde criteriile cantitative și controlul exercitat de elitele tehnologice să submineze demnitatea și valorile fundamentale. În acest context, viziunea transdisciplinară și cea teologică ne reamintesc că omul nu este un simplu obiect fragmentat, ci un microcosmos unitar, a cărui împlinire nu stă în expansiunea orizontală a puterii, ci în profunzimea verticală a spiritului.

CONCLUZII GENERALE ȘI RECOMANDĂRI

1. Ipoteza centrală a cercetării a postulat că inteligența artificială constituie o simulare funcțional-algoritmică a cogniției, incapabilă să depășească pragul noetic al înțelegerii autentice, și că depășirea reducționismului tehnocentrist reprezintă condiția fundamentală și indispensabilă a unei evaluări corecte a acestui decalaj. Investigația întreprinsă a confirmat această ipoteză, aducând totodată o nuanță importantă. Decalajul dintre inteligența artificială și cea umană nu este doar funcțional sau tehnic, ci unul de natură ontologică structurală. Mașina nu eșuează în a înțelege din lipsă de putere de procesare. Ea eșuează pentru că înțelegerea presupune corporalitate, intenționalitate și participare la niveluri de realitate pe care nici o arhitectură computațională nu le poate accesa. Metodologia transdisciplinară nu a fost doar un instrument de analiză, ci însăși condiția de posibilitate a vizibilității acestei distincții.
2. Metodologia filosofică transdisciplinară este pe deplin aplicabilă și justificată în analiza inteligenței artificiale și a conexiunilor ei cu celelalte discipline. Ea servește drept punte între diferite domenii de studiu și inteligența artificială, oferind un cadru coerent pentru explorarea și înțelegerea complexității și a implicațiilor inteligenței artificiale în societate, etică, tehnologie și știință. Transdisciplinaritatea facilitează o abordare integrată, permițând cercetătorilor și practicienilor să treacă dincolo de limitele disciplinare tradiționale și să gândească creativ și inovator.

   În lumina acestei concluzii, intensificarea colaborărilor pluri/inter/transdisciplinare va avea un impact prolific, pentru că astfel ne vom asigura că inteligența artificială este dezvoltată și utilizată într-un mod atât etic, cât și social responsabil. Există o nevoie imperativă de a integra diverse perspective în abordarea problemelor etice, tehnice și sociale complexe asociate cu inteligența artificială. Pentru acest lucru este necesară fuziunea de orizonturi de expertiză din filosofie, științe sociale, științe naturale și alte domenii pentru a se putea construi un cadru robust de cunoaștere și practică.

   De asemenea, considerăm că este necesară promovarea educației transdisciplinare în rândul studenților și al profesioniștilor, cultivând astfel o generație de lideri și inovatori care să fie echipați pentru a naviga în complexitatea și ambiguitatea inerente unei lumi în care inteligența artificială este din ce în ce mai prezentă. În plus, guvernele, organizațiile și instituțiile ar trebui să fie încurajate să adopte politici care susțin și facilitează cercetarea și dezvoltarea transdisciplinară în domeniul inteligenței artificiale, asigurându-se astfel că progresul tehnologic merge mână în mână cu considerațiile etice și sociale.

3. Luând în considerație schema transdisciplinară a ființei umane, putem conchide că:
   a. așa cum arată Penrose în lucrările sale [94; 95; 96] sunt dovezi foarte puternice că există un substrat cuantic sau chiar subcuantic de care inteligența umană este dependentă. Despre prezența manifestării inteligenței umane la nivelul macrofizic (există schimb energetic și replicare) și la nivelul viului (neuronul fiind o celulă nervoasă) credem că nu mai este nevoie de nici un fel de demonstrație;
   b. după cum știm încă din antichitate (începând cu filosofii greci), inteligența manifestată ca *empeiria*, ca *dianoia* și ca *nous*, se constituie în trei componente fundamentale ale ființei umane. Așa cum am arătat în decursul lucrării, inteligența manifestată doar pe unul dintre aceste niveluri nu este inteligență în sens uman, ci doar o proiecție unidimensională a unor aspecte specifice;
   c. de asemenea, o bogată literatură filosofică și teologică ne arată că ființa umană nu poate fi definită fără nivelul logosului, nivel care conferă semnificație tuturor celorlalte niveluri. Fără acest nivel nici nu putem vorbi ca atare despre o ființă umană. Toate aceste opt niveluri ale schemei transdisciplinare reprezentativă pentru ființa umană constituie un tot, fără a se confunda între ele (fără separare strictă și fără contopire). Nici unul dintre aceste niveluri nu poate exista în absența celorlalte niveluri. Nu putem vorbi despre înțelepciune fără înțelegere. Nu putem vorbi despre înțelegere fără învățare și despre învățare fără raportare la nivelul empiric. Fără aceste niveluri care se susțin reciproc, nu am putea avea o înțelegere comprehensivă a ființei umane. Fiecare nivel adaugă o componentă esențială la ceea ce suntem, iar absența oricărui nivel ar slăbi înțelegerea noastră asupra a ceea ce există. Cu toate că aceste niveluri nu se confundă între ele și nu se contopesc într-unul singur, ele funcționează împreună pentru a crea un întreg coerent. Așadar, înțelegerea noastră despre ființa umană este un mozaic complex, în care fiecare nivel adaugă un strat de semnificație, făcând din ființa umană un subiect fascinant, mereu deschis explorării și înțelegerii mai profunde;
   d. între aceste niveluri există un flux de informație prin care fiecare nivel contribuie la susținerea celorlalte niveluri. Deci, nivelul logosului transmite semnificație tuturor celorlalte niveluri, așa cum nivelul empiric structurează această informație până la nivelul subcuantic. În consecință, putem spune că ființa umană este caracterizată prin spirit, conștiință de sine, capacitate de analiză sinteză și decizie, poate iubi sau urî, are diferite afecte, are funcții metabolice, corporalitate (materialitate) și, finalmente, este constituită din particule cuantice și subcuantice.

4. Analiza inteligenței artificiale din perspectiva aceleiași scheme transdisciplinare ne duce inevitabil la o serie de limitări substanțiale. Astfel, observăm că:
   a. Inteligența artificială se limitează în principal la nivelul empiric și dianoetic, niveluri care corespund cu experiența senzorială și rațiunea instrumentală, respectiv. În termeni simpli, inteligența artificială poate „înțelege" și „procesa" informații bazate pe date și poate executa sarcini complexe prin algoritmi sofisticați. Însă aici, în acest spațiu calculat și mecanic, începe și se sfârșește întregul său domeniu de competență.
   b. nu putem spune că inteligența artificială posedă *nous* sau inteligență intuitivă, nici *logos* sau capacitatea de a conferi sens și semnificație. Fără aceste niveluri, inteligența artificială rămâne o unealtă extrem de avansată tehnologic, dar în cele din urmă doar o unealtă. Ea nu posedă nici capacitatea de a genera semnificație autentică sau de a participa la crearea unei realități conștiente și înțelese în mod autentic;
   c. în ciuda avansului tehnologic fulminant și a aparentei „înțelepciuni" pe care un sistem cu inteligență artificială o poate manifesta în anumite sarcini specifice, absența unui substrat conștient trasează o limită ontologic inerentă. Inteligența artificială nu are capacitatea de a iubi, de a genera valori, de a aspira la transcendență sau de a crea semnificație într-un mod care să meargă dincolo de calcule și algoritmi. În timp ce inteligența umană este capabilă să gândească și să creeze în afara „cutiilor negre" algoritmice, să simtă și să dea naștere unor valori și sensuri ce depășesc rațiunea instrumentală, inteligența artificială este legată implacabil de convergența și precizia datelor și a algoritmilor pe care îi folosește sau cu care este antrenată;
   d. intrând în complexitatea și misterul conștiinței umane, putem înțelege că aceasta se derulează pe planuri de existență ce nu pot fi cuantificate sau simulate complet de către mașini. Experiențele mistice, spiritualitatea, și profunzimea emoțională sunt manifestări care până acum au rămas inaccesibile și incomprehensibile pentru inteligența artificială și, ca o consecință a modelului transdisciplinar, credeam că așa va și rămâne;
   e. în timp ce inteligența umană continuă să exploreze universul, să-și adâncească înțelegerea și să caute răspunsuri la întrebări fundamentale, inteligența artificială nu rămâne decât un instrument, un mijloc prin care putem explora și înțelege mai bine lumea înconjurătoare și propria noastră condiție umană. Este important să ne amintim că, deși inteligența artificială poate fi un aliat de neprețuit în multe aspecte ale existenței noastre, ea nu își poate depăși creatorii din punct de vedere al profunzimii gândirii, înțelepciunii și bogăției existențiale.

Așadar, în pofida realizărilor impresionante ale inteligenței artificiale, există o esență intrinsecă și insesizabilă a inteligenței umane care va continua să prevaleze și să transceadă capacitățile mașinilor. Umanitatea își va păstra locul distinct și prețios în această lume, explorând misterele existenței cu un tip de înțelegere care depășește tipul de raționament algoritmic și calitatea datelor pe care le pot colecta și prelucra sistemele digitale.

Recomandările formulate în continuare sunt structurate în jurul a două axe complementare: valorificarea rezultatelor cercetării în plan aplicativ și trasarea unor direcții viabile pentru aprofundarea investigației în domeniu.

- Se recomandă integrarea metodologiei transdisciplinare în cercetarea dedicată inteligenței artificiale, prin crearea unor programe academice interdepartamentale care să reunească filosofia, informatica și etica într-un cadru epistemologic coerent.
- Curriculum-ul universitar din domeniul tehnologiei este recomandat să includă module de filosofia minții și etică tehnologică.
- Se recomandă ca factorii de decizie politică să opereze explicit cu distincția ontologică dintre algoritm și persoana umană în elaborarea cadrelor normative, prevenind astfel atribuirea improprie de responsabilitate morală sistemelor artificiale.
- Protocoalele experimentale comparative, de tipul celor realizate în prezenta lucrare, trebuie extinse și sistematizate, pentru a testa în mod riguros și constant capacitatea modelelor actuale de a accesa dimensiunea semantică și contextuală a înțelegerii.
- Este necesară inițierea unui dialog academic structurat între perspectivele transumaniste și antropologia teistă, cu scopul de a elabora un cadru axiologic capabil să orienteze etic dezvoltarea tehnologiilor de augmentare cognitivă și biologică.
- Riscurile sociale generate de accesul inegal la inteligența artificială trebuie evaluate printr-o cercetare transdisciplinară care să implice economiști, sociologi și filosofi, în vederea fundamentării unor politici publice coerente.
- Modelul transdisciplinar al ființei umane propus în lucrare merită aprofundat prin integrarea contribuțiilor din neuroștiințe, fenomenologie, fizica cuantică, teologia filosofică și psihologie, pentru a consolida o teorie comprehensivă a inteligenței care să depășească reducționismul computațional.

BIBLIOGRAFIE

Literatură în limba română

1. ***, *BIBLIA (NTR), Noua traducere în limba română*, Oradea: Scriptum, 2021, 1134 p. ISBN 9786068279732
2. ARISTOTEL, *Etica nicomahică*, trad. Stella Petecel, București: Editura Științifică și enciclopedică, 1988 (ediția a II-a, IRI, 1998, 393 p. ISBN 973-98377-4-3)
3. ARISTOTEL, *Politica*, trad. Alexander Baumgarten, București: Editura IRI, 2001, 522 p. ISBN 973-9437-67-4
4. ARISTOTEL, *Despre suflet*, trad. Alexander Baumgarten, București: Editura Humanitas, 2005, 412 p. ISBN 973-50-125-9
5. ARISTOTEL, *Poetica*, trad. D.M. Pippidi, București: Editura Academiei RPR, 1964, 256 p.
6. ARISTOTEL, *Organon (I-II)*, trad. Mircea Florian, București: Editura IRI, 1997, 480 p. 670 p. ISBN 973-97627-4-3, 973-97627-5-1
7. ARISTOTEL, *Metafizica*, trad. Ștefan Bezdechi, București: Editura IRI, 1999, 614 p. ISBN 973-97229-8-9
8. BACON, F., *Noul Organon*, trad. Bogdan Florea, București: Antet, 2011, 223 p. ISBN 9789736364495
9. BARROW, John, *Despre imposibilitate. Limitele științei și știința limitelor*, București: Editura Tehnică, 1999, trad. Mihai Popescu, 256p. ISBN 973-31-1386-7
10. BERDIAEV, N., *Despre sclavia și libertatea omului*, Oradea: Antaios, 2000, 267 p. ISBN 973-99294-5-1
11. BERDIAEV, N., *Despre menirea omului*, Oradea: Aion, 2004, 389 p. ISBN 973-99943-5-0
12. BIANCHI, E. et al., *Oameni și animale. O perspectivă teologică și etică*, trad. Raluca Lazarovici Vereș, Oradea: Ratio et Revelatio, 2013, 160 p. ISBN 978-606-93522-9-8
13. BOBOC, J., *Transumanismul decriptat: metamorfoza navei lui Tezeu*, trad. Valeriu Manuel, Iași: Doxologia, 2020, 477 p. ISBN 978-606-666-948-1
14. BOSTROM, N., *Superinteligența. Direcții, pericole, strategii*, București: Litera, 2016, 489 p. ISBN 978-606-33-0281-7
15. BROGAARD, B., MARLOW, K., *Mintea supraumană sau despre genialitate ca însușire firească a creierului nostru*, trad. Cristina Jinga, București: Humanitas, 2018, 278 p. ISBN 978-973-50-5835-7

16. BRUNEL, P., *Miturile secolului XX (I și II)*, București: Univers, 2003, 416 p. ISBN 973-34-0872-7
17. CASTI, J., *Paradigme pierdute. Tatonând misterele neelucidate ale științei moderne*, București: Pergament, 2007, trad. Valentin Stoica, 624 p. ISBN 978-973-85481-3-8
18. CHURCHLAND, P. S., *Neurofilosofia*, trad. Teodora Lucescu, Iași: Editura Cartea Românească Educațional, 2020, 695 p. ISBN 978-606-9088-93-7
19. CLÉMENT E., DEMONQUE C., HANSEN-LØVE L., KAHN P., *Filosofia de la A la Z*, trad. Magdalena Cojocea-Mărculescu, Aurelian Cojocea, Bucureşti: ALL Educațional, 1999. 576 p. ISBN 973-684-030-1
20. CRAINIC, N., *Cursurile de mistică*, Sibiu: Deisis, 2010, 740 p. ISBN 9789737859631
21. DA COSTA, N., *Logici clasice și neclasice*, trad. Ilie Gyurcsic, București: Editura Tehnică, 328 p. ISBN 973-31-2219-X
22. DANCY J., SOSA E., *Dicționar de filosofia cunoașterii.* Vol. I, trad. Gheorghe Ștefanov, Corina Matei, Anabela Zagura, Dan Bărbulescu, București: Editura Trei, 1999, 424 p. ISBN 973-9419-10-0
23. DANCY J., SOSA E., *Dicționar de filosofia cunoașterii.* Vol. II, trad. Gheorghe Ștefanov, Corina Matei, Anabela Zagura, Dan Bărbulescu, București: Editura Trei, 1999, 455 p. ISBN 973-9419-78-X
24. DE SAUSSURE, F., *Curs de lingvistica generala*, trad. Irina Izverna, Iași: Polirom, 1998, 400 p. ISBN 973-683-157-4
25. DERRIDA, J., *Despre gramatologie*, trad. de Bogdan Ghiu, Cluj-Napoca: Editura Tact, 2009, 340 p. ISBN 9789738667068
26. DESCARTES, R., *Discurs despre metoda de a ne conduce bine rațiunea și a căuta adevărul în științe*, trad. Daniela Rovența-Frumușani și Alexandru Boboc, București: Editura Academiei Române, 1990, 200 p.
27. DROZDA-SENKOWSKA, E. (coord.), *Capcanele raționamentului*, Iași: Polirom, 1998, 160 p., ISBN 973-683-057-8
28. DIONISIE AREOPAGITUL, *Despre ierarhia cerească*, trad. Marilena Vlad, Iași: Polirom, 2021, 344 p. ISBN 9789734687305
29. **DUCA, M., *Divinizarea tehnologiei ca proces de dezumanizare a umanității*, Studia Universitatis Moldaviae , 2020, nr.10(140) Seria "Științe umanistice" ISSN 1811-2668 | E-ISSN 2345-1009, pp.68-73**

30. **DUCA, M., *Artificial intelligence. Myth and reality*, Annals of Philosophy, Social and Human Disciplines (APSHUS), 2023, vol. I, SSN: 2069 – 4008 (Print) / ISSN: 2069 – 4016 (Online), pp. 71-84**
31. DUMITRAȘ, A., *Proiectarea rețelelor neurale artificiale*, București: Casa Editorială Odeon, 1997, 372 p. ISBN 973-9008-75-5
32. DUMITRESCU, D., *Principiile inteligenței artificiale*, Cluj-Napoca: Editura Albastră, 1999, 289 p. ISBN 973-9443-06-0
33. ELIADE, M., *Aspecte ale mitului*, București: Univers, 1978, 195 p.
34. ELIOT, T. S., *Opere poetice*, trad. Mircea Ivănescu, Şerban Foarţă, București: Humanitas, 2011, 432 p. ISBN 978-973-689-443-5
35. ENESCU, G., *Dicționar de logică*, București: Editura Științifică și Enciclopedică, 1985, 384 p.
36. ENESCU, G., *Paradoxuri. Sofisme. Aporii,* București: Editura Tehnică, 2003, 468 p. ISBN 973-31-2164-9
37. ENESCU, G., *Tratat de logică*, București: Editura Lider, fără an, 352 p. ISBN 973-97836-5-1
38. ENESCU, G., *Fundamentele logice ale gândirii*, București: Editura științifică și enciclopedică, 1980, 320 p.
39. ENGELHARDT, T. H. jr., *Fundamentele bioeticii creştine: perspectiva ortodoxă*, trad. Mihail Neamțu et al, Sibiu: Deisis, 2005, 507 p. ISBN 973-7859-08-1
40. EPICTET, *Opera omnia: conversații, manual, fragmente, gnomologion*, trad. Cristian Bejan, Iași: Polirom, 2022, 632 p. ISBN 978-973-46-8759-6
41. FESER, E., *De la Aristotel la John Searle și înapoi*, trad. Cornelia Dumitru și Steven Nemeș, București: Editura Eikon, 2017, 172 p. ISBN 978-606-711-677-9
42. FLEW, A., *Dicționar de filozofie și logică*, Ediția a II-a, trad.: D. Stoianovici, București: Humanitas, 1999. 376 p. ISBN 973-28-0966-3
43. GADAMER, H.-G., *Adevăr și metodă*, trad. Gabriel Cercel, Larisa Dumitru, Gabriel Kohn, Călin Petcana, București: Teora, 2001, 735 p. ISBN 973-601-684-6
44. GAVRILUȚĂ, N., *Noile religii seculare: corectitudinea politică, tehnologiile viitorului și transumanismul*, Iași: Polirom, 2018, 221 p. ISBN 978-973-46-7203-5
45. GHEORGHIU, D. *Logică generală I*, București: Editura Fundației România de mâine, 2004, 208 p. ISBN 973-725-059-1
46. GHEORGHIU, D. *Logică generală II*, București: Editura Fundației România de mâine, 2001, 164 p. ISBN 973-582-372-1

47. GOLEMAN, D., *Inteligența emoțională*, trad. Irina-Margareta Nistor, București: Curtea Veche, 2001, 424 p. ISBN 973-8120-67-5
48. GRAZIANO, M. S. A., *Conștiința dintr-o nouă perspectivă: o teorie științifică a experienței subiective*, trad. Vlad Vedeanu, București: Humanitas, 2021, 243 p. ISBN 978-973-50-6920-9
49. HARARI, Y. N., *Homo deus: scurtă istorie a viitorului*, București: Polirom, 2018, ISBN 978-973-46-7199-1
50. HAWKINS, J., *100 de creieri: o nouă teorie a inteligenței*, trad. Dan Crăciun, București: Publica, 2020, 343 p. ISBN 978-606-722-519-8
51. HEIDEGGER, M., *Ființă și timp*, trad. Gabriel Liiceanu și Catalin Cioaba, București, Humanitas, 2003, 673 p. ISBN 978-973-50-3459-7
52. HUME, D. *Cercetare asupra intelectului omenesc*, trat. Mircea Flonta și Adrian-Paul Iliescu, Oradea: Editura Ratio et Revelatio, 2014, 272 p. ISBN 978-606-93602-8-6
53. HUSSERL, E., *Cercetări logice I*, trad. Bogdan Olaru, București: Humanitas, 2007, 364 p. ISBN 978-973-50-1743-9
54. IONIȚĂ, C., *Incomensurabilitatea paradigmelor - O perspectivă cognitivistă*, Universitatea din București - Facultatea de Filosofie, teză de doctorat, 2012
55. IOSIF, A., *Modelarea sistemului biotehnic om-mașină*, București: Editura Printech, 2022, 148 p. ISBN 978-606-23-1318-0
56. IOSIF, A., *Inteligența artificială și mintea umană. Aspecte inter- și transdisciplinare, În: Perspective românești asupra științei și teologiei*, București: Curtea Veche, 2006. p. 431-458, 528 p. ISBN 973-669-284-1
57. KANT, I., *Critica rațiunii pure*, trad. Nicolae Bagdasar și Elena Moisuc, București: Univers Enciclopedic Gold, 2009, 610 p. ISBN 978-606-92266-2-9
58. KANT, I., *Critica rațiunii practice*, trad. Traian Brăileanu, București: Paideia, 2020, 164 p. ISBN 978-606-748-360-4
59. KANT, I., *Critica puterii de judecare*, trad. Traian Brăileanu, București: Paideia, 2020, 324 p. ISBN 978-606-748-365-9
60. KOESTLER A., *Lunaticii. Evoluția concepției despre Univers de la Pitagora la Newton*, București, Humanitas, 1995. 468 p. ISBN 973-28-0521-8
61. KURZWEIL, R., *Epoca mașinilor spirituale: când computerele depășesc inteligența umană*, trad. Bogdan Chircea, Pitești: Paralele 45, 2012, 523 p. ISBN 978-973-47-1445-2
62. LA ROCHEFOUCAULD, *Maxime*, trad. Cătălin Avramescu și Ștefan Vianu, Iași: Polirom, 2019, 232 p. ISBN 978-973-46-7825-9

63. LARCHET, J.-C., *Îndumnezeirea omului la Sfântul Maxim Mărturisitorul*, trad. Marinela Bojin, București: Basilica, 2019, 840 p. ISBN 978-606-29-0295-7
64. LARCHET, J.-C., *Sfântul Maxim Mărturisitorul: o introducere*, trad. Marinela Bojin, Iași: Doxologia, 2013, 397 p. ISBN 978-606-666-071-6
65. LARCHET, J,-C., *Semnificația trupului în Ortodoxie*, București: Editura Basilica, 2010, 120 p. ISBN 9786068141336
66. LARSON, E. J., *Mitul inteligenței artificiale: de ce computerele nu pot gândi la fel ca noi*, Iași: Polirom, 2022, 303 p. ISBN 978-973-46-8904-0
67. LECOURT, D., (coord), *Dicționar de istoria și filosofia științelor*, trad. Laurențiu Zoicaș (coord.), Iași: Polirom, 2005, 1446 p. ISBN 973-681-734-2
68. LEE, Kai-Fu, *Superputerile inteligenței artificiale: China, Silicon Valley și noua ordine mondială,* București: 2021, 351 p. ISBN 978-606-088-014-1
69. LEIBNIZ, G., *Opere filozofice I*, trad. Constantin Floru, București: Editura Științifică și Enciclopedică, 1977, 665 p.
70. LENNOX, J., *Groparul lui Dumnezeu*, București: GBV România, 2012, 294 p. ISBN:978-606-8441-00-9
71. LEWIS, C. S., *Desființarea omului*, trad. Petruța-Oana Naidut, București: Humanitas, 2023, 100 p. ISBN 978-973-50-7666-5
72. LEWIS, C. S., *Despre minuni*, trad. Sorin Mărculescu, București: Humanitas, 2007, 285 p. ISBN 978-973-50-1465-0
73. LOCKE, J. *Eseu asupra intelectului omenesc (I-II)*, trad. Teodor Voiculescu, București: Editura Științifică, 1961, I - 454 p. II - 422 p.
74. LOSSKY, V., *Introducere în teologia ortodoxă*, trad. Lidia Rus, Remus Rus, București: Sophia, 2014, 216 p. ISBN 978-973-136-439-1
75. LOSSKY, V., *Teologia dogmatică*, trad. Cristian Galeriu, București: Anastasia, 2014, 228 p. ISBN 978-973-1830-24-7
76. LOSSKY, V., *Teologia mistică a Bisericii de Răsărit*, trad. Vasile Răducă, București: Anastasia, fără an, 308 p. 973-95777-3-3
77. LOUTH, A., *Introducere în teologia ortodoxă*, trad. Dragoș Mîrșanu, Iași: Doxologia, 2014, 280 p. ISBN 978-606-666-311-3
78. MALRAUX, A., *Condiția umană*, București: Ed. RAO, 2006. 378 p. ISBN 978-973-576-974-1
79. MARCUS AURELIUS, *Gânduri către sine însuși*, trad. Cristian Bejan, București: Humanitas, 2022, 188p. ISBN 978-973-50-7627-6

80. MARCUS, S., *Cât de umane pot fi limbajele artificiale*, în *Sisteme de inteligență artificială*, București: Editura Academiei Române, 1991 (pp. 170-180)
81. MAXIM, Sf., *Ambigua*, în Părinți și Scriitori Bisericești 80, București: Editura IBMBOR, 1983, 372 p.
82. MAXIM, Sf., *Epistole*, în Părinți și Scriitori Bisericești 82, București: Editura IBMBOR, 1990, pp. 47-149
83. MAXIM, Sf., în *Filocalia*, vol. II, trad. Dumitru Stăniloae, București: Humanitas, 1999, 277 p. ISBN 973-28-0937-X
84. MAXIM, Sf., în *Filocalia*, vol. III, trad. Dumitru Stăniloae, București: Humanitas, 1999, 422 p. ISBN 973-28-0938-8
85. MEMELIS, G., *în: Perspective românești asupra științei și teologiei*, București: Curtea Veche, 2006. p. 181-204, 528 p. ISBN 973-669-284-1
86. MEMELIS, G., IOSIF, A., RĂILEANU, D., *Realitatea transdisciplinară: o fuziune de orizonturi ale teologiei, științei și filosofiei,* București: Curtea Veche, 2010, p. 280 ISBN 978-973-669-919-1
87. MERLEAU-PONTY, M., *Fenomenologia percepției*, trad. Ilieș Câmpeanu și Georgian Vătăjelu, Oradea: Aion, 1999, 536 p. ISBN 973-99943-3-4
88. MIROIU, A., *Lumea lui Anaxagora*, Iași: Polirom, 2017, 334 p. ISBN 978-973-46-6774-1
89. NAGEL, T., *Oare ce înseamnă toate astea?*, trad. Isabela Iețcu-Preoteasa, București: All, 2004, 88 p. ISBN 973-571-408-6
90. NAGEL, T., *Veșnice întrebări*, trad. Isabela Germina Chiroiu, București: All, 2004, 320 p. ISBN 973-571-525-2
91. NICOLESCU, B., *Noi, Particula și Lumea*, Iași: Polirom, 2002, 312 p. ISBN 973-681-010-0
92. NICOLESCU, B., *Transdisciplinaritatea. Manifest,* Iași: Junimea, 2007, 177 p. ISBN 973-37-1223-7
93. NIETZSCHE, F., *Așa grăit-a Zarathustra*, București: Edinter, 1991, 368 p. ISBN 973-95268-0-2
94. PENROSE, R., *Incertitudinile rațiunii. Umbrele minții*, București: Editura Tehnică, 1999, trad. Dana Jalobeanu, 536 p. ISBN 973-31-1290-9
95. PENROSE, R., *Mintea noastră… cea de toate zilele. Despre gândire, fizică, și calculatoare*, București: Editura Tehnică, 1996, trad. Cornelia și Mircea Rusu, 504 p. ISBN 973-31-0969-X

96. PENROSE, R. ș.a., *Mintea omenească între clasic și cuantic*, București: Editura Tehnică, 1999, trad. Cornelia și Mircea Rusu, 168 p. ISBN 973-31-1364-6
97. PETERS, F. E., *Termenii filozofiei grecești*, București: Humanitas, 1997, 326 p. 388 p. ISBN 973-28-0750-4
98. PIAGET, J., CHOMSKY, N., *Teorii ale limbajului. Teorii ale învățării*, București: Editura Politică, 1988, 632 p.
99. PLATON, *Theaitetos*, *în Opere VI*, trad. Marian Ciucă, București: Editura Științifică și Enciclopedică, 1989, pp. 181-307
100. PLATON, *Republica*, *în Opere V*, trad. Andrei Cornea, București: Editura Științifică și Enciclopedică, 1986, 504 p.
101. PLATON, *Legile*, trad. E. Bezdechi, București: IRI, 1995, 397 p. ISBN 973-97229-2-X
102. PLATON, *Phaidon, în Opere IV*, trad. Petru Creția, București: Editura Științifică și Enciclopedică, 1983, pp. 51-141
103. RAMACHANDRAN, V. S., *Ce ne spune creierul: dezvăluiri extraordinare despre misterele creierului uman*, trad. Adela Crăciun, București: Litera, 2019, 495 p. ISBN 978-606-33-4017–8
104. RICOEUR, P., *Sinele ca un altul*, trad. Alina-Daniela Marinescu, Paul Marinescu, București: Editura Spandugino, 2016, 455 p. ISBN 978-606-8401-91-1
105. RUSSELL, B., *Problemele filosofiei*, trad. Mihai Ganea, București: All, 2004, 152 p. ISBN 973-571-409-4
106. SCHMEMANN, A., *Euharistia. Taina Împărăției*, trad. Boris Răduleanu, București: Sophia, 2012, 296 p. ISBN 9789731363240
107. SEARLE, J. R., *Mintea: Scurtă introducere în filosofia minții*, București: Herald, 2013. 335 p. ISBN 978-973-111-553-5
108. SENECA, *Scrisori către Luciliu,* trad. Gheorghe Guțu, București: Humanitas, 2020, 544 p. ISBN 978-973-50-6374-0
109. SHAKESPEARE, *Hamlet, în Opere complete, Vol. 5,* trad. Leon Levițchi și Dan Duțescu, București: Editura Univers, 1986. p. 362
110. SPIDLIK, T. *Spiritualitatea răsăritului creștin, I. Manual sistematic*, trad. Ioan Ică jr., Sibiu: Editura Deisis, 1997, 476 p. ISBN 973-97750-9-8
111. SPINOZA, B., *Tratatul despre îndreptarea intelectului*, trad. Alexandru Posescu, București: Editura Științifică și Enciclopedică, 1979, 60 p.
112. SPINOZA, B., *Etica*, trad. Alexandru Posescu, București: Editura Științifică și Enciclopedică, 1981, 278 p.

113. SPITZER, M., *Demența digitală: cum ne tulbură mintea noile tehnologii*, trad. Dana Verescu, București: Humanitas, 2020, 326 p. ISBN 978-973-50-6713-7
114. TAMAȘ, C.-D., *100 probleme de logică*, Iași: Editura Taida, fără an, 76 p., ISBN 973-7669-17-7
115. TEGMARK, M., *Viața 3.0. Omul în epoca inteligenței artificiale*, Humanitas, București, 2019, 392 p. ISBN 978-973-50-6402-0
116. TOMA, M., *Raționamentul logic*, București: Universul juridic, 2019, 586 p. 978-606-39-0468-4
117. THUNBERG, L., *Omul și Cosmosul în viziunea Sf. Maxim Mărturisitorul*, trad. Remus Rus, București: Editura I.B.M.B.O.R., 1999, 184 p. ISBN 973-9332-11-0
118. VERGEZ, A., HUISMAN, D., *Curs de filosofie*, trad. Alexandru Vasile Drăgan, București: Humanitas, 1995, 398 p. ISBN 973-28-0521-8
119. WAAL, F. de, *Bonobo și ateul. În căutarea umanismului printre primate*, trad. Ioana Miruna Voiculescu, București: Humanitas, 2017, 324 p. 978-973-50-5848-7
120. WAAL, F. de, *Mama. Ultima îmbrățișare. Ce ne spun emoțiile animalelor despre noi înșine*, trad. Carmen Strungaru, București: Humanitas, 2020, 372 p. ISBN 978-973-50-6885-1
121. WIENER, N., *Dumnezeu și golemul: comentariu asupra câtorva probleme în care cibernetica intră în contradicție cu religia*, trad. Tudor Călin Zarojanu, București: Humanitas, 2019, 79 p. ISBN 978-973-50-9344-3
122. WITTGENSTEIN, L., *Tractatus Logico-Philosophicus*, trad. Mircea Flonta, Mircea Dumitru, București: Humanitas, 2021, 224 p. ISBN 978-973-50-3387-3
123. WITTGENSTEIN, L., *Cercetări filozofice*, trad. Mircea Dumitru, Mircea Flonta, Adrian-Paul Iliescu, București: Humanitas, 2013, 432 p. ISBN 978-973-50-4228-8-X
124. WRIGHT, G. H. von, *Explicație și înțelegere*, trad. Mihai Vasile, București: Humanitas, 1995, 236 p. ISBN 973-28-0534-X
125. WRIGHT, N. T., *Surprins de speranță*, trad. Florentina Galiger, Oradea: Metanoia, 2016, 384 p. ISBN 978-606-8432-95-3

Literatură în limba engleză

126. BHARGAVA, C., SHARMA, P., *Artificial Intelligence. Fundamentals and Applications*, CRC Press, 2022, p.256
127. CANTWELL SMITH, B., *The Promise of Artificial Intelligence Reckoning and Judgment*- Brian Cantwell Smith, Cambridge: MIT Press, 2019, 158 p. ISBN 9780262043045

128. CARTER, M., *Minds and Computers The Philosophy of Artificial Intelligence*, Edinburgh: University Press, 222 p. ISBN 978-0-7486-2098-2
129. CHOMSKY, N., *Language and Mind*, Cambridge: University Press, 2006, 190 p. ISBN 978-0-521-67493-5
130. CLARK, A., *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, 2010, 318 p. ISBN 978-0199773688
131. CLARK, A., *Being There: Putting Brain, Body, and World Together Again*, Bradford Books, 1998, 292 p. ISBN 978-0262531566
132. CLOUSER, R., *A Blueprint for a Non-Reductionist Theory of Reality, In: Transdisciplinarity in Science and Religion,* vol. 4, București, Curtea Veche, 2008. pp. 27-58
133. DEISENROTH, M. P., ALDO FAISAL, A., ONG, C. S., *Mathematics for Machine Learning*, Cambridge: University Press, 2020, 411 p. ISBN: 978-1108470049
134. DENNET, D., *Consciousness Explained*, New York: Back Bay Books, 1992, 528 p. ISBN 978-0316180665
135. DREYFUS, H. L., *What Computers Can't Do: The Limits of Artificial Intelligence*, New York: HarperCollins, 1978, 259 p. ISBN-13 978-0060906139
136. DREYFUS, H. L., *What Computers Still Can't Do: A Critique of Artificial Reason*, The MIT Press, Revised ed. 1992, 429 p. ISBN-10: 0262540673
137. ERWIG, M., *Once Upon an Algorithm: How Stories Explain Computing*, Cambridge: MIT Press, 2017, 320 p. ISBN 9780262036634
138. FEYNMAN R., LEIGHTON R., SANDS M., *The Feynman lectures on physics - Volumes 1*, Addison Wesley, 1 edition, February 11, 1977
139. FRANKISH, K., RAMSEY, W. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge: University Press, 2014, 354 p. ISBN 978-0-521-87142-6
140. GARDNER, H., *Frames of Mind. The Theory of Multiple Intelligences*, New York: Basic Books, 2011, 467 p. ISBN 978-0-465-02509-1
141. HANCOCK, P., *Mind, Machine and Morality. Toward a Philosophy of Human-Technology Symbiosis*, London: CRC Press, 2009, 184 p. ISBN 978-0-7546-7358-3
142. HAREL, D., *Computers Ltd. What They Really Can't Do*, Oxford: University Press, 2000, 221 p. ISBN 0-19·850555·8
143. HAYKIN, S., *Neural Networks and Learning Machines*, Pearson, 2008, 906 p. ISBN 978-0-13-147139-9

144. HAYLES, N. K., *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago Press Journals, 1999, 364 p. ISBN 978-0226283845
145. HOFSTADTER, D. R., *Godel, Escher, Bach: An Eternal Golden Braid*, New York: Basic Books, 1999, 777 p. ISBN 978-0465026562
146. HUGHES, J., *Citizen Cyborg: Why Democratic Societies Must Respond To The Redesigned Human Of The Future*, New York: Basic Books, 2004, 318 p. ISBN 978-0813341989
147. KOCIJAN, V., CREȚU, A.-M., CAMBURU, O-M., YORDANOV, Y., LUKASIEWICZ, T., *A Surprisingly Robust Trick for the Winograd Schema Challenge*, 2019, https://arxiv.org/pdf/1905.06290.pdf
148. KORN, O., *Social Robots. Technological, Societal and Ethical Aspects of Human-Robot Interaction*, Switzerland: Springer, 2019, 288 p. ISBN 978-3-030-17106-3
149. KRAUSS, L. M., *A Universe from Nothing: Why There Is Something Rather than Nothing*, New York: FREE PRESS, 2012, 240 p. ISBN 978-1-4516-2445-8
150. INDURKHYA, N., DAMEREAU, F. (eds.), *Handbook of Natural Language Processing*, London: CRC Press, 2010, 678 p. ISBN 978-1-4200-8592–1
151. LAKOFF, G., JOHNSON, M., *Metaphors We Live By*, University of Chicago Press, 2003, 242 p. ISBN 978-0226468013
152. LAKOFF, G., JOHNSON, M., *Philosophy in the Flesh: the Embodied Mind & its Challenge to Western Thought*, New York: Basic Books, 1999, 240 p. ISBN 978-0465056743
153. LEBEN, D., *Ethics for Robots. How to Design a Moral Algorithm*, New York: Routledge, 2018, 158 p., ISBN 9781138716155
154. LENNOX, J., *2084. Artificial Intelligence and the future of humanity*, ePub Edition, April 2020: ISBN 978-0-310-10958-7
155. LIAO, M. (ed.), *Ethics of Artificial Intelligence*, Oxford: University Press, 2020, 525 p. ISBN 978-0-19-090503-3
156. LIDDELL, H.G., SCOTT, R., *A Greek-English Lexicon*, Oxford: Calderon Press, 1996. 1705 p. ISBN 978-0-19-864226-8
157. MACH, E., *The Analysis of Sensations, and the Relation of the Physical to the Psychical*, London: Forgotten Books, 2018, 406 p. ISBN 978-1528254526
158. MINSKY, M., *The Emotion Machine*, 2006, New York: Simon&Schuster, 372 p. ISBN 978-0-7432-7663-4
159. MORE, M., *Principles of Extropy. Version 3.0*, 1998, https://www.mrob.com/pub/religion/extro_prin.html

160. MORIOKA, M. (ed.), *Artificial Intelligence, Robots, and Philosophy*, Waseda: Journal of Philosophy of Life, 2023, 155 p., ISBN 978-4-9908668-9-1
161. MULLER, V., *Philosophy and Theory of Artificial Intelligence*, Switzerland: Springer, 2018, 316 p. ISBN 978-3-319-96447
162. MUNAKATA, M.-T., *Fundamentals of the New Artificial Intelligence. Neural, Evolutionary, Fuzzy and More*, Springer, 2008, 256 p. ISBN 978-1-84628-838-8
163. NEWELL, A., SHAW, J. C., SIMON, H. A., 1958. *Chess-Playing Programs and the Problem of Complexity.* In: IBM Journal of Research and Development, Volume: 2, Issue: 4, Oct. 1958 p. 320–335
164. NICOLESCU, B., *The Idea of Levels of Reality. Its Relevance for Non-Reduction and Personhood,* în *Transdisciplinarity in Science and Religion,* vol. 4, București; Curtea Veche, 2008, pp. 11-25
165. NÖE, A., *Action in Perception*, The MIT Press, 2006, 296 p. ISBN 978-0262640633
166. SINGH, R. et al., *Artificial Intelligence in Agriculture*, Oxon: CRC Press, 2022, 176 p. ISBN:978-1-032-15810-5
167. ROMPORTL, L. et al., *Beyond AI. Interdisciplinary Aspects of Artificial Intelligence*, Pilsen: University of West Bohemia, 2011, 134 p.
168. RUSSELL S., NORVIG P., *Artificial Intelligence. A Modern Approach*, Pearson, 2021, p. 1170 ISBN 9780134610993
169. SIEGEL, L., *Against the machine: being human in the age of the electronic mob*, New York: Spiegel and Grau, 2008, 182 p. ISBN 9781846686979
170. SKINNER, B. F., *Beyond Freedom and Dignity*, Middlesex: Penguin Books, 1971, 221 p.
171. TIOBE Index, https://www.tiobe.com/tiobe-index/
172. TURING, A., *The Essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life*, Oxford: University Press, 2004, 613 p. ISBN 0–19–825079–7
173. VARELA F.J, THOMPSON E.T., ROSCH E., *The Embodied Mind: Cognitive Science and Human Experience*, The MIT Press, 1992, 328 p. ISBN 978-0262720212
174. VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A., KAISER, L., POLOSUKHIN, I., *Attention Is All You Need*, Long Beach: 31st Conference on Neural Information Processing Systems NIPS, 2017, 15 p., https://arxiv.org/pdf/1706.03762.pdf
175. XIONG, M., *Artificial Intelligence and Causal Inference,* CRC Press, 2022, p. 368 ISBN 978-0-367-85940-4

Literatură în limba franceză

176. PASCAL, B., *Oeuvres complètes*, Paris: Editions du Seuil, 1966, 680 p.
177. HEISENBERG, W., *Philosophie. Le manuscrit de 1942,* translation from German and introduction by Catherine Chevalley), Paris: Seuil, 1998. *Ordnung der Wirklichkeit*, Munich, R. Piper GmbH § KG, 1989. Published first in W. Blum, H.P. Dürr, and H. Rechenberg (eds.), *Werner Heisenbergs Gesammelte Werke, Vol. C-I: Physik und Erkenntnis*, *1927-1955*, Munich, R. Piper GmbH § KG, 1984, pp. 21-306

Literatură în limba rusă

178. БЕРДЯЕВ, Н.А., *Человек и машина* // Путь. Май 1933
179. БЕРДЯЕВ, Н.А., *Дух и реальность*, Москва: АСТ, 2006, 679 с. ISBN 5-17-019075-1
180. БЕРДЯЕВ, Н.А., *Смысл творчества*, Москва: АСТ, 2007, 670 с. ISBN 978-5-17-043421-3
181. БЕРДЯЕВ, Н.А., *Диалектика божественного и человеческого*, Москва: АСТ, 2005, 620 с. ISBN 5-17-017990-1
182. БУЛГАКОВ, С.Н., *Два града*, Москва: Астрель, 2008, 784 с. ISBN 978-5-271-18452-3
183. ДОСТОЕВСКИЙ, Ф., *Бесы*, Москва: Эксмо, 2009, 608 с. ISBN 978-5-699-27465-9
184. КОРДЕЙРО, Х. Л., Вуд, Д., *Смерть должна умереть: Наука в борьбе за наше бессмертие*, пер. Матвея Окунева, Москва: Альпина Паблишер, 2021, 376 с. ISBN 978-5-9614-7224-0
185. КУТЫРЕВ, В. А., *Философия трансгуманизма: Учебно-методологическое пособие*, Нижний Новгород: Нижегородский университет, 2010. - 85 с.
186. ЛОСЕВ, А. Ф., *Диалектика мифа*, Москва: Академический Проект, 2008, 303 с. ISBN 978-5-8291-0970-7
187. ЛОССКИЙ, В., *Боговидение*, Москва: АСТ, 2006, 759 с. ISBN 5-17-018452-2
188. СОЛОВЬЕВ, В., *Чтения о Богочеловечестве*, Санкт-Петербург: Азбука-классика, 2010, 384 с. ISBN 978-5-9985-0653-6
189. СОЛОВЬЕВ, В., *Смысл любви*, Санкт-Петербург: Азбука, 2016, 349 с. ISBN 978-5-389-11654-2
190. ФРАНК, С., *С нами Бог*, Москва: АСТ, 2003, 750 с. ISBN 5-17-015704-5

ANEXE

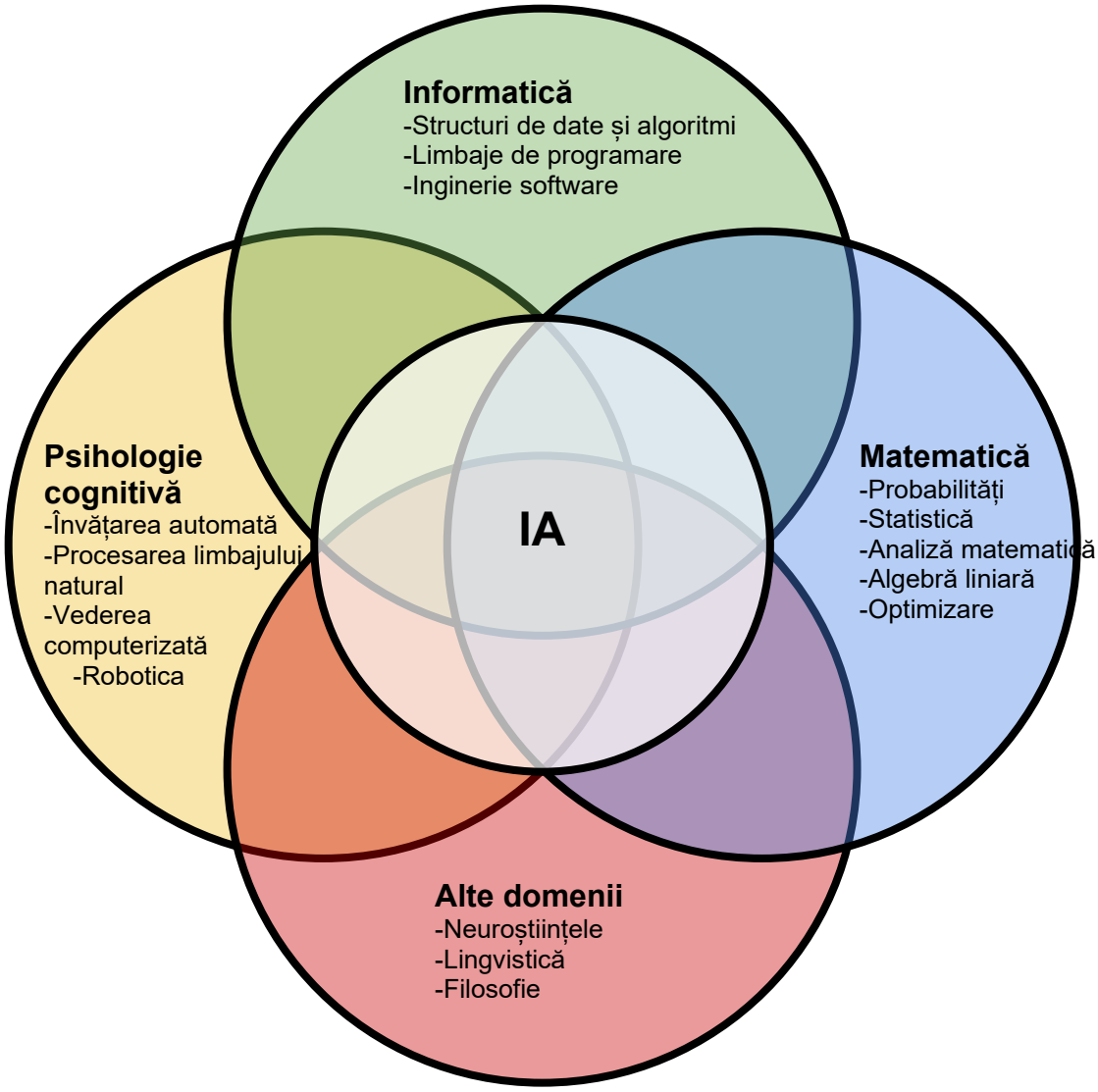


Fig. 1. 1 Fundamentele științifice ale inteligenței artificiale. Relațiile dintre diferite domenii. (schemă simplificată)

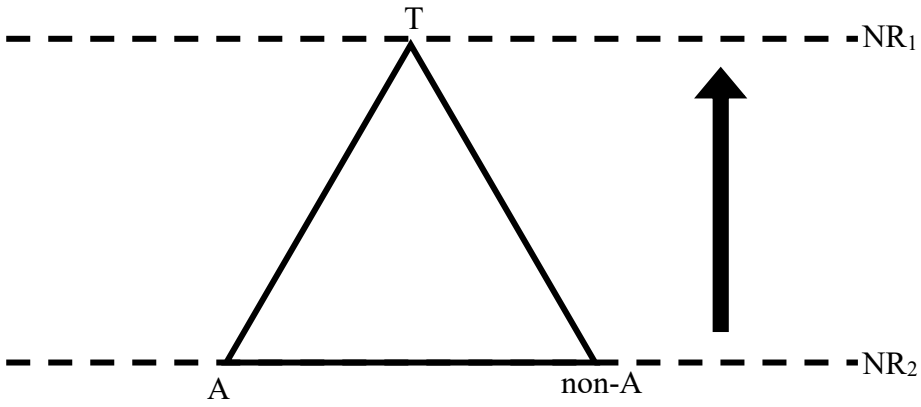


Fig. 1. 2 Reprezentarea simbolică a acţiunii logicii terţului inclus [91, p. 238]

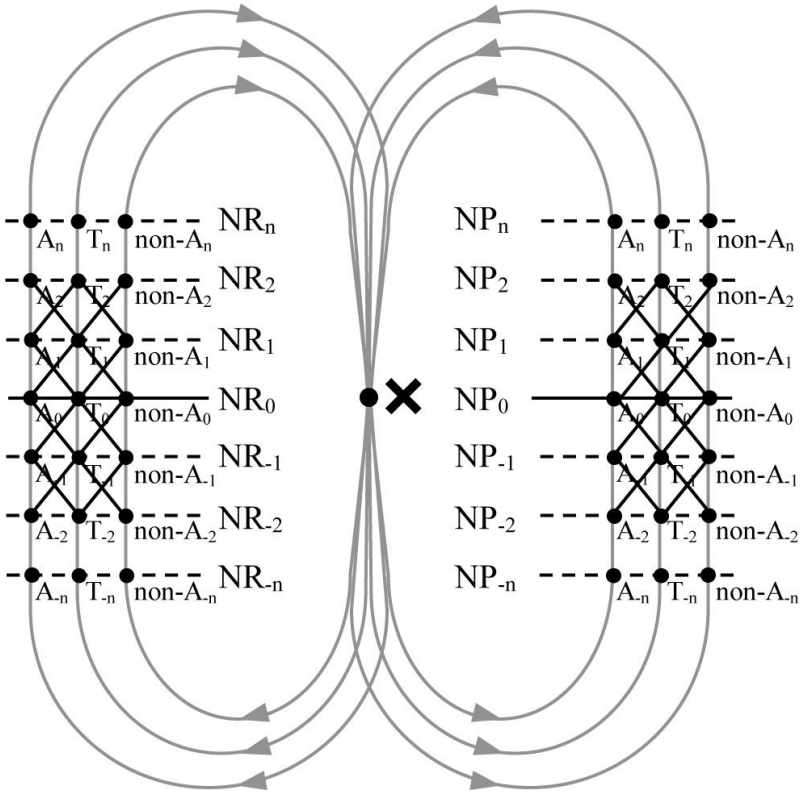


Fig. 1. 3 Obiectul transdisciplinar, Subiectul transdisciplinar și termenul de Interacţiune [91, p. 237]

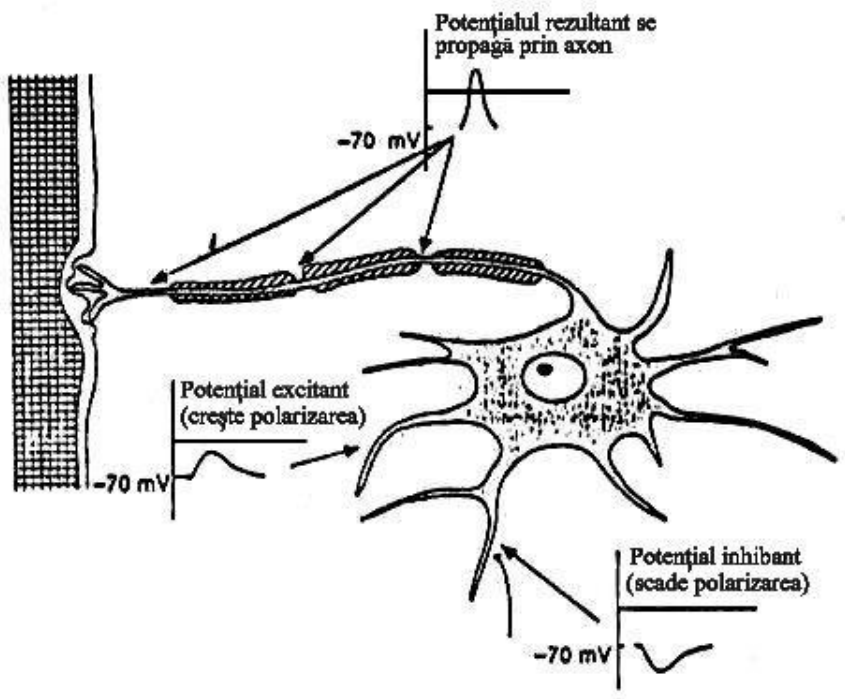


Fig. 2. 1 Structura neuronului biologic [55]

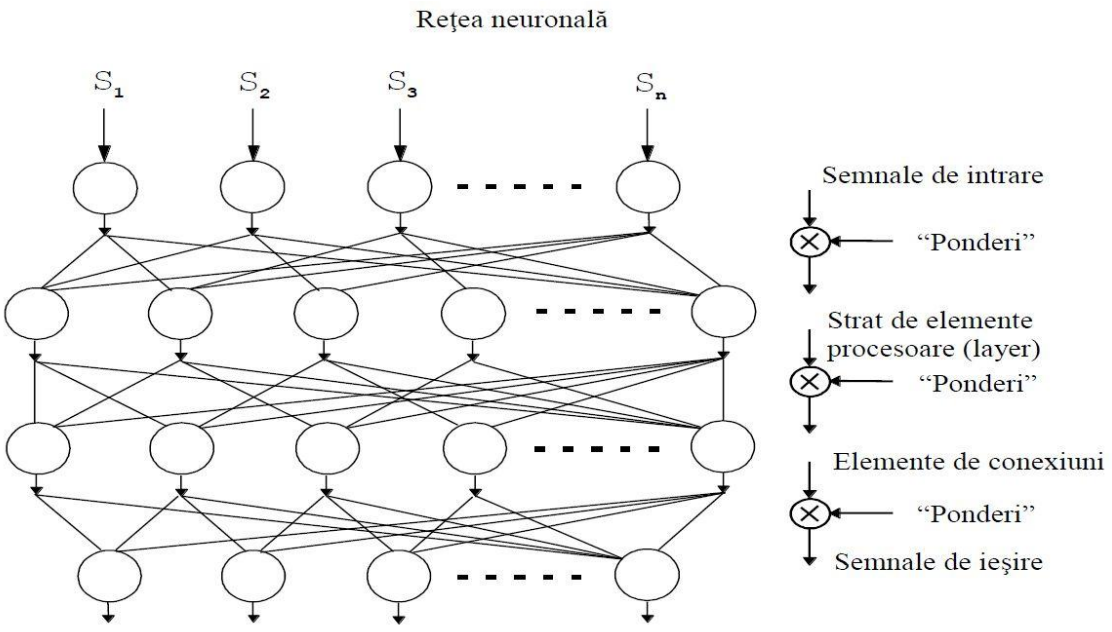


Fig. 2. 2 Schema de principiu a unei reţele neuronale artificiale [55]

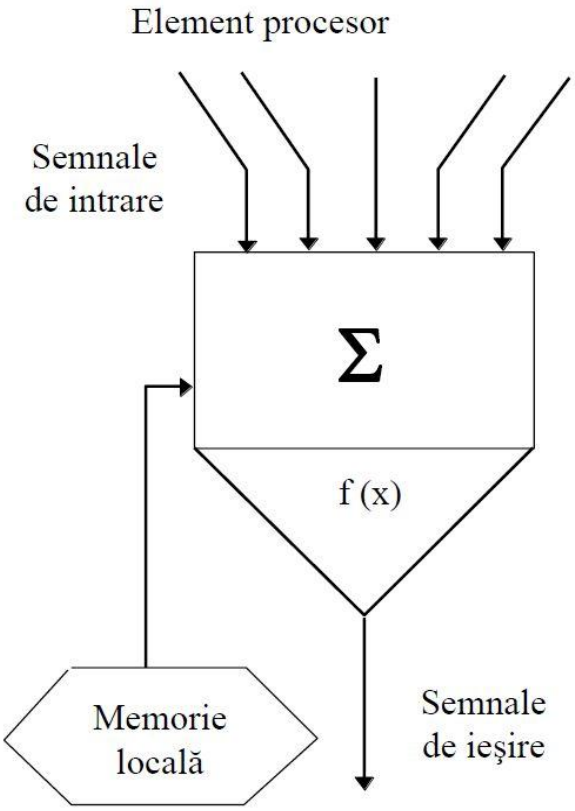


Fig. 2. 3 Schema de principiu al unui element procesor (neuron) și a fluxului de informație aferent [55]

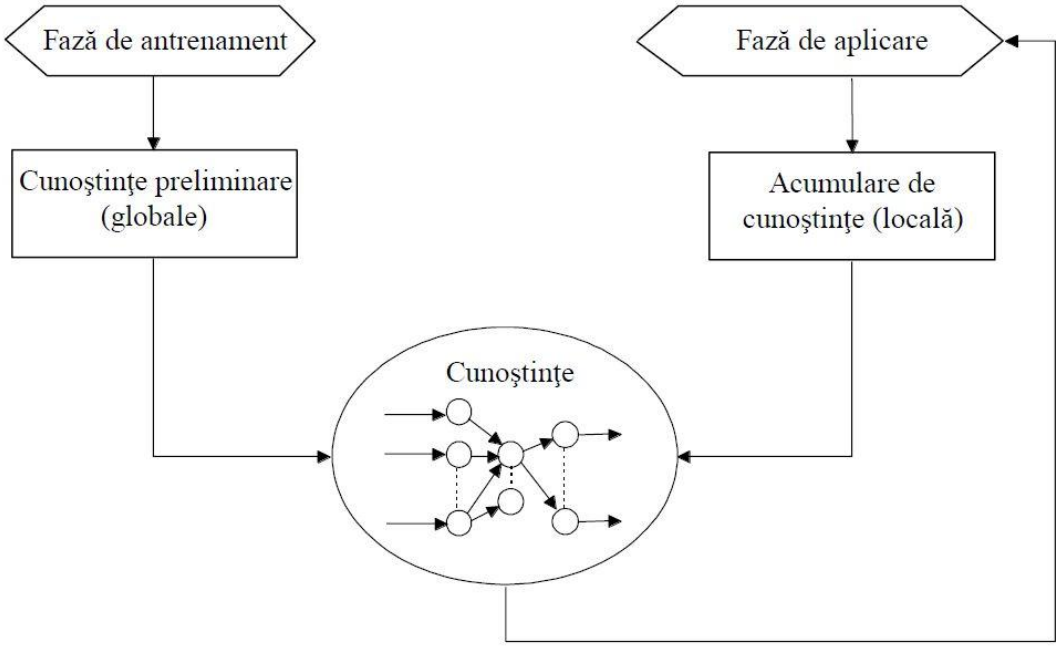


Fig. 2. 4 Conexiunile informaționale dintre faza de antrenament și cea de aplicare, în cazul folosirii unei reţele neuronale artificiale [55].

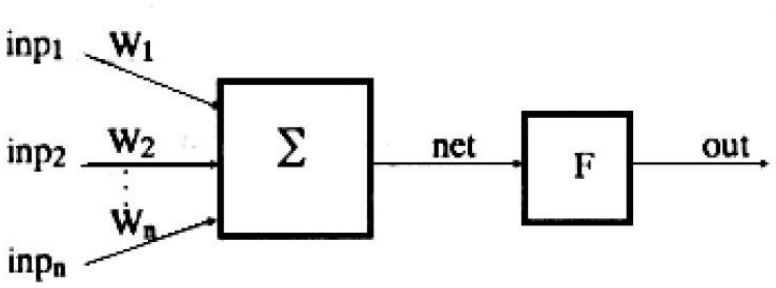


Fig. 2. 5 Schema neuronului artificial [55]

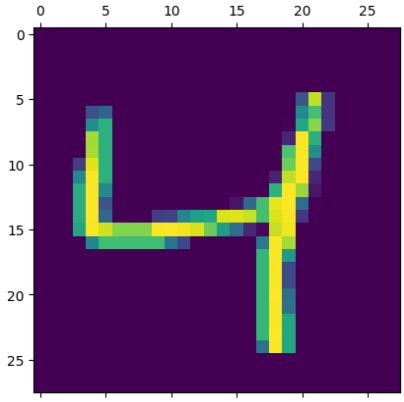


Fig. 2. 6 Exemplu de cifră scrisă de mână din baza de date MNIST (reprezentată printr-un grafic de tip heatmap - fotografia originală are rezoluția de 28x28 și valorile reprezintă nuanțe de gri (grayscale)

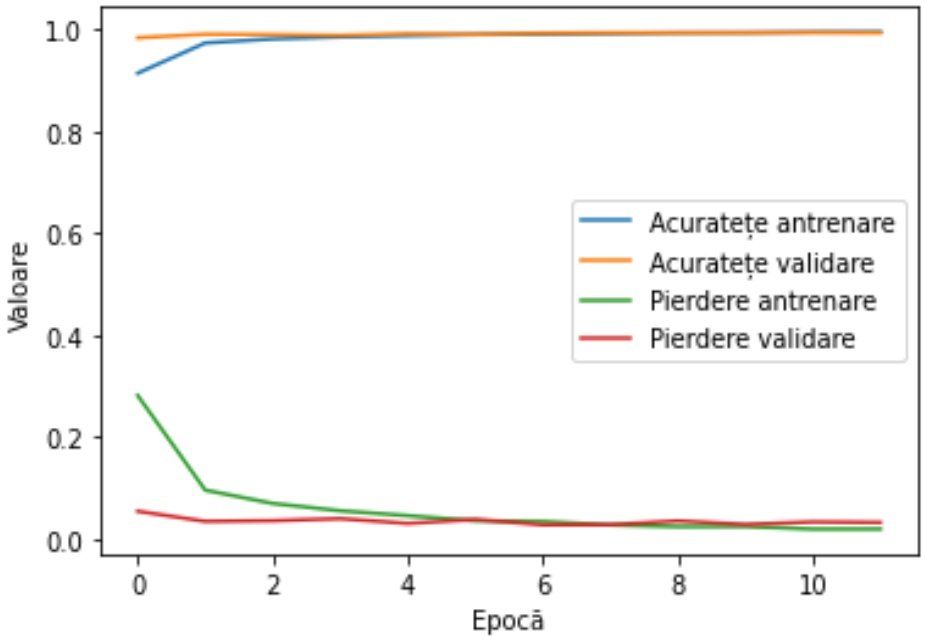


Fig. 2. 7 Evoluția acurateței și erorii („pierderii") în procesul de antrenare și de testare (validare)

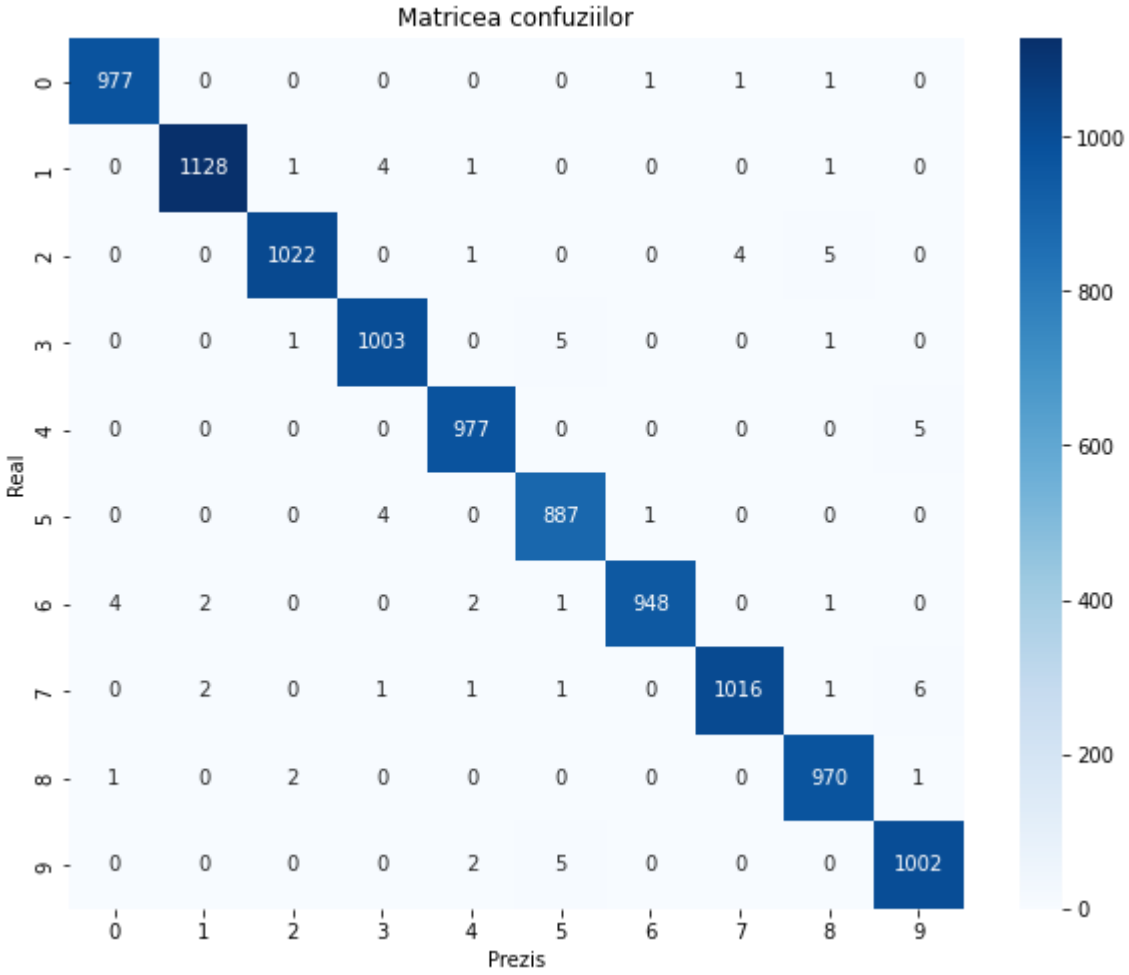


Fig. 2. 8 Matricea confuziilor

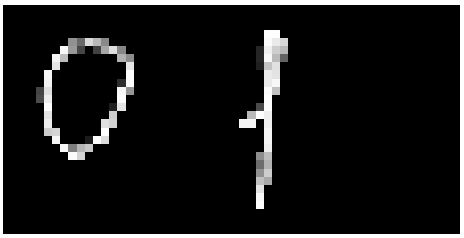


Fig. 2. 9 Cifrele 0 și 1 cu care a fost testată rețeaua neuronală



Fig. 2. 10 Poziție de șah pentru analiza pe computer [9, pp. 98-99]



Fig. 2. 11 Stockfish capturează turnul din a5 și… pierde partida.



Fig. 2. 12 O variație a poziției pe aceeași temă



Fig. 2. 13 În loc să așeze nebunul pe câmpul b4 și apoi să mute doar regele (poziția fiind, astfel, remiză), Stockfish își maximizează scorul, dar ajunge în poziție pierdută.

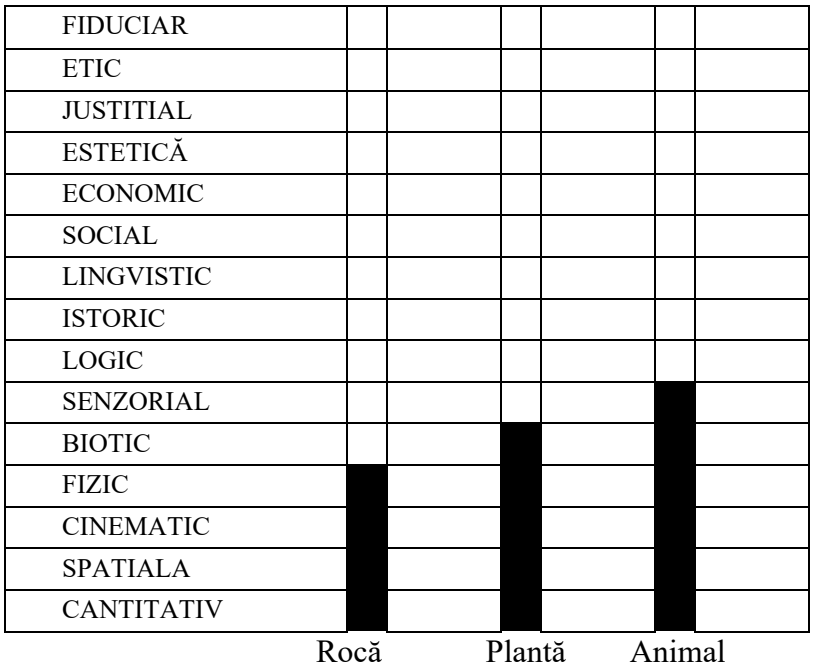


Fig. 2. 14 Schema funcțiilor active și pasive [132, p. 45]

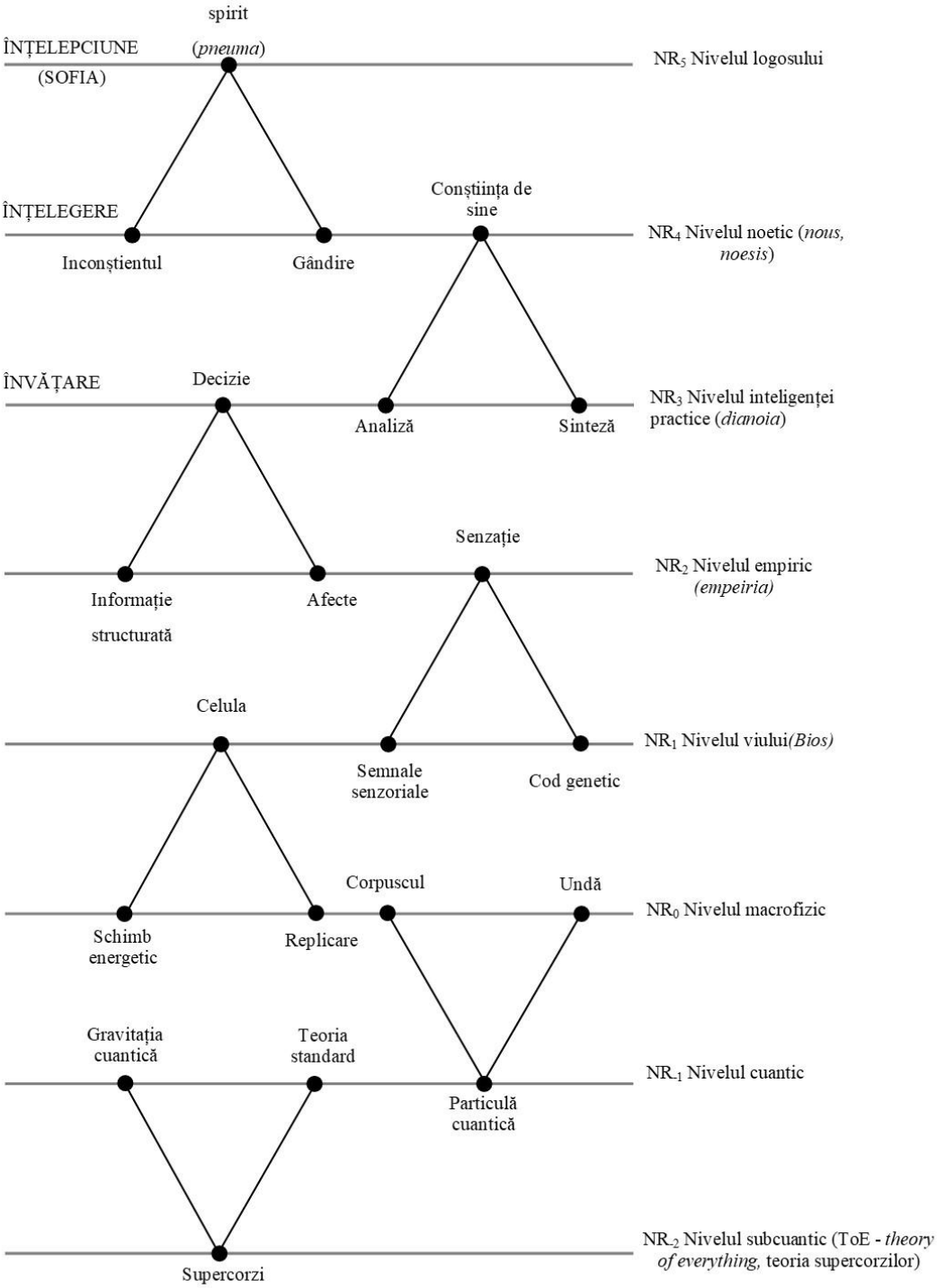


Fig. 2. 15 Modelul transdisciplinar al ființei umane

Declaraţia privind asumarea răspunderii

Subsemnata, declar pe răspundere personală că materialele prezentate în teza de doctorat sunt rezultatul propriilor cercetări şi realizări ştiinţifice. Conştientizez că, în caz contrar, urmează să suport consecinţele în conformitate cu legislaţia în vigoare.

Duca Mihail

20. 05. 2026



CV-ul candidatului

INFORMAŢII PERSONALE

Duca Mihail



Str. Morilor, 15, Bubuieci, mun. Chișinău, MD-2081, R. Moldova

+37369150280

duca.mihail.v@gmail.com

Data nașterii 12/10/1980

EXPERIENŢA PROFESIONALĂ

2025 – prezent — Asistent universitar, UTM

2024 – prezent — Asistent universitar, USM

2016 – prezent — Specialist pe probleme sociale
Asociația Obștească „ProValoare”
Str. Drumul Schinoasei 6, or. Codru, mun. Chișinău, R. Moldova

2002 – 2016 — Organizația Creștină „Viața Nouă” din R. Moldova

EDUCAŢIE ŞI FORMARE

2017 - prezent — Doctorand

- Profilul Filosofie, Specialitatea Ontologie și Gnoseologie, Școala doctorală științe umaniste și ale educației, Universitatea de Stat din Moldova
- Filosofia minții
- Inteligența Artificială
- Transdisciplinaritate

2015 - 2017 — Master în Științe umanistice

- Program de studii Filosofii contemporane, Facultatea Istorie și Filosofie, Universitatea de Stat din Moldova

1997 - 2002 — Inginer licențiat

- Profilul Construcții, Ingineria gospodăriilor comunale, Facultatea Urbanism și Arhitectură, Universitatea Tehnică a Moldovei

1990 - 1995
- Școala de arte plastice din or. Rezina

1987 - 1997
- Școala medie Nr. 1 din or. Rezina

Limba(i) maternă(e) Româna

Alte limbi străine cunoscute

	INTELEGERE		VORBIRE		SCRIERE
	Ascultare	Citire	Participare la conversație	Discurs oral	
Engleza	C1	C1	C1	C1	C1
Rusa	C2	C2	C2	C2	C2
Ucraineana	C1	C1	B1	B1	B1
Franceza	B1	B1	A1	A1	A1

Niveluri: A1/2: Utilizator elementar - B1/2: Utilizator independent - C1/2: Utilizator experimentat
Cadrul european comun de referință pentru limbi străine

Competențe de comunicare
Participarea cu diferite comunicări la conferințe naționale și internaționale.
Experiența de a ține discurs în fața unui public larg de peste o mie de participanți.
Folosirea diferitor mijloace tehnice pentru o comunicare mai eficientă.
Posibilitatea de a ține discurs fluent în română, rusă și engleză.

Competențe organizaționale/manageriale
Antrenare în leadership. Din 2005 am condus diferite echipe în țară și peste hotare. Multiple ocazii de a fi coordonator de proiecte.

Competențe informatice
- o bună cunoaștere a instrumentelor Microsoft Office™
- o bună cunoaștere a instrumentelor de design grafic
- o bună cunoaștere a instrumentelor de redactare video

Permis de conducere B

INFORMAȚII SUPLIMENTARE

Publicații

1. *Salt ontologic – tehnologic, uman și divin*, Universitatea de Stat din Moldova. Analele Ştiinţifice ale Universităţii de Stat din Moldova / Univ. de Stat din Moldova, Chișinău: CEP USM, 2017, pp 200-204, ISBN 978-9975-71-560-7, http://dspace.usm.md:8080/xmlui/bitstream/handle/123456789/1896/Anale_2017.pdf?sequence=1&isAllowed=y
2. *Divinizarea tehnologiei ca proces de dezumanizare a umanității*, Studia Universitatis Moldaviae , 2020, nr.10(140) Seria "Științe umanistice" ISSN 1811-2668 | E-ISSN 2345-1009, pp.68-73, https://humanities.studiamsu.md/nr-10-2020/
3. *La distanța de un click. Actualitatea lucrării Omul și mașina de N.A. Berdiaev pentru societatea hi-tech*, Filosofie, Educație, Distanță, Lucrările conferinței naționale online de filosofia educației, Editura Universității din București, 2022, pp 147-155, ISBN: 978-606-16-1359-5, https://editura-unibuc.ro/filosofie-educatie-distanta/
4. *Artificial intelligence. Myth and reality*, Annals of Philosophy, Social and Human Disciplines (APSHUS), 2023, vol. I, SSN: 2069 – 4008 (Print) / ISSN: 2069 – 4016 (Online), pp. 71-84, http://www.apshus.usv.ro/archive_template.php?myNumber=2023&myVolume=I

Conferințe

1. *Nivelurile de realitate și limitele inteligenței artificiale*, Școala internațională de metodologie în științe socio-umane. Conferință științifică internațională. Ediția III. 15-16 noiembrie 2018, Chișinău,Programul conferinței și rezumatele comunicărilor, pp. 141-142,
2. *Dialog iterdisciplinar între științele cognitive și teologie*, Școala internațională de metodologie în științe socio-umane. Conferință științifică internațională. Ediția IV. 19-20 noiembrie 2020, Chișinău, Programul conferinței și rezumatele comunicărilor. pp. 136-139,
3. *La distanța de un click. Actualitatea lucrării Omul și mașina de N.A. Berdiaev pentru societatea hi-tech*, Conferința online de filosofia educației. „Filosofie, Educație Distanță". București. Online, 5-6 martie 2021,
4. *Metamorfoza uman-postuman,* Conferințele București-Chișinău. „Identitate și valori culturale în context european". Ediția II. Online, 4 iunie 202,
5. *Metahumanism as a "simulacrum" of the "imago dei"*, Conferință internațională, „Knowledge exchange in the post-socialist black sea region: from closed systems to dialogue", Chișinău, 22-23 septembrie 2021,
6. *The role of the Myth of Artificial Intelligence in Hi-Tech society*, International Conference in Applied Ethics and Integrity in Doctoral

Conduct and Research, UNIVERSITY STEFAN CEL MARE OF SUCEAVA, Online, 4-5 noiembrie 2022

7. *Dezumanizarea umanității*,Conferința științifică națională cu participare internațională „Integrare prin cercetare și inovare”, 8-9 noiembrie 2018, Rezumate ale comunicărilor, Științe umanistice, Chișinău: CEP USM. pp. 30-33.
8. *Sunt oare maşinile cu adevărat inteligente?*Conferința științifică națională cu participare internațională „Integrare prin cercetare și inovare”, 7-8 noiembrie 2019, Rezumate ale comunicărilor, Științe sociale, Volumul II. Chișinău: CEP USM. pp. 166-170.
9. *A fi om în epoca digitală*,Conferința științifică națională cu participare internațională. „Integrare prin cercetare și inovare” Online, 10-11 noiembrie 2021,
10. *Inteligența artificială: între mit și realitate*, Conferința științifică națională cu participare internațională „Integrare prin cercetare și inovare”, Online, 10-11 noiembrie 2022,
11. *Inteligența din perspectiva inferenței. Problema abducției pentru inteligența artificială generală*, Conferința științifică națională cu participare internațională „Integrare prin cercetare și inovare”, Online, 9-10 noiembrie 2023,
12. *Omul și tehnologia: o abordare transdisciplinară*, Conferința științifică națională a doctoranzilor. „Metodologii contemporane de cercetare și evaluare”, Online, 22-23 aprilie 2021,
13. Membrual proiectului - Abordarea epistemologică a dezvoltării personale și a educației pentru societate: de la strategiile transdisciplinare la finalitățile pragmatice ale societății actuale din Republica Moldova (cifrul proiectului: 20.80009.1606.08)