

Cursul 1

Absolut. Conform materialului din curs, definirea **Inteligenței Artificiale (AI)** este complexă și este abordată din patru perspective majore.

Aceste perspective clasifică obiectivele AI fie pe baza **proceselor de gândire (raționament)**, fie pe baza **comportamentului (acțiune)**, și fie pe baza fidelității față de **om**, fie pe baza fidelității față de un standard **rațional (logic)**.

Cele Patru Perspective Asupra Inteligenței Artificiale

1. Acționând Uman (Acting Humanly)

Această abordare se concentrează pe măsurarea performanței unui sistem AI în raport cu performanța umană.

- **Scop:** Să creeze sisteme care se comportă (acționează) într-un mod indistinguibil de cel uman.
- **Măsură cheie: Testul Turing** (propus de Alan Turing în 1950).
 - **Descriere:** Un om (interogator) interacționează cu două entități invizibile (un om și un computer) printr-un terminal text. Dacă interogatorul nu poate distinge în mod fiabil care este omul și care este computerul, se spune că programul AI a trecut testul.
- **Capabilități necesare pentru a trece testul:**
 - **Procesarea Limbajului Natural (NLP):** Pentru a comunica eficient.
 - **Reprezentarea Cunoștințelor:** Pentru a stoca informații pe care le poate folosi.
 - **Raționament Automat:** Pentru a răspunde la întrebări și a trage concluzii.
 - **Învățarea Automată (Machine Learning):** Pentru a se adapta și a dobândi noi cunoștințe.
 - *(Pentru Testul Turing complet ar fi nevoie și de Viziune și Robotică, dar cel original se baza doar pe text.)*

2. Gândind Uman (Thinking Humanly)

Această abordare este o ramură a **Științei Cognitive** și își propune să înțeleagă modul în care funcționează mintea umană și să replice acele mecanisme de gândire.

- **Scop:** Să creeze sisteme care gândesc ca oamenii, replicând procesele cognitive biologice.
- **Metodă:** Necesită o modalitate de a determina cum gândesc oamenii, cum ar fi:
 - **Introspecția:** Încercarea de a ne observa propriile gânduri.
 - **Experimente Psihologice:** Măsurarea reacțiilor și luarea deciziilor în diverse scenarii.
 - **Scanări ale creierului (fMRI, EEG):** Observarea activității cerebrale.
- **Provocare:** A crea o teorie suficient de precisă a minții pentru a fi implementată pe un computer.

3. Gândind Rațional (Thinking Rationally)

Această abordare se bazează pe **Logică** și urmărește să creeze sisteme care folosesc un lanț de raționament corect (deductiv).

- **Scop:** Să creeze sisteme care raționează logic, conform regulilor inferenței.
- **Bază: Legile Gândirii (Laws of Thought)**, formulate în logica formală (ex. silogismele lui Aristotel).
- **Sistem:** Un program folosește cunoștințe codificate în notăție logică și aplică reguli de inferență pentru a trage concluzii.
- **Avantaj:** Garanția că, dacă premisele sunt adevărate, concluzia este, de asemenea, adevărată.
- **Provocări:**
 - Nu este simplu să traduci cunoștințele informale în termeni logici formali.
 - Există o diferență între a rezolva o problemă **în principiu** (logic) și a o rezolva **în practică** (problema este tractabilă din punct de vedere computațional).

4. Acționând Rațional (Acting Rationally)

Aceasta este abordarea care domină AI-ul modern și se concentrează pe construirea de **Agenți Raționali**.

- **Scop:** Să creeze sisteme care acționează pentru a-și atinge cel mai bine obiectivele (sau pentru a maximiza **utilitatea** așteptată).
- **Concept cheie:** Un **Agent Rațional** este un sistem care acționează pentru a obține cel mai bun rezultat sau, în condiții de incertitudine, cel mai bun rezultat așteptat.
- **Diferența față de "Gândind Rațional":** A acționa rațional nu necesită întotdeauna un raționament logic complet. Uneori, un reflex sau o acțiune rapidă este mai rațională decât o analiză îndelungată (ex. a te feri de o mașină).
- **Focus:** Pe **inferență corectă** (tragerea de concluzii) combinată cu **acțiune corectă** (luarea deciziei optime).

Concepte Filozofice Conexe

Strong AI (AI Puternică) vs. Weak AI (AI Slabă)

- **Weak AI (AI Slabă):** Premisa că mașinile pot acționa ca și cum ar fi inteligente, dar nu posedă o inteligență, conștientizare sau înțelegere reală (cum ar fi un om). Majoritatea sistemelor AI de astăzi sunt considerate Weak AI.
- **Strong AI (AI Puternică):** Premisa că un computer nu doar simulează inteligența, ci este de fapt o minte capabilă de gândire, conștiință și înțelegere reală.

Argumentul Camerei Chineze (The Chinese Room Argument)

Acesta este un contra-argument celebru la conceptul de Strong AI, propus de John Searle.

- **Idee:** O persoană care se află într-o cameră și urmează un set de reguli (un algoritm) pentru a manipula simboluri chinezești (fără a le înțelege) poate răspunde perfect la întrebările în chineză.
- **Concluzie:** Deși persoana (sistemul AI) acționează inteligent, ea nu înțelege limba (nu gândește uman). Astfel, simpla manipulare a simbolurilor (programare) nu este echivalentă cu înțelegerea sau inteligența reală (Strong AI).

Connectionism vs. Computationalism

Discută natura inteligenței.

- **Connectionism:** Inteligența este o consecință a unei structuri simple de bază, emergentă din interacțiunile neuronilor (de exemplu, Rețelele Neurale).
- **Computationalism (Simbolism):** Inteligența este un scop, realizat prin îndeplinirea anumitor funcții (de exemplu, manipularea simbolurilor logice).

Absolut. Continuăm cu o explicație detaliată a **Pilonilor AI (Componente) și Abordărilor (Modele de Decizie)**, conform materialului de curs.



Pilonii AI Moderne (The Three Pillars of Modern AI)

Inteligența Artificială modernă este construită pe o bază solidă de trei domenii complementare, fiecare esențial pentru funcționarea unui sistem inteligent.

1. Modelare (Modeling)

Acest pilon se referă la capacitatea sistemului AI de a reprezenta și de a lucra cu datele și informațiile din lumea reală. Un motor AI trebuie să poată descrie datele lumii reale.

- **Scop:** Crearea unei reprezentări structurate (un **model**) a lumii cu care lucrează AI.
- **Domenii de bază utilizate:**
 - **Structuri de Date (Data Structures):** Modul în care informațiile sunt organizate eficient.
 - **Baze de Date (Databases):** Stocarea și regăsirea informațiilor la scară mare.
 - **Algoritmi de Grafuri (Graph Algorithms):** Reprezentarea relațiilor complexe (de exemplu, hărți, rețele sociale, ontologii).
 - **Limbaje Formale, Automate:** Pentru definirea regulilor și a sintaxei.
 - **Sisteme Bazate pe Cunoștințe (Knowledge-Based Systems):** Reprezentarea cunoștințelor explicite despre domeniu.

2. Raționament (Reasoning)

Acest pilon reprezintă capacitatea sistemului AI de a manipula modelul (cunoștințele) pentru a deduce și a descoperi informații noi care nu erau disponibile explicit la început.

- **Scop:** Descoperirea de date sau concluzii **noi** din datele și regulile disponibile.
- **Domenii de bază utilizate:**
 - **Logica pentru Informatică (Logic for Computer Science):** Utilizarea regulilor de inferență pentru a trage concluzii.

- **Design de Algoritmi (Algorithm Design):** Dezvoltarea de metode eficiente pentru a căuta soluții sau a deduce informații.
- **Algoritmi de Grafuri:** Folosiți, de asemenea, pentru a naviga și a face inferențe în structurile de cunoștințe.
- **Exemplu:** Dacă sistemul știe că "Toți oamenii sunt muritori" și "Socrate este un om," el trebuie să poată **raționa** că "Socrate este muritor."

3. Învățare (Learning)

Acest pilon se referă la capacitatea sistemului AI de a se îmbunătăți, de a-și adapta modelul și de a dobândi cunoștințe noi pe baza experienței (datelor).

- **Scop:** Adaptarea modelului pentru contexte particulare sau necunoscute anterior, pe baza datelor empirice.
- **Domenii de bază utilizate:**
 - **Probabilități și Statistică:** Fundația matematică pentru a face predicții și a gestiona incertitudinea.
 - **Machine Learning (Învățarea Automată):** Algoritmi care permit computerului să se îmbunătățească pe baza datelor (Supervizat, Nesupervizat, prin Consolidare).
 - **Rețele Neurale (Neural Networks):** Modele inspirate din structura creierului, esențiale în Deep Learning.



Abordări (Sisteme de Decizie)

Materialul de curs clasifică abordările AI în funcție de modul în care sistemul ia decizii și rezolvă problemele.

1. Modele Bazate pe Stări (State-based Models)

Aceste modele se concentrează pe **starea** curentă a problemei și pe modul în care o succesiune de **acțiuni** transformă această stare pentru a atinge un **obiectiv**.

- **Idee Principală:** Sistemul își reprezintă lumea ca un set de stări posibile și o hartă a tranzițiilor (acțiunilor) între ele.
- **Întrebarea Centrală:** Cum se schimbă starea curentă pentru a ajunge la starea finală dorită (obiectivul)?
- **Domenii Acoperite:**
 - **Strategii de Căutare (Search Strategies):** Căutarea în lățime, căutarea în adâncime, căutarea A*.
 - **AI pentru Jocuri (Game Playing AI):** Jocuri pe tablă (șah, Go), unde fiecare mutare schimbă starea jocului.
- **Exemplu:** Găsirea celei mai scurte rute de la Orașul A la Orașul B. Stările sunt orașele, iar acțiunile sunt drumurile.

2. Modele Bazate pe Variabile (Variable-based Models)

Aceste modele se concentrează pe relația matematică sau logică dintre **variabilele** de intrare (input) și variabila de ieșire (output).

- **Idee Principală:** Sistemul caută o funcție (f) care, atunci când este aplicată datelor de intrare (X), produce rezultatul dorit (Y), adică $Y = f(X)$.
- **Întrebarea Centrală:** Ce funcție/relație poate fi aplicată datelor pentru a obține rezultatul?
- **Domenii Acoperite:**
 - **Machine Learning (Învățarea Automată):** (Regresie, Clasificare) unde funcția este "învățată" din date.
 - **Probleme de Satisfacere a Constrângerilor (Constraint Satisfaction Problems - CSP):** Unde funcția este o mulțime de constrângeri logice (ex. Sudoku, planificarea orarelor).
- **Exemplu:** Prezicerea prețului unei case pe baza variabilelor (mărime, locație, număr de camere). Funcția este modelul de regresie.

Abordările (Sistemele de Decizie) în Inteligența Artificială clasifică modul în care un sistem rezolvă problemele și ajunge la un obiectiv. Materialul de curs evidențiază două tipuri principale de modele: **Modelele Bazate pe Stări** și **Modelele Bazate pe Variabile**.

1. Modele Bazate pe Stări (State-based Models)

Această abordare se concentrează pe o secvență de acțiuni care modifică starea problemei până când se ajunge la starea dorită (obiectivul).

- **Întrebarea Centrală:** În ce mod ar trebui să schimb starea curentă a problemei pentru a mă apropia sau a ajunge la obiectiv?
 - **Mecanismul de Rezolvare:** Calcularea unei soluții ca o **secvență de tranziții** (un algoritm) care pornește de la o stare inițială și se termină în starea scop.
 - **Domenii Acoperite:** Este abordarea majoritară pentru **strategiile de căutare, sistemele de raționament și AI pentru jocuri**.
 - **Concepte asociate (Săptămânile 2 și 3):**
 - Probleme (de decizie).
 - Reprezentarea stărilor.
 - Strategii de căutare.
 - Spații de probleme deterministe și non-deterministe.
 - **Exemplu (vizual):** Un arbore sau un graf care arată stările și tranzițiile posibile de la o stare la alta.
-

2. Modele Bazate pe Variabile (Variable-based Models)

Această abordare se concentrează pe găsirea unei formule sau funcții care poate fi aplicată datelor problemei pentru a genera direct rezultatul dorit (obiectivul).

- **Întrebarea Centrală:** Ce formulă (funcție) poate fi aplicată datelor problemei pentru a genera obiectivul?

- **Mecanismul de Rezolvare:** Se pornește de la formule arbitrare, se testează pe rezultatele așteptate și se ajustează în consecință.
 - **Domenii Acoperite:** Este abordarea majoritară pentru **machine learning** și **probleme de satisfacere a constrângerilor** (Constraint Satisfaction Problems - CSP).
 - **Concepte asociate (Săptămânile 4 și 6):**
 - Probleme de Satisfacere a Constrângerilor (Week 4).
 - Rețele Neurale (Week 6).
 - **Exemplu:** O ecuație matematică sau o rețea neuronală care primește date (variabile) ca intrare și returnează o valoare (obiectivul) ca ieșire .
-

Sistemul de Decizie Inteligent

Pentru a fi un sistem de decizie rezonabil și practic, acesta trebuie să fie capabil să îndeplinească următoarele cerințe:

- **Să înțeleagă un model și un obiectiv.**
- **Să recunoască și să poată folosi mijloacele** pentru a atinge obiectivul.
- **Să decidă dacă și când să utilizeze acele mijloace.**
- **Să furnizeze un răspuns satisfăcător** pentru obiectivul stabilit.

Doriți să trecem la subiectul următor, "**Brief history of conversational agents**" (Scurt istoric al agenților conversaționali)?

Continuăm cu secțiunea despre **Agenții Conversaționali** (Conversational Agents) și evoluția lor, așa cum este prezentată în materialul de curs.

Scurt Istoric al Agenților Conversaționali

Istoria agenților conversaționali poate fi împărțită în trei etape majore, reflectând progresul tehnologic de la simple potriviri de șabloane la modele complexe de învățare profundă.

1. Anii '50 - '70: Template-uri și Expresii Regulate

În această etapă timpurie, agenții foloseau reguli predefinite și șabloane de text pentru a genera răspunsuri.

- **Mecanism:** Utilizarea de **template-uri** (șabloane) și **expresii regulate (regex)** pentru a detecta anumite structuri în intrarea utilizatorului și a genera un răspuns corespunzător
- **Exemplu AIML:** Un exemplu dat este cel din **AIML (Artificial Intelligence Markup Language)**, unde un șablon simplu asociază un model de intrare cu un șablon de ieșire⁴:
 - **Input:** HELLO ALICE
 - **Output:** Hello User

- **Exemplu Regex:** Un alt exemplu este o expresie regulată care ia o frază interogativă și o transformă într-o afirmație
 - Input: "Are you sentient?"
 - Output: "Indeed, I am sentient."
- **Limitare:** Acești agenți erau foarte rigizi și nu aveau o "înțelegere" reală, bazându-se doar pe potrivirea de șiruri de caractere.

2. Anii '80 - 2000: Modele de Limbaj și Lanțuri Markov

Această etapă a introdus conceptul de probabilitate în generarea de text, permițând răspunsuri mai fluide și mai puțin rigide.

- **Mecanism:** Utilizarea **Modelelor de Limbaj** bazate pe **Lanțuri Markov**
- **Lanț Markov:** Un model statistic care prezice următorul cuvânt (sau stare) dintr-o secvență pe baza cuvântului (sau stării) curent.
- **Calcul:** Se calculează **probabilitatea** ca un cuvânt să vină după un alt cuvânt, pe baza frecvenței aparițiilor lor într-un corpus de antrenament
 - *Exemplu:* Dacă după cuvântul 'I' (Eu) apare doar 'am', probabilitatea este 1. Dacă după 'a' (un/o) apar trei cuvinte ('Boy', 'Girl', 'Star'), probabilitatea fiecăruia de a apărea este $\frac{1}{3}$

3. Starea Actuală: Transformeri și Învățare prin Consolidare

Aceasta reprezintă faza actuală, dominată de arhitecturi avansate de rețele neurale.

- **Mecanism:** Utilizarea arhitecturilor **Transformer** (care includ **Encoder** și **Decoder**) și integrarea **Învățării prin Consolidare (Reinforcement Learning)**¹⁶.
- **Transformer:** Este o arhitectură de rețea neuronală care folosește mecanismul de **atenție** pentru a pondera importanța diferitelor cuvinte dintr-o intrare, generând stări ascunse (Hidden States) care codifică sensul.
- **Învățare prin Consolidare (RL):** Este un tip de învățare în care agentul învață să ia decizii prin interacțiunea cu un mediu, primind **recompensă** pentru acțiunile dorite și **penalizare** pentru cele nedorite (deși RL este adesea utilizată pentru **alinarea** modelelor mari de limbaj la preferințele umane, nu neapărat pentru generarea textului în sine).
- **Exemplu:** Traducerea sau generarea de text, unde encoderul procesează intrarea (ex. "Optimus Prime is a cool robot") în stări ascunse, iar decodorul generează ieșirea în altă limbă (ex. japoneză).

● Concluzii (Interludiu)

Materialul de curs trage o concluzie clară cu privire la starea actuală a agenților conversaționali:

- **ChatGPT-urile sunt agenți conversaționali, nu AI sau LLM-uri** (Large Language Models) în sensul că sunt doar un tip de agent
- Agenții conversaționali actuali **nu gândesc**, ci doar **refolosesc raționamentul uman aproximativ**
- Ei trebuie să **"halucineze"** (să inventeze informații) pentru a acoperi lipsa inevitabilă de acoperire a cunoștințelor
- Ar trebui folosiți ca **motoare puternice de căutare și rezumare, editori și asistenți virtuali**
- Pentru proiectul de anul acesta, studenții sunt **obligați să le folosească**

Doriți să continuați cu următoarea secțiune, **"Defining AI"** (Definirea AI) care se concentrează pe **Cunoaștere vs. Inteligență** și **Connectionism vs. Computationalism**?