

Introducere în Reinforcement Learning

Cursul #1

2025 – 2026 (Sem. I)

Drd. Ștefan Iordache

Conf. Dr. Ciprian Păduraru

Facultatea de Matematică și Informatică

Universitatea din București



Cuprins



01

Organizatorice & Evaluare
Desfășurare & Examinare

02

Introducere
Ce înseamnă Reinforcement Learning (RL)?

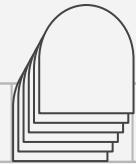
01

Organizatorice & Evaluare

Domnu' profesor, cum luăm
notă?



Structura cursului



Structură

→ Curs

- 2 ore/săptămână
- Miercuri: 18-20

→ Laborator

- 4 laboratoare/săptămână – hibrid (3 online, 1 fizic)

Detalii

- Prezență obligatorie?
Nu!
- Activitate cât mai mare?
Da!
- Examen teoretic?
Nu!
- Proiect?
Da!
- Când & cum?
La finalul semestrului (ultima săptămână), în echipe (3-5)

Cum luăm nota? (#1)

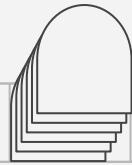
- Nota finală va fi formată 100% din nota proiectului!
- Echipe de minim 3 și maxim 5 persoane.
- Lucrul individual nu este permis, datorită structurii materiei și complexității proiectelor.
- Va fi deschisă o listă pentru notarea componenței echipelor și pentru propunerea unui subiect / unei teme de proiect.
- Prezentările vor fi stabilite pe parcursul mai multor sesiuni în ultima săptămână a primului semestru, pe durata laboratoarelor și cursului.

Cum luăm nota? (#2)

În ce constă realizarea proiectului:

- Alegerea unei teme de proiect (ex.: reinforcement learning într-un joc 2D / Unity);
- Implementarea sau editarea unui mediu existent;
- Implementarea a minim 3 algoritmi / agenți și compararea lor în același mediu;
- Rularea unor experimente multiple pentru calibrarea agenților;
- Evidențierea rezultatelor obținute;
- Documentarea implementării și experimentelor efectuate (PowerPoint, LaTeX, Word, etc.);
- Prezentarea proiectului la final de semestru, în echipă. Se urmărește evidențierea parcursului vostru de la idee la rezultate.

Tehnologii utilize



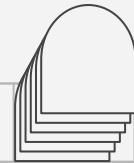
Avem libertate totală! 😊

Dar și câteva recomandări:

- Python 3.10+
- Jupyter Notebook
- Câteva librării de bază: PyTorch / TensorFlow / Stable-Baselines3
- Platforme de lucru: Colab / VS Code / PyCharm / ~~tablițe sumeriene~~



Scurtă bibliografie



- "*Biblia*" Reinforcement Learning
"Reinforcement Learning – An Introduction" – Richard S. Sutton & Andrew G. Barto
- Resurse suplimentare:
 1. Stanford CS234 (Emma Brunskill)
 2. Berkeley CS285 (Sergey Levine)
 3. DeepMind x UCL – RL Lectures (David Silver) - YouTube
 4. RLHF Papers (OpenAI 2022-2024)

02

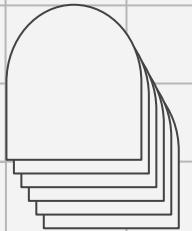
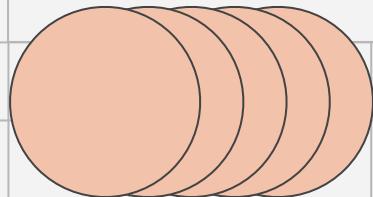
Introducere

Ce înseamnă Reinforcement Learning (RL)?



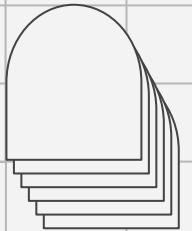
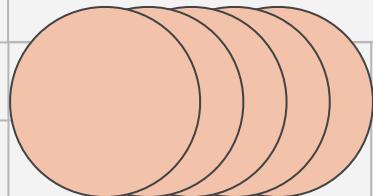
Premisa

Provocarea cea mai mare în inteligență artificială este **deducerea** unor **decizii bune** sub spectrul **incertitudinii**.



De ce RL?

- Multe sisteme reale nu au soluții perfecte, ci doar **acțiuni** mai bune în funcție de **experiența acumulată**.
- **Un agent RL încearcă să învețe cum să actioneze optim** într-un mediu necunoscut, prin **feedback (recompensă)**.



Cum funcționează deciziile?

Impact imediat sau întârziat?

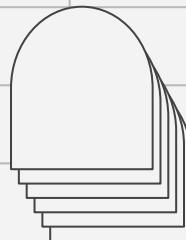
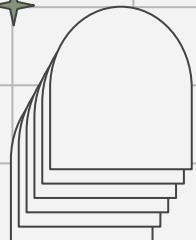
Ambele! În cazul oricărei decizii luate impactul va fi atât *imediat* cât și pe *termen lung*. Este necesar să cântărим beneficiile acțiunilor în ambele cazuri!

Ce înseamnă o decizie bună?

Problemele din lumea reală **nu** au întotdeauna o "cea mai bună soluție", în practică având nevoie de să definim **calitatea** unei *acțiuni* sau a unei *decizii*.

Avem la dispozitie toate datele?

Niciodată! În cazul problemelor de Reinforcement Learning **nu** avem un set de date complet, prestabilit, ci acesta **este dedus din interacțiunea cu mediul**.



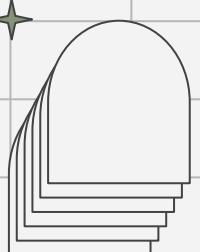
Exemple de decizii & impact



Impact imediat

**Apăsarea pedalei de acceleratie la
semafor**

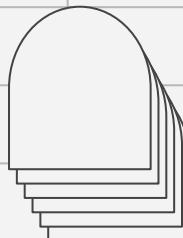
Care sunt efectele posibile?



Impact întârziat

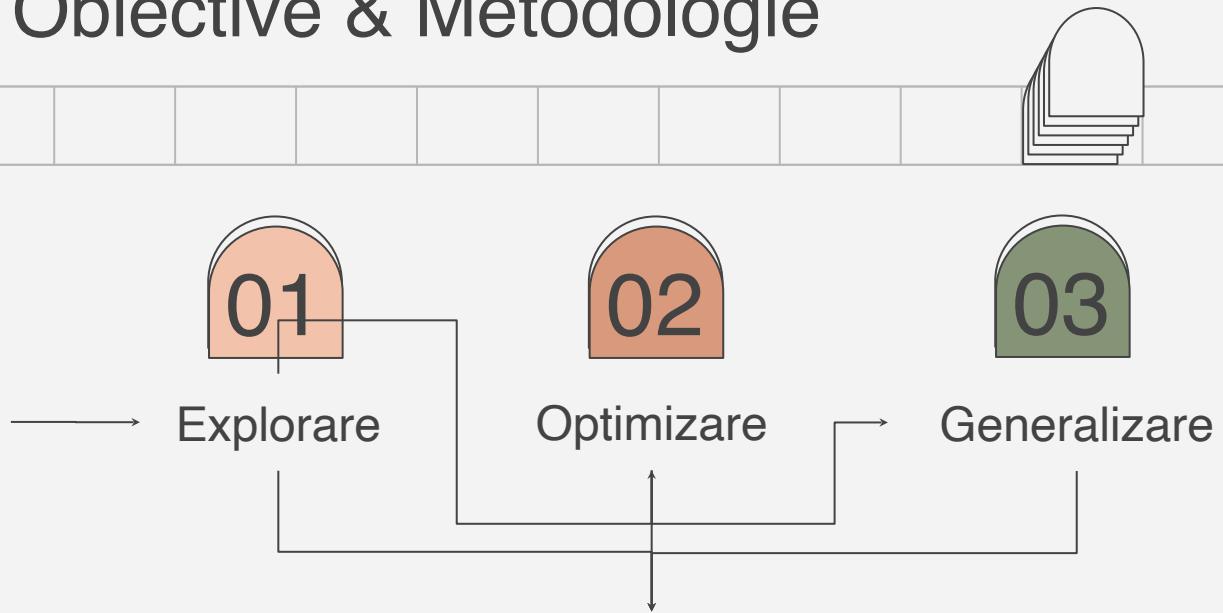
Planificarea unei cariere

Există recompensă imediată?



Obiective & Metodologie

Orice problemă
de RL



Metodologie:

- *Explorarea mediului*
- *Folosirea experienței pentru decizii viitoare*

RL vs. Supervised vs. Unsupervised

Supervised Learning

- Învățăm din date etichetate (labels)
- Scopul: prezicerea unor noi etichete
- Utilizări: sănătate, finanțial, marketing

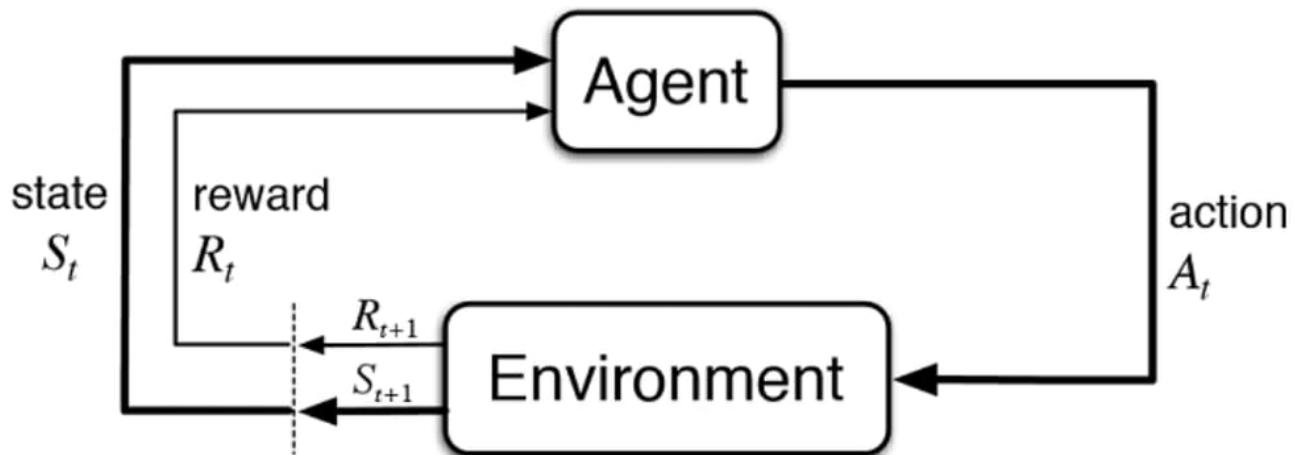
Unsupervised Learning

- Învățăm din date neetichetate
- Scopul: identificarea unor şabloane (patterns)
- Utilizări: securitate, retail, NLP

Reinforcement Learning

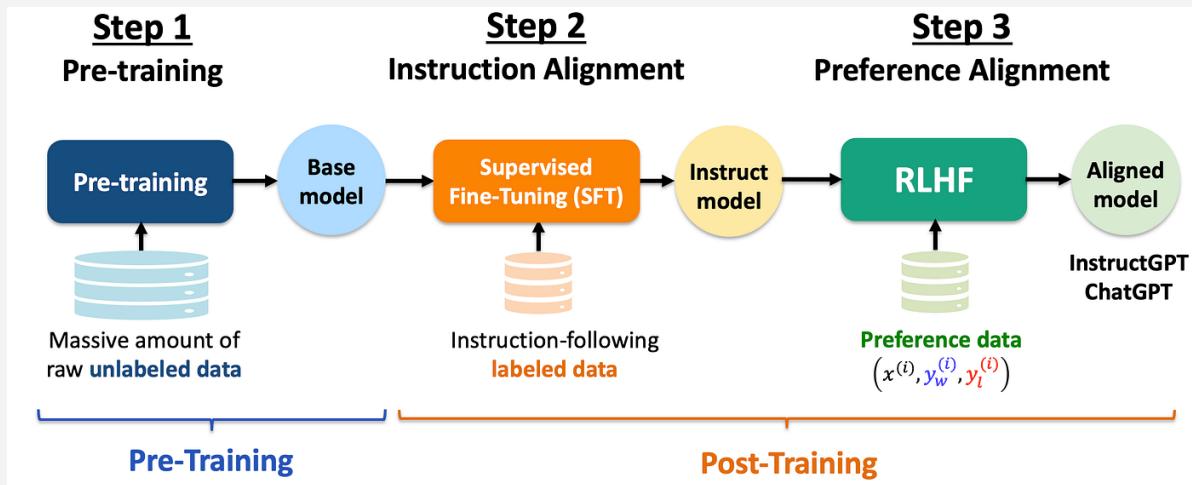
- Învățăm din interacțiune și recompense
- Scopul: optimizarea unui motor decizional
- Utilizări: robotică, gaming, finanțial

Diagrama generică - RL



RL în epoca noastră (LLM-uri)

RLHF – Reinforcement Learning from Human Feedback

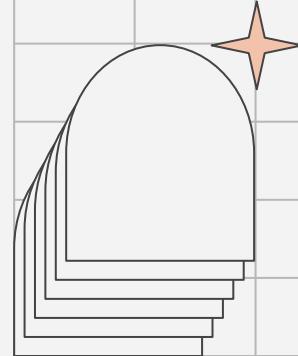
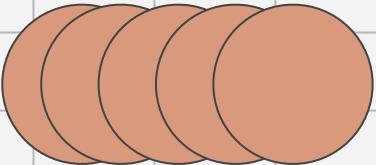


Thanks!

Este timpul pentru întrebări!!!

Acum...sau pe email:

stefan.iordache10@s.unibuc.ro



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#) and infographics & images by [Freepik](#)

