

Activitatea 3.1.2 - Principalele provocări și riscuri societale în utilizarea IA

Durata estimată	40 de minute
Cerințe	Modulele 1 și 2. Activitatea 3.1.1. Dacă este făcut fizic - spațiu deschis moderat mare, unde elevii pot discuta și prezenta în grupuri
Instrumente Hardware și Software utilizate	La distanță: Software - Software pentru conferințe video. Editor de text colaborare online în timp real și software de prezentare. Hardware - Calculator desktop / laptop
Tipul evaluării	Evaluarea se bazează pe înțelegerea problemelor discutate în clasă afișate prin reprezentarea argumentelor care evidențiază impactul negativ al IA și prin discuții critice care pun la îndoială aceste argumente.

Obiectivul acestei activități este de a facilita înțelegerea modului în care utilizarea tehnologiilor IA poate avea un impact negativ asupra lumii reale prin reprezentarea argumentelor și exemplelor justificative. Comentariile și întrebările care contestă aceste argumente vor fi, de asemenea, încurajate și explorate.

Procedura (pași)

Set anticipativ

- Profesorii le pot reaminti elevilor despre sistemul de inteligență artificială studiat anterior (ex. chatboți, asistenți digitali, motoare de căutare, recunoaștere facială precum și urmărire pe baza trăsăturilor faciale) Profesorii fac din nou referire la ODD-urile prezentate în activitatea anterioară pentru a explica faptul că utilizarea IA poate avea efecte opuse și ne poate îndepărta de la atingerea acestor obiective. Profesorii pot prezenta câteva exemple pentru a explica acest punct.

- Cazul prejudecăților de algoritm. Profesorul explică faptul că oamenii iau uneori decizii asistate de algoritmi (un set de acțiuni sau pași necesari pentru a rezolva o problemă); aceste decizii sunt, prin urmare, parțial automate. Recomandările computerizate au un aer de raționalitate sau infailibilitate, iar uneori oamenii le urmează orbește. Însă, în funcție de modul în care a fost instruit să funcționeze și de informațiile introduse de oameni în PC (vezi învățarea automată) algoritmul poate sugera decizii care discriminează, de exemplu poate favoriza anumite tipuri de persoane în defavoarea altora. Acest lucru se numește „prejudecată IA” (prejudecată = o distorsiune a procesului de luare a deciziilor pe baza unei prejudecăți).
- Profesorul poate da un exemplu: pe baza unei evaluări a clienților făcută automat de un algoritm, un angajat al unei bănci poate decide dacă un client poate solicita un împrumut de la bancă sau nu. Dacă algoritmul se bazează doar pe date colectate într-un moment în care femeile nu erau eligibile pentru a solicita un împrumut bancar, acesta poate da o indicație conform căreia cei mai potriviți clienți pentru acordarea împrumuturilor sunt bărbații. Acest lucru dă naștere la discriminarea sau tratamentul inegal al femeilor. Prin urmare, contrazice ODD 5 (Egalitatea de gen).
- Profesorul explică faptul că alte probleme posibile sunt: probleme de confidențialitate în utilizarea algoritmilor de recunoaștere facială în spațiile publice sau de securitate în utilizarea sistemelor care creează voce și sunete precum asistenții digitali. Alte exemple includ dificultăți ale algoritmilor de recunoaștere vocală de a identifica cu precizie comenzi provenite de la vocile unor adulți în vârstă sau mai tineri pentru că algoritmii sunt instruiți cu ajutorul unei populații de adulți de vârstă medie. Software-urile de recunoaștere a imaginilor sau motoarele de căutare bazate pe imagini pot genera rezultate părtinitoare (vezi <https://en.unesco.org/artificial-intelligence/ethics/cases#biasedai>)

Instrucțiuni directe

- Profesorii împart elevii în grupuri.
- Fiecare grup de elevi discută un exemplu de implicații negative sau riscuri cauzate de aplicațiile IA în lumea reală.
- Fiecare grup se gândește la implicațiile posibile ale exemplului respectiv în alte domenii/privind alte aspecte cheie

Practică ghidată și înțelegere

- Fiecare grup, utilizând materialul furnizat de profesor, se gândește la o metodă de a reprezenta exemplul discutat pentru a sublinia modurile în care tehnologiile IA creează riscuri și au un impact negativ asupra problemei vizate, precum și la contraargumentele față de aplicația respectivă. Fiecare elev poate reprezenta un aspect, agent sau funcționalitate a întregului sistem.

Practică independentă

- Elevii din fiecare grup interpretează reprezentarea exemplului de aplicație IA și argumentele împotriva utilizării acesteia în fața clasei.

Închidere

- Restul clasei contestă argumentele împotriva utilizării IA în domeniul respectiv prin comentarii critice și întrebări. Profesorii moderează discuția și clarifică orice interpretări greșite sau îndoieli pe care le pot avea ceilalți elevi.

Soluționarea problemelor

Problemă posibilă	Soluție posibilă
Exemplele IA de aplicații ale IA în sistemele din lumea reală sunt neclare pentru elevi sau elevii nu pot relaționa conceptele pe care le-au învățat anterior cu aplicațiile sistemelor IA pe care încearcă să le explice.	Asigurarea din partea profesorului și re-explicarea conceptelor de nivel înalt, dacă este necesar.
Elevii care nu doresc să joace fizic o aplicație a tehnologiilor IA în lumea reală.	Parteneriatul cu un profesor de teatru și transformarea lecției într-un mediu mai adecvat, cum ar fi un teatru școlar sau un spațiu mare.
Elevii care nu doresc să conteste exemplele și argumentele prezentate de un grup	Profesorul pune întrebări care încurajează dezbateră critică privind aplicarea specifică a IA

Resurse pentru profesori

Resurse care pot fi utilizate în clasă pentru a descrie exemple de aplicații IA și dimensiunea etică

Titlu	Descriere	Link	Limbi
Etică și IA: Confidențialitatea și viitorul pieței muncii	Video care explică modul în care etica și confidențialitatea sunt legate în cadrul IA și cum IA poate influența locurile de muncă în	https://www.youtube.com/watch?v=zNxw5gJtHLc	ENGLEZĂ cu subtitrări autogenerate

	viitor.		
Etică și IA: Acces egal și subiectivism algoritmic	Video care descrie potențialul IA de a ajuta societatea cu accent pe etică, acces și corectitudine	https://www.youtube.com/watch?v=tJQSyzBUAew	ENGLEZĂ cu subtitrări autogenerate
ως λειτουργεί η Τεχνητή Νοημοσύνη - Τεχνητή Νοημοσύνη: Δεδομένα και Μεροληψία	Un video scurt care explică cum sunt conectate datele utilizate la instruirea IA și prejudecățile	https://www.youtube.com/watch?v=7a2aNF-dRo8&list=PLIzOpQKPcu89NB4AdYYuixHPnN_2QocuZ&index=3	ENGLEZĂ cu subtitrări în GREACĂ
Ce se întâmplă cu recunoașterea facială?	Un video scurt care explică cum poate fi utilizată recunoașterea facială pentru a ajuta societatea sau pentru a încălca dreptul oamenilor la viață privată	https://youtu.be/BqQT4slOYA0	ENGLEZĂ cu subtitrări autogenerate

Impactul negativ al IA / riscurile reprezentate de resursele IA:

Poluarea algoritmică

- [Amazon elimină instrumentul de recrutare secret IA care arăta părtinire împotriva femeilor](#)
- [Google va plăti 2.6 milioane de dolari muncitorilor pentru că susține că practicile sale de angajare și de plată au fost părtinitoare împotriva femeilor și asiaticilor](#)
- [Instrumente de angajare bazate pe algoritmi și discriminarea persoanelor cu dizabilități](#)
- Köchling, A., Wehner, M. C. Discriminat de un algoritm: o revizuire sistematică a discriminării și corectitudinii prin luarea deciziilor algoritmice în contextul recrutării și dezvoltării resurselor umane. *Bus Res* **13**, 795–848 (2020).
<https://doi.org/10.1007/s40685-020-00134-w>
- [Conectat. „Mașinile învățate de fotografii învață o viziune sexistă asupra femeilor: algoritmii au arătat tendința de a asocia femeile cu cumpărăturile și bărbații cu tragerea”](#)
- [în italiană] Festival dei Diritti Umani 2021 - Algoritmocrazia "Nuove tecnologie, vecchi bias" - sesiune educațională
Include resurse privind utilizarea imaginilor și algoritmilor, cetățenia digitală, discursul de ură (de asemenea, modulul 4)
<https://festivaldirittiumani.it/event/nuove-tecnologie-vecchi-bias/>
Vezi și: <https://festivaldirittiumani.it/irony-di-radheya-jegatheva/>
- [Centrul pentru Democrație și tehnologie. utilizarea responsabilă a datelor și tehnologiei în educație: implicarea comunității pentru a se asigura că elevii și familiile sunt ajutați, nu să sufere](#)

Probleme de confidențialitate

- [Centrul pentru Democrație și Tehnologie, modulul de instruire privind protecția vieții private a elevilor](#)
- Stoilova, Mariya; Livingstone, Sonia; Khazbak, Rana (2021). Investigarea riscurilor și oportunităților pentru copii într - o lume digitală: o revizuire rapidă a dovezilor privind utilizarea și rezultatele internetului pentru copii, Innocenti Discussion Papers no. 2021-01, UNICEF Office of Research-Innocenti, Florența.
<https://www.unicef-irc.org/publications/1183-investigating-risks-and-opportunities-for-children-in-a-digital-world.html>
Pentru mai multe resurse UNICEF în domeniul drepturilor copiilor în era digitală:
<https://www.unicef-irc.org/research/child-rights-in-the-digital-age/>
- Salvați Copiii(2020). Salvagardare digitală pentru copiii migranți și strămutați: o prezentare generală a contextului actual și a tendințelor, riscurilor potențiale și pașii următori practici.
https://resourcecentre.savethechildren.net/node/18477/pdf/mdi_digital_final_rgb_rev_091220.pdf
[exemplu pozitiv]
Misiunea proiectului Tor este de a promova înțelegerea populară a tehnologiilor de confidențialitate → instruiți în Sudul Global
<https://blog.torproject.org/reaching-people-where-they-are>

Supraveghere

- Fișier BBC pe 4 Podcast (2020). "Recunoașterea facială se întinde în toată Marea Britanie. Dar este un instrument vital împotriva criminalității - sau o amenințare la adresa celor nevinovați?" <https://www.bbc.co.uk/programmes/m000dpkt>
- Utilizarea recunoașterii faciale în școli:
<https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal>
<https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/>
(a se vedea, de asemenea, raportul privind automatizarea societății 2020)
- Există două proiecte pilot în Suedia care își concentrează atenția asupra utilizării IA și a recunoașterii faciale a elevilor în scopuri administrative, permițând profesorilor să-și dedice timpul predării
<https://www.electronicshpecifier.com/products/artificial-intelligence/facial-recognition-tested-in-swedish-high-school>
- Schimbarea structurală și cognitivă a școlilor și a școlii, a se vedea Andrejevic și Selwyn, Tehnologia recunoașterii faciale în școli: întrebări și preocupări,
<https://www.tandfonline.com/doi/full/10.1080/17439884.2020.1686014>

Repercusiunile economice și de mediu ale sistemelor bazate pe IA

- BBC documentarul Podcast (2019). "Va ucide IA dezvoltarea? Robotizarea va împiedica țările mai sărace să urmeze calea tradițională spre prosperitate?"
<https://www.bbc.co.uk/programmes/p075ngyy>

-
- Sahota, Neil. "Dincolo de IA: Nevoia de soluții în crearea unei lumi durabile". Forbes. 24 aprilie 2019.
<https://www.forbes.com/sites/cognitiveworld/2019/04/24/beyond-a-i-the-need-for-solutionists-in-creating-a-sustainable-world/>
 - Jones, Phil. „Refugiații ajută la progresul învățării automate la Microsoft, Facebook și Amazon”. Restul lumii. 22 septembrie 2021.
<https://restofworld.org/2021/refugees-machine-learning-big-tech/>
 - Documentarul Deutsche Welle (2021) "Invizibili – exploatarea în lumea digitală a muncii" <https://www.youtube.com/watch?v=o-HphMeZR3k>
 - Strubell și colab. 2019. "Considerații energetice și politice pentru învățarea profundă în NLP." Lucrările celei de-a 57-a reuniuni anuale a asociației pentru lingvistică computațională. <https://www.aclweb.org/anthology/P19-1355>.
DOI:10.18653/v1/P19-1355
 - Schwartz și colab. 2019 "IA verde." <https://arxiv.org/abs/1907.10597>
 - De ce ar trebui să ne pese de impactul IA asupra mediului
<https://www.forbes.com/sites/forbestechcouncil/2020/08/17/why-we-should-care-about-the-environmental-impact-of-ai/?sh=766825ef56ee>