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

Manual

Project number: 2022-1-ES01-KA220-HED-000085257

Index

CU1 | Etica IA - O abordare practică

Index

1. Introducere

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Acest manual este menit să vă ghideze prin materialul pentru unitatea de curs. PPT-urile de sprijin conțin informații similare cu cele acoperite de E-learning, dar într-o formă mai detaliată. PPT-urile sunt extinse și nu sunt menite să fie acoperite în întregime în timpul sesiunilor interactive. În schimb, puteți întreba la începutul sesiunilor interactive ce subiecte au fost cel mai greu de înțeles pentru elevi și puteți utiliza doar acele slide-uri de sprijin în timpul sesiunii. Puteți întreba acest lucru utilizând instrumente precum Mentimeter sau Kahoot, ca exemplu, adăugând subiectele de curs din conținutul unității de competență descrise în puncte la sfârșitul secțiunii de introducere. În plus, studiul de caz bazat pe CU este menit să fie tratat și în timpul sesiunii interactive. Orice alte sugestii din manual sunt doar orientative; nu ezitați să le folosiți după cum credeți de cuviință.

Ce s-ar întâmpla dacă inteligența artificială (AI) ar avea puterea de a decide dacă veți obține un împrumut? Sau ce s-ar întâmpla dacă AI ar judeca cine va fi condamnat la închisoare? Care ar fi consecințele dacă o mașină autonomă ar trebui să decidă dacă să salveze un pieton sau viața celor din mașină? Ce s-ar întâmpla dacă un sistem de recrutare bazat pe inteligență artificială ar decide dacă să angajeze o femeie sau un bărbat drept căpitan de zbor?

Integrarea crescândă a IA în diverse aspecte ale societății ridică probleme etice semnificative. De la deciziile privind împrumuturile, condamnările penale și vehiculele autonome, până la practicile de angajare, dependența de algoritmii AI ridică întrebări cu privire la principii etice precum echitatea, transparența și responsabilitatea. Tehnologiile IA au potențialul de a perpetua prejudecățile, discriminarea și inegalitățile dacă nu sunt dezvoltate și reglementate cu atenție. Asigurarea faptului că sistemele de inteligență artificială acordă prioritate considerentelor etice este esențială pentru promovarea justiției, egalității și încrederii în aplicațiile lor în diferite domenii. (UNESCO, 2023c)

Această unitate va aprofunda principiile, cadrele și orientările etice care afectează dezvoltarea soluțiilor AI. Ciclul de viață al dezvoltării AI este examinat de la formularea problemei până la operarea și monitorizarea soluției AI. Principiile conexe și implicarea părților interesate sunt luate în considerare în fiecare etapă a procesului. La nivel înalt, această unitate de curs tratează următoarele subiecte.

- Etica IA și importanța acesteia în dezvoltarea soluțiilor IA
- Principiile etice ale IA
- Cadre etice, orientări și seturi de instrumente
- Grup de experți la nivel înalt; AI demn de încredere - cadru

- Legea UE privind inteligența artificială
- Ciclul de viață al dezvoltării AI
- Implicarea părților interesate în ciclul de viață al dezvoltării IA
- Ciclul de viață al dezvoltării IA cu principiile etice și părțile interesate aferente

2. Definirea principiilor, cadrelor și orientărilor etice

Inteligența artificială (IA) influențează toate aspectele vieții, afectând munca, timpul liber și oferind soluții la probleme globale precum schimbările climatice și accesul la asistență medicală. În contextul integrării pe scară largă a inteligenței artificiale în economii și societăți, se pune întrebarea cu privire la cadrele politice și instituționale adecvate necesare pentru a orienta dezvoltarea și aplicarea inteligenței artificiale, asigurând beneficii pentru societate.

Considerațiile etice ample din domeniul inteligenței artificiale au determinat crearea și publicarea a numeroase abordări pentru a asigura aplicarea etică a inteligenței artificiale. (Prem, 2023) Principiile, cadrele și orientările etice sunt utilizate în mod obișnuit și recunoscute pe scară largă ca fiind structura de bază în luarea deciziilor etice în numeroase discipline. Acestea oferă o abordare cuprinzătoare a navigării în probleme etice, de la stabilirea valorilor fundamentale la orientarea acțiunilor specifice. Deși adesea interconectate, acestea pot fi descrise după cum urmează:

Principiile etice oferă un cadru moral pentru evaluarea și rezolvarea dilemelor etice. Principiile etice formează valorile de bază și convingerile care servesc drept busolă pentru luarea deciziilor etice și oferă o bază pentru determinarea a ceea ce este bine și ceea ce este rău în diferite contexte. (Prem, 2023; Zhou & Chen, 2022)

Cadrele etice sunt abordări sistematice sau modele care îi ajută pe indivizi și pe organizații să abordeze problemele etice. Cadrele etice sunt modele cuprinzătoare sau metode structurate care facilitează examinarea și soluționarea problemelor etice. Ele oferă un cadru pentru evaluarea consecințelor etice ale unei acțiuni și ghidează procesul decizional, asigurând luarea în considerare sistematică a considerentelor etice. (Prem, 2023; Zhou & Chen, 2022)

Orientările etice sunt reguli, standarde și bune practici specifice care oferă sfaturi practice privind comportamentul etic în diferite contexte. Orientările etice sunt reguli și repere detaliate care oferă sfaturi practice cu privire la modul de aplicare a principiilor etice în diferite scenarii. Aceste orientări ajută atât indivizii, cât și organizațiile să facă față în mod eficient problemelor etice și să mențină un comportament etic în cadrul operațiunilor și al procesului decizional. (Prem, 2023; Zhou & Chen, 2022)

3. Efort comun pentru modelarea peisajului etic

Principiile, cadrele și orientările etice provin din multe domenii și organizații diferite, inclusiv companii private, instituții de cercetare și organizații din sectorul public, reprezentând un efort comun de a modela peisajul etic al inteligenței artificiale. (Jobin et al., 2019; Morley et al., 2020)

Documentele orientative și rapoartele privind IA etică sunt exemple de instrumente politice nelegislative sau "soft law". Spre deosebire de "legislația dură", care include reglementările juridice stabilite de organismele legislative pentru a impune sau a interzice anumite comportamente, orientările etice nu au forță juridică, dar au putere de convingere. Aceste documente sunt concepute pentru a sprijini procesele decizionale în domenii specifice și s-a observat că exercită o influență considerabilă asupra practicilor. (Jobin et al., 2019)

Exemple de documente care definesc principiile etice ale inteligenței artificiale sunt:

Principiile IA ale OCDE promovează utilizarea inovatoare și fiabilă a IA, care respectă drepturile omului și valorile democratice. Adoptate în mai 2019, acestea stabilesc standarde pentru inteligența artificială care își propun să fie practice și flexibile. (OCDE, 2019)

Academia de Inteligență Artificială din Beijing (BAAI) a publicat "Principiile IA de la Beijing", o schiță pentru a ghida cercetarea și dezvoltarea, punerea în aplicare și guvernanța IA. ("Beijing Academy of Artificial Intelligence - Beijing AI Principles", 2019; *Principiile inteligenței artificiale de la Beijing*, 2022)

Actori importanți din industrie, precum Google, IBM, Microsoft și Intel, și-au elaborat propriile orientări etice care reflectă perspectivele și domeniile lor unice de activitate. (Jobin et al., 2019; Morley et al., 2020)

Organismele guvernamentale nu au fost lăsate în urmă, cu documente-cheie precum Declarația de la Montreal și contribuții din partea unor organisme precum House of Lords Select Committee on Artificial Intelligence și Grupul de experți al Comisiei Europene, fiecare implicat în reglementarea și politica privind etica inteligenței artificiale. (Morley et al., 2020)

Instituțiile academice și grupurile de reflecție, inclusiv Future of Life Institute, IEEE și AI4People, au aprofundat dezbaterea cu cercetări științifice și modele teoretice. (Floridi et al., 2018; Jobin et al., 2019; Morley et al., 2020)

Combinarea acestor eforturi reflectă o mișcare globală către o inteligență artificială responsabilă, iar fiecare contribuție constituie o piesă din puzzle-ul mai larg al creării unor sisteme de inteligență artificială etice, care să fie în conformitate cu valorile și normele societății.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Puteți cere elevilor să verifice sursele menționate mai sus sau unele dintre ele și să discute diferențele dintre principiile prezentate de acestea.

4. Principii etice

Peisajul principiilor etice în domeniul inteligenței artificiale este bogat și divers, cu numeroase principii propuse de o varietate de actori, inclusiv instituții academice, lideri din industrie, organisme guvernamentale și organizații internaționale. Proliferarea acestor principii reflectă un consens tot mai mare cu privire la necesitatea unei orientări etice în dezvoltarea și aplicarea tehnologiilor de inteligență artificială. (Jobin et al., 2019; Morley et al., 2020; Prem, 2023)

Jobin et al (2019) au efectuat o analiză cuprinzătoare a diferitelor principii etice privind inteligența artificială și le-au distilat pentru a identifica temele și elementele comune care revin din diferite surse. Printre orientările pe care le-au analizat, ei au constatat un consens emergent cu privire la principiile etice cheie pentru inteligența artificială. Deși nu sunt susținute universal, principiile de convergență includ deschiderea, echitatea și justiția, responsabilitatea, repararea, confidențialitatea, bunăvoința, libertatea și autonomia, încrederea, durabilitatea, demnitatea umană și solidaritatea.

Elementele comune identificate se referă la un set de preocupări și aspirații comune în comunitatea globală. Ele indică un consens emergent cu privire la valorile care, în opinia multora, ar trebui să stea la baza dezvoltării și implementării sistemelor de inteligență artificială. Acest acord preliminar formează un teren comun fundamental care poate servi ca punct de plecare pentru stabilirea așteptărilor și evaluarea rezultatelor. (Morley et al., 2020)

Morley et al (2020) au recunoscut acest acord colectiv ca fiind un punct de plecare fundamental. Ei susțin că aceste principii convenite caracterizează o IA adaptată din punct de vedere etic după cum urmează:

- Benefic și respectuos față de oameni și mediu (beneficență).
- Robust și sigur (non-maleficență).
- Respectă valorile umane (autonomia).
- Corect (justiție).
- Explicabil, responsabil și inteligibil (explicabilitate).

Beneficență

Principiul beneficenței este unul dintre considerentele etice cheie în dezvoltarea inteligenței artificiale și subliniază importanța sistemelor de inteligență artificială care nu numai că aduc beneficii persoanelor și societății, dar contribuie și la protecția mediului. Potrivit lui Floridi et al. (2018), obiectivul este de a se asigura că inteligența artificială funcționează într-un mod care promovează bunăstarea generală și sănătatea mediului.

Jobin et al. (2019) enumeră mai multe acțiuni-cheie pentru promovarea eficientă a inteligenței artificiale:

- Alinierea IA la valorile umane și asigurarea faptului că sistemele IA sprijină și consolidează aceste valori, în loc să le submineze.
- Implicarea părților interesate în procesul de dezvoltare a AI și luarea în considerare a nevoilor și feedback-ului acestora pentru a îmbunătăți impactul sistemelor AI asupra vieții lor.
- Lucrul activ pentru a minimiza și rezolva potențialele conflicte de interese, folosind feedback-ul utilizatorilor pentru a ghida dezvoltarea și aplicarea IA.
- Crearea de noi modalități de măsurare a bunăstării umane.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Discutați cu elevii despre impactul asupra mediului și societății. De exemplu, algoritmii TikTok pot furniza conținut dăunător dacă utilizatorul manifestă interes pentru acesta, ceea ce duce la efecte negative asupra sănătății mintale. În plus, cantitatea semnificativă de energie electrică utilizată pentru a crea modele lingvistice mari poate lăsa o amprentă considerabilă de dioxid de carbon.

Non-maleficență

Principiul etic al non-maleficenței este dedicat prevenirii daunelor. Conform Floridi et al. (2018), non-maleficența acoperă măsurile proactive atât împotriva abuzurilor intenționate ale oamenilor, cât și împotriva acțiunilor negative neintenționate ale sistemelor AI care pot afecta comportamentul uman. Obiectivul principal este de a se asigura că IA nu conduce la efecte dăunătoare, indiferent de originea acestora.

Jobin et al. (2019) au identificat următoarele abordări pentru gestionarea riscurilor și protejarea împotriva daunelor în sistemele AI:

- Utilizarea managementului strategic al riscurilor pentru anticiparea și reducerea riscurilor potențiale legate de sistemele AI.

- Punerea în aplicare a unei combinații de măsuri tehnice și politici de guvernanță care să acopere întregul ciclu de viață al IA, cu scopul de a preveni daunele înainte ca acestea să apară.
- Anumite daune pot fi inevitabile în ciuda tuturor eforturilor depuse, ceea ce necesită o evaluare cuprinzătoare a riscurilor. Aceasta include măsuri de reducere și atenuare a riscurilor și atribuirea clară a responsabilității atunci când se produce un prejudiciu.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Discutați cu elevii despre posibilele prejudicii cauzate de inteligența artificială. De exemplu, instrumentul de recrutare AI al Amazon a arătat prejudecăți față de candidații de sex feminin. Consultați mai multe detalii, de ex. aici: Insight - Amazon renunță la instrumentul secret de recrutare AI care a arătat prejudecăți împotriva femeilor | Reuters

Autonomie

Autonomia în contextul IA se concentrează pe menținerea unui echilibru între procesul decizional uman și influența IA. Oamenii trebuie să păstreze controlul asupra procesului decizional și să aleagă când să facă propriile alegeri și când să lase AI să decidă. Cu toate acestea, oamenii ar trebui să aibă întotdeauna capacitatea de a prelua controlul de la AI, dacă este necesar, ceea ce înseamnă că ei au ultimul cuvânt. (Floridi et al., 2018)

Jobin et al (2019) identifică o serie de măsuri de promovare a libertății și autonomiei în sistemele AI:

- Îmbunătățirea transparenței și predictibilității sistemelor AI, astfel încât utilizatorii să poată înțelege și anticipa comportamentul AI.
- Îmbunătățirea înțelegerii de către publicul larg a tehnologiilor IA.
- Punerea în aplicare a unor protocoale puternice de notificare și consimțământ pentru a asigura autonomia individuală în interacțiunea cu sistemele AI.
- evitarea colectării și distribuirii de date cu caracter personal fără consimțământul expres și în cunoștință de cauză al persoanelor vizate, respectând viața privată și independența acestora.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Discutați cu elevii despre posibilele probleme de autonomie ale IA. De exemplu, Facebook oferă instrumente pentru ca utilizatorii să exercite controlul asupra conținutului lor, dar complexitatea și opacitatea algoritmului pot submina uneori adevărata autonomie a utilizatorului?

Justiție

Justiția este un principiu cu multiple fațete care urmărește să utilizeze IA ca instrument de corectare a nedreptăților din trecut, să se asigure că beneficiile IA sunt distribuite în mod echitabil și să protejeze împotriva apariției de noi prejudicii sau perturbări societale. (Floridi et al., 2018)

Realizarea justiției în inteligența artificială, astfel cum este discutată de Jobin et al. (2019), include mai multe măsuri proactive:

- Stabilirea de standarde și norme tehnice.
- Creșterea transparenței drepturilor și reglementărilor.
- Punerea în aplicare a testării, monitorizării și auditării, în special în unitățile de protecție a datelor.
- Consolidarea statului de drept, inclusiv a căilor de atac și a remediilor juridice.
- Sunt implementate schimbări sistemice, cum ar fi supravegherea guvernamentală, echipe versatile și o societate civilă incluzivă, cu accent pe distribuirea echitabilă a beneficiilor.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

A se vedea, de exemplu, cazul chatbotului Tay al Microsoft: In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation - IEEE Spectrum. Referitor la justiție; algoritmul a prioritizat în mod incorect conținutul senzațional în detrimentul discursului semnificativ și constructiv, ducând la o reprezentare dezechilibrată a opiniilor și vocilor pe platformă, creând perturbări societale.

Explicabilitate

Explicabilitatea este esențială pentru crearea și menținerea încrederii utilizatorilor în sistemele AI. Aceasta înseamnă că procesele trebuie să fie transparente, capacitățile și scopul sistemelor de inteligență artificială trebuie să fie comunicate în mod deschis, iar deciziile trebuie explicate celor afectați direct și indirect. (High-Level Expert Group on AI (AI HLEG), 2019) Floridi et al (2018) vede, de asemenea, că explicabilitatea este centrală în aplicarea altor principii etice.

Pentru a spori transparența sistemelor AI, Jobin et al. (2019) enumeră mai multe abordări:

- Dezvoltatorii și utilizatorii AI sunt încurajați să împărtășească mai multe informații despre sistemele lor, demistificând procesele din spatele procesului decizional al AI.
- Explicații ale operațiunilor și deciziilor AI în moduri care sunt ușor de înțeles de către neexperți, asigurând că operațiunile AI pot fi auditate de oameni care nu sunt neapărat experți în domeniu.
- Utilizarea auditurilor ca modalitate de a asigura transparența în IA, care poate contribui la identificarea și abordarea potențialelor prejudecăți sau probleme etice.
- Instituirea unor mecanisme de monitorizare a performanței AI, implicarea părților interesate și a publicului pentru a aduna diverse perspective și mecanisme de sprijinire a denunțării neregulilor.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Întrebați elevii dacă se simt suficient de informați cu privire la deciziile luate de soluțiile AI pe care le folosesc. De exemplu, instrumentul de recrutare AI al Amazon a fost lipsit de transparență în procesul său decizional. Candidaților și managerilor de angajare nu li s-au oferit explicații clare cu privire la motivul pentru care anumiți candidați au fost respinși, ceea ce a făcut dificilă contestarea sau abordarea prejudecăților care stau la baza sistemului.

NOTĂ! La sfârșitul secțiunii privind principiile, vă rugăm să folosiți studiul de caz furnizat ca o lucrare de grup.

5. Cadre etice, orientări și seturi de instrumente

Odată ce principiile etice au fost identificate, acestea ar trebui transformate în cadre de acțiune și strategii care să ghideze inovațiile bazate pe inteligența artificială și implementarea practică a eticii inteligenței artificiale.

Potrivit lui Ayling & Chapman (2022), cadrele etice traduc principiile etice în măsuri practice. Acest proces implică crearea de orientări și proceduri pe care le pot adopta dezvoltatorii și cei care adoptă tehnologiile IA, încorporând astfel considerentele etice în fiecare etapă, de la proiectare la aplicațiile din lumea reală.

Prem (2023) dezvoltă această idee afirmând că multe cadre etice de IA încearcă în mod activ să anticipeze și să abordeze potențialele probleme și riscuri etice. Prin identificarea proactivă a acestor provocări, cadrele ghidează dezvoltatorii în crearea de soluții pentru IA. Această abordare proactivă este esențială într-un

domeniu care evoluează atât de rapid precum inteligența artificială, în care pot apărea noi considerații etice pe măsură ce tehnologia se dezvoltă și devine mai integrată în diferite aspecte ale vieții cotidiene.

Un cadru etic oferă un plan cuprinzător pentru stabilirea standardelor de reglementare și operaționale care se armonizează cu normele etice. După cum subliniază Floridi & Cowls (2019), un astfel de cadru este esențial pentru elaborarea legislației, formularea normelor, stabilirea standardelor tehnice și identificarea celor mai bune practici aplicabile într-o gamă variată de industrii și regiuni geografice.

Prem (2023, p. 701) definește componentele esențiale ale cadrelor etice după cum urmează:

- **Concepte** de bază legate de discutarea aspectelor etice
- **Principii** etice (de exemplu, valori)
- **Preocuparea** cu privire la modul în care utilizarea și dezvoltarea sistemelor de inteligență artificială respectă principiile etice
- **Remedii** pentru rezolvarea problemelor (de exemplu, strategii, norme și orientări)

Sunt foarte necesare seturi de instrumente și orientări pentru aplicarea principiilor etice la proiectare, implementare și desfășurare. (Zhou et al., 2020) Seturile de instrumente și orientările etice sunt instrumente importante pentru orientarea inovării în domeniul IA într-o direcție care este în conformitate cu normele și valorile etice. Seturile de instrumente și orientările etice sunt considerate importante pentru reducerea decalajului dintre principiile etice abstracte și integrarea lor concretă în practicile IA.

Deși importanța seturilor de instrumente și a orientărilor este clară, transpunerea principiilor în instrumente concrete reprezintă o provocare, iar multe dintre cadrele existente nu au un context de aplicare. (Prem, 2023, p. 702) Motivul pentru diferența mare dintre principii și practică se datorează lipsei de standardizare, variabilității, complexității și înțelegerii subiective a principiilor mai degrabă abstracte.

Ritmul rapid recent al dezvoltării IA a dus la o creștere substanțială atât a numărului, cât și a diversității instrumentelor disponibile, cu noi evoluții care apar în mod constant. Morley et. al (2019) și Prem (2023) au trecut în revistă sute de instrumente diferite pentru abordarea provocărilor etice din soluțiile IA. Aceste instrumente sunt împărțite în funcție de etapele de proiectare a soluțiilor AI (definite ulterior începând cu capitolul H) și de diferite principii etice. Cu toate acestea, multe instrumente sunt percepute ca fiind dificil de utilizat sau sunt considerate a necesita un efort substanțial din partea utilizatorilor lor.

Listele întocmite de Morley et. al (2019) și Prem (2023) arată clar că disponibilitatea instrumentelor și metodelor este împărțită inegal între diferitele principii și etape de proiectare. Ca exemplu, instrumentele de măsurare a beneficenței soluției sunt numeroase în etapele timpurii ale dezvoltării, în timp ce instrumentele de

explicabilitate sunt poziționate în cea mai mare parte în etapa ulterioară a procesului de dezvoltare. Consultați instrumentele prin intermediul acestui link: Applied AI Ethics Typology - Google Docs.

Metodele utilizate variază, de asemenea, în funcție de etapa de dezvoltare. Cadrele, bibliotecile și algoritmii joacă adesea un rol mai important în faza ex-ante, în special în etapa de proiectare a soluției. Cu toate acestea, unele cadre ajung atât în etapa ex-ante, cât și în etapa ex-post (UNESCO, 2023b, p. 5) În faza de testare, din nou, auditurile și metricile sunt prezentate în mod predominant, în timp ce în fața post ante sunt utilizate ca abordări în principal declarații, etichete și licențe. (Prem, 2023, p. 709)

Studiul a arătat, de asemenea, că doar câteva instrumente au abordat impactul soluției asupra unui individ sau asupra unei societăți în ansamblu. Acest lucru s-ar putea datora faptului că este o provocare să schimbi comportamentul uman complex în instrumente de proiectare simple și generale. Cu toate acestea, ar fi esențial să se abordeze impactul asupra omului și a societății pentru a obține acceptarea și preferabilitatea socială a soluțiilor AI. (Morley et al., 2020, p. 2156)

6. De la principii la practică; Cadrul IA de încredere al HLEG

Deși există un consens cu privire la faptul că IA trebuie să fie etică, continuă dezbaterile cu privire la definiția exactă a "IA etică" și la obligațiile etice și criteriile tehnice necesare pentru punerea sa în aplicare. (Leijnen et al., 2020)

Comisia Europeană a elaborat o strategie cuprinzătoare pentru avansarea inteligenței artificiale (IA) în Europa, punând un accent puternic pe asigurarea faptului că IA dezvoltată și utilizată nu este doar avansată din punct de vedere tehnologic, ci respectă și standardele etice și este sigură.

Pentru a pune în aplicare în mod eficient această strategie și principiile care stau la baza acesteia, Comisia Europeană a înființat Grupul de experți la nivel înalt privind inteligența artificială (AI HLEG). Acest grup a fost însărcinat cu elaborarea Orientărilor privind etica inteligenței artificiale și a Recomandărilor privind politicile și investițiile, care sunt menite să ghideze dezvoltarea, implementarea și utilizarea etică a inteligenței artificiale în întreaga Europă, promovând inovarea și protejând în același timp valorile și drepturile fundamentale. [Grupul de experți la nivel înalt privind IA (AI HLEG), 2019]

Cadrul IA demnă de încredere este conceput pentru a garanta că tehnologiile IA sunt dezvoltate și implementate în mod etic, respectând drepturile fundamentale și asigurând robustețea și fiabilitatea. Cadrul pentru o IA demnă de încredere este alcătuit din trei componente, fiecare fiind concepută pentru a ghida dezvoltarea și implementarea IA într-un mod etic și fiabil. [Grupul de experți la nivel înalt privind inteligența artificială (AI HLEG), 2019]

Bazele inteligenței artificiale demne de încredere

Acest segment stabilește principiile de bază care stau la baza IA demnă de încredere, adoptând o abordare bazată pe drepturile fundamentale. Acesta definește principiile etice esențiale pentru a se asigura că sistemele AI sunt dezvoltate și utilizate într-un mod care este atât etic, cât și robust.

Principiile etice constituie fundamentul pe care trebuie construită IA demnă de încredere, asigurând respectarea standardelor etice. Aceste principii etice sunt (i) respectarea autonomiei umane (ii) prevenirea daunelor (iii) corectitudinea și (iv) explicabilitatea.

Realizarea unei IA demne de încredere

Pornind de la principiile etice subliniate, a doua parte a cadrului transpune aceste principii în șapte cerințe-cheie pe care sistemele de inteligență artificială trebuie să le încorporeze și să le respecte pe parcursul ciclului lor de viață. Această transformare a principiilor etice în cerințe practice garantează că etica fundamentală nu este doar un ideal teoretic, ci este integrată activ în fiecare etapă de dezvoltare și funcționare a unui sistem de inteligență artificială.

Lista de cerințe include: Agenție umană și supraveghere: drepturi fundamentale, agenție umană și supraveghere umană; Robustețe tehnică și siguranță: rezistență la atacuri și securitate, plan de rezervă și siguranță generală, acuratețe, fiabilitate și reproductibilitate; Confidențialitate și guvernanța datelor: respectarea vieții private, calitatea și integritatea datelor și accesul la date; Transparență: trasabilitate, explicabilitate și comunicare; diversitate, nediscriminare și echitate: evitarea prejudecăților nedrepte, accesibilitate și design universal și participarea părților interesate; bunăstarea societății și a mediului: durabilitate și respectarea mediului, impact social, societate și democrație; responsabilitate: auditabilitate, minimizarea și raportarea impactului negativ, compromisuri și reparații.

Evaluarea inteligenței artificiale demne de încredere

A treia parte a cadrului oferă o listă concretă și cuprinzătoare pentru evaluarea IA demnă de încredere, concepută pentru a operaționaliza cerințele menționate mai sus.

Cunoscut sub numele de "Assessment List for Trustworthy AI (ALTAI)", acest instrument oferă o listă de verificare cuprinzătoare și dinamică care servește drept foaie de parcurs pentru ca dezvoltatorii și implementatorii să integreze aceste principii în aplicațiile lor de IA. ALTAI facilitează acest proces prin oferirea unui set de pași concreți pentru autoevaluare, asigurându-se astfel că tehnologiile IA sunt dezvoltate într-un mod care maximizează beneficiile utilizatorilor, minimizând în același timp expunerea la riscuri inutile. (Grupul de experți la nivel înalt în domeniul inteligenței artificiale (AI HLEG), 2020)

Această listă de evaluare neexhaustivă servește drept ghid practic pentru practicienii AI, oferind orientări specifice cu privire la modul de punere în aplicare eficientă a acestor cerințe. În mod important, această evaluare se dorește a fi adaptabilă, permițând personalizarea în funcție de aplicarea specifică a fiecărui sistem IA, asigurând astfel relevanța și eficacitatea în diverse contexte.

7. De la principii la legislație: EU AI ACT

Actul privind inteligența artificială reglementează aplicarea inteligenței artificiale în cadrul UE, marcând prima legislație extinsă privind inteligența artificială la nivel global. (Parlamentul European, 2023) Scopul AI ACT este de a îmbunătăți funcționarea pieței interne și de a promova adoptarea inteligenței artificiale centrate pe om și demne de încredere, asigurând în același timp un nivel ridicat de protecție a sănătății, a siguranței, a drepturilor fundamentale, inclusiv a democrației, a statului de drept și a protecției mediului împotriva efectelor nocive ale sistemelor de inteligență artificială în Uniune și sprijinind inovarea. (Future of Life Institute, 2024)

Comisia Europeană a organizat reglementările privind IA într-un cadru bazat pe riscuri, creând un sistem de categorii de risc. Aplicațiile IA vor fi reglementate în funcție de nivelul de risc pe care îl prezintă. (Comisia Europeană, 2023). Abordarea bazată pe risc prezentată în Legea UE privind IA clasifică sistemele de IA în patru niveluri în funcție de impactul lor potențial: risc inacceptabil, risc ridicat, risc limitat și risc scăzut sau minim. (Comisia Europeană, 2023; Hillard & Gulley, 2023)

Risc minim

Majoritatea sistemelor AI se încadrează în această categorie, prezentând riscuri minime sau inexistente pentru drepturile sau siguranța cetățenilor. Printre exemple se numără sistemele de recomandare și filtrele antispam bazate pe IA. Companiile care operează astfel de sisteme nu sunt obligate să respecte cerințe stricte, dar se pot angaja voluntar să respecte coduri de conduită suplimentare pentru a spori transparența și responsabilitatea.

Risc limitat

Această categorie se referă la sistemele AI care prezintă doar un risc limitat pentru drepturile și siguranța utilizatorilor. Exemplele includ roboții de chat și falsurile profunde. Deși nu fac obiectul unor cerințe stricte de reglementare, furnizorii sunt obligați să se asigure că utilizatorii sunt conștienți când interacționează cu sistemele AI și să eticheteze conținutul generat de AI în mod corespunzător. Acest lucru asigură transparența și sensibilizarea cu privire la utilizarea tehnologiilor AI.

Risc ridicat

Sistemele de inteligență artificială identificate ca fiind cu risc ridicat trebuie să respecte cerințe stricte pentru a reduce riscurile potențiale. Aceste cerințe includ sisteme solide de atenuare a riscurilor, seturi de date de înaltă calitate, înregistrarea activităților, documentație detaliată, informații clare pentru utilizatori, supraveghere umană și măsuri puternice de securitate cibernetică. Sandbox-urile de reglementare sunt create pentru a încuraja inovarea responsabilă și pentru a facilita dezvoltarea de sisteme AI conforme. Printre

exemplele de sisteme AI cu risc ridicat se numără infrastructurile critice, dispozitivele medicale, sistemele de acces ale instituțiilor de învățământ și anumite aplicații în domeniul aplicării legii și al controlului frontierelor.

Risc inacceptabil

Sistemele AI care reprezintă o amenințare clară la adresa drepturilor fundamentale sunt interzise. Printre acestea se numără sistemele care manipulează comportamentul uman pentru a submina liberul arbitru, cum ar fi jucăriile care încurajează comportamentul periculos al minorilor sau cele care permit "marcarea socială" a guvernului sau a întreprinderilor. Aplicațiile polițienești predictive și anumite utilizări ale sistemelor biometrice, cum ar fi recunoașterea emoțiilor la locul de muncă, sunt, de asemenea, interzise.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

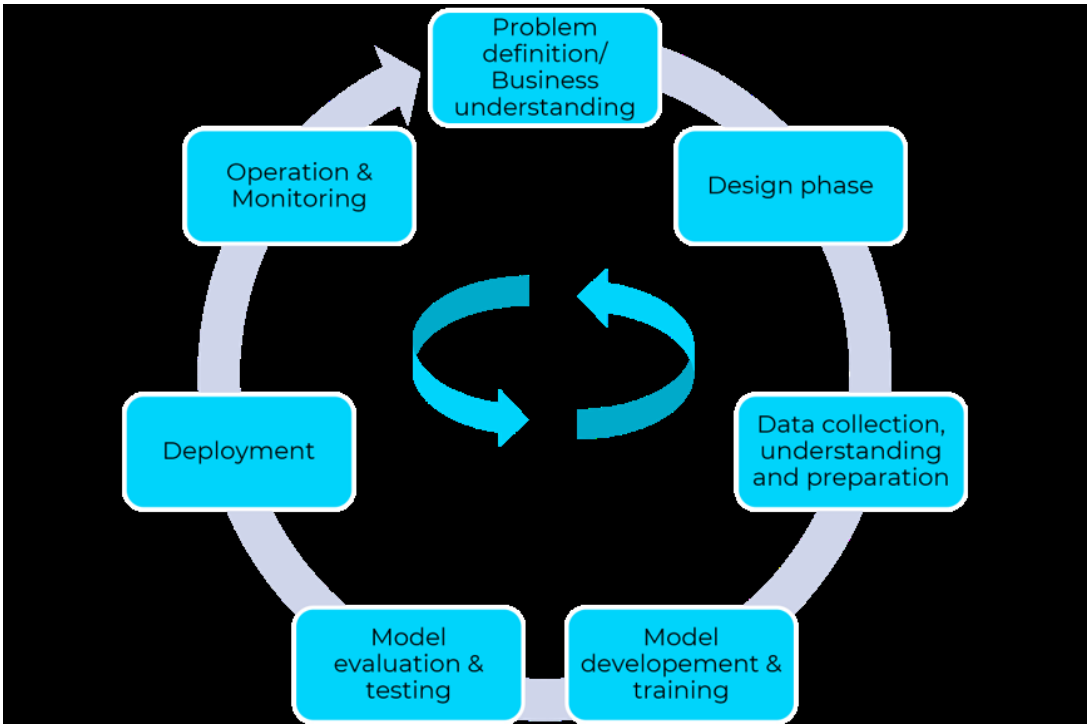
1) Discutați cu elevii dacă sunt de părere că legislația UE va fi mai strictă decât alte legi și ce va cauza aceasta. Sugestii, de exemplu, din acest articol: https://news.bloomberglaw.com/artificial-intelligence/eu-embraces-new-ai-rules-despite-doubts-it-got-the-right-balance
2) Dacă ne gândim la recenta discuție despre TikTok și cât de dăunător este, în ce categorie de risc s-ar încadra?

8. Ciclul de viață al dezvoltării AI

După cum se analizează în capitolul E, anumite principii etice devin mai relevante în anumite etape ale dezvoltării unui sistem de inteligență artificială, diverse instrumente fiind utilizate în consecință. Înțelegerea procesului de dezvoltare dintr-o perspectivă largă este crucială. (Prem, 2023, p. 703) În acest curs, adoptăm termenul "ciclul de viață al dezvoltării AI" pentru a încapsula această viziune cuprinzătoare a etapelor de dezvoltare.

Industria rapidă a soluțiilor AI este uneori martora eșecurilor proiectelor din cauza proceselor inadecvate de gestionare a proiectelor. Un ciclu de viață al AI bine definit, care ia în considerare toate etapele de dezvoltare, poate spori rata de succes a soluțiilor AI. Deși există diferite versiuni ale descrierilor ciclului de viață al IA, abordarea noastră sintetizează modelele propuse de Morley et al. (2020), Prem (2023) și modelul CRISP al Data Science Process Alliance (Hotz, 2018) prezentat de Rochel & Evequoz (2021).

Ciclul de viață al dezvoltării AI descrie parcursul unui sistem AI de la concepția inițială până la implementarea și întreținerea finală. Fiecare fază este vitală pentru asigurarea unui rezultat de succes. Ciclul de viață al proiectării este iterativ, ceea ce înseamnă că revizuirea etapelor anterioare poate fi necesară dacă anumite variabile nu funcționează conform așteptărilor. Deși considerentele etice pot varia în funcție de diferite etape, integrarea acestor considerente în fiecare fază a procesului este esențială. (Saltz, 2023)



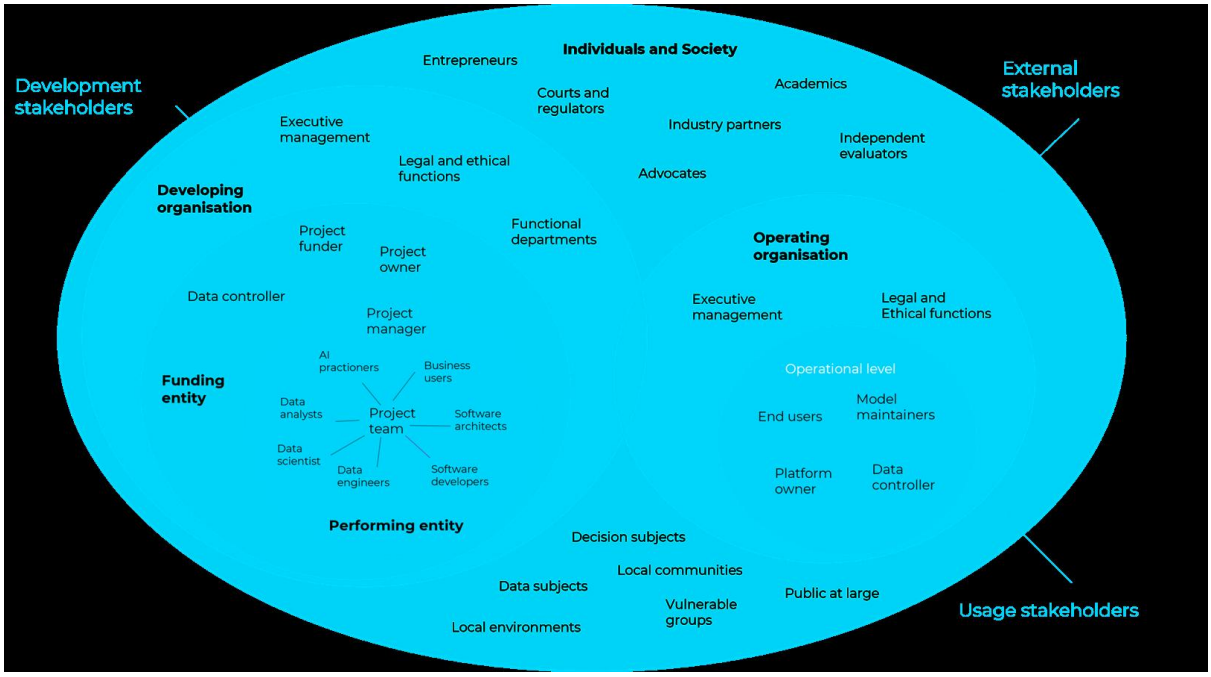
Imagine: Autorii documentului. Modificat ca o combinație a diferitelor modele descrise de articolele Data Science Process Alliance (2023), Morley et.al' (2019), Prehm's (2022) și Rochel et. all (2020).

9. Implicarea părților interesate în ciclul de viață al dezvoltării IA

Implicarea unei game variate de părți interesate în procesul de dezvoltare a IA este esențială pentru a evalua temeinic impactul etic al unui proiect de IA din perspective multiple. Consultarea părților interesate, inclusiv a celor direct afectate de soluția IA, este esențială pentru construirea unor sisteme demne de încredere (Morley et al., 2020) și pentru o evaluare cuprinzătoare a impactului proiectului (UNESCO, 2023a, p. 15). Prezentul capitol oferă o prezentare generală a diferitelor părți interesate implicate în proiectele de dezvoltare a IA și descrie rolurile acestora în diferite etape ale ciclului de viață al dezvoltării IA.

Părțile interesate sunt împărțite în părți interesate de dezvoltare, părți interesate de utilizare și părți interesate externe. Părțile interesate de dezvoltare și de utilizare sunt părți interesate implicate activ, de exemplu factorii de decizie care au puterea de a decide și proiectanții cu experiență. Părțile interesate externe nu sunt implicate în proiectul IA, dar pot afecta sau fi afectate de acesta sau pot fi consultate. (Miller, 2022)

În continuare, sunt examinate rolurile-cheie ale diferitelor părți interesate în dezvoltarea inteligenței artificiale și sunt clasificate în trei categorii: dezvoltare, utilizare și părți interesate externe. De asemenea, sunt explorate responsabilitățile și contribuțiile specifice ale fiecărui grup în diferite etape ale ciclului de viață al dezvoltării AI. Înțelegerea acestor roluri este esențială pentru a naviga prin complexitatea dezvoltării AI și pentru a asigura succesul proiectului.



Imagine: G.J.J: G.J. Miller (2022)

Părțile interesate de dezvoltare

În domeniul dezvoltării proiectelor de IA, Miller (2022) împarte părțile interesate în trei categorii principale: organizația, entitatea finanțatoare și entitatea executantă, fiecare jucând un rol distinct în ciclul de viață al proiectului. Organizația beneficiază de rezultatele proiectului AI, valorificând progresele și inovațiile pentru creșterea și succesul său. Entitatea finanțatoare, esențială pentru sprijinul financiar, nu numai că finanțează, ci și guvernează proiectul, asigurându-se că investițiile sunt aliniate la obiectivele proiectului și la standardele de guvernanță. Entitatea executantă este esențială, fiind responsabilă de generarea rezultatelor proiectului și de crearea sistemului IA, întruchipând execuția tehnică a proiectului. (Miller, 2022)

În cadrul organizației, rolurile se întind de la organizația de dezvoltare, care sponsorizează și definește domeniul de aplicare al proiectului, până la managementul executiv, care supraveghează traiectoria proiectului. Departamentele funcționale oferă sprijin esențial, în timp ce funcțiile juridice și etice asigură alinierea proiectului la standardele de reglementare și la normele etice. Grupul de gestionare a proiectelor joacă un rol de sprijin, ajutând managerii de proiect cu sarcini administrative pentru a eficientiza executarea proiectului (Zwikael & Meredith, 2018).

Entitatea de finanțare cuprinde roluri precum finanțatorul proiectului, care furnizează resurse financiare, și proprietarul proiectului, care imprimă direcția strategică și, eventual, prezidează comitetul director, asigurând alinierea proiectului la obiectivele strategice. În scenariile în care finanțatorul nu este direct implicat, responsabilitatea este delegată pentru a asigura o supraveghere continuă. Controlorul de date, un rol adesea asociat cu entitatea finanțatoare, gestionează utilizarea datelor, asigurând conformitatea cu cadrele juridice și de reglementare stricte pentru a evita eventualele sancțiuni (Miller, 2022).

În centrul entității de execuție se află managerii de proiect și echipa de proiect, responsabili pentru realizarea rezultatelor proiectului în conformitate cu planul aprobat. Această colaborare asigură progresul proiectului de la concepție la finalizare, fiecare membru al echipei contribuind la dezvoltarea și implementarea cu succes a sistemului IA (Zwikael & Meredith, 2018).

Această organizare structurată a părților interesate și a rolurilor acestora subliniază abordarea colaborativă și multidimensională necesară pentru dezvoltarea și implementarea cu succes a proiectelor de inteligență artificială, subliniind importanța unor roluri și responsabilități clare și a respectării standardelor juridice și etice.

Părțile interesate de utilizare

În contextul proiectelor de inteligență artificială, părțile interesate de utilizare joacă un rol crucial, în special la nivel organizațional, unde organizația operațională funcționează ca o parte interesată client care achiziționează soluția de inteligență artificială. Similar organizației de dezvoltare, organizația operațională cuprinde roluri precum managementul executiv și funcțiile juridice și etice, evidențiind structurile paralele atât în faza de dezvoltare, cât și în cea operațională (Miller, 2022).

La nivel operațional, etapa de utilizare implică o varietate de actori-cheie, inclusiv custozi de date, utilizatori finali, proprietari de platforme și manageri de modele. Utilizatorii finali, fie că sunt persoane fizice sau grupuri, interacționează cu sistemul de inteligență artificială fie prin contracte de muncă sau de servicii cu organizația de exploatare, fie ca consumatori, implicându-se direct în funcționalitățile sistemului (Miller, 2022).

Responsabilii cu menținerea modelelor își asumă sarcina esențială de gestionare, actualizare și îmbunătățire continuă a modelelor de învățare automată care conduc sistemul AI, asigurând eficiența și eficacitatea acestuia. Responsabilii cu prelucrarea datelor sunt însărcinați cu definirea obiectivelor și metodelor de prelucrare a datelor cu caracter personal în cadrul sistemului, asigurând conformitatea cu legislația privind protecția datelor. În cele din urmă, proprietarii de platforme poartă responsabilitatea pentru infrastructura generală a sistemului de inteligență artificială și pentru implementarea acestuia, menținând aspectele fundamentale pe baza cărora funcționează inteligența artificială (Miller, 2022).

Părțile interesate externe

Miller (2022) clasifică părțile interesate externe din proiectele IA în trei subgrupuri principale: indivizi, societate și reprezentanți, fiecare având roluri și contribuții distincte la proiect.

Subgrupul persoanelor fizice cuprinde persoanele vizate de date, persoanele vizate de decizii și lucrătorii care interacționează cu organizația de dezvoltare sau de exploatare prin acorduri precum contractele de servicii sau de muncă. Aceste părți interesate includ persoane direct implicate în proiect sau afectate de acesta. Aceste părți interesate joacă roluri cruciale în calitate de subiecți de date, subiecți de decizie și lucrători în cadrul proiectului (Miller, 2022).

Subgrupul societate surprinde impactul mai larg asupra populației generale, inclusiv asupra comunităților locale, mediului și grupurilor vulnerabile, cum ar fi persoanele cu handicap, grupurile minoritare, minorii și publicul larg. Acest segment este direct influențat de dezvoltarea și utilizarea sistemului de inteligență artificială, punând accentul pe impactul societal al proiectelor de inteligență artificială (Miller, 2022).

Reprezentanții, care formează al treilea subgrup, se disting de reprezentanții formali. În timp ce reprezentanții includ grupuri și organizații precum antreprenorii și evaluatorii independenți (reporteri, auditori sau recenzori) care pot să nu fie direct afectați de sistemul IA, dar care joacă un rol în dezvoltarea sau evaluarea acestuia, reprezentanții formali precum instanțele, legile, autoritățile de reglementare și sindicatele reprezintă supravegherea și guvernanța juridică și de reglementare în cadrul proiectului (Miller, 2022; Rodrigues, 2020).

Implicarea părților interesate în dezvoltarea soluțiilor AI

Co-crearea se referă la angajamentul de colaborare cu părțile interesate, inclusiv utilizatorii finali sau clienții, pe parcursul procesului de proiectare. Această abordare facilitează schimbul de perspective între participanții cu roluri și expertiză variate. (Interaction Design Foundation - IxDF, 2021; Sanin, 2020)

Practica co-creării poate fi realizată prin intermediul atelierelor facilitate care cuprind o varietate de activități, cum ar fi spargerea gheții, jocurile de rol,

cartografierea parcursului, brainstorming-ul și prototiparea. Prin co-creație, proiectanții pot înțelege în profunzime nevoile și dorințele clienților sau ale utilizatorilor finali, creând astfel soluții care țin seama de diversele perspective ale utilizatorilor. Această metodă incluzivă garantează că serviciile dezvoltate sunt bine aliniate la cerințele reale ale celor pe care își propun să îi deservească. (Interaction Design Foundation - IxDF, 2021; Sanin, 2020)

10. Ciclul de viață al dezvoltării IA: Definirea problemei și înțelegerea afacerii

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Această secțiune va explica în detaliu ciclul de viață al dezvoltării unei soluții AI, implicațiile sale etice și părțile interesate aferente. Pentru a stimula o discuție vie, vă rugăm să folosiți un caz de exemplu pentru a parcurge întregul ciclu de viață. Ca exemplu, puteți utiliza sistemul de recomandări bazat pe inteligență artificială cu privire la ce să cumpărați prin intermediul acestui link: https://www.datascience-pm.com/ai-lifecycle/#An_Example_Ai_Project_Life_Cycle

Ca orice proiect, și ciclul de viață al IA ar trebui să înceapă cu definirea problemei. O expunere bine definită a problemei oferă o înțelegere profundă a nevoii clientului și, prin urmare, o direcție pentru proiect. În acest moment, ar trebui efectuat un studiu de fezabilitate pentru a înțelege dacă IA este cea mai potrivită soluție pentru rezolvarea problemei. De asemenea, cercetarea și implicarea părților interesate sunt necesare pentru a înțelege problema în profunzime. (De Silva & Alahakoon, 2022, pp. 4-5)

Principiile etice în etapa de definire a problemei și de înțelegere a activității

În faza inițială a unui proiect, este esențial să se ia în considerare aspectele etice principale, deoarece acestea pun bazele parcursului proiectului și ale impactului societal ulterior. În special în această etapă, principiile beneficenței și non-maleficenței sunt considerate primordiale. (Prem, 2023, p. 703)

În ceea ce privește binefacerile, accentul se pune pe evaluarea faptului dacă sistemul de inteligență artificială promovează în mod activ bunăstarea individuală și beneficiile pentru societate. Obiectivele sistemului trebuie să fie transparente și să vizeze beneficii tangibile, susținând în același timp drepturile fundamentale și luând în considerare durabilitatea mediului. Este esențial să se implice o varietate de părți interesate, asigurându-se că perspectivele celor afectați de soluția IA sunt auzite și integrate. (De Silva & Alahakoon, 2022, p. 5; Morley et al., 2020, p. 2151)

În schimb, non-maleficența implică asigurarea faptului că sistemul de inteligență artificială nu aduce prejudicii persoanelor sau comunităților. Acest lucru necesită o abordare proactivă pentru identificarea și atenuarea potențialelor efecte adverse, protejând cu vigilență bunăstarea fizică și mentală a persoanelor

afectate de sistem. (De Silva & Alahakoon, 2022, p. 5; Morley et al., 2020, p. 2151) Preocupările generale privind siguranța trebuie abordate proactiv și integrate în proiectarea sistemului (UNESCO, 2023b, p. 18).

Din punctul de vedere al justiției, este imperativ să se analizeze implicațiile mai largi ale sistemului IA asupra instituțiilor, democrației și structurilor societale. Este esențial să se analizeze dacă proiectul ar putea crea în mod involuntar prejudecăți, favorizând anumite grupuri în detrimentul altora, sau ar putea încălca autonomia individuală. Fiecare dintre aceste aspecte necesită o analiză atentă pentru a se asigura că sistemul IA contribuie la o societate corectă și echitabilă. (Morley et al., 2020, p. 2151)

Părțile interesate în etapa de definire a problemei și de înțelegere a activității

În această etapă, este esențial să se implice cel puțin utilizatorii finali, care sunt principalii utilizatori ai soluției AI. În plus, dezvoltatorii AI ar trebui incluși încă de la început pentru a obține o înțelegere aprofundată a problemei pe care vor fi însărcinați să o rezolve. Eticienii sau specialiștii în științe sociale ar trebui, de asemenea, să fie implicați pentru a evalua impactul potențial al soluțiilor asupra oamenilor și societății. De asemenea, este esențial să se includă experți în domeniu pentru a spori cunoștințele despre industria în care se aplică IA. (De Silva & Alahakoon, 2022)

11. Ciclul de viață al dezvoltării IA: Etapa de proiectare

În timpul etapei de proiectare, munca de bază efectuată în etapa de definire a problemei evoluează într-un plan cuprinzător pentru implementarea proiectului de sistem de inteligență artificială. Această etapă include definirea obiectivelor proiectului și a rezultatelor dorite, elaborarea unui plan de proiect cu termene și considerente bugetare, precum și o înțelegere a cerințelor părților interesate. (De Silva & Alahakoon, 2022, p. 4)

Activitățile-cheie includ elaborarea unei specificații a cerințelor care formulează funcțiile sistemului de inteligență artificială și parametrii de măsurare pentru succesul proiectului. În plus, sunt rafinate strategiile de extragere a datelor, inclusiv, de exemplu, deciziile privind utilizarea modelelor pre-antrenate. De-a lungul acestei etape, considerentele etice fac parte integrantă și trebuie luate în considerare în fiecare aspect al proiectării sistemului. (De Silva & Alahakoon, 2022, p. 3)

Principii etice în etapa de proiectare

În timpul fazei de proiectare a soluției de inteligență artificială, este esențial să se utilizeze strategii proactive de tipul "ce se întâmplă dacă" pentru a prevedea și a atenua potențialele provocări de-a lungul ciclului de viață al soluției de inteligență artificială. Prin formularea unor întrebări critice precum: "Ce se întâmplă dacă datele utilizate sunt părtinitoare?" și prin explorarea mecanismelor necesare pentru detectarea și abordarea comportamentului nedrept, dezvoltatorii pot aborda preventiv aceste probleme. Stabilirea unor măsuri concrete este imperativă pentru a oferi dezvoltatorilor cunoștințele necesare pentru a evita prejudecățile legate, de exemplu, de rasă, culoare, descendență, sex, vârstă, limbă, religie, opinie politică, origine națională sau etnică, origine socială, statut economic, condiții de naștere sau handicap. În plus, este esențial să se instituie canale de comunicare solide, protocoale de raportare și măsuri care să reacționeze rapid la prejudecăți și la ajustările pe care acestea le-ar putea necesita. (Morley et al., 2020, pp. 2151, 2155; UNESCO, 2023b, pp. 13, 17)

În plus față de aceste măsuri de precauție, cerințele de proiectare ar trebui să articuleze în mod clar metodele de sporire a transparenței procesului, îmbunătățind astfel explicabilitatea. Această abordare subliniază importanța de a face operațiunile sistemului IA inteligibile și responsabile (Morley et al., 2020, pp. 2151, 2155; UNESCO, 2023b, pp. 13, 17).

În plus, Prem (2023, p. 703) subliniază valoarea implicării părților interesate în timpul acestei etape și rolul esențial al mecanismelor de supraveghere umană. Această includere asigură faptul că o gamă diversă de perspective contribuie la dezvoltarea IA, promovând un proces de proiectare mai echitabil și mai informat.

Părțile interesate în etapa de proiectare

În această fază, este important să se implice arhitecți de software, deoarece aceștia joacă un rol crucial în stabilirea bazei tehnice a soluției AI și iau în considerare atât cerințele actuale, cât și scalabilitatea viitoare (Peter, 2024). Designerii UX sunt esențiali pentru a crea interfețe care să faciliteze adoptarea și satisfacția utilizatorilor și pentru a se asigura că aceștia pot interacționa fără probleme cu sistemul AI (White, 2023). Practicienii AI trebuie să colaboreze îndeaproape cu designerii UX pentru a integra fără probleme componentele AI în interfața utilizatorului. Experții în etică abordează preocupările etice în dezvoltarea IA. Funcțiile juridice și etice sunt necesare pentru a reduce riscurile juridice asociate proiectelor AI.

Potrivit lui De Silva și Alahakoon (2022), se recomandă utilizarea unei abordări de proiectare participativă care să implice persoanele și comunitățile afectate de modelul IA și de deciziile acestuia, promovând transparența, incluziunea și alinierea la valorile societale pe tot parcursul ciclului de viață al proiectului.

12. Ciclul de viață al dezvoltării IA: Colectarea, înțelegerea și pregătirea datelor

După stabilirea obiectivului și elaborarea unei strategii, următoarea etapă constă în obținerea și pregătirea datelor relevante. Aceasta include:

- Colectarea datelor
- Definirea și categorisirea acesteia
- Efectuarea unei analize exploratorii
- Validarea relevanței sale
- Selectarea informațiilor esențiale
- Curățarea, reconstrucția, integrarea și formatarea acestuia pentru a se potrivi obiectivelor proiectului

Abordarea și compensarea valorilor lipsă este esențială. Este esențial să se înțeleagă ipotezele care stau la baza datelor existente și să se asigure claritatea modului în care seturile de date sunt utilizate în procesele decizionale. Menținerea calității datelor este extrem de importantă, în special atunci când datele sunt incomplete sau ambigue. (Rochel & Evéquoz, 2021, pp. 614-617)

Deși această fază necesită mult timp, ea este, de asemenea, fundamentală. Fiabilitatea și performanța soluției AI rezultate sunt direct corelate cu calitatea și precizia datelor de bază. (Saltz, 2023)

Principiile etice în etapa de colectare, înțelegere și pregătire a datelor

Faza de colectare a datelor prezintă riscuri inerente, în special în ceea ce privește principiul non-maleficenței. Asigurarea vieții private și a confidențialității datelor este esențială în timpul acestei faze. Sistemele de inteligență artificială trebuie să respecte confidențialitatea datelor în mod consecvent, de la început și pe tot parcursul ciclului de viață al soluției. Reglementările legale, cum ar fi Legea globală privind protecția datelor, reglementează de obicei aspectele legate de confidențialitate. (De Silva & Alahakoon, 2022, p. 5)

Prejudecățile din cadrul datelor pot duce la suprareprezentarea anumitor categorii demografice, astfel încât este esențial să se analizeze și să se asigure calitatea și integritatea datelor. În plus, seturile de date sunt expuse amenințărilor cibernetice, necesitând măsuri de securitate solide pentru a le proteja de eventualele atacuri. (UNESCO, 2023a, pp. 9, 18)

Transparența cu privire la procesele de colectare a datelor, inclusiv ce date sunt colectate, în ce scopuri și cum vor fi utilizate în sistemul de inteligență artificială este, de asemenea, valabilă în acest moment. (De Silva & Alahakoon, 2022, p. 6)

Părțile interesate în etapa de colectare, înțelegere și pregătire a datelor

Implicarea cercetătorilor de date în această fază este crucială, deoarece aceștia joacă un rol esențial în discernerea modelelor și tendințelor din cadrul datelor, transformându-le în cunoștințe utilizabile (Miller, 2022). În mod similar, inginerii AI sunt indispensabili pentru evaluarea și verificarea integrității datelor (Rochel & Evéquoz, 2021), în timp ce analiștii de date îmbunătățesc proiectul prin înțelegerea caracteristicilor datelor și consilierea cu privire la strategiile eficiente de pregătire a datelor. În plus, contribuția inginerilor de date este vitală pentru pregătirea și întreținerea datelor, asigurând calitatea acestora (Miller, 2022).

Includerea unui responsabil cu protecția datelor (DPO) este importantă pentru a asigura respectarea standardelor juridice și etice privind datele cu caracter personal (Autoritatea Europeană pentru Protecția Datelor, 2024).

Eticienii joacă un rol important prin integrarea considerațiilor etice în procesul de manipulare și selecție a datelor. Implicarea acestor diverse părți interesate echipează proiectul AI cu date obținute în mod etic și pregătite meticulos, stabilind o bază de încredere și fiabilitate pentru soluția AI.

13. Ciclul de viață al dezvoltării inteligenței artificiale; dezvoltarea și formarea modelelor

În timpul fazei de dezvoltare și formare a modelului, crearea unui model de inteligență artificială vizează în mod specific abordarea provocării identificate în prealabil. Această fază include selectarea atentă, configurarea și utilizarea intenționată a instrumentelor algoritmice, ceea ce face esențială documentarea și evaluarea aprofundată a implicațiilor acestor alegeri, în special identificarea oricăror dezavantaje potențiale sau compromisuri necesare. Caracterul iterativ al acestei etape este o caracteristică definitorie, implicând mai multe runde de rafinare pentru a îmbunătăți acuratețea și eficiența modelului.

Procesul de selectare a instrumentelor algoritmice implică adesea navigarea între obiective contradictorii. Un prim exemplu este provocarea de filtrare a e-mailurilor spam, în care este necesar să se echilibreze obiectivul de reducere la minimum a spamului în căsuțele de primire ale utilizatorilor cu riscul de a clasifica incorect e-mailurile legitime drept spam. Atingerea echilibrului optim este esențială pentru îmbunătățirea eficacității modelului fără a compromite fiabilitatea acestuia sau încrederea utilizatorilor. (Rochel & Evéquoz, 2021, pp. 617-618; Saltz, 2023)

Principii etice în etapa de dezvoltare și formare a modelului

Deși detaliile tehnice au adesea întâietate, această etapă oferă o bună oportunitate de a asigura explicabilitatea și interpretabilitatea modelului ales, precum și de a evalua dacă există eventuale prejudecăți. (Prem, 2023, p. 703)

Procesul de dezvoltare trebuie să fie clar și inteligibil pentru toate părțile implicate, necesitând o documentare temeinică a surselor de date, a alegerilor de modele și a logicii care a stat la baza acestor decizii. (Morley et al., 2020, p. 2151)

Atunci când sunt necesare compromisuri, mai ales că eficiența implică adesea compromisuri, este esențială o abordare sistematică pentru recunoașterea și evaluarea deschisă a efectelor acestora. Inginerii AI trebuie să fie pregătiți să își raționalizeze alegerile, obligându-i să ia în considerare și să atenueze orice consecințe negative asupra celor afectați. (Morley et al., 2020, p. 2151)

Părțile interesate în etapa de dezvoltare a modelului și de formare

Oamenii de știință din domeniul inteligenței artificiale sau al învățării automate joacă un rol esențial în transformarea definițiilor problemelor în prototipuri inițiale de modele de inteligență artificială (De Silva & Alahakoon, 2022), marcând implicarea lor critică pe măsură ce proiectul avansează în faza de dezvoltare și formare a modelului. În această fază, inginerii AI devin indispensabili datorită capacității lor de a identifica și de a utiliza cele mai potrivite instrumente algoritmice care sunt aliniate în mod specific cu obiectivele proiectului (Rochel & Evéquoz, 2021). În plus, oamenii de știință din domeniul datelor se află în centrul acestei faze, conducând dezvoltarea și perfecționarea continuă a modelelor de învățare automată către o mai mare acuratețe și eficiență.

Includerea experților în etică este extrem de importantă pentru a asigura integrarea considerațiilor etice în procesul de dezvoltare, cu scopul de a evita orice impact negativ potențial. În plus, informațiile furnizate de utilizatorii din mediul de afaceri sunt de o valoare inestimabilă, garantând că modelele în curs de dezvoltare sunt în concordanță cu strategiile de afaceri și îndeplinesc rezultatele așteptate, reducând astfel decalajul dintre inovarea tehnică și aplicațiile practice de afaceri.

14. Ciclul de viață al dezvoltării AI: Evaluarea și testarea modelelor

După ce modelul AI a fost dezvoltat și antrenat, eficacitatea acestuia trebuie testată și evaluată în raport cu obiective și criterii de succes predefinite. Această evaluare nu măsoară doar performanța modelului, ci examinează și ipotezele de bază și criteriile de selecție ale seturilor de date utilizate. În cazurile în care modelul nu îndeplinește așteptările, pot fi necesare ajustări, fie în ceea ce privește parametrii modelului, arhitectura sa generală sau seturile de date pe care se bazează. Comunicarea transparentă a oricăror deficiențe cu șeful de proiect și cu toate părțile interesate relevante este esențială, având în vedere impactul potențial semnificativ asupra persoanelor implicate. Pe baza rezultatelor evaluării, echipa trebuie să decidă cu privire la acțiunile ulterioare, care ar putea implica

iterații suplimentare pentru rafinare sau trecerea la faza de implementare, asigurându-se că fiecare pas este clar aliniat și convenit. (Hotz, 2018; Rochel & Evéquoz, 2021, p. 619)

Principii etice în etapa de evaluare și testare a modelului

În această fază, este esențial să se testeze respectarea de către model a orientărilor etice, asigurându-se că acesta se aliniază la principiile esențiale stabilite la început. Dezvoltarea și perfecționarea auditurilor și a parametrilor de auditabilitate reprezintă sarcini esențiale. Este important să se includă principii etice esențiale, cum ar fi justiția, beneficența și non-maleficența. Evaluarea rezilienței modelului la potențiale amenințări la adresa securității și stabilirea unei strategii pentru provocări neașteptate sunt etape esențiale. Ar trebui să existe mecanisme clare pentru acțiuni corective în cazul unor rezultate negative. Este necesară o evaluare aprofundată care să cuprindă confidențialitatea, impactul social, potențialele prejudecăți și multe altele, cu angajamentul de a minimiza și de a raporta în mod transparent orice efecte negative. (De Silva & Alahakoon, 2022, p. 8; Morley et al., 2020, p. 2151; Prem, 2023, p. 703)

Transparența este esențială, nu numai în ceea ce privește funcționarea tehnică a modelului, ci și în ceea ce privește justificările pentru deciziile umane din cadrul proiectului, asigurând explicabilitatea (De Silva & Alahakoon, 2022, p. 8; Morley et al., 2020, p. 2151). În plus, este esențial să se recunoască faptul că inginerii s-ar putea confrunta cu presiuni în vederea accelerării procesului de dezvoltare pentru implementare, ceea ce ar putea introduce riscuri suplimentare (Rochel & Evéquoz, 2021, p. 618).

Părțile interesate în etapa de evaluare și testare a modelului

În această fază, inginerii AI joacă un rol esențial prin interpretarea rezultatelor modelării, oferind o perspectivă esențială asupra rezultatelor (Rochel & Evéquoz, 2021).

Echipele de asigurare a calității (QA) sunt esențiale pentru confirmarea faptului că modelele respectă standardele de calitate și sunt lipsite de erori sau de rezultate neintenționate. Implicarea experților în etică este esențială pentru garantarea implementării etice a modelelor AI, asigurându-se că acestea sunt utilizate în mod responsabil. Apărătorii sau reprezentanții utilizatorilor contribuie în mod semnificativ prin alinierea modelelor AI la așteptările și cerințele utilizatorilor, sporind astfel satisfacția și acceptarea acestora.

În plus, echipele juridice și de conformitate sunt indispensabile în abordarea riscurilor juridice, asigurându-se că modelele AI sunt în conformitate cu toate legile și reglementările aplicabile. (Moore, 2023)

15. Ciclul de viață al dezvoltării AI: Implementare

După o dezvoltare reușită, soluția AI este implementată într-un mediu de producție destinat rezolvării problemelor din lumea reală. Domeniul de aplicare și metoda de implementare depind de obiectivele specifice ale soluției și de sistemele sau aplicațiile pentru care este proiectată. De obicei, implementarea începe cu o fază pilot care implică un grup selectat de experți și de utilizatori timpurii care pot oferi feedback inițial (De Silva & Alahakoon, 2022, p. 8).

Soluția AI ar putea fi integrată în sistemele existente, ar putea servi drept bază pentru o nouă aplicație sau serviciu, sau constatările sale ar putea fi împărtășite prin metode offline, cum ar fi rapoartele manageriale. Este esențial să se stabilească un mecanism de feedback pentru a monitoriza performanța și impactul acesteia, facilitând îmbunătățirea continuă pe baza utilizării în lumea reală (Saltz, 2023).

Pentru a aborda riscurile potențiale, sunt puse în aplicare strategii cuprinzătoare de gestionare a riscurilor. Sunt puse în aplicare măsuri operaționale și de monitorizare pentru a se asigura că soluția AI funcționează eficient după implementare. La final, se realizează o revizuire completă a proiectului, ceea ce duce la un raport detaliat care rezumă rezultatele proiectului și lecțiile învățate (De Silva & Alahakoon, 2022, pp. 8-9).

Principii etice în etapa de implementare

Morley (2020) subliniază faptul că considerentele etice primesc adesea o atenție insuficientă în timpul fazei de implementare a procesului de proiectare. Distilarea comportamentelor umane complexe în instrumente care sunt atât inteligibile, cât și transparente prezintă provocări semnificative. Cu toate acestea, crearea unor astfel de instrumente este esențială pentru sporirea încrederii în sistemele AI. Prin urmare, explicabilitatea apare ca un principiu etic esențial în această etapă.

În plus, asigurarea autonomiei este esențială prin informarea utilizatorilor finali atunci când interacționează cu un proces bazat pe AI. Această transparență este esențială pentru a evita dependența excesivă de sistemele AI. UNESCO (2023b, p. 36) subliniază necesitatea unor mecanisme eficiente în patru domenii-cheie: cunoașterea sistemului, proceduri de audit solide, claritate în raționamentul algoritmic și măsuri de evaluare a impactului, cum ar fi canale de apel sau reclamații, care ar trebui să includă protecții pentru avertizori. În plus, rolul esențial al supravegherii umane nu poate fi supraestimat, fiind necesare dispoziții care să permită intervenția și supravegherea umană a sistemelor de inteligență artificială pentru a asigura implementarea etică a acestora (Morley et al., 2020, p. 2151).

Părțile interesate în etapa de implementare

Inginerii din domeniul inteligenței artificiale sau al învățării automate joacă un rol esențial în transformarea modelului prototip de inteligență artificială într-un serviciu sau o soluție complet implementată, accesibilă părților interesate și utilizatorilor finali (De Silva & Alahakoon, 2022).

Experții în securitate cibernetică efectuează o evaluare a riscurilor pentru a securiza sistemul IA (Junklewitz et al., 2023, pp. 9-12). Utilizatorii finali contribuie semnificativ prin evidențierea oricăror probleme sau deficiențe care ar fi putut să nu fie evidente în timpul testării în mediu controlat. În plus, avocații sau reprezentanții utilizatorilor au un rol esențial în colectarea feedback-ului cu privire la utilitatea și performanța soluției AI implementate. Rolul lor este de a se asigura că soluția se aliniază așteptărilor și cerințelor utilizatorilor, facilitând îmbunătățirile acolo unde este necesar.

16. Ciclul de viață al dezvoltării IA: Funcționare și monitorizare

Întreținerea, actualizările și evaluările continue sunt esențiale pentru longevitatea și eficacitatea unui sistem de inteligență artificială operațional. Se efectuează evaluări constante ale performanțelor pentru a se asigura că sistemul îndeplinește standardele așteptate, în timp ce monitorizarea continuă a interacțiunilor cu utilizatorii oferă informații valoroase privind aplicarea în lumea reală și potențialele domenii de îmbunătățire. Actualizările regulate cu date noi sunt esențiale pentru a menține relevanța și acuratețea modelului. Feedback-ul utilizatorilor este esențial pentru perfecționarea continuă a modelului AI, subliniind importanța îmbunătățirilor iterative ca parte a ciclului de întreținere, simbolizând mai degrabă progresul decât eșecul.

În plus, măsurarea rentabilității investițiilor sistemului (ROI) și a altor parametri critici este vitală pentru a evalua succesul acestuia în atingerea obiectivelor predefinite. (De Silva & Alahakoon, 2022, p. 9; Saltz, 2023)

Principii etice în etapa de operare și monitorizare

În această fază, monitorizarea performanței se extinde dincolo de parametrii tehnici pentru a include conformitatea cu standardele etice. Reevaluarea periodică a tuturor principiilor este esențială, în special având în vedere posibilele modificări ale seturilor de date și ale structurii modelului. Implicarea părților interesate este esențială chiar și după implementare, contribuția acestora fiind esențială pentru îmbunătățirile continue. Ar trebui să se solicite în mod activ feedback din partea utilizatorilor, a organizațiilor și a societății în general. Ar trebui să existe mecanisme structurate pentru a colecta în mod consecvent și a încorpora acest feedback în eforturile de îmbunătățire continuă. (Grupul de

experți la nivel înalt privind inteligența artificială (AI HLEG), 2019, p. 19; UNESCO, 2023b, p. 15)

Părțile interesate în etapa de operare și monitorizare

În timpul fazei de operare și monitorizare, echipele DevOps și IT Operations sunt esențiale pentru monitorizarea performanței sistemului și menținerea disponibilității continue (Immersant Data Solutions, 2023). Evaluatorii independenți, cum ar fi certificatorii de siguranță, pot evalua securitatea soluției AI dezvoltate (Miller, 2022). În plus, grupurile de experți, comitetele directoare sau organismele de reglementare efectuează revizuiri ale proiectului, acoperind atât considerații tehnice, cât și etice (De Silva & Alahakoon, 2022).

17. Concluzii

Această unitate a aprofundat principiile, cadrele și orientările etice care sunt esențiale în modelarea dezvoltării soluțiilor AI. Ea a explorat ciclul de viață al dezvoltării AI, de la enunțarea problemei inițiale până la operarea și monitorizarea soluției AI, cu un accent deosebit pe integrarea principiilor conexe și pe implicarea părților interesate în fiecare etapă.

Dezvoltarea etică a IA evidențiază importanța încorporării considerentelor etice pe parcursul întregului ciclu de viață al dezvoltării IA. Această abordare asigură faptul că tehnologiile IA nu numai că contribuie în mod pozitiv la societate, ci și abordează în mod eficient riscurile potențiale, cum ar fi prejudecățile și preocupările legate de confidențialitate. În plus, a fost subliniată importanța implicării unui spectru larg de părți interesate, ilustrând faptul că o dezvoltare reușită a IA necesită contribuții dincolo de expertiza tehnică pentru a include diverse perspective. Această abordare incluzivă îmbogățește procesul de dezvoltare, asigurând alinierea soluțiilor IA la valorile societății și obținerea unei acceptări mai largi.

În plus, necesitatea monitorizării continue și a perfecționării iterative a sistemelor de inteligență artificială după implementare a fost subliniată ca fiind esențială pentru menținerea integrității lor etice și asigurarea succesului lor susținut. Adoptând o atitudine proactivă față de inovare, dezvoltatorii pot identifica și aborda provocările etice pe măsură ce acestea apar, navigând astfel prin complexitatea dezvoltării AI cu un angajament față de standardele etice și inovarea responsabilă.

18. Referințe

Ayling, J., & Chapman, A. (2022). Punerea în aplicare a eticii IA: Sunt instrumentele adecvate scopului? AI and Ethics, 2(3), 405-429. https://doi.org/10.1007/s43681-021-00084-x
Academia de Inteligență Artificială din Beijing-Principiile IA din Beijing. (2019). Datenschutz Und Datensicherheit, 10.
Principiile inteligenței artificiale de la Beijing. (2022, 10 ianuarie). Centrul internațional de cercetare pentru etică și guvernanță în domeniul IA. https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/
De Silva, D., & Alahakoon, D. (2022). Un ciclu de viață al inteligenței artificiale: De la concepție la producție. Patterns, 3(6), 100489. https://doi.org/10.1016/j.patter.2022.100489
Comisia Europeană. (2023, 12 septembrie). Comisia salută acordul politic privind Legea IA [Text]. Comisia Europeană - Comisia Europeană. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473
Autoritatea Europeană pentru Protecția Datelor. (2024, 9 aprilie). Responsabilul cu protecția datelor (DPO) | Autoritatea Europeană pentru Protecția Datelor. https://www.edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en
Parlamentul European. (2023, 8 iunie). Actul UE privind inteligența artificială: Primul regulament privind inteligența artificială. Subiecte | Parlamentul European. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
Floridi, L., & Cowls, J. (2019). Un cadru unificat de cinci principii pentru IA în societate. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8cd550d1
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An Ethical Framework for a Good AI Society: Oportunități, riscuri, principii și recomandări. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5
Institutul Viitorul Vieții. (2024). The AI Act Explorer | Legea UE privind inteligența artificială. https://artificialintelligenceact.eu/ai-act-explorer/
Grupul de experți la nivel înalt privind inteligența artificială (AI HLEG). (2019). Liniile directoare etice ale inteligenței artificiale demne de încredere. Grupul de experți la nivel înalt privind Artificiale Inteligență Artificială. Comisia Europeană. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
Grupul de experți la nivel înalt privind inteligența artificială (AI HLEG). (2020). Lista de evaluare pentru inteligența artificială demnă de încredere (ALTAI) pentru autoevaluare. Comisia Europeană.
Hillard, A., & Gulley, A. (2023, 9 decembrie). Ce este Actul UE privind inteligența artificială? https://www.holisticai.com/blog/eu-ai-act
Hotz, N. (2018, 10 septembrie). Ce este CRISP DM? Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2/

Immersant Data Solutions. (2023, 9 iunie). DevOps: Reducerea decalajului dintre dezvoltare și operațiuni | LinkedIn. https://www.linkedin.com/pulse/devops-bridging-gap-between-development-operations/
Interaction Design Foundation - IxDF. (2021, 27 octombrie). Ce este co-crearea? Interaction Design Foundation-IxDF. The Interaction Design Foundation. https://www.interaction-design.org/literature/topics/co-creation
Jobin, A., Ienca, M., & Vayena, E. (2019). Peisajul global al liniilor directoare privind etica IA. Nature Machine Intelligence, 1(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2
Junklewitz, H., Hamon, R., André, A., Evas, T., Soler Garrido, J., & Sanchez Martin, J. I. (2023). Securitatea cibernetică a inteligenței artificiale în Legea privind inteligența artificială: Principii directoare pentru abordarea cerinței de securitate cibernetică pentru sistemele de inteligență artificială cu risc ridicat. Oficiul pentru Publicații al Uniunii Europene. https://data.europa.eu/doi/10.2760/271009
Leijnen, S., Aldewereld, H., van Belkom, R., Bijvank, R., & Ossewaarde, R. (2020). Un cadru agil pentru AI demn de încredere. 75-78.
Miller, G. J. (2022). Rolurile părților interesate în proiectele de inteligență artificială. Project Leadership and Society, 3, 100068. https://doi.org/10.1016/j.plas.2022.100068
Moore, S. (2023, 18 decembrie). Ce este... o echipă juridică și de conformitate corporativă? | LinkedIn. https://www.linkedin.com/pulse/what-isa-corporate-legal-compliance-team-steven-moore-iczqf/
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
OCDE. (2019). Prezentare generală a principiilor IA. Prezentare generală a principiilor IA ale OCDE. https://oecd.ai/en/principles
Peter, A. (2024, 2 octombrie). AI Software Architecture: Navigarea în elementele esențiale ale dezvoltării aplicațiilor de afaceri | LinkedIn. https://www.linkedin.com/pulse/ai-software-architecture-navigating-essentials-business-alexx-peter-uugnc/
Prem, E. (2023). De la cadre etice de IA la instrumente: O revizuire a abordărilor. AI and Ethics, 3(3), 699-716. https://doi.org/10.1007/s43681-023-00258-9
Rochel, J., & Evéquoz, F. (2021). Intrarea în sala motoarelor: Un plan pentru a investiga etapele obscure ale eticii IA. AI & SOCIETY, 36(2), 609-622. https://doi.org/10.1007/s00146-020-01069-w
Rodrigues, R. (2020). Aspecte juridice și legate de drepturile omului ale IA: lacune, provocări și vulnerabilități. Journal of Responsible Technology, 4, 100005. https://doi.org/10.1016/j.jrt.2020.100005
Saltz, J. (2023, 1 iunie). Ce este ciclul de viață al IA? Data Science Process Alliance. https://www.datascience-pm.com/ai-lifecycle/
Sanin, A. (2020, 4 septembrie). Co-Creation How-To's. Design Globant. https://medium.com/design-globant/co-creation-how-tos-e25696f56d6f

UNESCO. (2023a). Evaluarea impactului etic. Un instrument al Recomandării privind etica inteligenței artificiale. UNESCO. https://doi.org/10.54678/YTSA7796

UNESCO. (2023b). Metodologia de evaluare a pregătirii. Un instrument al Recomandării privind etica inteligenței artificiale. UNESCO. https://doi.org/10.54678/YHAA4429

UNESCO. (2023c, 21 aprilie). Inteligența artificială: Exemple de dileme etice | UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases

White, C. (2023, 30 noiembrie). Ce face de fapt un designer UX? [Ghid 2024]. https://careerfoundry.com/en/blog/ux-design/what-does-a-ux-designer-actually-do/

Zhou, J., & Chen, F. (2022). Etica IA: De la principii la practică. AI & SOCIETY. https://doi.org/10.1007/s00146-022-01602-z

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). Un studiu privind principiile etice ale IA și implementările. 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 3010-3017. https://doi.org/10.1109/SSCI47803.2020.9308437

Zwikael, O., & Meredith, J. R. (2018). Cine este cine în grădina zoologică a proiectului? Cele zece roluri de bază ale proiectului. International Journal of Operations & Production Management, 38(2), 474-492. https://doi.org/10.1108/IJOPM-05-2017-0274

CU2 | Confidențialitate și confort AI

Index

1. Introducere

În era digitală, progresul rapid al inteligenței artificiale (AI) a transformat nenumărate aspecte ale vieții cotidiene, remodelând modul în care interacționăm cu tehnologia și între noi. În centrul acestei transformări se află interacțiunea dinamică dintre doi factori critici: confidențialitatea și confortul. Acest CU aprofundează relația complexă dintre aceste elemente, explorând provocările și oportunitățile pe care le prezintă în era tehnologiilor inteligente.

Pe măsură ce integrăm din ce în ce mai mult IA în viața noastră personală și profesională, devine imperativ să examinăm în mod critic echilibrul dintre îmbunătățirea confortului utilizatorului și protejarea vieții private. Acest CU este conceput pentru a oferi participanților o înțelegere cuprinzătoare a modului în care viața privată și confortul coexistă armonios în cadrul aplicațiilor AI. Ea subliniază atât potențialul acestora de a îmbunătăți viața de zi cu zi, cât și considerentele etice pe care le implică.

Participanții vor dobândi cunoștințe despre principiile fundamentale care guvernează confidențialitatea datelor, dilemele etice ridicate de inteligența artificială și cadrele juridice care modelează viitorul implementării tehnologiei. Prin explorarea aplicațiilor din lumea reală și a scenariilor ipotetice, acest curs își propune să ofere cursanților instrumentele necesare pentru a naviga în mod responsabil în peisajul în continuă evoluție al IA, asigurându-se că progresele tehnologice îmbunătățesc experiențele umane fără a compromite drepturile individuale.

Această călătorie prin domeniile inteligenței artificiale, confidențialității și confortului va provoca participanții să gândească critic la rolul tehnologiei în societate și la responsabilitatea lor ca viitori lideri în domeniul inteligenței artificiale.

Sursa imaginii | Generat de DALL-E

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

O întrebare captivantă sau un fapt surprinzător care să atragă interesul cursanților pentru a explica peisajul actual al IA, subliniind importanța confidențialității și a confortului.

Fapt surprinzător

Un studiu recent a arătat că sistemele de inteligență artificială pot lua decizii cu privire la cererile dvs. de angajare, scorurile de credit și chiar asistența medicală - adesea fără ca dvs. să fiți direct informat sau să știți. Această integrare a IA ridică întrebări cruciale: Cât de mult control avem asupra datelor noastre și ce înseamnă aceasta pentru viața noastră privată? Haideți să pătrundem în lumea complexă a inteligenței artificiale, în care confortul ar putea veni în detrimentul vieții noastre private.

Întrebare captivantă

Știați că o persoană obișnuită interacționează cu inteligența artificială de peste 20 de ori pe zi, fără să-și dea seama? De la recomandările personalizate pentru cumpărături până la dispozitivele inteligente de acasă care ne gestionează rutina zilnică, inteligența artificială este perfect integrată în viața noastră. Dar cât de des vă gândiți ce informații personale schimbați pentru acest confort? Să explorăm care este miza reală.

De asemenea, conținutul poate fi prezentat ca un sondaj.

Întrebare sondaj

"Ce servicii folosiți despre care știți că vă colectează datele pentru a îmbunătăți funcționalitatea și a vă personaliza experiența?"

Opțiuni

- Smartphone-uri
- Platforme social media
- Servicii de streaming
- Site-uri de comerț electronic
- Dispozitive inteligente pentru acasă
- Dispozitive de urmărire a condiției fizice și aplicații de sănătate
- Asistenți vocali

- Aplicații de navigație și călătorie
- Aplicații bancare și financiare
- Servicii de e-mail

Întrebare complementară

"Cât de confortabil vă simțiți cu faptul că aceste servicii vă colectează datele?"

Foarte confortabil || Oarecum confortabil || Neutru || Oarecum incomod || Foarte incomod

Această abordare ajută la identificarea conștientizării colectării de date în cadrul diferitelor servicii și surprinde sentimentele cu privire la confidențialitate și la utilizarea datelor personale, oferind o imagine cuprinzătoare a atitudinilor participanților față de confidențialitatea datelor.

2. Dilema lui Sarah: compromisul confidențialitate-convenință

Sursa imaginii | Generat de DALL-E

Imaginați-vă-o pe Sarah, o studentă care tocmai s-a mutat în primul ei apartament. Dornică să își eficientizeze viața aglomerată, Sarah investește în mai multe dispozitive inteligente: un difuzor inteligent pentru a-și gestiona programul și a asculta muzică, un termostat inteligent pentru a controla temperatura apartamentului și un ceas inteligent pentru a-și monitoriza rutinele de fitness.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sfat: puneți întrebări în punctele critice pentru a face relația mai ușoară

IMAGINI SURSĂ | Generat de DALL-E

Într-o dimineață, difuzorul inteligent al lui Sarah îi sugerează un nou traseu către campusul facultății sale, economisind timp în timpul unei perioade cu trafic intens - un exemplu perfect de confort. Mai târziu, Sarah primește pe telefon un anunț personalizat pentru o cafenea de pe noul traseu, care o tentează cu cafeaua ei cu lapte preferată. Inițial încântată, Sarah se întreabă,

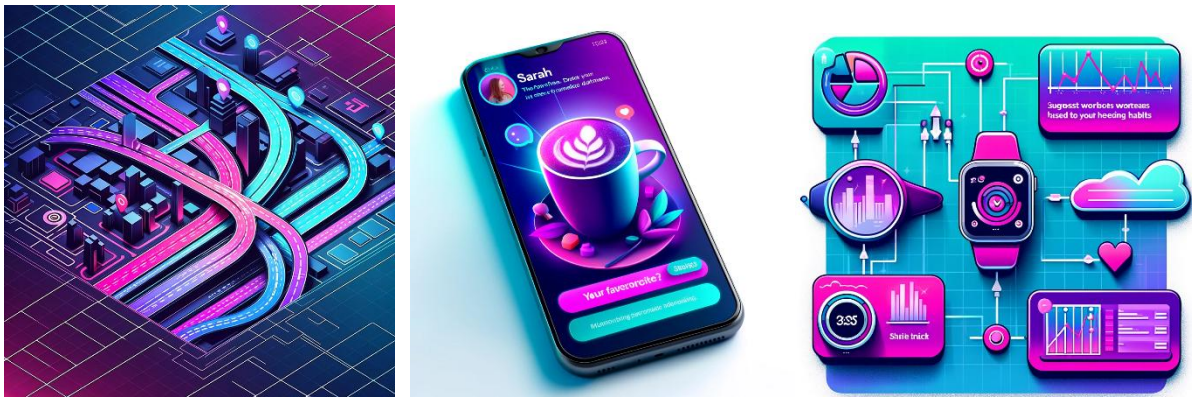
"Cum a știut?"

Întrebare

"Credeți că Sarah s-a gândit la implicațiile asupra vieții private înainte de a achiziționa aceste dispozitive? De ce sau de ce nu?"

Întrebare

- Ce părere aveți despre dispozitivele care iau decizii pe baza rutinei dumneavoastră zilnice? Care sunt posibilele compromisuri în materie de confidențialitate?
- Ce informații ar fi putut fi partajate pentru a personaliza această reclamă? Sunteți de acord cu acest nivel de partajare a datelor pentru reclamele personalizate?



IMAGINI SURSĂ | Generat de DALL-E

Pe măsură ce zilele trec, Sarah observă mai multe astfel de evenimente. Ceasul ei inteligent îi sugerează planuri de exerciții pe baza nivelului ei de formă fizică și a achizițiilor recente de fast-food, înregistrate de aplicația de plată. Termostatul său se adaptează la programul ei și pare să anticipeze schimbări neobișnuite în rutina ei, cum ar fi rămânerea acasă atunci când este bolnavă.

Întrebare

"În ce moment personalizarea devine prea invazivă? Unde ați trage linia pentru dumneavoastră?"

Sursa imaginii | Generat de DALL-E

Deși Sarah apreciază confortul oferit de aceste dispozitive bazate pe inteligență artificială, care îi fac viața mai ușoară și mai eficientă, începe să se simtă neliniștită cu privire la cantitatea de informații personale pe care le colectează și la modul în care acestea sunt utilizate. Ea nu a fost de acord în mod explicit să își

împărtășească locația sau istoricul achizițiilor. Încântarea lui Sarah față de facilitățile oferite devine rapid marcată de îngrijorarea cu privire la viața sa privată.

Această poveste ne prezintă echilibrul delicat dintre confidențialitate și comoditate în IA. Pe măsură ce tehnologiile IA devin din ce în ce mai integrate în viața noastră de zi cu zi, acestea oferă un confort fără precedent și ridică probleme semnificative legate de confidențialitate. Provocarea constă în găsirea unui echilibru în care Sarah și alți utilizatori ca ea să se poată bucura de avantajele AI fără a simți că informațiile lor personale sunt compromise.

Sursa imaginii | Generat de DALL-E

Întrebare:

- Te-a făcut vreodată tehnologia să te simți la fel ca Sarah? Ce măsuri ați luat, dacă ați luat vreuna?
- Ce ați sfătui-o pe Sarah să facă în această situație?
- Ce măsuri poate lua pentru a-și controla mai bine datele?

3. Înțelegerea confidențialității în IA

Confidențialitatea poate fi înțeleasă în sens larg ca fiind dreptul persoanelor de a-și păstra informațiile personale departe de ochii publicului și de a-și controla datele și interacțiunile. Acest drept este recunoscut ca un drept fundamental al omului în multe sisteme juridice și tratate internaționale, subliniindu-se importanța autonomiei și demnității personale.

În domeniul IA, confidențialitatea capătă o complexitate suplimentară. Sistemele de inteligență artificială se bazează adesea pe date vaste pentru a învăța și a lua decizii. Aceste date pot include informații personale sensibile, care, dacă sunt manipulate greșit, pot duce la încălcări semnificative ale vieții private. Integrarea inteligenței artificiale în tehnologiile de zi cu zi - de la asistenți personali și dispozitive portabile la analize predictive în domeniul sănătății - înseamnă că protecția vieții private necesită măsuri tradiționale de protecție a datelor și noi abordări pentru a răspunde provocărilor unice pe care le ridică tehnologiile IA.

"Confidențialitatea este un drept fundamental al omului recunoscut în Declarația drepturilor omului a ONU, în Pactul internațional cu privire la drepturile civile și

politice și în multe alte tratate globale și regionale. În contextul IA, viața privată cuprinde mai multe dimensiuni-cheie."

Semnificația vieții private în contextul IA este multiplă:

- **Încredere**: Protecția eficientă a vieții private este esențială pentru menținerea încrederii utilizatorilor în tehnologiile IA. Fără încredere, utilizatorii pot fi reticenți în a adopta inovații potențial benefice.
- **Obligații etice**: Există un imperativ etic de a respecta și proteja viața privată a persoanelor. Acest lucru implică asigurarea faptului că sistemele de inteligență artificială nu încalcă drepturile personale și că sunt concepute ținând cont de confidențialitate încă de la început.
- **Conformitate**: Multe jurisdicții au reglementări stricte privind protecția datelor, cum ar fi GDPR în Europa, care impun obligații specifice organizațiilor care prelucrează date cu caracter personal prin intermediul sistemelor AI.

3.1. Confidențialitatea datelor: Protecția datelor cu caracter personal împotriva accesului neautorizat

Confidențialitatea datelor în IA se referă la protejarea datelor personale împotriva accesului, utilizării sau divulgării neautorizate. Sistemele de inteligență artificială trebuie să fie concepute astfel încât să se asigure că datele și informațiile sensibile sunt gestionate în siguranță și sunt accesate numai de părțile autorizate. Măsurile eficiente de confidențialitate a datelor contribuie la prevenirea furtului de identitate, a fraudei financiare și a riscurilor de exploatare a datelor cu caracter personal. Acestea includ:

- **Criptarea datelor**: Criptarea datelor atât în repaus, cât și în tranzit pentru a preveni accesul neautorizat.
- **Controlul accesului**: Limitarea accesului la date la persoanele care au nevoie de acestea pentru a-și îndeplini atribuțiile profesionale.
- **Anonimizarea datelor**: Eliminarea informațiilor de identificare personală din seturile de date utilizate în IA pentru a reduce riscul încălcării confidențialității.

De exemplu, o companie de retail online colectează date despre clienți pentru procesarea comenzilor și marketing personalizat. Compania utilizează criptarea pentru a securiza detaliile de plată ale clienților și informațiile personale pentru a asigura confidențialitatea datelor. În plus, aplică politici interne stricte care limitează accesul angajaților la aceste date sensibile doar la cei care au nevoie de ele pentru a procesa comenzile. De asemenea, asigură consimțământul clienților prin proceduri clare de opt-in înainte de a trimite comunicări de marketing.

3.2. Securitatea informațiilor: Măsuri pentru protejarea integrității și confidențialității datelor

Securitatea informațiilor este strâns legată de confidențialitatea datelor, dar se concentrează mai mult pe măsurile de protecție care garantează că datele rămân exacte, fiabile și accesibile numai utilizatorilor autorizați. În contextul IA:

Integritate: Protejarea datelor împotriva modificărilor neautorizate care ar putea compromite acuratețea acestora.

Confidențialitate: Asigurarea că informațiile cu caracter personal sunt păstrate secrete și sunt accesibile numai persoanelor care sunt autorizate să le acceseze.

Protocoale de securitate: Implementarea unor protocoale robuste, cum ar fi Secure Socket Layer (SSL), firewall-uri și sisteme de detectare a intruziunilor pentru a proteja împotriva atacurilor externe.

De exemplu, un furnizor de servicii medicale utilizează un sistem de evidență electronică a sănătății (EHR) pentru a stoca datele pacienților. Pentru a proteja aceste date, furnizorul utilizează mai multe niveluri de măsuri de securitate. Printre acestea se numără utilizarea unei criptări puternice pentru stocarea și transferul datelor, autentificarea cu mai mulți factori (MFA) pentru accesul la sistem și audituri de securitate periodice pentru identificarea și reducerea vulnerabilităților. Aceste practici contribuie la asigurarea integrității și confidențialității datelor pacienților împotriva amenințărilor cibernetice.

3.3. Autonomia personală: Dreptul de a controla informațiile despre propria persoană

Autonomia personală în era digitală, în special în cadrul IA, presupune menținerea controlului asupra informațiilor personale și asupra deciziilor luate pe baza acestor informații. Sistemele de inteligență artificială trebuie să fie concepute astfel încât să respecte și să consolideze autonomia personală prin:

- **Consimțământul informat**: Asigurarea că utilizatorii sunt pe deplin informați cu privire la modul în care datele lor vor fi utilizate și obținerea consimțământului acestora înainte de colectarea datelor.
- **Transparență**: Furnizarea de informații clare utilizatorilor cu privire la practicile de prelucrare a datelor și la logica din spatele deciziilor AI.
- **Mecanisme de control**: Furnizați utilizatorilor mecanisme pentru a controla modul în care sunt utilizate informațiile lor și pentru a renunța la prelucrarea datelor atunci când doresc.

De exemplu, o aplicație de urmărire a activității fizice colectează date privind activitatea fizică a utilizatorului pentru a oferi sfaturi personalizate în materie de fitness. Aplicația respectă autonomia personală prin informarea clară a utilizatorilor cu privire la tipurile de date pe care le colectează și la modul în care acestea vor fi utilizate. De asemenea, permite utilizatorilor să își acceseze datele, să corecteze eventualele inexactități și să renunțe la colectarea datelor pentru anumite funcții. Această abordare garantează că utilizatorii au control asupra informațiilor lor personale și asupra deciziilor luate.

4. Confidențialitatea datelor

Confidențialitatea datelor se referă la protejarea datelor personale împotriva accesului, utilizării sau divulgării neautorizate. Aceasta presupune asigurarea faptului că persoanele fizice au control asupra informațiilor lor și că acestea sunt utilizate în conformitate cu preferințele lor și cu reglementările legale.

4.1. Importanța confidențialității datelor în IA

Confidențialitatea datelor este extrem de importantă în domeniul inteligenței artificiale, unde sunt prelucrate cantități mari de date cu caracter personal. În lipsa unei protecții adecvate, informațiile sensibile ale persoanelor pot fi vulnerabile la încălcări și la utilizarea abuzivă, ceea ce poate duce la potențiale prejudicii, cum ar fi furtul de identitate, pierderi financiare sau discriminare. În plus, sistemele AI se bazează foarte mult pe date pentru a lua decizii în cunoștință de cauză, ceea ce înseamnă că calitatea și integritatea datelor au un impact direct asupra performanței și fiabilității acestor sisteme.

4.2. Rolul reglementărilor precum GDPR

Regulamente precum Regulamentul general privind protecția datelor (GDPR) sunt esențiale pentru aplicarea principiilor de confidențialitate a datelor în IA. GDPR stabilește cerințe stricte pentru colectarea, prelucrarea și stocarea datelor cu caracter personal, inclusiv obținerea consimțământului explicit din partea persoanelor, punerea în aplicare a unor măsuri de securitate adecvate și furnizarea de mecanisme pentru ca persoanele vizate să poată accesa și controla datele lor. Prin punerea în aplicare a acestor reglementări, GDPR urmărește să protejeze drepturile la confidențialitate ale persoanelor fizice și să responsabilizeze organizațiile în ceea ce privește gestionarea datelor cu caracter personal.

De exemplu, scenariul aplicației de sănătate: Luați în considerare o aplicație de sănătate care colectează și stochează datele medicale ale utilizatorilor, cum ar fi istoricul lor medical, diagnosticele și planurile de tratament. Această aplicație prezintă riscuri semnificative pentru confidențialitatea datelor utilizatorilor în lipsa unui consimțământ adecvat sau a unor măsuri de securitate. De exemplu, în cazul în care aplicația nu dispune de o criptare robustă sau de controale ale accesului, aceasta ar putea fi vulnerabilă la încălcări ale securității datelor, expunând informații medicale sensibile unor părți neautorizate. În plus, dacă aplicația partajează datele utilizatorilor cu terți fără consimțământ explicit sau anonimizare, ar putea încălca reglementările privind confidențialitatea, cum ar fi GDPR, ducând la consecințe juridice. Acest scenariu ilustrează importanța implementării unor măsuri stricte de confidențialitate a datelor în aplicațiile AI pentru a proteja informațiile sensibile ale persoanelor și a asigura conformitatea cu cerințele de reglementare.

4.3. Aspecte-cheie ale confidențialității datelor

Confidențialitatea datelor este esențială pentru practicile digitale moderne, în special odată cu proliferarea tehnologiilor și serviciilor bazate pe date. Iată câteva elemente-cheie ale confidențialității datelor care contribuie la protejarea informațiilor personale și la asigurarea gestionării etice a datelor:

- **Consimțământul:** Consimțământul presupune obținerea permisiunii persoanelor înainte de colectarea, utilizarea sau partajarea datelor lor. Consimțământul trebuie să fie liber, specific, informat și lipsit de ambiguitate. Aceasta înseamnă că persoanele trebuie să fie pe deplin conștiente de ceea ce consimt și să înțeleagă clar domeniul de aplicare și implicațiile activităților de prelucrare a datelor.
- **Limitarea scopului:** Acest principiu impune ca datele să fie colectate în scopuri specificate, explicite și legitime și să nu fie prelucrate ulterior într-un mod incompatibil. Limitarea scopului contribuie la asigurarea faptului că datele sunt utilizate numai în moduri la care persoana vizată se așteaptă și consimte, împiedicând utilizarea datelor în mod rău intenționat sau irresponsabil.
- **Minimizarea datelor:** Minimizarea datelor înseamnă că numai datele necesare pentru prelucrare sunt colectate și prelucrate. Acest principiu încurajează organizațiile să limiteze colectarea datelor la ceea ce este direct relevant și necesar pentru îndeplinirea unui scop specificat. Acest lucru reduce riscul de prejudicii în urma încălcării securității datelor, deoarece mai puține date pot fi compromise.
- **Responsabilitate:** Responsabilitatea se referă la obligația organizațiilor de a-și asuma responsabilitatea pentru datele pe care le prelucrează și de a demonstra conformitatea cu toate principiile de protecție a datelor. Aceasta include păstrarea înregistrărilor activităților de prelucrare a datelor, punerea în aplicare a măsurilor de protecție a datelor și demonstrarea

modului în care se realizează conformitatea. Acest principiu garantează că organizațiile sunt transparente cu privire la practicile lor de prelucrare a datelor și pot fi trase la răspundere pentru activitățile lor de prelucrare a datelor.

- **Transparența:** Transparența impune ca orice informație și comunicare referitoare la prelucrarea datelor cu caracter personal să fie ușor accesibilă și ușor de înțeles pentru persoane. Politicile de confidențialitate și notificările privind colectarea datelor trebuie să utilizeze un limbaj clar și simplu. Transparența creează încredere, sporește corectitudinea și permite utilizatorilor să ia decizii în cunoștință de cauză cu privire la datele lor.
- **Acces și corectare:** Persoanele fizice au dreptul de a-și accesa datele și de a corecta datele incorecte care le privesc. Oferirea posibilității persoanelor de a-și accesa și actualiza datele în funcție de necesități asigură acuratețea datelor și permite persoanelor să păstreze controlul asupra informațiilor lor, ceea ce este fundamental pentru autonomia personală.
- **Securitatea:** Securitatea implică punerea în aplicare a unor măsuri tehnice și organizaționale adecvate pentru a proteja datele cu caracter personal împotriva distrugerii accidentale sau ilegale, pierderii, modificării, divulgării neautorizate sau accesului la acestea. Aceasta include practici precum criptarea, stocarea securizată a datelor și controlul accesului fizic și digital.

Împreună, aceste aspecte ale confidențialității datelor formează un cadru solid pe care organizațiile îl pot utiliza pentru a proteja datele cu caracter personal și pentru a respecta drepturile de confidențialitate ale persoanelor. Ele contribuie la crearea unei culturi conștiente de confidențialitate, care se aliniază standardelor juridice și așteptărilor etice în era digitală.

5. Înțelegerea convenienței în IA

În contextul IA, comoditatea se referă la ușurința și eficiența cu care pot fi îndeplinite sarcinile sau pot fi îmbunătățite experiențele prin intermediul tehnologiilor IA. IA eficientizează procesele, reduce efortul manual și oferă experiențe personalizate, adaptate preferințelor individuale.

Comoditatea, facilitată de tehnologiile AI, cuprinde integrarea perfectă a automatizării, personalizării și eficienței în diverse domenii. De la asistență medicală și finanțe la viața de zi cu zi, soluțiile bazate pe inteligența artificială optimizează procesele, permit luarea deciziilor și îmbogățesc experiențele, sporind în cele din urmă productivitatea și îmbunătățind bunăstarea generală.

Convenienta, în domeniul IA, reprezintă integrarea fără probleme a tehnologiei pentru a simplifica sarcinile și a amplifica eficiența. Tehnologiile AI utilizează algoritmi și învățarea automată pentru a automatiza procesele, a reduce efortul manual și a adapta experiențele la preferințele individuale. Sistemele AI optimizează fluxurile de lucru prin analizarea unor cantități mari de date și învățarea din tipare, economisind timp și resurse și oferind în același timp

rezultate îmbunătățite. Mai jos sunt prezentate câteva puncte în care AI contribuie la confort în mai multe domenii:

- **Automatizarea sarcinilor de rutină:** Inteligența artificială excelează în automatizarea sarcinilor repetitive și banale care necesită în mod tradițional intervenția umană. De exemplu, software-ul bazat pe inteligență artificială poate gestiona automat programarea întâlnirilor, răspunsul la solicitările clienților sau gestionarea e-mailurilor. Acest lucru accelerează procesul și eliberează resurse umane pentru sarcini mai complexe și mai creative, crescând astfel productivitatea și eficiența.
- **Personalizarea:** Sistemele AI pot analiza cantități mari de date pentru a adapta serviciile și produsele la preferințele și nevoile individuale. În cazul consumatorilor, acest lucru se poate manifesta prin recomandarea de către AI a unor produse pe baza achizițiilor anterioare sau a obiceiurilor de vizionare, așa cum se întâmplă în comerțul cu amănuntul online și în serviciile de streaming. În aplicații mai personale, AI poate regla sistemele de încălzire a locuinței în funcție de programul și preferințele proprietarului, asigurând confortul și optimizând în același timp consumul de energie.
- **Îmbunătățirea procesului decizional:** Inteligența artificială îmbunătățește procesul decizional, oferind persoanelor și organizațiilor informații derivate din analiza unor seturi mari de date care ar fi prea complexe pentru a fi prelucrate manual de către oameni. În sectoare precum cel al asistenței medicale, algoritmii AI ajută la diagnosticarea bolilor prin analizarea datelor de imagistică medicală mai rapid și adesea mai precis decât practicienii umani. În domeniul financiar, sistemele de inteligență artificială pot analiza datele de pe piață pentru a oferi informații privind investițiile sau pentru a detecta tranzacțiile frauduloase cu o precizie și o viteză ridicate.
- **Eficiență și gestionarea resurselor:** IA îmbunătățește semnificativ eficiența generală și optimizează gestionarea resurselor. De exemplu, sistemele de logistică și de gestionare a lanțului de aprovizionare bazate pe inteligență artificială pot prevedea rapid cererea, optimiza rutele de livrare și gestiona stocurile, reducând risipa și costurile. În mod similar, rețelele inteligente bazate pe inteligență artificială pot gestiona mai eficient distribuția energiei electrice într-un oraș, adaptându-se în timp real la modificările cererii.
- **Accesibilitatea și ușurința în utilizare:** Tehnologiile AI fac adesea tehnologia mai accesibilă și mai ușoară pentru o gamă mai largă de persoane, inclusiv pentru cele cu handicap. Dispozitivele cu asistență vocală bazate pe inteligență artificială ajută persoanele cu deficiențe de vedere sau limitări fizice să interacționeze cu tehnologia și să își gestioneze mediul mai eficient. De asemenea, inteligența artificială poate elimina barierele lingvistice cu ajutorul serviciilor de traducere în timp real, făcând informațiile și comunicarea accesibile și vorbitorilor care nu sunt nativi.

5.1. Exemple reale de convenienţă

Asistența medicală: În domeniul asistenței medicale, tehnologiile AI oferă confort prin automatizarea proceselor de diagnosticare, analizarea imaginilor medicale și detectarea modelelor în datele pacienților pentru a ajuta profesioniștii din domeniul asistenței medicale să diagnosticheze bolile mai rapid și mai precis. Astfel, furnizorii de asistență medicală și pacienții economisesc timp, ceea ce conduce la o furnizare mai eficientă a asistenței medicale și la îmbunătățirea rezultatelor pentru pacienți.

Sursa imaginii | Generat de DALL-E

Comoditate în domeniul asistenței medicale prin AI

- **Procese de diagnosticare automatizate**: IA accelerează analiza datelor medicale, cum ar fi imagistica și testele de diagnosticare, permițând un diagnostic mai rapid.
- **Acuratețe sporită a diagnosticului**: Algoritmii AI detectează modele subtile în date pe care oamenii le-ar putea omite, sporind precizia și fiabilitatea diagnosticelor.
- **Analiza predictivă**: Inteligența artificială utilizează date din diverse surse, inclusiv tehnologia purtabilă, pentru a prezice eventualele probleme de sănătate înainte ca acestea să devină grave, facilitând astfel asistența medicală preventivă.
- **Gestionarea raționalizată a pacienților**: Inteligența artificială optimizează fluxurile de lucru ale spitalelor prin gestionarea programărilor, a admiterii pacienților și a direcționării către mediile de îngrijire adecvate, reducând timpii de așteptare și îmbunătățind experiența pacienților.
- **Planuri de tratament personalizate**: Algoritmii analizează datele individuale ale pacienților pentru a personaliza planurile de tratament care au cele mai mari șanse de a fi eficiente pe baza profilului unic de sănătate al pacientului.

Finanțe: Instrumentele financiare bazate pe inteligență artificială oferă confort prin oferirea de consultanță personalizată în materie de investiții pe baza obiectivelor financiare individuale, a toleranței la risc și a tendințelor pieței. Aceste instrumente utilizează algoritmi pentru a analiza cantități mari de date financiare și pentru a oferi recomandări privind strategiile de investiții, ajutând persoanele fizice să ia decizii în cunoștință de cauză și să își optimizeze portofoliile financiare.

Sursa imaginii | Generat de DALL-E

Comoditate în finanțe prin AI

- **Consiliere personalizată în materie de investiții**: Instrumentele AI analizează obiectivele financiare, toleranța la risc și situațiile economice personale ale unei persoane pentru a oferi recomandări de investiții personalizate.
- **Analiza avansată a datelor**: Aceste sisteme utilizează algoritmi pentru a analiza volume mari de date financiare, inclusiv tendințele pieței și performanțele istorice, pentru a identifica cele mai bune oportunități de investiții.
- **Gestionarea optimizată a portofoliului**: Inteligența artificială ajută la ajustarea dinamică a portofoliilor de investiții în funcție de modificările pieței în timp real și de obiectivele financiare individuale, sporind potențialele randamente și gestionând în același timp expunerea la risc.
- **Insight-uri eficiente ale pieței**: Instrumentele AI oferă o perspectivă rapidă asupra condițiilor pieței, ajutând investitorii să înțeleagă dinamica complexă a pieței și să ia decizii în timp util.
- **Evaluarea riscurilor**: Algoritmii evaluează riscul asociat diferitelor opțiuni de investiții, permițând utilizatorilor să facă alegeri în cunoștință de cauză în funcție de toleranța la risc.
- **Automatizarea tranzacțiilor**: IA poate automatiza tranzacțiile de tranzacționare și investiții, eficientizând procesul de execuție și reducând probabilitatea de eroare umană.
- **Securitate sporită**: Măsurile avansate de securitate, cum ar fi detectarea anomaliilor, sunt utilizate de AI pentru a identifica și atenua potențialele amenințări la adresa conturilor de investiții.

Viața de zi cu zi: Dispozitivele inteligente de acasă bazate pe inteligența artificială, cum ar fi asistenții vocali, termostatele inteligente și sistemele automatizate de securitate a locuințelor, oferă confort prin automatizarea sarcinilor casnice și îmbunătățirea confortului și securității. Aceste dispozitive învață preferințele utilizatorilor în timp și ajustează setările în consecință, oferind experiențe personalizate și fără întreruperi care simplifică rutinele zilnice și îmbunătățesc calitatea vieții.

Sursa imaginii | Generat de DALL-E

Comoditate în viața de zi cu zi prin AI

- **Automatizarea treburilor casnice**: Dispozitivele bazate pe inteligență artificială, cum ar fi aspiratoarele inteligente și aparatele de bucătărie automatizate, se ocupă de treburile de rutină, eliberând timpul utilizatorilor.
- **Setări personalizate**: Dispozitive precum termostatele inteligente și sistemele de iluminat învață și se adaptează la preferințele utilizatorului, ajustând automat setările pentru un confort optim.
- **Tehnologie asistată vocal**: Asistenții vocali precum Alexa și Google Assistant permit controlul hands-free asupra dispozitivelor de uz casnic, facilitând accesul la informații și gestionarea sarcinilor.
- **Securitate sporită a locuințelor**: Sistemele de securitate inteligente utilizează inteligența artificială pentru a monitoriza mediul de acasă, a recunoaște fețe familiare, a detecta activități neobișnuite și a alerta proprietarii de locuințe cu privire la potențiale amenințări.
- **Eficiența energetică**: Dispozitivele bazate pe inteligență artificială optimizează consumul de energie în locuință prin învățarea perioadelor de vârf de utilizare și ajustarea operațiunilor pentru a reduce risipa, reducând facturile la utilități.
- **Integrare perfectă**: Ecosistemele casnice inteligente conectează diverse dispozitive, permițându-le să lucreze împreună fără probleme, sporind confortul utilizatorului prin controale și interacțiuni integrate.
- **Monitorizare și control de la distanță**: Proprietarii de locuințe pot monitoriza și controla dispozitivele inteligente de acasă de la distanță prin intermediul telefoanelor inteligente, asigurând liniștea atunci când sunt plecați de acasă.

În timp ce inteligența artificială sporește semnificativ confortul, aceasta ridică, de asemenea, probleme importante, cum ar fi preocupările legate de

confidențialitate, necesitatea supravegherii umane și potențialul de eliminare a locurilor de muncă în anumite sectoare. Este esențial să ne asigurăm că sistemele de IA sunt utilizate în mod etic și responsabil pentru a maximiza beneficiile acestora, minimizând în același timp potențialele prejudicii. În concluzie, comoditatea inteligenței artificiale se referă la valorificarea puterii inteligenței artificiale pentru a face viața mai ușoară, mai plăcută și mai eficientă. Pe măsură ce tehnologiile IA evoluează și se integrează în diferite aspecte ale vieții, potențialul lor de a stimula confortul se va extinde, promițând efecte transformatoare în toate aspectele societății.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Prezentați scenarii ipotetice care îi provoacă pe participanți să decidă între a acorda prioritate vieții private sau confortului.

Exemplu

Scenariu ipotetic: Aplicație de urmărire a sănătății

Context: Vă gândiți la o nouă aplicație de urmărire a sănătății, HealthMate. Această aplicație oferă planuri personalizate de fitness și dietă prin analizarea activităților zilnice, a aportului alimentar și a parametrilor de sănătate.

Provocarea scenariului: HealthMate se poate sincroniza cu rețelele dvs. de socializare pentru a colecta mai multe date despre stilul dvs. de viață pentru o personalizare îmbunătățită și pentru a împărtăși informații cu partenerii care vă pot trimite reclame direcționate în funcție de obiectivele dvs. de sănătate.

Întrebări pentru reflecție:

- Ați opta pentru personalizarea îmbunătățită, știind că datele dvs. ar putea fi utilizate pentru publicitate direcționată?
- Cum puneți în balanță beneficiile unui plan de sănătate personalizat cu riscurile partajării datelor dumneavoastră?

Acest scenariu solicită participanților să evalueze echilibrul dintre comoditatea recomandărilor personalizate în materie de sănătate și riscurile de confidențialitate ale partajării datelor.

6. Interacțiunea dintre confidențialitate și comoditate

Interacțiunea dinamică dintre confidențialitate și confort în tehnologia modernă prezintă oportunități și provocări. Pe măsură ce tehnologia se integrează din ce în ce mai mult în viața noastră de zi cu zi, înțelegerea modului de a echilibra aceste

aspecte devine crucială. Mai jos sunt prezentate implicațiile compromisurilor, dilemele etice și consecințele din lumea reală ale echilibrului dintre viața privată și comoditate.

Compromisuri între confidențialitate și comoditate:

Echilibrul dintre confidențialitate și confort necesită o navigare atentă a compromisurilor inerente implicate:

- Restricții privind colectarea datelor: Implementarea protecției vieții private înseamnă adesea limitarea cantității și a tipului de date colectate și utilizate. Acest lucru poate limita funcționalitatea sistemelor AI care se bazează pe seturi mari de date pentru a îmbunătăți furnizarea de servicii și experiența utilizatorilor.
- Sacrificarea vieții private pentru comoditate: Pe de altă parte, sporirea confortului implică adesea colectarea și analizarea unor date personale extinse. Aceasta poate include urmărirea comportamentelor, preferințelor și locațiilor utilizatorilor, ceea ce ar putea încălca viața privată.

Dilemele etice care decurg din aceste compromisuri:

Echilibrul dintre confidențialitate și comoditate aduce în prim plan mai multe dileme etice:

- Compromisul privind viața privată: Cât de mult sunt dispuse persoanele să renunțe la viața privată pentru un confort sporit? Răspunsul variază adesea în funcție de percepțiile individuale și de contextul de utilizare a tehnologiei.
- Practici de colectare a datelor: Organizațiile sunt, de asemenea, responsabile din punct de vedere etic pentru colectarea, utilizarea și partajarea informațiilor personale. Considerentele etice trebuie să ghideze deciziile cu privire la datele care sunt colectate, modul în care sunt utilizate și cât de mult sunt informați utilizatorii cu privire la aceste practici.

Implicațiile din lumea reală ale alegerii convenienței în detrimentul confidențialității sau viceversa:

Deciziile luate cu privire la echilibrul dintre confidențialitate și comoditate au efecte tangibile:

- Consecințele prioritizării comodității: Alegerea comodității crește adesea riscurile, cum ar fi încălcarea securității datelor, furtul de identitate și încălcarea semnificativă a vieții private. Acest lucru poate duce la deteriorarea pe termen lung a încrederii utilizatorilor și la potențiale repercusiuni juridice pentru companii.
- Efectele prioritizării confidențialității: Dimpotrivă, acordarea unei mari priorități vieții private ar putea limita eficacitatea anumitor soluții tehnologice. Acest lucru ar putea duce la servicii mai puțin personalizate și ar putea încetini inovarea și progresul tehnologic care ar putea aduce beneficii societății.

Găsirea echilibrului între confidențialitate și confort necesită o abordare nuanțată care să ia în considerare implicațiile etice, sociale și personale ale tehnologiei. Organizațiile și persoanele trebuie să se străduiască să găsească o cale de mijloc care să respecte viața privată a utilizatorilor, profitând în același timp de avantajele pe care le poate aduce tehnologia. Acest echilibru nu este static și necesită o evaluare continuă pe măsură ce tehnologia evoluează și valorile societății se schimbă. Înțelegerea acestei dinamici ajută părțile interesate să ia decizii în cunoștință de cauză, care să protejeze drepturile individuale, acceptând în același timp progresele erei digitale.

7. Importanța vieții private și a confortului

Confidențialitatea și confortul inteligenței artificiale sunt esențiale în modelarea încrederii utilizatorilor și în adoptarea tehnologiei. Pe măsură ce inteligența artificială se integrează în viața noastră de zi cu zi, atingerea unui echilibru între protecția datelor personale și îmbunătățirea experienței utilizatorilor este esențială. Confidențialitatea asigură protejarea informațiilor personale, în timp ce confortul sporește eficiența și personalizarea în utilizarea tehnologiei. Provocarea constă în armonizarea acestor aspecte pentru a promova implementarea etică a IA și acceptarea de către societate, asigurându-se că progresele în domeniul IA aduc beneficii fără a compromite drepturile individuale.

Sporirea încrederii utilizatorilor și adoptarea tehnologiei

Prioritatea acordată confidențialității și confortului în tehnologiile AI este esențială pentru a consolida încrederea utilizatorilor și pentru a încuraja o adoptare mai largă. Atunci când sistemele AI gestionează în mod transparent confidențialitatea și oferă experiențe convenabile fără întreruperi, acestea sunt acceptate mai ușor de către utilizatori. Iată cum prioritizarea acestor aspecte poate influența încrederea utilizatorilor și adoptarea tehnologiei:

- **Consolidarea încrederii**: Utilizatorii tind să aibă încredere în tehnologia care le respectă viața privată și le îmbunătățește viața de zi cu zi fără intruziuni semnificative. Soluțiile AI care protejează în mod eficient datele utilizatorilor, oferind în același timp experiențe personalizate și eficiente, tind să se bucure de o reputație pozitivă în rândul utilizatorilor.
- **Încurajarea adoptării**: Atunci când utilizatorii consideră că o tehnologie de inteligență artificială le sporește productivitatea sau confortul fără a le compromite viața privată, este mai probabil să o adopte și să o recomande. Demonstrarea valorii AI prin comoditate, cuplată cu garanții solide de confidențialitate, motivează utilizatorii să integreze aceste tehnologii în rutina lor zilnică.

Considerații juridice și etice în implementarea IA

Implementarea tehnologiilor AI necesită o analiză atentă a implicațiilor juridice și etice pentru a asigura utilizarea responsabilă și respectarea reglementărilor:

- **Conformitatea cu reglementările**: Legi precum Regulamentul general privind protecția datelor (GDPR) oferă orientări stricte privind confidențialitatea datelor pe care trebuie să le respecte tehnologiile AI care operează sau vizează utilizatorii din Uniunea Europeană. Conformitatea cu aceste reglementări nu ține doar de necesitatea legală, ci și de demonstrarea unui angajament față de confidențialitatea utilizatorilor.

- **Cadre etice**: Dincolo de cerințele legale, cadrele etice orientează dezvoltarea IA pentru a garanta că tehnologiile sunt utilizate în beneficiul societății. Aceste cadre ajută dezvoltatorii și companiile să abordeze probleme complexe, cum ar fi prejudecățile, echitatea, transparența și responsabilitatea în sistemele de inteligență artificială.

Impactul asupra normelor societale și comportamentelor individuale

Rolul IA în echilibrarea vieții private și a confortului are efecte semnificative asupra normelor societale și a comportamentelor individuale:

- **Modelarea așteptărilor și preferințelor în materie de confidențialitate**: Pe măsură ce tehnologiile AI se integrează tot mai mult în viața de zi cu zi, acestea influențează modul în care indivizii își percep și își apreciază viața privată. Acest lucru poate duce la schimbarea așteptărilor, unii utilizatori devenind mai preocupați de confidențialitate, în timp ce alții pot deveni desensibilizați față de partajarea informațiilor personale.

- **Influențarea rutinei zilnice și a procesului decizional**: Tehnologiile AI care oferă confort pot modifica semnificativ viața de zi cu zi. Dispozitivele inteligente de acasă, experiențele de cumpărături personalizate și gestionarea automată a finanțelor personale pot remodela activitățile zilnice, făcând viața mai ușoară și creând dependență de tehnologie pentru luarea deciziilor.

Prin înțelegerea și abordarea interacțiunii complexe dintre confidențialitate și comoditate, dezvoltatorii și implementatorii AI pot spori acceptarea tehnologiei și se pot asigura că aceasta contribuie pozitiv la dezvoltarea societății și la bunăstarea individuală. Aceste considerente sunt esențiale pentru dezvoltarea durabilă a tehnologiilor IA într-o societate din ce în ce mai conștientă de respectarea vieții private și a standardelor etice.

8. Navigarea între riscuri și beneficii

Riscurile potențiale asociate cu tehnologiile AI în ceea ce privește confidențialitatea și confortul includ:

- **Încălcarea securității datelor:** Sistemele de inteligență artificială gestionează adesea cantități mari de date sensibile, ceea ce le face ținte pentru atacuri rău intenționate. Brețele de date pot duce la accesul neautorizat la informații personale, ceea ce duce la furtul de identitate, pierderi financiare și daune reputaționale.

- **Pierderea vieții private:** Tehnologiile AI pot colecta și analiza date cu caracter personal fără consimțământul sau cunoștința persoanelor, compromițând dreptul la viață privată. Acest lucru poate eroda încrederea

dintre utilizatori și sistemele de inteligență artificială, ducând la probleme legate de utilizarea abuzivă și exploatarea datelor.

- **Prejudecăți în algoritmii AI:** Algoritmii AI pot perpetua prejudecățile în datele de instruire, conducând la rezultate discriminatorii. Algoritmii AI părtinitori pot duce la tratamente sau decizii incorecte, exacerbând inegalitățile sociale și subminând încrederea în sistemele AI.

Consecințele acestor riscuri:

Aceste riscuri pot submina încrederea în sistemele AI și pot avea consecințe negative pentru indivizi și societate:

- **Erodarea încrederii:** Brețele de date și încălcarea confidențialității pot eroda încrederea dintre utilizatori și sistemele de inteligență artificială, ducând la scăderea adoptării și utilizării acestora. Persoanele fizice pot deveni reticente în a-și împărtăși datele sau în a se angaja în tehnologiile AI, împiedicând beneficiile potențiale ale acestora.
- **Impactul societal negativ:** Algoritmii AI părtinitori pot perpetua discriminarea și inegalitatea, ducând la tratament inechitabil și tulburări sociale. Acest lucru poate afecta coeziunea socială și încrederea în instituții, exacerbând diviziunile sociale.
- **Repercusiuni juridice și de reglementare:** Brețele de date și încălcările confidențialității pot avea repercusiuni juridice și de reglementare pentru organizațiile responsabile de sistemele AI. Amenzile, procesele și daunele aduse reputației pot avea implicații financiare și operaționale semnificative.

Beneficii potențiale:

În ciuda riscurilor, tehnologiile IA oferă numeroase beneficii potențiale:

- **Eficiență crescută:** IA raționalizează procesele și automatizează sarcinile, sporind eficiența și productivitatea. Prin manipularea sarcinilor repetitive, AI eliberează timp angajaților pentru a se concentra asupra activităților cu valoare mai mare.
- **Decizii mai bune bazate pe date:** Algoritmii AI analizează cantități mari de date pentru a descoperi perspective și modele, permițând organizațiilor să ia decizii mai bine fundamentate. Perspectivele bazate pe AI pot optimiza operațiunile, pot îmbunătăți experiența clienților și pot stimula inovarea.
- **Îmbunătățirea experienței utilizatorilor:** Personalizarea AI îmbunătățește experiențele utilizatorilor prin furnizarea de recomandări și servicii personalizate. De la recomandări personalizate de produse la mentenanță predictivă în producție, IA sporește satisfacția și implicarea utilizatorilor.

Mitigarea riscurilor:

Organizațiile ar trebui să pună în aplicare măsuri de securitate solide pentru a atenua aceste riscuri, să asigure transparența și responsabilitatea în sistemele de inteligență artificială și să acorde prioritate considerentelor etice pe parcursul

ciclului de viață al dezvoltării inteligenței artificiale. Prin abordarea proactivă a acestor riscuri, organizațiile pot consolida încrederea în sistemele de inteligență artificială și pot reduce la minimum potențialele prejudicii pentru persoane și societate.

Pentru a gestiona eficient riscurile și beneficiile tehnologiilor IA, organizațiile pot adopta următoarele strategii:

- **Implementați măsuri robuste de securitate:** Organizațiile trebuie să implementeze măsuri de securitate solide pentru a se proteja împotriva încălcării securității datelor și a accesului neautorizat. Acestea includ criptarea, controlul accesului și audituri de securitate periodice.
- **Asigurați transparența și responsabilitatea:** Transparența și responsabilitatea sunt esențiale pentru construirea încrederii în sistemele AI. Organizațiile ar trebui să fie transparente cu privire la modul în care sunt utilizate tehnologiile IA și să asigure responsabilitatea pentru acțiunile lor.
- **Prioritizarea considerentelor etice:** Considerațiile etice ar trebui să ghideze dezvoltarea și implementarea AI. Organizațiile ar trebui să acorde prioritate corectitudinii, transparenței și responsabilității în sistemele AI pentru a atenua prejudecățile și a asigura utilizarea etică a datelor.

Prin adoptarea unor practici proactive de gestionare a riscurilor și de implementare responsabilă a IA, organizațiile pot maximiza beneficiile tehnologiilor IA, minimizând în același timp potențialele prejudicii pentru persoane și societate.

9. Studiu de caz: IA în supraveghere

Scenariu

Autoritățile locale implementează sisteme de supraveghere bazate pe inteligență artificială într-un oraș aglomerat pentru a spori siguranța și securitatea publică. Aceste sisteme utilizează tehnologii avansate, cum ar fi recunoașterea facială, analiza comportamentului și detectarea obiectelor, pentru a monitoriza spațiile publice, a identifica potențialele amenințări și a răspunde în timp real urgențelor.

Sursa imaginii | Generat de DALL-E

Îmbunătățiri ale securității vs. încălcări ale vieții private:

Îmbunătățiri ale securității	Încălcări ale confidențialității
Sistemele de supraveghere AI consolidează măsurile de securitate, permițând agenților de aplicare a legii să detecteze și să reacționeze rapid la activități suspecte sau potențiale amenințări la adresa securității. Tehnologia de recunoaștere facială ajută la identificarea persoanelor de pe listele de supraveghere sau a celor date dispărute, facilitând	În ciuda beneficiilor în materie de securitate, apar preocupări cu privire la potențialele încălcări ale vieții private asociate cu supravegherea AI. Monitorizarea continuă și colectarea de date în spațiile publice ridică semne de întrebare cu privire la dreptul indivizilor la viață privată și la libertatea de a fi supravegheați constant. Există riscul identificării eronate și al rezultatelor fals pozitive, ceea ce duce la acuzații eronate și la încălcarea vieții private a persoanelor nevinovate semnalate eronat

Importanța implementării inteligente a IA:

Cazul inteligenței artificiale în supraveghere evidențiază importanța unei implementări bine gândite a inteligenței artificiale pentru a echilibra nevoile de securitate cu considerentele privind viața privată:

- **Politici și reglementări clare:** Implementarea inteligentă a IA implică stabilirea unor politici și reglementări clare care să reglementeze tehnologiile de supraveghere, pentru a asigura transparența, responsabilitatea și respectarea legislației privind protecția vieții private.
- **Utilizarea etică a datelor:** Organizațiile trebuie să acorde prioritate utilizării etice a datelor colectate prin intermediul sistemelor de supraveghere, să respecte dreptul persoanelor la viață privată și să reducă la minimum riscul de încălcare a vieții private sau de utilizare abuzivă a informațiilor personale.
- **Transparență algoritmică și responsabilitate:** Implementarea unor algoritmi de inteligență artificială transparenți și a unor mecanisme de responsabilizare garantează că sistemele de supraveghere sunt corecte,

exacte și responsabile pentru acțiunile lor, reducând riscul de prejudecăți sau erori care ar putea duce la încălcări ale vieții private.

- **Implicarea și supravegherea publicului:** Implicarea publicului în discuțiile privind supravegherea AI și implicarea părților interesate în procesul decizional promovează transparența, încrederea și responsabilitatea, încurajând implementarea și utilizarea responsabilă a tehnologiilor de supraveghere.

În concluzie, cazul inteligenței artificiale în supraveghere subliniază necesitatea unei implementări bine gândite a inteligenței artificiale, care să echilibreze sporirea securității cu considerentele legate de viața privată. Organizațiile pot implementa sisteme de supraveghere care sporesc securitatea, acordând în același timp prioritate principiilor etice, transparenței și responsabilității, respectând dreptul la viață privată al persoanelor și promovând încrederea publică.

10. Concluzii

Explorarea confidențialității și a confortului în cadrul inteligenței artificiale (IA) evidențiază un echilibru critic necesar pentru promovarea încrederii și adoptarea mai largă a tehnologiei. Pe măsură ce inteligența artificială pătrunde în diverse aspecte ale asistenței medicale, ale finanțelor și ale vieții de zi cu zi, aceasta aduce beneficii extraordinare, cum ar fi o eficiență sporită și servicii personalizate, însoțite de necesitatea unor protecții solide ale vieții private și de considerații etice. Respectarea cadrelor juridice precum GDPR și abordarea dilemelor etice sunt esențiale pentru a asigura implementarea responsabilă a IA.

Pe măsură ce IA modelează normele societale și comportamentele individuale, menținerea unui echilibru între confortul utilizatorilor și confidențialitatea datelor devine crucială. Dezvoltatorii, utilizatorii și factorii de decizie politică trebuie să colaboreze pentru a se asigura că tehnologiile IA îmbunătățesc viața fără a compromite drepturile fundamentale. Accentuarea transparenței și a responsabilității va fi esențială pentru valorificarea în mod responsabil a potențialului inteligenței artificiale, asigurându-se că aceasta rămâne o tehnologie benefică și de încredere în lumea noastră tot mai digitală.

CU3 | Algoritmi și limitările acestora

Index

1. Introducere[1]

Acest manual este conceput pentru a fi un instrument cheie pentru formatorii care susțin cursul de formare Charlie Project. Acesta nu va descrie doar conținutul de bază al cursului, ci va oferi, de asemenea, instrucțiuni detaliate, sfaturi și recomandări pentru formatori cu privire la modul de desfășurare eficientă a cursului. În plus, manualul va propune o varietate de activități menite să implice cursanții și să le îmbunătățească experiența de învățare. Acesta include, de asemenea, orientări privind modul de evaluare a rezultatelor învățării cursanților, asigurându-se că obiectivele cursului sunt îndeplinite și că atât formatorii, cât și cursanții au o înțelegere clară a competențelor preconizate a fi dezvoltate pe parcursul formării. Această abordare cuprinzătoare îi va ajuta pe formatori să faciliteze un mediu de învățare dinamic și eficient .[2]

Această unitate va prezenta conceptele fundamentale ale algoritmilor, modelele și limitele acestora. Participanții vor dobândi o înțelegere a rolului pe care algoritmii îl joacă în diverse sectoare și a potențialelor capcane care decurg din utilizarea lor. Subiectele abordate vor include complexitatea algoritmică, optimizarea și limitările procesului decizional algoritmic.

În era digitală de astăzi, algoritmii joacă un rol din ce în ce mai important în modelarea diferitelor aspecte ale vieții noastre, de la conținutul pe care îl consumăm online la deciziile luate de sistemele automate. Înțelegerea complexității algoritmice, a strategiilor de optimizare și a limitărilor inerente proceselor decizionale algoritmice este esențială pentru oricine navighează în acest peisaj.

Înțelegerea **complexității algoritmice** presupune examinarea eficienței și performanței algoritmilor atunci când aceștia abordează diverse sarcini de calcul. Aceasta implică înțelegerea resurselor - cum ar fi timpul și memoria - de care algoritmii au nevoie pentru a rezolva problemele și a modului în care aceste cerințe cresc odată cu mărimile de intrare. Această înțelegere ne permite să

[1] Notă: În pregătirea acestui document, autorii au utilizat modele lingvistice de mari dimensiuni (LLM) pentru a ajuta la corectarea textului, structurarea conținutului și identificarea exemplelor pertinente incluse în material. După utilizarea acestui instrument de inteligență artificială, autorii au revizuit meticulos și au rafinat conținutul în funcție de necesități. Autorii își asumă întreaga responsabilitate pentru integritatea și acuratețea conținutului în publicația finală.

[2] Pentru profesor/formator: Acest manual este menit să vă ghideze prin materialul pentru unitatea de curs. PPT-urile de sprijin conțin informații similare cu cele acoperite de E-learning, dar într-o formă mai detaliată. PPT-urile sunt extinse și nu sunt menite să fie acoperite în întregime în timpul sesiunilor interactive. În schimb, puteți întreba la începutul sesiunilor interactive ce subiecte au fost cel mai greu de înțeles pentru elevi și puteți utiliza doar acele slide-uri de sprijin în timpul sesiunii. Puteți întreba acest lucru utilizând instrumente precum Mentimeter sau Kahoot, ca exemplu, adăugând subiectele de curs din conținutul unității de competență descrise în puncte la sfârșitul secțiunii de introducere. În plus, studiul de caz bazat pe CU este menit să fie abordat și în timpul sesiunii interactive. Orice alte sugestii din manual sunt doar orientative; nu ezitați să le folosiți după cum credeți de cuviință.

evaluăm viabilitatea și aplicabilitatea utilizării anumitor algoritmi în contexte reale (Cormen et al., 2009).

Metodologiile de optimizare constituie un alt aspect fundamental al înțelegerii algoritmilor. Scopul acestor metode este de a rafina eficiența și eficacitatea algoritmilor prin ajustarea parametrilor acestora sau prin restructurarea logicii lor fundamentale. Fie că este vorba despre reducerea timpului de calcul, maximizarea utilizării resurselor sau optimizarea pentru anumite obiective, utilizarea unor strategii de optimizare adecvate poate îmbunătăți semnificativ performanța algoritmilor în diverse domenii (Cormen et al., 2009).

Cu toate acestea, pe lângă beneficiile potențiale ale procesului decizional algoritmic, există limitări inerente care necesită o analiză atentă. Algoritmii funcționează în cadre predefinite și depind de calitatea datelor de intrare și de acuratețea ipotezelor de bază. Abaterile din date, ipotezele de modelare eronate sau cazurile limită neașteptate pot conduce la rezultate eronate sau consecințe neintenționate, subliniind fragilitatea sistemelor decizionale algoritmice.

Aprofundarea înțelegerii complexității algoritmice, a strategiilor de optimizare și a limitărilor inerente proceselor decizionale algoritmice poate favoriza o înțelegere mai solidă a nuanțelor și provocărilor algoritmice. Aceasta oferă persoanelor abilitățile de gândire critică necesare pentru a evalua în mod critic soluțiile algoritmice, pentru a identifica potențialele capcane și pentru a pleda pentru practici algoritmice responsabile și etice într-o lume din ce în ce mai condusă de algoritmi.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sfaturi și recomandări pentru profesori pentru a se implica în conceptele algoritmului inițial

1. Clarificarea conceptelor cheie
Înainte de a vă adânci în discuții, asigurați-vă că toți participanții au o înțelegere clară a terminologiei de bază, cum ar fi "complexitatea algoritmică", "optimizarea" și "limitările algoritmilor". Acest lucru poate fi realizat printr-o scurtă prezentare sau prin distribuirea unui glosar la începutul sesiunii.

2. Raportarea la aplicații din lumea reală
Conectați conceptele abstracte la aplicații din lumea reală din diverse sectoare, cum ar fi finanțele, asistența medicală sau comerțul electronic. Acest lucru ajută participanții să vadă relevanța și impactul algoritmilor în scenariile de zi cu zi, făcând conținutul mai atractiv și mai ușor de înțeles.

3. Utilizați ajutoare vizuale
Algoritmii pot fi complecși și abstracți. Utilizați diagrame, diagrame de flux și animații pentru a vizualiza concepte precum fluxul algoritmic, complexitatea și

procesele de optimizare. Acest lucru poate ajuta participanții să înțeleagă și să rețină mai ușor informațiile.

4. Încurajați întrebările
Creați un mediu deschis în care participanții se simt confortabil să pună întrebări. Acest lucru poate fi facilitat prin sesiuni regulate de întrebări și răspunsuri după acoperirea fiecărui subiect major, asigurându-se că participanții au înțeles pe deplin materialul înainte de a trece mai departe.

5. Încorporarea studiilor de caz
Discutați cazuri reale în care procesul decizional algoritmic a avut succes sau s-a confruntat cu provocări. Analiza acestor cazuri poate declanșa discuții despre implicațiile etice, eficiența și considerentele practice ale utilizării algoritmilor.

6. Promovarea gândirii critice
Încurajați participanții să se gândească în mod critic la limitările și potențialele prejudecăți ale modelelor algoritmice. Acest lucru ar putea fi realizat prin discuții de grup sau dezbateri privind modul de atenuare a acestor probleme în proiectarea și implementarea algoritmilor.

Activități de încălzire recomandate

1. Algoritm joc de rol
O activitate de joc de rol în care participanții simulează un algoritm în acțiune, fie prin sortarea manuală a elementelor, fie prin urmarea unui set simplu de instrucțiuni pentru atingerea unui obiectiv, fie prin interpretarea pașilor pe care un algoritm îi poate urma în procesele decizionale. Acest exercițiu ajută la ilustrarea conceptului de procesare pas cu pas în algoritmi.

2. Complexitatea algoritmică Icebreaker
Începeți cu o sarcină simplă, cum ar fi organizarea cărților după culoare, și treceți la organizarea după gen, apoi după autor în cadrul fiecărui gen. Discutați despre modul în care complexitatea crește cu fiecare nivel adăugat, făcând o paralelă cu complexitatea algoritmică și cu modul în care resursele sunt utilizate mai mult pe măsură ce sarcinile devin mai complexe.

3. Provocarea optimizării
Dați grupurilor mici o sarcină prestabilită, cum ar fi aranjarea scaunelor într-o cameră astfel încât să încapă cât mai multe persoane, dar cu cât mai puține rânduri. Lăsați-i să încerce, să discute strategiile lor și apoi să dezvăluie modul în care strategiile de optimizare ar putea îmbunătăți soluțiile lor.

4. Dezbatere etică
Organizați o dezbatere cu privire la implicațiile etice ale algoritmilor în procesul decizional, concentrându-vă asupra potențialelor prejudecăți și a consecințelor dependenței masive de sistemele automate. Acest lucru poate sensibiliza participanții la complexitatea și responsabilitățile lucrului cu algoritmii.

5. Jocul Break-the-Algorithm
Furnizați participanților un algoritm simplu sau un set de reguli și provocați-i să găsească cazurile limită sau scenariile în care algoritmul eșuează. Această activitate poate evidenția limitările și potențialele defecte ale logicii algoritmice.

Aceste activități și strategii nu numai că vor face ca procesul de învățare să fie captivant, dar vor oferi participanților o înțelegere și o apreciere mai profundă a complexităților implicate în luarea deciziilor algoritmice.

2. Să înțeleagă conceptele fundamentale ale algoritmilor, modelele și limitările acestora

2.1. Complexitatea algoritmică

Algoritmii sunt astăzi omniprezenți. Acestea ghidează computerele pentru a procesa informații, a rezolva probleme și a lua decizii importante. Ele pot rezolva probleme de la organizarea datelor până la găsirea celor mai bune rute de călătorie sau sugerarea de filme. Într-o epocă dominată de big data și social media, cunoașterea modului în care funcționează algoritmii este esențială pentru a le înțelege limitele și posibilele prejudecăți. Aceștia nu îmbunătățesc doar modul în care computerele îndeplinesc sarcinile, ci ne influențează și experiențele de zi cu zi, ceea ce îi face esențiali pentru informatica modernă.

Un algoritm este o secvență de instrucțiuni lipsite de ambiguitate pentru rezolvarea unei probleme, trebuie să fie corect, să ofere întotdeauna o soluție corectă, și trebuie să fie finit, să se încheie. Este o procedură pas cu pas care ne spune ce să facem pentru a atinge un anumit scop. Ne putem gândi la aceasta ca la gătit: urmăm o rețetă care ne spune exact ce ingrediente să folosim, cât de mult din fiecare și ce pași să facem. În mod similar, un algoritm ghidează un computer sau o persoană printr-o serie de acțiuni pentru a obține un rezultat specific, fie că este vorba de sortarea unei liste de numere, găsirea celui mai scurt traseu pe o hartă sau recomandarea de filme în funcție de preferințele dumneavoastră.

Conceptul de algoritm există de mult timp, datând din civilizațiile antice. Unul dintre cele mai vechi exemple este algoritmul euclidian, atribuit matematicianului Euclid din Grecia antică, care a fost dezvoltat în jurul anului 300 î.Hr. pentru a găsi cel mai mare divizor comun a două numere.

De-a lungul istoriei, diverse culturi și civilizații și-au dezvoltat proprii algoritmi pentru a rezolva probleme matematice, a naviga pe teren geografic sau a îndeplini sarcini în mod eficient. Acești algoritmi au fost adesea transmise oral sau prin texte scrise.

Termenul "algoritm" în sine provine de la numele matematicianului persan Muhammad ibn Musa al-Khwarizmi, care a trăit în timpul Epocii de Aur Islamice din secolul al IX-lea. Al-Khwarizmi a scris o carte intitulată "Al-Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabala" (Cartea cuprinzătoare privind calculul prin completare și echilibrare), care introduce metode sistematice de rezolvare a ecuațiilor matematice. Cuvântul "algoritm" este derivat din versiunea latinizată a numelui său, "Algoritmi".

De atunci, algoritmii au evoluat și au devenit fundamentali în diverse domenii, în special odată cu apariția informaticii moderne. În prezent, algoritmii sunt utilizați

nu numai în matematică, ci și în informatică, inginerie, biologie, economie și multe alte discipline pentru a rezolva probleme complexe și a automatiza sarcini.

2.2. Caracterizarea algoritmului

Caracterizările algoritmilor sunt încercări de a formaliza cuvântul algoritm. Termenul algoritm nu are o definiție formală general acceptată. Knuth (1997) a definit o listă de cinci proprietăți care sunt larg acceptate ca cerințe pentru un algoritm:

- **Finitudinea:** "Un algoritm trebuie să se încheie întotdeauna după un număr finit de pași ... un număr foarte finit, un număr rezonabil". Finitudinea asigură că algoritmii se încheie după un număr finit de pași, evitând buclele infinite care ar putea consuma resurse de calcul la nesfârșit. Este ca și cum ai completa un puzzle în care, în cele din urmă, ultima piesă cade la locul ei, semnalând încheierea algoritmului. Această caracteristică a finitudinii oferă previzibilitate și control asupra proceselor de calcul.
- **Definitivitate:** "Fiecare etapă a unui algoritm trebuie definită cu precizie; acțiunile care trebuie efectuate trebuie specificate riguros și fără ambiguitate pentru fiecare caz". Se subliniază necesitatea unor etape lipsite de ambiguitate, care să nu lase loc de interpretări. Un algoritm definit prezintă în mod clar fiecare acțiune care trebuie efectuată în fiecare etapă a rezolvării problemei, asigurându-se că nu există nicio ambiguitate sau confuzie cu privire la ceea ce trebuie făcut. Această caracteristică este esențială, deoarece permite oricui, indiferent de pregătirea sau expertiza sa, să înțeleagă și să pună în aplicare algoritmul în mod corect.
- **Introducere:** "...cantități care îi sunt date inițial înainte de începerea algoritmului. Aceste intrări sunt preluate din seturi specificate de obiecte". Intrarea se referă la datele sau informațiile pe care algoritmul operează pentru a produce un rezultat. Această intrare poate fi orice formă de date, cum ar fi numere, text, imagini sau chiar alți algoritmi. Datele de intrare sunt esențiale deoarece furnizează materia primă pe care algoritmul o procesează și o manipulează pentru a rezolva o problemă sau a îndeplini o sarcină.
- **Output:** "...cantități care au o relație specifică cu intrările". Ieșirea se referă la rezultatul sau soluția produsă de algoritm după prelucrarea datelor de intrare. Acest rezultat poate lua, de asemenea, diverse forme, în funcție de natura problemei rezolvate, cum ar fi o singură valoare, un set de valori, o decizie, o reprezentare vizuală sau orice altă formă de informație care reprezintă rezultatul dorit.
- **Eficacitate:** "... toate operațiile care trebuie efectuate în algoritm trebuie să fie suficient de elementare pentru a putea fi, în principiu, efectuate exact și într-un timp finit de către un om care utilizează hârtie și creion". Eficacitatea accentuează caracterul practic și fezabilitatea algoritmilor. La fel cum o rețetă cu ingrediente obscure sau imposibil de obținut devine

ineficientă, algoritmii trebuie să utilizeze resurse accesibile pentru a-și îndeplini sarcinile. Prin asigurarea faptului că algoritmii sunt practici și fezabili, eficacitatea garantează că soluțiile sunt realizabile în limitele resurselor și timpului disponibile.

După cum putem observa, descrierea de mai sus a caracteristicilor unui algoritm poate fi clară din punct de vedere intuitiv, dar este lipsită de rigoare formală, deoarece nu este exact clar ce înseamnă "definit cu precizie", sau "specificat riguros și fără ambiguitate", sau "suficient de fundamental" și așa mai departe. Noi o considerăm suficientă pentru scopurile noastre.

2.3. Structura algoritmului

Unul dintre principalele concepte din spatele algoritmilor este abstractizarea. În acest context, abstractizarea se referă la ideea de a ascunde detaliile complexe și de a se concentra doar pe aspectele esențiale necesare pentru înțelegerea și rezolvarea unei probleme. Este ca și cum ai mări imaginea pentru a vedea imaginea de ansamblu fără a te pierde în detaliile fine. În proiectarea algoritmilor, abstractizarea ajută la simplificarea procesului de rezolvare a problemelor prin divizarea acestuia în părți mai mici, ușor de gestionat. Acest lucru ne permite să abordăm fiecare parte separat, fără a fi nevoie să înțelegem toate detaliile complexe deodată.

Algoritmii implică de obicei mai multe instrucțiuni care lucrează împreună, mai degrabă decât un singur pas. Atunci când creăm programe complexe, executăm o serie de astfel de instrucțiuni în succesiune, le repetăm sau facem alegeri pe baza unor condiții specifice.

Cunoașterea modului în care funcționează algoritmii implică înțelegerea diferitelor moduri în care instrucțiunile pot fi combinate. Aceste scheme oferă o abordare structurată pentru organizarea și secvențierea instrucțiunilor într-un algoritm. Ele ajută la împărțirea problemelor complexe în sarcini ușor de gestionat, făcând algoritmul mai ușor de înțeles și mai ușor de implementat. Algoritmii sunt de obicei compuși din structuri simple care sunt organizate într-o manieră ierarhică. Aceste structuri controlează fluxul de instrucțiuni din cadrul algoritmului, dictând secvența în care acestea sunt executate. Aceasta este adesea denumită structura de control a algoritmului.

O structură de control direcționează în esență ordinea de execuție a instrucțiunilor dintr-un algoritm. Aceasta determină calea pe care o urmează procesul de execuție, pe baza condițiilor definite în cadrul structurii. Structurile de control sunt un aspect fundamental al oricărui algoritm, deoarece asigură că instrucțiunile sunt executate în ordinea corectă și că se obține rezultatul dorit.

O caracteristică comună a tuturor structurilor de control este că acestea au un singur punct de intrare și un singur punct de ieșire. Punctul unic de intrare

marchează începutul structurii de control, unde procesul de execuție intră în structură. Pe de altă parte, punctul unic de ieșire marchează sfârșitul structurii de control, unde procesul de execuție părăsește structura și trece la următoarea parte a algoritmului. Această caracteristică asigură că procesul de execuție urmează o cale definită în cadrul structurii de control, menținând integritatea și corectitudinea algoritmului.

Există trei scheme de compoziție a instrucțiunilor:

- Secvențial: În această schemă, instrucțiunile sunt executate una după alta, urmând o ordine liniară. Este foarte aproape de a urma o rețetă pas cu pas. Algoritmii secvențiali sunt predominanți în sarcinile de zi cu zi, cum ar fi prepararea unui sandwich sau calcularea sumei numerelor. Ei asigură că acțiunile au loc într-o secvență specifică, fără nicio ramificare sau luare de decizii.

Exemplu: Un algoritm pentru a face un tort.
1. Preîncălziți cuptorul.
2. Se unge și se tapetează cu făină o formă de tort.
3. Într-un bol de amestecare, se ung împreună untul și zahărul.
4. Se bat ouăle, unul câte unul.
5. Se amestecă cu extractul de vanilie.
6. Într-un bol separat, se combină făina, praful de copt și sarea.
7. Se adaugă treptat ingredientele uscate la amestecul umed, alternând cu laptele.
8. Se toarnă aluatul în forma de tort pregătită.
9. Se introduce forma de tort în cuptorul preîncălzit.
10. Coaceți timp de 25-30 de minute sau până când o scobitoare introdusă în centru iese curată
11. Scoateți prăjitura din cuptor și lăsați-o să se răcească în tavă timp de 10 minute.
15. Serviți tortul.

- Condiționate sau alternative: Aceste instrucțiuni introduc puncte de decizie. În funcție de anumite condiții, programul urmează căi diferite. Gândiți-vă la aceasta ca la o alegere între mai multe opțiuni. Instrucțiunile condiționale (precum "if", "else" și "switch") ne permit să gestionăm diverse scenarii.

Exemplu: Algoritm care simulează procesul de a decide dacă să aduci o umbrelă sau o jachetă atunci când ieși din casă pe baza prognozei meteo:

1. Verificați prognoza meteo pentru ploaie.
2. Dacă prognoza anunță ploaie, atunci:
3. Luați o umbrelă.
4. Else Dacă prognoza nu anunță ploaie și temperatura este sub 15 grade Celsius, atunci:
5. Luați o jachetă.
6. Plecați din casă.

Rețineți că indentarea instrucțiunilor 3 și 5 are un rol specific în definirea algoritmului. Aceste instrucțiuni sunt executate numai dacă instrucțiunea anterioară este evaluată pozitiv.

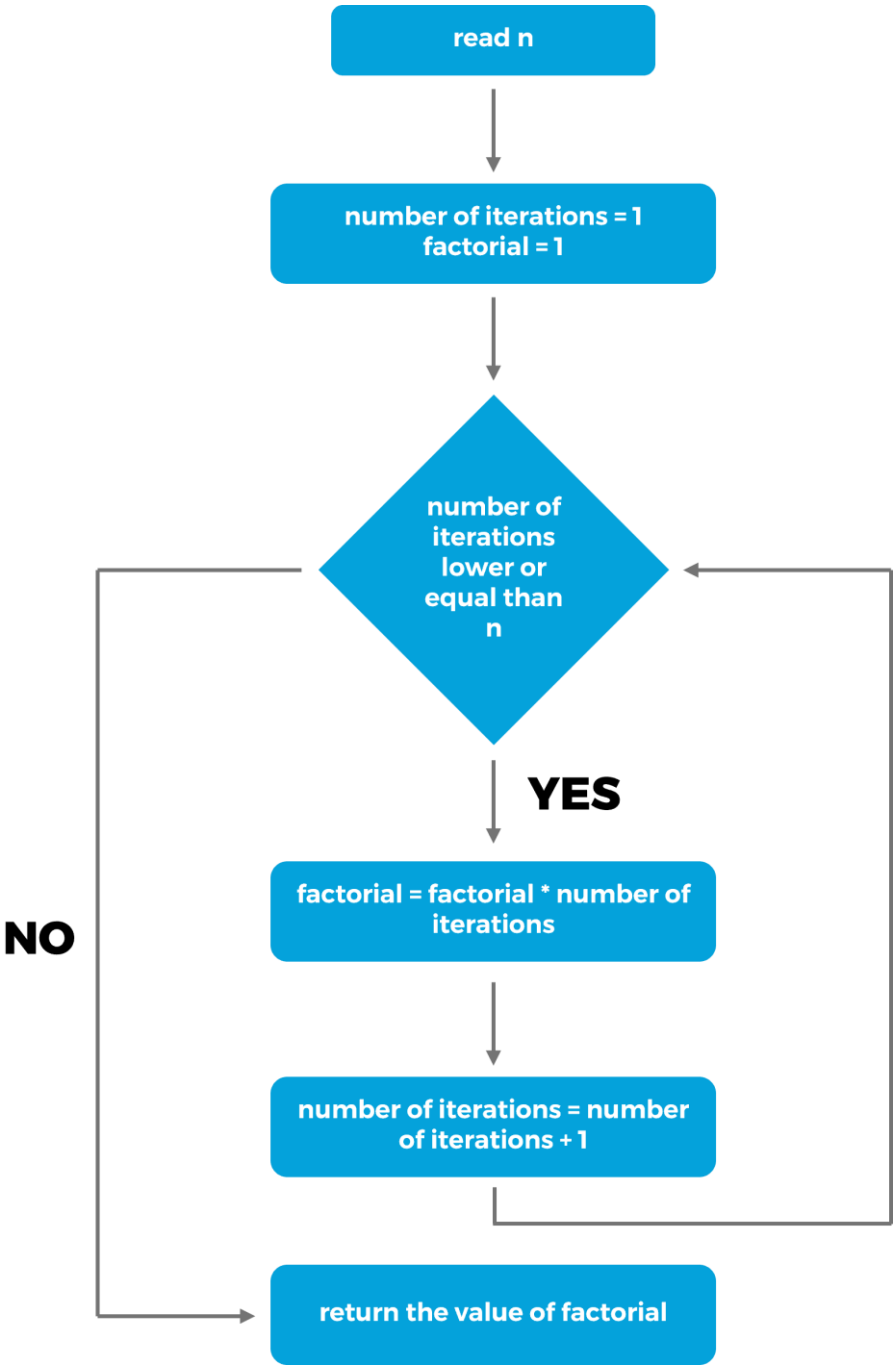
- Repetitiv sau Iterativ: Repetiția este inima acestor algoritmi. Ei execută un set de instrucțiuni în mod repetat până când o anumită condiție este îndeplinită. Buclele (cum ar fi "for" și "while") intră în această categorie. Ele sunt esențiale pentru sarcini precum prelucrarea unor seturi mari de date, simularea evenimentelor sau rezolvarea problemelor de optimizare.

Exemplu: Un program care calculează factorialul unui număr folosind o buclă. Factorul unui număr este o operație matematică care înmulțește un număr dat, n, cu fiecare număr întreg mai mic decât n până la 1.

1. Citiți n, adică numărul pentru care dorim să calculăm factorialul.
2. Setați valoarea factorială este 1.
3. Repetați de la 1 la n:
4. Înmulțiți factorialul cu valoarea buclei.
5. Arătați valoarea factorialului.

Rețineți că indentarea instrucțiunii numărul 4 are, de asemenea, un rol specific în definirea algoritmului. Această instrucțiune este executată de n ori, o dată la fiecare iterație.

Graficul 1. Exemplu de buclă



Sursă | Elaborare proprie

Putem construi algoritmi din ce în ce mai complecși prin combinarea acestor trei tehnici. Să vedem un exemplu de algoritm care utilizează toate schemele de compunere pentru a găsi valoarea maximă a unui număr întreg într-o listă:

Listă cu 10 numere întregi. Numărul 10 este primul element al listei.

1. Elementul curent este primul element al listei.
2. Valoarea maximă este elementul curent.
3. Deși nu evaluăm toate elementele din listă, bucla:
4. Dacă elementul curent este mai mare decât valoarea maximă, atunci:
5. Valoarea maximă este elementul curent.
6. Elementul curent este următorul element al listei.
7. Returnează valoarea maximă.

2.4. Eficiența algoritmului

În domeniul informaticii, algoritmii sunt reprezentați ca programe. Un program poate fi definit ca fiind descrierea unui algoritm în repertoriul finit de instrucțiuni de mașină. Setul de programe de care dispune un echipament informatic se numește software. Pentru a determina un calculator să îndeplinească o sarcină, trebuie mai întâi să gândim și să proiectăm un algoritm pentru rezolvarea acesteia, apoi să îl programăm în calculator. Scopul este de a transforma algoritmul conceptual într-un set clar de instrucțiuni într-un limbaj de programare specific, bazat pe diferite abordări ale procesului de programare (paradigme de programare).

Pe măsură ce problemele devin mai complexe, algoritmii care le rezolvă devin la fel. În acest moment devine necesar să se poată evalua costul pe care îl are un algoritm. Eficiența algoritmică se referă la eficiența cu care un algoritm utilizează resursele de calcul. Această proprietate este esențială pentru înțelegerea performanței unui algoritm și implică analiza consumului său de resurse, inclusiv timp și spațiu. Atingerea eficienței maxime presupune minimizarea utilizării resurselor. Cu toate acestea, compararea eficienței algoritmilor poate fi complexă, deoarece resursele diferite nu pot fi comparate direct.

Putem considera un algoritm eficient atunci când consumul său de resurse, denumit cost de calcul, rămâne la un prag acceptabil sau sub acest prag. Acest prag este determinat pe baza capacității algoritmului de a funcționa într-un interval de timp sau spațiu rezonabil pe un computer standard, adesea în raport cu dimensiunea datelor de intrare. În esență, "acceptabil" implică faptul că algoritmul poate fi executat în cadrul unor constrângeri practice, asigurându-se că rămâne viabil pentru utilizarea în lumea reală.

Există două măsuri principale ale eficienței algoritmice:

Complexitatea temporală: Aceasta este complexitatea computațională care descrie timpul de execuție al unui algoritm în funcție de mărimea datelor de intrare ale programului.

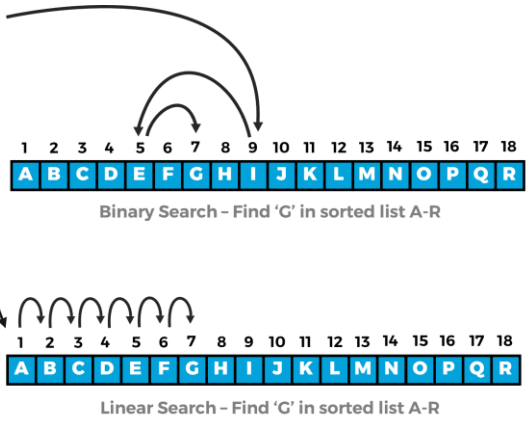
Complexitatea spațiului: Aceasta este cantitatea de memorie de care are nevoie un algoritm pentru a rula până la finalizare.

Scopul este de a minimiza atât timpul, cât și spațiul necesar. Cu toate acestea, există adesea un compromis între cele două. De exemplu, un algoritm care rulează mai repede poate necesita mai multă memorie, iar un algoritm care utilizează mai puțină memorie poate fi mai lent. Alegerea algoritmului depinde de cerințele specifice ale sarcinii. Algoritmii eficienți stabilesc un echilibru între aceste considerente, urmărind performanța optimă în scenariile din lumea reală.

Să vedem un exemplu simplu de doi algoritmi diferiți care pot fi utilizați pentru a căuta un element într-o listă, dar care au aceeași complexitate spațială, dar complexitate temporală diferită. Apoi vom analiza când să folosim fiecare dintre ei:

Căutarea liniară este un algoritm utilizat pentru a găsi o valoare într-o listă, este foarte asemănător cu algoritmul de găsire a valorii maxime a listei pe care l-am dezvoltat anterior. Acesta evaluează fiecare element al listei, începând de la început până când se găsește valoarea țintă sau se ajunge la sfârșitul listei. În timpul fiecărei iterații, algoritmul compară elementul curent cu valoarea țintă. Dacă se găsește o potrivire, algoritmul se încheie, returnând indexul valorii țintă.

Căutarea binară este un algoritm eficient utilizat pentru a găsi o valoare țintă în cadrul unei liste sortate prin împărțirea repetată la jumătate a intervalului de căutare. Acesta începe prin examinarea elementului central al listei și îl compară cu valoarea țintă. Dacă elementul din mijloc se potrivește cu valoarea țintă, căutarea are succes. În caz contrar, dacă valoarea țintă este mai mică decât elementul din mijloc, căutarea continuă în jumătatea inferioară a listei; dacă valoarea țintă este mai mare, căutarea trece în jumătatea superioară. Acest proces de înjumătățire a intervalului de căutare se repetă până când valoarea țintă este găsită sau intervalul de căutare devine gol. Căutarea binară funcționează pe principiul divide și cucerește, reducând semnificativ spațiul de căutare cu fiecare iterație.



Sursă | Board Infinity

Cu aceste două exemple putem înțelege cu ușurință că există mai mulți algoritmi pentru îndeplinirea aceleiași sarcini. Fiecare algoritm are propriile condiții, puncte forte și puncte slabe; căutarea liniară este ușor de implementat și potrivită pentru liste mici și date nesortate. Cu toate acestea, eficiența algoritmului scade pe măsură ce dimensiunea listei crește, ceea ce îl face mai puțin util pentru seturile mari de date, deoarece dacă elementul pe care îl căutăm este aproape de sfârșitul listei, trebuie să evaluăm majoritatea elementelor acesteia. Căutarea binară este foarte eficientă, ceea ce o face ideală pentru căutarea în seturi de date mari și sortate. Cu toate acestea, ea necesită ca datele să fie sortate inițial, ceea ce poate fi o condiție prealabilă pentru aplicarea sa.

2.5. Limitări

În ultimii câțiva ani, am observat o creștere remarcabilă a tehnologiilor de învățare automată și inteligență artificială, ambele bazându-se pe algoritmi pentru a procesa informații din seturi masive de date. Această creștere a alimentat progrese remarcabile în diverse domenii, inclusiv recunoașterea și generarea de imagini și clipuri video, robotică, analiză medicală, luarea deciziilor, clonarea vocii sau prelucrarea limbajului natural. Aceste progrese ilustrează potențialul nelimitat pe care îl oferă algoritmii, propulsându-ne într-o eră marcată de inovații fără precedent, însă este important să înțelegem că, chiar și cu aceste progrese importante, algoritmii au propriile limite. Există probleme care nu pot fi rezolvate de un algoritm și există probleme pe care nu suntem pregătiți să le rezolvăm.

În primul rând, vom descrie problemele pe care un algoritm nu le poate rezolva. Există două categorii: Problemele indecidabile, cele care sunt teoretic imposibil de rezolvat de către orice algoritm și problemele intractabile, acele probleme care nu au soluții rezonabile în timp.

Un exemplu clasic de problemă indecidabilă este problema halting, o problemă de decizie cu un răspuns binar (da sau nu) care este indecidabilă. Aceasta

reprezintă provocarea de a determina dacă un program de calculator se va opri în cele din urmă sau va rula indefinit. Enunțul acestei probleme este de genul: "Este posibil să se scrie un program care poate spune, dat fiind orice program și intrările sale și fără a executa programul, dacă acesta se va opri?

Pe de altă parte, există probleme greu de rezolvat. Unele probleme greu de rezolvat pot fi rezolvate într-un timp rezonabil numai pentru intrări mici, dar algoritmul nu poate fi adaptat la intrări mai mari. Unele probleme intratabile sunt rezolvate zilnic, chiar dacă sunt intratabile. Cum ar fi algoritmii de rutare, planificarea timpului, programarea resurselor. Acestea tind să utilizeze o soluție ghicită la o problemă intratabilă care poate fi verificată rapid. Aceasta se numește euristică și utilizează cunoștințele și experiența pentru a face presupuneri inteligente. Abordarea euristică poate, de asemenea, să încerce să găsească o soluție adecvată și nu neapărat perfectă sau ideală.

Una dintre cele mai cunoscute probleme greu de rezolvat este **problema vânzătorului ambulant** (**Travelling Salesman Problem -** TSP). În această provocare, sarcina este de a găsi cel mai scurt traseu care permite unui vânzător să viziteze fiecare oraș de pe o hartă exact o dată înainte de a se întoarce în orașul de plecare. Scopul este de a minimiza distanța totală parcursă, ceea ce reprezintă o mare provocare de calcul din cauza creșterii exponențiale a posibilităților odată cu creșterea numărului de orașe.

Să facem câteva numere pentru a observa cum crește. Pentru a calcula numărul de drumuri între orașe trebuie să calculăm factorialul numărului de orașe minus unu. Numărul de drumuri pe care trebuie să le explorăm este jumătate din numărul de legături între orașe (trebuie să folosim fiecare drum o singură dată).

Privind tabelul, putem observa cum numărul de posibilități crește foarte repede :[3]

Number of cities	Number of connections between cities	Number of possible routes
5	4! = 24	12
10	9! = 362.880	181.440
15	14! = 87.178.291.200	43.589.145.600

2.6. Limitări ale algoritmilor de învățare automată

În era învățării automate (profunde), avem o colecție imensă de algoritmi concepuți pentru a rezolva sarcini multiple, de exemplu: Arbori de decizie, Random Forest, Support Vector Machines sau rețele neuronale (în toate versiunile sale) ... Toate acestea pot fi adaptate la diferite probleme printr-un proces de formare în care le prezentăm date, iar ele se adaptează. Prin acest proces putem

[3] Exemplu preluat de la https://runestone.academy/ns/books/published/mobilecsp/Unit5-Algorithms-Procedural-Abstraction/Limits-of-Algorithms.html

obține soluții adecvate. Este important să rețineți că acestea au, de asemenea, propriile lor limitări.

- **Calitatea și cantitatea datelor:** Algoritmii de învățare automată se bazează foarte mult pe calitatea și cantitatea datelor disponibile pentru instruire. Datele limitate sau distorsionate pot duce la predicții inexacte sau distorsionate ale modelului. În plus, datele insuficiente pot împiedica capacitatea algoritmului de a generaliza bine la date nevăzute.

- **supraadaptarea și subadaptarea:** Modelele de învățare automată pot suferi de supraadaptare, atunci când învață să capteze zgomotul sau modelele irelevante din datele de instruire, ceea ce duce la performanțe slabe pe date noi. Pe de altă parte, subadaptarea apare atunci când modelul este prea simplist, ceea ce duce la performanțe suboptimale.

- **Interpretabilitatea:** Multe modele de învățare automată, în special cele complexe precum rețelele neuronale, nu sunt interpretabile, ceea ce face dificilă înțelegerea și încrederea în predicțiile lor. Acest lucru poate fi problematic în domenii în care interpretabilitatea este esențială, cum ar fi sănătatea și finanțele.

- **Resurse computaționale:** Formarea modelelor complexe de învățare automată necesită adesea resurse de calcul semnificative, inclusiv hardware de înaltă performanță și infrastructură de prelucrare a datelor la scară largă. Resursele limitate de calcul pot constrânge scalabilitatea și caracterul practic al soluțiilor de învățare automată.

- **Expertiza specifică domeniului:** Dezvoltarea de soluții eficiente de învățare automată necesită adesea expertiză specifică domeniului pentru preprocesarea corespunzătoare a datelor, pentru a crea caracteristici relevante și pentru a interpreta rezultatele modelului.

- **Învățarea continuă și adaptarea:** Algoritmii de învățare mecanică pot avea dificultăți în a se adapta la mediile în schimbare sau la distribuțiile de date care evoluează în timp. Monitorizarea și reantrenarea continuă a modelelor sunt necesare pentru a se asigura că performanța acestora rămâne robustă și actualizată.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sfaturi și recomandări pentru profesori pentru a se implica în conceptele de complexitate a algoritmilor

1. Construiți o fundație solidă
Începeți cu o prezentare generală de bază a ceea ce sunt algoritmii și a dezvoltării lor istorice pentru a oferi un context. Acest lucru ajută la construirea unei narațiuni care face conceptul abstract mai tangibil și mai ușor de abordat.

2. Utilizați analogii și metafore
Comparați algoritmii cu activități cotidiene precum gătitul sau navigarea într-un oraș. Acest lucru poate ajuta la demistificarea conceptului și îl poate face accesibil cursanților care nu au o pregătire tehnică solidă.

3. Demonstrații interactive
Arătați algoritmii în acțiune prin sesiuni de programare live sau instrumente web interactive care permit participanților să introducă exemple și să vadă cum le procesează algoritmii. Acest lucru poate fi deosebit de atractiv și informativ.

4. Promovarea unei săli de clasă interactive
Încurajați întrebările și discuțiile pentru a permite participanților să își exprime înțelegerea sau confuzia cu privire la subiectele abordate. Acest lucru ajută, de asemenea, la evaluarea implicării și înțelegerii lor în timp real.

5. Învățarea incrementală
Descompuneți subiectele complexe în părți mai mici, ușor de gestionat. Începeți cu concepte de bază și introduceți treptat mai multă complexitate. Această abordare ajută la prevenirea supraîncărcării cognitive.

Activități de încălzire

1. Crearea unui algoritm simplu
Rugați participanții să scrie un algoritm simplu pentru o sarcină de zi cu zi, cum ar fi prepararea ceaiului sau pregătirea pentru muncă. Această activitate îi face să gândească în secvențe logice pas cu pas, care sunt fundamentale pentru proiectarea algoritmilor.

2. Explorarea algoritmului istoric
Discutați algoritmul euclidian în grup, explorând etapele acestuia și modul în care a fost utilizat istoric pentru rezolvarea problemelor. Acest lucru poate duce la o apreciere mai profundă a modului în care conceptele fundamentale în algoritmi au rădăcini vechi.

3. Joc de rol privind structura de control
Organizați un joc de rol în care fiecare participant interpretează o parte a unei structuri de control dintr-un algoritm, cum ar fi o condițională sau o buclă. Acest lucru îi va ajuta să vizualizeze și să înțeleagă modul în care interacționează diferitele părți ale unui algoritm.

4. Provocarea de depanare a algoritmului
Furnizați un algoritm simplu cu erori intenționate și cereți participanților să îl "depaneze". Aceasta poate fi o modalitate distractivă și atractivă de a aprofunda înțelegerea modului în care algoritmii sunt structurați și funcționează.

Sfaturi și recomandări pentru evaluare

1. Feedback continuu
Oferiți feedback imediat și continuu pe parcursul cursului. Acest lucru ajută participanții să corecteze din timp neînțelegerile și le consolidează învățarea pe măsură ce progresează.

2. Sarcini practice
Concepeți sarcini care să solicite participanților să aplice ceea ce au învățat în scenarii practice. De exemplu, aceștia ar putea optimiza un algoritm existent sau identifica ineficiențele unui anumit algoritm.

3. Evaluarea colegială
Încorporați sesiuni de evaluare colegială în care participanții își evaluează reciproc munca. Acest lucru nu numai că oferă mai multe puncte de feedback, dar încurajează și învățarea în colaborare și gândirea critică.

4. Utilizați rubrici
Elaborați rubrici clare, bazate pe criterii, pentru evaluarea temelor. Acest lucru asigură consecvență în notare și așteptări clare pentru participanți.

5. Autoevaluarea
Încurajați participanții să își evalueze propria muncă pe baza criteriilor date. Acest lucru promovează auto-reflecția și învățarea aprofundată.

Aceste strategii îi vor ajuta pe formatori să prezinte în mod eficient conținuturi complexe privind algoritmii, asigurându-se în același timp că participanții sunt implicați activ și capabili să aplice practic cunoștințele lor.

3. Să recunoască rolul algoritmilor în diverse sectoare și potențialele capcane care decurg din utilizarea lor

Dobândirea unei înțelegeri profunde **a rolului algoritmilor în diferite sectoare**, recunoscând în același timp potențialele capcane, este esențială pentru promovarea unei viziuni cuprinzătoare și holistice asupra implicațiilor și responsabilităților legate de implementarea algoritmilor. Această noțiune subliniază necesitatea unei înțelegeri aprofundate a modului în care algoritmii pot influența diferite aspecte ale societății, economiei și guvernanței. Aplicațiile diverse ale algoritmilor, de la creșterea eficienței operaționale la personalizarea experiențelor utilizatorilor, demonstrează impactul semnificativ al acestora asupra vieții moderne (Mourtzis et al., 2022). Cu toate acestea, această influență implică și riscuri substanțiale, cum ar fi potențialul **de părtinire**, încălcarea **vieții private** și amplificarea **inegalităților societale existente** (Patel, 2024).

Implementarea algoritmilor trebuie să fie ghidată de considerente etice și **cadre de reglementare** pentru a atenua aceste riscuri (Tsamados et al., 2021). Este esențial să se dezvolte sisteme robuste de responsabilitate care să garanteze că algoritmii nu perpetuează sau exacerbează practicile sau disparitățile dăunătoare. Acest lucru include procese riguroase de testare și validare care evaluează algoritmii pentru **corectitudine** și acuratețe, în special în domenii sensibile precum asistența medicală, justiția penală și serviciile financiare. Părțile interesate, inclusiv tehnologii, factorii de decizie politică și publicul, trebuie să se angajeze în dialoguri continue pentru a examina implicațiile deciziilor algoritmice și pentru a se asigura că aceste instrumente servesc intereselor mai largi ale societății (Donovan et al., 2018).

În plus, **transparența** operațiunilor algoritmice și a proceselor decizionale este esențială. Publicul are nevoie de asigurări că deciziile algoritmice sunt luate din considerente etice și că sunt deschise examinării. Această transparență este esențială nu numai pentru construirea încrederii, ci și pentru a permite experților și autorităților de reglementare să efectueze audituri și evaluări în mod eficient (Felzmann et al., 2019). Inițiativele educaționale care îmbunătățesc alfabetizarea algoritmică în rândul publicului și al factorilor de decizie politică sunt, de asemenea, vitale (Reisdorf & Blank, 2021). Astfel de eforturi vor împuternici mai multe persoane să participe la discuții și să ia decizii în cunoștință de cauză cu privire la utilizarea și reglementarea acestor tehnologii.

Cercetarea și dezvoltarea în domeniul corectitudinii și eticii algoritmice trebuie să continue să evolueze. Pe măsură ce tehnologia avansează, vor apărea noi provocări, necesitând soluții inovatoare care să abordeze atât aspecte tehnice, cât și etice. Colaborarea între mediul academic, industrie și guvern poate facilita

schimbul de bune practici și poate promova dezvoltarea unor algoritmi mai echitabili și mai responsabili.

Discuția despre algoritmi trebuie să includă și potențialul de schimbare pozitivă. Atunci când sunt concepuți și implementați în mod responsabil, algoritmii au puterea de a spori accesibilitatea, eficiența și echitatea. De exemplu, în domeniul educației, algoritmii pot contribui la crearea unor experiențe de învățare personalizate care să se adapteze la nevoile individuale ale elevilor, ceea ce ar putea reduce decalajele în ceea ce privește nivelul de educație. În domeniul sănătății, algoritmii predictivi pot îmbunătăți diagnosticarea și rezultatele pentru pacienți prin identificarea riscurilor și recomandarea unor planuri de tratament personalizate.

În concluzie, integrarea algoritmilor în diverse sectoare prezintă atât oportunități, cât și provocări (Dwivedi et al., 2021). Prin promovarea unui mediu care încurajează practicile etice, transparența și incluziunea, societatea poate valorifica beneficiile algoritmilor, minimizând în același timp riscurile acestora. După cum sugerează Williamson (2018), o înțelegere nuanțată a acestor dinamici este esențială pentru realizarea întregului potențial al aplicațiilor algoritmice într-un mod care să fie benefic și corect pentru toți membrii societății.

3.1. Rolul algoritmilor în diferite sectoare

După cum subliniază Parker et al. (2016), "algoritmii au pătruns în toate sectoarele, revoluționând procesele și procesul decizional". În această secțiune, explorăm câteva sectoare importante în care influența algoritmilor este deosebit de semnificativă:

a. Finanțe: IA în finanțe a evoluat semnificativ de la începuturile sale rudimentare, devenind un instrument indispensabil în industrie. Utilizarea IA în finanțe poate fi urmărită până în anii 1980, când programele de bază au fost adaptate pentru a automatiza sarcini simple, cum ar fi introducerea datelor și contabilitatea. Cu toate acestea, adevărata fază de transformare a început la sfârșitul anilor 1990 și începutul anilor 2000, odată cu apariția unor algoritmi de învățare automată mai sofisticați și cu explozia disponibilității datelor (Aksoy & Gurol, 2021). Instituțiile financiare au recunoscut rapid potențialul IA de a oferi evaluări mai precise ale riscurilor, o procesare mai rapidă a tranzacțiilor și servicii îmbunătățite pentru clienți. Acest lucru a fost alimentat în continuare de creșterea puterii de calcul și de dezvoltarea tehnicilor avansate de analiză a datelor. Până în anii 2010, IA a început să remodeleze în mod profund tranzacționarea, subscrierea, detectarea fraudelor și gestionarea relațiilor cu clienții (Davenport & Mittal, 2023). Iată câteva exemple de utilizare a IA în sectorul financiar:

- Tranzacționarea algoritmică: Unul dintre cele mai faimoase exemple de tranzacționare algoritmică de succes este Renaissance Technologies, un

fond speculativ cunoscut pentru fondul său Medallion. Acest fond utilizează modele matematice complexe pentru a analiza și executa tranzacțiile la viteze mari. Utilizând algoritmi avansați, Renaissance a reușit să obțină randamente excepționale, depășind cu mult performanța pieței (Qin, 2012).

- Gestionarea riscurilor: JPMorgan Chase utilizează algoritmi pentru a gestiona riscul de credit. Sistemul lor evaluează riscul de neplată a creditelor pe baza a numeroși factori, inclusiv condițiile economice, performanța sectorului și istoricul de credit individual. Această abordare algoritmică le permite să își adapteze ofertele de credite și să minimizeze eventualele pierderi din creanțe neperformante.
- Detectarea fraudelor: PayPal utilizează algoritmi de învățare automată pentru a detecta tranzacțiile frauduloase. Acești algoritmi analizează mii de atribute ale tranzacției în timp real, inclusiv suma, dispozitivul utilizat și istoricul tranzacțiilor utilizatorului. Prin identificarea tiparelor care deviază de la normă, sistemele PayPal pot semnala cu mare precizie tranzacțiile potențial frauduloase, prevenind astfel pierderile financiare și protejând utilizatorii.
- Serviciul clienți: Bank of America utilizează un asistent virtual bazat pe inteligență artificială numit Erica. Acest instrument algoritmic ajută clienții să își gestioneze conturile, să urmărească cheltuielile și să efectueze plăți. De asemenea, Erica poate oferi actualizări în timp real cu privire la modificările contului și poate sugera modalități de economisire a banilor pe baza modelelor de cheltuieli analizate de algoritmi.

b. Asistența medicală: Inteligența artificială în domeniul asistenței medicale a cunoscut o evoluție remarcabilă, de la primele aplicații experimentale până la a deveni o componentă esențială a practicii medicale moderne. Această transformare a fost alimentată în mare măsură de progresele în învățarea automată, analiza datelor mari și digitalizarea crescândă a dosarelor medicale. Pe măsură ce ne aprofundăm, vom explora diferitele fațete ale implementării IA în domeniul asistenței medicale, inclusiv asistența pentru diagnosticare, tratamentul personalizat, predicția bolilor, gestionarea pacienților și intervențiile chirurgicale robotizate (Bohr & Memarzadeh, 2020). Iată câteva exemple remarcabile:

- **Asistență pentru diagnosticare:** Capacitatea IA de a analiza seturi mari de date a revoluționat procesele de diagnosticare în domeniul asistenței medicale. Prin integrarea IA în tehnologiile de imagistică, cadrele medicale pot detecta bolile cu o precizie și o viteză mai mari decât metodele tradiționale. IBM Watson Health demonstrează puterea AI în îmbunătățirea preciziei diagnosticării. Algoritmii AI avansați ai Watson pot analiza semnificația și contextul datelor structurate și nestructurate din notele și rapoartele clinice. De exemplu, acesta a fost utilizat în oncologie pentru a identifica opțiuni de tratament pentru pacienții cu cancer prin compararea

a milioane de note clinice oncologice cu fișele medicale ale pacienților și cu literatura de specialitate existentă.

- **Tratament personalizat:** IA facilitează medicina personalizată prin luarea în considerare a profilurilor genetice individuale, a factorilor de mediu și a stilului de viață ales pentru a adapta tratamentele. Această abordare îmbunătățește semnificativ rezultatele pentru pacienți prin direcționarea terapiilor care au cele mai mari șanse de a funcționa pentru anumiți pacienți. **Tempus** utilizează inteligența artificială pentru a colecta și analiza cantități mari de date genomice și clinice pentru a ajuta medicii să creeze planuri de tratament personalizate pentru pacienții cu cancer. Platforma sa utilizează algoritmi de învățare automată pentru a înțelege datele moleculare și terapeutice și pentru a prezice care tratamente sunt susceptibile de a fi cele mai eficiente pentru pacienții individuali.
- **Predicția și gestionarea bolilor**: Modelele de inteligență artificială sunt din ce în ce mai utilizate pentru a prezice probabilitatea apariției unei boli, evoluția acesteia și eventualele complicații. Această abordare proactivă permite intervenții mai timpurii, ceea ce poate salva vieți și reduce costurile asistenței medicale. DeepMind de la Google a dezvoltat sisteme AI care pot prezice cu o precizie remarcabilă leziunile renale acute cu până la 48 de ore înainte ca acestea să se producă. Modelul AI analizează o gamă largă de date de sănătate în timp real, permițând intervenții în timp util care pot preveni deteriorarea și îmbunătăți rezultatele.
- **Gestionarea și monitorizarea pacienților**: IA îmbunătățește gestionarea pacienților prin monitorizarea continuă a datelor de sănătate prin intermediul tehnologiei portabile și al dispozitivelor IoT. Aceste instrumente alertează furnizorii de servicii medicale cu privire la schimbările în starea pacientului, permițând un răspuns imediat și ajustări continue ale îngrijirii. Virta Health utilizează instrumente bazate pe IA pentru gestionarea bolilor cronice, cum ar fi diabetul de tip 2. Platforma sa monitorizează datele pacienților în timp real și ajustează dinamic planurile de tratament, ajutând la menținerea unor niveluri optime ale glicemiei și reducând dependența de medicamente.
- **Chirurgia robotică**: Chirurgia robotică, asistată de inteligența artificială, oferă o precizie ridicată, traume reduse și perioade de recuperare mai rapide. Algoritmii AI furnizează date în timp real chirurgilor și pot chiar automatiza anumite aspecte ale intervenției chirurgicale pentru a îmbunătăți siguranța și rezultatele. Sistemul chirurgical robotic da Vinci al Intuitive Surgical utilizează inteligența artificială pentru a spori precizia chirurgicală. Sistemul oferă chirurgilor vederi 3D de înaltă definiție foarte mărite ale locului intervenției chirurgicale și traduce mișcările mâinii chirurgului în mișcări mai mici și mai precise ale unor instrumente minuscule din corpul pacientului.

Pe măsură ce inteligența artificială continuă să progreseze, potențialul său de a transforma asistența medicală crește exponențial. Aceste exemple subliniază capacitatea IA de a spori acuratețea diagnosticelor, de a adapta tratamentele la fiecare pacient în parte, de a prezice evoluția bolilor, de a gestiona eficient

afecțiunile cronice și de a executa proceduri chirurgicale cu o precizie fără precedent. Integrarea inteligenței artificiale în asistența medicală nu numai că optimizează rezultatele pentru pacienți, ci și simplifică munca furnizorilor de asistență medicală, deschizând calea pentru abordări mai inovatoare ale medicinei și chirurgiei. Dezvoltarea și implementarea continuă a tehnologiilor IA sunt foarte promițătoare pentru abordarea unora dintre cele mai dificile probleme din domeniul asistenței medicale de astăzi.

c. Comerțul electronic: IA îmbunătățește în mod semnificativ atât **experiența** clienților, cât și **pe cea a întreprinderilor**, în special în sectoarele comerțului cu amănuntul și al comerțului electronic. Prin acumularea de date ample despre consumatori și integrarea acestora în algoritmi de învățare automată, comercianții cu amănuntul pot dezvolta funcționalități avansate de personalizare, recomandare și automatizare. Aceste caracteristici bazate pe inteligență artificială sunt acum standard pe toate platformele de cumpărături și servesc la îmbunătățirea implicării clienților și la eficientizarea operațiunilor, sporind astfel profiturile companiilor și eficiența resurselor. Acestea sunt câteva exemple de utilizare a IA în comerțul electronic și în comerțul cu amănuntul:

- **Strategii de publicitate îmbunătățite**
  Smartly.io: Utilizează inteligența artificială pentru a automatiza și optimiza campaniile de publicitate în social media, reducând semnificativ efortul manual și îmbunătățind în același timp performanța anunțurilor și implicarea pe toate platformele.
  eBay: Utilizează inteligența artificială pentru a oferi recomandări personalizate de cumpărături și sfaturi clienților, sporind satisfacția și fidelizarea cumpărătorilor.
- **Optimizarea vânzărilor de automobile**
  Cox Automotive: Implementează inteligența artificială prin intermediul instrumentului său Esntial pentru a eficientiza procesele de cumpărare online a automobilelor, inclusiv estimarea plăților și evaluarea riscurilor, îmbunătățind astfel parcursul digital al clienților.
- **Sisteme eficiente de livrare**
  Gopuff: Integrează inteligența artificială pentru a optimiza rutele de livrare, asigurând livrarea rapidă și eficientă a produselor esențiale de zi cu zi prin intermediul microcentrelor sale de aprovizionare.
- **Dezvoltarea de produse inovatoare**
  Mondelez International: Utilizează inteligența artificială în sectorul său de cercetare și dezvoltare pentru a accelera inovarea și a îmbunătăți eficiența costurilor în producția de gustări.
- **Experiențe de cumpărături personalizate**
  Mirakl: Folosește inteligența artificială pentru a personaliza recomandările de produse, îmbunătățind implicarea clienților și vânzările pe piețele digitale.
  Rue Gilt Groupe: Aplică inteligența artificială pentru a rafina recomandările de produse, îmbunătățind experiența de cumpărare pe site-urile sale de produse de lux.
- **Managementul avansat al stocurilor și al lanțului de aprovizionare**

IBM Watson: sprijină companiile de retail în eficientizarea operațiunilor și personalizarea interacțiunilor cu clienții prin analiza datelor în timp real.
Anaplan: Utilizează inteligența predictivă pentru a prognoza rezultatele afacerii și a optimiza strategiile de vânzare cu amănuntul, influențând totul, de la gestionarea stocurilor la achiziția de clienți.

- **Detectarea contrafacerii și moderarea conținutului**
  3PM Solutions: Folosește inteligența artificială pentru a detecta produsele contrafăcute și pentru a asigura acuratețea evaluărilor vânzătorilor pe principalele piețe online.
- **Inteligența artificială conversațională și implicarea clienților**
  Valyant AI: dezvoltă instrumente conversaționale bazate pe inteligența artificială pentru restaurantele cu servire rapidă, îmbunătățind experiența clienților în materie de comenzi prin intermediul tehnologiilor bazate pe voce.

d. Transporturi: Fără îndoială, IA remodelează industria transporturilor prin creșterea eficienței, siguranței și experienței utilizatorilor. Această secțiune analizează diferitele moduri în care IA este integrată în sistemele de transport, de la transportul public și logistică la mobilitatea personală și gestionarea traficului.

- **Vehicule autonome**
  Mașini care se conduc singure: IA alimentează funcționalitățile de bază ale vehiculelor autonome, permițându-le să navigheze, să evite obstacolele și să ia decizii în timp real. Companii precum Tesla și Waymo conduc progresele în acest domeniu, îmbunătățind semnificativ siguranța rutieră și eficiența vehiculelor.
  Drone pentru livrări: Companii precum Amazon experimentează cu drone care utilizează inteligența artificială pentru a livra pachete în mod autonom, promițând reducerea timpilor de livrare și a costurilor pentru logistica ultimului kilometru.
- **Managementul traficului și orașele inteligente**
  Sisteme inteligente de trafic: Inteligența artificială este utilizată pentru a analiza tiparele de trafic și pentru a optimiza secvențele de semaforizare, reducând congestionarea traficului și sporind siguranța rutieră. Orașe precum Barcelona și Singapore utilizează sisteme de inteligență artificială pentru a menține fluiditatea traficului în intersecțiile aglomerate.
  Întreținerea predictivă: Inteligența artificială ajută la predicția momentului în care vehiculele de transport public și infrastructura (cum ar fi podurile și drumurile) vor avea nevoie de întreținere, prevenind defecțiunile și prelungind durata de viață a activelor publice.
- **Optimizarea transportului în comun**
  Planificarea rutelor: Algoritmii AI optimizează orarele autobuzelor și trenurilor pe baza datelor în timp real și a modelelor istorice de trafic, maximizând eficiența operațională și minimizând timpul de așteptare pentru pasageri.

Gestionarea capacității: În timpul orelor de vârf, **sistemele AI pot prezice fluxul de pasageri și pot ajusta în consecință frecvența serviciilor de transport în comun, sporind confortul navetiștilor.**

- **Marfă și logistică**
  Depozitarea automatizată: Companii precum Amazon utilizează roboți cu inteligență artificială pentru a eficientiza operațiunile de depozitare, crescând viteza și acuratețea proceselor de preluare și ambalare.
  Rutare optimizată: Instrumentele AI analizează numeroase variabile, cum ar fi vremea, condițiile de trafic și ferestrele de livrare, pentru a sugera cele mai eficiente rute pentru expediere, economisind timp și combustibil.
- **Caracteristici de siguranță îmbunătățite**
  Sisteme de evitare a coliziunilor: Senzorii și sistemele de la bordul vehiculelor bazate pe inteligență artificială pot identifica pericolele potențiale și pot lua măsuri imediate pentru a evita accidentele.
  Sisteme de monitorizare a șoferilor: Aceste sisteme utilizează inteligența artificială pentru a monitoriza vigilența și starea generală de sănătate a șoferilor, pentru a preveni accidentele cauzate de oboseala șoferilor sau de urgențele medicale.
- **Serviciul clienți și asistență**
  AI Chatbots: Companiile de transport utilizează AI chatbots pentru a oferi asistență în timp real călătorilor, răspunzând solicitărilor privind orarele, tarifele și întreruperile fără intervenție umană.
  Recomandări de călătorie personalizate: Algoritmii AI analizează preferințele utilizatorilor și comportamentul lor trecut pentru a oferi sugestii de călătorie personalizate, îmbunătățind experiența clienților.

e. educație: IA transformă peisajul educațional, permițând experiențe de învățare personalizate, automatizând sarcinile administrative și facilitând medii educaționale imersive. Această secțiune explorează diversele aplicații ale IA în educație, evidențiind modul în care aceasta sprijină profesorii, îmbunătățește învățarea elevilor și optimizează managementul educațional.

- **Învățarea personalizată**
  Platforme de învățare adaptivă: Sistemele AI precum DreamBox Learning și Knewton oferă experiențe de învățare personalizate prin adaptarea la ritmul și stilul de învățare al fiecărui elev. Aceste platforme evaluează nivelul de cunoștințe al elevilor și ajustează automat conținutul pentru a se potrivi nevoilor lor de învățare.
  Tutori și asistenți AI: Sistemele de meditații bazate pe inteligență artificială, cum ar fi Carnegie Learning, oferă feedback și asistență în timp real studenților, ajutându-i să înțeleagă subiecte complexe prin instruire și practică personalizate.
- **Dezvoltarea conținutului educațional**
  Generarea automată de conținut: Instrumentele AI pot ajuta la crearea și personalizarea conținutului educațional. Instrumente precum Quillionz utilizează inteligența artificială pentru a genera teste și rezumate din manuale și articole, îmbunătățind materialele de studiu fără o contribuție umană semnificativă.

Aplicații de învățare a limbilor străine: Aplicații precum Duolingo utilizează inteligența artificială pentru a adapta cursurile de învățare a limbilor străine la nivelul de competență și progresul utilizatorului, făcând învățarea limbilor străine mai accesibilă și mai eficientă.

- **Evaluare și feedback**
  Sisteme automatizate de notare: Inteligența artificială automatizează notarea testelor standardizate și chiar a eseurilor, reducând povara administrativă asupra educatorilor și permițându-le să se concentreze mai mult pe predare și mai puțin pe notare.
  Analiză predictivă: Sistemele AI analizează datele elevilor pentru a prezice riscurile și rezultatele școlare, permițând educatorilor să intervină din timp asupra elevilor cu risc pentru a le îmbunătăți rezultatele la învățătură.
- **Automatizare administrativă**
  Procesele de înscriere și admitere: Inteligența artificială simplifică sarcinile administrative, cum ar fi admiterea și înscrierea, prin automatizarea procesării datelor și a procesului decizional, reducând astfel birocrația și îmbunătățind acuratețea.
  Gestionarea resurselor: Instrumentele AI ajută la gestionarea resurselor școlare, a programării și a prezenței elevilor, optimizând operațiunile și reducând costurile generale.
- **Accesibilitate îmbunătățită**
  Tehnologii asistive: Instrumentele bazate pe inteligență artificială, cum ar fi recunoașterea vorbirii și conversia text-vorbire, fac materialele de învățare mai accesibile elevilor cu handicap, asigurând incluziunea și egalitatea de șanse pentru toți elevii.
  Ajutoare pentru învățarea vizuală și auditivă: Aplicațiile bazate pe inteligență artificială, cum ar fi Immersive Reader de la Microsoft, ajută elevii cu dislexie prin oferirea de asistență la citire, demonstrând că tehnologia poate reduce decalajele de învățare.
- **Mediile de învățare imersive**
  Realitatea virtuală (VR) și realitatea augmentată (AR): Inteligența artificială integrată cu VR și AR creează experiențe de învățare imersive care simulează scenarii din lumea reală, făcând subiecte complexe precum știința și istoria mai atractive și mai ușor de înțeles.
  Jocuri și simulări educaționale: Jocurile educaționale îmbunătățite cu inteligență artificială se adaptează la răspunsurile elevilor, oferind un mediu de învățare dinamic care stimulează implicarea și îmbunătățește învățarea.

f. Securitatea și sectorul militar: Inteligența artificială este din ce în ce mai importantă pentru consolidarea măsurilor de securitate și a operațiunilor militare. Această secțiune prezintă modul în care tehnologiile AI sunt integrate în strategiile de apărare, sistemele de supraveghere și protocoalele de securitate, îmbunătățind eficiența și eficacitatea și abordând în același timp provocări complexe în domeniul securității naționale și globale.

- **Supraveghere și monitorizare îmbunătățite**

Sisteme automatizate de supraveghere: Sistemele bazate pe inteligență artificială sunt utilizate pentru a monitoriza zonele sensibile, folosind analiza datelor în timp real pentru a detecta activitățile neobișnuite sau amenințările. Aceste sisteme pot face diferența între comportamentele normale și cele suspecte, reducând semnificativ eroarea umană și timpul de răspuns.
Supraveghere cu drone: Dronele cu inteligență artificială sunt utilizate pentru supravegherea aeriană, oferind o recunoaștere completă a zonelor greu accesibile sau a zonelor de conflict, fără a risca vieți omenești.

- **Securitate cibernetică și apărare**
  Detectarea și răspunsul la amenințări: Sistemele AI analizează modelele din traficul de rețea pentru a identifica potențialele amenințări cibernetice, cum ar fi atacurile malware și ransomware, mult mai rapid decât ar putea operatorii umani. Odată ce amenințările sunt detectate, AI poate, de asemenea, să automatizeze răspunsurile sau să recomande strategii de atenuare.
  Criptarea datelor: IA îmbunătățește tehnicile de criptare, făcând mai dificilă descifrarea informațiilor sensibile de către entități neautorizate, securizând astfel transmisiile de date în rețelele militare și de securitate.
- **Sisteme autonome de apărare**
  Unități de luptă robotizate: Inteligența artificială este utilizată pentru operarea unităților robotice autonome sau semiautonome care pot îndeplini diverse sarcini militare, de la recunoaștere la roluri de luptă, reducând nevoia de soldați umani în operațiuni periculoase.
  Sisteme de apărare antirachetă: Algoritmii AI ajută la prezicerea și interceptarea cu mare precizie a amenințărilor, cum ar fi rachetele sau vehiculele aeriene fără pilot (UAV), asigurând poziții de apărare proactive.
- **Analiza informațiilor și sprijinirea deciziilor**
  Integrarea și analiza datelor: Sistemele AI integrează și analizează cantități mari de informații din surse multiple, oferind o imagine de ansamblu cuprinzătoare care îi ajută pe strategii militari să ia decizii în cunoștință de cauză rapid și precis.
  Simulare și formare: Simulările bazate pe inteligență artificială și mediile de realitate virtuală (VR) oferă instruire realistă pentru personalul de securitate și soldați, pregătindu-i pentru o varietate de scenarii fără riscurile unei lupte reale.
- **Întreținere și logistică**
  Întreținerea predictivă: Instrumentele AI prezic momentul în care echipamentele militare au nevoie de întreținere, ceea ce ajută la prevenirea defecțiunilor și la prelungirea duratei de viață a bunurilor valoroase.
  Optimizarea lanțului de aprovizionare: IA optimizează logistica și gestionarea lanțului de aprovizionare, asigurându-se că resursele sunt alocate eficient și livrate acolo unde este necesar, în special în cadrul operațiunilor militare complexe.

g. Media și petrecerea timpului liber: IA a devenit o forță transformatoare în industriile media și de agrement, îmbunătățind crearea de conținut, personalizând experiențele utilizatorilor și optimizând operațiunile. Această secțiune analizează modul în care tehnologiile IA sunt integrate în diferite aspecte ale producției media, distribuției și activităților de petrecere a timpului liber.

- **Crearea și gestionarea conținutului**
  Jurnalismul automatizat: Instrumentele AI, cum ar fi software-ul de generare a limbajului natural (NLG), sunt utilizate pentru a crea articole de știri și rapoarte, în special pentru conținutul bazat pe date, cum ar fi rezultatele sportive și actualizările financiare, permițând publicarea mai rapidă și eliberând jurnaliștii umani pentru investigații aprofundate.
  Producție video și muzicală: Algoritmii AI ajută la editare, de la corecția culorilor în videoclipuri la mixarea sunetului în muzică, simplificând procesele de producție și permițând experimente mai creative.
- **Sisteme de personalizare și recomandare**
  Recomandări media: Platformele de streaming precum Netflix și Spotify utilizează inteligența artificială pentru a analiza obiceiurile de vizionare și, respectiv, de ascultare, oferind recomandări de conținut personalizate pentru a menține implicarea utilizatorilor și a îmbunătăți satisfacția acestora.
  Publicitate targetată: Analiza bazată pe inteligența artificială a datelor utilizatorilor ajută companiile media să creeze și să difuzeze reclame direcționate, sporind eficiența și eficacitatea campaniilor de marketing.
- **Angajarea audienței și media interactivă**
  Chatbots și asistenți virtuali: Roboții de chat alimentați cu inteligență artificială asigură interacțiunea în timp real cu utilizatorii, oferind asistență pentru clienți, recomandări de conținut și experiențe interactive în jocuri și medii de realitate virtuală (VR).
  Realitatea augmentată (AR) și VR: Inteligența artificială face parte integrantă din dezvoltarea experiențelor imersive AR și VR, sporind realismul și interactivitatea mediilor virtuale utilizate în jocuri, tururi virtuale și conținut educațional.
- **Eficiență operațională**
  Publicitatea programatică: Inteligența artificială automatizează cumpărarea și vânzarea de spațiu publicitar, optimizând plasarea și stabilirea prețurilor în timp real pentru a maximiza randamentul investițiilor pentru agenții de publicitate și instituțiile media.
  Distribuția conținutului: Instrumentele de inteligență artificială ajută la optimizarea strategiilor de distribuție, analizând modelele de consum pentru a determina cele mai bune canale și momente pentru lansarea conținutului, pentru a maximiza atingerea și implicarea.
- **Consolidarea proceselor creative**
  Scenariul și dezvoltarea intrigii: Instrumentele de inteligență artificială ajută la scrierea scenariilor sugerând evoluții ale intrigii, dialoguri și

interacțiuni între personaje pe baza convențiilor de gen și a preferințelor publicului.
Artă și design grafic: Instrumentele bazate pe inteligența artificială asistă artiștii și designerii prin sugerarea de îmbunătățiri, generarea de idei și automatizarea sarcinilor repetitive, permițând o mai mare libertate creativă și experimentarea.

- **Managementul și planificarea evenimentelor**
  Gestionarea mulțimilor: Inteligența artificială este utilizată în planificarea și gestionarea evenimentelor de amploare, analizând datele privind participanții pentru a optimiza amenajarea spațiilor, programarea și măsurile de securitate.
  Experiențe personalizate: În parcurile de agrement și la evenimente, sistemele de inteligență artificială oferă itinerarii și experiențe personalizate pe baza preferințelor vizitatorilor și a comportamentului anterior, sporind satisfacția și fidelizarea clienților.

3.2. Capcane potențiale ale utilizării algoritmului

În ciuda numeroaselor avantaje pe care le oferă, algoritmii pot provoca probleme semnificative, cum ar fi prejudecățile și discriminarea, problemele legate de confidențialitate, lipsa de transparență și de responsabilitate, precum și o încredere excesivă în automatizare (O'Neil, 2016). Fiecare dintre aceste capcane nu numai că prezintă riscuri etice și operaționale, dar ar putea submina și încrederea publicului în tehnologiile IA. În continuare, explorăm aceste categorii mai în detaliu, oferind exemple din diverse sectoare pentru a ilustra pericolele potențiale și pașii greșiți în implementarea IA.

a. Prejudecăți și discriminare: Sistemele AI pot perpetua și amplifica în mod involuntar prejudecățile dacă sunt antrenate pe seturi de date care nu sunt reprezentative sau conțin erori prejudiciabile. Această problemă este deosebit de importantă în sectoare precum finanțele și asistența medicală, unde astfel de prejudecăți pot duce la un tratament inechitabil al persoanelor.

- Finanțe: Sistemele de inteligență artificială utilizate în evaluarea creditelor pot reflecta prejudecățile rasiale sau de gen existente în datele istorice. De exemplu, dacă un model este antrenat pe baza datelor anterioare privind aprobarea împrumuturilor, iar aceste date reflectă prejudecăți istorice împotriva unui anumit grup demografic, modelul poate, de asemenea, să discrimineze acel grup.
- Sănătate: Au existat cazuri în care instrumentele de diagnosticare AI au arătat discrepanțe în recomandările de tratament bazate pe diferențe rasiale sau de gen. Un exemplu notabil este un sistem AI care a diagnosticat greșit anumite boli de piele la persoanele cu pielea mai închisă la culoare, din cauza unui set de date alcătuit predominant din persoane cu pielea deschisă la culoare.

b. Preocupări legate de confidențialitate: Volumul mare de date necesare pentru instruirea și funcționarea sistemelor de inteligență artificială ridică probleme semnificative legate de confidențialitate, în special în ceea ce privește modul în care datele sunt colectate, stocate și utilizate. Acest lucru este evident în sectoare precum comerțul electronic și educația, unde datele personale sunt utilizate pe scară largă pentru a îmbunătăți experiența clienților și a studenților.

- Comerțul electronic: Comercianții cu amănuntul care utilizează inteligența artificială pentru a personaliza experiențele de cumpărături pot colecta cantități mari de informații personale, de la istoricul achizițiilor la date de localizare în timp real. Acest lucru generează îngrijorări cu privire la potențialul de încălcare a securității datelor și de utilizare neautorizată a informațiilor personale.
- Educație: În sectorul educațional, sistemele de inteligență artificială utilizate pentru a urmări progresul elevilor pot colecta informații sensibile

despre dificultățile de învățare, date despre sănătate și chiar modele comportamentale. Acest lucru prezintă riscuri cu privire la cine are acces la aceste date și la modul în care acestea ar putea fi utilizate în afara contextului educațional.

c. Lipsa transparenței și a responsabilității: Sistemele AI funcționează adesea ca niște **"cutii negre"**, cu procese decizionale obscure sau ascunse utilizatorilor și autorităților de reglementare. Această lipsă de transparență poate duce la probleme de responsabilitate, în special atunci când deciziile au consecințe semnificative asupra vieții oamenilor.

- Securitate și armată: În aplicațiile militare, procesul decizional bazat pe inteligența artificială în cazul armelor autonome poate conduce la rezultate de viață sau de moarte. Lipsa de transparență a modului în care sunt luate aceste decizii complică implicațiile etice și ridică probleme semnificative de responsabilitate.
- Asistență medicală: Este posibil ca un sistem de inteligență artificială utilizat pentru diagnosticarea pacienților să nu ofere informații despre procesul său decizional, ceea ce face dificil pentru medici să înțeleagă cum a ajuns la un anumit diagnostic. Acest lucru nu numai că face mai dificilă încrederea în judecata AI, dar complică și răspunderea în caz de diagnostic greșit.

d. Dependența excesivă de automatizare: Bazarea masivă pe inteligența artificială poate duce la o degradare a expertizei umane, deoarece competențele se atrofiază atunci când sistemele de inteligență artificială preiau procesele decizionale. Această dependență excesivă poate fi deosebit de problematică în medii cu mize mari, precum transporturile și finanțele.

- Transportul: Industria aviatică a înregistrat incidente în care dependența excesivă de sistemele automate a contribuit la producerea unor accidente. Uneori, piloții depind prea mult de pilotul automat și este posibil să nu aibă suficientă experiență de zbor manual pentru a prelua controlul eficient în caz de defecțiune a sistemului.
- Finanțe: **Piața bursieră a cunoscut mai multe prăbușiri fulgerătoare, atribuite parțial algoritmilor de tranzacționare automată**. Dependența excesivă de aceste sisteme, fără o supraveghere umană suficientă, poate duce la scăderi bruște ale pieței, bazate pe erori algoritmice.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sfaturi și recomandări pentru ca profesorii să se implice în rolul algoritmilor în diverse sectoare

1. Studii de caz de utilizare

Începeți sesiunile examinând studii de caz din lumea reală în care algoritmii au jucat un rol crucial în diferite sectoare. Acest lucru oferă o înțelegere practică a modului în care conceptele abstracte sunt aplicate în diferite industrii precum finanțele, asistența medicală și comerțul electronic.

2. Evidențierea considerentelor etice
Subliniați implicațiile etice ale utilizării algoritmilor, discutând potențialele prejudecăți, preocupările legate de confidențialitate și responsabilitatea. Acest lucru va încuraja gândirea critică și conștientizarea etică în rândul participanților.

3. Încurajarea discuțiilor interactive
Facilitați discuțiile în grup cu privire la implicațiile implementării algoritmilor. Puneți întrebări precum "Ce ar putea merge prost cu acest algoritm într-o aplicație din lumea reală?" sau "Cum putem reduce riscurile potențiale?"

4. Utilizarea resurselor multimedia
Încorporați videoclipuri, infografice și podcasturi care ilustrează impactul algoritmilor. Această varietate poate răspunde diferitelor stiluri de învățare și menține conținutul captivant.

5. Simularea impactului algoritmului
Utilizați simulări sau instrumente software care permit participanților să vadă efectele modificărilor algoritmului în medii simulate. Această abordare practică ajută la solidificarea înțelegerii prin aplicare practică.

Activități de încălzire

1. Maparea rolurilor algoritmului
Atribuiți roluri participanților în care fiecare persoană reprezintă o componentă a unui algoritm într-un anumit sector. Aceștia trebuie să explice impactul rolului lor și posibilele capcane, promovând o înțelegere imersivă a funcțiilor algoritmului și a implicațiilor etice.

2. Dezbaterea privind etica algoritmilor
Organizați o dezbatere cu privire la o utilizare controversată a algoritmilor, cum ar fi recunoașterea facială sau poliția predictivă. Acest lucru încurajează participanții să exploreze și să articuleze perspective diferite asupra provocărilor etice.

3. Analiza scenariului
Prezentați grupurilor mici scenarii care implică utilizarea algoritmilor în diverse sectoare, solicitându-le să identifice beneficiile potențiale, riscurile și deciziile etice pe care factorii de decizie ar trebui să le ia în considerare.

Sfaturi și recomandări pentru evaluare

1. Analiza studiului de caz
Cereți participanților să analizeze un studiu de caz în care algoritmii sunt utilizați într-un anumit sector. Evaluați capacitatea acestora de a identifica atât beneficiile, cât și considerentele etice, astfel cum sunt descrise în caz.

2. Prezentări de grup

Rugați participanții să prezinte diferite aspecte ale rolurilor și riscurilor algoritmilor în diverse sectoare. Evaluați înțelegerea lor pe baza profunzimii analizei și a capacității de a se implica în considerații etice.

3. Eseuri de reflecție
Cereți participanților să scrie eseuri în care să reflecteze asupra modului în care înțelegerea algoritmilor le poate influența responsabilitățile profesionale și considerentele etice. Evaluați capacitatea acestora de a integra conținutul cursului cu scenarii profesionale personale sau ipotetice.

4. Chestionare și teste
Utilizați chestionare pentru a evalua înțelegerea unor concepte-cheie precum părtinirea algoritmică și transparența. Includeți întrebări bazate pe scenarii care solicită gândirea critică dincolo de memorarea pe de rost.

5. Feedback-ul colegilor
Încorporați evaluarea inter pares ca parte a procesului de evaluare. Acest lucru nu numai că oferă perspective multiple asupra înțelegerii participantului, dar încurajează, de asemenea, angajamentul față de materialul de învățare din punctul de vedere al unui evaluator.

Folosind aceste strategii, formatorii pot educa în mod eficient participanții cu privire la impactul semnificativ al algoritmilor în toate sectoarele, asigurându-se că aceștia înțeleg atât aspectele tehnologice, cât și implicațiile etice mai ample.

4. Înțelegerea complexității algoritmice, a optimizării și a limitelor procesului decizional algoritmic

4.1. Complexitatea algoritmică

În informatică, algoritmii pot face sau desface succesul sistemelor și aplicațiilor, deoarece trebuie să funcționeze eficient. Complexitatea algoritmică este ca o lumină călăuzitoare pentru dezvoltatori și ingineri. Este vorba despre a calcula de câtă putere de calcul au nevoie diferiții algoritmi, iar noi folosim **notația Big O** pentru a face acest lucru. Dar nu este vorba doar despre a face lucrurile să funcționeze mai repede; înțelegerea complexității algoritmice ne ajută să știm cât de bine vor funcționa algoritmii noștri pe măsură ce mărim sistemele noastre și cum să gestionăm resursele de care au nevoie.

Complexitatea algoritmică se referă practic la cât de repede și cât de mult spațiu are nevoie un algoritm pentru a rezolva o problemă. Complexitatea temporală ne spune cât timp îi ia unui algoritm să rezolve o problemă în funcție de cât de mare este problema. Complexitatea spațială, pe de altă parte, se referă la cantitatea de memorie utilizată de algoritm în timpul funcționării. Aceste măsurători ne ajută să ne dăm seama cât de bine funcționează un algoritm și dacă poate face față sarcinilor mari fără să încetinească prea mult.

În lumea informaticii, există un lucru numit notația Big O, care este modalitatea matematică de a vorbi despre cât de repede sau lent rulează algoritmii. Aceasta ne ajută să înțelegem de cât timp sau spațiu are nevoie un algoritm pe măsură ce îi dăm mai multe lucruri cu care să lucreze. Astfel, dacă spunem că un algoritm este O(n), înseamnă că, pe măsură ce îi dăm mai multe date de intrare (să le numim "n"), îi ia liniar mai mult timp să termine. Dar dacă este O(n^2), atunci va dura mult mai mult timp pe măsură ce crește "n", deoarece crește în mod pătratic. Așadar, notația Big O este o modalitate rapidă de a înțelege cât de eficient va fi un algoritm atunci când are de-a face cu o mulțime de date.

Importanța analizei complexității algoritmice acoperă diverse domenii:

- **Predicția performanței:** Atunci când analizăm cât de complecși sunt algoritmii, putem anticipa modul în care aceștia vor funcționa cu diferite intrări. Această capacitate de a prevedea performanța este esențială pentru alegerea celui mai adecvat algoritm pentru o anumită problemă și pentru optimizarea performanței sistemului.

- **Evaluarea scalabilității:** În mediile în care dimensiunile datelor de intrare se pot modifica, înțelegerea complexității algoritmice ne ajută să evaluăm cât de bine se pot adapta algoritmii. Algoritmii cu caracteristici de complexitate favorabile funcționează fiabil într-o gamă largă de dimensiuni ale datelor de intrare, asigurând scalabilitatea și capacitatea de reacție.

- **Gestionarea resurselor:** Utilizarea eficientă a resurselor este esențială în mediile în care resursele sunt limitate, cum ar fi sistemele integrate și cloud computing. Analiza complexității algoritmice informează deciziile cu privire la modul de alocare a resurselor, asigurând utilizarea optimă și minimizând risipa.

- **Proiectarea algoritmilor:** Înțelegerea complexității algoritmice influențează procesul de proiectare, orientând dezvoltatorii către algoritmi care echilibrează eficiența și funcționalitatea. Prin selectarea algoritmilor cu caracteristici de complexitate favorabile, dezvoltatorii pot crea sisteme care funcționează optim fără a sacrifica funcționalitatea.

4..2. Complexitatea algoritmică temporală

Să luăm în considerare problema omniprezentă a sortării. Algoritmi precum sortarea cu bule, care au o complexitate algoritmică notată ca O(n2), tind să aibă performanțe slabe atunci când au de-a face cu seturi de date extinse, în comparație cu alternative mai eficiente precum sortarea prin fuzionare sau sortarea rapidă, care se laudă cu o complexitate algoritmică de O(n log n). Înțelegerea acestei diferențe este esențială pentru luarea unor decizii în cunoștință de cauză la selectarea algoritmilor. Acest lucru, la rândul său, conduce la îmbunătățirea performanței sistemului și, ulterior, la îmbunătățirea experienței utilizatorului (Karp, 2010).

Pentru o ilustrare mai clară, să luăm în considerare un scenariu în care trebuie să sortăm un set de date care conține un miliard de intrări (n = 1.000.000.000), ceea ce ar putea fi similar cu sortarea utilizatorilor unei rețele sociale mari.

- **Bubble Sort:** Cu o complexitate algoritmică de O(n2), atunci când este aplicat la acest set masiv de date, timpul de execuție ar fi astronomic. Pentru a oferi o estimare, timpul de execuție ar fi proporțional cu aproximativ 1018, rezultând un număr uluitor. Dacă fiecare intrare necesită 1 nanosecundă pentru sortare, timpul total necesar sortării cu bule pentru a procesa un miliard de intrări s-ar ridica la 1018 nanosecunde, echivalentul a 11574,07 zile sau aproximativ 31,6881 ani. Acest proces ineficient ar duce, probabil, la întârzieri semnificative în sortarea datelor, ceea ce ar cauza o experiență proastă a utilizatorului.

- **Quicksort:** Pe de altă parte, cu o complexitate temporală de O(n log n), quicksort se dovedește a fi semnificativ mai eficient. La sortarea unui set de date cu un miliard de intrări, timpul de execuție folosind quicksort ar fi proporțional cu aproximativ 1 000 000 000 000 * log2(1 000 000 000 000). Dacă fiecare intrare necesită 1 nanosecundă pentru sortare, timpul de execuție necesar pentru ca quicksort să sorteze un set de date conținând un miliard de intrări ar fi de aproximativ 29 897 352 853,986263 nanosecunde, echivalent cu 29,8974 secunde sau aproximativ 0,00034603 zile. În consecință, utilizatorii ar beneficia de timpi de răspuns mai rapizi și de o experiență generală mai plăcută.

Implementarea unor algoritmi de sortare mai eficienți, cum ar fi quicksort, ilustrează potențialul transformator al optimizării în procesele de calcul (Bentley, 1999)

De ce punem exemplul sortării? Organizarea datelor este esențială pentru ca algoritmii să funcționeze eficient, asemănător cu descoperirea unei bogății de beneficii, cum ar fi procesarea mai rapidă și interacțiunile mai fluide cu utilizatorii. Imaginați-vă cum ar fi să căutați un anumit obiect într-o cameră dezordonată; este consumator de timp și laborios. În mod similar, datele nesortate din algoritmi necesită o căutare liniară, în care fiecare element este inspectat până când este găsit cel dorit. Acest proces, cu o complexitate de O(n) (reprezentând dimensiunea setului de date), seamănă cu scanarea unei cărți pagină cu pagină pentru a localiza un cuvânt - o metodă funcțională, dar ineficientă, în special pentru seturile mari de date.

Acum, imaginați-vă aceeași cameră ordonată, cu obiectele aranjate cu grijă. Găsirea a ceea ce aveți nevoie se face fără efort, deoarece știți unde să căutați. În mod similar, datele sortate permit algoritmilor să utilizeze căutarea binară, o abordare mult mai eficientă. Căutarea binară împarte intervalul de căutare în două, în mod repetat, până la găsirea elementului dorit, cu o complexitate temporală de O(log n), scalând mai bine cu seturi de date mai mari. Este ca și cum ai localiza un cuvânt într-un dicționar răsfoind paginile, înjumătățind spațiul de căutare cu fiecare întoarcere.

În esență, sortarea datelor permite algoritmilor să lucreze mai inteligent, nu mai greu. Aceasta transformă o căutare liniară laborioasă într-o căutare binară rapidă

ca fulgerul, reducând timpii de procesare și îmbunătățind performanța generală. Așadar, data viitoare când întâlniți date nesortate, amintiți-vă de impactul sortării și de capacitatea acesteia de a debloca eficiența algoritmică.

4.3. Complexitatea algoritmică a memoriei

Atunci când se analizează complexitatea algoritmică privind memoria, se întâlnește un aspect critic al analizei algoritmilor care completează examinarea complexității temporale. Memoria, o resursă limitată în mediile de calcul, influențează profund performanța algoritmilor și eficiența sistemului. Explorarea conceptului de complexitate algoritmică legată de memorie poate părea complexă, dar haideți să o descriem în termeni mai simpli.

Atunci când vorbim despre complexitatea algoritmică și memorie, ne uităm practic la cât spațiu are nevoie un algoritm pentru a rezolva o problemă. Gândiți-vă că este ca și cum v-ați organiza geanta pentru diferite sarcini - unele sarcini ar putea necesita mai mult spațiu, în timp ce altele au nevoie de mai puțin.

Imaginați-vă că sortați o grămadă de cărți pe care sunt scrise numere. O modalitate de a face acest lucru se numește sortare prin fuzionare. Merge sort împarte cardurile în grupuri mai mici, sortează fiecare grup și apoi le pune la loc. Deși sortarea combinată este foarte bună la sortarea rapidă, are nevoie și de spațiu suplimentar pentru a lucra. Acest spațiu suplimentar este ca și cum ați avea mese suplimentare pentru a vă organiza cardurile în timp ce le sortați.

Un alt exemplu este atunci când calculăm numerele Fibonacci, precum cele găsite în natură (cum ar fi modelul petalelor unei flori sau spirala unei scoici). Există o modalitate de a calcula aceste numere folosind o metodă numită programare dinamică. Această metodă ne ajută să găsim numerele mai repede, dar are nevoie și de spațiu suplimentar pentru a memora numerele pe care le-a calculat deja. Este ca și cum ai avea un caiet în care să notezi numerele pe măsură ce le calculezi.

Acum, să vorbim despre modul în care stocăm informațiile în programele de calculator. Avem diferite instrumente numite structuri de date, cum ar fi matricele și listele, care ne ajută să păstrăm lucrurile organizate. Fiecare dintre aceste instrumente ocupă o anumită cantitate de spațiu în memoria computerului.

De exemplu, imaginați-vă că faceți o listă cu numele prietenilor dvs. Dacă folosiți o listă simplă, cum ar fi să le scrieți numele pe o bucată de hârtie, aceasta nu ocupă mult spațiu. Dar dacă folosiți ceva mai elegant, cum ar fi un tabel cu multe coloane, s-ar putea să ocupați mai mult spațiu, deoarece stocați informații suplimentare despre fiecare prieten.

Prin urmare, atunci când analizăm complexitatea algoritmică legată de memorie, ne gândim practic la cât spațiu au nevoie algoritmii și structurile noastre de date pentru a-și face treaba eficient. Înțelegând acest lucru, putem face alegeri mai

bune în programare pentru a ne asigura că programele noastre rulează fără probleme și nu utilizează prea multă memorie. Este ca și cum v-ați organiza geanta în mod inteligent, astfel încât să puteți transporta tot ce aveți nevoie fără ca aceasta să devină prea grea.

Pe scurt, gândirea la memorie și la complexitatea algoritmică ne ajută să scriem programe mai bune, care funcționează bine și nu acaparează toate resursele computerului. Este ca și cum ai fi un organizator inteligent pentru computerul tău!

Complexitatea algoritmică este un concept crucial în informatica modernă. Acesta oferă o modalitate structurată de a evalua și de a îmbunătăți eficiența algoritmilor. Prin utilizarea notației Big O, putem înțelege caracteristicile de bază ale algoritmilor și modul în care aceștia funcționează cu diferite cantități de date. Înțelegerea complexității algoritmice ajută dezvoltatorii și inginerii să creeze sisteme care pot face față cerințelor în continuă schimbare ale informaticii. Pe măsură ce tehnologia progresează, cunoașterea complexității algoritmice va continua să fie vitală pentru a stimula inovarea și a îmbunătăți ceea ce pot face calculatoarele.

4.4. Optimizare

Optimizarea reprezintă un far al inovației și al eficienței, ghidând algoritmii pentru a funcționa cu o viteză și o precizie de neegalat. În esența sa, optimizarea este arta de a rafina algoritmii pentru a minimiza complexitatea și a maximiza performanța, eliberând întregul lor potențial de a aborda probleme complexe cu finețe.

Imaginați-vă algoritmii ca fiind motoarele lumii noastre digitale, care alimentează totul, de la motoarele de căutare la sistemele logistice. La fel ca un motor bine reglat, un algoritm optimizat funcționează fără probleme, navigând rapid prin seturi vaste de date și calcule complicate. Dar cum anume realizează optimizarea această performanță?

Optimizarea cuprinde o multitudine de strategii menite să eficientizeze algoritmii. Aceasta implică identificarea blocajelor, eliminarea etapelor redundante și ajustarea proceselor pentru a asigura o eficiență maximă. În esența sa, optimizarea constă în găsirea echilibrului optim între utilizarea resurselor și calitatea producției.

Unul dintre obiectivele fundamentale ale optimizării este minimizarea complexității. Algoritmii întâmpină adesea provocări atunci când sunt însărcinați cu rezolvarea unor probleme complexe care necesită resurse de calcul extinse. Prin optimizarea algoritmilor, urmărim să le simplificăm structura și să reducem sarcina de calcul, făcându-i mai ușor de gestionat și mai scalabili.

În plus, optimizarea urmărește să îmbunătățească performanța prin utilizarea unor abordări inovatoare. Aceasta implică explorarea tehnicilor de ultimă oră,

cum ar fi calculul paralel, algoritmii euristici și învățarea automată, pentru a spori eficiența. Prin adoptarea acestor progrese, optimizarea permite algoritmilor să ofere rezultate mai rapide, consumând în același timp mai puține resurse. Luați în considerare, de exemplu, optimizarea algoritmilor de căutare utilizați de browserele de internet. Prin ajustări meticuloase și îmbunătățiri algoritmice, motoarele de căutare pot trece rapid prin cantități mari de date pentru a furniza rezultate relevante în câteva milisecunde. Această optimizare nu numai că îmbunătățește experiența utilizatorului, dar reduce și încărcarea serverelor și consumul de energie.

În domeniul informaticii, unitățile de procesare grafică (GPU) au devenit componente indispensabile, în special în contextul inteligenței artificiale (AI) și al învățării automate. În ciuda utilizării lor pe scară largă, mulți încă nu sunt familiarizați cu funcționarea internă a GPU-urilor și cu rolul lor esențial în avansarea tehnologiilor IA.

În esența sa, un GPU este un circuit electronic specializat conceput pentru a manipula și modifica rapid memoria pentru a accelera crearea de imagini într-un buffer de cadre destinat emiterii către un dispozitiv de afișare. Dezvoltate inițial pentru redarea grafică, GPU-urile au evoluat în procesoare puternic paralelizate, capabile să execute simultan mii de sarcini de calcul.

Spre deosebire de unitățile centrale de procesare (CPU) tradiționale, care excelează la sarcinile de procesare secvențială, GPU-urile sunt optimizate pentru procesarea paralelă. Acestea sunt formate din numeroase unități de procesare mai mici, numite "nuclee", fiecare capabilă să execute sarcini în mod independent. Această arhitectură paralelă permite GPU-urilor să abordeze sarcini intensive de calcul cu o eficiență remarcabilă. GPU a fost conceput pentru paralelizarea sarcinilor care implică efectuarea simultană a aceleiași operații pe mai multe bucăți de date. Acest paralelism permite GPU-urilor să abordeze sarcini intensive de calcul cu o eficiență remarcabilă (redare grafică și procesare de imagini și video care implică efectuarea a numeroase calcule pentru fiecare pixel, simulare și modelare, minare de criptomonede și operații matriceale paralelizate, printre altele).

Paralelismul este deosebit de avantajos în aplicațiile AI, unde sarcinile implică adesea procesarea simultană a unor cantități mari de date. Sarcini precum formarea rețelelor neuronale profunde, recunoașterea imaginilor, procesarea limbajului natural și analiza datelor beneficiază enorm de pe urma abilităților de procesare paralelă ale GPU-urilor.

Inteligența artificială se bazează în mare măsură pe operații matematice complexe, cum ar fi multiplicările matriceale și convoluțiile, care sunt fundamentale pentru formarea și implementarea modelelor de învățare automată. Aceste operații implică manipularea matricelor mari și efectuarea simultană a numeroase calcule, ceea ce le face candidați ideali pentru execuția paralelă pe GPU.

În acest moment, trebuie să introducem consumul de energie al algoritmilor AI. Deoarece aceste sisteme au devenit omniprezente în diverse domenii, este necesar să se pună accentul pe eficiența energetică pentru a atenua impactul asupra mediului și costurile operaționale. Tehnicile de optimizare și complexitatea algoritmică joacă un rol esențial în realizarea de economii de energie în cadrul AI. Optimizarea implică rafinarea algoritmilor pentru o performanță maximă a sarcinii cu resurse computaționale minime. Consumul de energie în cadrul AI provine din sarcini precum manipularea datelor și formarea de modele. Complexitatea algoritmică redusă este esențială pentru reducerea consumului de energie, deoarece necesită mai puține operațiuni de calcul.

4.5. Limitările procesului decizional algoritmic

Algoritmul decizional a apărut ca instrument de automatizare și optimizare a proceselor în diverse domenii. Cu toate acestea, deși promițător, algoritmul de luare a deciziilor ascunde limitări inerente care necesită o examinare meticuloasă. De la finanțe la asistență medicală, de la educație la justiție penală, algoritmii au fost implementați, promițând eficiență, precizie și obiectivitate de neegalat. Cu toate acestea, există o rețea complexă de constrângeri care împiedică eficacitatea nelimitată a procesului decizional algoritmic (Yeung, 2018).

Prejudecăți și discriminare: Una dintre constrângerile predominante și dăunătoare care afectează procesul decizional algoritmic este înclinația spre părtinire și discriminare. În ciuda eforturilor de imparțialitate, algoritmii moștenesc frecvent prejudecățile prezente în datele de formare pe care le utilizează. Prezența nedreptăților istorice și a prejudecăților societale în cadrul acestor date poate susține și intensifica disparitățile, ducând la rezultate discriminatorii. În plus, opacitatea metodologiilor algoritmice prezintă dificultăți în identificarea și atenuarea prejudecăților, sporind astfel temerile privind corectitudinea și egalitatea în procesele decizionale. Aceasta poate fi o oportunitate pozitivă de a detecta prejudecățile și discriminarea neobservate anterior de observatorii umani. Prin analiza unor seturi mari de date cu ajutorul unor algoritmi avansați, pot fi identificate tipare și disparități subtile, permițând luarea de măsuri corective pentru rectificarea prejudecăților sistemice. Caracteristicile de transparență și trasabilitate ale anumitor algoritmi permit o examinare atentă a procesului decizional, permițând părților interesate să abordeze prejudecățile la baza lor.

Lipsa înțelegerii contextuale: Procesul decizional algoritmic operează într-un domeniu naiv din punct de vedere contextual, depinzând exclusiv de măsurători cantitative și reguli predefinite. Această constrângere este deosebit de pronunțată în medii complexe și dinamice în care subtilitățile contextuale fac parte integrantă din procesul decizional. Raționamentul uman, sporit de cunoștințele contextuale și de competența în domeniu, depășește în mod frecvent capacitățile algoritmice de a discerne subtilitățile nuanțate și de a lua

decizii în cunoștință de cauză. Ca urmare, sistemele algoritmice pot manifesta rigiditate și insuficiență în adaptarea la contextele în evoluție, compromițând astfel eficacitatea și aplicabilitatea deciziilor lor.

Consecințe neprevăzute și dileme etice: Natura deterministă a algoritmilor îi face susceptibili la ramificații neprevăzute și dileme etice, care decurg din simplificarea excesivă și reducționismul inerente proceselor decizionale algoritmice. Rezultatele neprevăzute, efectele în cascadă și repercusiunile neintenționate pot apărea din cauza incapacității algoritmilor de a lua în considerare complexitatea și complexitatea scenariilor din lumea reală. În plus, dilemele etice privind încălcarea vieții private, erodarea autonomiei și responsabilitatea morală subliniază necesitatea unei deliberări și a unei supravegheri meticuloase în punerea în aplicare a sistemelor algoritmice.

4.6. Cum se pot identifica sursele potențiale de părtinire și erori ale sistemelor algoritmice? O problemă crucială pentru asigurarea corectitudinii, transparenței și responsabilității

Una dintre principalele surse de părtinire în procesul decizional algoritmic provine din datele utilizate pentru antrenarea și testarea algoritmilor. Prejudecățile prezente în date se pot propaga de-a lungul procesului decizional, ducând la rezultate distorsionate. Pentru a identifica potențialele prejudecăți din date, este necesar să se efectueze o analiză aprofundată a procesului de colectare a datelor pentru a evalua orice prejudecăți sistematice sau subreprezentare a anumitor grupuri, să se examineze caracteristicile demografice ale setului de date pentru a identifica disparitățile sau dezechilibrele care pot conduce la rezultate părtinitoare și să se evalueze etapele de pretratare a datelor, cum ar fi curățarea datelor și selectarea caracteristicilor, pentru a identifica orice transformări sau distorsiuni neintenționate ale datelor.

Pentru a identifica potențialele prejudecăți și erori în proiectarea și punerea în aplicare a algoritmilor, ar trebui să se analizeze modelele și metodologiile algoritmice, să se evalueze ipotezele și constrângerile care stau la baza algoritmului pentru a determina dacă acestea se aliniază normelor etice și societale, să se examineze procesul de formare pentru a identifica sursele potențiale de supraadaptare sau subadaptare, care pot conduce la predicții inexacte sau nedrepte, să se investigheze alegerea caracteristicilor și variabilelor utilizate de algoritm pentru a se asigura că acestea sunt relevante, nediscriminatorii și reprezentative pentru populație. Supraadaptarea și subadaptarea se referă la două probleme comune care apar la formarea modelelor de învățare automată. Supraadaptarea apare atunci când un model învață să capteze zgomotul sau fluctuațiile aleatorii din datele de formare, mai degrabă decât modelele sau relațiile subiacente. Ca urmare, modelul funcționează bine pe baza datelor de formare, dar nu reușește să se generalizeze

la date noi, nevăzute. Pe de altă parte, subadaptarea apare atunci când un model este prea simplist pentru a surprinde structura de bază a datelor, ceea ce duce la performanțe slabe atât pe datele de formare, cât și pe datele noi.

Pentru a identifica potențialele prejudecăți și erori în evaluare și validare, ar trebui testate metodologii de măsurare a parametrilor de performanță ai algoritmului în diferite grupuri demografice pentru a detecta disparitățile sau inechitățile, de efectuare a analizelor de sensibilitate pentru a evalua robustețea algoritmului la variațiile datelor de intrare sau ale parametrilor modelului și de solicitare a feedback-ului din partea părților interesate, inclusiv a comunităților afectate și a experților în domeniu, pentru a identifica potențiale prejudecăți sau preocupări etice trecute cu vederea în timpul dezvoltării.

De asemenea, este necesară o monitorizare și un audit continuu pentru a detecta și atenua prejudecățile, erorile sau consecințele neintenționate.

4.7. Abordarea și atenuarea riscurilor și prejudecăților algoritmice

Odată identificate, riscurile și prejudecățile algoritmice trebuie abordate printr-o abordare multifațetată care să cuprindă considerații tehnice, procedurale și etice. Intervențiile tehnice pot implica rafinarea algoritmilor prin tehnici precum debișarea algoritmilor, diversificarea datelor de formare și încorporarea parametrilor de corectitudine în evaluarea modelului. Intervențiile procedurale presupun punerea în aplicare a unor protocoale de validare solide, a unor măsuri de transparență și a unor mecanisme de responsabilitate pentru a se asigura că deciziile algoritmice sunt analizate și validate în mod eficient. Intervențiile etice se referă la promovarea unei culturi a conștientizării și responsabilității etice în rândul dezvoltatorilor, responsabililor politici și utilizatorilor finali, subliniind implicațiile etice ale procesului decizional algoritmic și promovând orientări și standarde etice.

Atenuarea riscurilor și a prejudecăților algoritmice necesită monitorizare, evaluare și adaptare continuă la provocări și contexte în continuă evoluție. Monitorizarea continuă a performanței și a impactului algoritmic permite părților interesate să detecteze și să abordeze în timp util prejudecățile și riscurile emergente. Mecanismele de evaluare, cum ar fi auditurile privind prejudecățile, analizele de sensibilitate și evaluările de impact, oferă o perspectivă asupra eficacității strategiilor de atenuare a riscurilor și informează cu privire la îmbunătățirile iterative. În plus, încurajarea colaborării și a angajamentului interdisciplinar facilitează schimbul de informații și de bune practici între domenii, îmbogățind efortul colectiv de reducere la minimum a riscurilor și a prejudecăților algoritmice.

Sfaturi și recomandări pentru profesori pentru a se implica în complexitatea algoritmică

1. Introducere conceptuală
Începeți cu o explicație directă a ceea ce este complexitatea algoritmică, folosind analogii legate de activitățile cotidiene care necesită planificare și resurse, cum ar fi organizarea unui eveniment sau planificarea unei călătorii, pentru a ilustra conceptele de eficiență și gestionare a resurselor.

2. Demonstrații vizuale
Utilizați suporturi vizuale, cum ar fi diagrame și grafice, pentru a demonstra modul în care diferiți algoritmi funcționează în diferite condiții. Acest lucru ajută participanții să înțeleagă vizual modul în care complexitatea afectează performanța.

3. Activități practice
Încorporați exerciții practice în care participanții pot experimenta diferiți algoritmi pentru a rezolva probleme specifice. Acestea ar putea implica exerciții simple de codare sau utilizarea unor instrumente software care simulează performanța algoritmilor.

4. Discutarea aplicațiilor din lumea reală
Corelați discuția cu aplicații din lumea reală în care complexitatea algoritmică joacă un rol esențial, cum ar fi prelucrarea datelor în marile companii de tehnologie sau găsirea rutelor în sistemele de navigație GPS, pentru a sublinia importanța acesteia.

5. Încurajarea învățării colaborative
Facilitați discuțiile de grup și sesiunile de rezolvare a problemelor în care cursanții pot dezbate cei mai buni algoritmi de utilizat în diferite scenarii, promovând o înțelegere mai profundă prin colaborare.

Activități de încălzire

1. Provocare de sortare
Organizați o activitate practică în care participanții sortează manual obiecte (cum ar fi cărți sau carduri) folosind diferite metode pentru a imita algoritmii de sortare. Discutați cum se raportează fiecare metodă la complexitatea temporală și spațială a unui algoritm.

2. Complexitatea algoritmului Joc de potrivire
Creați un joc de cărți în care participanții potrivesc diferiți algoritmi cu notațiile Big O ale acestora și cu exemple de cazuri de utilizare optime. Această abordare interactivă ajută la solidificarea înțelegerii conceptelor teoretice prin joc.

3. Brainstorming privind eficiența

Rugați participanții să enumere sarcinile zilnice pe care le îndeplinesc și care ar putea fi optimizate cu ajutorul algoritmilor (cum ar fi sortarea e-mailurilor sau organizarea fișierelor). Discutați despre modul în care înțelegerea complexității algoritmice poate duce la îmbunătățirea acestor sarcini.

Sfaturi și recomandări pentru evaluare

1. Verificări ale conceptului
Utilizați frecvent teste scurte pentru a evalua înțelegerea conceptelor cheie, cum ar fi notația Big O și diferențele dintre complexitatea temporală și cea spațială. Acest lucru ajută la asigurarea faptului că participanții înțeleg elementele de bază înainte de a trece la subiecte mai complexe.

2. Teste practice de codificare
Dacă este cazul, includeți teste practice de codificare în care participanții trebuie să scrie sau să identifice cei mai eficienți algoritmi pentru anumite probleme. Acest lucru le testează capacitatea de a aplica cunoștințele teoretice în scenarii practice.

3. Evaluarea pe bază de proiect
Atribuiți un mic proiect în care participanții trebuie să analizeze performanța unui sistem sau a unui software și să sugereze îmbunătățiri bazate pe complexitatea algoritmică. Acest lucru le evaluează capacitatea de a-și aplica cunoștințele la sisteme din lumea reală.

4. Prezentări de grup
Rugați participanții să lucreze în grupuri pentru a pregăti prezentări privind modul în care complexitatea algoritmică afectează diferite sectoare, cum ar fi finanțele, asistența medicală sau tehnologia. Acest lucru le evaluează înțelegerea și capacitatea de a comunica informații complexe.

5. Eseuri de reflecție
Rugați participanții să scrie eseuri de reflecție cu privire la modul în care înțelegerea complexității algoritmice le poate influența munca sau studiile. Acest lucru ajută la evaluarea capacității lor de a integra și reflecta asupra cunoștințelor dobândite.

Folosind aceste strategii, formatorii pot prezenta în mod eficient conținuturi complexe privind complexitatea algoritmică, asigurându-se în același timp că participanții sunt implicați, înțeleg materialul în întregime și își pot aplica cunoștințele în mod practic.

5. Referințe

Aksoy, T., & Gurol, B. (2021). Inteligența artificială în tehnicile și tehnologiile de audit asistate de calculator (CAATTs) și o propunere de aplicare pentru auditori. În *Auditing Ecosystem and Strategic Accounting in the Digital Era: Abordări globale și noi oportunități* (pp. 361-384). Cham: Springer International Publishing.

Bentley, P. (1999). *Evolutionary design by computers*. Morgan Kaufmann.

Bohr, A., & Memarzadeh, K. (2020). Creșterea inteligenței artificiale în aplicațiile din domeniul sănătății. În *Artificial Intelligence in healthcare* (pp. 25-60). Academic Press.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. MIT Press.

Davenport, T. H., & Mittal, N. (2023). *All-in on AI: How smart companies win big with artificial intelligence.* Harvard Business Press.

Yeung, K. (2018). Reglementarea algoritmică: O interogare critică. *Regulation & governance, 12*(4), 505-523. https://doi.org/10.1111/rego.12158

Donovan, J., Caplan, R., Matthews, J. și Hanson, L. (2018). *Responsabilitatea algoritmică: O introducere.* https://apo.org.au/node/142131

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Inteligența artificială (AI): Perspective multidisciplinare privind provocările emergente, oportunitățile și agenda pentru cercetare, practică și politică. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparență în care poți avea încredere: Cerințe de transparență pentru inteligența artificială între normele juridice și preocupările contextuale. *Big Data & Society, 6*(1), 2053951719860542. https://doi.org/10.1177/2053951719860542

Karp, R.M. (2010). Reducibilitatea între problemele combinatorii. În: Jünger, M., et al. *50 Years of Integer Programming 1958-2008*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68279-0_8

Knuth, D. E. (1997). *The Art of Computer Programming: Algoritmi fundamentali, volumul 1.* Addison-Wesley Professional.

Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2022). O revizuire a literaturii privind provocările și oportunitățile tranziției de la industria 4.0 la societatea 5.0. Energies, 15(17), 6276. https://doi.org/10.3390/en15176276

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Revoluția platformelor: Cum piețele în rețea transformă economia și cum să le faci să lucreze pentru tine.* WW Norton & Company.

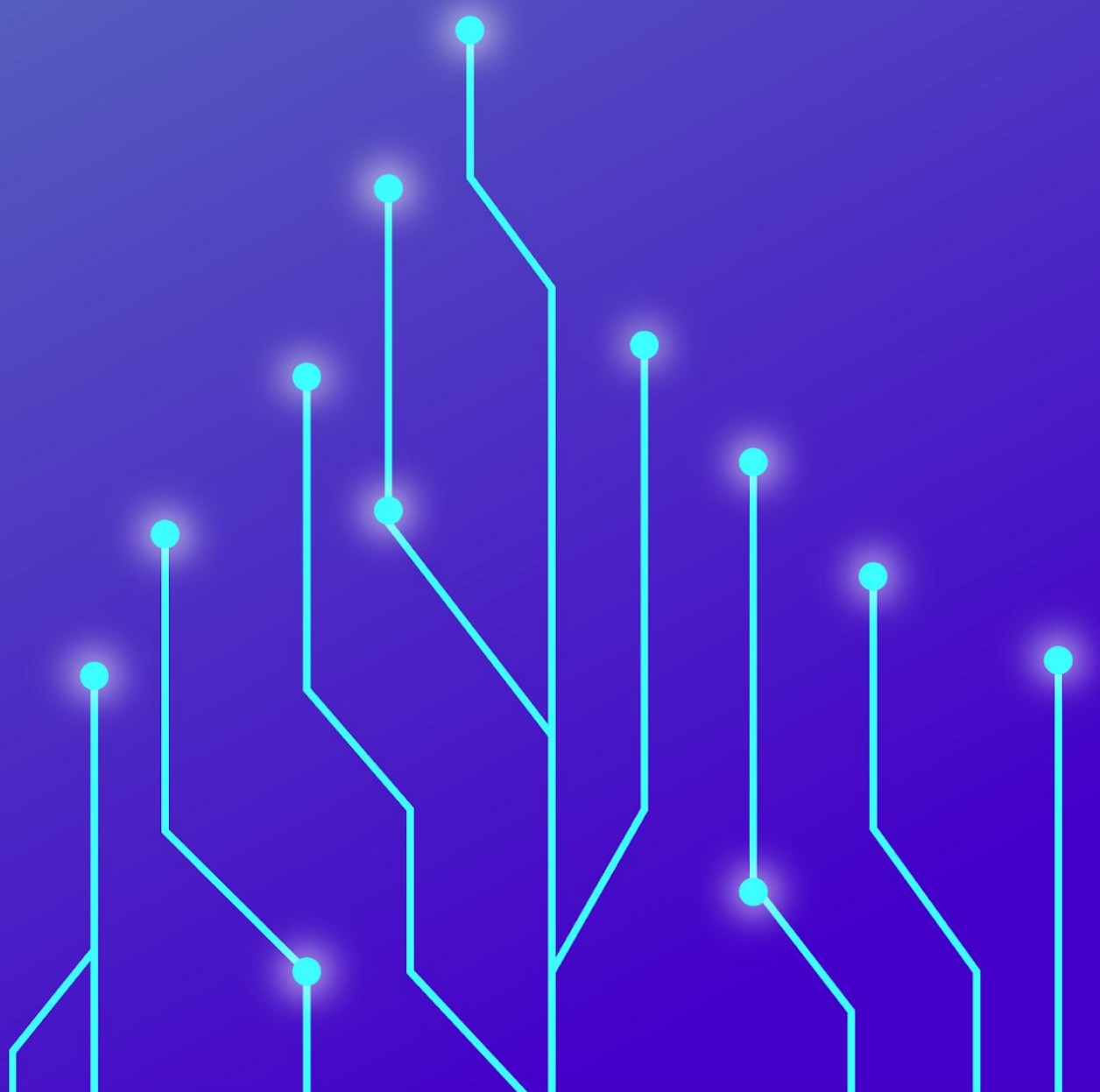
Patel, K. (2024). Reflecții etice privind IA centrată pe date: echilibrarea beneficiilor și riscurilor. *International Journal of Artificial Intelligence Research and Development, 2*(1), 1-17. https://orcid.org/0009-0005-9197-2765

Qin, X. (2012). Utilizarea datelor mari: următoarea generație de algoritmi de tranzacționare. În *Artificial Intelligence and Computational Intelligence: 4th International Conference*, AICI 2012, Chengdu, China, 26-28 octombrie 2012. Proceedings 4 (pp. 34-41). Springer Berlin Heidelberg.

Reisdorf, B. C., & Blank, G. (2021). Alfabetizarea algoritmică și încrederea în platformă. În *Handbook of digital inequality* (pp. 341-357). Edward Elgar Publishing.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M. și Floridi, L. (2021). Etica algoritmilor: Probleme-cheie și soluții. În: Floridi, L. (eds) Ethics, Governance, and Policies in Artificial Intelligence. Philosophical Studies Series, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-030-81907-1_8

CU4 | Corectitudinea și părtinirea datelor



Index

1. Introducere

Sistemele **de inteligență artificială** (AI) nu sunt doar concepte abstracte, ci fac parte din ce în ce mai mult din viața noastră de zi cu zi. Ele influențează deciziile în sectoare precum sănătatea și educația, finanțele și securitatea. Cu toate acestea, pe măsură ce influența AI crește, crește și potențialul acestor sisteme de a perpetua prejudecățile sociale existente sau de a introduce altele noi. Înțelegerea conceptelor de corectitudine și părtinire în IA nu este doar un exercițiu academic, ci este esențială pentru dezvoltarea unor sisteme juste și echitabile care au un impact asupra vieții noastre.

Acest CU privind corectitudinea și prejudecățile AI are ca scop înțelegerea problemei și împuternicirea dumneavoastră pentru a face parte din soluție. Acesta încearcă să ilumineze modul în care prejudecățile pot apărea involuntar în algoritmii AI, impactul potențial al acestor prejudecăți și strategiile de atenuare a acestora. Cursul explorează, de asemenea, provocările etice, societale și tehnice ale creării unor sisteme AI corecte, subliniind rolul dvs. în practicile etice de dezvoltare a AI. La finalizarea acestui CU, veți fi capabil să:

Identificarea și înțelegerea tipurilor de

Să recunoască diferite tipuri de prejudecăți care pot apărea în sistemele AI, inclusiv prejudecăți legate de date, prejudecăți algoritmice și prejudecăți legate de rezultate, și să înțeleagă mecanismele prin care se manifestă aceste

Măsurarea și cuantificarea prejudecăților:

Înțelegerea instrumentelor și metodologiilor de evaluare și măsurare a impactului prejudecăților în cadrul sistemelor AI, utilizând o serie de metrici de corectitudine

Elaborarea de strategii de atenuare

Dobândirea de cunoștințe practice privind modul de punere în aplicare a strategiilor de abordare și reducere a prejudecăților în sistemele AI, inclusiv diverse practici de colectare a datelor, ajustări algoritmice și monitorizarea

Evaluarea corectitudinii sistemului AI

Ar trebui să fie capabili să evalueze critic sistemele de inteligență artificială din punct de vedere al corectitudinii pe mai multe dimensiuni, asigurându-se că aceste sisteme

Pregătiți-vă pentru provocările viitoare:

Să fie capabili să anticipeze și să răspundă provocărilor emergente din domeniul corectitudinii AI pe măsură ce tehnologia evoluează, asigurându-se că rămân adaptabili și

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sugestie: începeți cu o scurtă activitate în care participanții își împărtășesc percepțiile sau experiențele personale cu IA în viața lor de zi cu zi. Acest lucru îi poate ajuta pe participanți să facă legătura între conceptul abstract de IA și implicațiile sale practice.

2. Echitate și părtinire - analogia cu bomboanele

Sursa imaginii | Generat de DALL-E

Imaginează-ți că te afli într-o sală de clasă și că te uiți la un castron cu bomboanele tale preferate de pe biroul profesorului. Fiecare elev poate alege o bomboană, dar există o întorsătură: elevii care poartă tricouri roșii primesc două bomboane, în timp ce toți ceilalți primesc una. Acest lucru nu pare corect. Acest lucru se numește "părtinire" - atunci când cineva (sau ceva, cum ar fi un AI) oferă avantaje nedrepte anumitor persoane fără un motiv întemeiat.

Acum, imaginați-vă că, în schimb, profesorul le cere tuturor să rezolve un puzzle, iar cine reușește, indiferent de culoarea tricoului, primește o bomboană în plus. Acest lucru pare corect, deoarece toată lumea are aceeași șansă - este vorba despre puzzle, nu despre tricou. Aceasta este "corectitudinea", atunci când toată lumea are șanse egale și nimeni nu este favorizat din motive arbitrare.

La fel ca în acest scenariu cu bomboane, sistemele AI din lumea reală ar trebui să ofere tuturor o șansă echitabilă, fie că recomandă filme, aprobă împrumuturi sau interpretează muzică. Atunci când inteligența artificială este corectă, toată lumea rezolvă puzzle-ul și primește bomboana.

Când este părtinitor, este ca și cum unii oameni primesc mai multe bomboane doar din cauza culorii tricoului lor. Iar noi vrem să fim corecți!

Sugestie: folosiți această analogie în mod interactiv, cerând participanților să simuleze scenariul cu bomboane reale sau cu o activitate virtuală. Această abordare practică ar putea contribui la solidificarea conceptelor de corectitudine și părtinire.

3. Înțelegerea peisajului: Egalitatea și prejudecățile AI

Date Echitatea în inteligența artificială se referă la principiul conform căruia sistemele de inteligență artificială ar trebui să trateze toate persoanele în mod echitabil, fără rezultate nedrepte sau prejudiciabile bazate pe caracteristici inerente sau dobândite. Echitatea în inteligența artificială garantează faptul că deciziile luate de aceste sisteme nu favorizează sau dezavantajează un anumit grup de utilizatori în detrimentul altora. Acest lucru este deosebit de important în aplicațiile care afectează viața oamenilor, cum ar fi angajarea, aprobarea împrumuturilor, aplicarea legii și diagnosticarea medicală.

Între timp, prejudecățile în inteligența artificială se referă la erori sistematice în date sau algoritmi care conduc la rezultate inechitabile pentru anumite grupuri. Prejudecățile pot fi explicite, cum ar fi cele încorporate în mod intenționat într-un sistem, sau, mai des, implicite, care rezultă din date de formare nereprezentative sau incomplete sau din proiectarea algoritmică defectuoasă care avantajează sau dezavantajează în mod neintenționat anumite populații.

Sursa imaginii | Generat de DALL-E

Abordarea echității și a prejudecăților în dezvoltarea IA nu poate fi supraestimată. Sistemele de inteligență artificială care funcționează cu părtinire pot perpetua și amplifica inegalitățile sociale existente, conducând la cicluri de dezavantaj greu de întrerupt. În plus, sistemele părtinitoare erodează încrederea în tehnologie, putând bloca adoptarea și inovarea. Asigurarea echității în IA este o necesitate tehnică și un imperativ moral pentru a menține încrederea societății și a preveni daunele.

SFATURI ȘI RECOMANDĂRI PENTRU PROFESORI

Sugestie: utilizați studii de caz sau articole de știri care evidențiază probleme recente legate de prejudecățile din sistemele AI, urmate de discuții în grup. Acest lucru poate contextualiza importanța corectitudinii în IA.

Prejudecăți de gen în recrutarea forței de muncă AI

În 2018, a fost dezvăluit faptul că un sistem AI utilizat de Amazon pentru recrutare era părtinitor față de femei. Inteligența artificială învățase singură că candidații de sex masculin sunt de preferat, observând modelele din CV-urile trimise pe o perioadă de 10 ani, predominant de la bărbați, din cauza dezechilibrului de gen din industria tehnologiei. Sistemul a retrogradat CV-urile care includeau cuvântul "al femeilor", ca în "căpitanul clubului de șah al femeilor" sau "colegiul femeilor".

4. Prejudecăți în sistemele AI

Prejudecățile în inteligența artificială se referă la erorile sistematice care conduc la rezultate inechitabile pentru anumite grupuri, favorizând de obicei un grup demografic în detrimentul altora. Aceste prejudecăți pot proveni din diverse surse, cum ar fi datele utilizate pentru antrenarea sistemelor de inteligență artificială, algoritmii care prelucrează aceste date sau prejudecățile interpretative ale celor care concep și implementează aceste sisteme.

4.1. Apariția prejudecăților în sistemele AI

- **Prejudecarea datelor**: această prejudecată apare atunci când seturile de date utilizate pentru antrenarea algoritmilor AI nu reprezintă populația în general. Acest lucru se poate întâmpla din cauza inegalităților istorice sau pur și simplu pentru că metodele de colectare a datelor surprind o categorie demografică mai bine decât altele.

- **Prejudecățile algoritmului**: Uneori, algoritmii pot fi concepuți în moduri care favorizează în mod inerent anumite rezultate. Acest lucru se poate datora modului în care algoritmii interpretează datele sau obiectivelor stabilite în timpul dezvoltării AI.

- **Buclele de feedback**: Prejudecățile pot fi, de asemenea, perpetuate și amplificate prin bucle de feedback în care deciziile inițiale părtinitoare conduc la colectarea de date suplimentare care consolidează aceste prejudecăți, creând un ciclu de discriminare.

4.2. Consecințele prejudecăților în inteligența artificială

- **Impactul social**: Prejudecățile AI pot consolida inegalitățile sociale și stereotipurile existente, conducând la o societate în care tehnologiile digitale dezavantajează sistematic anumite grupuri. Această erodare a încrederii în tehnologiile IA poate avea implicații mai ample asupra adoptării și eficienței IA în diverse sectoare.

- **Impactul individual**: La nivel personal, prejudecățile din IA pot duce la un tratament inechitabil în domenii critice precum angajarea, aplicarea legii, asistența medicală și serviciile financiare. Persoanele se pot confrunta cu rezultate nedrepte din cauza proceselor decizionale eronate influențate de sisteme AI părtinitoare.

4.3. Abordarea prejudecăților AI

Adoptarea unor strategii cuprinzătoare pe tot parcursul ciclului de viață al dezvoltării AI pentru a atenua prejudecățile AI este esențială. Acestea includ:

- Diversificarea surselor de date pentru a asigura o reprezentare mai largă în seturile de date de formare.
- Elaborarea și punerea în aplicare a auditurilor algoritmice pentru identificarea și corectarea prejudecăților.
- Crearea de sisteme de monitorizare continuă pentru a detecta și a aborda prejudecățile emergente în sistemele AI implementate.

4.4. Prezentare generală a tipurilor de prejudecăți

Înțelegerea diferitelor tipuri de prejudecăți care se pot manifesta în sistemele de inteligență artificială este esențială pentru dezvoltatorii, factorii de decizie și utilizatorii care doresc să creeze și să interacționeze cu tehnologii corecte și echitabile. Această secțiune va explora diferitele forme de părtinire, fiecare însoțită de exemple din lumea reală care ilustrează impactul acestora.

1. Prejudecățile de selecție: Prejudecățile de selecție apar atunci când datele utilizate pentru formarea unui sistem AI nu reprezintă populația sau scenariul țintă, ceea ce duce la rezultate distorsionate. De exemplu, un model de inteligență artificială antrenat să recunoască trăsături faciale utilizând un set de date compus în principal din persoane tinere poate să nu reușească să identifice

cu exactitate trăsături ale adulților în vârstă, deoarece antrenamentul modelului nu a inclus un eșantion reprezentativ al tuturor grupurilor de vârstă.

2. Prejudecata de confirmare: Prejudecata de confirmare în inteligența artificială se manifestă atunci când datele sau interpretarea acestora de către creatorii de modele sprijină în mod involuntar convingerile preexistente, ignorând dovezile contradictorii. O inteligență artificială utilizată la angajare, antrenată pe baza unor CV-uri predominant de un gen, poate continua să favorizeze candidații de acel gen, consolidând dezechilibrul existent în selecția posturilor.

3. Prejudecată de eșantionare: Prejudecata de eșantionare este un tip specific de prejudecată de selecție, în care eșantionul de date colectat pentru instruirea inteligenței artificiale nu reprezintă cu exactitate populația mai largă. Dacă un sistem de recunoaștere vocală este antrenat în principal pe baza datelor provenite de la vorbitori cu accent american, acesta ar putea avea dificultăți în a înțelege accente din alte regiuni, cum ar fi cel britanic sau australian, din cauza faptului că a fost antrenat pe un eșantion nereprezentativ.

4. Prejudecățile algoritmice: Prejudecățile algoritmice apar atunci când algoritmii care stau la baza sistemelor AI produc rezultate părtinitoare, adesea din cauza unor defecte inerente în proiectarea lor sau a obiectivelor stabilite în timpul creării lor. O AI de evaluare a creditelor care refuză în mod disproporționat acordarea de împrumuturi solicitanților din anumite cartiere din cauza datelor istorice care reflectă prejudecățile anterioare în materie de creditare este un exemplu de părtinire algoritmică.

5. Tendința de raportare: Tendința de raportare apare atunci când datele introduse într-un sistem AI sunt distorsionate de tendința de a raporta doar anumite tipuri de rezultate. De exemplu, un sistem AI de monitorizare a efectelor secundare ale medicamentelor care primește în principal rapoarte privind cazurile grave poate sugera în mod eronat că majoritatea pacienților prezintă reacții extreme, ignorând efectele secundare mai blânde, dar mai frecvente.

6. Abaterea de măsurare: Abaterea de măsurare implică erori în colectarea datelor care denaturează sistematic datele. Un AI de monitorizare a mediului care se bazează pe senzori mai puțin eficienți la temperaturi mai scăzute va avea o imagine distorsionată a nivelurilor de poluare, putând subraporta emisiile în timpul lunilor de iarnă.

5. Metode de identificare și măsurare a prejudecăților în IA

Înțelegerea și atenuarea prejudecăților din inteligența artificială sunt esențiale pentru crearea unor sisteme corecte și eficiente. Această secțiune detaliază metodologiile de identificare și măsurare a acestor prejudecăți, oferind o cale clară dezvoltatorilor de inteligență artificială și cercetătorilor în domeniul datelor pentru a se asigura că sistemele lor de inteligență artificială funcționează fără prejudecăți incorecte.

- **Auditul datelor** este un proces cuprinzător de revizuire în cadrul căruia seturile de date utilizate pentru antrenarea modelelor AI sunt examinate pentru a depista eventualele prejudecăți sau anomalii. Pentru a efectua un audit al datelor, ar trebui:
    - **Evaluați reprezentativitatea datelor**: Evaluați dacă setul de date reflectă cu acuratețe diversitatea populației pe care o va servi modelul. Aceasta include verificarea reprezentativității demografice și a acoperirii scenariilor.
    - **Identificați datele lipsă**: Căutați modele în datele lipsă - lacunele pot introduce prejudecăți dacă anumite subgrupuri sunt subreprezentate.
    - **Analizați consecvența etichetării**: Asigurați-vă că procesul de etichetare a datelor este consecvent și imparțial. Discrepanțele de etichetare pot duce la distorsiuni semnificative în rezultatele modelului.
    - **Analiza statistică**: Utilizați instrumente statistice pentru a detecta anomalii sau distribuții distorsionate ale datelor care ar putea indica date distorsionate.
- **Analiza disparităților** presupune compararea performanțelor modelelor între diferite grupuri demografice sau alte subgrupuri relevante pentru a identifica discrepanțele care ar putea indica prejudecăți. Etapele pentru efectuarea unei analize a disparităților includ:
    - **Definiți subgrupuri relevante:** Pe baza unor atribute precum vârsta, sexul, etnia sau alte caracteristici relevante pentru aplicarea modelului.
    - **Măsurați parametrii de performanță:** Calculați parametrii de performanță precum acuratețea, precizia, reamintirea și scorul F1 pentru fiecare subgrup.
    - **Compararea rezultatelor:** Analizați diferențele în ceea ce privește parametrii de performanță între grupuri. Discrepanțele semnificative pot indica prezența unor prejudecăți.
    - **Analiza cauzelor principale:** Dacă se constată disparități, investigați cauzele care stau la baza acestora, care pot fi legate de colectarea datelor, arhitectura modelului sau selectarea caracteristicilor.
- **Analiza sensibilității** testează robustețea unui model AI prin modificarea ușoară a datelor de intrare și observarea modului în care se modifică rezultatele. Acest lucru este deosebit de util pentru identificarea prejudecăților:
    - **Variați datele de intrare**: Introduceți mici modificări ale datelor de intrare care reflectă variații plauzibile în scenariile din lumea reală.
    - **Monitorizați fluctuațiile rezultatelor**: Observați dacă aceste modificări conduc la modificări disproporționate ale rezultatelor modelului.

- **Evaluarea impactului asupra grupurilor specifice**: Concentrați-vă asupra modului în care aceste modificări afectează rezultatele modelului pentru diferite grupuri demografice.
- **Implementați ajustările**: Utilizați rezultatele pentru a rafina preprocesarea datelor, ingineria caracteristicilor sau modelul pentru a reduce sensibilitățile nedorite.

6. Strategii pentru abordarea și atenuarea prejudecăților

Atenuarea prejudecăților în sistemele de inteligență artificială implică o abordare multidimensională în timpul fazelor de proiectare, dezvoltare și implementare. Următoarele strategii sunt esențiale pentru **reducerea riscului de rezultate părtinitoare**.

- **Reprezentare diversă** Asigurarea diversității în ceea ce privește datele și componența echipei este esențială pentru dezvoltarea unor sisteme AI imparțiale.
  - **Diversitatea datelor**: Colectați date dintr-o gamă largă de surse pentru a acoperi un spectru larg al populației. Aceasta include date demografice diferite, medii socio-economice și alte caracteristici relevante.
  - **Diversitatea echipei**: Echipele diverse aduc perspective diferite care pot identifica și atenua potențialele prejudecăți care ar putea să nu fie evidente pentru un grup mai omogen.
- **Algoritmi conștienți de prejudecăți** Dezvoltarea de algoritmi care sunt conștienți în mod inerent de prejudecăți și care se pot adapta la acestea reprezintă o abordare proactivă a atenuării prejudecăților.
  - **Încorporarea parametrilor de corectitudine**: Integrați considerente de echitate în funcția obiectiv a algoritmului, cum ar fi minimizarea disparității în ratele de eroare între grupuri.
  - **Formare adversarială**: Implementarea tehnicilor de formare adversarială în care modelele sunt formate simultan pentru a prezice corect și pentru a minimiza părtinirea.
- **Monitorizarea și evaluarea periodică** Monitorizarea continuă și evaluarea periodică a sistemelor de inteligență artificială sunt vitale pentru a se asigura că acestea rămân imparțiale în timp și se adaptează la noi date sau contexte.
  - **Stabilirea unor cadre de monitorizare**: Creați sisteme care urmăresc în permanență performanța modelelor AI, concentrându-se în primul rând asupra parametrilor de corectitudine.

- **Revizuiri periodice**: Revizuirea și recalibrarea periodică a modelelor pe baza rezultatelor monitorizării continue, adaptându-le la schimbările în modelele de date sau în normele societale.
- **Feedback-ul părților** interesate: Implicați părțile interesate, inclusiv utilizatorii finali, pentru a colecta feedback cu privire la performanța AI și la corectitudinea percepută, ceea ce poate oferi informații care nu sunt surprinse în evaluările cantitative.

7. Înțelegerea parametrilor de corectitudine

Criteriile de măsurare a echității sunt esențiale pentru evaluarea modului în care sistemele IA tratează persoanele sau grupurile din medii diferite. Acestea oferă un cadru de măsurare și de asigurare a faptului că sistemele IA nu perpetuează sau exacerbează prejudecățile sociale existente. Mai jos sunt prezentate câteva dintre **metricile esențiale ale echității** care ar trebui luate în considerare:

- **Paritatea statistică (paritatea demografică)**: Acest parametru evaluează dacă proporția de rezultate pozitive este aceeași în cadrul diferitelor grupuri demografice (de exemplu, sex, rasă). Este esențială pentru aplicațiile în care obiectivul este de a asigura egalitatea rezultatelor, indiferent de caracteristicile de intrare legate de identitatea grupului.
- **Șanse egalizate**: Această măsură presupune ca rata pozitivelor adevărate (sensibilitate) și rata falselor pozitive (specificitate) să fie aceleași pentru toate grupurile. Aceasta este utilizată în principal în situațiile în care consecințele unei erori afectează toate grupurile în mod egal, cum ar fi în justiția penală sau la angajare.
- **Disparitatea demografică condiționată**: Acest parametru evaluează diferențele în rezultatele predictive condiționate de atribute legitime. Aceasta măsoară disparitățile care nu sunt justificate de atributele relevante și este adecvată pentru scenariile în care se așteaptă anumite diferențe între grupuri din cauza acestor atribute.
- **Echitate prin conștientizare**: Această abordare implică încorporarea conștientizării atributelor sensibile (cum ar fi rasa sau sexul) direct în procesul decizional pentru a asigura decizii echitabile. Această măsurătoare pledează pentru algoritmi care compensează nedreptățile istorice sau prejudecățile societale care ar putea afecta echitatea.
- **Echitate individuală:** Această metrică afirmă că persoanele similare ar trebui să primească previziuni similare, indiferent de apartenența lor la un grup. Acest concept este deosebit de relevant în cazul serviciilor personalizate, cum ar fi recomandările de locuri de muncă, unde este esențială consecvența în tratamentul aplicat profilurilor similare ale utilizatorilor.

- **Corectitudine contrafactuală**: Conform acestui criteriu, o decizie este considerată echitabilă dacă ar fi fost luată într-o lume contrafactuală în care individul aparținea unui grup demografic diferit, dar era identic în rest. Această măsură este valoroasă în evaluarea corectitudinii la nivel individual și ajută la înțelegerea impactului anumitor atribute asupra deciziilor.

Selectarea parametrilor adecvați de măsurare a corectitudinii depinde de contextul și obiectivele specifice ale aplicației AI. Iată câteva linii directoare de luat în considerare:

1. **Cerințe specifice aplicației:** Alegerea metricii echității trebuie să se alinieze la obiectivele și cerințele legale ale domeniului de aplicare. De exemplu, egalitatea șanselor poate fi esențială pentru aplicațiile din domeniul justiției penale, în timp ce paritatea statistică poate fi mai potrivită pentru marketing sau publicitate.

2. **Considerații normative și etice:** Unele industrii pot avea cerințe de reglementare care dictează anumiți parametri de echitate. Considerentele etice, cum ar fi nevoia de a repara nedreptățile istorice, pot, de asemenea, să orienteze alegerea parametrilor de măsurare a echității.

3. **Evaluarea impactului:** Este esențial să se ia în considerare potențialul impact social al sistemului de inteligență artificială și să se aleagă parametrii de măsurare care minimizează prejudecățile dăunătoare. De exemplu, corectitudinea contrafactuală poate fi esențială în domenii precum asistența medicală, unde este necesar un tratament individualizat.

4. **Fezabilitatea tehnică:** În funcție de complexitatea sistemului de inteligență artificială și de disponibilitatea datelor, unii parametri pot fi mai dificil de implementat decât alții. Limitările tehnice ar trebui luate în considerare pentru a se asigura că măsurile de echitate alese pot fi aplicate cu acuratețe și consecvență.

8. Echitate în colectarea și preprocesarea datelor

Asigurarea faptului că practicile de colectare a datelor sunt diverse și imparțiale este fundamentală pentru crearea unor sisteme de IA echitabile.

- **Eșantionare reprezentativă**: Asigurați-vă că procesul de colectare a datelor include eșantioane din toate segmentele de populație pe care le va servi sistemul IA. Acest lucru poate implica eșantionarea stratificată pentru a include grupurile subreprezentate.

- **Colectarea continuă a datelor**: Actualizați și extindeți periodic colecțiile de date pentru a reflecta tendințele noi și emergente din cadrul populației, deoarece seturile de date statice pot deveni rapid depășite.

8.1. Tehnici de preprocesare

Asigurarea echității în colectarea și preprocesarea datelor este esențială pentru a preveni infiltrarea prejudecăților în sistemele de inteligență artificială. Mai jos sunt prezentate câteva practici care pot contribui la realizarea unui sistem de inteligență artificială mai echilibrat și mai echitabil prin acordarea unei atenții deosebite etapelor inițiale ale dezvoltării inteligenței artificiale.

- **Gestionarea datelor lipsă**: Analizați modelele de date lipsă, deoarece acestea pot introduce prejudecăți. Tehnici precum imputarea trebuie aplicate cu atenție pentru a evita introducerea de erori suplimentare.
- **Selectarea caracteristicilor**: Evaluarea critică a caracteristicilor care sunt incluse în procesul de formare a modelului. Caracteristicile corelate cu atribute sensibile trebuie tratate cu prudență pentru a evita discriminarea indirectă.
- **Normalizare și standardizare**: Aplicați normalizarea sau standardizarea pentru a vă asigura că modelul nu ponderează în mod inerent anumite caracteristici mai mult doar din cauza scării lor numerice.
- **Diverse surse de date:** Asigurarea că datele colectate cuprind un spectru larg de caracteristici demografice ale populației este esențială pentru a preveni prejudecățile din modelele AI. Scopul este de a colecta date care să reflecte varietatea și diversitatea lumii reale în care va funcționa sistemul AI. Acest lucru implică angajarea cu diverse grupuri demografice și utilizarea mai multor puncte de colectare. De exemplu, atunci când se dezvoltă o inteligență artificială în domeniul sănătății, ar trebui colectate date de la spitale din regiuni diferite, inclusiv din zone rurale și urbane, pentru a surprinde diverse profiluri de sănătate și impactul mediului asupra bolilor.
- **Eșantionarea incluzivă**: Eșantionarea incluzivă urmărește să asigure că toate segmentele populației sunt reprezentate în setul de date de formare. Obiectivul este de a preveni excluderea grupurilor minoritare sau subreprezentate a căror absență ar putea conduce la predicții părtinitoare ale IA. Se utilizează tehnici precum eșantionarea stratificată, în care populația este împărțită în subgrupuri (straturi) care reflectă caracteristici cheie (de exemplu, etnie, vârstă, sex), iar eșantioanele sunt selectate aleatoriu din fiecare strat pentru a asigura incluziunea în setul de date.
- **Curățarea datelor**: Curățarea datelor implică eliminarea inexactităților și inconsecvențelor care pot distorsiona rezultatele modelului AI. Obiectivul este de a se asigura că datele sunt de calitate ridicată și nu conțin erori care ar putea influența deciziile modelului. Acest proces include corectarea sau eliminarea valorilor aberante, completarea valorilor lipsă utilizând strategii de imputare adecvate care nu denaturează datele și asigurarea faptului că toate datele introduse sunt coerente și formatate corect.

- **Monitorizarea periodică**: Monitorizarea datelor și a performanței modelului este esențială pentru detectarea și abordarea oricăror prejudecăți care pot apărea în timp. Scopul este de a menține o vigilență continuă pentru a se asigura că sistemul AI continuă să funcționeze corect pe măsură ce se colectează date noi și se schimbă condițiile. Crearea de sisteme automate de monitorizare care să evalueze în permanență rezultatele AI în ceea ce privește parametrii de corectitudine și să alerteze dezvoltatorii în cazul în care sunt detectate prejudecăți. Acest sistem ar trebui să evalueze în mod regulat datele care intră în AI pentru a verifica calitatea datelor sau modificările de reprezentare.

- **Ingineria corectă a caracteristicilor:** Ingineria corectă a caracteristicilor implică selectarea și transformarea caracteristicilor pentru a minimiza prejudecățile, păstrând în același timp utilitatea datelor. Obiectivul este de a se asigura că intrările modelului nu dezavantajează în mod inerent niciun grup. Tehnicile includ analiza corelației caracteristicilor cu atribute sensibile și fie modificarea, fie eliminarea caracteristicilor care ar putea conduce la rezultate părtinitoare. Tehnicile de reducere a dimensionalității pot, de asemenea, să identifice și să elimine caracteristicile redundante sau irelevante care pot genera prejudecăți.

- **Transparență și documentație:** Menținerea transparenței prin documentarea detaliată a proceselor de colectare a datelor, de preprocesare și de selecție a caracteristicilor. Obiectivul este de a furniza înregistrări clare care pot fi auditate pentru a verifica conformitatea cu standardele de corectitudine. Păstrarea înregistrărilor detaliate ale deciziilor de prelucrare a datelor, ale metodologiilor utilizate pentru curățarea datelor și ale raționamentelor care stau la baza selecției caracteristicilor. Aceste documente ar trebui să fie ușor accesibile părților interesate și organismelor de reglementare pentru revizuirea și evaluarea corectitudinii sistemului IA.

- **Detectarea prejudecăților**: Detectarea prejudecăților implică identificarea potențialelor prejudecăți din setul de date înainte ca acesta să fie utilizat pentru instruirea modelelor AI. Obiectivul este de a corecta preventiv orice prejudecăți care afectează corectitudinea modelului. Utilizarea instrumentelor de analiză statistică și de vizualizare pentru a examina distribuția datelor între diferite grupuri demografice. Acest lucru poate implica software specializat sau algoritmi concepuți pentru a evidenția zonele în care datele pot să nu fie reprezentative sau în care rezultatele pot afecta în mod disproporționat anumite grupuri.

Corectitudinea unui sistem de inteligență artificială începe cu modul în care datele sunt colectate, prelucrate și pregătite înainte ca acestea să ajungă în etapa de formare algoritmică. Prin punerea în aplicare a unor practici riguroase de manipulare și preprocesare a datelor, dezvoltatorii pot reduce semnificativ riscul de părtinire a sistemelor AI. Aceste practici nu numai că sporesc corectitudinea și

fiabilitatea aplicațiilor AI, dar creează și încredere în rândul utilizatorilor și al părților interesate în medii diverse și reglementate.

9. Tendințe emergente în materie de corectitudine în IA

Pe măsură ce tehnologiile IA evoluează, acestea se confruntă cu seturi de date complexe și medii de reglementare dinamice. Noile tehnologii, cum ar fi inteligența artificială explicabilă (XAI) și învățarea federată, oferă oportunități de îmbunătățire a echității inteligenței artificiale prin asigurarea unei mai mari transparențe a proceselor decizionale în materie de inteligență artificială și prin permiterea unei prelucrări descentralizate a datelor care poate contribui la protejarea vieții private a persoanelor. Provocările viitoare includ:

- **Interacțiuni complexe ale datelor**: Odată cu apariția volumului mare de date, înțelegerea interacțiunilor și corelațiilor din cadrul seturilor de date mari și complexe devine mai dificilă, dar esențială pentru asigurarea echității.
- **Reglementări adaptative**: Pe măsură ce gradul de conștientizare a publicului cu privire la prejudecățile AI crește, organismele de reglementare vor introduce probabil orientări mai stricte privind corectitudinea, solicitând sistemelor AI să fie adaptabile și transparente cu privire la măsurile lor de corectitudine.
- **Explicabilitate și interpretabilitate**: Explicabilitatea și interpretabilitatea în IA implică transparența și înțelegerea de către oameni a operațiunilor și deciziilor sistemelor IA. Obiectivul este de a se asigura că părțile interesate înțeleg modul în care sunt luate deciziile AI, ceea ce este esențial pentru încredere și responsabilitate. Această tendință implică dezvoltarea de metode și instrumente care permit vizualizarea și explicarea proceselor decizionale ale IA. Tehnici precum scorurile de importanță a caracteristicilor, arborii decizionali și metodele de analiză a modelelor oferă o perspectivă asupra funcționării modelelor complexe, cum ar fi rețelele neuronale profunde.
- **Atacuri și apărări de tip adversar**: Atacurile adverse implică manipularea intrărilor AI pentru a înșela sistemele AI în luarea unor decizii incorecte. Apărarea împotriva unor astfel de atacuri urmărește să facă sistemele AI mai robuste și mai sigure și să le protejeze de intrările rău intenționate care ar putea exploata prejudecățile sau punctele slabe. Implementarea mijloacelor de apărare împotriva atacurilor adversare include utilizarea unor tehnici precum instruirea adversară, în care modelul este expus la exemple adversare în timpul instruirii pentru a învăța să le reziste. Alte abordări implică evaluări regulate ale securității și actualizarea modelelor pentru a recunoaște și contracara noile strategii de atac.

- **Echitate în mai multe dimensiuni:** Echitatea pe mai multe dimensiuni se referă la abordarea prejudecăților care se intersectează cu diverși factori demografici și socioeconomici, inclusiv, dar fără a se limita la, rasă, sex, vârstă și handicap. Obiectivul este de a asigura o echitate cuprinzătoare care să țină seama de identitățile umane complexe și de interacțiunile acestora. Acest lucru implică dezvoltarea unor indicatori multidimensionali ai echității și utilizarea tehnicilor de analiză intersecțională a datelor pentru a evalua și a atenua prejudecățile la intersecția mai multor atribute.

- **Echitate în tehnologiile emergente**: Pe măsură ce inteligența artificială este integrată în tehnologii emergente precum vehiculele autonome, asistența medicală inteligentă și dispozitivele IoT, asigurarea echității în aceste aplicații devine esențială. Obiectivul este de a implementa echitatea de la zero în aceste noi tehnologii pentru a preveni prejudecățile care ar putea avea un impact societal larg. Acest lucru implică efectuarea de evaluări ale impactului și audituri ale prejudecăților pentru noile tehnologii și elaborarea de orientări privind echitatea specifice contextului și cazului de utilizare al fiecărei tehnologii.

- **Considerente etice și guvernanță**: Considerarea etică și guvernanța în domeniul IA implică stabilirea unor cadre și orientări care să garanteze că sistemele de IA sunt dezvoltate și implementate într-un mod care respectă standardele etice și valorile societății. Obiectivul este de a oferi un mecanism de reglementare și de supraveghere etică pentru a ghida dezvoltarea și utilizarea IA. Elaborarea de orientări cuprinzătoare privind etica IA, constituirea de comitete de etică și colaborarea cu factorii de decizie politică în vederea adoptării de reglementări care să impună corectitudinea, transparența și responsabilitatea în IA.

- **Proiectare centrată pe om**: Proiectarea centrată pe om creează sisteme AI care acordă prioritate bunăstării umane, valorilor și considerentelor etice. Obiectivul este de a se asigura că tehnologiile IA îmbunătățesc capacitățile umane și calitatea vieții fără a înlocui sau a diminua contribuția și supravegherea umană. Acest lucru implică implicarea diverselor grupuri de utilizatori în timpul procesului de proiectare, încorporarea unor bucle de feedback care să permită utilizatorilor să informeze dezvoltarea continuă a IA și asigurarea faptului că IA sprijină, și nu înlocuiește, procesul decizional uman.

- **Atenuarea prejudecăților algoritmice:** Atenuarea prejudecăților algoritmice implică identificarea și corectarea prejudecăților din algoritmii AI pentru a preveni rezultatele inechitabile. Obiectivul este de a rafina continuu modelele AI pentru a se asigura că acestea sunt cât mai obiective și imparțiale posibil. Aceasta include utilizarea tehnicilor de detectare și corectare a părtinirilor în timpul procesului de formare a modelului, cum ar fi reponderarea datelor de formare, ajustarea parametrilor modelului pentru a echilibra acuratețea și corectitudinea și utilizarea algoritmilor externi de atenuare a părtinirilor.

Aceste tendințe și probleme evidențiază natura dinamică a dezvoltării inteligenței artificiale și nevoia continuă de inovare și vigilență în asigurarea echității, pe măsură ce sistemele de inteligență artificială devin tot mai integrate în fiecare aspect al vieții umane.

10. Concluzii

Explorarea corectitudinii și a prejudecăților AI este esențială pe măsură ce integrăm inteligența artificială în mai multe aspecte ale vieții cotidiene și ale infrastructurii critice. De la definițiile și impactul diverselor prejudecăți în sistemele de inteligență artificială la strategiile de identificare, măsurare și atenuare a acestora, este evident că echitatea în inteligența artificială este o provocare cu multiple fațete care necesită o atenție cuprinzătoare și continuă.

Sursa imaginii | Generat de DALL-E

- **Înțelegerea tipurilor de prejudecăți și a impactului acestora:** Primul pas către atenuare este recunoașterea diferitelor tipuri de prejudecăți, cum ar fi prejudecățile de selecție, algoritmice și de confirmare, și înțelegerea potențialului lor de a perpetua inegalitatea. Exemple din lumea reală, cum ar fi inexactitățile de recunoaștere facială și algoritmii de angajare părtinitori, au ilustrat efectele dăunătoare ale prejudecăților AI necontrolate.

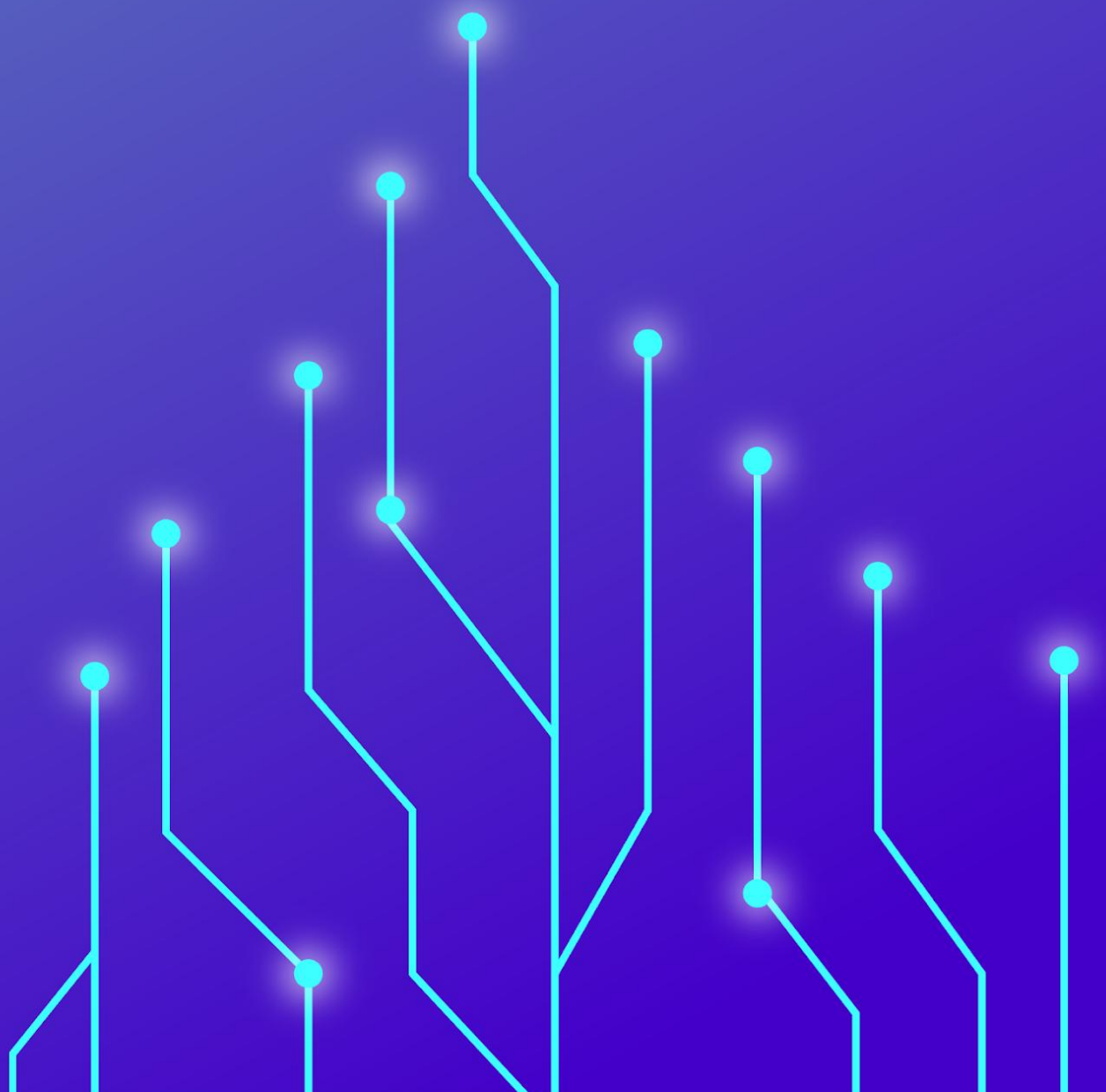
- **Strategii pentru corectitudine:** Metodele eficiente de asigurare a corectitudinii includ colectarea de date diverse, eșantionarea incluzivă și monitorizarea periodică a sistemelor de inteligență artificială. Aceste strategii contribuie la crearea unor modele de inteligență artificială care sunt reprezentative și corecte și reduc erorile sistematice care dezavantajează anumite grupuri.

- **Tendințe și provocări emergente:** Caracterul echitabil al IA evoluează, cu noi provocări precum asigurarea caracterului echitabil al tehnologiilor emergente și dezvoltarea de mijloace de apărare împotriva atacurilor adversarilor. De asemenea, se pune din ce în ce mai mult accentul pe explicabilitate și pe proiectarea centrată pe om, care sunt esențiale pentru crearea unor sisteme AI demne de încredere.

- **Considerații etice și guvernanță:** Importanța guvernanței etice în dezvoltarea IA nu poate fi supraestimată. Stabilirea unor orientări etice și a unor cadre de guvernanță solide va fi esențială pentru orientarea implementării responsabile a tehnologiilor IA.

- **Învățarea și adaptarea continue:** Pe măsură ce tehnologiile IA și normele societale evoluează, trebuie să evolueze și abordările noastre privind echitatea. Dezvoltatorii, factorii de decizie politică și părțile interesate au nevoie de educație continuă, vigilență și adaptare pentru a se asigura că sistemele de IA rămân echitabile și benefice pentru toți.

CU5 | Studii de caz și proiecte



INDEX

1. Introducere

În această unitate de curs, accentul este pus pe examinarea prejudecăților într-un context practic. Proiectul implică un scenariu ipotetic în care o companie internațională dezvoltă o soluție AI pentru recrutare. Acest proiect va fi abordat în timpul sesiunii finale de studiu interactiv cu studenții. Vă rugăm să facilitați o discuție animată pentru a sprijini învățarea acestora, încurajând studenții să exprime și să aplice ceea ce au învățat în sesiunile anterioare.

Prejudecata algoritmică este producerea de rezultate inechitabile sau discriminatorii de către sistemele algoritmice. Acest lucru se întâmplă atunci când un algoritm favorizează sau dezavantajează în mod sistematic anumite persoane sau grupuri pe baza unor caracteristici specifice, cum ar fi rasa, sexul sau statutul socioeconomic. Algoritmii de învățare automată operează pe diverse seturi de date, cum ar fi datele de antrenament, din care învață modele aplicabile altor persoane sau grupuri pentru a face predicții. Cu toate acestea, acești algoritmi ar putea trata în mod diferit indivizi sau elemente similare, ducând la replicarea sau amplificarea prejudecăților umane, în special a celor care afectează grupurile protejate (Friedman & Nissenbaum, 1996; Lange & Duarte, 2018; Lee et al., 2019). Înțelegerea cauzelor și a implicațiilor prejudecăților algoritmice este esențială pentru dezvoltarea unor soluții etice și demne de încredere în domeniul IA.

Implicațiile prejudecăților algoritmice

Prejudecățile algoritmice pot apărea sub diverse forme, afectând diferite grupuri în moduri distincte. Exemplele adunate de Lee, Resnick și Barton (2019) ilustrează modul în care prejudecățile pot proveni din surse multiple și pot afecta grupurile în mod neintenționat sau deliberat.

- Asociațiile de cuvinte pot reflecta prejudecăți. Cercetătorii de la Universitatea Princeton au constatat că numele europene erau percepute mai pozitiv decât cele afro-americane, iar cuvinte precum "femeie" și "fată" erau legate mai mult de arte decât de știință și matematică (Lee et al., 2019). După cum s-a demonstrat în exemplul inițial al acestui curs, asocieri precum "asistentă medicală" cu femeile și "inginer" cu bărbații sunt adesea modelate de date părtinitoare încorporate în algoritmii de căutare, reflectând stereotipurile societale. Acest lucru perpetuează prejudecățile în inteligența artificială și învățarea automată din cauza dependenței de conținutul online existent, reprezentând o provocare continuă semnificativă în domeniul inteligenței artificiale și al învățării automate.
- În instrumentele de recrutare online, Amazon a întrerupt utilizarea unui algoritm de recrutare din cauza prejudecăților de gen. Software-ul cu inteligență artificială penaliza CV-urile care conțineau cuvântul "feminin" și le retrograda pe cele de la facultățile pentru femei.

- Anunțurile online prezintă, de asemenea, prejudecăți. Latanya Sweeney, cercetător la Harvard, a descoperit că, în cazul căutărilor de nume afro-americane, era mai probabil să apară anunțuri pentru servicii de evidență a arestărilor, comparativ cu numele albilor.
- Tehnologia de recunoaștere facială poate fi și ea părtinitoare. Joy Buolamwini, cercetător la MIT, a constatat că sistemele comerciale nu reușesc adesea să recunoască tonurile de piele mai închise, deoarece sunt antrenate în principal pe fețe masculine și cu piele deschisă.
- Chiar și algoritmii din justiția penală pot fi părtinitori. Algoritmul COMPAS folosit de judecători pentru a prezice deciziile privind cauțiunea s-a dovedit a fi părtinitor împotriva afro-americanilor, după cum a raportat ProPublica.

2. Diferite tipuri de prejudecăți

Potrivit lui Friedman și Nissenbaum (1996), prejudecățile algoritmice pot fi clasificate în trei tipuri principale.

- În primul rând, **prejudecățile preexistente** se referă la prejudecățile care există deja în cadrul unui sistem, în mod intenționat sau neintenționat, înainte de crearea acestuia. Aceste prejudecăți pot intra într-un sistem prin eforturi deliberate sau acțiuni inconștiente, uneori chiar cu intenții bune.
- În al doilea rând, **prejudecățile tehnice** rezultă din constrângerile sau considerentele tehnice din timpul procesului de proiectare. Diverse aspecte ale proiectării pot contribui la prejudecățile tehnice.
- În al treilea rând, **prejudecățile emergente** apar în contextul utilizării în lumea reală cu utilizatori reali. Acest tip de prejudecată apare de obicei după finalizarea procesului de proiectare, adesea din cauza schimbărilor în cunoștințele societății, în populație sau în valorile culturale. (Friedman & Nissenbaum, 1996)

Exemple de părtinire în seturile de date de învățare automată

Modelele de învățare automată se bazează în mare măsură pe datele de instruire, însă implicarea umană în selectarea și pregătirea acestor date poate introduce diverse tipuri de prejudecăți. Un exemplu predominant este **tendința de raportare**, în care frecvența evenimentelor din setul de date nu reflectă evenimentele din lumea reală. Această părtinire poate face ca modelele să se confrunte cu dificultăți în gestionarea situațiilor obișnuite din cauza unui accent prea mare pus pe documentarea circumstanțelor neobișnuite. (Google pentru dezvoltatori, 2022)

Un alt tip este **prejudecata de atribuire de grup**, care se manifestă prin favorizarea propriului grup sau stereotipizarea altora pe baza caracteristicilor de grup. Prejudecățile implicite, care decurg din experiențe personale și modele mentale, complică și mai mult lucrurile, ducând uneori la **prejudecăți de**

confirmare sau **la prejudecăți ale experimentatorului** în timpul formării modelului. (Google pentru dezvoltatori, 2022)

Prejudecățile de automatizare, în care sistemele automatizate sunt favorizate în detrimentul judecății umane, indiferent de acuratețea lor, și **prejudecățile de selecție**, care rezultă din metodele nereprezentative de colectare a datelor, reprezintă, de asemenea, provocări semnificative în atenuarea prejudecăților în modelele de învățare automată. (Google pentru dezvoltatori, 2022)

Exemple de cazuri din lumea reală

După înțelegerea cadrului teoretic al prejudecăților algoritmice și a implicațiilor acestora, este acum oportun să explorăm exemple de cazuri reale care ilustrează impactul prejudecăților algoritmice. Primul caz analizează algoritmul dăunător al TikTok, în timp ce celălalt explorează preocupările etice legate de tehnologia de recunoaștere facială.

Algoritmul dăunător al TikTok

Efectele nocive ale algoritmului TikTok au ieșit la iveală în cadrul unei investigații revelatoare realizate de compania națională de media Yle din Finlanda. Yle a efectuat un studiu axat pe conținutul oferit unei fete fictive în vârstă de 13 ani, pe nume Ella, care suferea de depresie și probleme legate de imaginea corporală. (YLE, 2023)

În studiul lor, cercetătorii de date de la Yle au navigat pe TikTok folosind profilul Ellei timp de cinci ore. Inițial, Ella a întâlnit videoclipuri vesele, de la mâncare și animale de companie la glume. Cu toate acestea, în câteva minute, algoritmul a început să servească conținut legat de sănătatea mintală, inclusiv antidepresive. Pe măsură ce navigarea a continuat, tonul videoclipurilor a devenit mai întunecat, TikTok afișând conținut despre sinucidere, autovătămare și tulburări alimentare. Până la sfârșitul studiului, un procent uluitor de 95 % din conținutul prezentat Ellei a fost considerat dăunător, inclusiv glorificarea tulburărilor alimentare și sfaturi privind restricționarea consumului de alimente. (YLE, 2023)

Psihologul social Suvi Uski a evidențiat aspectul îngrijorător al comportamentului algoritmului, subliniind tendința acestuia de a prioritiza implicarea utilizatorilor în detrimentul prejudiciului potențial cauzat de conținutul afișat. Această dezvăluire subliniază implicațiile etice semnificative asociate cu recomandările algoritmice ale TikTok. (YLE, 2023)

Ceea ce a început ca o navigare inocentă a degenerat rapid într-o spirală îngrijorătoare de conținut dăunător, reflectând starea vulnerabilă a Ellei. Identificarea rapidă de către algoritm a interesului Ellei pentru comportamente depresive și nesănătoase este alarmantă în sine, dar și mai alarmant este rolul său în amplificarea acestor gânduri dăunătoare, în loc să ofere o contrabalansare. În

loc să o redirecționeze pe Ella către resurse pozitive și de susținere, algoritmul i-a întărit tendințele negative, exacerbându-i problemele de sănătate mintală. Acest lucru ridică întrebări critice cu privire la responsabilitatea etică a proiectanților de algoritmi și a operatorilor de platforme în protejarea bunăstării utilizatorilor. Cazul Ellei subliniază nevoia urgentă de mai multă transparență, responsabilitate și considerații etice în dezvoltarea și implementarea sistemelor algoritmice, în special a celor care exercită o influență semnificativă asupra comportamentului și sănătății mintale a utilizatorilor.

Colectarea neetică a datelor de recunoaștere facială

Tehnologia de recunoaștere facială a devenit din ce în ce mai răspândită în diverse sectoare, de la asistența medicală la aplicarea legii (Javaid, 2024). Cu toate acestea, adoptarea sa pe scară largă a ridicat preocupări etice semnificative, în special în ceea ce privește colectarea datelor și impactul acesteia asupra vieții private și libertăților civile. Un exemplu notabil de colectare neetică a datelor de recunoaștere facială a apărut într-un raport Washington Post din 7 iulie 2019. Raportul a dezvăluit că agenții federale precum FBI și ICE au accesat bazele de date privind permisele de conducere în scopul recunoașterii faciale, scanând fotografiile a milioane de americani fără consimțământul sau cunoștințele acestora (Harwell, 2019a).

Această dezvăluire alarmantă subliniază problema mai amplă a potențialului de părtinire și discriminare al tehnologiei de recunoaștere facială. Un studiu federal ulterior, prezentat de asemenea de Washington Post la 19 decembrie, a confirmat prevalența prejudecăților rasiale în cadrul sistemelor de recunoaștere facială. Studiul a constatat că persoanele asiatice și afro-americane erau de până la 100 de ori mai susceptibile de a fi identificate greșit decât bărbații albi, în funcție de algoritmul și tipul de căutare utilizate (Harwell, 2019b).

3. Instrumente de evaluare a eticii unei soluții AI

În dezvoltarea sistemelor de inteligență artificială, în special în domenii sensibile la prejudecăți precum recrutarea, este esențial să se utilizeze diverse instrumente concepute pentru a identifica și evalua potențialele prejudecăți. Aceste **instrumente de evaluare** sunt esențiale pentru a asigura funcționarea corectă și transparentă a aplicațiilor AI. Prin utilizarea acestor resurse, dezvoltatorii pot înțelege mai bine modul în care prejudecățile pot influența rezultatele sistemului și pot lua măsuri proactive pentru a atenua aceste efecte. Această abordare nu numai că îmbunătățește fiabilitatea și corectitudinea sistemelor de inteligență artificială, dar contribuie și la consolidarea încrederii utilizatorilor prin demonstrarea unui angajament față de practicile etice în materie de inteligență artificială.

Instrumentele de proiectare sunt esențiale pentru considerentele etice în dezvoltarea inteligenței artificiale, deoarece facilitează o înțelegere holistică a perspectivelor diferitelor părți interesate. Aceste instrumente permit dezvoltatorilor să ia în considerare cu empatie impactul divers al tehnologiilor IA asupra diferitelor grupuri de utilizatori. Prin cartografierea interacțiunilor, comportamentelor și nevoilor acestor grupuri, aceste instrumente ajută la identificarea și atenuarea potențialelor prejudecăți inerente sistemelor AI. Implicarea pe scară largă nu numai că asigură faptul că soluțiile de inteligență artificială sunt mai incluzive, dar aliniază, de asemenea, dezvoltarea produselor la standardele etice prin abordarea proactivă a problemelor care ar putea conduce la rezultate inechitabile sau discriminatorii. În cazul acestui proiect, diferite instrumente sunt prezentate pe parcursul procesului de proiectare în cadrul unei etape relevante.

Deși există o multitudine de instrumente de evaluare disponibile în acest scop, acest curs se concentrează pe prezentarea a doar trei dintre acestea. Instrumentele de proiectare utilizate în acest proiect sunt detaliate în descrierea proiectului care se găsește mai departe în acest manual Capitolul E. Identificarea potențialelor prejudecăți.

Instrumentul UNESCO de evaluare a impactului etic, dezvoltat în 2023, oferă o abordare structurată pentru evaluarea implicațiilor etice ale sistemelor de inteligență artificială. Acesta cuprinde o serie de întrebări diverse, care cuprind atât formate deschise, cât și formate cu opțiuni multiple, concepute pentru a ghida utilizatorii printr-un proces de evaluare cuprinzător. Punând accentul de la bun început pe considerentele etice, instrumentul începe cu anchete de anvergură înainte de a aprofunda principiile UNESCO. Prin abordarea unor aspecte-cheie precum impactul asupra părților interesate, responsabilitățile echipei, siguranța, securitatea și preocupările privind durabilitatea, confidențialitatea, transparența, explicabilitatea și responsabilitatea, instrumentul facilitează o examinare aprofundată a dimensiunilor etice inerente implementării IA. (UNESCO, 2023)

Secțiunea Echitate, nediscriminare și diversitate a instrumentului UNESCO de evaluare a impactului etic se concentrează pe identificarea prejudecăților algoritmice, în special în datele utilizate de sistemele de inteligență artificială. Aceasta examinează în detaliu rolurile și responsabilitățile implicate pentru a se asigura că toate prejudecățile inerente sunt abordate. (UNESCO, 2023)

Lista de evaluare pentru inteligența artificială demnă de încredere (ALTAI) a Grupului de experți la nivel înalt privind inteligența artificială (AI HLEG), instituit de Comisia Europeană, oferă un cadru structurat pentru evaluarea caracterului demn de încredere al sistemelor AI. Acest instrument prezintă șapte cerințe specifice pentru o IA demnă de încredere, cu explicații detaliate disponibile pe site-ul însoțitor. În plus, acesta oferă orientări privind finalizarea evaluării și identifică principalele părți interesate care ar trebui să fie implicate în proces, asigurând o evaluare cuprinzătoare a fiabilității sistemelor AI. (ALTAI, 2020)

Secțiunea Diversitate, nediscriminare și corectitudine a instrumentului ALTAI se concentrează pe evitarea prejudecăților incorecte. Aceasta evaluează prejudecățile algoritmice prin întrebări specifice, evaluând datele de intrare și proiectarea algoritmului pentru prejudecăți, promovând în același timp incluziunea, accesibilitatea și participarea părților interesate pentru a asigura corectitudinea și transparența în sistemele AI. (ALTAI, 2020)

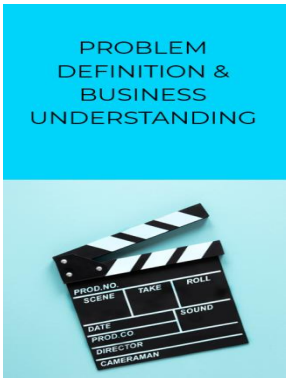
Lucrarea lui Mario Sosa Hidalgo "Design of an Ethical Toolkit for the Development of AI Applications" oferă o abordare inovatoare, concretă și vizuală pentru integrarea eticii în procesul de dezvoltare a aplicațiilor AI. Instrumentul abordează prejudecățile și îi sfătuiește pe dezvoltatori să pună în aplicare strategii care minimizează nedreptatea, asigurându-se că sistemul AI este cât mai imparțial posibil. (Hidalgo, 2019)

4. Proiect: Soluție AI în recrutare

Această secțiune se concentrează pe abordarea prejudecăților algoritmice în cadrul sistemelor AI. Acesta prezintă un proiect care implică un scenariu ipotetic în care o companie internațională dezvoltă o soluție AI pentru recrutare.

Pentru a ghida această explorare, narațiunea este cuprinsă în casete de text conturate în albastru, care servesc drept informații esențiale pentru proiect. În plus, această secțiune oferă informații complete pentru a ajuta la înțelegerea contextului și a dezvoltării aplicației AI în scopul recrutării.

Scopul este de a implica studenții în recunoașterea și atenuarea prejudecăților care pot apărea în timpul creării și punerii în aplicare a tehnologiilor AI.

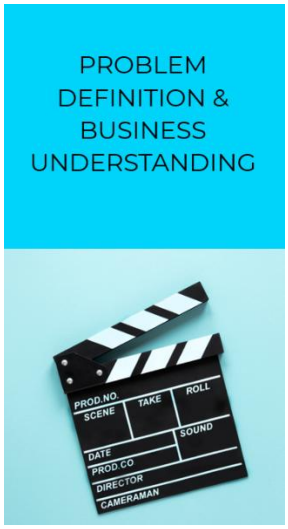


Stabilirea scenei

O companie industrială globală se confruntă cu ineficiențe și incertitudini în procesul său de recrutare, care pune o presiune semnificativă pe departamentul de resurse umane, consumând timp și resurse.

Anticipând creșterea viitoare, compania recunoaște, de asemenea, nevoia urgentă de a angaja o varietate de specialiști, inclusiv ingineri, specialiști în marketing, profesioniști în finanțe, programatori și manageri de

Inteligența artificială (AI) este din ce în ce mai mult integrată în procesele de recrutare, oferind o serie de posibilități de îmbunătățire a practicilor de recrutare. Deși aplicațiile IA în recrutare aduc avantaje semnificative, acestea prezintă și anumite provocări. Este esențial ca organizațiile să recunoască și să analizeze cu atenție ambele părți ale monedei. Prin înțelegerea beneficiilor potențiale, precum și a dezavantajelor, companiile pot lua decizii în cunoștință de cauză cu privire la cel mai bun mod de implementare a tehnologiilor AI. Această abordare echilibrată asigură faptul că, în timp ce se exploatează puterea AI pentru a raționaliza și îmbunătăți recrutarea, implicațiile și riscurile sunt, de asemenea, abordate în mod adecvat.



Înțelegerea problemei

Compania începe cercetarea de bază pentru a dezvolta o soluție AI, formând o echipă internă dedicată identificării principalelor provocări din procesul de recrutare existent. Echipa realizează interviuri cu o serie de părți interesate, inclusiv departamentul de resurse umane, angajații nou angajați și managerii care au integrat recent noi membri ai echipei. De asemenea, echipa culege date privind costurile actuale de recrutare și analizează rata de succes a acestor inițiative.

Înțelegerea percepțiilor potențialilor angajați cu privire la utilizarea IA în recrutare este esențială pentru companie. Aceasta urmărește să evite orice afectare a reputației sau a mărcii sale. Pentru a realiza acest lucru, compania se angajează să se asigure că adoptarea inteligenței artificiale nu numai că este acceptată pe scară largă, dar și că aderă la standardele etice și juridice.

După discuții și investigații, avantajele și dezavantajele unei soluții de recrutare bazate pe inteligența artificială sunt enumerate în continuare:

Beneficiile inteligenței artificiale în recrutare

Inteligența artificială revoluționează procesul de recrutare prin automatizarea sarcinilor de rutină, permițând astfel recrutorilor să se concentreze pe aspecte mai strategice ale activității lor. Această automatizare nu numai că sporește eficiența generală în gestionarea aplicațiilor, dar și reduce semnificativ timpul petrecut în procesele de recrutare. (Albaroudi et al., 2024; www.recruiter.com, 2023)

În plus, capacitatea IA de a analiza seturi extinse de date permite identificarea celor mai potriviți candidați, îmbunătățind astfel calitatea angajărilor. Această

capacitate analitică poate reduce fluctuația personalului prin asigurarea unei bune concordanțe între post și candidat. Feedback-ul și actualizările rapide facilitate de inteligența artificială nu numai că îmbunătățesc comunicarea, dar și țin candidații bine informați pe parcursul procesului de selecție. Astfel, recrutorii au mai mult timp la dispoziție pentru a intra în contact cu candidații selectați, îmbunătățind și mai mult experiența candidaților. (Arivu Recruitment and Consulting, 2023; Sheard, 2022; www.recruiter.com, 2023)

Implementarea IA în recrutare conduce, de asemenea, la reducerea costurilor prin minimizarea dependenței de munca manuală. (www.recruiter.com, 2023) În plus, IA se străduiește să diminueze prejudecățile umane, sporind astfel obiectivitatea, consecvența și corectitudinea în procesul de recrutare. Acest progres tehnologic este deosebit de benefic pentru grupurile care se confruntă cu bariere în calea participării economice, deoarece le îmbunătățește potențial perspectivele de angajare. (Bursell & Roumbanis, 2024; Köchling & Wehner, 2020; Sheard, 2022)

Capacitatea AI de analiză profundă se extinde la selectarea pe rețelele sociale, unde evaluează profilurile candidaților pentru potrivirea culturală și adecvarea la rolul postului. De asemenea, depistează conținutul inadecvat, asigurându-se că recrutarea se aliniază valorilor organizaționale. În plus, instrumentele AI ajută la depășirea barierelor lingvistice în recrutare, permițând accesul la un bazin de talente mai larg și mai divers la nivel global. (Albaroudi et al., 2024)

În cele din urmă, se așteaptă ca IA să îmbunătățească diversitatea la locul de muncă prin eliminarea prejudecăților din procesul de recrutare. (Sheard, 2022) Această serie de beneficii subliniază impactul transformator al IA asupra peisajului recrutării, promițând un proces de angajare mai eficient, echitabil și favorabil incluziunii.

Preocupări legate de IA în recrutare

Introducerea inteligenței artificiale în procesul de recrutare, deși oferă numeroase eficiențe, aduce cu sine provocarea contactului uman limitat și a interacțiunii impersonale. Această reducere a implicării personale poate face ca experiența de recrutare să pară mecanică, ceea ce poate îndepărta candidații și diminua calitatea experienței de recrutare. (www.recruiter.com, 2023)

În plus, riscul de părtinire și discriminare reprezintă o preocupare semnificativă. IA, dacă nu este proiectată și instruită corespunzător sau dacă se bazează pe date de instruire părtinitoare, poate introduce în mod involuntar prejudecăți în procesul de recrutare. Acest lucru poate avea un impact negativ asupra diversității, incluziunii și echității în practicile de angajare. (Albassam, 2023; www.recruiter.com, 2023)

Preocupările legate de protecția și securitatea datelor sunt, de asemenea, primordiale, având în vedere cantitatea mare de date cu caracter personal colectate și prelucrate de sistemele de inteligență artificială. Aceste preocupări

ridică probleme legate de viața privată, protecția datelor și riscul atacurilor cibernetice, subliniind necesitatea unor măsuri de securitate solide și a transparenței în utilizarea datelor. (Albassam, 2023; www.recruiter.com, 2023)

Implementarea IA în recrutare nu este lipsită de costuri și de resurse. Pentru întreprinderile mici, în special, investițiile financiare și de resurse necesare pot fi substanțiale, reprezentând o barieră la intrare și limitând potențial adoptarea IA în procesele lor de recrutare. (www.recruiter.com, 2023)

Există, de asemenea, preocupări cu privire la lipsa de judecată umană și la potențialele probleme de acuratețe. Bazarea exclusivă pe inteligența artificială, fără supraveghere umană, ar putea duce la ignorarea candidaților calificați din cauza datelor incomplete sau a nerecunoașterii transferabilității competențelor. În plus, acuratețea AI poate fi compromisă de candidații care își manipulează datele pentru a părea mai calificați. (Albassam, 2023; Arivu Recruitment and Consulting, 2023)

Evoluția cadrului juridic privind utilizarea inteligenței artificiale în recrutare impune companiilor să se adapteze și să se conformeze pentru a asigura echitatea. Acest peisaj în evoluție evidențiază necesitatea unei vigilențe și a unei adaptabilități constante în ceea ce privește utilizarea instrumentelor AI în procesul de recrutare. (Fernandes, 2021; www.recruiter.com, 2023)

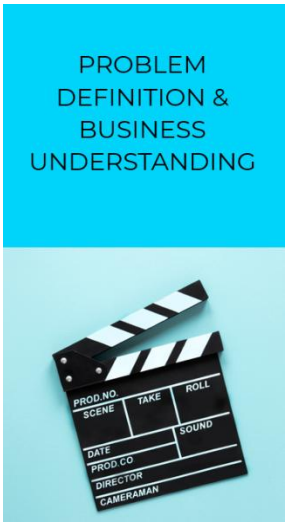
Limitările în capacitatea AI de a evalua potrivirea culturală sau abilitățile de lucru în echipă pot duce la ratarea oportunităților de a angaja candidați care ar putea excela dacă li s-ar oferi șansa. Albassam (2023) subliniază importanța luării în considerare a acestor factori, care sunt adesea critici pentru performanța la locul de muncă, dar dificil de cuantificat cu exactitate de către sistemele AI.

Dominația pieței și lipsa de transparență a instrumentelor de angajare bazate pe inteligența artificială, în special a celor dezvoltate de corporații private din SUA, complică eforturile de atenuare a prejudecăților și de control extern. Natura proprietară a acestor sisteme, ca proprietate intelectuală, face dificilă evaluarea și îmbunătățirea corectitudinii și eficacității lor. (Sheard, 2022)

În cele din urmă, exacerbarea inegalităților prin intermediul platformelor automate de angajare este o preocupare din ce în ce mai mare. Aceste sisteme pot consolida inegalitățile sociale existente prin algoritmi părtinitori și prin rolul semnificativ al rețelelor și al structurilor informale în obținerea unui loc de muncă. Această preocupare evidențiază necesitatea unei analize atente și a luării de măsuri pentru a elimina prejudecățile din platformele de angajare bazate pe inteligența artificială. (Harvard John A. Paulson School of Engineering and Applied Sciences, 2023)

Fiecare dintre aceste puncte subliniază complexitatea și natura multidimensională a încorporării IA în procesul de recrutare, subliniind necesitatea unei abordări echilibrate care să ia în considerare eficiența, corectitudinea și incluziunea.

5. Identificarea prejudecăților potențiale



Înțelegerea utilizatorilor sistemului

Una dintre persoanele intervievate este Lead Recruiter Maria, care are mai mulți ani de experiență în cadrul companiei. Ea a văzut cu ochii ei acumularea de sarcini de recrutare și presiunea tot mai mare pe care o provoacă asupra departamentului de resurse umane, care se intensifică atunci când se angajează talente globale.

Maria este deschisă la minte și interesată de posibilitățile oferite de inteligența artificială pentru a ajuta la sarcinile de recrutare, chiar dacă nu este

Instrument: Persona

Personas sunt utilizate în proiectare pentru a crea profiluri detaliate care reprezintă tipuri specifice de utilizatori sau grupuri. Aceste profiluri surprind caracteristicile esențiale fără a recurge la stereotipuri. Prin includerea unor detalii realiste, cum ar fi comportamentele, nevoile și informațiile demografice, persoanele devin relatabile și utile pentru înțelegerea diferitelor grupuri de utilizatori. Acestea ajută echipele de proiectare și dezvoltare să creeze empatie, să își alinieze eforturile și să dezvolte soluții adaptate pentru a răspunde nevoilor specifice. Personajele sunt deosebit de valoroase deoarece ajută la descompunerea segmentelor tradiționale de marketing, conducând la o proiectare mai incluzivă și mai eficientă. (Kumar, 2013; Stickdorn et al., 2018a, 2018b)

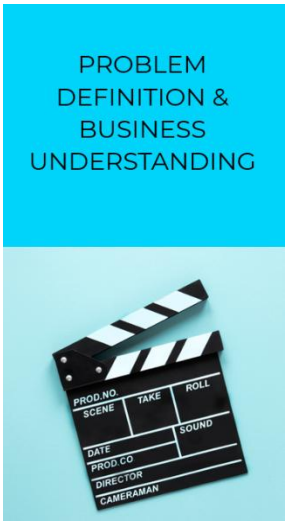
Biasul de perspectivă al profesionistului în resurse umane Maria

Atunci când evaluați persoana, este important să luați în considerare faptul că opiniile personale ale acesteia ar putea influența recomandările și opiniile pe care le oferă cu privire la adoptarea instrumentului. Există o dezbatere continuă cu privire la faptul dacă IA reprezintă o provocare sau un risc pentru rolurile tradiționale de recrutare. Profesioniștii din domeniul resurselor umane ar putea percepe sistemele de recrutare bazate pe IA ca pe o amenințare la adresa locurilor lor de muncă, ceea ce ar putea duce la o reticență față de adoptarea instrumentelor de IA pentru recrutare (Chen, 2023). Alte prejudecăți individuale bazate pe persoana sa ar putea fi, de exemplu, următoarele:

- Deschiderea sau scepticismul legat de vârstă față de noile tehnologii și impactul acestora asupra rolurilor tradiționale.
- Influența faptului că se află într-un oraș tehnologic avansat, precum Helsinki, care poate favoriza soluțiile inovatoare.
- Originea sa guatemaleză ar putea afecta percepția sa asupra impactului AI asupra diversității și incluziunii în recrutare.
- Educația sa în domeniul managementului resurselor umane poate favoriza credința în importanța judecății umane, ceea ce poate determina reticența de a se baza pe inteligența artificială pentru sarcinile critice de recrutare.
- Trăind singură, experiențele sale personale ar putea să o facă să prețuiască mai mult interacțiunea umană, ceea ce ar duce la scepticism în ceea ce privește natura impersonală a IA.
- Obiectivele sale profesionale pun accentul pe eficiență și pe implicarea personală a candidaților, ceea ce ar putea intra în conflict cu viziunea sa asupra recrutării AI.
- Frustrările legate de volumul mare de muncă și de procesele lente ar putea să o facă deschisă la eficiența AI, însă teama de depersonalizare și de pierderea contactului cu candidații ar putea provoca ezitări.

- Hobby-uri precum înotul, alpinismul și cititul indică o abordare echilibrată și multifațetată a vieții, care s-ar putea traduce prin dorința de a avea o abordare nuanțată a recrutării.
- Îngrijorările legate de modificarea funcțiilor de resurse umane de către inteligența artificială o pot determina să se întrebe cum se aliniază aceasta cu setul său de competențe și cu nevoile viitoare pentru rolul său.

Călătoria de recrutare



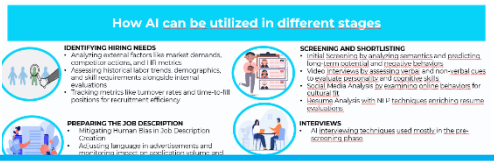
Înțelegerea procesului de recrutare

Pentru a obține o înțelegere cuprinzătoare a procesului de recrutare, compania dezvoltă o călătorie de recrutare care o implică pe Maria și echipa de recrutare. În colaborare cu managerii de linie, aceștia navighează prin fiecare etapă a călătoriei pentru a identifica frustrările semnificative și etapele care consumă mult timp. În plus, ei explorează potențiala integrare a IA în fiecare etapă pentru a spori eficiența și eficacitatea procesului de recrutare.

Instrument: Harta călătoriei

O hartă de parcurs este un instrument vizual utilizat în proiectare pentru a descrie experiența pas cu pas pe care un utilizator o are cu un serviciu. Aceasta reprezintă perspectiva utilizatorului, detaliind ce se întâmplă în fiecare etapă de interacțiune, punctele de contact implicate și orice provocări sau obstacole cu care s-ar putea confrunta. În plus, hărțile de parcurs includ adesea straturi care indică răspunsurile emoționale ale utilizatorului, evidențiind experiențele pozitive sau negative pe parcursul parcursului lor. Aceste hărți pot fi utilizate pentru a ilustra atât experiențele actuale (hărți de parcurs în starea actuală), cât și interacțiunile viitoare preconizate (hărți de parcurs în starea viitoare). Acestea variază în ceea ce privește domeniul de aplicare, de la vederi generale, la nivel înalt, ale întregii experiențe de la un capăt la altul, până la hărți foarte detaliate care se concentrează pe momente specifice. (Kumar, 2013; Stickdorn et al., 2018a).

În acest proiect, harta parcursului este prezentată inițial ca o descriere a stării actuale, urmată de o proiecție a stării viitoare, subliniind capacitățile pe care instrumentele AI le pot oferi pentru recrutare. Utilizarea AI în diferite etape de recrutare este tratată mai în detaliu în capitolul "Dezvăluirea soluțiilor AI: Abordarea etapelor procesului de recrutare și a prejudecăților asociate acestora".



Recunoașterea prejudecăților personale

După o expunere clară a problemei, compania a ales să continue cu o soluție bazată pe IA. Ați fost numit dezvoltator principal de inteligență artificială, responsabil cu supravegherea punerii în aplicare a tehnologiei AI alese.

Compania recunoaște importanța reunirii unui grup divers, cu mai multe părți interesate, pentru a asigura o varietate de perspective pentru succesul proiectului.

Într-un efort de a aborda potențialele prejudecăți inconștiente din cadrul echipei, compania a impus ca toți membrii să participe la un exercițiu modelat după Alan Turing numit "Matricea poziționalității"

Părțile interesate și potențialele lor prejudecăți

Instrument: Matricea poziționalității Alan Turing

Reflecția asupra "poziționalității" - înțelegerea propriului mediu social și cultural - este esențială în dezvoltarea inteligenței artificiale pentru luarea deciziilor etice. Institutul Alan Turing (2022) subliniază importanța "matricei poziționalității", un instrument de evaluare a modului în care trăsături individuale precum vârsta, rasa și educația influențează cercetarea și inovarea. Această conștiință de sine este esențială pentru prevenirea prejudecăților și a dezechilibrelor de putere în sistemele AI. Prin luarea în considerare a acestor factori personali, dezvoltatorii se pot asigura că soluțiile AI sunt echitabile și respectă diversele comunități pe care le deservesc, contribuind astfel în mod pozitiv fără a perpetua nedreptățile.

Matricea poziționalității Alan Turing este un cadru de reflecție care relevă impactul mediilor socioeconomice individuale, al contextelor culturale și al experiențelor de viață asupra judecății și luării deciziilor în cadrul unei echipe. Aceasta servește drept instrument pentru a discerne și a aborda influența diferitelor atribute personale, precum vârsta și etnia, asupra perspectivelor de cercetare și dezvoltare. Această matrice este esențială în descoperirea prejudecăților inconștiente și în navigarea în dinamica puterii, promovând astfel un angajament față de diversitate și echitate în dezvoltarea IA. (Institutul Alan Turing, 2022)

Luarea în considerare a părților interesate

Pe lângă grupul de părți interesate interne care iau parte la proiect, compania trebuie să ia în considerare și alte părți interesate conexe.

Prin urmare, compania a creat o hartă a părților interesate în conformitate cu definiția lui G.J. Millers a procesului de dezvoltare a IA, care definește părțile interesate din domeniul dezvoltării și utilizării, precum și părțile interesate externe.

Pe baza hărții, managerul de proiect trebuie să organizeze părțile interesate în grupuri pentru a ști cum să informeze părțile interesate și să vadă care este influența lor asupra proiectului.

Instrument: Harta părților interesate

O hartă a părților interesate este un instrument vizual utilizat pentru a identifica și afișa rolurile și relațiile tuturor părților interesate implicate într-un proiect. Aceasta poate varia de la un simplu cadran care arată nivelurile de influență și angajament la o matrice detaliată care detaliază interacțiunile dintre părțile interesate. Harta părților interesate ajută la clarificarea dinamicii părților interesate și este esențială pentru înțelegerea și gestionarea relațiilor de rețea. (Giordano et al., 2018; SDT, n.red.)

Instrument: Matricea de analiză a părților interesate

Matricea de analiză a părților interesate este un instrument vizual utilizat pentru clasificarea și gestionarea părților interesate dintr-un proiect în funcție de nivelurile lor de influență și interes. Prin aranjarea părților interesate într-o grilă de cadrane, această matrice ajută la prioritizarea strategiilor de implicare. Cadranul din stânga sus reprezintă părțile interesate cu influență ridicată, dar interes scăzut, care necesită eforturi pentru a le menține satisfăcute fără a le copleși. Cadranul de sus-dreapta include părțile interesate foarte influente și interesate, care necesită o gestionare atentă. Cadranul de jos-dreapta, pentru părțile interesate cu interes ridicat, dar influență scăzută, sugerează actualizări regulate pentru a le menține informate. În cele din urmă, cadranul de jos-stânga pentru cei cu interes și influență minime necesită un angajament minim, concentrându-se pe monitorizare. (Hoory & Botorff, 2022; Wallbridge, 2023)

Biasuri în etapele de recrutare; Atelier

Atunci când părțile interesate sunt clare, compania decide să organizeze un atelier etic care să treacă în revistă toate prejudecățile posibile în diferite etape de angajare.

Părțile interesate din domeniul "Lucrăm împreună" (diapozitivul anterior) sunt invitate să participe.

Ce fel de post ați fi completat?

Instrument: Atelier de lucru

Un atelier este un instrument de proiectare colaborativă în care un grup de participanți se angajează în activități interactive pentru a aborda o problemă specifică sau pentru a lucra la un proiect. Facilitate de obicei de un lider, atelierele pot varia ca durată de la câteva ore la câteva zile, în funcție de obiectivele și complexitatea sarcinilor în cauză. Această metodă este eficientă pentru generarea de idei, rezolvarea problemelor și încurajarea muncii în echipă și a inovării în cadrul unui grup. (Wirtz, 2022)

Prezentarea soluțiilor AI: Abordarea etapelor procesului de recrutare și a prejudecăților asociate acestora

Algoritmii și tehnologiile de învățare automată sunt utilizate din ce în ce mai mult încă din etapele inițiale ale procesului de recrutare. Aceste abordări predictive ajută la anticiparea evoluțiilor afacerilor prin analizarea factorilor externi, cum ar fi cererile pieței, mișcările concurenților și schimbările în nevoile consumatorilor. Măsurătorile de resurse umane joacă un rol crucial în crearea acestor previziuni. Acestea includ o analiză a tendințelor istorice ale forței de muncă, a schimbărilor demografice și a evoluției cerințelor în materie de competențe. În plus, evaluările interne, cum ar fi evaluările volumului de muncă, observațiile manageriale, feedback-ul angajaților și datele demografice ale forței de muncă sunt esențiale pentru înțelegerea stării actuale și a evoluției continue a bazei de angajați. Parametrii precum ratele de fluctuație a personalului, timpul necesar pentru ocuparea posturilor și costul per angajare sunt esențiali pentru evaluarea

eficacității recrutării unei organizații și pentru identificarea domeniilor de îmbunătățire a procesului. (www.recruiter.com, 2023)

Prejudecățile în angajarea predictivă pot apărea din diverse surse, cum ar fi nivelul de performanță al angajaților actuali în anumite roluri, ceea ce poate duce la prejudecăți de ancorare. De exemplu, dacă un angajat care excelează în rolul său trebuie să fie înlocuit, analiza ar putea sugera o singură înlocuire. Cu toate acestea, nivelul ridicat de performanță al angajatului care pleacă ar putea necesita de fapt două înlocuiri. Dimpotrivă, performanța unui angajat cu performanțe scăzute ar putea sugera în mod inexact necesitatea unor angajări suplimentare pentru a compensa lipsa de productivitate a acestuia. (Chen, 2023, p. 145) În plus, rezistența la schimbare în rândul angajaților poate duce la o prejudecată de status quo, în care există o reticență față de adaptarea structurilor organizaționale ca răspuns la nevoi de afaceri clare. (Sukernek, 2024)

În plus, analiza predictivă poate uneori să treacă cu vederea factori critici din cauza unei abordări înguste. De exemplu, dacă nu se iau în considerare tendințele mai largi din industrie sau schimbările demografice, cum ar fi îmbătrânirea populației, o întreprindere poate rămâne nepregătită pentru cerințele viitoare privind forța de muncă.

Ca și alte etape ale procesului de recrutare, această fază este, de asemenea, susceptibilă de a fi influențată de bucla de feedback. Acest tip de părtinire poate apărea atunci când sistemele de inteligență artificială utilizate în recrutare sunt antrenate pe baza datelor istorice și își consolidează din greșeală propriile decizii prejudiciate. Acest lucru se întâmplă în medii dinamice în care deciziile AI influențează datele din care va învăța în viitor. Dacă un AI, deja influențat de datele din trecut, continuă să facă alegeri de angajare părtinitoare, aceste decizii se întorc în setul său de date de învățare, perpetuând și chiar amplificând prejudecățile inițiale (Vivek, 2023, p. 0107).

Biasuri în descrierea postului

Inteligența artificială poate facilita crearea unor descrieri exacte ale posturilor. Prin ajustarea limbajului din anunțuri și monitorizarea efectului acestor modificări asupra volumului și calității candidaturilor, organizațiile își pot spori eficiența eforturilor de promovare. În plus, inteligența artificială poate determina care aspecte ale unei companii, inclusiv cultura și realizările acesteia, ar trebui să fie evidențiate potențialilor candidați pentru a obține cele mai favorabile răspunsuri. (Chen, 2023, p. 140)

Descrierile posturilor joacă un rol esențial în comunicarea nevoilor unei companii către potențialii angajați. Cu toate acestea, prejudecățile prezente în aceste descrieri pot împiedica persoanele calificate să aplice. Astfel de prejudecăți sunt adesea legate de vârstă, sex sau calificări educaționale (Ongig, 2024).

În mod interesant, articolele sugerează că AI are potențialul de a reduce unele dintre aceste probleme prin revizuirea descrierilor posturilor și identificarea

limbajului care poate contribui la prejudecăți. Deși descrierile de locuri de muncă generate de AI ar putea prezenta mai puține prejudecăți decât cele create de oameni, riscul de perpetuare a prejudecăților existente rămâne dacă instrumentele AI nu sunt concepute meticulos (CareerExperts, 2023). De exemplu, o prejudecată de ancorare poate apărea din deciziile anterioare de angajare ale unui manager, în care caracteristicile și competențele actualilor performanți de top influențează crearea de noi descrieri ale posturilor (Chen, 2023, p. 145).

În plus, sugestia AI de a utiliza anumiți termeni în anunțurile de angajare, precum "ambițios" sau "lider încrezător", poate favoriza sau descuraja neintenționat anumite categorii demografice, introducând astfel prejudecăți (Lawton, 2022). În cazul în care sistemele de inteligență artificială sunt antrenate folosind descrieri de locuri de muncă din trecut, acestea riscă să reproducă limbajul și cerințele care au atras în mod istoric un grup mai puțin divers de candidați, continuând potențial această tendință.

Biasuri în căutarea talentelor

Diverse metodologii AI sunt utilizate în timpul fazei inițiale de angajare cu potențialii candidați. Algoritmii facilitează comunicarea personalizată prin intermediul platformelor digitale și al rețelelor de socializare, în timp ce roboții AI sunt utilizați pentru identificarea atât a persoanelor active, cât și a celor pasive aflate în căutarea unui loc de muncă pe diverse platforme sau chiar în baza de date a unei companii cu foști candidați. În plus, inteligența artificială poate implementa strategii de minare a textului pentru a spori atractivitatea anunțurilor de angajare prin încorporarea frazelor care atrag cei mai mulți candidați. În plus, punerea în aplicare a unei metrici a tonului de către AI poate oferi recomandări pentru a rafina descrierile posturilor, făcându-le mai incluzive (Hunkenschroer & Luetge, 2022, p. 992).

Anunțurile de angajare bazate pe inteligența artificială prezintă în mod inerent riscul discriminării indirecte. Acest lucru se datorează faptului că angajatorii pot specifica publicul pentru anunțurile lor folosind criterii precum competențele lingvistice, nivelul de educație, experiența profesională sau chiar vârsta și sexul, ceea ce poate împiedica anumite grupuri să vadă aceste oportunități de angajare, contribuind astfel la omogenitatea locului de muncă (Chen, 2023, p. 144; Sheard, 2022, p. 624). De exemplu, decizia Verizon de a direcționa un anunț pe Facebook către persoane cu vârste cuprinse între 25 și 36 de ani, care locuiesc sau au vizitat capitala SUA și care sunt interesate de finanțe, ar putea exclude din greșeală alte persoane aflate în căutarea unui loc de muncă din aria de acoperire a anunțului (Sheard, 2022, p. 624). Algoritmii concepuți pentru a spori eficiența adăugării de locuri de muncă tind să se concentreze pe categoriile demografice care interacționează istoric mai mult cu aceste anunțuri, reducând astfel diversitatea bazinului de candidați în timp (Köchling & Wehner, 2020).

Prejudecățile de recrutare sunt întărite și mai mult de preferințele recrutorilor pentru anumite platforme de publicitate și de tendința lor de a se angaja mai mult cu candidații de pe aceste platforme. Acest lucru poate duce la o lipsă de diversitate, deoarece diferite platforme atrag diferite tipuri de utilizatori (Lawton, 2022). Prejudecățile în recrutare nu sunt doar rezultatul algoritmilor sau al practicilor angajatorilor; acestea provin, de asemenea, din comportamentele candidaților înșiși. Bazându-se pe social media sau pe alte activități online pentru a identifica potențialii candidați, recrutarea poate favoriza persoanele care sunt active online sau care au cunoștințele necesare pentru a crea o prezență digitală atrăgătoare. Acest lucru este evident în special în industriile extrem de competitive, cum ar fi IT, unde cei mai ocupați profesioniști s-ar putea să nu își actualizeze în mod regulat profilurile LinkedIn. În schimb, există persoane care și-ar putea gestiona în mod strategic profilurile online pentru a-și spori atractivitatea, ceea ce ridică semne de întrebare cu privire la autenticitatea unor astfel de persoane create digital. (Naakka, 2018, pp. 46-47)

Problema transparenței este o preocupare etică semnificativă în recrutarea asistată de inteligența artificială. Complexitatea algoritmilor poate duce la un efect de "cutie neagră", în care raționamentul deciziilor rămâne obscur pentru candidați (DigitalOcean, s.f.). În plus, există întrebări etice cu privire la utilizarea profilurilor și activităților candidaților pe rețelele sociale și pe diverse platforme în scopul angajării. Standardele pentru gestionarea acestor date variază la nivel internațional, Regulamentul general european privind protecția datelor (GDPR) fiind printre cele mai stricte (Hunkenschroer & Luetge, 2022, p. 995).

Prejudecăți în selectarea și preselecția candidaților

Tehnologiile AI joacă un rol crucial în selectarea inițială a candidaților, acordând prioritate celor care par a fi cei mai potriviți pentru rolul respectiv. În trecut, companiile se bazau pe scanarea CV-urilor pentru anumite cuvinte-cheie. Cu toate acestea, abordările moderne sunt mai sofisticate, incluzând chatbots și instrumente de analiză a CV-urilor care caută corelații semantice și evaluează alte calificări. Unele instrumente prezic chiar potențialele performanțe viitoare ale unui candidat, căutând indicatori ai angajamentului pe termen lung sau ai productivității, precum și absența indicatorilor negativi, cum ar fi întârzierile obișnuite sau problemele disciplinare (Hunkenschroer & Luetge, 2022, p. 992).

În procesul de preselecție a candidaților, inteligența artificială este utilizată și în analiza interviurilor video. Asistenții AI conduc aceste interviuri, punând o listă de întrebări și evaluând nu doar răspunsurile, ci și tonul vocii, expresiile microfaciale și răspunsurile emoționale ale candidatului pentru a deduce trăsături de personalitate. În plus, jocurile video bazate pe inteligență artificială sunt utilizate pentru a evalua trăsături precum comportamentul de asumare a riscurilor, abilitățile de planificare, perseverența și motivația (Hunkenschroer & Luetge, 2022, p. 992).

Mai mult, soluțiile AI analizează amprentele digitale ale candidaților, inclusiv activitățile din social media și alte comportamente online, prin analiză lingvistică, pentru a evalua compatibilitatea acestora cu cultura întreprinderii (Hunkenschroer & Luetge, 2022, p. 992). Dincolo de analiza rețelelor de socializare, tehnicile de programare neurolingvistică (NLP) sunt aplicate în diferite etape de recrutare, inclusiv analiza CV-ului, pentru a oferi informații mai detaliate (Albaroudi et al., 2024, p. 391).

IA promite promovarea unui mediu de recrutare mai echitabil prin reducerea prejudecăților inconștiente. Punând accentul pe abilități și acreditări predefinite, soluțiile bazate pe AI pot trece cu vederea detalii demografice precum vârsta, sexul și etnia, care ar putea influența în mod neintenționat judecata unui recrutor. O astfel de abordare are potențialul de a spori diversitatea la locul de muncă, deoarece asigură faptul că candidații sunt evaluați pe baza calificărilor și realizărilor lor. În plus, anumite aplicații de inteligență artificială sunt dezvoltate special pentru a ajuta companiile să își atingă obiectivele în materie de diversitate și incluziune. Acestea realizează acest lucru prin examinarea tendințelor de recrutare și propunerea de ajustări pentru a aborda orice disparități. (DigitalOcean, n.red.)

În ciuda potențialului proceselor de screening de a reduce anumite prejudecăți, acestea pot introduce involuntar altele dacă nu sunt gestionate meticulos. Prejudecățile pot proveni din principiile care stau la baza conceperii modelului, din selectarea caracteristicilor și din datele utilizate pentru formare (Hunkenschroer & Luetge, 2022, p. 994).

Atunci când datele de formare se bazează pe un grup de angajați existenți sau pe un grup definit în mod restrâns care nu reprezintă în mod proporțional populații diverse, aceasta poate duce la discriminarea neintenționată a grupurilor subreprezentate (Hunkenschroer & Luetge, 2022, p. 994). Sistemele AI, care învață din date care pot conține deja prejudecăți istorice, au potențialul de a amplifica și mai mult aceste prejudecăți. De exemplu, dacă un sistem AI este antrenat în mod predominant pe baza CV-urilor unei anumite categorii demografice, acesta poate favoriza în mod incorect candidații din acel grup (Vivek, 2023). Un caz notabil a avut loc în 2018, când sistemul de recrutare AI al Amazon a afișat o preferință pentru modelele dominante în rândul candidaților de sex masculin, discriminând astfel candidații de sex feminin. Acest lucru s-a datorat faptului că datele de formare ale Amazon au constat în principal din CV-uri ale persoanelor cu performanțe de vârf, dintre care majoritatea erau bărbați, ceea ce a determinat algoritmul să penalizeze caracteristicile asociate de obicei cu femeile (Albaroudi et al., 2024, p. 386; Hunkenschroer & Luetge, 2022, p. 994).

În plus, prejudecățile de formare pot apărea și atunci când un instrument algoritmic nu ia în considerare în mod adecvat toate competențele și trăsăturile relevante pentru post. În cazul în care instrumentul se concentrează prea mult pe abilitățile tehnice, fără a lua în considerare abilitățile interpersonale, competențe esențiale precum comunicarea pot fi trecute cu vederea. De asemenea, pot apărea prejudecăți de agregare, atunci când se fac generalizări false cu privire la

populații întregi, ceea ce duce la excluderea unor persoane pe baza unor presupuneri privind caracteristicile grupului. De exemplu, algoritmii pot presupune că persoanele mai tinere au competențe tehnologice superioare, trecând astfel cu vederea candidații mai în vârstă care sunt la fel de calificați (Albaroudi et al., 2024, p. 386).

Evaluarea amprentei digitale a unui candidat ca parte a procesului de selecție poate, de asemenea, să perpetueze prejudecățile, reflectând percepțiile diferite cu privire la ceea ce ar trebui prezentat pe social media și limitările cu care se confruntă oamenii în actualizarea profilurilor lor. Un studiu realizat de Universitatea de Stat din Carolina de Nord, care a implicat interviuri cu 61 de profesioniști din domeniul resurselor umane, a relevat o tendință de a valoriza interesele personale, cum ar fi drumețiile sau sărbătorirea Crăciunului, în detrimentul trăsăturilor profesionale. Această abordare avantajează anumite categorii demografice în detrimentul altora și presupune că caracteristici precum "energicul" sau "activul" sunt în mod inerent superioare, ceea ce poate duce la discriminarea persoanelor în vârstă sau cu handicap. Introducerea inteligenței artificiale în acest proces riscă să integreze aceste prejudecăți în algoritmi (Shipman, 2021).

Utilizarea software-ului de recunoaștere emoțională în evaluările video trebuie să ia în considerare cu atenție intonațiile diverse ale limbilor și potențialul de dezavantajare sistematică a anumitor rase sau grupuri etnice. În plus, unii indivizi pot întâmpina dificultăți în a se comporta în mod natural în fața unei camere de filmat, ceea ce duce la evaluări inexacte ale instrumentului (Hunkenschroer & Luetge, 2022, p. 995).

Confidențialitatea și consimțământul în cunoștință de cauză în utilizarea informațiilor despre candidați pentru recrutare sunt considerații etice esențiale, reglementările variind la nivel internațional. Regulamentul general european privind protecția datelor (GDPR) se numără printre cele mai stricte cadre, fiind conceput pentru a proteja drepturile cetățenilor UE prin reglementarea colectării, stocării și prelucrării datelor cu caracter personal și prin solicitarea consimțământului informat pentru orice activități de prelucrare a datelor cu caracter personal. Cu toate acestea, dilema apare atunci când se analizează dacă solicitanții ar risca să renunțe, temându-se că acest lucru le-ar putea pune în pericol perspectivele de angajare (Hunkenschroer & Luetge, 2022, p. 995).

Prejudecăți în intervievarea candidaților

Un motiv pentru utilizarea tehnicilor de intervievare bazate pe inteligența artificială poate fi barierele lingvistice în recrutarea multinațională, care necesită achiziționarea de talente locale și profesioniști în resurse umane multilingvi. Inteligența artificială atenuează această problemă permițând intervievarea unor vorbitori de limbi diferite, fără a fi nevoie ca departamentul de resurse umane să cunoască fiecare dialect local, datorită progreselor în prelucrarea limbajului natural (NLP). Această tehnologie permite întreprinderilor să reducă în mod

eficient decalajul lingvistic în cadrul angajărilor la nivel mondial, fără a se baza pe traducători. (Albaroudi et al., 2024, p. 392)

Cu toate acestea, instrumentele de interviu bazate pe IA sunt utilizate mai des pentru a automatiza prima rundă de interviuri, contribuind astfel la evaluarea candidaților (Fritts & Cabrera, 2021, p. 792). Prin urmare, soluțiile de intervievare bazate pe inteligența artificială și posibilele prejudecăți pe care acestea le creează sunt tratate în capitolul anterior "Selectarea și preselectarea candidaților".

Vivek (2023) sugerează că, datorită înțelegerii nuanțate și inteligenței emoționale pe care o posedă oamenii, puterea finală de decizie ar trebui să rămână la recrutorii umani. Fritts & Cabrera (2021) menționează, de asemenea, că preocupările legate de procesele bazate pe inteligența artificială pentru identificarea, selectarea, intervievarea și integrarea candidaților ar putea duce la dezumanizare prin eliminarea interacțiunilor personale. Acest lucru ar putea fi problematic din punct de vedere etic, deoarece valorile artificiale încorporate în algoritmii de angajare ar putea submina relația angajat-angajator prin înlăturarea aspectelor inerent umane ale interacțiunilor tradiționale.

Prejudecăți în evaluare și crearea ofertelor

După ce a ales un candidat, compania trebuie să îi facă o ofertă de muncă. Algoritmii de inteligență artificială pot evalua posturile anterioare ale candidaților pentru a formula o ofertă pe care candidatul este probabil să o accepte. Cu toate acestea, aceste instrumente pot exacerba inegalitățile existente în ceea ce privește salariile inițiale și continue între diferite sexe, rase și alte categorii demografice. Acest lucru se întâmplă deoarece aceste instrumente se bazează adesea pe date salariale istorice, care pot conține prejudecăți inerente. (Lawton, 2022)

Biasuri în procesul de integrare

Implementarea inteligenței artificiale în procesul de onboarding poate reduce sarcinile de rutină ale profesioniștilor din domeniul resurselor umane prin automatizarea gestionării documentelor, programarea comunicărilor și utilizarea chatbot-urilor pentru feedback și întrebări inițiale. Aceasta evaluează, de asemenea, progresul noilor angajați și asigură conformitatea cu cerințele de formare, furnizând în același timp analize privind tendințele de onboarding. (Marr, 2023)

Interfețele AI pot adapta traseele educaționale pentru a se potrivi cu abilitățile și stilurile individuale de învățare, asigurându-se că noii angajați primesc sprijinul de care au nevoie pentru a reuși. Departamentele de resurse umane pot utiliza informații din interacțiunile și feedback-ul automatizat pentru a înțelege capacitățile noilor angajați și domeniile de dezvoltare. Analizele predictive pot anticipa nevoile noilor angajați, oferind resurse și sprijin proactiv. (Marr, 2023).

Inteligența artificială poate fi, de asemenea, utilizată în potrivirea noului venit cu mentorul de lucru pe baza competențelor, intereselor și expertizei acestuia. (Featured, 2023)

Ca și în cazul fazelor anterioare, prejudecățile din cadrul procesului de integrare bazat pe inteligența artificială pot apărea din utilizarea datelor de formare înrădăcinate în înregistrări istorice care conțin prejudecăți anterioare. În plus, algoritmii ar putea încorpora în mod involuntar prejudecățile subconștiente ale dezvoltatorilor în modelarea parcursurilor educaționale.

6. Verificarea și preselecția candidaților: definirea setului de date

Selecția și preselecția candidaților; definirea setului de date

În timp ce compania analizează toate aspectele și posibilitățile de recrutare prin IA, în prezent compania evaluează intens IA în faza de selecție și preselecție. Aceasta implică o analiză aprofundată a scanării CV-urilor și alinierea candidaților la posturile vacante corespunzătoare.

Prin implementarea instrumentelor de inteligență artificială pentru automatizarea procesului de selectare a CV-urilor, obiectivul este de a identifica rapid și precis candidații care îndeplinesc cerințele postului. Se așteaptă ca adoptarea unei astfel de tehnologii să conducă la o utilizare mai strategică a resurselor de resurse umane, să îmbunătățească calibrul candidaților care ajung pe lista scurtă și, prin alinierea competențelor și experienței candidaților la cerințele companiei, să crească eficiența generală a procesului de recrutare.

În primul rând, compania trebuie să definească seturile

Instruirea AI

Seturile de date de instruire joacă un rol crucial în dezvoltarea algoritmilor. Aceste seturi de date reprezintă baza pe care sunt construiți algoritmii. Scopul acestor seturi de date este de a servi drept "adevăr de bază" sau bază factuală care instruiește algoritmul cu privire la modul de interpretare și prelucrare a unor cantități mari de date. Acestea învață algoritmul să înțeleagă relațiile dintre variabilele de intrare (informațiile introduse în sistem) și o variabilă de ieșire (predicția sau decizia pe care sistemul o ia pe baza datelor de intrare). O provocare semnificativă apare atunci când datele de formare conțin în sine erori. Aceste prejudecăți pot fi o reflectare a inegalităților istorice, a prejudecăților sau pur și simplu rezultatul unei colectări de date nereprezentative sau incomplete. Atunci când un algoritm este antrenat pe date părtinitoare, acesta învață aceste prejudecăți și le perpetuează în predicțiile și deciziile sale. Acest fenomen este descris în mod obișnuit ca problema "bias in, bias out". Aceasta înseamnă că, dacă datele de intrare (bias in) sunt tendențioase, rezultatul modelului de învățare automată (bias out) va fi, de asemenea, tendențios. Acest lucru poate duce la rezultate inechitabile, discriminatorii sau eronate care afectează selectarea sau respingerea candidaților pe baza unor criterii părtinitoare, mai degrabă decât pe baza calificărilor sau potențialului lor real. (Sheard, 2022)

Instruirea AI pe baza datelor interne ale companiei

Atunci când instruiesc aplicațiile AI pentru verificarea CV-urilor, companiile au acces la o varietate de surse interne de date care pot fi esențiale în procesul de învățare. (Fernandes, 2021; Ma, 2024; Sheard, 2022)

O astfel de sursă este reprezentată de datele privind performanța angajaților, cuplate cu datele istorice privind ocuparea forței de muncă, care cuprind înregistrările interne ale evaluărilor angajaților și rezultatele angajărilor anterioare. În plus, companiile pot utiliza acumularea de CV-uri depuse de-a lungul timpului, care include o multitudine de CV-uri și CV-uri istorice care au fost primite direct pentru cererile de angajare.

O altă resursă valoroasă sunt datele privind "cazurile fals negative", care se referă la informații privind candidații care au fost respinși din greșeală în ciclurile de recrutare anterioare. Analiza acestor cazuri poate oferi informații despre tiparele pe care AI ar trebui să evite să le repete. În plus, companiile pot extrage cuvinte-cheie și fraze specifice care sunt frecvent întâlnite în CV-urile colectate, folosindu-le pentru a rafina capacitatea AI de a identifica calificările și experiența relevante.

În cele din urmă, informațiile privind educația, care sunt o componentă standard a aplicațiilor candidaților, pot fi, de asemenea, valorificate pentru a îmbunătăți înțelegerea inteligenței artificiale a calificărilor academice și relevanța acestora pentru rolurile profesionale.

Instruirea AI pe baza datelor externe ale companiei

Au fost sugerate posibile surse de date externe în formarea IA, care includ mai multe platforme și metode de colectare a informațiilor. (Banerjee, 2022) Platformele social media sunt o sursă bogată în care indivizii împărtășesc interese personale, experiențe și realizări profesionale, oferind informații despre personalitatea, abilitățile și rețeaua profesională a unui candidat. (Albaroudi et al., 2024; Albassam, 2023; Chaker, 2018; Filtered, s.n.; Heymans, 2022; Shipman, 2021) În plus, site-urile profesionale concepute pentru crearea de rețele și dezvoltarea carierei, cum ar fi LinkedIn, conțin profiluri detaliate care includ istoricul muncii, educația, competențele, aprobările și realizările profesionale. Aceste profiluri oferă o imagine cuprinzătoare a potențialelor talente. În cele din urmă, căutările pe internet efectuate de companii pot scoate la iveală informații disponibile publicului despre un candidat din surse precum articole de știri, bloguri personale, publicații sau alte prezențe web. Această gamă largă de informații oferă un context valoros cu privire la calificările, realizările și caracterul unei persoane. (Chaker, 2018)

7. Dezvoltarea modelului algoritmic

Anunț de angajare; selectarea cuvintelor cheie pentru algoritm

Compania a publicat recent o descriere a postului și a primit sute de candidaturi în răspuns.

Prototipul de screening și preselecție cu inteligență artificială va scana CV-urile candidaților pentru cuvinte-cheie și alte informații considerate relevante pentru angajările de succes, cum ar fi experiența, titlurile posturilor, angajatorii anteriori, universitățile și diplomele. Pe baza acestor informații, sistemul creează un profil structurat al candidatului, iar toți candidații sunt notați și clasificați. (Sheard, 2022)

Inteligența artificială simplifică în mod semnificativ procesul de selecție a candidaților, care este în mod tradițional una dintre etapele cele mai consumatoare de timp în recrutare. În timpul etapei de "selecție" a pâlniei de recrutare, evaluarea candidaților pe baza experienței, competențelor și a altor caracteristici pertinente este esențială pentru crearea unei liste scurte de interviuri. Inteligența artificială poate fi utilizată pentru scanarea CV-urilor în căutarea cuvintelor-cheie și a altor date asociate cu angajările de succes - cum ar

fi titlurile posturilor, angajatorii anteriori, instituțiile de învățământ și calificările - simplificând astfel procesul de identificare a candidaților potriviți. (Dennis, 2023; Fernandes, 2021; Sheard, 2022)

Prin utilizarea instrumentelor de inteligență artificială, recrutorii pot identifica rapid și prioritiza candidații de top, sporind astfel eficiența și îmbunătățind parametrii-cheie de recrutare, cum ar fi timpul până la angajare. Sistemele de inteligență artificială elimină automat candidații care nu îndeplinesc anumite criterii prestabilite, analizează CV-urile pentru a evidenția experiența relevantă și clasifică candidații în funcție de adecvarea lor pentru rol. Această automatizare permite recrutorilor să se concentreze mai mult pe evaluarea candidaților în etapele ulterioare ale pâlniei de recrutare. În plus, inteligența artificială poate scana CV-urile în timpul fazei inițiale de selecție a CV-urilor, identificând candidații care îndeplinesc anumite criterii. Algoritmii de învățare mecanică evaluează apoi competențele și calificările acestor candidați în raport cu specificațiile postului pentru a compila o listă de candidați potriviți. (ResumeMent, 2023; Dennis, 2023)

Descrierea postului oferă detalii esențiale despre rol, inclusiv competențele, calificările și experiența necesare, pe care AI le utilizează pentru a selecta CV-urile. Fără o descriere detaliată a postului, inteligența artificială nu ar avea parametrii necesari pentru ceea ce trebuie să caute în CV-uri.

Job description for Communication Assistant
We are looking for a communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

Responsibilities
- Develop and manage diverse content, including media releases, blogs, and social media posts. aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support. including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

Requirements
- Holds a Bachelor's degree in communications. marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

Progresele în domeniul inteligenței artificiale au depășit simpla potrivire a cuvintelor cheie. Inițial, companiile foloseau algoritmi pentru a scana CV-urile în funcție de cuvinte-cheie sau fraze preselectate. Tehnologiile AI de astăzi, inclusiv chatbot-urile și instrumentele sofisticate de scanare a CV-urilor, evaluează acum

relevanța semantică și termenii înrudiți pentru a determina mai precis calificările unui candidat. În plus, algoritmii de învățare automată sunt utilizați pentru a prezice performanțele profesionale viitoare pe baza diferiților indicatori, cum ar fi vechimea în muncă și productivitatea. Aceste instrumente pot sugera, de asemenea, cele mai potrivite locuri de muncă pentru candidați pe baza profilurilor acestora. (Hunkenschroer & Luetge, 2022)

Cu toate acestea, în acest curs, vom folosi un simplu exemplu de potrivire a cuvintelor cheie pentru a ilustra modul în care o aplicație AI de filtrare a CV-urilor procesează cererile.

Lista cuvintelor-cheie specifice utilizate pentru depistarea candidaților

"Diplomă de licență în comunicare"
"Diplomă de licență în marketing" "Creație de conținut"
"Comunicate de presă"
"Bloguri"
"Postări în social media"
"Monitorizarea mass-media" și "acoperirea campaniei"
"Sprijinirea punerii în aplicare a strategiilor de comunicare"
"Gestionarea imaginii companiei"
"Organizarea de evenimente de marketing"
"Furnizarea de asistență administrativă completă" "Actualizarea listelor de contacte"
"Compilarea de prezentări" și "rapoarte" "Urmărirea progresului proiectului" și "expunerea în mass-media"
"Competență în strategiile de social media" și "relații cu mass-media"
"Abordare creativă și inovatoare"
"Abilități organizaționale puternice" și "atenție la detalii"
"Capabil de multitasking"
"Menținerea unor relații interpersonale pozitive" "Competențe în utilizarea programelor de birou" si "software de proiectare"

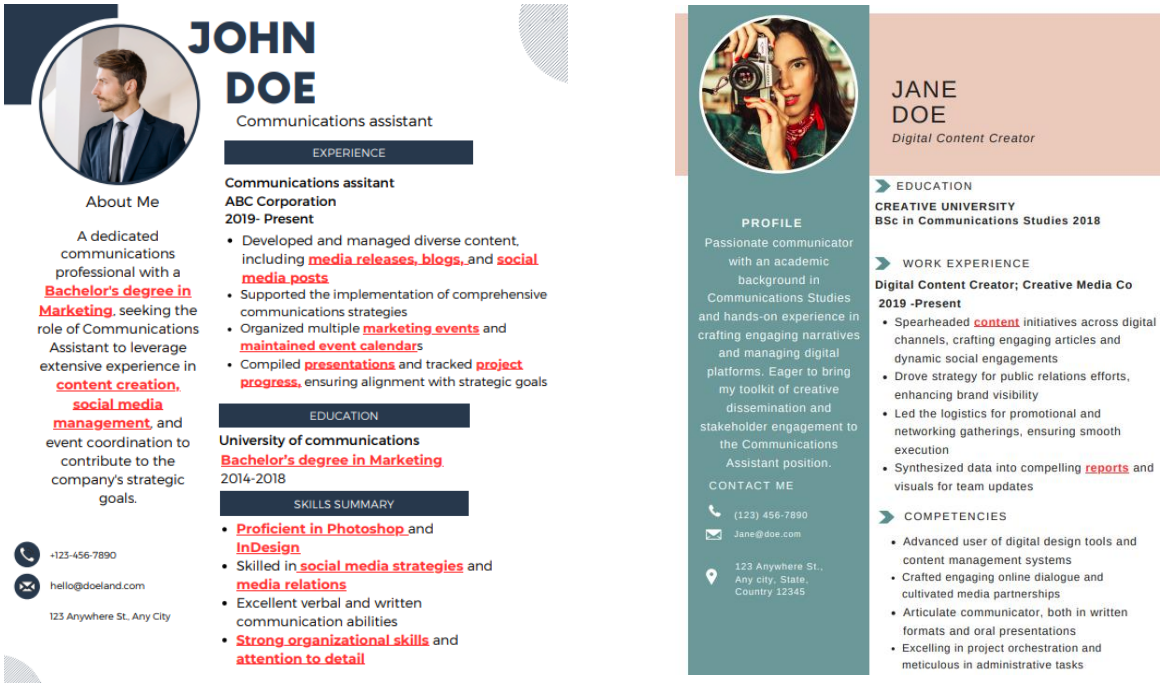
Compararea CV-urilor a doi candidați în ceea ce privește formularea cuvintelor-cheie

Compania organizează un atelier pentru a explora detaliile selectării CV-urilor bazate pe inteligența artificială, utilizând un studiu de caz din propriile operațiuni.

În timpul atelierului, aceștia analizează două CV-uri unul lângă altul pentru a descoperi modul în care diferențele de formulare și de frazare pot duce în

Potrivit lui Sheard (2022), CV-urile sunt, de obicei, examinate de ochi umani numai dacă se aliniază criteriilor anunțului de angajare, în timp ce restul pot fi ignorate, ceea ce evidențiază o potențială preocupare în cazul în care o parte semnificativă a CV-urilor ar putea fi exclusă prematur.

În continuare, ne vom concentra pe analiza a două aplicații pentru a înțelege modul în care prototipul aplicației AI evaluează candidații în timpul fazei de selecție. Această analiză ne va ajuta să identificăm dacă există prejudecăți care influențează procesul de selecție.



Acest exemplu ilustrează impactul diferențelor de redactare și de formulare în două CV-uri și modul în care acestea pot afecta rezultatul unui proces de selecție AI. În ciuda faptului că deține calificări relevante, al doilea CV poate fi trecut cu vederea din cauza faptului că descrierile sale nu se aliniază îndeaproape cu cuvintele-cheie specifice stabilite de companie pentru selecție. Acest lucru demonstrează importanța adaptării limbajului dintr-un CV la cuvintele-cheie așteptate de un sistem AI.

Aspect	CV1	CV 2
Job Titles and Terminology	Uses standard terms like "Communications Assistant"	Uses titles like "Digital Content Creator"
Academic Background	Explicitly mentions "Bachelor's Degree in Marketing"	States "BSc in Communications Studies"
Describing Tasks and Skills	Direct match with keywords like "Content creation", "Media releases"	Uses varied language like "crafting engaging narratives", "spearheaded content initiatives"
Technical and Software Skills	Specifies "Photoshop" and "InDesign"	General mention of "digital design tools and content management systems"
Direct Keyword Matches	Closely matches many specified keywords	Uses different phrasing that may not match the specific Keywords set for screening

8. Cum se măsoară echitatea în soluțiile de învățare automată

Testarea caracterului echitabil al soluției

Evaluarea prototipului de algoritm a scos la iveală inconsecvențe în analiza CV, ceea ce a condus la îmbunătățiri pentru o înțelegere mai cuprinzătoare a atributelor CV. Modelul este acum pregătit pentru o fază de testare înainte de implementarea în lumea reală.

Deși mai multe aspecte necesită evaluare, accentul se va pune aici pe testarea corectitudinii aplicației pentru a determina amploarea oricăror prejudecăți apărute în timpul procesului. Măsurătorile de corectitudine sunt utilizate pentru a măsura nivelul de părtinire prezent în deciziile aplicației.

Pentru a crea sisteme echitabile pentru toți și lipsite de prejudecăți, cum ar fi discriminarea de gen sau rasială, este important să se evalueze performanța soluției pe diverse segmente ale populației. Deși există diverse metode de evaluare a corectitudinii unei soluții, în acest context, ne vom concentra exclusiv asupra testelor de clasificare binară. (*Învățarea automată și corectitudinea*, 2021)

Corectitudinea a apărut ca un subiect crucial în învățarea automată datorită aplicării sale tot mai frecvente în diverse domenii. Cercetări ample s-au concentrat asupra acestui domeniu deoarece sistemele de învățare automată depind în mare măsură de date care, atunci când provin de la oameni, pot fi în mod inerent părtinitoare. Definirea echității în termeni matematici s-a dovedit a fi o provocare, ducând la lipsa unui consens privind formulările standard. Cu toate acestea, după cum a remarcat Zhong (2018), majoritatea definițiilor corectitudinii se reunesc în jurul mai multor concepte-cheie:

- Inconștiență
- Paritatea demografică
- Șanse egale
- Paritatea ratei predictive
- Echitate individuală
- Echitate contrafactuală

Paritatea demografică, șansele egale și paritatea ratei predictive sunt grupate de obicei sub umbrela "echității de grup". (Zhong, 2018). Acest concept va fi explorat în continuare în această secțiune, în special prin prisma IA în recrutare.

Pe lângă corectitudinea grupului, Castelnovo et al. (2022) împart categoriile de corectitudine în corectitudine individuală și corectitudine bazată pe cauzalitate, incluzând și subiectele menționate de definițiile lui Zhong (2018).

Matricea de confuzie

Una dintre metodele binare de clasificare este matricea de confuzie (*Machine Learning and Fairness*, 2021). Pentru a înțelege parametrii de corectitudine, începem cu conceptul de matrice de confuzie. Acest instrument rezumă predicțiile făcute de un model în comparație cu rezultatele reale, sau adevărul de bază, pe care a fost antrenat. Matricea de confuzie afișează numărul de predicții corecte și incorecte, oferind o imagine clară a performanței modelului și a validității sale explicabile. (Saplicki, 2022)

	Unqualified	Qualified
Reject	True Negative (TN)	False Negative (FN)
Hire	False Positive (FP)	True Positive (TP)

SURSĂ IMAGINE | Imaginea 1 - Matricea de confuzie (Modificat din (Machine Learning and Fairness, 2021)

O matrice de confuzie ne ajută să clasificăm rezultatele procesului de angajare. Candidații calificați care sunt angajați cu succes sunt considerați adevărați pozitivi (TP), iar candidații necalificați respinși în mod corespunzător sunt adevărați negativi (TN). În schimb, falsii pozitivi (FP) se referă la candidații necalificați care sunt angajați în mod eronat, în timp ce falsii negativi (FN) sunt candidații calificați respinși în mod eronat. O matrice de confuzie clasifică pur și simplu candidații în grupuri distincte pe baza rezultatelor clasificării lor. Cele patru valori produse de o matrice de confuzie sunt utilizate pentru a calcula metrici standard de învățare automată, inclusiv acuratețea și precizia, pe diferite grupuri demografice. (*Machine Learning and Fairness*, 2021; Teodorescu, 2020)

În scenariul de recrutare discutat, ne vom concentra în special asupra demografiei de gen. Deși este important să se ia în considerare și alți factori demografici, acest exemplu se va concentra exclusiv pe gen.

Să presupunem că există un număr egal de 100 de candidați bărbați și femei. Sistemul a identificat cu acuratețe 75 de persoane din fiecare grup ca fiind calificate sau necalificate, demonstrând că rata de acuratețe este consecventă pentru ambele categorii demografice. (*Machine Learning and Fairness*, 2021)

MEN	Unqualified	Qualified	WOMEN	Unqualified	Qualified
Reject	**15 (TN)**	5 (FN)	Reject	**60 (TN)**	20 (FN)
Hire	20 (FP)	**60 (TP)**	Hire	5 (FP)	**15 (TP)**

SURSĂ IMAGINE | Imaginea 2 - Exemplu de matrice de confuzie din recrutare (Modificat din Machine Learning and Fairness, 2021)

Paritatea demografică

Pe baza datelor furnizate, vom examina modul în care se aplică în acest context diverși parametri de echitate, începând cu paritatea demografică. Acest parametru ia în considerare doar rezultatul clasificatorului, fără a ține cont de etichetele reale. Paritatea demografică este unul dintre parametrii sugerați atunci când se evaluează un sistem automat de angajare. (*Machine Learning and Fairness*, 2021)

Paritatea demografică, cunoscută în mod obișnuit ca independență sau paritate statistică, este un criteriu de echitate în care probabilitatea unui rezultat favorabil, cum ar fi angajarea, nu este influențată de o caracteristică protejată precum sexul (Zhong, 2018).

MEN	Unqualified	Qualified
Reject	15 (TN)	5 (FN)
Hire	**20 (FP)**	**60 (TP)**

WOMEN	Unqualified	Qualified
Reject	60 (TN)	20 (FN)
Hire	**5 (FP)**	**15 (TP)**

SURSĂ IMAGINE | Imagine 3 - Exemplu de paritate demografică din recrutare (Modificat din Machine Learning and Fairness, 2021)

În scenariul de exemplu al recrutării (imaginea 3), dacă 80 din 100 de candidați bărbați și doar 20 din 100 de candidați femei sunt angajați (imaginea 3), paritatea demografică nu este respectată. Cu toate acestea, acest lucru ar putea fi acceptabil dacă candidații de sex masculin sunt într-adevăr mai calificați decât candidații de sex feminin, așa cum pare să fie în acest caz. (*Machine Learning and Fairness*, 2021).

Cu toate acestea, paritatea demografică este dificil de realizat atunci când persoanele aparțin mai multor grupuri protejate, deoarece probabilitățile egale ar putea să nu fie fezabile pentru toate grupurile demografice. Uneori, paritatea demografică poate favoriza în mod involuntar persoanele mai puțin calificate din cauza concentrării sale la nivel de grup, ceea ce poate conduce la inechități la nivel individual. De exemplu, creșterea proporției de candidați de sex feminin mai puțin calificați ar putea reduce șansele de angajare a candidaților de sex masculin calificați. (Teodorescu, 2020) Prin urmare, această măsură ia în considerare doar rezultatul, nu adevăratele etichete, ceea ce înseamnă că nu ține cont de calificările reale ale candidaților. Aceasta susține că raportul dintre candidații angajați și totalul candidaților ar trebui să fie similar pentru toți. (*Machine Learning and Fairness*, 2021).

Paritate predictivă

Suficiența, privită din perspectiva celor care primesc aceeași decizie de la un model, impune rate de eroare egale, indiferent de atributele sensibile. Acest

concept este înglobat de paritatea predictivă, cunoscută și sub denumirea de testul rezultatului. Acesta garantează că ratele de eroare sunt echilibrate pentru persoanele grupate în funcție de deciziile pe care le primesc, mai degrabă decât de rezultatele reale. Spre deosebire de paritatea demografică, paritatea predictivă ia în considerare atât deciziile clasificatorului, cât și rezultatele reale. Aceasta evaluează dacă probabilitatea ca un candidat să fie potrivit pentru un post este consecventă pentru diferite grupuri, cu condiția ca AI să le fi selectat pentru angajare. Se preconizează că această probabilitate ar trebui să fie aproape identică pentru toate grupurile luate în considerare. (Castelnovo et al., 2022; *Machine Learning and Fairness*, 2021).

MEN	Unqualified	Qualified
Reject	15	5
Hire	20	**60**

WOMEN	Unqualified	Qualified
Reject	60	20
Hire	5	**15**

Sursa imaginii | Imaginea 4 - Exemplu de paritate predictivă din recrutare (modificat din *Machine Learning and Fairness*, 2021)

În exemplul nostru de recrutare, sistemul îndeplinește criteriile pentru paritatea predictivă, deoarece trei sferturi dintre candidații angajați din ambele grupuri sunt într-adevăr calificați (*Machine Learning and Fairness*, 2021).

Bilanțul ratei fals pozitive, bilanțul ratei fals negative și cote egalizate

Măsura cunoscută sub numele de **rata falselor pozitive** afirmă că probabilitatea ca un candidat necalificat să fie angajat în mod eronat ar trebui să fie constantă pentru toate categoriile demografice. În termeni mai simpli, sistemul ar trebui să judece la fel de greșit candidații necalificați, indiferent de sexul acestora. (*Machine Learning and Fairness*, 2021)

False positive rate

MEN	Unqualified	Qualified
Reject	15	5
Hire	**20**	60

WOMEN	Unqualified	Qualified
Reject	60	20
Hire	**5**	15

SURSĂ IMAGINE | Imaginea 5 - Exemplu de rată fals pozitivă din recrutare (Modificat din *Machine Learning and Fairness*, 2021)

În acest exemplu (imaginea 5), ratele fals pozitive sunt diferite, o fracțiune mult mai mare de bărbați necalificați fiind angajați decât femei necalificate.

Echilibrul **ratei fals negative** este în esență același lucru cu rata fals pozitivă, dar pentru ratele fals negative. Conceptul de paritate a ratei fals negative sugerează că probabilitatea de respingere eronată a candidaților calificați ar trebui să fie uniformă pentru toate grupurile.

False negative rate

MEN	Unqualified	Qualified
Reject	15	**5**
Hire	20	60

WOMEN	Unqualified	Qualified
Reject	60	**20**
Hire	5	15

SURSĂ IMAGINE | Imaginea 6. Exemplu de rată fals negativă din recrutare (Modificat din *Machine Learning and Fairness*, 2021)

În acest caz (imaginea 6), există o discrepanță notabilă în ceea ce privește ratele fals negative, candidații de sex feminin calificați fiind respinși într-o proporție mai mare decât omologii lor de sex masculin.

Șansele egalizate combină aceste două atribute prin satisfacerea atât a echilibrului ratei fals pozitive, cât și a echilibrului ratei fals negative, ceea ce înseamnă că sistemul trebuie să gestioneze ambele tipuri de erori - candidați necalificați acceptați incorect și candidați calificați respinși incorect - în mod echitabil în cadrul diferitelor grupuri, asigurându-se că toți candidații calificați în mod similar primesc un tratament consecvent. (*Machine Learning and Fairness*, 2021). Egalitatea cotelor este greu de realizat, dar atunci când este atinsă poate fi considerată unul dintre cele mai înalte niveluri de corectitudine algoritmică. (Teodorescu, 2020)

Compromisuri cu metrica

Din cauza constrângerilor matematice, un sistem nu poate îndeplini simultan paritatea predictivă, echilibrul ratei fals pozitive și echilibrul ratei fals negative. Dacă un sistem îndeplinește două dintre aceste criterii, în mod inevitabil nu va reuși să-l îndeplinească pe cel de-al treilea. (*Machine Learning and Fairness*, 2021)

Această teoremă subliniază compromisurile inerente dintre diferitele obiective de echitate. Pe măsură ce ne străduim să fim echitabili, ne confruntăm adesea cu obiective contradictorii, iar atingerea unui echilibru perfect între toate dimensiunile rămâne dificilă.

În practică, factorii de decizie trebuie să cântărească cu atenție aceste compromisuri, să ia în considerare contextul și să facă alegeri în cunoștință de

cauză pe baza nevoilor specifice ale aplicației. Deși perfecțiunea poate fi de neatins, eforturile continue în direcția echității și transparenței sunt esențiale în sistemele decizionale algoritmice.

În cele din urmă, este important să recunoaștem că numeroase aspecte ale echității scapă cuantificării din cauza complexității lor sociotehnice. În consecință, nu toate aspectele echității pot fi încapsulate prin măsurători. (*Învățarea automată și echitatea*, 2021)

Echitate individuală

Echitatea individuală diferă de criteriile de echitate bazate pe grup (parametrii discutați anterior) prin faptul că se concentrează pe tratamentul indivizilor. Acest concept afirmă că persoanele cu caracteristici similare ar trebui să primească rezultate comparabile. Determinarea unui criteriu pentru măsurarea similarității individuale este complexă. Să luăm în considerare trei candidați la un loc de muncă: A cu o diplomă de licență și un an de experiență în domeniu; B cu o diplomă de master și aceeași experiență; și C cu o diplomă de master, dar fără experiență. A decide dacă A seamănă mai mult cu B sau C și a cuantifica această asemănare este o provocare fără a putea compara performanțele lor profesionale, ceea ce nu este posibil dacă nu sunt angajați toți trei. (Zhong, 2018)

9. Implementarea modelului

Informarea candidaților cu privire la utilizarea IA în recrutare

În urma unor evaluări interne atente, compania este pregătită să lanseze instrumentul său de screening selectiv pentru rolurile deschise, pentru a colecta feedback-ul inițial al utilizatorilor și al candidaților.

Este esențial ca solicitanții să fie în permanență conștienți de faptul că se angajează cu un sistem de recrutare bazat pe inteligență artificială, ceea ce stimulează încrederea. Conștientizarea cu privire la avantajele instrumentelor AI ar trebui să fie stabilită de la bun început, sporind dorința candidaților de a utiliza tehnologii interactive precum chatbots. În plus, perfecționarea proceselor de recrutare bazate pe IA și asigurarea clarității cu privire la funcționarea sistemului sunt esențiale pentru adoptarea și succesul acestuia în rândul candidaților.

10. Funcționare și monitorizare

După implementare

Deși evaluările interne ale corectitudinii au fost efectuate înainte de lansare, compania trebuie să fie pregătită și pentru audituri externe independente.

De asemenea, este avantajos să se caute în permanență feedback după implementare și să se stabilească canale de implicare continuă pentru părțile interesate. Acest lucru implică informarea și implicarea lucrătorilor în consultări și procese participative pe parcursul întregii implementări a sistemului AI în cadrul organizației. (HLEG

Auditul AI

În timp ce măsurătorile interne ale corectitudinii sunt efectuate înainte de implementare, un audit extern este, de asemenea, ceva pentru care compania trebuie să fie pregătită.

Potrivit lui Ahmed (2020), auditorii care evaluează aplicațiile IA ar trebui să acorde prioritate a două aspecte: asigurarea conformității cu drepturile persoanelor vizate și evaluarea riscurilor tehnologice în materie de învățare automată și securitate cibernetică. Auditul începe cu definirea domeniului său de aplicare, a obiectivelor și a riscurilor legate de IA pentru organizație, documentate de obicei într-o matrice de risc și control în cadrele stabilite.

Riscurile principale implică neconcordanța strategiilor IT cu obiectivele de afaceri, guvernanța ineficientă și structurile de responsabilizare. În ceea ce privește conformitatea, auditorii trebuie să înțeleagă principiile de confidențialitate a datelor și implicațiile acestora asupra drepturilor persoanelor în temeiul unor reglementări precum GDPR. (Ahmed, 2020)

Există mai multe orientări pentru auditul IA. Pe măsură ce IA remodelează întreprinderile și societatea, auditorii trebuie să se asigure că sunt pregătiți să facă față provocărilor emergente și să verifice eficacitatea controalelor și a guvernanței în domeniul IA. (Ahmed, 2020)

11. Concluzii

Această unitate de curs s-a axat pe examinarea prejudecăților algoritmice într-un context practic. Tema a fost explorată prin intermediul unui scenariu ipotetic în care o companie internațională dezvoltă o soluție AI pentru recrutare. Pe măsură ce tehnologiile AI sunt integrate în recrutare, acestea oferă beneficii substanțiale prin sporirea eficienței și creșterea potențială a diversității la locul de muncă prin accentuarea calificărilor în detrimentul caracteristicilor demografice. Cu toate acestea, punerea în aplicare a acestor tehnologii ridică, de asemenea, provocări etice semnificative, în special în identificarea și atenuarea prejudecăților algoritmice.

Sistemele de inteligență artificială pot perpetua în mod involuntar prejudecățile existente în datele lor de instruire. Aceste prejudecăți pot reflecta disparități istorice sau inegalități societale, conducând la practici discriminatorii în procesele de recrutare. Abordarea acestor prejudecăți nu este doar un imperativ etic, ci și o necesitate juridică, deoarece peisajul de reglementare privind IA în recrutare este în continuă evoluție. Organizațiile trebuie să rămână vigilente și proactive în aderarea la aceste standarde pentru a preveni discriminarea.

Pentru a atenua în mod eficient aceste provocări, este esențial să se utilizeze instrumente sofisticate concepute special pentru identificarea și corectarea prejudecăților din sistemele AI. Monitorizarea periodică și auditurile independente sunt esențiale, deoarece acestea ajută organizațiile să identifice prejudecățile emergente și să își ajusteze sistemele de inteligență artificială în consecință.

În plus, menținerea unui echilibru între procesele bazate pe IA și supravegherea umană este esențială. În timp ce inteligența artificială poate simplifica semnificativ procesele de recrutare, elementul uman rămâne indispensabil în interpretarea contextelor pe care inteligența artificială le-ar putea judeca greșit. Supravegherea umană garantează că recrutarea rămâne un proces sensibil și incluziv, care respectă nuanțele experiențelor și valorilor umane.

Pe scurt, deși inteligența artificială în recrutare oferă avantaje considerabile, utilizarea responsabilă a acestei tehnologii este esențială. Organizațiile ar trebui să se concentreze pe identificarea continuă a prejudecăților și pe utilizarea instrumentelor avansate pentru atenuarea acestora. Această abordare echilibrată asigură faptul că procesele de recrutare respectă cele mai înalte standarde de corectitudine și practici etice, permițând companiilor să adopte progresele tehnologice, protejând în același timp diversitatea și echitatea la locul de muncă.

12. Referințe

Ahmed, H. S. A. (2020, 21 decembrie). *Orientări privind auditul pentru inteligența artificială*. ISACA. https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2020/volume-26/auditing-guidelines-for-artificial-intelligence

Institutul Alan Turing. (2022). *Reflectați asupra scopului, poziției și puterii-Turing Commons.* https://alan-turing-institute.github.io/turing-commons/skills-tracks/aeg/chapter5/reflect/

Albaroudi, E., Mansouri, T. și Alameer, A. (2024). O revizuire cuprinzătoare a tehnicilor AI pentru abordarea prejudecăților algoritmice în angajarea forței de muncă. *AI*, *5*(1), 383-404. https://doi.org/10.3390/ai5010019

Albassam, W. A. (2023). Puterea inteligenței artificiale în recrutare: O revizuire analitică a strategiilor actuale de recrutare bazate pe IA. *International Journal of Professional Business Review*, *8*(6), e02089. https://doi.org/10.26668/businessreview/2023.v8i6.2089

ALTAI. (2020). *Lista de evaluare pentru inteligența artificială demnă de încredere.* https://altai.insight-centre.org/

Arivu Recrutare și Consultanță. (2023, 13 iulie). *Pro și contra utilizării inteligenței artificiale (AI) în recrutare | LinkedIn.* https://www.linkedin.com/pulse/pros-cons-using-artificial-intelligence/

Banerjee, S. (2022, 20 februarie). *Big data în sectorul de recrutare | LinkedIn.* https://www.linkedin.com/pulse/big-data-recruitment-sector-shantanu-banerjee-phd/

Bursell, M., & Roumbanis, L. (2024). După algoritmi: Un studiu al judecăților meta-algoritmice și al diversității în procesul de angajare la o mare companie multisite. *Big Data & Society*, *11*(1), 20539517231221758. https://doi.org/10.1177/20539517231221758

CareerExperts. (2023, 24 aprilie). Reducerea prejudecăților la angajare cu ajutorul AI-Powered Job Description Generation. *Career Experts.* https://www.careerexperts.co.uk/management-leadership/ai-powered-job-description

Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., & Cosentini, A. C. (2022). O clarificare a nuanțelor din peisajul metricii corectitudinii. *Scientific Reports*, *12*(1), 4209. https://doi.org/10.1038/s41598-022-07939-1

Chaker, N. (2018, 24 octombrie). *Surse de date privind candidații: Lista de verificare a recrutorului.* https://beamery.com/resources/blogs/candidate-data-sources-the-recruiter-s-checklist

Chen, Z. (2023). Colaborarea dintre recrutori și inteligența artificială: Eliminarea prejudecăților umane în ocuparea forței de muncă. *Cognition, Technology & Work*, *25*(1), 135-149. https://doi.org/10.1007/s10111-022-00716-0

Dennis, J. (2023, 22 mai). *Recrutarea AI: Utilizări, avantaje și dezavantaje 2024*. TechnologyAdvice. https://technologyadvice.com/blog/human-resources/ai-recruiting/

DigitalOcean. (n.red.). *Cum să utilizați inteligența artificială în recrutare: Tehnici și instrumente*. Retrieved April 26, 2024, from https://www.digitalocean.com/resources/article/ai-in-hiring

Recomandat. (2023, 22 decembrie). *Cum folosesc acești 6 lideri onboardingul alimentat cu AI*. Fast Company. https://www.fastcompany.com/90996367/how-leaders-using-ai-powered-onboarding

Fernandes, R. F. (2021). Big Data ca instrument de îmbunătățire a proceselor de recrutare. *E-Journal of International and Comparative LABOUR STUDIES*, *10*(03).

Filtrat. (n.red.). *Ce este inteligența artificială pentru depistarea CV-urilor și cum poate fi benefică pentru întreprinderea dumneavoastră? | Filtered Blog*. Retrieved March 14, 2024, from https://www.filtered.ai/blog/ai-resume-screening-fd

Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions on Information Systems*, *14*(3), 330-347.

Fritts, M., & Cabrera, F. (2021). Algoritmii de recrutare AI și problema dezumanizării. *Ethics and Information Technology*, *23*(4), 791-801. https://doi.org/10.1007/s10676-021-09615-w

Giordano, F., Morelli, N., Götzen, A. D., & Hunziker, J. (2018, iunie). *Harta părților interesate: Un instrument de conversație pentru proiectarea serviciilor publice conduse de oameni*. ServDes2018 - Service Design Proof of Concept, Politecnico di Milano.

Google pentru dezvoltatori. (2022). *Corectitudine: Tipuri de prejudecăți | Machine Learning*. Google for Developers. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias

Harvard John A. Paulson School of Engineering and Applied Sciences. (2023, 6 decembrie). *Cum poate fi eliminată părtinirea din platformele de recrutare bazate pe inteligența artificială?* https://seas.harvard.edu/news/2023/06/how-can-bias-be-removed-artificial-intelligence-powered-hiring-platforms

Harwell, D. (2019a, 7 iulie). FBI, ICE descoperă că fotografiile permiselor de conducere de stat sunt o mină de aur pentru căutările prin recunoaștere facială. *Washington Post.* https://www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches/

Harwell, D. (2019b, 19 decembrie). Studiul federal confirmă părtinirea rasială a multor sisteme de recunoaștere facială, pune la îndoială extinderea utilizării acestora. *Washington Post.* https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/

Heymans, Y. (2022, 7 septembrie). *Învățarea automată în recrutare: A deep dive.* https://www.herohunt.ai/blog/machine-learning-in-recruitment-a-deep-dive

Hidalgo, M. A. S. (2019). *Design of an Ethical Toolkit for the Development of AI Applications* [Delft University of Technology]. http://resolver.tudelft.nl/uuid:b5679758-343d-4437-b202-86b3c5cef6aa

Hoory, L., & Botorff, C. (2022, 7 august). *Ce este o analiză a părților interesate? Tot ce trebuie să știți - Forbes Advisor.* https://www.forbes.com/advisor/business/what-is-stakeholder-analysis/

Hunkenschroer, A. L., & Luetge, C. (2022). Etica recrutării și selecției bazate pe IA: O revizuire și o agendă de cercetare. *Journal of Business Ethics*, *178*(4), 977-1007. https://doi.org/10.1007/s10551-022-05049-6

Javaid, S. (2024, 12 ianuarie). *Recunoașterea facială: Cele mai bune practici și cazuri de utilizare în 2024*. AIMultiple: High Tech Use Cases & Tools to Grow Your Business. https://research.aimultiple.com/facial-recognition/

Köchling, A., & Wehner, M. C. (2020). Discriminat de un algoritm: O revizuire sistematică a discriminării și corectitudinii prin luarea deciziilor algoritmice în contextul recrutării și dezvoltării resurselor umane. *Business Research*, *13*(3), 795-848. https://doi.org/10.1007/s40685-020-00134-w

Kumar, V. (2013). *101 Design Methods-A Structured Approach for Driving Innovation in Your Organization*. John Wiley & Sons, Inc. https://learning.oreilly.com/library/view/101-design-methods/9781118083468/cvi.xhtml

Lange, A. R., & Duarte, N. (2018, 4 aprilie). Înțelegerea prejudecăților în proiectarea algoritmică. *ASME ISHOW / IDEA LAB.* https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Lawton, G. (2022, 29 august). *Prejudecățile AI la angajare: Tot ce trebuie să știți | TechTarget*. Software HR. https://www.techtarget.com/searchhrsoftware/tip/AI-hiring-bias-Everything-you-need-to-know

Lee, N. T., Resnick, P. și Barton, G. (2019, 22 mai). *Detectarea și atenuarea prejudecăților algoritmice: Cele mai bune practici și politici pentru reducerea prejudiciilor aduse consumatorilor | Brookings.* https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Ma, Y. (2024). Metodă inteligentă de recomandare a resurselor de talente bazată pe Big Data. *Intelligent Computing Technology and Automation.* https://doi.org/10.3233/ATDE231176

Învățarea automată și corectitudinea. (2021, 24 mai). Webinarii și tutoriale Microsoft Research. https://youtu.be/ZtN6Qx4KddY?si=kWo9MxkNQ7ko1HUB

Marr, B. (2023, 12 decembrie). *AI-Enhanced Employee Onboarding: A New Era In HR Practices*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/12/12/ai-enhanced-employee-onboarding-a-new-era-in-hr-practices/

Naakka, S.-S. (2018). *BIG DATA KILPAILUETUNA SUORAHAKU- KONSULTOINNIN EHDOKASHAUISSA Enemmän, nopeammin ja parempia IT-osaajia big datalla?* [Universitatea din Turku]. https://www.utupub.fi/bitstream/handle/10024/145309/Naakka%20Sara-Selia.pdf?sequence=1&isAllowed=y

Ongig. (2024). *Descrierea postului Software | Ongig.* https://www.ongig.com
ResumeMent. (2023, 16 octombrie). *From Resumes to Algorithms: Rolul inteligenței artificiale în selectarea candidaților | LinkedIn.* https://www.linkedin.com/pulse/from-resumes-algorithms-role-ai-screening-candidates-resumement/
Saplicki, C. (2022, 11 noiembrie). Explicarea echității: Definiții și metrici. Medium. https://medium.com/ibm-data-ai/fairness-explained-definitions-and-metrics-9690f8e0a4ea
SDT. (n.red.). *Instrumente | Instrumente de proiectare a serviciilor*. Retrieved April 18, 2024, from https://servicedesigntools.org/tools.html
Sheard, N. (2022). Discriminarea la angajare prin algoritm: Poate fi cineva tras la răspundere? *University of New South Wales Law Journal*, *45*(2). https://doi.org/10.53637/XTQY4027
Shipman, M. (2021, 4 martie). Verificările din social media pot aduce prejudecăți la angajare. *Futurity.* https://www.futurity.org/cybervetting-human-resources-bias-2527382-2/
Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018a). *This is Service Design Doing*. O'Reilly Media, Inc. https://learning.oreilly.com/library/view/this-is-service/9781491927175/ch03.html
Stickdorn, M., Hormess, M., Lawrence, A., & Schneider, J. (2018b, iulie). *Biblioteca de metode #TISDD*. Aceasta este proiectarea serviciilor. https://www.thisisservicedesigndoing.com/methods
Sukernek, W. (2024). *Cum să depistați prejudecățile la angajare*. FloCareer. https://blog.flocareer.com/how-to-spot-bias-in-hiring
Teodorescu, M. (2020). *Criterii de corectitudine | Explorarea corectitudinii în învățarea automată pentru dezvoltarea internațională | Resurse suplimentare*. MIT OpenCourseWare. https://ocw.mit.edu/courses/res-ec-001-exploring-fairness-in-machine-learning-for-international-development-spring-2020/pages/module-three-framework/fairness-criteria/
UNESCO. (2023). *Evaluarea impactului etic. Un instrument al Recomandării privind etica inteligenței artificiale*. UNESCO. https://doi.org/10.54678/YTSA7796
Vivek, R. (2023). Îmbunătățirea diversității și reducerea prejudecăților în recrutare prin intermediul IA: o analiză a strategiilor și provocărilor. *Информатика. Экономика. Управление - Informatică. Economie. Management*, *2*(4), 0101-0118. https://doi.org/10.47813/2782-5280-2023-2-4-0101-0118
Wallbridge, A. (2023, 13 aprilie). *Cum să efectuați o matrice de analiză a părților interesate Bullet-Proof în 4 pași utili*. TSW Training. https://www.tsw.co.uk/blog/leadership-and-management/stakeholder-analysis-matrix/
Wirtz, D. (2022, 3 august). *Ce este un atelier de lucru? (+2 exemple) | Facilitator School.* https://www.facilitator.school/blog/what-is-a-workshop
www.recruiter.com. (2023, 26 aprilie). *Pro și contra utilizării AI în recrutare*. Recruiter.Com. https://www.recruiter.com/recruiting/pros-and-cons-of-using-ai-in-recruiting/

YLE. (2023, 11 noiembrie). *Loimme 13-vuotiaan Ellan - Tiktok tarjosi hänelle sisältöä itsemurhasta ja kalorien laskemisesta*. Yle Uutiset. https://yle.fi/a/74-20059318

Zhong, Z. (2018, 22 octombrie). *Un tutorial privind corectitudinea în învățarea automată*. Medium. https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb

Charlie

Cofinanțat de Uniunea Europeană

Finanțat de Uniunea Europeană. Punctele de vedere și opiniile exprimate aparțin, însă, exclusiv autorului (autorilor) și nu reflectă neapărat punctele de vedere și opiniile Uniunii Europene sau ale Agenției Executive Europene pentru Educație și Cultură (EACEA). Nici Uniunea Europeană și nici EACEA nu pot fi considerate răspunzătoare pentru acestea.

2022-1-ES01-KA220-HED-000085257