

Bazele algoritmice și etica în IA: de la teorie la practică

Set de instrumente pentru sesiuni
sincrone

Exerciții bazate pe proiecte

Exerciții bazate pe proiecte



Exerciții bazate pe proiect

	Descriere
Descrierea sarcinii	În echipe de 4-5, veți lucra la evaluarea uneia dintre cele șase soluții AI pe parcursul a cinci unități de curs. Această sarcină cuprinzătoare implică o analiză detaliată a soluției AI din mai multe perspective, urmând structura cursului și subiectele abordate în fiecare unitate de curs.
Descrierea modului de realizare a sarcinii	Instrucțiuni • Echipa dvs. va lucra la o soluție AI din lista de la pagina 4. • Analizați soluția aleasă pe baza celor cinci perspective furnizate la pagina 6 • Documentați-vă constatările într-un raport structurat, abordând în detaliu fiecare criteriu de evaluare. Fiecare punct trebuie să fie prezentat în max. 2 diapozitive • Prezentați analiza în fața clasei în ultima sesiune sincronă, subliniind principalele concluzii și recomandări pentru îmbunătățirea soluției. • Prezentarea dvs. către echipă ar trebui să dureze între 5-10 minute
Timpul estimat pentru realizarea sarcinii	Aproximativ 1-2 ore / UC (5-10 ore în total)
Sugestii de surse pentru realizarea sarcinii	Broșuri interactive și alte materiale de curs + orice alte surse
Descrierea detaliată a modului de realizare a sarcinii	A se vedea paginile următoare.
Informații privind termenul limită pentru predarea sarcinii	Prezentat în ultima sesiune sincronă.
Informații de contact sau modul de clarificare a dubiilor	Profesorul trebuie să furnizeze o formă de contact (ar putea fi o adresă de e-mail, un număr de telefon...).

Exerciții bazate pe proiect

Cronologie și instrucțiuni

CALENDARUL EXERCITIULUI DE PROIECT

Prezentarea proiectului și formarea echipelor.

- De preferință, începeți să lucrați la proiect după prima sesiune sincronă, conveniți cu echipa asupra întâlnirilor pentru a lucra la proiect.
- Alternativ, începeți după a 2-a sesiune sincronă care acoperă CU 1 și 2.

Analizați progresul

- La a 3-a și a 4-a sesiune sincronă posibilitatea de a revizui progresul și de a discuta orice provocări cu profesorul.

Prezentarea finală

- În cea de-a 5-a sesiune sincronă prezentați exercițiul proiectului în echipă.
 - Prezentare 5-10 minute.
 - Puteți alege aspectul PowerPoint sau Canva, nu uitați să faceți o prezentare atractivă.

NOTĂ | În cazul în care studentul nu este disponibil pentru lucrul în grup (sau nu participă activ la proiect), exercițiul de proiect trebuie să fie prezentat individual profesorului (10 slide-uri).

Instrucțiuni suplimentare

- Subiectul dvs. este prezentat pe următorul diapozitiv. Ca echipă, veți ajunge să decideți numele, scopul și conținutul mai detaliat al serviciului dumneavoastră. Cu toate acestea, cererea trebuie să se afle în câmpul care vă este dat în pagina următoare.
- Continuați să lucrați la proiect în ritmul convenit cu echipa de proiect, cu toate acestea, conținutul este întotdeauna legat de unitățile de curs, astfel încât este de preferat să lucrați după ce membrii grupului au finalizat e-learningul pe subiect.
- Puteți decide asupra modului în care vă construiți prezentarea; astfel, întrebările sunt acolo ca îndrumare, dar ar trebui să creați un conținut care să fie atractiv pentru public.

Exerciții bazate pe proiecte

Soluții AI împărțite pe echipe

Soluția AI la care echipa ta ar trebui să lucreze:

Instrument de diagnosticare medicală bazat pe IA

- Un instrument care asistă medicii oferind diagnostice preliminare bazate pe simptomele și istoricul medical al pacientului.
 - XX grupuri care lucrează la acest subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Sistem de recomandare a serviciilor de streaming

- O aplicație care sugerează utilizatorilor filme și seriale pe baza istoricului de vizionare și a preferințelor acestora.
 - XX de grupuri care lucrează la subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Sistem inteligent de automatizare a locuinței

- O soluție care automatizează dispozitivele de acasă, cum ar fi luminile, termostatele și sistemele de securitate, pe baza rutinelor și preferințelor utilizatorului.
 - XX grupuri care lucrează la subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Consilier pentru investiții financiare bazat pe inteligență artificială

- Un consilier care oferă recomandări de investiții pe baza datelor de pe piață și a obiectivelor financiare ale utilizatorului.
 - XX de grupuri care lucrează la subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Aplicație instrument de cumpărături online

- O aplicație care recomandă produse utilizatorilor pe baza istoricului lor de navigare și cumpărare.
 - XX de grupuri care lucrează la subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Platformă personalizată de e-Learning

- O platformă care personalizează traseele de învățare și conținutul pentru studenți pe baza stilurilor de învățare, progreselor și preferințelor acestora.
 - XX grupuri care lucrează la subiect (adăugați aici numele elevilor, împărțiți în grupuri de 4-5 de către profesor)

Exerciții bazate pe proiecte

CU 1 conținut asociat

Începeți prin a defini următoarele (această sarcină nu ar trebui să necesite mai mult de 2 diapozitive de prezentare)

Enunțarea problemei: Ce problemă rezolvă aplicația dumneavoastră?

Scopul: Care este scopul aplicației dvs. de inteligență artificială? Cum se numește aplicația? Care sunt criteriile de succes ale soluției dumneavoastră?

Creați o hartă a părților interesate, care ar trebui să definească cel puțin părțile interesate de utilizare, părțile interesate externe și părțile interesate de dezvoltare (utilizați materialul CU1 ca îndrumare).

Luați în considerare măsura în care următoarele principii etice sunt respectate în cadrul aplicației dumneavoastră, accentul trebuind să fie pus în principal pe aspectele societale și de mediu:

Beneficentă

Asigurați-vă că soluția AI sporește bunăstarea individuală și binele societății. Domenii cheie de luat în considerare.

- Protecția drepturilor fundamentale: protejarea drepturilor individuale, consolidând integritatea etică.
- Durabilitatea mediului: minimizarea impactului asupra mediului prin practici durabile.
- Justificare: definiți și legați în mod clar scopul sistemului de beneficiile societale tangibile.

Non-maleficentă.

- Asigurați-vă că identificarea problemei și soluția propusă nu dăunează persoanelor sau comunităților.
- Luați în considerare cu atenție și monitorizați impactul sistemului asupra bunăstării fizice și mentale.
- Evaluarea potențialelor riscuri de siguranță și a strategiilor de atenuare.

Justiție.

Evaluati impactul sistemului asupra instituțiilor, democrației și societății.

Exerciții bazate pe proiect

CU 1 (Diapozitiv de ajutor)

Utilizați acest tabel pentru a vedea mai în detaliu posibilele considerente din cadrul fiecărui principiu etic, astfel cum sunt combinate de Morley et. al (2020)

Table 5 showing the connection between high-level ethical principles and tangible system requirements as adapted from the methodology outlined in Chapter II of the European Commission's "Ethics Guidelines for Trustworthy AI"

Principle	Beneficence	Non-Maleficence	Autonomy	Justice	Explicability
Requirements	Stakeholder participation: to develop systems that are trustworthy and support human flourishing, those who will be affected by the system should be consulted Protection of fundamental rights Sustainable and environmentally friendly AI: the system's supply chain should be assessed for resource usage and energy consumption Justification: the purpose for building the system must be clear and linked to a clear benefit—system's should not be built for the sake of it	Resilience to attack and security: AI systems should be protected against vulnerabilities that can allow them to be exploited by adversaries. Fallback plan and general safety: AI systems should have safeguards that enable a fallback plan in case of problems. Accuracy: for example, documentation that demonstrates evaluation of whether the system is properly classifying results. Privacy and Data Protection: AI systems should guarantee privacy and data protection throughout a system's entire lifecycle. Reliability and Reproducibility: does the system work the same way in a variety of different scenarios. Quality and integrity of the data: when data is gathered it may contain socially constructed biases, inaccuracies, errors and mistakes—this needs to be addressed Social Impact: the effects of system's on people's physical and mental wellbeing should be carefully considered and monitored	Human agency: users should be able to make informed autonomous decisions regarding AI systems Human oversight: may be achieved through governance mechanisms such as human-on-the-loop, human-in-the-loop, human-in-command	Avoidance of unfair bias Accessibility and universal design Society and democracy: the impact of the system on institutions, democracy and society at large should be considered Auditability: the enablement of the assessment of algorithms, data and design processes. Minimisation and reporting of negative impacts: measures should be taken to identify, assess, document, minimise and respond to potential negative impacts of AI systems Trade-offs: when trade-offs between requirements are necessary, a process should be put in place to explicitly acknowledge the trade-off, and evaluate it transparently Redress: mechanism should be in place to respond when things go wrong	Traceability: the data sets and the processes that yield the AI system's decision should be documented Explainability: the ability to explain both the technical processes of an AI system and the related human decisions Interpretability

SURSA | Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). De la ce la cum: O revizuire inițială a instrumentelor, metodelor și cercetărilor disponibile public în domeniul eticii IA pentru a transpune principiile în practici. Science and Engineering Ethics, 26(4), 2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>

Exerciții bazate pe proiect

CUs conținut asociat

CU 2

Preocupări legate de confidențialitate

- Identificați problemele de confidențialitate asociate cu soluția.
- Luați în considerare modul în care datele sunt colectate, stocate și utilizate.
- Sugerați măsuri de protecție a confidențialității utilizatorilor și de asigurare a securității datelor.

CU 3

Algoritmul utilizat

- Descrieți algoritmul care alimentează soluția.
- Explicați modul în care algoritmul este construit și antrenat pe baza datelor colectate (cu2).
- Discutați orice detalii tehnice specifice care sunt importante pentru înțelegerea soluției.

CU 4

Posibile prejudecăți

- Identificați posibilele distorsiuni ale sistemului.
- Discutați ce ar putea cauza aceste distorsiuni (de exemplu, datele de formare, proiectarea algoritmului).
- Sugerați modalități de prevenire sau atenuare a acestor distorsiuni.

CU 5

- TU delft a creat o listă de verificare pentru dezvoltarea etică a IA. Verificați instrumentul prin intermediul acestui [link](#) și utilizați lista de verificare a proiectului AI APP din paginile 12-13 și completați lista de verificare din pagina 13. Prezentați lista de verificare la sfârșitul prezentării proiectului dumneavoastră.
- Diapozitivul final ar trebui să prezinte ce ați învățat ca echipă și ce perspective noi ați dobândit prin această prezentare.

VĂ MULTUMIM

Numărul proiectului: 2022-1-ES01-KA220-HED-000085257



Universitat
de les Illes Balears



Cofinanțat de
Uniunea Europeană

Sprijinul acordat de Comisia Europeană pentru realizarea acestei publicații nu constituie o garanție a conținutului, care reflectă numai opiniile autorilor, iar Comisia nu poate fi considerată responsabilă pentru utilizarea informațiilor conținute în aceasta.

