

Inteligență artificială și sisteme expert



P-ța Victoriei nr. 2
RO 300006 - Timișoara
Tel: +4 0256 403000
Fax: +4 0256 403021
rector@rectorat.upb.ro
www.upb.ro

Domeniul de studii: Informatică / Specializarea (Programul de studii): Tehnologii Informaticice

IASE – Cursul 1

Ș.I. dr.ing. Adriana ALBU

adriana.albu@upb.ro

<http://www.aut.upb.ro/~adrianaa>

Cine sunt eu?

- Adriana ALBU
- Departamentul de Automatică și Informatică Aplicată
 - Șef de lucrări
 - Director Departament
- Contact
 - adriana.albu@upt.ro
 - www.aut.upt.ro/~adrianaa
- Localizare – Clădirea Electro
 - birou B621



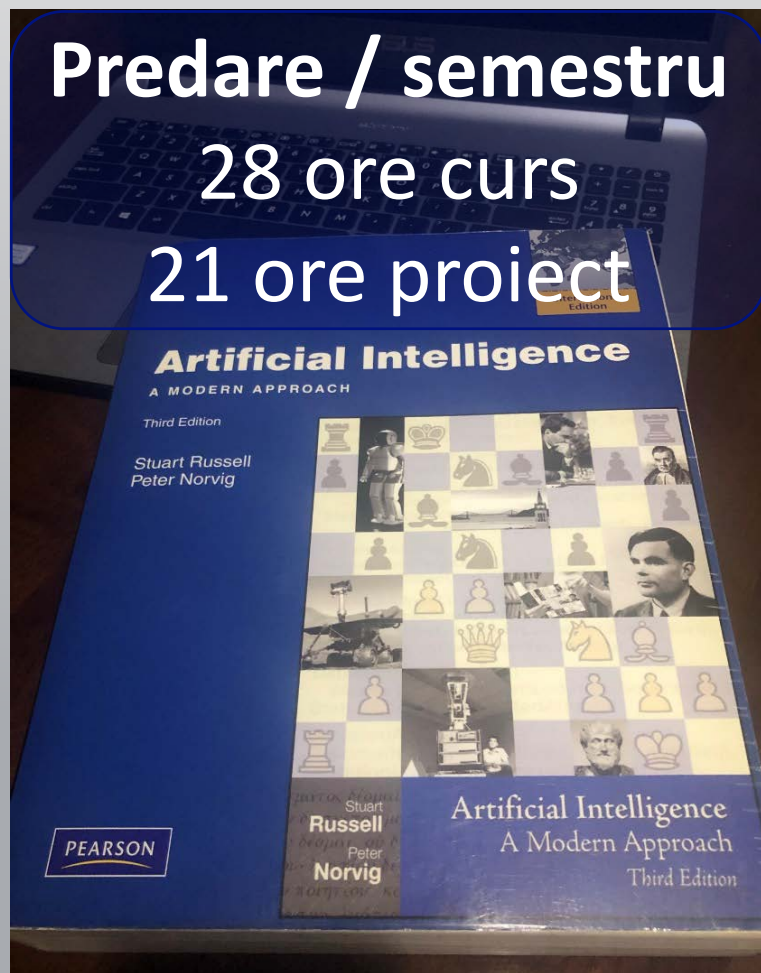
Organizare

www.cv.upt.ro

Predare / semestru

28 ore curs

21 ore proiect



Evaluare curs (NE)

Pregătirea și susținerea unei prezentări
pe o temă din domeniu IASE

Pondere 2/3 din nota finală

Activitate pe parcurs (NP)

Participarea la susțineri și evaluarea
prezentărilor (1.5 puncte / săpt.)

Pondere 1/3 din nota finală

Nota finală (NF) = $(2 \cdot NE + NP) / 3$

Bibliografie

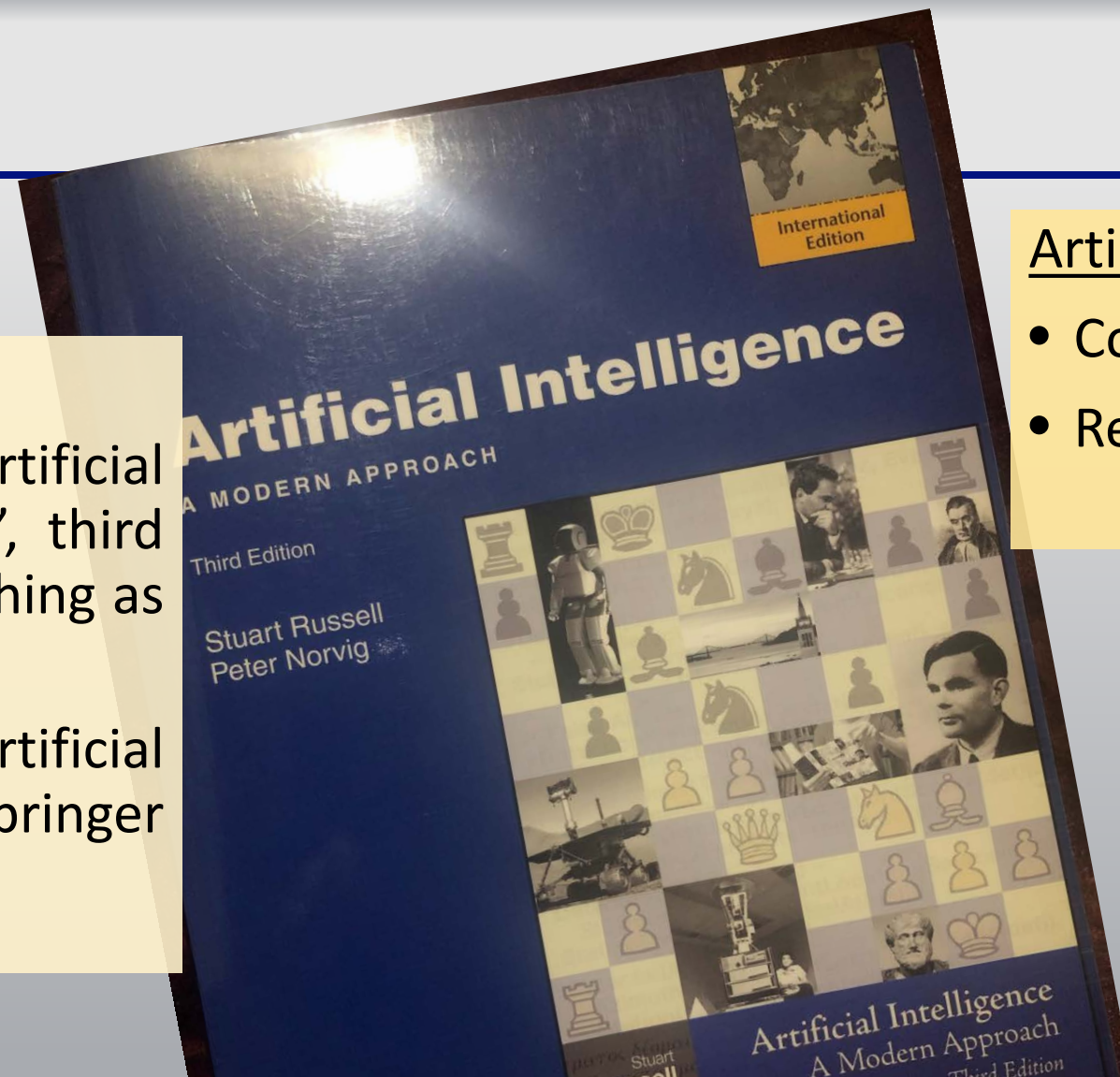
Cărți

- Stuart RUSSELL, Peter NORVIG, „Artificial Intelligence: A Modern Approach”, third ed., Pearson Education, Inc., Publishing as Prentice Hall, United States, 2010
- Wolfgang ERTEL, „Introduction to Artificial Intelligence”, second ed., Springer International Publishing AG, 2017

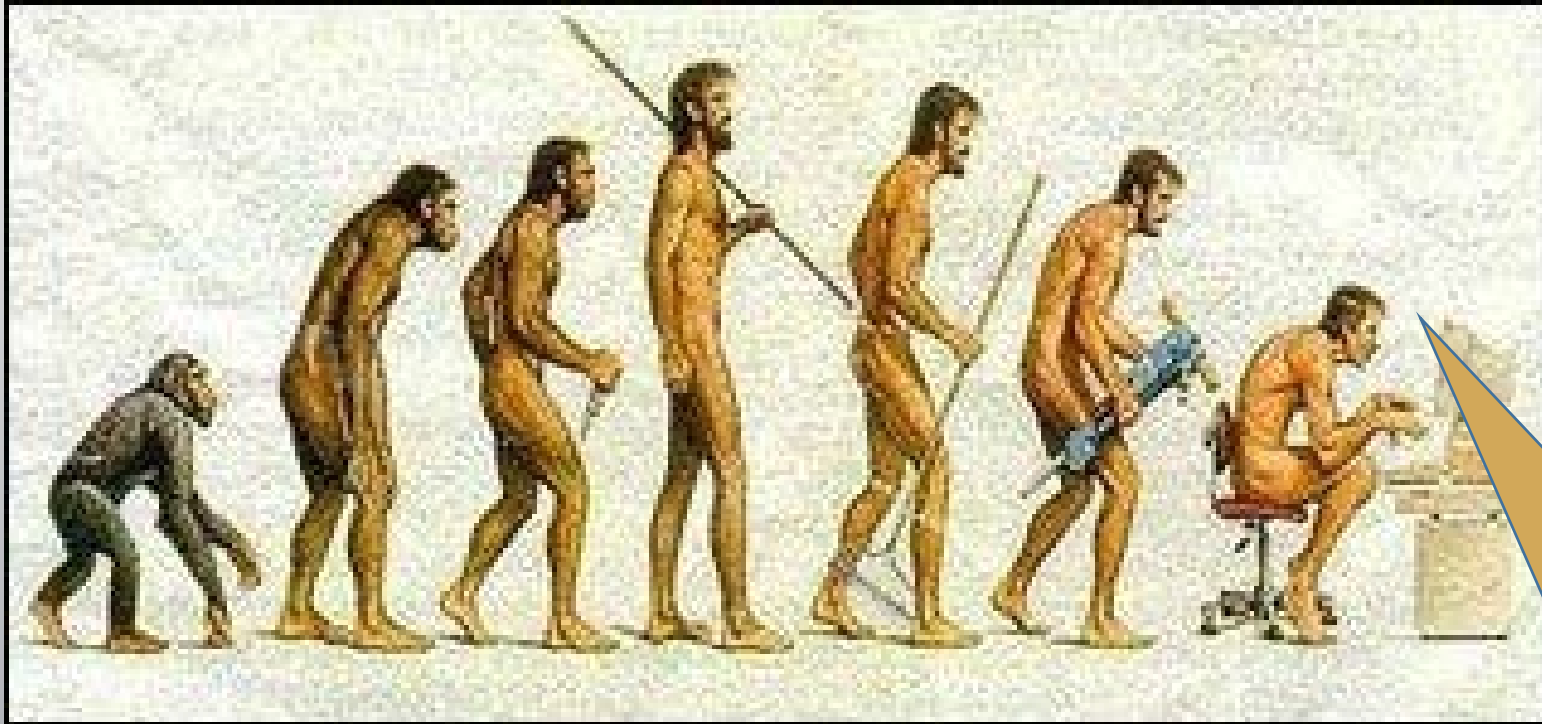
- MIT OpenCourseWare, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/>
- Alte www...

Articole

- Conferințe
- Reviste



Justificarea disciplinei



Somewhere, something went wrong? (Anonymous)

Homo sapiens

Inteligența umană
vs.
Inteligența artificială

Inteligența umană vs. inteligența artificială

- Homo sapiens
 - Homo = om, ființă umană
 - Sapiens = înțelept, rațional, inteligent
- Inteligența noastră – foarte importantă pentru noi
- Mii de ani am încercat să înțelegem felul în care gândim
 - cum o simplă mână de materie poate percepe, înțelege și manipula ceva mult mai mare și mai complex
- Inteligența artificială (IA) – merge mai departe
 - nu doar înțelege
 - dar și construiește entități inteligente

Definiții ale IA



John McCarthy (1927-2011)
introduce termenul IA în 1956

Thinking Humanly

“The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense.” (Haugeland, 1985)

“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)

Acting Humanly

“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)

“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)

Thinking Rationally

“The study of mental faculties through the use of computational models.”
(Charniak and McDermott, 1985)

“The study of the computations that make it possible to perceive, reason, and act.”
(Winston, 1992)

Acting Rationally

“Computational Intelligence is the study of the design of intelligent agents.” (Poole *et al.*, 1998)

“AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)

Example



Simptome
Analyze



Cunoștințe
Experiență



Diagnostic



„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Piața muncii în 2050 – influențată de învățare automată și robotică
 - disponibilizări sau generarea de noi locuri de muncă?
 - temerea exista și la începutul revoluției industriale (sec. XIX)
- Abilitățile umane – fizice și cognitive (învățare, analiză, comunicare, înțelegerea emoțiilor umane)
 - revoluția industrială a concurat abilitățile fizice ale omului
 - IA începe să-l întreacă în ceea ce privește abilitățile cognitive
 - alimentată și se inovațiile științelor naturale și sociale
 - înțelegerea mecanismelor biochimice aflate la baza emoțiilor, dorințelor și alegerilor umane conduc la creșterea performanțelor computerelor în:
 - analiza comportamentului uman
 - anticiparea deciziilor umane
 - => înlocuirea oamenilor (șoferi, medici etc.)

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Alegerile noastre
 - nu se datorează liberului arbitru sau intuiției
 - ci câtorva miliarde de neuroni care calculează probabilități într-o fracțiune de secundă
 - algoritmi biochimici ai creierului uman nu sunt perfecți
- Șoferii, de exemplu
 - nu au intuiții legate de trafic
 - identificând tipare recurente, observă și încearcă să evite pietonii neatenți
 - mai fac și greșeli
- IA nu concurează cu sufletul omenesc, ci cu rețele neuronale care calculează probabilități și identifică tipare
- IA dotată cu senzori adecvați ar putea lua decizii cu mai multă acuratețe și precizie decât Homo sapiens

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Caracteristici IA non-umane
- conectarea
 - => rețea integrată
 - nu comparăm entități separate (șofer uman – mașină autonomă)
 - ci grupuri (doi șoferi care-și comunică greșit intențiile în trafic vs. două mașini autonome care fac parte din același sistem și folosesc același algoritm)
- updatarea
 - schimbarea unei reguli de circulație va fi implementată ușor și în același timp de toate mașinile autonome
 - identificarea unei noi boli sau unui nou medicament
 - imposibil de informat și actualizat instantaneu toți medicii umani

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Înlocuire totală a indivizilor cu computere?
- Dezavantaje:
 - dacă un medic uman ia o decizie greșită nu omoară toți pacienții
 - într-un sistem automat rezultatele ar fi catastrofale
 - greu de implementat abilități motorii și emoționale specifice asistentelor medicale
 - realizarea unei injecții dureroase, schimbarea unui bandaj etc.
 - există domenii în care e nevoie de creativitate, care de regulă e asociată cu emoțiile umane (artă)
 - arta e evaluată în funcție de impactul emoțional asupra publicului

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Totuși

- vehiculele autonome nu vor face niciodată greșelile pe care le fac oamenii (din cauza consumului de alcool, neatenției, oboselii)
 - reducerea numărului de morți și răniți în accidente rutiere cu 90%
- emoțiile (de exemplu generate de muzică) sunt rezultatul unui proces biochimic
 - senzori
 - învățare automată
 - => tip de personalitate, dispoziție, impactul pe care o secvență muzicală l-ar produce (rezonare cu tristețea, înveselire etc.)

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Apare o problemă: codificarea eticii
 - Cum se decide dacă o melodie e bine să ne alimenteze tristețea sau să ne înveselească?
 - implicarea unui psiholog
 - Cum se pun în balanță viețile oamenilor care ar putea fi implicați într-un accident?
 - doi copii (alergând după o minge) sar în fața unei mașini autonome
 - pentru evitare e necesară virarea pe contrasens, unde se apropie un camion
 - acest lucru ar conduce la uciderea pasagerului din mașina autonomă
 - care va fi decizia algoritmului?

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

- Există o temere: Va rămâne IA supusă?
 - filme SF cu roboți răzvrățiți
 - greu de implementat conștiința, sentimentele
 - cui îi este supusă IA?
 - caracteristici ale unor roboți militari
 - respectarea regulilor
 - neimplicarea emoțională (pozitivă sau negativă)
 - algoritmul implementat poate fi
 - moderat și inofensiv
 - nemilos și crud
- Problema nu e inteligența artificială a roboților, ci cruzimea, răutatea și prostia umană

„21 de lecții pentru secolul XXI” – Yuval Noah HARARI

Pericolul

Dacă investim prea mult în dezvoltarea IA și prea puțin în dezvoltarea conștiinței umane, inteligența artificială deosebit de sofisticată a mașinilor (computerelor) ar putea sluji doar la manipularea prostiei naturale a oamenilor.

Soluția

Pentru fiecare dolar și pentru fiecare minut pe care le investim în îmbunătățirea IA, ar fi înțelept să investim câte un dolar și câte un minut în dezvoltarea conștiinței umane

Vă mulțumesc!

UP