

Inteligența Artificială (IA) reprezintă un domeniu al științei calculatoarelor care s-a constituit în scopul emulării comportamentului inteligent la mașini. La o rapidă trecere în revistă a publicațiilor ce abordează subiecte din domeniul IA devine evident faptul că IA este un complex de discipline ce acoperă subiecte de natură foarte diferită, de la logica formală până la învățarea automată. În acest context pare surprinzătoare la prima vedere introducerea unor termeni ce fac referire la o nouă IA (new AI), termen introdus de Prem (1992). Desigur, pentru a putea motiva termenii introduși, trebuie să caracterizăm pe scurt IA în sens clasic. IA urmărește ca scop, și aceasta nu diferă de noua abordare, obținerea calculatoarelor inteligente. IA încearcă să permită calculatoarelor să realizeze tot ceea ce oamenii pot face, restrictiv însă la anumite domenii specifice bine definite ale activității umane. IA clasică enunță că astfel de sisteme trebuie să fie bazate pe cunoștințe. În consecință, acestea trebuie să posede o structură pentru reprezentarea cunoștințelor și un mecanism pentru generarea a noi cunoștințe (sau a unei comportări dorite). Aceste cunoștințe sunt partiționate în unități mici, inteligibile, care conlucrează conform unor moduri ce pot fi descrise lingvistic pentru a conduce la rezultatele dorite. În extenso, metoda constă în analiza problemei și sinteza sistemului prin construcții necesar restrictive. Cel mai utilizat termen pentru o astfel de construcție este *regula*, utilizată atât pentru analiză cât și pentru descrierea constructivă a soluției. Regula însăși este enunțată pe baza *simbolurilor*. Alte construcții relative la această abordare sunt cadrele, rețelele semantice ș.a. Criticile față de această abordare a IA privesc în special eforturile de a utiliza astfel de sisteme pentru a explica fenomenele cognitive umane. IA a impus, prin specificul său, crearea unor limbaje adaptate acestui specific, primele două fiind LISP (McCarthy, 1960) și PROLOG (Roussel, Ph., 1975). IA este un domeniu important tocmai datorită capacității sale de a aborda noi clase de probleme, diferite de cele tratate de informatica clasică, cum ar fi: percepția, luarea de decizii, planificarea, diagnoza, interpretarea semnalelor, înțelegerea limbajului natural, concepția. Aceste probleme acoperă activități umane dintre cele mai variate și prezintă particularități comune care se bazează în mod fundamental pe exploatarea "*inteligentă*" a unor mari cantități de cunoștințe, specifice domeniului studiat. Unul din proiectele importante ale perioadei actuale este generația a șasea de sisteme de calcul. Acest proiect a fost lansat în anii '90 în Japonia și se bazează pe rețelele conexiunilor, conexiuni optice și paralelism masiv. Proiectul e împărțit în trei mari domenii: știința, tehnologiile și aplicațiile. Știința trebuie să caute răspunsuri la probleme precum: Ce sunt inteligența, mintea, cunoașterea? Cum gândim, vedem, auzim? Care este arhitectura și care este modul de funcționare a creierului? Ce coordonează comportamentul uman? Cum lucrează memoria umană? Tehnologia trebuie să proceseze informațiile furnizate de știință și să dezvolte noi tehnici de învățare, recunoaștere de forme și luarea deciziilor. Se caută noi arhitecturi hardware, noi algoritmi și noi materiale. Aplicațiile se suprapun cu cele ale sistemelor clasice de calcul, cu multe noi dimensiuni adăugate. De exemplu, un robot al zilelor noastre este doar un manipulator mecanic programat, pe când un robot din generația a șasea este echipat cu senzori preciși conectați la o bază de cunoștințe, robotul având un comportament mult mai apropiat de cel al unui sistem viu.

Ingenieria cunoștințelor (IC) a marcat generația a cincea, ocupându-se de manipularea simbolică și raționamente simbolice. Ingineria psihologică, inteligența neuronală a cunoștințelor (PINC), este definitorie pentru generația a șasea și este mult mai largă decât ingineria cunoștințelor, cuprinzând raționamente avansate, adaptative, de timp real, simbolice, precum și raționamente fuzzy, bazate pe o sinteză globală, masiv paralelă. Caracteristicile PINC includ printre altele: interfețe naturale cu utilizatorul bazate pe dialoguri nerestricționate, prin limbaj și imagine; interfețe prin unde cerebrale și semnale neuronale; comportări cvasiumane oferind utilizatorului senzații bogate și foarte complexe; interfațarea cu mediul prin senzori foarte preciși, datele furnizate de aceștia permițând autoorganizarea, adaptarea și învățarea; programarea automată (de exemplu programarea genetică).

Un pas nou început în anul 1995 îl constituie IA „invizibilă”, care produce sisteme ce funcționează, iar inginerii software folosesc diagnoze pe bază de model, module bazate pe reguli și agenți pentru

interfețe om-calculator inteligente. Tehnicile de IA sunt într-un mod invizibil interconectate cu sistemele existente. Implicarea rezultatelor IA în mai toate domeniile este un lucru vizibil. Unele exemple sunt: mica agrafă de ajutor (asistentul) din pachetul Office al firmei Microsoft folosește un model de rețea Bayse; instrumentul care subliniază ortografia incorectă din produsul Microsoft Word este realizat pe baza cercetărilor în domeniul procesării limbajului natural; programul pentru calculul impozitelor utilizează metode ale IA bazate pe reguli, televizorul și multe alte produse electrocasnice includ mecanisme de control realizate în laboratoarele de Inteligență artificială, etc.

În domeniul procesării limbajului natural s-au obținut importante rezultate cum ar fi: traducerea unor limbi naționale, înțelegerea unui text, etc. Realitatea virtuală și jocurile sunt o zonă de avânt în care fabricanții și-au dat seama de rezultatele pe care le pot obține utilizând facilitățile produselor IA, cum ar fi: jucătorii artificiali. În unele din aceste jocuri problema este că jucătorii artificiali sunt atât de buni încât jocul devine plictisitor pentru jucătorii umani care nu pot câștiga.

Alte realizări sunt comenzi orale pentru calculator, care sunt deja existente pe piață în domeniul serviciului telefonic. Tendința în acest domeniu este de a realiza traducerea limbii în același timp cu recunoașterea vorbirii. Deja sunt implementate unele programe demonstrative în acest domeniu în unele corporații importante. Totuși mai există încă unele probleme ce trebuiesc rezolvate. Un lucru interesant este acela că tehnologia roboților devine mai ieftină și mult mai răspândită, unele aplicații fiind roboții jucării. O mare provocare în domeniul roboticii este interacțiunea cu calculatorul, cu Internet-ul astfel încât utilizatorii să ajungă mai repede la informațiile dorite, și mergând mai departe, ideea revoluționară a implicării tehnologiei informațiilor și a roboților în procesul educațional.

Un exemplu de aplicație bazată pe cunoștințe este un motor de căutare pe Internet care are câteva rezultate captivante, și își folosește baza de cunoștințe pentru căutări eficiente; de exemplu dacă i se cere să caute după cuvântul „mouse”, motorul întoarce o întrebare: „animalul sau dispozitivul de intrare?”, și în funcție de răspuns va genera o nouă căutare. Domeniul roboților aduce importante descoperiri și utilizări și în ceea ce privește roboții utilizați de NASA, capabili să exploreze suprafața planetei Marte și să realizeze cu succes părți ale misiunii în spațiu, fără intervenție umană. Următorii zece ani se preconizează că rezultatele IA vor fi orientate către aplicațiile Internet: recunoașterea vorbirii, traducerea limbajului, căutări mai eficiente și multe altele. De asemenea se lucrează mult în domeniul roboticii, care va avea rezultate în jucăriile pentru copii. Viitorul cel mai promițător aparține sistemelor neuronale, care se inspiră din structurile creierului uman și a căror originalitate ține seama de distribuția cunoașterii. Software-ul integrabil în sistemele neuronale va fi disponibil ca instrumente hibride de concepere a sistemelor care asociază tehnici orientate obiect celor din programarea clasică, tehnici fuzzy, etc.

Sistemele de IA în general au următoarele caracteristicile: realizarea unor operații în medii ostile, domenii mari și cuprinzătoare de cunoștințe, abilitatea de a interacționa natural cu oamenii și un anumit grad al auto-înțelegerii și integritate internă. Din această perspectivă, subliniem:

- *Mediul înconjurător este un mediu ostil care nu poate fi anticipat. Pentru un robot sau un agent inteligent confruntat cu cereri care se schimbă în timp este vital ca acel sistem să poată învăța rapid din experiența sa și să se adapteze la schimbările din mediu pentru a nu eșua.*
- *Sistemul va necesita largi cantități de cunoștințe despre mediu. El trebuie să cuprindă aceste informații într-o formă convenabilă pentru a fi utilizată în diferite sarcini, pentru a realiza performanțe și pentru a-și însuși informații din propria sa experiență de rezolvare a problemelor prin cooperare cu oameni sau cu alți agenți. Explorarea unor porțiuni mari de informație slab structurată, cum este World Wide Web, va necesita eficiență și dexteritate în manevrarea cantităților foarte mari de informație complexă.*
- *Pentru a interacționa cu colaboratorii umani, sistemul trebuie să se angajeze în dialoguri extinse ce clarifică și îmbogățesc progresiv adâncimea cunoașterii. Acest lucru necesită utilizarea limbajului natural și dispozitive de afișare corespunzătoare precum și modalități de comunicare tactile, decât simplul schimb de propoziții corecte așa cum cer sistemele limitate din ziua de astăzi. Interacțiunea cu colaboratorii lor*

artificiali atrage după sine economie și claritate și nu necesită obligatoriu utilizarea aceleiași limbi pentru comunicarea cu oamenii.

- *Sistemul trebuie de asemenea, să se înțeleagă în aceeași măsură în care îl înțeleg colaboratorii săi, cu atât mai bine cu cât integrează diversele sale componente, facilitează interacțiunea lor omogenă, menține și îmbogățește cunoștințele și pricepera, și adaptează dinamic utilizarea resurselor.*

Construirea sistemelor cu astfel de caracteristici aduce o provocare pentru IA și fiecare zonă de investigație: reprezentarea și procesarea cunoștințelor, învățarea și adaptarea, dezbateră, planificarea și acțiunea, procesarea limbajelor și vorbirii, înțelegerea și sinteza imaginii, manevrarea și locomoția, agenți autonomi și roboți, sisteme multiagent, modelarea cognitivă, și fundamente matematice – suportă un efort de cercetare riguros contribuind la îndeplinirea acestor provocări. Aceste direcții constau în analiza sistematică și integrarea intelectuală printre care construirea roboților (fizic și computațional) și modelarea rațională (în principal în ceea ce privește luarea deciziei) formează două cazuri speciale și cuprinzătoare. Sprijinirea colaborării, intensificarea comunicării, obținerea unor cunoștințe importante cerute pentru a acționa inteligent și aprofundarea fundamentelor matematice corespunzătoare sunt alte direcții majore ale IA. Direcția roboticii constituie de asemenea o zonă tehnică, dar una vastă care atinge și celelalte zone.

1.2 Prezentare în paralel a IA simbolică și IA conexionistă

IASimbolică poate aborda o mare varietate de domenii:

- **vederea artificială** - ce presupune recunoașterea formelor, identic cu vederea umană;
- **robotica** - focalizează producerea dispozitivelor mecanice capabile să reproducă mișcarea;
- **prelucrarea vocii** - ce privește constituirea și sinteza vocii umane;
- **prelucrarea în limbaj natural** - înțelegerea și vorbirea în limbaj natural;
- **demonstrarea teoremelor** - în matematică și logică;
- **“General Problem Solving”** - rezolvarea unei clase generale de probleme exprimate în limbaje formale;
- **recunoașterea formelor** - recunoașterea și clasificarea diferitelor forme;
- **teoria jocurilor**;
- **învățarea automată** - mașini ce acumulează cunoștințe prin observarea exemplilor;
- **sisteme bazate pe cunoștințe.**

Nucleul comun al acestor domenii diverse îl constituie emularea comportamentelor intelectuale umane, folosind un formalism simbolic bazat pe logica formală. Acest nucleu comun al IA (manipularea formală a simbolurilor pe principii logice) o diferențiază de informatica clasică și de alte domenii conexe cum ar fi psihologia).

Dacă se consideră IA drept un model cognitiv, se pot evidenția următoarele sale caracteristici:

- **reprezentare simbolică clară** (logico-lingvistică);
- **compoziționalitate** (modularitate și structuri de reprezentare, care o diferențiază de abordările conexioniste);
- **funcționalism** (se elimină interesul față de creier);
- **învățarea nu este o motivație de bază** (aceasta fiind considerată o caracteristică separată).

Tehnologia sistemelor inteligente de tip sisteme expert este cea mai „veche” și cea mai bine pusă la punct, în special pentru sisteme expert bazate pe reguli. Există deja câteva sute de sisteme expert operaționale și comerciale pentru o gamă largă de aplicații. Toate folosesc o cunoaștere numită

expertiză, provenită de la experții umani, iar procesul de colectare se numește achiziția de cunoștințe. Cercetătorii sunt de acord că sub denumirea de sisteme expert se află acele programe de inteligență artificială sau mașini bazate pe cunoașterea de nivel înalt, comparabilă cu a celor mai competenți specialiști într-un domeniu aplicativ și care pot realiza performanțe de gândire și intuiție similare experților umani. Sistemele expert reproduc raționamentul experților umani asupra cunoștințelor puse la dispoziția lor. Structura unui sistem expert trebuie să conțină obligatoriu trei module principale care formează așa-numitul *sistem esențial*:

- *Baza de cunoștințe* este formată din ansamblul cunoștințelor specializate introduse de expertul uman. Cunoștințele stocate aici sunt în principal descrierile obiectelor și ale relațiilor dintre acestea. Baza de cunoștințe face parte din sistemul cognitiv, cunoașterea fiind memorată într-un spațiu special organizat. Forma de stocare trebuie să asigure căutarea pieselor de cunoaștere specificate direct prin simboluri identificatoare sau indirect, prin proprietăți asociate sau inferențe care pornesc de la alte piese de cunoaștere;
- *Mecanismul (sau motorul) de inferență* reprezintă noutatea sistemelor expert. El preia din baza de cunoștințe faptele utilizate pentru construirea raționamentului. Mecanismul de inferență urmărește o serie de obiective majore cum ar fi: alegerea strategiei de control în funcție de problema curentă, elaborarea planului de rezolvare a problemei după necesități, comutarea de la o strategie de control la alta, executarea acțiunilor prevăzute în planul de rezolvare. Deși mecanismul de inferență este constituit dintr-un ansamblu de proceduri în sensul obișnuit al termenului, modul în care utilizează cunoștințele nu este prevăzut prin program, ci depinde de cunoștințele pe care le are la dispoziție;
- *Baza de fapte* reprezintă o memorie auxiliară care conține toate datele utilizatorului, (faptele inițiale care descriu enunțul problemei de rezolvat) și rezultatele intermediare produse în cursul raționamentului. Conținutul bazei de fapte este stocat în general în memoria volatilă (RAM), însă la cererea utilizatorului, el poate fi conservată.

Pe lângă acestea, un sistem expert mai conține o serie de module ce asigură comunicarea cu operatorul și expertul uman. *Avantajele* sistemelor expert sunt:

- sunt colecții de informații valoroase;
- sunt indispensabile în lipsa expertizei umane;
- în unele situații, pot fi mai ieftine și mai eficiente decât experții umani;
- pot fi mai rapide decât experții umani;
- dacă sunt flexibile, pot fi actualizate cu ușurință;
- pot fi folosite pentru instruirea de noi experți umani;
- la cerere, pot explica premisele și linia de raționament;
- tratează incertitudinea într-o manieră explicită, care, spre deosebire de cazul experților umani, poate fi inspectată și verificată;

De asemenea, trebuie amintite și câteva *limite* ale acestor sisteme:

- nu pot raționa pe baza intuiției sau bunului simț („common sense”), deoarece acestea nu sunt ușor reprezentabile;
- sunt limitate la un domeniu restrâns; cunoștințe din alte domenii nu pot fi ușor integrate, nici nu pot genera în mod convingător;
- procesul de învățare nu este automat; pentru actualizările cunoștințelor este nevoie de intervenția umană;
- în prezent, nu pot raționa pe baza teoriilor sau analizelor;
- cunoștințele înmagazinate în baza de cunoștințe depind foarte mult de expertul uman care le exprimă și articulează.

Sistemul expert este un intermediar între expertul uman și utilizatorul uman în sensul că interacționează cu expertul uman la momentul „achiziției cunoștințelor” și cu utilizatorul la momentul „consultației”. Un sistem expert are posibilitatea să explice modul în care a ajuns la o anumită concluzie pe baza cunoștințelor pe care le posedă. Adevărata putere a sistemelor expert stă în motoarele de inferență și în sistemele de explicații de care dispun. Totuși, sistemele expert nu sunt capabile să învețe să se instruiască automat și din acest motiv nu se pot adapta ușor la schimbările

intervenite în mediul în care acționează. Cercetările recente se concentrează pe dezvoltarea unor generatoare de sisteme expert mai puternice, comercializabile, de genul instrumentelor generale hibride care combină mai multe reprezentări ale cunoașterii sau stiluri de programare, inclusiv reguli bazate pe obiecte. Sistemele expert sunt un instrument modern în mâna oamenilor de afaceri pentru obținerea de soluții acceptabile, aducătoare de profit la problemele cu care se confruntă zilnic. Sistemele expert integrate în aplicațiile multimedia constituie cea mai spectaculoasă realizare pentru afaceri pentru că acestea permit urmărirea videoconferințelor de la rețele specializate. Sistemele expert industriale pot oferi servicii de diagnostic de proces pentru întreținerea și repararea liniilor de producție robotizate sau instruirea la timp pentru siguranța și securitatea instalațiilor industriale. Principalul obstacol în perfecționarea și răspândirea sistemelor expert se află în faza achiziției de cunoștințe și constă în dificultatea stabilirii conceptelor, relațiilor dintre ele și regulilor de deducție specifice domeniului investigat.

1.3 Caracteristicile Sistemelor Neuronale Artificiale

Sistemele neuronale artificiale (rețelele neuronale), sunt numite adesea sisteme conexioniste, deoarece modelează soluțiile problemelor simulând conectarea neuronilor într-o rețea. Un sistem neuronal este un calculator analogic care folosește elemente simple de procesare conectate într-un nivel înalt de paralelism. Elementele de procesare execută funcții booleene simple sau funcții aritmetice la intrările lor. Cheia funcționării unei rețele este *pondera* asociată fiecărui element. Ponderile reprezintă informația memorată în sistem. Neuronal poate avea mai multe intrări dar numai o ieșire. Semnalele de intrare către neuroni sunt înmulțite cu ponderile și însumate pentru a produce intrarea totală a neuronului. Ponderile pot fi reprezentate într-o matrice și identificate prin indici. Ieșirea neuronului este adesea o funcție sigmoidă a intrării. Fiecare neuron are asociată o valoare de prag, care se scade din intrarea totală. Rețeaua învață prin adaptarea automată a ponderilor care în rețea conectează neuronii, ponderi care împreună cu valorile prag determină propagarea datelor în cadrul rețelei, deci determină răspunsul ei corect la datele de antrenament. Determinarea rețelei să furnizeze răspunsurile corecte este un proces care poate dura ore sau zile, în funcție de numărul de modele pe care rețeaua trebuie să le învețe, de hardware și de software. Cu toate acestea, o dată ce procesul de învățare s-a finalizat, rețeaua reacționează foarte rapid.

Arhitectura SNA este foarte diferită de arhitectura clasică a unui computer, în cadrul căreia este posibilă corelarea informațiilor discrete cu celulele memoriei, un SNA fiind modelat după teoriile cerebrale contemporane în care informația este reprezentată prin ponderi. Cu toate acestea, nu există o corelație directă între o pondere specifică și un anumit element din informația stocată. Această reprezentare distribuită a informației este similară unei holograme, ale cărei linii acționează ca o rețea de difracție în care imaginea stocată este reconstituită la trecerea razei laser.

O rețea neuronală reprezintă o bună alegere atunci când nu există decât date empirice numeroase și nici un algoritm care să furnizeze suficientă acuratețe și viteză. În acest caz un SNA oferă anumite *avantaje*:

- sistemul de stocare este tolerant la erori, prin faptul că se pot înlătura porțiuni din rețea și astfel se deteriorează într-o manieră limitată doar calitatea datelor stocate, datorită manierei de distribuire a informației;
- calitatea imaginii stocate se degradează proporțional cu dimensiunile rețelei detașate. Cu alte cuvinte nu există o pierdere catastrofală de informație, aceasta fiind o caracteristică a hologramei;
- datele sunt stocate în mod natural în forma memoriei asociative. Aceasta reprezintă un tip de memorie în care datele parțiale sunt suficiente pentru a actualiza toată informația stocată. Aceasta contrastează cu memoria convențională prin care datele sunt actualizate pe baza specificării locației datelor care trebuiesc actualizate: o intrare parțială sau bruiată determină astfel omiterea unei părți din informația originală;

- rețelele pot interpola și extrapola din propria lor informație memorată. Prin antrenament se poate învăța o rețea să observe caracteristici specifice sau relații în date;
- rețelele au plasticitate. Dacă un număr de neuroni este înlăturat, rețeaua poate fi reantrenată la nivelul său original de pricepere, dacă au rămas în rețea suficienți neuroni.

Aceste caracteristici fac rețelele neuronale deosebit de atractive pentru roboți spațiali, dispozitive subacvatice, controlul proceselor și alte aplicații care necesită funcționarea timp îndelungat în mediu ostil, fără reparații. Pe lângă cerința de fiabilitate, rețelele neuronale oferă posibilitatea unui cost scăzut de mentenanță datorită plasticității. Dacă hardware-ul trebuie reparat, este probabil să fie mai costisitor de reprogramat rețeaua decât de înlocuit. Rețelele neuronale nu sunt în general recomandate pentru aplicații mari consumatoare de calcule numerice sau care cer soluții optime. Dacă pentru problema respectivă există soluție algoritmică, o rețea neuronală nu este cea mai bună alegere. Rețelele neuronale își dovedesc în principal utilitatea în rezolvarea unor probleme dificile, cum sunt cele de estimare, identificare și predicție, sau de optimizare complexă. Datorită independenței efectuării operațiilor în interiorul componentelor față de celelalte componente din sistem, modelele conexiunilor au un potențial mare de paralelism. Modul de memorare și procesare a datelor diferențiază rețelele neuronale artificiale de programele clasice, care urmează instrucțiunile într-o ordine secvențială predefinită, informația fiind memorată în zone bine definite. Datorită capacității lor de a rezolva probleme complexe pe baza unei mulțimi consistente de exemple, sistemele conexiunilor au un spectru larg de aplicabilitate: de la sisteme de recunoaștere de forme (caractere, semnături) sau de semnale sonore, până la sisteme pentru controlul unor procese complexe, cum ar fi sistemele de auto-reglare sau piloții automați.

În prezent, în domeniul IA se înregistrează o nouă deplasare a accentului asupra prelucrărilor numerice, mai eficiente în execuție. Atât calculul neuronal cât și cel simbolic se bazează pe prelucrări preponderent numerice. Superioritatea sistemului nervos față de procesoarele artificiale de informații se datorează unor caracteristici precum: dimensiune, conectivitate, paralelism. Sistemele bazate pe calcul simbolic au posibilități reduse de învățare. Aceste sisteme sunt în majoritate sisteme deductive, ce trag concluzii din datele furnizate, dar nu pot decât cu mare greutate să achiziționeze sau să genereze noi cunoștințe. Creșterea complexității sarcinii de rezolvat reclamă creșterea cantității de cunoștințe ce trebuie încorporată în sistem. Operarea cu volume mari de cunoștințe ridică dificultăți deosebite. Așa se explică de ce sistemele expert care au limitat foarte mult domeniul acoperit pentru a reduce volumul de cunoștințe necesar au reprezentat un succes al IA. Sistemele expert operează foarte bine în domeniile înguste pentru care au fost proiectate. Calculul neuronal depășește dificultățile întâmpinate de calculul simbolic în realizarea învățării automate, prin asigurarea unei codificări distribuite difuze a cunoștințelor, în întreaga rețea neuronală, ca valori ale intensităților conexiunilor dintre neuroni. Această reprezentare distribuită stă la baza definirii unor algoritmi eficienți de instruire. Calculul simbolic nu reprezintă un bun instrument pentru emularea proceselor de învățare, majoritatea bazate pe procese inductive, de natură preponderent intuitivă. Metodele mixte presupun combinarea învățării simbolice cu învățarea prin calcul neuronal așa numită învățare empirică.

Noua tendință în IA folosește noi metode (specifice rețelelor neurale) pentru a modela sistemele cognitive și, în final, sistemele inteligente. Noile caracteristici provin din modalitățile de construire a modelului cunoașterii, adică obținerea unui model al comportamentului complex care este dificil de simulat și prezis dar urmărește anumite obiective predefinite. Trebuie subliniat că noua tendință în IA urmărește aspectul modelării cognitive, având ca surse biologia și teoriile evoluționiste în contextul ipotezei că inteligența umană este bazată istoric pe cea a animalelor. Reducerea problemelor (în sensul clasic al IA) către un set corect și complet de reguli este necesară datorită condițiilor de realizabilitate impuse de utilizarea arhitecturii von Neumann. Dar această modelare prin reducere implică și reducerea, în însăși substanța sa, a sensului domeniilor IA (spre exemplu, demonstrarea automată a teoremelor, procesarea limbajului natural ș.a.). Toate aceste discipline tradiționale tind să excludă subiectul propriilor investigații prin includerea obiectivității instrumentului matematic, logic și

lingvistic. Principala caracteristică a unei probleme de IA este explozia de combinații și diversitatea acțiunilor și soluțiilor posibile, cât și a intrărilor. Aceasta este o cauză determinantă a necesității unui mecanism viabil de reducere a complexității problemei, urmărind modularizarea diversității în cazuri, decizia corectă în fiecare caz și validarea completitudinii (acoperirea tuturor cazurilor). Regulile, în sensul folosirii predicatelor, au în acest sens un rol comprehensiv, urmărind reducerea complexității problemei în funcție de un anumit obiectiv, cu prețul reducerii sensurilor, deci a diversității.

Dimpotrivă, noul sens dat IA urmărește un tratament nereductiv al cât mai multor aspecte ale cunoașterii umane fără sau cu minime constrângeri predicative ale diversității intrărilor unui model (plecând de la premisa că proiectantul nu poate anticipa toate aspectele necesare modelării). Singura abordare rămasă este *învățarea*, prin reafirmarea istoriei și subiectivității ca și condiții necesare ale actului inteligent. Se evidențiază astfel experiența subiectului și calitățile contextuale. Una din principalele probleme ale unei astfel de abordări constă în proiectarea, respectiv specificarea obiectivelor. Un astfel de obiectiv implică probleme similare proiectării unui set de reguli: determinarea obiectivelor rezonabile este comparabilă ca dificultate cu stabilirea corectitudinii și completitudinii setului de reguli ce descriu comportarea modelului. Un posibil răspuns practic la această problemă poate fi tot *învățarea*.

Problema corelării metodelor simbolice, bazate pe reguli, cu procesarea asociativă, bazată pe conexiunism, trebuie să combine avantajele aduse de cele două abordări. Rezultatele obținute prin integrare neuro-simbolică pot fi aplicate atât pentru explicarea proceselor cognitive umane (identificare, asociere, generalizare, raționament, caracterizare a diferitelor stări mentale) cât și pentru construirea unor sisteme adaptive care să recunoască, să clasifice, să învețe din exemple, să se auto-organizeze, să raționeze, să folosească limbajul natural ș.a.

În acest context, rețelele neurale (Artificial Neural Networks, ANN), privite ca model al creierului uman, pot fi interpretate, datorită aproximărilor și simplificărilor folosite, ca o posibilitate de abordare cognitivă de nivel „înalt”. Procesele cognitive sub-simbolice, dependente de caracteristicile reprezentărilor neurale interne (unele invizibile lumii externe), sunt implementate prin scheme și modele neurale. Aceste caracteristici pot reprezenta diferite tipuri de date: semnale senzoriale, informații numerice, variabile lingvistice. Se poate imagina un sistem de coordonate bazat pe aceste caracteristici care să definească spațiul multidimensional mental: în acest spațiu, o funcție mentală se definește prin descrierea obiectelor mentale ca domenii fuzzy în care funcția are valori nenule. Valorile active ale reprezentării interne la un moment dat reprezintă starea mentală, căreia îi corespunde un punct în spațiul mental. Dacă în această regiune există un obiect mental, acesta va fi recunoscut (activat). Evoluția stărilor mentale este echivalentă cu o succesiune de activări ale obiectelor în spațiul mental. Aceste obiecte sunt create și poziționate folosind metode de învățare atât supravegheată cât și prin auto-organizare. Asociațiile între obiectele mentale se bazează pe distanța dintre ele, considerând nu doar proprietățile reprezentărilor dar și corelațiile spațio-temporale. În aceste condiții, intuiția este dependentă de topografia spațiului mental. Obținem astfel pentru raționamentul logic o evoluție dinamică în spațiul mental (activarea succesivă a unor obiecte mentale). Raționamentul logic și cel bazat pe reguli devin astfel o aproximare a dinamicii stărilor mentale.

1.4 Elaborarea sistemelor de IA integrate în structuri distribuite

Idea distribuirii reprezintă un curent novator în inteligența artificială, datorită reducerii complexității, prin descompunerea unei probleme, modularitate, siguranță în funcționare. Conceptul de inteligență artificială distribuită se bazează fundamental pe cel puțin trei abordări neexcluzive: paralelizarea sistemelor de inteligență artificială, (algoritmi și mașini paralele), sisteme multiagenți sau multibaze de cunoștințe, distribuirea totală a cunoștințelor și raționamentelor. La acest nivel apar doua

mecanisme foarte importante: cooperarea și competiția. IA distribuită este un subcâmp al IA care cercetează metodele de sinteză ale bazelor de cunoștințe distribuite, tehnicile de comunicare și raționament necesare pentru a participa la societăți alcătuite din oameni și calculatoare.

IA distribuită se preocupă cu situații în care o serie de sisteme interacționează pentru a rezolva o problemă comună. Obiectivul acestei discipline este de a construi cadrul teoretic și instrumentele pentru modelarea agenților cu capacități de planificare și comunicare. Termenul de „Inteligență Artificială Distribuită” este utilizat la prima manifestare științifică din domeniu în 1980, la MIT (The Workshop on Distributed AI). Aria IA distribuite este delimitată de două domenii de cercetare: rezolvitorii distribuiți de probleme (RDP – Distributed Problem Solvers) și sisteme multiagent. RDP au ca strategie în rezolvarea unor probleme de mari dimensiuni împărțirea sa în subprobleme până la nivelul task-urilor, care sunt ulterior alocate unor agenți (noduri și module).

Agenții formează o echipă în care componentele cooperează în utilizarea cunoștințelor despre problemă și evoluția modului de obținere a soluției. Cooperarea și coordonarea sunt încorporate ca funcții în sistem din momentul proiectării (abordare top-down). Sistemele multiagent au în centrul preocupărilor coordonarea comportamentului unei mulțimi de agenți autonomi, care în afara unor obiective individuale proprii, au ca obiectiv rezolvarea unei probleme comune (bottom-up). Un sistem multiagent poate fi definit ca o rețea slab cuplată de rezolvitori bazați pe cunoștințe, care lucrează împreună pentru rezolvarea unei probleme a cărei soluționare depășește capacitatea individuală a oricărui participant. Rezolvitorii individuali de probleme (numiți agenți) sunt autonomi sau eterogeni (au capacități diferite de soluționare), fiecare având informații incomplete. Cooperarea este impusă de faptul că resursele (de prelucrare, memorare, comunicare), expertiza și informațiile necesare soluționării problemei au localizări diferite, iar prelucrările sunt asincrone. Evoluția indivizilor este determinată în mare măsură de apartenența lor la colectivități. Activitățile de colaborare din aceste colectivități conduc la evoluția grupului și oferă avantaje indivizilor care alcătuiesc comunitatea. Acest mediu social este studiat și simulat de Sistemele Multiagent, care studiază agentul ca o parte a spațiului social al altor agenți și nu numai lumea fizică sau un agent izolat. Informația socială include informația despre valori, strategii și intenții precum și informații despre starea lumii și reprezintă baza pentru introducerea conceptelor fundamentale pentru știința socială a informației și a sistemelor multiagent. Informația este activă prin participarea la crearea interacțiunilor dintre agenți. Agenții sunt entități care participă activ la spațiul social al interacțiunilor. Una din preocupările actuale ale IA este dezvoltarea de sisteme software care să trateze probleme din aria capacităților legate de „societăți” de agenți. Comportamentul inteligent presupune coordonare și comunicare. Prin coordonare se înțelege adaptarea la mediu (în care acționează o societate de agenți) realizată prin modificarea comportamentului agenților. Mediul este suport pentru comunicare și mediază transferul informației de coordonare. Agenții sunt rezolvitori autonomi de probleme, având obiective, intenții, capacități și cunoștințe proprii, care își desfășoară acțiunile conform unor planuri. Planurile sunt mulțimi de acțiuni raționale ale agenților. Teoriile formale ale sistemelor multiagent sunt specificații ale agenților care oferă o bază pentru proiectare, urmată de implementarea și verificarea acestora. Abordările multiagent reprezintă o direcție nouă, distinctă de ingineria software, care iau în considerare sisteme distribuite complexe ce conțin societăți de agenți autonomi cooperanți.

Diferitele abordări specifice s-au făcut în raport cu:

- Nivelul ridicat pe care îl oferă tehnologia actuală. Nivelul tehnologic atins în capacitatea de prelucrare, memorare și comunicare dă posibilitatea cuplării asincrone a unui număr mare de agenți. Exploatarea eficientă a acestor structuri abordări teoretice și practice specifice în cazul în care tehnicile consacrate nu dau rezultate satisfăcătoare.
- Distribuția inerentă. Majoritatea aplicațiilor de inteligență artificială sunt distribuite prin natura lor. Utilizarea sistemelor multiagent oferă o serie de avantaje comparativ cu sistemele monolitice singulare: rapiditate (prin exploatarea paralelismului), reducerea traficului (prin comunicarea de

rezultate parțiale și nu a datelor primare către un centru de prelucrare), creșterea flexibilității (agenți cu capacități diferite pot forma rapid echipe pentru rezolvarea de tipuri de probleme, obișnuit diferite) și creșterea fiabilității (prin preluarea de către alți membri ai echipei a sarcinilor agenților care cad).

- Rațiuni epistemologice (modele cognitive ale cooperării). Cooperarea și coordonarea sunt fenomene complexe. IA distribuită poate oferi o mai bună înțelegere asupra interacțiunilor dintre oameni, asupra modului cum aceștia se organizează în grupuri pentru a rezolva probleme.

IA distribuită consideră că entitățile componente dispun de inteligență spre deosebire de sistemele conexionate în care componentele individuale de calcul se combină pentru a forma un întreg inteligent. Acest domeniu nu tratează probleme referitoare la procesarea paralelă sau distribuită ci se ocupă de problema coordonării dintre procese concurente la diferite niveluri de reprezentare și rezolvare. Ca urmare, inteligență artificială distribuită nu este interesată de procesarea paralelă pentru IA din rațiuni de eficiență a calcului. Eforturile sunt îndreptate spre rezolvarea problemelor de eficiență a sistemelor de IA și spre înțelegerea conceptuală a raționamentului și comportamentului inteligent într-o mulțime de agenți. În IA distribuită nu sunt tratate probleme legate de arhitecturi de calcul paralel sau limbaje de programare paralelă.

Domeniul IA distribuite impune abordări pentru structurarea și organizarea efortului de cercetare, din:

- Perspectiva agentului care grupează acele elemente ce caracterizează agentul implicat în sistemul multiagent;
- Perspectiva grupului (se referă la elementele ce caracterizează grupul de agenți) poate fi abordată din mai multe puncte de vedere: organizarea grupului; coordonarea; cooperarea; negocierea; coerența comportamentului; planificarea; comunicarea.
- Perspectiva unor abordări specifice – se referă la probleme privind sistemele deschise, reflecție, agenți autonomi și sisteme informatice organizaționale.
- Perspectiva proiectantului – se referă la metode și tehnici de implementare referitoare la construirea sistemelor multiagent, standarde de test și proiectarea aplicațiilor de inteligență artificială distribuită.

Pentru crearea unui agent inteligent reprezentarea cunoștințelor este o problemă crucială. Ca orice aplicație de inteligență artificială, ceea ce se așteaptă de la un agent să facă și în ce domeniu, va avea un impact semnificativ asupra tipului de reprezentare a cunoștințelor. Dacă agentul are un număr limitat de situații la care trebuie să răspundă, atunci inteligența sa poate fi dată de programe procedurale. Dacă în schimb, agentul are de construit sau de folosit modele sofisticate ale domeniului problemei și de rezolvat probleme la diferite niveluri de abstractizare, atunci se impune folosirea rețelelor semantice. Dacă agentul are de răspuns la întrebări sau trebuie să genereze noi fapte din datele existente, atunci trebuie folosite logica predicatelor sau reguli if-then. Dacă agentul va interacționa cu alți agenți și trebuie să partajeze cunoștințe, atunci va trebui probabil să citească și să scrie date de tip Knowledge Interchange Format (KIF). În multe aplicații se folosesc simultan mai multe reprezentări ale cunoștințelor. Dacă pentru mediile comerciale de software agenții inteligenți sunt o noutate, ei au fost în atenția cercetătorilor de mai mult timp, fiind propuse mai multe moduri de clasificare a agenților: prin încadrarea lor în contextul inteligenței, agentificării și mobilității; prin focalizarea asupra strategiei primare de procesare a agentului; după funcțiile pe care le îndeplinesc. O descriere relevantă a noțiunii de agent poate fi dată prin enumerarea caracteristicilor de bază pe care un agent trebuie să le posede:

1. **Autonomie:** un agent operează fără intervenție umană, are inițiative și exercită control asupra propriilor lui acțiuni. Agentul acceptă cereri de la utilizatori dar este cel care decide cum și unde va satisface aceste cereri. În același timp agentul nu răspunde orbește comenzilor primite ci colaborează cu utilizatorul sau alți agenți pentru satisfacerea cerințelor acestora, putând să propună modificări sau să ceară clarificări suplimentare.
2. **Reactivitate:** agenții percep mediul în care evoluează (care poate fi lumea fizică, un utilizator prin intermediul unei interfețe, o colecție de alți agenți sau rețeaua Internet) și răspund în timp util la schimbările sesizate în acest mediu.
3. **Comunicare și abilitate socială:** un agent trebuie să fie capabil să se angajeze în comunicări și dialoguri complexe cu alți agenți sau persoane pentru a obține informații sau ajutor în satisfacerea scopurilor.

Pentru anumiți cercetători, în special cei din domeniul IA, termenul de agent implică și alte caracteristici suplimentare cum ar fi:

- noțiuni mentale de atitudine (părerii, dorințe, obligații, opțiuni);
- noțiuni mentale epistemice (convingerii, cunoștințe);
- raționalitate;
- adaptabilitate sau capacitate de învățare.

Aceste caracteristici nu sunt exhaustive și, de altfel, nu există agenți inteligenți care să le înglobeze pe toate. Deocamdată nu există un consens referitor la importanța relativă a diverselor proprietăți ale agenților, dar cei mai mulți cercetători recunosc că cele amintite ar fi caracteristicile definitorii ale unui program agent față de un program obișnuit. Componentele cheie care apar în următoarele două scenarii pot fi considerate a fi agenți inteligenți. Sistemul de control al traficului aerian din țara X s-a defectat din cauza condițiilor meteorologice. Din fericire, sistemele computerizate de control al traficului aerian din țările învecinate sesizează pericolul (reactivitate), negociază între ele (abilitate socială) pentru a coordona toate zborurile afectate și situații potențial periculoase sunt evitate fără intervenția omului (autonomie). Un utilizator se conectează la un terminal în contul lui și i se prezintă de îndată o listă a mesajelor electronice, sortată în ordinea importanței, de către agentul asistent personal al utilizatorului (Personal Digital Assistant). De asemenea, i se prezintă o listă a noilor articole apărute, în care PDA evidențiază un articol particular (reactivitate) care descrie rezultate recente relevante domeniului de cercetare al utilizatorului (domeniu indicat sau învățat de agent pe parcursul cooperării cu utilizatorul). După o discuție electronică cu alți PDA (abilitate socială și comunicare), PDA-ul utilizatorului în cauză a obținut deja textul acestui articol, anticipând astfel dorința utilizatorului (autonomie). Când se vorbește de agenți software există trei dimensiuni de evaluare a abilităților lor: agentificare, inteligență și mobilitate. Agentificarea reprezintă gradul de autonomie pe care agentul software îl are în reprezentarea cunoștințelor și intereselor utilizatorului în fața altor agenți, aplicații și sisteme de calcul. Un agent reprezintă utilizatorul, îl ajută, îl ghidează și în unele cazuri acționează în numele utilizatorului. Inteligența se referă la abilitatea agentului de a capta și de aplica cunoștințe specifice domeniului aplicației pentru a rezolva probleme. Modelul sau arhitectura agenților poate varia de la structuri foarte simple, gen automate, cum este cazul agenților *reactivi*, până la structuri de cunoștințe complexe, cu reprezentări sofisticate, capacități de raționament și decizie, structuri întâlnite în cazul agenților *cognitivi*, numiți și agenți *raționali*.

Agenții cognitivi conțin o reprezentare simbolică explicită a lumii și sunt capabili să ia decizii (ce acțiuni să execute) pe baza unui raționament simbolic. Ei au scopuri fixate și sunt capabili să-și modifice acțiunile în funcție de schimbările din mediu și interacțiunea cu alți agenți, în vederea realizării acestor scopuri. Fiecare agent cognitiv este un sistem bazat pe cunoștințe, cu un grad mai mare sau mai mic de rezolvare și care include întreaga metodologie asociată lor. În cazul sistemelor de

agenți cognitivi, inteligența sistemului este dată atât de suma comportamentelor inteligente individuale ale fiecărui agent cât și de inteligența rezultată din comportamentul colectiv bazat pe interacțiuni de tip social.

Un sistem de agenți cognitivi este, de exemplu, un planificator automat de ocupare a unei săli de reuniuni aparținând unei mari companii. Apariția sistemelor deschise și descentralizarea proiectării distribuite a impus o nouă abordare a rezolvării distribuite a problemelor, respectiv *sistemele multiagent* (SMAg). Un sistem multiagent este un sistem distribuit format dintr-o colecție de agenți autonomi care interacționează într-un mediu comun, fiecare agent având cunoștințe, capacități de acțiune și scopuri proprii. De cele mai multe ori, sistemele multiagent sunt sisteme inteligente. Se pot identifica două categorii de sisteme multiagent inteligente: SMAg cu agenți cognitivi, numite și sisteme multiagent cognitive, și SMAg cu agenți reactivi, numite și sisteme multiagent reactive.

Sistemele multiagent cognitive încearcă să simuleze elemente ale modelului uman de comportare prin includerea noțiunilor de scop, cooperare, competiție, organizare în structuri sociale și stabilirea relațiilor de dependență în aceste structuri, capacitate de învățare și auto-perfecționare. Un sistem multiagent cognitiv poate fi văzut ca un sistem particular bazat pe cunoștințe ce conține o reprezentare simbolică a cunoștințelor despre lumea în care evoluează și care este capabil să ia decizii (de exemplu ce acțiune să execute) prin intermediul unui proces inferențial (de raționament). Spre deosebire de un sistem bazat pe cunoștințe clasic, un agent într-un SMAg reprezintă simbolic lumea atât prin convingeri, care sunt păreri despre lume, eventual incomplete sau eronate, cât și prin cunoștințe, ce reprezintă fapte adevărate despre acea lume. În plus, un agent trebuie să fie capabil să raționeze atât despre propriile lui convingeri și cunoștințe cât și despre cele ale celorlalți agenți cu care interacționează într-un anumit mediu. Sistemele multiagent (SMAg) reactive sunt sisteme în care agenții sunt unități simple de prelucrare, capabile să reacționeze la schimbările de mediu și să execute acțiuni simple în consecință. Modelul este în principal inspirat din structura comunităților biologice de insecte: o albină nu poate fi considerată inteligentă dar comportamentul stupului ca întreg și chiar modul de organizare a albinelor în acesta prezintă cert elemente de inteligență, executând acțiuni comune și coordonate în vederea realizării scopului. Conform adepților SMAg reactive, inteligența reală este situată în lumea exterioară și nu la nivelul componentelor de prelucrare individuală, agenții, comportarea inteligentă fiind un rezultat al interacțiunii dintre agenți și mediu, iar inteligența o proprietate emergentă a sistemului ca întreg.

Multe din modelele agenților specifici IA conțin o reprezentare internă a lumii, existând o stare mentală explicită care poate fi modificată, utilizând diferite forme ale raționamentului simbolic. Agenții care păstrează o reprezentare internă a domeniului problemei și care realizează o transformare explicită în spațiul de stare al sistemului, utilizând anumite forme de raționament simbolic, se numesc agenți deliberativi, a căror modelare se bazează pe încredere, dorința și intenții. Aceasta abordare descrie starea internă a agentului prin intermediul unei mulțimi de categorii mentale și definește o arhitectură de control prin care agentul selectează în mod rațional evoluția sa în spațiul de stare. Categoriile mentale sunt încredere, dorința și intenții, care au fost suplimentate din punct de vedere practic cu noțiunile de scopuri și planuri. Unul dintre cele mai simple tipuri de agenți sunt agenți reactivi sau reflexi, care răspund în sensul eveniment – condiție – acțiune. Agenți reflexi nu au modele interne ale lumii. Ei răspund doar la stimuli externi și la informația disponibilă lor de la senzorii pentru mediu. Ca și rețelele neuronale, agenți reactivi desfășoară un comportament emergent, care este rezultatul interacțiunilor acestor agenți individuali simpli. Atunci când agenți reactivi interacționează, ei schimbă date de nivel scăzut. Aplicațiile agenților s-au limitat la roboți care folosesc senzori pentru percepere lumea. Iar conceperea agenților reactivi, care iau decizii în timpul execuției lor, fiind bazați pe o serie de situații curente este parțial ghidată de ipoteza lui Simon, conform căreia complexitatea comportării unui agent poate fi o reflectare a complexității mediului în care agentul operează, mai degrabă decât o reflectare a complexității interne de concepție a acestuia.