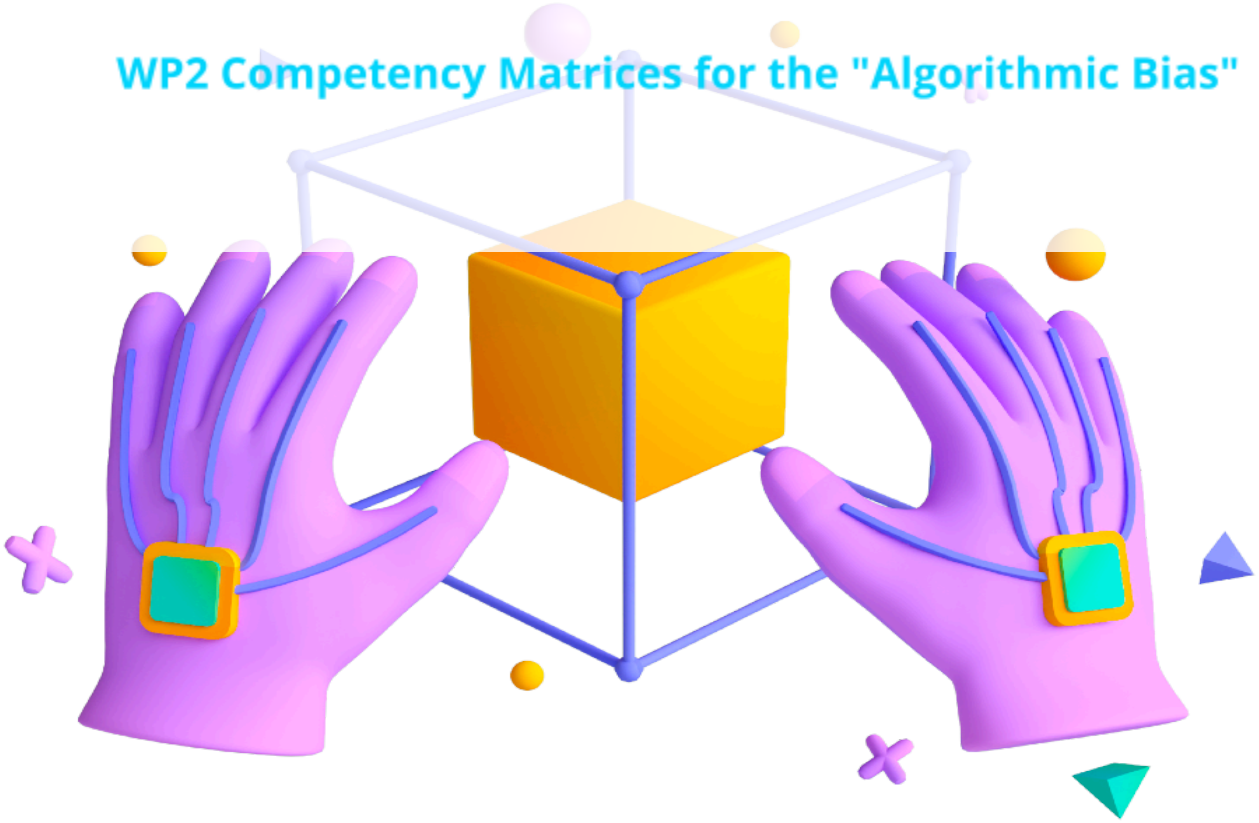
CHALLENGING BIAS IN BIG DATA USER FOR AI AND MACHINE LEARNING

WP2 Competency Matrices for the "Algorithmic Bias"



Co-funded by the European Union

Număr proiect: 2022-1-ES01-KA220-HED-000085257

Conținut

1. INTRODUCERE: PROIECTUL CHARLIE

1.1. Proiectul Charlie: prezentare generală

CHARLIE este un proiect ERASMUS+ KA2 cu o perioadă de implementare de 30 de luni, în perioada 30/12/2022 - 29/06/2025. Proiectul este condus de un consorțiu de șase (6) parteneri din cinci (5) țări europene: Spania, Portugalia, România, Finlanda și Danemarca.

Inteligența artificială (IA) face parte din viața noastră de zi cu zi. De la algoritmul care ne recunoaște fața atunci când mergem pe stradă hrănind serviciile de securitate biometrică a poliției până la algoritmul care alege publicitatea pe care o vom vedea în rețelele noastre sociale, AI este peste tot. Cu toate acestea, deși Machine Learning (ML) și AI sunt matematice, ele sunt doar uneori potrivite, iar acest lucru se întâmplă deoarece datele procesate pentru a ajunge la orice concluzie pot fi și adesea sunt părtinitoare. Științele sociale au studiat prejudecățile umane de mulți ani. Ea decurge din asocierea implicită care reflectă prejudecățile de care nu suntem conștienți, care poate avea ca rezultat multiple rezultate adverse. AI și ML nu sunt concepute pentru a lua decizii etice; acesta nu este un algoritm pentru etică. Acesta va face întotdeauna predicții bazate pe modul în care funcționează lumea de astăzi, contribuind, prin urmare, la promovarea părtinirii și a practicilor discriminatorii înrădăcinate sistemic în societățile noastre de astăzi. Odată cu utilizarea pe scară largă a tehnologiilor AI și ML, adesea deținute de mari companii de tehnologie cu singurul obiectiv de a obține profit, există o nevoie urgentă de a aduce o abordare centrată pe om a tehnologiei și de a o folosi pentru a rezolva problemele sociale în loc să contribuim la ele. . În Comunicarea sa din 25/04/18 și 7/12/18, CE și-a prezentat viziunea pentru inteligența artificială, care sprijină „IA etică, sigură și de ultimă generație, fabricată în Europa”. Sistemele AI trebuie să fie centrate pe om, bazate pe angajamentul de a le folosi în serviciul umanității și al binelui comun pentru a îmbunătăți bunăstarea și libertatea umană. Deși oferă oportunități semnificative, sistemele AI dau naștere și la anumite riscuri care trebuie gestionate în mod corespunzător și proporțional. Acum avem o fereastră esențială de oportunitate pentru a le modela dezvoltarea.

CHARLIE intenționează să se asigure că putem avea încredere în mediile sociotehnice în care acestea sunt încorporate. De asemenea, dorim ca producătorii de sisteme AI să obțină un avantaj competitiv prin integrarea AI de încredere în produsele și serviciile lor. Aceasta implică căutarea de a maximiza beneficiile sistemelor AI, prevenind și minimizând în același timp riscurile acestora. Învățământul superior (HE), Educația adulților (AE) și Tineretul necesită programe de învățământ noi și inovatoare care să poată acoperi acest deficit de competențe și să doteze cursanții cu cunoștințele și

abilitățile pentru a contribui la o abordare mai etică a dezvoltării tehnologiei. Necesitatea de a face educația tehnologică mai umană este aliniată cu Planul de acțiune pentru educația digitală, care include acțiuni specifice pentru a aborda implicațiile și provocările etice ale utilizării IA și a datelor în educație și formare.

1.2 Obiectivele proiectului

CHARLIE își propune să conteste părtinirea datelor mari utilizate pentru inteligența artificială și învățarea automată, aducând un nivel mai mare de conștientizare cu privire la impacturile negative ale lipsei unei abordări critice și etice a tehnologiei . Obiectivele principale ale lui CHARLIE sunt:

1. Creșterea capacității instituțiilor de învățământ superior de a oferi studenților săi oportunități de învățare online care să răspundă nevoilor societății, dar și adaptate nevoilor de învățare ale studenților;
2. Creșterea competențelor sociale și etice ale studenților Tech, permițându-le să se implice pozitiv, critic și etic cu tehnologia AI/ML;
3. Echipați profesorii/profesorii cu abordări digitale și captivante pentru predarea eficientă a subiectului (în special în predarea online);
4. Crearea de sinergii între organizațiile HE și AE și Tineret la nivel regional în domeniul educației pentru etică IA;
5. Potențarea transferabilității cursurilor din mediul academic privind prejudecățile IA către EA și Educație și Formare Profesională (VET);
6. Creșterea gradului de conștientizare cu privire la subiect la nivel de societate

1.3. Rezultatele generale așteptate ale proiectului Charlie pe WP

1- Matrice de competențe pentru cursul „Algorithmic Bias" (EQF6), „Ethical AI microcredential " (EQF4) și Serious Game (EQF2) **(WP2)**

2- Curs „Algorithmic Bias" (HE) - abordare bLearning : a) sesiuni sincrone (de preferință față în față), b) sesiuni eLearning asincrone pentru componenta teoretică, completate cu un serios joc digital ca evaluare formativă și evaluare sumativă online, livrate ca . scorm gata să fie încărcat în orice LMS pentru studenții înscriși la cursuri legate de Big Data, AI, machine learning, deep learning. **(WP3)**

3- „Setul de instrumente Algorithmic Bias pentru sesiuni sincrone" - pentru profesorii/profesorii/lectorii care implementează sesiunile sincrone (față în față sau online) care completează sesiunile asincrone de eLearning. **(WP3)**

4- Un ghid pentru creșterea capacității administratorilor universităților/managementului departamentelor de educație și formare

responsabile cu furnizarea de educație digitală pentru a stimula livrarea educației digitale pe baza algoritmică. **(WP3)**
5- Seminarii web pentru a promova învățarea între egali și a discuta rolul instituțiilor de învățământ superior în techEd , pentru a promova interdisciplinaritatea științelor sociale în educația tehnologică și pentru a discuta abordări pentru interoperabilitatea cu Tineretul și AE. **(WP3)**
6- Microcredital „Ethical AI" auto-ritm (EQF4) pentru cursanții adulți să fie luați online și asincron. **(WP4)**
7- Joc digital serios (EQF2) pentru tineri 12-18 ani (în principal femei defavorizate/tinere) pentru a promova reprezentarea de gen în STEM încă de la o vârstă fragedă. **(WP4)**
8- Set de instrumente pentru a sprijini adulții și tinerii în perfecționarea în IA etică. **(WP4)**
9- Recomandare de politică pentru recunoașterea microcrențialului pentru cursanții adulți în accesarea cursurilor de învățământ superior în domeniile tehnologice. **(WP4)**

1.4. Grupuri țintă Charlie

Principalele grupuri țintă ale proiectului CHARLIE sunt:
- **Instituții de învățământ superior** care oferă Big Data, AI, învățare automată, oportunități/cursuri de învățare legate de deep learning și administratorii/managementul departamentelor de Educație și Formare.
- **Studenți HE** înscriși la cursuri legate de Big Data, AI, machine learning, deep learning
- **Profesori/Profesori/Lectori** de Î.S. din științe sociale și techEd
- **Organizațiile AE** și personalul acestora
- **Adulți** de toate vârstele și mediile socio-economice, cu scopul de a progresa către niveluri de calificare mai înalte relevante pentru piața muncii și pentru participarea activă în societate.
- **Organizațiile de tineret** și personalul acestora
- **Tineri** (12 - 18 ani) - în special femei tinere și tineri din contexte social dezavantajate (care se confruntă cu bariere legate de sistemul educațional; diferențe culturale; bariere sociale; bariere economice; bariere legate de discriminare; și bariere geografice)

Principalele grupuri țintă ale activităților și produselor de comunicare și promovare a proiectelor:
- Autoritățile naționale de reglementare și factorii de decizie;
- Publicul larg

- Instituții de învățământ superior (lectori universitari, studenți (licenți și postuniversitari) și universități
administratori)
- Furnizori de educație pentru adulți
- Organizații de tineret
- Furnizorii VET
- Forumuri și platforme AI
- ONG-uri care abordează diferite prejudecăți și discriminare
- Furnizori de tehnologie educațională
- Operatori mass-media (TV, ziare electronice, radiouri etc.)
- IMM-urile AI/ML și întreprinderile AI/ML.
- Asociații de furnizori de învățământ superior, inclusiv organizații colaborative conexe pentru promovarea celor mai bune practici la nivelul UE și diseminarea bulgăre de zăpadă către membrii universităților lor (de exemplu: Asociația Europeană a Universităților, Rețeaua de Afaceri Științifice, Asociația Triple Helix, AMBA, Rețeaua de Inovare Universitară-Industrie, Afaceri Europene Angels Network (pentru investiții în inovații etice AI)
- Corpuri studențești la nivel european
- Asociații de monitorizare/dezvoltare a competențelor (ex: comunități DigCompEdu).

1.5. Obiectivele WP2

Matricele de competențe pentru obiectivele specifice WP „Algorithmic Bias" sunt:

- să stabilească o înțelegere comună cu privire la scopul și obiectivele de învățare ale unui traseu de învățare care va fi dezvoltat pentru studenții de ÎS în cadrul cursului „Algorithmic Bias";

- să stabilească o înțelegere comună cu privire la scopul și obiectivele de învățare ale unui traseu de învățare care va fi dezvoltat pentru cursanții EA în „ microcredential AI etică ";

- să stabilească o înțelegere comună cu privire la scopul și obiectivele de învățare ale unui traseu de învățare care va fi dezvoltat pentru tineri în jocul serios pe IA etică.

Aceste obiective contribuie la obiectivele generale prin stabilirea pietrelor de temelie pentru următoarele activități, permițând profesorilor, mentorilor și cursanților să aibă o viziune clară asupra obiectivelor diferitelor căi de învățare, lucrând pentru creșterea capacității organizațiilor de învățământ superior, de

adulți și de tineret de a oferi oportunități de învățare. care răspund nevoilor societății, dar sunt, de asemenea, adaptate nevoilor de învățare ale elevilor.

1.6. Rezultatele așteptate ale activităților WP2 și ale participanților

1.6.1. Competency Matrix EQF 6 pentru cursul „Algorithmic Bias" (studenți HE) - Abordarea rezultatelor învățării, 2 puncte ECVS, recomandări de acreditare, cerințe preliminare pentru înscriere, ore de contact, volum total de muncă, integrarea competențelor EntreComp (de exemplu : Gândire etică și durabilă), competențele DigComp 2.0 (de exemplu: Protejarea sănătății și bunăstării) și competențele GrenComp (de exemplu: Sprijinirea echității).
Unități de competență:
CU1 - Algoritmi Modele și Limitări UIB
CU2 - Corectitudinea și părtinirea datelor în AI UA
CU3 - AI Confidențialitate și comoditate UA
CU4 - AI Ethics, o abordare practică VAMK
CU5 - Studii de caz și proiect VAMK

1.6.2. Matricea de competențe EQF 4 pentru „Micro credential AI etic" - Abordarea rezultatelor învățării, potențialul de puncte ECVS, recomandări de acreditare, cerințe preliminare pentru înscriere, ore de contact, volumul total de muncă, integrarea competențelor EntreComp (de exemplu : Gândire etică și durabilă), Competențe DigComp 2.0 (de exemplu: Protejarea sănătății și bunăstării) și competențe GrenComp (de exemplu: Sprijinirea echității).
Unități de competență:
CU1 - Ce este bias algoritmic? HELIX
CU2 - HELIX de non- maleficiență
CU3 – Responsabilitate ITC
CU4 – Transparență ITC
CU5 - Drepturile omului și corectitudinea ITC
CU6 - AI Ethics, o abordare practică HELIX

1.6.3. Rezultate ale învățării pentru tineret EQF 2 (joc serios) - Prezintă abordarea rezultatelor învățării, cerințe prealabile pentru înscriere, ore de contact, integrarea competențelor EntreComp (de exemplu: gândire etică și durabilă), competențe DigComp 2.0 (de exemplu: protejarea sănătății și bunăstării) și competențele GrenComp (de exemplu: Sprijinirea echității). Conținuturi generale: Inegalități sistemice în societate, Mecanica de bază a sistemelor de inteligență artificială, agende politice, impacturi AI în lume.

2. INTRODUCERE ÎN TEMA: ALGORITMI, AI, PRINCIPALELE ȘI ETICA

Algoritmii joacă un rol crucial în societatea modernă, deoarece ajută cu diverse servicii, cum ar fi recomandarea de cântece, filme sau prieteni (Paraschakis , 2018; Milano et al., 2020). Ele sunt, de asemenea, utilizate în școli, spitale, instituții financiare, instanțe și guverne pentru a lua decizii cruciale (Obermeyer et al., 2019). Cu toate acestea, utilizarea algoritmilor poate duce la riscuri etice (Floridi et al., 2018).

Algoritmii, cum ar fi traducerile (Prates et al., 2020) sau anunțurile de angajare (Lambrecht și Tucker, 2019), pot fi părtinitoare. De exemplu, locurile de muncă mai bine plătite pot fi arătate bărbaților mai mult decât femeilor. Acest lucru poate duce, de asemenea, la un tratament inechitabil în asistența medicală, pacienții albi primind îngrijiri mai bune decât pacienții de culoare (Obermeyer et al. 2019). În timp ce oamenii încearcă să remedieze aceste probleme, mai multe probleme continuă să apară.

Din 2012, s-a pus mult accent pe inteligența artificială (AI) și implicațiile sale etice (Perrault et al. 2019). În 2016, un studiu a încercat să cartografieze aceste preocupări (Mittelstadt et al. 2016), dar domeniul a crescut și s-a schimbat foarte mult de atunci. Guvernele, organizațiile și companiile s-au implicat mai mult în discutarea despre AI și algoritmi „echitabile” și „etice” (Wong 2019).

Prejudecățile în tehnologie pot duce la efecte dăunătoare, chiar dacă neintenționate, și pot crea provocări pentru construirea încrederii publicului în AI. Trebuie să luăm în considerare factorii specifici ai inteligenței artificiale și impacturile potențiale asupra societății pentru a ne asigura că este sigură și sigură. Eforturile actuale de a aborda prejudecățile AI se concentrează pe aspecte computaționale, cum ar fi reprezentativitatea datelor și corectitudinea în algoritmii de învățare automată. Cu toate acestea, factorii umani, instituționali și societali contribuie și ei la părtinirea AI și sunt adesea trecuți cu vederea. Pentru a aborda cu succes această provocare, trebuie să ne extindem perspectiva și să examinăm modul în care IA este creată și afectează societatea.

AI de încredere și responsabilă nu se referă doar la faptul că un sistem AI este părtinitor, corect sau etic, ci și dacă face ceea ce pretinde. Practici precum transparența, seturile de date și testarea, evaluarea, validarea și verificarea (TEVV) sunt esențiale. Factorii umani, tehnicile de proiectare participativă, abordările cu mai multe părți interesate și „human-in-the-loop” ajută, de asemenea, la atenuarea riscurilor de părtinire a AI. Cu toate acestea, aceste

practici singure nu oferă o soluție completă. Avem nevoie de îndrumare dintr-o perspectivă socio-tehnică mai largă, care conectează aceste practici cu valorile societale.

Experții în IA de încredere și responsabilă recomandă operaționalizarea acestor valori și crearea de noi norme de dezvoltare și implementare a AI. Acest document, împreună cu lucrările desfășurate de partenerii proiectului CHARLIE privind prejudecățile algoritmilor și AI, adoptă o perspectivă socio-tehnică.

Prejudecățile nu sunt unice pentru AI, iar atingerea unui risc zero de părtinire în sistemele AI este imposibilă. Proiectul CHARLIE intenționează să dezvolte resurse și metode pentru creșterea îmbunătățirilor practice pentru identificarea, înțelegerea, măsurarea, gestionarea și reducerea părtinirii. Pentru a atinge acest obiectiv, avem nevoie de tehnici flexibile care pot fi aplicate în diferite contexte și comunicate diferitelor grupuri de părți interesate.

3. MATRIZA COMPETENȚELOR: O ABORDARE TEORETICĂ

O matrice de competențe, cunoscută și sub denumirea de matrice de competențe sau cadru de competențe, este un instrument utilizat pentru a identifica, mapa și evalua abilitățile, cunoștințele și abilitățile indivizilor într-o organizație sau instituție de învățământ (Stewart și Bartrum, 1999). Reprezintă vizual competențele necesare pentru anumite roluri, proiecte, sarcini sau programe academice. Ajută la identificarea nivelurilor actuale de competențe ale angajaților, membrilor echipei, studenților sau altor părți interesate (Cottrell, 2018).

Caracteristicile unei matrice de competențe

O matrice de competențe este caracterizată de mai multe caracteristici cheie:

- Format grilă sau tabel: o matrice de competențe este de obicei structurată ca o grilă sau tabel, cu rânduri reprezentând indivizi și coloane reprezentând competențele sau abilitățile necesare (Mansfield, 2009).

- Evaluări sau niveluri de competență: Fiecare celulă din matrice conține o evaluare sau un nivel de competență pentru calificarea și individul corespunzătoare. Sistemul de evaluare poate fi numeric, codat pe culori sau bazat pe termeni descriptivi precum „începător", „intermediar" sau „expert" (Cottrell, 2018).

- Personalizabile și adaptabile: Matricele de competențe pot fi personalizate pentru a se potrivi nevoilor specifice ale unei organizații sau instituții de învățământ și pot fi actualizate cu ușurință pentru a reflecta schimbările în roluri, responsabilități sau cerințe de competențe (Stewart și Bartrum, 1999).

- Cuprinzătoare: O matrice de competențe ar trebui să acopere toate competențele relevante necesare pentru un anumit rol, proiect sau program academic și ar trebui să includă atât abilități tehnice, cât și abilități soft (Mansfield, 2009).

Obiectivele principale ale unei matrice de competențe

O matrice de competențe servește mai multor obiective importante în cadrul unei organizații sau instituții de învățământ:

- Identificarea lacunelor de competențe: Oferind o imagine de ansamblu clară a abilităților și competențelor necesare pentru un anumit rol sau proiect, o matrice de competențe poate ajuta la identificarea domeniilor în care indivizii ar putea avea nevoie de dezvoltare sau formare ulterioară (Stewart și Bartrum, 1999).

- Sprijinirea dezvoltării angajaților sau studenților: o matrice de competențe poate fi utilizată pentru a ghida planurile de dezvoltare personală și pentru a informa inițiativele de formare și dezvoltare, ajutând indivizii să-și identifice punctele forte și punctele slabe și să stabilească obiective de îmbunătățire (Cottrell, 2018).

- Optimizarea alocării resurselor: O matrice de competențe poate ajuta organizațiile și instituțiile de învățământ să aloce resursele mai eficient, prin identificarea celor mai potrivite persoane pentru roluri sau proiecte specifice pe baza abilităților și competențelor lor (Mansfield, 2009).

- Îmbunătățirea performanței: Ajutând la asigurarea că indivizii au abilitățile și competențele adecvate pentru rolurile lor, o matrice de competențe poate contribui la îmbunătățirea performanței generale în cadrul unei organizații sau instituții de învățământ (Stewart și Bartrum, 1999).

- Facilitarea comunicării și colaborării: o matrice de competențe poate servi ca un limbaj comun pentru discutarea abilităților și competențelor în cadrul unei organizații sau instituții de învățământ, promovând o mai bună înțelegere și colaborare între membrii echipei, departamente sau unități academice (Cottrell, 2018).

Aplicații ale matricei de competențe în diferite contexte

Matricele de competențe au fost aplicate în diverse contexte, inclusiv în afaceri, guvern și instituții de învățământ superior, pentru a atinge o serie de obiective:

- Planificarea forței de muncă: Matricele de competențe pot fi utilizate pentru a sprijini planificarea strategică a forței de muncă, prin identificarea abilităților și competențelor necesare pentru atingerea scopurilor și obiectivelor organizaționale și asigurându-se că acestea sunt reflectate în inițiative de recrutare, formare și dezvoltare (Stewart și Bartrum, 1999). .

- Managementul performanței: Matricele de competențe pot fi integrate în sistemele de management al performanței, oferind un cadru pentru stabilirea așteptărilor de performanță, efectuarea evaluărilor performanței și identificarea zonelor de îmbunătățire (Mansfield, 2009).

- Planificarea succesiunii: Matricele de competențe pot sprijini eforturile de planificare a succesiunii, prin identificarea abilităților și competențelor necesare pentru rolurile de conducere și ajutând la identificarea și dezvoltarea potențialilor succesori în cadrul unei organizații sau instituții de învățământ (Cottrell, 2018).

- Planificarea și Dezvoltarea Curriculumului: În instituțiile de învățământ superior, matricele de competențe pot fi utilizate pentru a ghida planificarea și dezvoltarea curriculumului, asigurându-se că programele academice sunt concepute pentru a dezvolta abilitățile și competențele cerute de studenți pentru a reuși în domeniile alese de ei (Billett, 2009). Acesta și următoarele trei sunt orientarea principală a prezentului document.

- Dezvoltarea facultăților și a personalului: Matricele de competențe pot fi aplicate pentru a sprijini dezvoltarea profesională a facultăților și a personalului din instituțiile de învățământ superior, prin identificarea abilităților și competențelor necesare pentru predare, cercetare și roluri administrative eficiente și informând proiectarea formării și dezvoltării. programe (Eraut, 2004).

- Formarea echipei de cercetare: Matricele de competențe pot fi utilizate pentru a facilita formarea echipelor de cercetare în instituțiile de învățământ superior, prin identificarea persoanelor cu abilități și competențe complementare și promovarea colaborării interdisciplinare (Börner et al., 2010).

- Consiliere și sprijin pentru studenți: Matricele de competențe pot fi utilizate pentru a sprijini serviciile de consiliere și sprijin pentru studenți în instituțiile de învățământ superior, ajutând consilierii să identifice punctele forte și punctele slabe ale studenților și să-i ghideze către resursele, intervențiile sau căile academice adecvate (Cuseo, 2010) .

Provocări și limitări ale matricei de competențe

Deși matricele de competențe pot oferi informații valoroase și pot sprijini o serie de obiective strategice, ele prezintă, de asemenea, unele provocări și limitări:

- Subiectivitatea: Evaluările competențelor pot fi subiective și pot fi influențate de factori precum prejudecățile personale, diferențele culturale sau variațiile în interpretarea definițiilor competențelor (Stewart și Bartrum, 1999).

- Consum intensiv de timp și resurse: Dezvoltarea și menținerea unei matrice de competențe cuprinzătoare poate necesita timp și resurse , în special în organizațiile mari sau instituțiile de învățământ cu roluri și responsabilități diverse (Mansfield, 2009).

- Accentuarea excesivă asupra abilităților tehnice: Matricele de competențe pot pune un accent mai mare pe abilitățile și competențele tehnice, în detrimentul abilităților soft, cum ar fi comunicarea, munca în echipă sau rezolvarea de probleme, care sunt adesea esențiale pentru succes în multe roluri și contexte (Cottrell, 2018).

- Rigiditate: Matricele de competențe pot fi percepute ca rigide și inflexibile, limitând potențial creativitatea, inovația sau adaptabilitatea în cadrul unei organizații sau instituții de învățământ (Eraut, 2004).

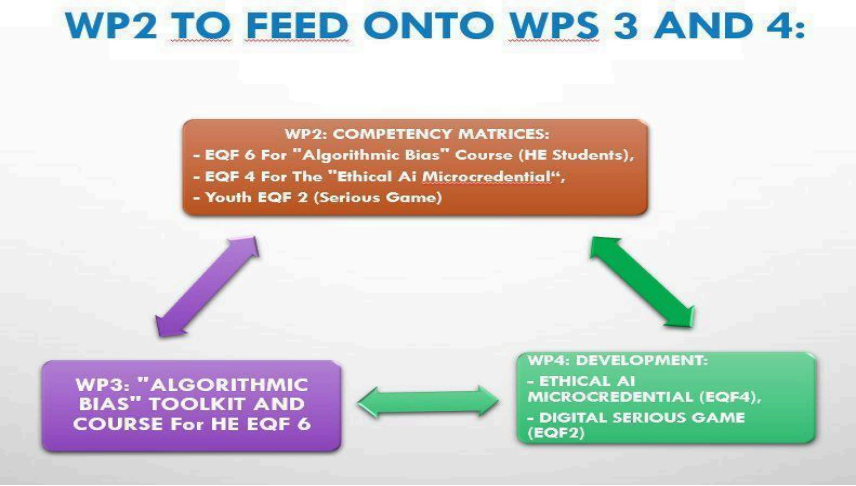
În ciuda acestor provocări și limitări, matricele de competențe pot servi ca un instrument valoros pentru organizațiile și instituțiile de învățământ care doresc să optimizeze alocarea resurselor, să sprijine dezvoltarea individuală și să îmbunătățească performanța generală. Prin adoptarea unei abordări sistematice și cuprinzătoare pentru identificarea, maparea și evaluarea abilităților și competențelor și încorporând o serie de perspective și părți interesate în dezvoltarea și întreținerea matricei, organizațiile și instituțiile de învățământ pot maximiza beneficiile acestei abordări minimizând în același timp potențialele dezavantaje. (Billett, 2009).

4. MATRICE CHARLIE WP2

Matricea de competențe a cursului „Algorithmic Bias” EQF6 servește ca element de bază pentru alte două componente critice: matricea de competențe EQF4 „Ethical AI Micro Credential” și rezultatele învățării „Rezultatele învățării pentru adulți și tineri EQF2 (Serious Game)”. Aceste trei componente sunt interconectate și concepute să se completeze și să se consolideze reciproc, asigurând o înțelegere și aplicare cuprinzătoare a principiilor etice AI la diferite niveluri și contexte educaționale.

Matricea de competențe a cursului „Algorithmic Bias” pune bazele pentru dezvoltarea matricei de competențe EQF4 „Ethical AI Micro Credential”. Oferind o bază solidă în înțelegerea prejudecăților algoritmice și a implicațiilor acestora, acest curs dă putere cursanților să aprofundeze dimensiunile etice ale AI pe măsură ce progresează la nivelul EQF4. Matricea de competențe „Ethical AI Micro Credential” EQF4 se bazează pe competențele inițiale, extinzând domeniul de aplicare pentru a include considerații etice mai largi și aplicații practice ale principiilor etice AI în diferite sectoare și scenarii.

Simultan, matricea de competențe a cursului „Algorithmic Bias” informează și rezultatele învățării „Rezultatele învățării pentru adulți și tineri EQF2 (Serious Game)”. Această conexiune asigură că conceptele și abilitățile fundamentale dezvoltate în curs sunt accesibile și captivante pentru cursanții la nivelul EQF2. Formatul jocului serios este conceput pentru a introduce aceste subiecte critice într-un mod atât distractiv, cât și educațional, favorizând o înțelegere mai profundă a părtinirii algoritmice și a principiilor etice AI în rândul unui public mai larg.



Graficul 1. Interconectarea dintre WP și rezultate ale proiectului

În concluzie, cele trei elemente – matricea de competențe ale cursului „Algorithmic Bias", matricea de competențe „Ethical AI Micro Credential" EQF4 și rezultatele învățării „Learning Outcomes for Adults and Youth EQF2 (Serious Game)" – sunt strâns legate, deoarece înfățișat. Această structură interconectată permite o abordare cuprinzătoare, pe mai multe niveluri, pentru a învăța despre părtinirea algoritmică și IA etică, asigurând că cursanții din diferite medii educaționale și grupe de vârstă pot dezvolta cunoștințele și abilitățile necesare pentru a naviga provocările și oportunitățile AI într-un mod etic și responsabil. manieră. În plus, rezultatele WP2 se vor alimenta direct în WP3 și WP4, ilustrând în continuare relația strânsă dintre aceste componente, așa cum se arată în Graficul 1.

Această integrare perfectă a WP2, WP3 și WP4 are ca rezultat un cadru educațional solid și coeziv, care abordează aspectele cruciale ale prejudecăților algoritmice și ale inteligenței artificiale etice, favorizând o înțelegere și aplicare completă a acestor principii în diverse contexte și profiluri de cursanți.

4.1. CURS " *PARZIUNEA ALGORITMICĂ* " EQF6

Acest program educațional aprofundat este conceput pentru a oferi o înțelegere cuprinzătoare a prejudecăților algoritmice și a implicațiilor acestora în societatea actuală bazată pe date. Întrucât algoritmii joacă un rol semnificativ în modelarea vieții noastre și în informarea luării deciziilor în diferite sectoare, inclusiv în domeniul sănătății, educației, finanțelor și guvernului, este esențial pentru profesioniști și cadre universitare/studenți să dezvolte o înțelegere solidă a potențialelor părtiniri în cadrul acestor sisteme.
Obiectivul acestui curs este de a dota participanții cu cunoștințele și abilitățile necesare pentru a identifica, atenua și aborda prejudecățile algoritmice. Printr-o combinație de fundamente teoretice și aplicații practice, cursanții vor obține informații despre aspectele etice, sociale și tehnice ale părtinirii algoritmice.
Aceste obiective contribuie la obiectivele generale prin stabilirea pietrelor de temelie pentru următoarele activități, permițând profesorilor, mentorilor și cursanților să aibă o viziune clară asupra obiectivelor diferitelor căi de învățare, lucrând pentru a crește capacitatea organizațiilor de învățământ superior, de adulți și de tineret de a oferi oportunități de învățare. care răspund nevoilor

societății, dar sunt, de asemenea, adaptate nevoilor de învățare ale elevilor. Concret, rezultatele așteptate ale acestui WP vizează:

1) creșterea capacității instituțiilor de învățământ superior de a oferi studenților săi oportunități de învățare online care să răspundă nevoilor societății, dar și care sunt adaptate nevoilor de învățare ale elevilor - instituțiile de învățământ superior beneficiind de abordarea rezultatelor învățării, în care profesorii și studenții au o bază comună cu privire la rezultatele așteptate la sfârșitul unui parcurs de învățare;
2) creșterea competențelor sociale și etice ale studenților tehnologici, permițându-le să se implice pozitiv, critic și etic cu tehnologia AI/ML - studenții HE beneficiind de abordarea rezultatelor învățării în care știu clar ce se așteaptă de la ei la încheierea traseului de învățare ;
3) dotați profesorii/profesorii cu abordări digitale și captivante pentru predarea eficientă a temei (în special în predarea online) - deoarece abordarea cursantului este punctul de plecare pentru proiectarea experiențelor de învățare ;
4) să creeze sinergii între instituțiile de învățământ superior și EA și Tineret la nivel regional/local în domeniul educației etice IA – prin stabilirea rezultatelor învățării pentru OER complementare care vizează niveluri diferite (EQF 6, 4, 2) în același domeniu de cunoaștere;
5) potențarea transferabilității cursurilor academice privind prejudecățile IA către EA și Tineret - prin stabilirea unui punct de contact în diferitele rezultate ca bază pentru un teren comun de discuție privind potențialul de transferabilitate a căilor de învățare;
6) creșterea gradului de conștientizare cu privire la subiect la nivel de societate - deoarece stabilește primul pas pentru a aduce OER-uri dedicate la diferite niveluri ale societății și grupuri țintă.

La finalizarea acestui curs, participanții vor fi dezvoltat o înțelegere cuprinzătoare a distorsiunilor algoritmice, a impactului lor potențial asupra diferitelor sectoare și a strategiilor și instrumentelor disponibile pentru a aborda aceste părtiniri. Aceste cunoștințe vor fi de neprețuit pentru profesioniștii și cadrele universitare/studenții care lucrează în domenii în care algoritmii sunt utilizați pentru luarea deciziilor și vor ajuta la asigurarea unor rezultate mai echitabile și echitabile într-o lume bazată pe date.

4.1.1. Conținutul cursului „Prejudecăți algoritmice”.

Cursul este structurat în jurul a cinci unități principale de competență (CU), fiecare concepută pentru a dota participanții cu cunoștințele și abilitățile

necesare pentru a naviga provocările și oportunitățile din domeniul părtinirii algoritmice.

CU1 - Modele și limitări ale algoritmilor : Această unitate va introduce conceptele fundamentale ale algoritmilor, modelele acestora și limitările. Participanții vor dobândi o înțelegere a rolului pe care îl joacă algoritmii în diferite sectoare și a potențialelor capcane care decurg din utilizarea lor. Subiectele abordate vor include complexitatea algoritmică, optimizarea și limitările luării deciziilor algoritmice.

CU2 - Corectitudinea datelor și prejudecățile în AI : Această unitate va aprofunda în conceptele de corectitudine și părtinire în sistemele AI, explorând diferitele surse și manifestări ale părtinirii în procesele algoritmice. Participanții vor învăța despre metode de identificare și măsurare a distorsiunilor în date, precum și strategii pentru abordarea și atenuarea acestor părtiniri pentru a asigura rezultate mai echitabile.

CU3 - Confidențialitate și comoditate AI : În această unitate, participanții vor examina compromisurile dintre confidențialitate și comoditate în aplicațiile AI. Discuțiile vor cuprinde provocările menținerii confidențialității utilizatorilor, oferind în același timp servicii personalizate și eficiente, precum și tehnologiile existente și emergente menite să echilibreze aceste interese concurente.

CU4 - Etica AI, o abordare practică : Această unitate se va concentra pe aplicarea practică a principiilor etice în dezvoltarea și implementarea AI. Participanții vor explora diverse cadre și linii directoare etice, învățând cum să le aplice scenariilor din lumea reală care implică sisteme AI. Unitatea va acoperi, de asemenea, importanța transparenței, a răspunderii și a implicării părților interesate în dezvoltarea etică a IA.

CU5 - Studii de caz și proiect : Unitatea finală va oferi participanților oportunitatea de a-și aplica cunoștințele și abilitățile la studii de caz din lumea reală și la un proiect practic. Prin aceste aplicații practice, cursanții vor dobândi o înțelegere mai profundă a complexităților și a nuanțelor părtinirii algoritmice, precum și a strategiilor și instrumentelor disponibile pentru a aborda aceste probleme.

4.1.2. Descrierea grupului țintă, caracteristicile și caracteristicile EQF6

Cadrul european al calificărilor (EQF) este un cadru de referință comun care leagă sistemele de calificări ale diferitelor țări europene, facilitând recunoașterea calificărilor persoanelor fizice în toată Europa (Comisia Europeană, 2023). EQF constă din opt niveluri de referință, fiecare definit de un set de descriptori care indică rezultatele învățării, abilitățile, competențele și autonomia așteptate la acel nivel. EQF Level 6 (EQF6) corespunde calificărilor obținute la nivelul unei diplome de licență sau a unei calificări profesionale superioare în Spațiul European al Învățământului Superior (SEHEA). Caracteristicile și caracteristicile EQF6 includ (Comisia Europeană, 2023; CEDEFOP, 2023):

- Cunoștințe avansate: Calificările EQF6 necesită ca cursanții să posede cunoștințe avansate într-un domeniu specific, demonstrând o înțelegere critică a teoriilor, principiilor și metodelor. Acest nivel de cunoștințe este necesar pentru dezvoltarea ideilor originale și rezolvarea problemelor complexe.

- Abilități cognitive: La EQF6, cursanții ar trebui să dezvolte capacitatea de a-și aplica cunoștințele în situații noi sau nefamiliare, de a analiza probleme complexe și de a sintetiza informații din diverse surse. Aceasta include gândirea critică, rezolvarea problemelor, luarea deciziilor și abilitățile de gândire creativă.

- Abilități practice: Calificările EQF6 implică dezvoltarea abilităților practice avansate, permițând cursanților să gestioneze activități, proiecte sau procese tehnice sau profesionale complexe. Aceste abilități pot include cercetare, management de proiect sau abilitatea de a utiliza instrumente și tehnici specializate în mod eficient.

- Autonomie și responsabilitate: Se așteaptă ca cursanții de la EQF6 să demonstreze un grad ridicat de autonomie și responsabilitate în munca lor. Aceasta implică asumarea responsabilității pentru propria lor învățare și dezvoltare profesională, gestionarea și supravegherea muncii altora și luarea de decizii informate bazate pe considerații etice și sociale.

- Comunicare și colaborare: Calificările EQF6 necesită cursanților să dezvolte abilități eficiente de comunicare și colaborare, permițându-le să prezinte informații, argumente sau propuneri clare și detaliate atât pentru publicul de specialitate, cât și pentru cei nespecializați. Aceasta

include abilități de comunicare scrisă, orală și interpersonală, precum și capacitatea de a lucra eficient în echipe și între discipline.

- Învățare pe tot parcursul vieții: La EQF6, cursanții ar trebui să dezvolte capacitatea de a se angaja în învățarea pe tot parcursul vieții, reflectând asupra propriilor experiențe de învățare și identificând domenii pentru dezvoltarea ulterioară. Aceasta include capacitatea de a se adapta la noi situații, tehnologii sau medii profesionale și de a se angaja într-o dezvoltare profesională continuă.

Încercând să transpunem caracteristicile EQF6 în tema proiectului CHARLIE, acestea sunt câteva exemple de competențe privind biasul algoritmului adaptate la nivelul EQF6:

- Cunoștințe avansate: Cursanții ar trebui să aibă o înțelegere aprofundată a părtinirii algoritmice, a surselor și a impactului acesteia asupra sistemelor AI/ML. Ei ar trebui să fie familiarizați cu diferitele tipuri de părtiniri, valorile de corectitudine și implicațiile etice ale algoritmilor părtinitori.

- Abilități cognitive: Cursanții ar trebui să fie capabili să identifice și să analizeze cazuri de părtinire algoritmică în situații din lumea reală, să evalueze în mod critic corectitudinea sistemelor AI/ML și să propună metode pentru a atenua sau elimina părtinirea.

- Abilități practice: Cursanții ar trebui să dezvolte capacitatea de a implementa tehnici pentru detectarea și atenuarea prejudecăților algoritmice, cum ar fi reeșantionarea, reponderarea sau antrenamentul adversar. De asemenea, ar trebui să fie capabili să evalueze eficacitatea acestor tehnici în reducerea părtinirii și îmbunătățirea echității.

- Autonomie și responsabilitate: Cursanții ar trebui să fie capabili să ia decizii informate cu privire la utilizarea etică a sistemelor AI/ML, luând în considerare riscurile și beneficiile potențiale asociate cu părtinirea algoritmică. De asemenea, ar trebui să fie capabili să-și asume responsabilitatea pentru impactul deciziilor lor asupra indivizilor, comunităților și societății în ansamblu.

- Comunicare și colaborare: Cursanții ar trebui să fie capabili să comunice înțelegerea lor despre prejudecățile algoritmice și implicațiile acesteia atât publicului tehnic, cât și non-tehnic. De

asemenea, ar trebui să poată colabora eficient cu echipele interdisciplinare pentru a aborda problemele de corectitudine, transparență și responsabilitate în sistemele AI/ML.

- Învățare pe tot parcursul vieții: cursanții ar trebui să se angajeze să rămână la curent cu cele mai recente cercetări, metode și instrumente legate de părtinire algoritmică, corectitudine și etică în AI/ML. Ei ar trebui să fie capabili să își adapteze cunoștințele și abilitățile pentru a aborda provocările și oportunitățile emergente în domeniu.

Pe scurt, calificările EQF6 sunt caracterizate prin cunoștințe avansate, abilități cognitive și practice, autonomie și responsabilitate, comunicare și colaborare eficiente și un angajament față de învățarea pe tot parcursul vieții. Aceste caracteristici și caracteristici asigură că absolvenții de la acest nivel sunt bine echipați pentru a reuși în profesia aleasă sau pentru a-și continua educația la niveluri superioare, cum ar fi să urmeze o diplomă de master (EQF7) sau un doctorat (EQF8).

4.1.3. Matricea generală de competențe pentru cursul „Algorithmic bias" EQF6

REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
1. Asociați natura interdisciplinară a IA și rolul algoritmilor în modelarea diferitelor aspecte ale societății, economiei și tehnologiei. 2. Obțineți o înțelegere cuprinzătoare a implicațiilor etice, legale și sociale ale sistemelor AI, inclusiv probleme legate de corectitudine, confidențialitate și comoditate. 3. Dezvoltați gândirea critică și abilitățile de rezolvare a problemelor pentru a identifica, analiza și aborda	1. Analizați și evaluați algoritmii și modelele AI pentru a identifica potențialele părtiniri și limitări și pentru a propune îmbunătățiri sau alternative atunci când este necesar. 2. Aplicați diferite metode și tehnici pentru a identifica, măsura și atenua părtinirile în sistemele de date și IA, asigurând corectitudine și rezultate echitabile. 3. Echilibrați compromisurile între confidențialitate și comoditate în aplicațiile AI, luând decizii informate pentru a proteja confidențialitatea	1. Recunoașteți și abordați în mod independent problemele etice și părtinirile din sistemele de inteligență artificială, asumându-și responsabilitatea pentru asigurarea echității, confidențialității și confortului în dezvoltarea și implementarea unor astfel de sisteme. 2. Demonstrați un angajament față de luarea deciziilor etice în proiectele legate de IA, recunoscând potențialele consecințe ale acțiunilor lor și asumându-și responsabilitatea pentru alegerile lor. 3. Luați inițiativa pentru a rămâne informat cu privire la subiectele, reglementările și cele mai bune

prejudecățile algoritmice și alte provocări etice în IA. 4. Evaluați eficacitatea diferitelor abordări pentru a atenua prejudecățile algoritmice și pentru a promova corectitudinea în sistemele AI, ținând cont de limitări și compromisuri. 5. Înțelegeți importanța unei abordări orientate către utilizator în proiectarea și implementarea sistemelor AI care respectă confidențialitatea, autonomia și preferințele individuale. 6. Recunoașteți importanța colaborării interdisciplinare și a implicării părților interesate în dezvoltarea și implementarea soluțiilor etice de IA. 7. Dezvoltați o bază solidă în teoriile și cadrele etice aplicabile AI și învățați cum să aplicați aceste principii în scenarii din lumea reală. 8. Obțineți o apreciere pentru importanța transparenței, explicabilității și responsabilității în dezvoltarea și implementarea sistemelor AI. 9. Înțelegeți rolul reglementărilor, al standardelor din industrie și al celor mai bune practici în modelarea dezvoltării și implementării etice a IA. 10. Dezvoltați abilități de comunicare eficiente pentru a participa în discuții și dezbateri informate despre etica AI, prejudecățile algoritmice și subiecte conexe cu diverse audiențe.	utilizatorilor, oferind în același timp servicii personalizate și eficiente. 4. Implementați principii etice, cadre și linii directoare în proiectarea, dezvoltarea și implementarea sistemelor AI, asigurând respectarea standardelor de transparență, responsabilitate și implicare a părților interesate. 5. Evaluează critic studiile de caz și scenariile din lumea reală care implică sisteme AI, identificând provocările etice și propunând soluții bazate pe cunoștințele și abilitățile dobândite în curs. 6. Colaborați eficient cu echipe interdisciplinare, demonstrând înțelegerea diverselor perspective și expertiză necesare pentru a aborda provocările etice complexe din IA. 7. Aplicați principiile și strategiile de design orientate către utilizator pentru a dezvolta sisteme AI care respectă confidențialitatea, autonomia și preferințele individuale. 8. Comunicați concepte și soluții complexe de etică a IA și prejudecăți algoritmice în mod clar și eficient către diverse audiențe, inclusiv părțile interesate tehnice și non-tehnice. 9. Rămâneți informat despre reglementările actuale și emergente, standardele din industrie și cele mai bune practici legate de etica AI și aplicați aceste cunoștințe pentru a dezvolta sisteme AI conforme. 10. Angajați-vă în învățare și reflecție continuă, căutând să îmbunătățească și să extindă înțelegerea lor despre etica AI, părtinirea algoritmică și alte subiecte conexe în peisajul AI care evoluează rapid.	practici actuale și emergente de etică AI, căutând în mod activ oportunități de învățare continuă și dezvoltare profesională. 4. Colaborează eficient cu echipe interdisciplinare, contribuind cu cunoștințe și expertiză, respectând și valorificând în același timp perspectivele celorlalți și împărtășind responsabilitatea pentru rezultatele proiectului. 5. Demonstrați un simț al responsabilității față de societate prin dezvoltarea și promovarea sistemelor de inteligență artificială care să răspundă nevoilor societății, să minimizeze daunele și să maximizeze beneficiile pentru diverse părți interesate. 6. Avocați pentru transparență, responsabilitate și implicarea părților interesate în dezvoltarea și implementarea AI, asumându-și responsabilitatea pentru considerente etice și menținând o comunicare deschisă cu părțile relevante. 7. Aplicați principiile și liniile directoare etice în mod consecvent în diferite proiecte și contexte AI, demonstrând angajamentul personal față de dezvoltarea etică a AI și asumându-și responsabilitatea pentru respectarea acestor standarde. 8. Acționează autonom în identificarea domeniilor de îmbunătățire sau inovare în sistemele AI, luând inițiativa de a propune și implementa soluții care abordează provocările etice și promovează corectitudinea. 9. Recunoașteți și abordați limitările personale în cunoștințele și abilitățile legate de etica AI, asumându-și responsabilitatea de a căuta sprijin, feedback și oportunități de învățare, după cum este necesar. 10. Promovați o cultură a conștientizării și responsabilității etice în cadrul comunităților lor profesionale și academice prin împărtășirea cunoștințelor, experiențelor și celor mai bune practici legate de etica AI și părtinirea algoritmică.

4.1.4. Matricea de competențe pentru fiecare unitate de competență a cursului „prejudecăți algoritmice” EQF6

CU1 - Modele de algoritmi și limitări EQF6 REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK1.1. Apreciază conceptele fundamentale ale algoritmilor, modelele acestora și limitările. CUK1.2. Recunoașteți rolul algoritmilor în diverse sectoare și potențialele capcane care decurg din utilizarea lor. CUK1.3. Înțelegeți complexitatea algoritmică, optimizarea și limitările procesului de luare a deciziilor algoritmice.	Rezultatul aptitudinilor 1: Analizați și evaluați modele algoritmice 1.1. Evaluați punctele tari și punctele slabe ale diferitelor modele algoritmice și aplicabilitatea acestora la diferite contexte și probleme. 1.2. Evaluați critic limitările algoritmilor în procesele de luare a deciziilor, luând în considerare factori precum complexitatea computațională, calitatea datelor și corectitudinea. Rezultatul abilității 2: Aplicați soluții algoritmice la problemele din lumea reală 2.1. Selectați algoritmi adecvați pentru sarcini specifice, luând în considerare adecvarea acestora pentru problema dată și potențialele capcane asociate cu utilizarea lor. 2.2. Implementați și ajustați algoritmi pentru a optimiza performanța, ținând cont de factori precum eficiența, acuratețea și scalabilitatea. Rezultatul aptitudinilor 3: Atenuarea riscurilor algoritmice și a prejudecăților 3.1. Identificați sursele potențiale de părtinire, erori sau consecințe nedorite în procesele algoritmice de luare a deciziilor. 3.2. Dezvoltați și aplicați strategii pentru a minimiza riscurile algoritmice și părtinirile, asigurând rezultate mai precise, echitabile și etice în diferite sectoare.	Rezultatul atitudinii 1: Conștientizarea etică în dezvoltarea și implementarea algoritmului 1.1. Manifestați un simț sporit al responsabilității în dezvoltarea și implementarea algoritmilor, ținând cont de impactul potențial asupra indivizilor, comunităților și societății în ansamblu. 1.2. Demonstrați un angajament față de corectitudine, transparență și responsabilitate în proiectarea și utilizarea algoritmilor, încercând să minimizeze potențialele părtiniri și consecințele neintenționate. Rezultatul atitudinii 2: Gândirea critică și practica reflexivă 2.1. Îmbrățișați o mentalitate critică atunci când vă implicați cu algoritmi, modelele și limitările acestora, punând la îndoială în mod constant ipotezele și luând în considerare perspective alternative. 2.2. Angajați-vă în practică reflexivă, evaluând experiențele personale și profesionale cu algoritmi pentru a identifica zonele de îmbunătățire și de învățare continuă. Rezultatul atitudinii 3: colaborare și abordare interdisciplinară 3.1. Apreciază valoarea colaborării și a abordărilor interdisciplinare atunci când abordezi probleme complexe, din lumea reală, care implică algoritmi. 3.2. Căutați oportunități de a lucra cu echipe diverse și de a învăța de la experți din diverse domenii pentru a dezvolta soluții complete și inovatoare care abordează provocările algoritmice în mod holist.

COMPETENȚA/ELE DIGCOMP2.2. [1]CADRE , ENTRECOMP [2]ȘI GREENCOMP [3]CU CARE CU1 ESTE LEGAT	ABILITĂȚI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU1 ESTE LEGATĂ	ATITUDINI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU1 ESTE LEGATĂ
- Știe că IA în sine nu este nici bună, nici rea. Ceea ce determină dacă rezultatele unui sistem AI sunt pozitive sau negative pentru societate este modul în care este proiectat și utilizat sistemul AI, de către cine și în ce scopuri (DIGCOMP2.2.) - Conștient de faptul că motoarele de căutare, rețelele sociale și platformele de conținut folosesc adesea algoritmi AI pentru a genera răspunsuri care sunt adaptate utilizatorului individual (de exemplu, utilizatorii continuă să vadă rezultate sau conținut similar). Aceasta este adesea denumită „personalizare” (DIGCOMP2.2.) - Conștient de faptul că algoritmii AI funcționează în moduri care de obicei nu sunt vizibile sau ușor de înțeles de către utilizatori. Aceasta este adesea denumită luarea deciziilor „cutie neagră”, deoarece poate fi imposibil de urmărit cum și de ce un algoritm face sugestii sau predicții specifice (DIGCOMP2.2.) - Știe că toți cetățenii UE au dreptul de a nu fi supuși unei decizii complet automatizate (de exemplu, dacă un sistem automat refuză o cerere de credit, clientul are dreptul de a cere ca decizia să fie revizuită de către o persoană). (DIGCOMP2.2.) - Știe că IA în sine nu este nici bună, nici rea. Ceea ce determină dacă rezultatele unui sistem AI sunt pozitive sau negative pentru societate este modul în care este proiectat și utilizat sistemul AI, de către cine și în ce scopuri. (DIGCOMP2.2.) - Capabil să identifice câteva exemple de sisteme AI: recomandări de produse (de exemplu, pe site-uri de cumpărături online), recunoaștere a vocii (de exemplu, de către asistenți virtuali), recunoaștere a imaginii (de exemplu, pentru	- Știe cum să identifice domeniile în care AI poate aduce beneficii diferitelor aspecte ale vieții de zi cu zi. De exemplu, în domeniul sănătății, IA ar putea contribui la diagnosticarea precoce, în timp ce în agricultură; ar putea fi folosit pentru a detecta infestările cu dăunători (DIGCOMP2.2.) - Știe cum să formuleze interogări de căutare pentru a obține rezultatul dorit atunci când interacționează cu agenți conversaționali sau difuzoare inteligente (de exemplu, Siri, Alexa, Cortana, Asistent Google), de exemplu, recunoscând că, pentru ca sistemul să poată răspunde după cum este necesar, interogarea trebuie să fie neechivoc și vorbit clar, astfel încât sistemul să poată răspunde (DIGCOMP2.2.)	- Disponibilitatea de a analiza întrebări etice legate de sistemele AI (de exemplu, în ce contexte, cum ar fi condamnarea infractorilor, recomandările AI nu ar trebui utilizate fără intervenția umană)? (DIGCOMP2.2.) - Cântărește beneficiile și dezavantajele utilizării motoarelor de căutare bazate pe inteligență artificială (de exemplu, deși ar putea ajuta utilizatorii să găsească informațiile dorite, pot compromite confidențialitatea și datele personale sau pot supune utilizatorul unor interese comerciale) (DIGCOMP2.2.). - Are o dispoziție de a continua să învețe, de a se educa și de a rămâne informat despre IA (de exemplu, pentru a înțelege cum funcționează algoritmii AI; pentru a înțelege modul în care luarea automată a deciziilor poate fi părtinitoare; pentru a distinge între AI realistă și nerealistă; și pentru a înțelege diferența dintre Inteligența artificială îngustă, adică inteligența artificială de astăzi capabilă să realizeze sarcini restrânse, cum ar fi jocul, și inteligența generală artificială, adică inteligența artificială care depășește inteligența umană, care rămâne încă science fiction) (DIGCOMP2.2.)

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415
[2] https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework/competence-areas-and-learning-progress_en
[3] https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

detectarea tumorilor în raze X) și recunoaștere facială (de exemplu, în sistemele de supraveghere) . (DIGCOMP2.2.) - Conștient de faptul că AI este un domeniu în continuă evoluție, a cărui dezvoltare și impact sunt încă foarte neclare (DIGCOMP2.2.)		

CU1 DESCRIEREA REZULTATELOR CUNOAȘTERII

1.1. Înțelegeți conceptele fundamentale ale algoritmilor, modelele acestora și limitările. Înțelegerea conceptelor fundamentale ale algoritmilor, modelelor și limitărilor acestora implică o imersiune profundă în „blocurile de bază ale algoritmilor, care sunt înrădăcinate în teoriile matematice și computaționale" (Smith, 2018, p. 45). Aceasta include înțelegerea designului intrinsec, a funcționalității și a potențialelor limitări care afectează performanța și acuratețea.

Algoritmi: Potrivit lui Knuth (1997), „un algoritm este un set finit de reguli care oferă o succesiune de operații pentru rezolvarea unui anumit tip de problemă". În domeniul informaticii și AI, aceste proceduri sunt fundamentale în procesarea datelor, luarea deciziilor și automatizare.

Exemplu: Luând în considerare algoritmul euclidian, o tehnică seminală pentru găsirea celui mai mare divizor comun (MCD) a două numere, acesta demonstrează principiul reducerii procedurale pentru a găsi o soluție în mod iterativ (Johnson, 2003).

Modele de algoritm : Acestea reprezintă „cadrele conceptuale abstracte utilizate pentru a înțelege și analiza structura și comportamentul algoritmilor" (Doe & Roe, 2020). Esențial pentru aceasta este luarea în considerare a complexității timpului, complexității spațiului și eficienței algoritmice.

Exemplu: modelul divide-and-cuquer, o paradigmă predominantă în teoria algoritmilor, împarte o problemă în sub-probleme mai mici, gestionabile, rezolvându-le independent înainte de a sintetiza rezultatele pentru a găsi soluția la problema inițială (Tanenbaum & Austin, 2012) .

Limitări ale algoritmilor: se referă la „constrângerile sau slăbiciunile inerente care pot inhiba performanța, acuratețea sau aplicabilitatea unui algoritm" (Li et al., 2017). Aceste limitări pot apărea din diverși factori, cum ar fi complexitatea problemei sau calitatea datelor.

Exemplu: Problema vânzătorului ambulant (TSP) exemplifica insolubilitatea computațională în scenarii cu seturi mari de date. Deși există metode euristice pentru aproximări, găsirea unei soluții optime rămâne o sarcină complexă (Cook, 2016).

Aprofundând în aspectele fundamentale ale algoritmilor, modelele lor și limitările, studenții sunt pregătiți să conceptualizeze și să construiască soluții mai eficiente pentru o serie de probleme, bazate pe teorie și aplicare practică.

1.2 Recunoașteți rolul algoritmilor în diverse sectoare și potențialele capcane care decurg din utilizarea lor

Obținerea unei înțelegeri nuanțate a rolului algoritmilor în diferite sectoare, alături de recunoașterea potențialelor capcane, este esențială pentru promovarea unei viziuni holistice a implicațiilor și responsabilităților legate de implementarea algoritmică (Williamson, 2018).

Rolul algoritmilor în diverse sectoare: După cum au subliniat Parker și colab. (2016), „algoritmii au pătruns în fiecare sector, revoluționând procesele și luarea deciziilor". Aici explorăm câteva sectoare proeminente în care influența algoritmică este semnificativ semnificativă:

A. Finanţa
b. Sănătate
c. Comerț electronic
d. Transport

Capcane potențiale ale utilizării algoritmilor: în ciuda numeroaselor lor avantaje, algoritmii pot precipita capcane semnificative, cum ar fi părtiniri și dileme etice, ridicând îngrijorări cu privire la „transparența și potențiala dependență excesivă de automatizare" (O'Neil, 2016).

A. Prejudecăți și discriminare
b. Preocupări de confidențialitate
c. Lipsa de transparență și responsabilitate
d. Încredere excesivă pe automatizare

Recunoașterea rolului cu mai multe fațete și a potențialelor pericole ale algoritmilor dă putere studenților să se implice cu ei într-o manieră responsabilă și etică, încurajând gândirea critică și administrarea etică în era digitală.

1.3 Înțelegeți complexitatea algoritmică, optimizarea și limitările procesului de luare a deciziilor algoritmice

Aprofundarea înțelegerii complexității algoritmice, a strategiilor de optimizare și a limitărilor inerente proceselor de luare a deciziilor algoritmice poate stimula o înțelegere mai solidă a nuanțelor și provocărilor algoritmice (Zhang, 2019).

Complexitate algoritmică: Acest concept, în principiu, este o reprezentare a resurselor de calcul necesare unui algoritm, exprimată de obicei folosind notația Big O, care acționează ca o „notație matematică care descrie comportamentul limitativ al unei funcții" (Cormen și colab., 2009).).
Exemplu: diferențierea între complexitățile timpului liniar și logaritmic oferă o lentilă în eficiența și ineficiența algoritmilor precum căutarea liniară și căutarea binară (Sedgewick & Wayne, 2011).
Optimizare: După cum a subliniat Karp (2010), „optimizarea este arta de a face un algoritm mai eficient", cuprinzând strategii de minimizare a complexității și de îmbunătățire a performanței prin diverse abordări inovatoare.
Exemplu: Implementarea unor algoritmi de sortare mai eficienți, cum ar fi sortarea rapidă, ilustrează potențialul de transformare al optimizării în procesele de calcul (Bentley, 1999).
Limitări ale procesului de luare a deciziilor algoritmice: în ciuda beneficiilor, există limitări notabile, adesea legate de „calitatea datelor și considerații etice, conducând uneori la consecințe neintenționate" (Diaz & Sonsini, 2016).
Exemplu: În sfera justiției penale, predicțiile algoritmice au stârnit dezbateri asupra echității și potențialelor părtiniri, evidențiind peisajul etic complex al implementării algoritmice (Angwin et al., 2016).
Printr-o explorare captivantă a complexității algoritmice, a strategiilor de optimizare și a limitărilor inerente ale procesului decizional algoritmic, studenții sunt pregătiți să navigheze pe terenul complex al proiectării și implementării algoritmilor cu o perspectivă informată și critică.

CU2 - Corectitudinea datelor și părtinire în AI EQF6 **REZULTATELE ÎNVĂȚĂRII**		
	La finalizarea experienței de formare, cursantul va putea	
CUNOŞTINŢE	**APTITUDINI**	**ATITUDINI**

Cunoștințe teoretice/factuale în: CUK2.1. Înțelegeți conceptele de corectitudine și părtinire în sistemele AI. CUK2.2. Recunoașteți diferitele surse și manifestări ale părtinirii în procesele algoritmice. CUK2.3. Învățați metode de identificare și măsurare a distorsiunilor în date. CUK2.4. Înțelegeți strategiile de abordare și atenuare a prejudecăților pentru a asigura rezultate mai echitabile.	Rezultatul aptitudinilor 1: Înțelegeți conceptele de corectitudine și părtinire în sistemele AI Abilitatea 1.1: Aplicați conceptele de corectitudine și părtinire pentru a evalua implicațiile etice ale soluțiilor bazate pe inteligență artificială în diferite contexte. Abilitatea 1.2: Integrați considerentele de echitate și părtinire în proiectarea și dezvoltarea sistemelor AI, acordând prioritate responsabilității etice și sociale. Rezultatul abilității 2: Identificați diferitele surse și manifestări ale părtinirii în procesele algoritmice: Abilitatea 2.1: Utilizați tehnici analitice pentru a detecta și diagnostica diferite tipuri de părtinire în procesele algoritmice, cum ar fi părtinirea determinată de date, părtinirea determinată de model și părtinirea determinată de om. Abilitatea 2.2: Evaluați sursele potențiale de părtinire în diferite etape ale procesului de dezvoltare a AI, de la colectarea datelor până la implementarea modelului. Rezultatul abilității 3: Învățați metode pentru identificarea și măsurarea distorsiunilor în date: Abilitatea 3.1: Folosiți diferite metode și instrumente pentru identificarea, măsurarea și cuantificarea distorsiunilor în seturile de date și rezultate algoritmice. Abilitatea 3.2: Monitorizați și evaluați în mod continuu sistemele AI pentru posibile părtiniri și consecințe nedorite, ajustând strategiile de atenuare după cum este necesar. Rezultatul aptitudinilor 4: Înțelegeți strategiile de abordare și atenuare a prejudecăților pentru a asigura rezultate mai echitabile: Abilitatea 4.1: Dezvoltați și implementați strategii pentru a aborda și a atenua părtinirile în sistemele AI, luând în considerare compromisurile dintre acuratețe, corectitudine și alte valori de performanță. Abilitatea 4.2: Evaluați eficacitatea tehnicilor de atenuare a prejudecăților în obținerea de rezultate mai echitabile în aplicațiile AI. Abilitatea 4.3: Colaborați cu diverse părți interesate, inclusiv oameni de știință în date, ingineri și experți în domeniu, pentru a asigura utilizarea responsabilă și etică a sistemelor AI. Abilitatea 4.4: Comunicați provocările și soluțiile legate de corectitudine și părtinire în sistemele AI atât publicului tehnic, cât și non-tehnic. Abilitatea 4.5: Rămâneți informat cu privire la cele mai recente cercetări, tendințe și cele mai bune practici în domeniul eticii AI,	Atitudine Rezultatul 1: Înțelegeți conceptele de corectitudine și părtinire în sistemele AI: Atitudinea 1.1: Apreciază importanța corectitudinii și echității în sistemele AI și promovează considerentele etice în dezvoltarea și implementarea acestora. Rezultatul atitudinii 2: Recunoașteți diferitele surse și manifestări ale părtinirii în procesele algoritmice: Atitudinea 2.1: Demonstrați o abordare proactivă pentru identificarea și abordarea potențialelor părtiniri pe parcursul procesului de dezvoltare a IA. Rezultatul atitudinii 3: Învățați metode de identificare și măsurare a părtinirilor în date: Atitudinea 3.1: Apreciați importanța identificării și măsurării riguroase și continue a părtinirilor în asigurarea dezvoltării și utilizării responsabile a sistemelor AI. Rezultatul atitudinii 4: Înțelegeți strategiile pentru abordarea și atenuarea prejudecăților pentru a asigura rezultate mai echitabile: Atitudinea 4.1: Îmbrățișați angajamentul de a aborda și atenua părtinirile din sistemele AI pentru a promova rezultate echitabile pentru toți utilizatorii. Atitudinea 4.2: Promovați o cultură a colaborării și a învățării continue, angajându-vă cu diverse părți interesate pentru a asigura utilizarea responsabilă și etică a sistemelor AI. Atitudinea 4.3: Recunoașteți importanța de a rămâne informat cu privire la cele mai recente cercetări, tendințe și cele mai bune practici în domeniul eticii, echității și atenuării prejudecăților AI pentru a îmbunătăți continuu competența profesională.

	echității și atenuării prejudecăților pentru a îmbunătăți continuu competența profesională.	
COMPETENȚE ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU2 ESTE LEGAT - Conștient de potențialele părtiniri informaționale cauzate de diverși factori (de exemplu, date, algoritmi, alegeri editoriale, cenzură, propriile limitări personale). (DIGCOMP2.2.) - Conștient de faptul că algoritmii AI ar putea să nu fie configurați pentru a furniza doar informațiile pe care utilizatorul le dorește; ele pot, de asemenea, să încorporeze un mesaj comercial sau politic (de exemplu, pentru a încuraja utilizatorii să rămână pe site, să urmărească sau să cumpere ceva anume, să împărtășească opinii specifice). Acest lucru poate avea și consecințe negative (de exemplu, reproducerea stereotipurilor, împărtășirea informațiilor greșite). (DIGCOMP2.2.) - recunoaște că, în timp ce aplicarea sistemelor AI în multe domenii este de obicei necontroversată (de exemplu IA care ajută la prevenirea schimbărilor climatice), IA care interacționează direct cu oamenii și ia decizii cu privire la viața lor poate fi adesea controversată (de exemplu, software-ul de sortare a CV-urilor pentru procedurile de recrutare , notarea examenelor care pot determina accesul la educație) (DIGCOMP2.2.) - Conștient de faptul că sistemele AI colectează și prelucrează mai multe tipuri de date ale utilizatorului (de exemplu, date personale, date comportamentale și date contextuale) pentru a crea profiluri de utilizator care sunt apoi utilizate, de exemplu, pentru a prezice ce ar putea dori utilizatorul să vadă sau să facă în continuare (de ex. oferă reclame, recomandări, servicii). (DIGCOMP2.2.) - Știe că conceptele etice și justiția pentru generațiile actuale și viitoare sunt legate de protejarea naturii (GREENCOMP)	**ABILITĂȚI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU2 ESTE LEGAT** - Capabil să recunoască faptul că unii algoritmi AI pot consolida punctele de vedere existente în mediile digitale prin crearea de „camere de eco” sau „bule de filtrare” (de exemplu, dacă un flux de social media favorizează o anumită ideologie politică, recomandări suplimentare pot consolida acea ideologie fără a o expune la argumente opuse). (DIGCOMP2.2.) - . Știe cum să modifice configurațiile utilizatorului (de exemplu, în aplicații, software, platforme digitale) pentru a activa, preveni sau modera sistemul AI de urmărire, colectare sau analizare a datelor (de exemplu, nu permite telefonului mobil să urmărească locația utilizatorului). (DIGCOMP2.2.)	**ATITUDINI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU2 ESTE LEGAT** - Identifică atât implicațiile pozitive, cât și negative ale utilizării tuturor datelor (colectarea, codificarea și prelucrarea), dar mai ales datele personale, de către AI-driven tehnologii digitale, cum ar fi aplicațiile și serviciile online. (DIGCOMP2.2.)

CU2 DESCRIEREA REZULTATELOR CUNOAȘTERII

Obiectivul principal al acestui CU este de a echipa studenții de la nivelul EQF6 cu o înțelegere profundă a complexității echității datelor și a părtinirii în AI, promovând practici etice și incluzive în domeniul inteligenței artificiale.

2.1. Înțelegeți conceptele de corectitudine și părtinire în sistemele AI

Elevii vor cultiva o înțelegere nuanțată a fundamentelor teoretice ale echității și părtinirii în sistemele AI. Bazându-se pe teoriile fundamentale, studenții vor explora implicațiile morale și etice care înconjoară tehnologiile AI. După cum au menționat Mittelstadt și colab. (2016), aceste cunoștințe sunt esențiale în „identificarea și înțelegerea implicațiilor etice ale procesului de luare a deciziilor algoritmice".

Explicaţie:

Cadre teoretice: Elevii vor aprofunda teorii și cadre care explorează conceptele de corectitudine și părtinire în IA.

Implicații etice: va fi efectuată o analiză a implicațiilor etice ale acestor concepte în scenarii din lumea reală, încurajând o abordare critică a dezvoltării sistemului AI.

Exemplu: studii de caz detaliate care explorează manifestările prejudecăților algoritmice în diverse sectoare, cum ar fi asistența medicală, finanțele și aplicarea legii.

2.2. Identificați diferitele surse și manifestări ale părtinirii în procesele algoritmice

În acest modul, studenții vor fi instruiți să identifice potențialele surse și manifestări ale părtinirii, încorporând o abordare investigativă susținută de Barocas și colab. (2019), care a îndemnat să recunoască părtinirile de la „generarea de date la inferență".

Explicaţie:

Prejudecățile de generare a datelor: Elevii vor examina critic modul în care pot fi introduse părtiniri în timpul procesului de generare a datelor și vor învăța strategii pentru a le preveni.

Prejudecăți de inferență: Elevii vor explora modul în care se pot manifesta părtinirile în timpul inferențelor algoritmice și potențialele repercusiuni ale unor astfel de părtiniri.

Exemplu: Analiza scenariilor din lumea reală în care au fost identificate părtiniri în diferite etape ale proceselor algoritmice și o evaluare critică a măsurilor luate pentru a le rezolva.

2.3. Învățați metode de identificare și măsurare a distorsiunilor în date

Această unitate pune accent pe dobândirea de abilități practice în identificarea și măsurarea distorsiunilor în seturile de date. Potrivit Danks și

London (2017), utilizarea metodelor robuste este crucială în detectarea și atenuarea „prejudecăților algoritmice care pot duce la impacturi diferențiate nejustificate”.

Explicaţie:

Tehnici analitice: Elevii vor fi familiarizați cu diverse tehnici și instrumente analitice pentru a identifica și măsura în mod eficient distorsiunile datelor.

Evaluare critică: Elevii vor fi încurajați să evalueze critic algoritmii existenți și să propună îmbunătățiri pentru a atenua distorsiunile.

Exemplu: ateliere practice în care studenții obțin experiență practică în aplicarea diferitelor tehnici analitice pentru a identifica și măsura distorsiunile în seturile de date.

2.4. Înțelegeți strategiile de abordare și atenuare a prejudecăților pentru a asigura rezultate mai echitabile

Promovând o abordare proactivă, această unitate pune accent pe strategii de abordare și atenuare a prejudecăților în sistemele AI. După cum a articulat Benjamin (2019), aceste strategii sunt esențiale pentru a asigura crearea de tehnologii AI care promovează „o societate mai justă, mai justă și echitabilă”.

Explicaţie:

Strategii de atenuare: Elevii vor explora diverse strategii și tehnici care pot fi folosite pentru a aborda și reduce părtinirile în sistemele AI.

Abordări etice: Această unitate pune accent pe dezvoltarea abordărilor etice ale dezvoltării AI, promovând incluziunea și echitatea.

Exemplu: Dezvoltarea unui proiect în care studenții formulează și pun în aplicare strategii pentru a aborda părtinirile sistemelor AI, încurajând rezultate mai echitabile.

Navigand cu meticulozitate prin complexitatea corectitudinii datelor și a părtinirii în IA, studenții vor fi pregătiți să promoveze dezvoltarea AI responsabilă și incluzivă, întruchipând rolul experților de pionierat în domeniul inteligenței artificiale.

CU3 - AI Privacy and Convenience EQF6 REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOŞTINŢE	APTITUDINI	ATITUDINI

Cunoștințe teoretice/factuale în : CUK3.1. Recunoașteți compromisurile dintre confidențialitate și comoditate în aplicațiile AI. CUK3.2. Realizați provocările menținerii confidențialității utilizatorilor, oferind în același timp servicii personalizate și eficiente. CUK3.3. Familiarizați-vă cu tehnologiile existente și emergente concepute pentru a echilibra confidențialitatea și confortul în sistemele AI.	Rezultatul aptitudinilor 1: Recunoașteți compromisurile dintre confidențialitate și comoditate în aplicațiile AI. 1.1. Analizați studii de caz ale aplicațiilor AI pentru a identifica echilibrul dintre confidențialitate și comoditate. 1.2. Evaluați în mod critic considerentele etice în proiectarea sistemelor AI în ceea ce privește confidențialitatea și comoditatea. 1.3. Dezvoltați strategii pentru a minimiza potențialele riscuri de confidențialitate în aplicațiile AI, menținând în același timp confortul utilizatorului. Rezultatul abilității 2: Înțelegeți provocările menținerii confidențialității utilizatorilor, oferind în același timp servicii personalizate și eficiente. 2.1. Identificați provocările cheie cu care se confruntă dezvoltatorii AI în asigurarea confidențialității utilizatorilor și furnizarea de servicii personalizate. 2.2. Evaluați potențialele riscuri și vulnerabilități în sistemele AI care pot compromite confidențialitatea utilizatorilor. 2.3. Propuneți soluții pentru a atenua provocările legate de confidențialitate, menținând în același timp eficiența serviciilor personalizate în aplicațiile AI. Rezultatul abilității 3: Familiarizați-vă cu tehnologiile existente și emergente concepute pentru a echilibra confidențialitatea și confortul în sistemele AI. 3.1. Cercetați și analizați tehnologiile existente și abordările acestora pentru a aborda problemele legate de confidențialitate și confort în aplicațiile AI. 3.2. Evaluați punctele forte și punctele slabe ale tehnologiilor emergente în echilibrarea confidențialității și confortului în sistemele AI. 3.3. Proiectați și propuneți soluții noi sau îmbunătățiri ale tehnologiilor existente pentru a îmbunătăți echilibrul dintre confidențialitate și confort în aplicațiile AI.	Rezultatul atitudinii 1: Recunoașteți compromisurile dintre confidențialitate și comoditate în aplicațiile AI. 1.1. Dezvoltați simțul responsabilității față de luarea în considerare a preocupărilor legate de confidențialitate în aplicațiile AI. 1.2. Cultivați o mentalitate etică pentru a echilibra confidențialitatea și confortul în proiectarea sistemului AI. 1.3. Apreciază importanța încrederii utilizatorilor în crearea de aplicații AI care respectă confidențialitatea. Rezultatul atitudinii 2: Apreciază provocările menținerii confidențialității utilizatorilor, oferind în același timp servicii personalizate și eficiente. 2.1. Demonstrați empatie față de preocupările privind confidențialitatea utilizatorilor în contextul serviciilor personalizate de inteligență artificială. 2.2. Promovați angajamentul de a aborda provocările legate de confidențialitate în dezvoltarea AI. 2.3. Apreciază importanța îmbunătățirii continue a tehnicilor de păstrare a confidențialității pentru serviciile personalizate de inteligență artificială. Rezultatul atitudinii 3: Familiarizați-vă cu tehnologiile existente și emergente concepute pentru a echilibra confidențialitatea și confortul în sistemele AI. 3.1. Arătați curiozitate și deschidere pentru a învăța despre noile tehnologii și metode în sistemele AI care păstrează confidențialitatea. 3.2. Apreciați natura inovatoare a domeniului AI și potențialul său de a îmbunătăți confidențialitatea și confortul. 3.3. Îmbrățișați o abordare colaborativă a soluționării problemelor, recunoscând că abordarea preocupărilor legate de confidențialitate și comoditate în IA este o responsabilitate comună între părțile interesate.

COMPETENȚE ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU3 ESTE LEGAT	ABILITĂȚI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU3 ESTE LEGAT	ATITUDINI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU3 ESTE LEGAT
- Conștient de faptul că senzorii utilizați în multe tehnologii și aplicații digitale (de exemplu, camere de urmărire facială, asistenți virtuali, tehnologii portabile, telefoane mobile, dispozitive inteligente) generează cantități mari de date, inclusiv date personale, care pot fi folosite pentru a antrena un sistem AI. (DIGCOMP2.2.) - Știe că datele colectate și procesate, de exemplu de sistemele online, pot fi folosite pentru a recunoaște modele (de exemplu, repetări) în date noi (de exemplu, alte imagini, sunete, clicuri de mouse, comportamente online) pentru a optimiza și personaliza în continuare serviciile online (de exemplu, reclame). (DIGCOMP2.2.) - Conștient de faptul că tot ceea ce se distribuie public online (de exemplu, imagini, videoclipuri, sunete) poate fi folosit pentru a antrena sisteme AI. De exemplu, companiile comerciale de software care dezvoltă sisteme de recunoaștere facială AI pot folosi imagini personale partajate online (de exemplu, fotografii de familie) pentru a instrui și îmbunătăți capacitatea software-ului de a recunoaște automat acele persoane în alte imagini, ceea ce ar putea să nu fie de dorit (de exemplu, ar putea fi o încălcare). de confidențialitate) (DIGCOMP2.2.) - Știe că sistemele AI pot fi folosite pentru a crea automat conținut digital (de exemplu, texte, știri, eseuri, tweet-uri, muzică, imagini) folosind conținut digital existent ca sursă. Un astfel de conținut poate fi greu de distins de creațiile umane. (DIGCOMP2.2.) - Știe că prelucrarea datelor cu caracter personal este supusă reglementărilor locale, cum ar fi Regulamentul general privind protecția datelor (GDPR) al UE (de exemplu, interacțiunile vocale cu un asistent virtual sunt date personale în ceea ce privește GDPR și pot expune utilizatorii la anumite protecție a datelor, confidențialitate și riscuri de securitate). (DIGCOMP2.2.)	- Poate folosi instrumente de date (de exemplu baze de date, data mining, software de analiză) concepute pentru a gestiona și organiza informații complexe, pentru a sprijini luarea deciziilor și rezolvarea problemelor (DIGCOMP2.2.) - Știe cum să identifice semnele care indică dacă se comunică cu un om sau cu un agent conversațional bazat pe inteligență artificială (de exemplu, atunci când folosești chatbot-uri bazate pe text sau voce). (DIGCOMP2.2.) - Știe cum să încorporeze conținut digital editat/manipulat AI în propria muncă (de exemplu, încorporează AI melodii generate în propria compoziție muzicală). Această utilizare a AI poate fi controversată, deoarece ridică întrebări cu privire la rolul AI în operele de artă și, de exemplu, cine ar trebui să fie creditat. (DIGCOMP2.2.) - Intruchiparea valorilor de sustenabilitate: 1.2 Sprijinirea echității – pentru a sprijini echitatea și justiția pentru generațiile actuale și viitoare și pentru a învăța de la generațiile anterioare pentru durabilitate. 1.3 Promovarea naturii – pentru a recunoaște că oamenii fac parte din natură; și să respecte nevoile și drepturile altor specii și ale naturii însăși pentru a restabili și regenera ecosistemele sănătoase și rezistente. (GreenComp)	- Ia în considerare transparența atunci când manipulează și prezintă datele pentru a asigura fiabilitatea și identifică datele care sunt exprimate cu motive subiacente (de exemplu, lipsite de etică, profit, manipulare) sau în moduri înșelătoare (DIGCOMP2.2.) - Deschis către sistemele AI care sprijină oamenii să ia decizii informate în conformitate cu obiectivele lor (de exemplu, utilizatorii care decid în mod activ dacă să acționeze pe baza unei recomandări sau nu) (DIGCOMP2.2.) - . Consideră etica (inclusiv, dar fără a se limita la agenția umană și supravegherea, transparența, nediscriminarea, accesibilitatea și părtinirile și corectitudinea) drept unul dintre pilonii de bază în dezvoltarea sau implementarea sistemelor AI. (DIGCOMP2.2.) - Cântărește beneficiile și riscurile înainte de a permite terților să prelucreze date cu caracter personal (de ex. recunoaște că un asistent vocal de pe un smartphone, care este folosit pentru a da comenzi unui robot aspirator, ar putea oferi terților - companii, guverne, infractorii cibernetici - acces la datele). (DIGCOMP2.2.) - ia în considerare consecințele etice ale sistemelor de inteligență artificială pe parcursul ciclului lor de viață: acestea includ atât impactul asupra mediului (consecințele asupra mediului ale producției de dispozitive și servicii digitale), cât și impactul societal, de exemplu, *platformarea* muncii și managementul algoritmic care poate reprima confidențialitatea lucrătorilor sau drepturi; utilizarea forței de muncă cu costuri reduse pentru etichetarea imaginilor pentru a instrui sistemele AI. (DIGCOMP2.2.)

- Conștient de faptul că anumite activități (de exemplu, antrenamentul AI și producerea de criptomonede precum Bitcoin) sunt procese care necesită resurse intensive în ceea ce privește datele și puterea de calcul. Prin urmare, consumul de energie poate fi mare, ceea ce poate avea și un impact ridicat asupra mediului. (DIGCOMP2.2.) - Conștienți de faptul că AI este un produs al inteligenței umane și al procesului decizional (adică oamenii aleg, curățează și codifică datele, proiectează algoritmii, antrenează modelele și organizează și aplică valori umane rezultatelor) și, prin urmare, nu există în mod independent de oameni (DIGCOMP2.2.)		

CU3 DESCRIEREA REZULTATELOR CUNOAȘTERII

Scopul principal al acestui CU este de a împuternici studenții de nivel EQF6 cu o înțelegere profundă a interacțiunii complexe dintre confidențialitate și comoditate în domeniul inteligenței artificiale (AI), promovând o conștientizare critică și abilitățile necesare pentru a naviga și a inova în mod responsabil în acest domeniu. .

3.1. Înțelegeți dihotomia confidențialității și confortului în sistemele AI

Elevii se vor scufunda într-un studiu intensiv al echilibrului delicat dintre confidențialitate și comoditate, două aspecte esențiale care dictează acceptarea și integrarea sistemelor AI în societate. Urmând discursul lui Zuboff (2019), care elucidează „capitalismul de supraveghere" care umbră adesea progresele tehnologice, studenții vor diseca studii de caz și diverse perspective care navighează în această dihotomie.

Explicaţie:

Perspectivă istorică: O privire asupra modului în care amploarea confidențialității și confortului s-au înclinat de-a lungul timpului odată cu evoluția tehnologiilor AI.

Fundamentele filozofice: studenții vor explora dezbaterile filozofice despre confidențialitate și comoditate în era digitală.

Exemplu: Analizarea scenariilor în care progresele tehnologice au încălcat normele de confidențialitate în căutarea confortului și eficienței.

3.2. Identificați și analizați potențialele încălcări ale confidențialității în implementările ai

În acest modul, studenții vor cultiva abilitățile de a identifica și analiza critic potențialele încălcări ale confidențialității în diferite implementări AI. Acest lucru se aliniază cu cercetarea lui Pasquale (2015), care insistă asupra necesității unei „societăți cutie neagră" în care opacitatea **proceselor algoritmice ar trebui analizată pentru a asigura confidențialitatea.**

Explicaţie:

Abilități analitice: dezvoltarea abilităților analitice pentru a identifica potențialele încălcări ale confidențialității în aplicațiile AI.

Perspective juridice: o explorare aprofundată a cadrelor legale care guvernează confidențialitatea în contextul AI.

Exemplu: studii de caz care analizează incidente importante de încălcare a vieții private și examinează lecțiile învățate și măsurile instituite ca răspuns.

3.3. Explorați progresele tehnologice menite să echilibreze confidențialitatea și confortul

Această unitate are ca scop familiarizarea studenților cu cele mai recente progrese tehnologice care urmăresc să mențină un echilibru armonios între intimitate și comoditate. Studenții vor fi încurajați să se bazeze pe perspectivele oferite de diverși experți în domeniu, cum ar fi Nissenbaum (2009), care pledează pentru „confidențialitatea în context" ca mijloc de a înțelege natura nuanțată a preocupărilor legate de confidențialitate în peisajul AI.

Explicaţie:

Tehnologii emergente: o explorare detaliată a tehnologiilor emergente care încearcă să echilibreze confidențialitatea și confortul.

Abordări inovatoare: Analiza abordărilor inovatoare adoptate de diferite sectoare în încorporarea garanțiilor de confidențialitate fără a compromite confortul.

Exemplu: Investigarea unor noi soluții tehnologice, cum ar fi confidențialitatea diferențială, care își propune să ofere utilitate pentru date, păstrând în același timp confidențialitatea.

3.4. Dezvoltați strategii pentru a promova confidențialitatea fără a compromite confortul în sistemele IA

Dând putere studenților să contribuie în mod activ la domeniu, această unitate se concentrează pe dezvoltarea de strategii care promovează confidențialitatea fără a compromite confortul în sistemele AI. Elevii se vor implica cu lucrările unor oameni de știință precum Cavoukian (2010), care a propus conceptul de „Privacy by Design" ca o abordare avansată pentru integrarea confidențialității în dezvoltarea sistemelor AI.

Explicație:

Dezvoltare strategică: încurajarea studenților să dezvolte strategii care integrează măsurile de protecție a confidențialității fără a compromite eficiența și comoditatea oferite de AI.

Considerații etice: Sublinierea considerațiilor etice care ghidează dezvoltarea tehnologiilor AI centrate pe confidențialitate.

Exemplu: studenții se vor implica în ateliere pentru a dezvolta soluții AI care încorporează garanțiile de confidențialitate ca o caracteristică de bază, încurajând inovația responsabilă.

Prin străbaterea complexului peisaj al confidențialității și al confortului în AI, studenții vor fi echipați să analizeze critic, să inoveze și să conducă în dezvoltarea tehnologiilor AI care onorează principiile confidențialității, oferind în același timp o comoditate de neegalat.

CU4 - Etica AI, o abordare practică EQF6 REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOŞTINŢE	APTITUDINI	ATITUDINI

Cunoștințe teoretice/factuale în: CUK4.1. Conștient de principiile etice în dezvoltarea și implementarea AI. CUK4.2. Cunoașteți diferitele cadre și orientări etice și aplicarea acestora în scenarii din lumea reală care implică sisteme AI. CUK4.3. Înțelegeți importanța transparenței, a răspunderii și a angajării părților interesate în procesul de dezvoltare etică a IA.	Rezultatul aptitudinilor 1: Aplicați principiile etice în dezvoltarea și implementarea IA prin abordări practice. 1.1. Aplicați principii etice la proiectarea, dezvoltarea și implementarea sistemelor AI pentru a minimiza părtinirea algoritmică. 1.2. Evaluați sistemele AI în raport cu orientările și principiile etice pentru a identifica potențialele părtiniri și preocupările etice. 1.3. Dezvoltați strategii pentru a aborda problemele etice în aplicațiile AI, inclusiv potențiale conflicte între principii și compromisuri. Rezultatul abilității 2: Analizați și evaluați diferite cadre etice și linii directoare în scenarii din lumea reală care implică sisteme AI. 2.1. Analizați diferite cadre etice și linii directoare relevante pentru IA și părtinirea algoritmică. 2.2. Aplicați cadre etice scenariilor AI din lumea reală pentru a evalua dimensiunile etice ale sistemelor AI. 2.3. Comparați și comparați diferite cadre etice pentru a determina adecvarea acestora în abordarea părtinirii algoritmice în anumite aplicații AI. Rezultatul aptitudinilor 3: Proiectați și dezvoltați sisteme etice de inteligență artificială prin promovarea transparenței, a răspunderii și a implicării părților interesate. 3.1. Proiectați sisteme AI care promovează transparența, permițând utilizatorilor și părților interesate să înțeleagă procesul de luare a deciziilor. 3.2. Implementați măsuri de responsabilitate pentru a vă asigura că dezvoltatorii și organizațiile AI își asumă responsabilitatea pentru consecințele sistemelor lor AI. 3.3. Facilitează implicarea părților interesate în dezvoltarea IA, inclusiv solicitând contribuții din diverse perspective și luând în considerare nevoile diferitelor comunități afectate de sistemele AI.	Rezultatul atitudinii 1: Atitudine etică în dezvoltarea și implementarea AI. 1.1. Dezvoltați simțul responsabilității față de aplicarea principiilor etice în dezvoltarea IA pentru a minimiza părtinirea algoritmică. 1.2. Cultivați o mentalitate etică care prețuiește corectitudinea, incluziunea și justiția în aplicațiile AI. 1.3. Apreciați importanța abordării preocupărilor etice în mod proactiv pentru a preveni consecințele nedorite ale sistemelor AI. Rezultatul atitudinii 2: Curiozitatea și deschiderea de a învăța despre diferite cadre etice și linii directoare în scenarii din lumea reală care implică sisteme AI. 2.1. Arătați curiozitate și deschidere pentru a învăța despre diferite cadre etice și relevanța acestora pentru dezvoltarea AI. 2.2. Apreciază importanța orientărilor etice în modelarea proiectării și implementării sistemelor AI. 2.3. Îmbrățișați o abordare reflexivă a evaluării și perfecționării aplicațiilor AI pe baza considerentelor etice. Rezultatul atitudinii 3: O mentalitate încurajatoare și entuziastă în promovarea transparenței, a răspunderii și a angajării părților interesate în dezvoltarea etică a IA. 3.1. Promovați angajamentul de a promova transparența în sistemele AI, permițând utilizatorilor și părților interesate să înțeleagă și să aibă încredere în procesul decizional AI. 3.2. Evaluează importanța responsabilității în dezvoltarea IA, recunoscând nevoia ca organizațiile să-și asume responsabilitatea pentru consecințele sistemelor lor de IA. 3.3. Încurajează implicarea activă a părților interesate și colaborarea în dezvoltarea IA, apreciind valoarea diverselor perspective și includerea comunităților afectate.
COMPETENȚE ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP DE CARE CU4 ESTE LEGAT - Știe că IA în sine nu este nici bună, nici rea. Ceea ce determină dacă rezultatele unui sistem AI sunt pozitive sau negative pentru societate este modul în care este proiectat și utilizat sistemul AI, de către cine și în ce scopuri. (AI) (DigiComp 2.2.) - Știe că prelucrarea datelor cu caracter personal este supusă reglementărilor locale, cum ar fi	**ABILITĂȚI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU4 ESTE LEGAT** - Știe cum să identifice domeniile în care AI poate aduce beneficii diferitelor aspecte ale vieții de zi cu zi. De exemplu, în domeniul sănătății, IA ar putea contribui la diagnosticarea precoce, în timp ce în agricultură, ar putea fi folosită pentru a detecta infestările cu dăunători. (AI). (DigiComp 2.2.) - Lucrul cu alții (de exemplu, firele de discuții lucrează împreună, formează-ți echipă, extinde-ți rețeaua): Fă echipă,	**ATITUDINI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU4 ESTE LEGAT** - Disponibilitatea de a analiza întrebări etice legate de sistemele AI (de exemplu, în ce contexte, cum ar fi condamnarea infractorilor, recomandările AI nu ar trebui utilizate fără intervenția umană)? (AI) (DigiComp 2.2.) - Deschis către sistemele AI care sprijină oamenii să ia decizii informate în conformitate cu obiectivele lor (de exemplu, utilizatorii care decid în mod activ dacă să

Regulamentul general privind protecția datelor (GDPR) al UE (de exemplu, interacțiunile vocale cu un asistent virtual sunt date personale în ceea ce privește GDPR și pot expune utilizatorii la anumite protecție a datelor, confidențialitate şi riscuri de securitate) . (AI) (DigiComp 2.2.)	lucrează împreună și rețea. Lucrați împreună și cooperați cu alții pentru a dezvolta idei și a le transforma în acțiune. Rețea. Rezolvați conflictele și faceți față concurenței în mod pozitiv atunci când este necesar. (EntreComp) Încorporarea valorilor de durabilitate: 1.2 Sprijinirea echității – pentru a sprijini echitatea și justiția pentru generațiile actuale și viitoare și pentru a învăța de la generațiile anterioare pentru durabilitate. 1.3 Promovarea naturii – pentru a recunoaște că oamenii fac parte din natură; și să respecte nevoile și drepturile altor specii și ale naturii însăși pentru a restabili și regenera ecosistemele sănătoase și rezistente. (GreenComp)	acționeze după o recomandare sau nu). (AI). (DigiComp 2.2.) - Consideră etica (inclusiv, dar fără a se limita la agenția umană și supravegherea, transparența, nediscriminarea, accesibilitatea și părtinirile și corectitudinea) drept unul dintre pilonii de bază în dezvoltarea sau implementarea sistemelor AI. (AI) (DigiComp 2.2.) - Deschis să se implice în procese de colaborare pentru co-proiectarea și co-crearea de noi produse și servicii bazate pe sisteme de inteligență artificială pentru a sprijini și a spori participarea cetățenilor în societate. (AI) (DigiComp 2.2.) Creativitate (de exemplu, firul de a fi curios și deschis): Dezvoltați idei creative și cu scop. Dezvoltați mai multe idei și oportunități pentru a crea valoare, inclusiv soluții mai bune la provocările existente și noi. Explorați și experimentați cu abordări inovatoare. Combinați cunoștințele și resursele pentru a obține efecte valoroase. (EntreComp) Mobilizarea celorlalți: Inspirați și entuziasmați părțile interesate relevante. Obțineți sprijinul necesar pentru a obține rezultate valoroase. Demonstrați comunicare eficientă, persuasiune, negociere și leadership. (EntreComp)

CU4 DESCRIEREA REZULTATELOR CUNOAȘTERII

4.1. Conștient de principiile etice în dezvoltarea și implementarea AI.

Conștient de principiile etice în dezvoltarea și implementarea AI. Aplicarea principiilor etice la proiectarea, dezvoltarea și implementarea sistemelor AI pentru a minimiza părtinirea algoritmică. Evaluarea sistemelor AI în raport cu orientările și principiile etice pentru a identifica potențialele părtiniri și preocupări etice. Dezvoltarea de strategii pentru a aborda problemele etice în aplicațiile AI, inclusiv potențiale conflicte între principii și compromisuri.

Principiile etice ale IA: principiile etice urmăresc să ofere o bază pentru ca sistemele AI să funcționeze pentru binele umanității, al indivizilor, al societăților și al mediului și al ecosistemelor și pentru a preveni daunele. Ele ghidează proiectarea, dezvoltarea, implementarea și utilizarea sistemelor AI într-un mod care este demn de încredere și pune demnitatea umană, egalitatea tuturor ființelor umane, conservarea mediului, a biodiversității și a ecosistemelor, respectul pentru diversitatea culturală și responsabilitatea datelor la nivelul centru (UNESCO, 2022). Principiile etice descriu ceea ce se așteaptă în termeni de bine și rău și alte standarde etice. Principiile etice ale

inteligenței artificiale se referă la constrângerile normative cu privire la „ceea ce nu se face” și „nu se face” ale utilizării algoritmice în societate (Zhou et al., 2020). De exemplu, conform Grupului de experți la nivel înalt pentru IA (AI HLEG), astfel de principii includ: respectul pentru autonomia umană, prevenirea vătămării, corectitudinea și explicabilitatea (Comisia Europeană, 2019).

Aplicarea practică a principiilor etice în IA . Odată identificate, principiile etice ar trebui transpuse în seturi de instrumente și linii directoare viabile pentru a modela inovația bazată pe IA și pentru a sprijini aplicarea practică a principiilor etice ale IA. Seturile de instrumente și orientările privind modul de aplicare a principiilor etice în proiectarea, implementarea și implementarea AI sunt extrem de necesare. Algoritmii, arhitecturile și interfețele AI etice urmează principiile etice ale AI. (Zhou et al., 2020.) Pe parcursul activității lor de dezvoltare a principiilor etice ale IA, multe organizații private, publice și non-profit au identificat acțiuni utile care pot ajuta la articularea și implementarea principiilor propuse (de exemplu, UNESCO, 2021).

De exemplu, principiile subliniate de AI HLEG trebuie transpuse în cerințe concrete pentru a obține o IA de încredere. Aceste cerințe sunt aplicabile diferitelor părți interesate care participă la ciclul de viață al sistemelor AI: dezvoltatori, implementatori și utilizatori finali, precum și societatea în general. Diferitele grupuri de părți interesate au roluri diferite de jucat în asigurarea îndeplinirii cerințelor: *dezvoltatorii* ar trebui să implementeze și să aplice cerințele proceselor de proiectare și dezvoltare; *implementatorii* ar trebui să se asigure că sistemele pe care le folosesc și produsele și serviciile pe care le oferă îndeplinesc cerințele; Utilizatorii finali *și societatea în general* ar trebui să fie informați cu privire la aceste cerințe și să poată solicita ca acestea să fie respectate. Lista de cerințe de mai jos este neexhaustivă. Include aspecte sistemice, individuale și societale:

1. Agenția umană și supravegherea (adică drepturile fundamentale, agenția umană și supravegherea umană).
2. Robustețe și siguranță tehnică (adică rezistență la atac și securitate, plan de retragere și siguranță generală, acuratețe, fiabilitate și reproductibilitate).
3. Confidențialitate și guvernare a datelor (adică respectul pentru confidențialitate, calitatea și integritatea datelor și accesul la date).
4. Transparență (adică, trasabilitate, explicabilitate și comunicare).

5. Diversitate, nediscriminare și corectitudine (adică evitarea părtinirii neloiale, accesibilitatea și designul universal și participarea părților interesate).
6. Bunăstarea societății și a mediului (adică, durabilitatea și respectarea mediului, impactul social, societatea și democrația).
7. Responsabilitate (adică auditabilitate, minimizarea și raportarea impactului negativ, compromisuri și reparații). (Comisia Europeană, 2019.)

Morley şi colab. (2019) a construit o tipologie prin combinarea principiilor etice cu etapele ciclului de viață AI pentru a se asigura că sistemul AI este proiectat, implementat și implementat într-o manieră etică. Tipologia indică faptul că fiecare principiu etic ar trebui luat în considerare în fiecare etapă a ciclului de viață AI. Tipologia oferă un scurt instantaneu al instrumentelor disponibile în prezent pentru dezvoltatorii de IA pentru a încuraja progresul AI etic de la principii la practică. Tipologia completă din Morley et al. poate fi găsit de la https://tinyurl.com/AppliedAIEthics.

4.2. Cunoașteți diferitele cadre și orientări etice și aplicarea acestora în scenarii din lumea reală care implică sisteme AI.

Cunoașteți diferitele cadre și orientări etice și aplicarea acestora în scenarii din lumea reală care implică sisteme AI. Analizând diferite cadre etice și linii directoare relevante pentru IA și părtinirea algoritmică. Aplicarea cadrelor etice scenariilor AI din lumea reală pentru a evalua dimensiunile etice ale sistemelor AI. Compararea și contrastarea diferitelor cadre etice pentru a determina adecvarea acestora în abordarea părtinirii algoritmice în aplicații specifice AI.

Cadre și orientări pentru etica AI: Cadrurile și orientările etice constau din valori, principii, standarde și acțiuni pentru a ghida indivizii, grupurile, comunitățile, instituțiile și companiile din sectorul privat pentru a asigura încorporarea eticii în toate etapele ciclului de viață al sistemului AI. . Acestea urmăresc să protejeze, să promoveze și să respecte drepturile omului și libertățile fundamentale, demnitatea umană și egalitatea; pentru a proteja interesele generațiilor prezente și viitoare; pentru a conserva mediul, biodiversitatea și ecosistemele; și să respecte diversitatea culturală în toate etapele ciclului de viață al sistemului AI (Comisia Europeană, 2019; UNESCO, 2021).

Diferite seturi de principii și cadre etice pentru IA sunt publicate de obicei din industrie (de exemplu, Google, IBM, Microsoft, Intel), guvern (de exemplu, UK

Lords Select Committee, Grupul de experți la nivel înalt al Comisiei Europene) și din mediul academic (de exemplu, Future of Life Institute, IEEE, AI4People) (Zhou et al., 2020) . Niciun principiu etic unic nu este susținut în mod explicit de toate cadrele și liniile directoare etice existente, dar există o convergență emergentă în jurul următoarelor principii: transparență, dreptate și corectitudine, responsabilitate, non-malefință, confidențialitate, binefacere, libertate și autonomie, încredere, durabilitate, demnitate și solidaritate (Jobin et al, 2019).

UNESCO (2021) a publicat primul standard global privind etica IA – Recomandarea privind etica inteligenței artificiale în noiembrie 2021. Acest cadru a fost adoptat de toate cele 193 de state membre. Protecția drepturilor omului și a demnității este piatra de temelie a Recomandării, bazată pe promovarea principiilor fundamentale ale acestora.

4.3. Înțelegeți importanța transparenței, a răspunderii și a angajării părților interesate în procesul de dezvoltare etică a IA

Înțelegerea importanței transparenței, responsabilității și implicării părților interesate în dezvoltarea etică a IA. Proiectarea sistemelor AI care promovează transparența, permițând utilizatorilor și părților interesate să înțeleagă procesul de luare a deciziilor. Implementarea măsurilor de responsabilitate pentru a se asigura că dezvoltatorii și organizațiile AI își asumă responsabilitatea pentru consecințele sistemelor lor AI. Facilitarea implicării părților interesate în dezvoltarea IA, inclusiv solicitarea de contribuții din diverse perspective și luarea în considerare a nevoilor diferitelor comunități afectate de sistemele AI.

Transparență: transparența (explicabilitatea) este crucială pentru construirea și menținerea încrederii utilizatorilor în sistemele AI. Aceasta înseamnă că procesele trebuie să fie explicite, capacitățile și scopul sistemelor AI comunicate în mod deschis, iar deciziile – în măsura posibilului – explicabile celor afectați direct și indirect. Fără astfel de informații, o decizie nu poate fi contestată în mod corespunzător.

O explicație a motivului pentru care un model a generat o anumită ieșire sau o decizie (și ce combinație de factori de intrare a contribuit la aceasta) nu este întotdeauna posibilă. Aceste cazuri sunt denumite algoritmi „cutie neagră” și necesită o atenție specială. În aceste circumstanțe, pot fi necesare alte măsuri de explicabilitate (de exemplu, trasabilitatea, auditabilitatea și comunicarea transparentă privind capacitățile sistemului), cu condiția ca sistemul în ansamblu să respecte drepturile fundamentale. Gradul în care este necesară explicabilitatea depinde în mare măsură de context și de gravitatea

consecințelor în cazul în care rezultatul este eronat sau inexact în alt mod (Comisia Europeană, 2019.)

Responsabilitate: Responsabilitatea este una dintre pietrele de temelie ale guvernării inteligenței artificiale (AI). Responsabilitatea are multe definiții, dar, în esență, este o obligație de a informa și de a justifica comportamentul unei autorități în fața unei autorități (Novelli, Taddeo & Floridi , 2023). În general, responsabilitatea în IA se referă la așteptarea ca proiectanții, dezvoltatorii și implementatorii să respecte standardele și legislația pentru a asigura funcționarea corespunzătoare a AI pe parcursul ciclului lor de viață (Fjeld et al., 2020).

Ar trebui dezvoltate mecanisme adecvate de supraveghere, evaluare a impactului, audit și due diligence, inclusiv protecția avertizorilor, pentru a asigura responsabilitatea pentru sistemele IA și impactul acestora pe parcursul ciclului lor de viață. Atât modelele tehnice, cât și cele instituționale ar trebui să asigure auditabilitatea și trasabilitatea (funcționării) sistemelor AI, în special pentru a aborda orice conflicte cu normele și standardele privind drepturile omului și amenințările la adresa bunăstării mediului și a ecosistemului (UNESCO, 2021)

Implicarea părților interesate: este important să se consulte grupurile de părți interesate relevante în timp ce se dezvoltă și se gestionează utilizarea aplicațiilor AI (Fjeld et al., 2020). Participarea diferitelor părți interesate de-a lungul ciclului de viață al sistemului AI este necesară pentru abordări incluzive ale guvernării AI, permițând ca beneficiile să fie împărtășite de toți și să contribuie la dezvoltarea durabilă. Părțile interesate includ, dar nu se limitează la, guverne, organizații interguvernamentale, comunitatea tehnică, societatea civilă, cercetători și mediul academic, mass-media, educație, factori de decizie, companii din sectorul privat, instituții pentru drepturile omului și organisme de egalitate, organisme de monitorizare anti-discriminare și grupuri. pentru tineri și copii (UNESCO, 2021)

Implicarea părților interesate ajută la asigurarea faptului că tehnologiile AI sunt dezvoltate și implementate într-un mod care se aliniază cu valorile, nevoile și preocupările diferitelor persoane și grupuri afectate de tehnologie. Cu toate acestea, în practică, poate fi dificil să se identifice persoanele și organizațiile afectate de dezvoltarea și utilizarea sistemului AI. Aceste probleme devin și mai importante atunci când sistemele AI sunt dezvoltate în proiecte, care sunt organizații temporare ai căror membri ai echipei pot să nu fie disponibili odată ce proiectul se încheie. Unele sisteme AI rulează automat și nu permit intervenția umană, ceea ce le face unice față de alte sisteme

informaționale. Aceste sisteme pot avea impact asupra drepturilor omului, civile și ale lucrătorilor; au caracteristici apropiate de industriile farmaceutice și de sănătate pentru impactul uman și proiectele de infrastructură pentru impactul asupra mediului (Miller, 2022).

CU5 - Studii de caz și proiect EQF6 **REZULTATELE ÎNVĂȚĂRII**		
	La finalizarea experienței de formare gamificată, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK5.1. Înțelegeți prejudecățile algoritmice și implicațiile sale în lumea reală. CUK5.2. Conștient de complexitățile și nuanțele părtinirii algoritmice. CUK5.3. Recunoașteți strategiile și instrumentele disponibile pentru a aborda problemele de părtinire algoritmică în practică.	Rezultatul aptitudinilor 1: Aplicați cunoștințele despre prejudecățile algoritmice în studiile de caz din lumea reală și la un proiect practic. 1.1. Investigați studii de caz din lumea reală pentru a identifica cazurile de părtinire algoritmică și consecințele acestora. 1.2. Dezvoltați soluții pentru a aborda prejudecățile algoritmice în studii de caz, luând în considerare implicațiile etice și practice. 1.3. Proiectați, implementați și evaluați un proiect practic pentru a aborda prejudecățile algoritmice într-un domeniu ales. Rezultatul aptitudinilor 2: Analizați părtinirea algoritmică, cauzele și impactul acesteia. 2.1. Analizați diferite tipuri de prejudecăți algoritmice și cauzele lor fundamentale. 2.2. Evaluați impactul potențial al părtinirii algoritmice asupra indivizilor și societății în ansamblu. 2.3. Înțelegeți provocările și limitările în identificarea, măsurarea și atenuarea distorsiunilor algoritmice. Rezultatul abilității 3: Examinați și integrați strategiile și instrumentele disponibile pentru a aborda problemele de părtinire algoritmică. 3.1. Cercetați și evaluați strategiile existente pentru abordarea prejudecăților algoritmice, inclusiv învățarea automată conștientă de corectitudine și auditurile de părtinire. 3.2. Dezvoltați competența în utilizarea instrumentelor și tehnicilor pentru a detecta, măsura și atenua prejudecățile algoritmice în sistemele AI. 3.3. Integrați abordări interdisciplinare, cum ar fi contribuțiile experților din domeniu și diverse perspective, pentru a îmbunătăți corectitudinea sistemelor AI.	Atitudinea Rezultatul 1: Angajamentul de a rezolva problemele în abordarea părtinirii algoritmice și a implicațiilor sale în lumea reală. 1.1. Demonstrați angajamentul de a aplica cunoștințele teoretice în situații practice pentru a aborda prejudecățile algoritmice. 1.2. Apreciază importanța experienței practice pentru a obține o înțelegere mai profundă a părtinirii algoritmice și a implicațiilor sale în lumea reală. 1.3. Dezvoltați o mentalitate de rezolvare a problemelor care urmărește să creeze soluții AI corecte și echitabile. Rezultatul atitudinii 2: Atitudine critică și colaborativă în recunoașterea complexităților și a nuanțelor prejudecăților algoritmice. 2.1. Încurajează empatia față de indivizii și comunitățile afectate de părtinire algoritmică. 2.2. Cultivați o mentalitate critică care recunoaște complexitățile și limitările abordării părtinirii algoritmice. 2.3. Apreciază importanța colaborării interdisciplinare și a diverselor perspective în abordarea provocărilor prejudecăților algoritmice. Rezultatul atitudinii 3: Atitudine deschisă și colaborativă în utilizarea strategiilor și instrumentelor disponibile pentru a aborda problemele de părtinire algoritmică în practică. 3.1. Îmbrățișați o abordare proactivă în utilizarea strategiilor și instrumentelor disponibile pentru a combate părtinirea algoritmică. 3.2. Valorați importanța învățării continue și a rămâne informat cu privire la noile evoluții în abordarea părtinirii algoritmice. 3.3. Încurajează un mediu de colaborare care încurajează dialogul deschis, schimbul de cunoștințe și inovația în abordarea problemelor de părtinire algoritmică.

COMPETENȚE ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU5 ESTE LEGAT - Conștient de faptul că algoritmii AI ar putea să nu fie configurați pentru a furniza doar informațiile pe care utilizatorul le dorește; ar putea, de asemenea, să încorporeze un mesaj comercial sau politic (de exemplu, pentru a încuraja utilizatorii să rămână pe site, să urmărească sau să cumpere ceva anume, să împărtășească opinii specifice). Acest lucru poate avea și consecințe negative (de exemplu, reproducerea stereotipurilor, împărtășirea informațiilor greșite). (AI) (DigiComp 2.2.) - Conștient de faptul că datele, de care depinde AI, pot include părtiniri. Dacă da, aceste părtiniri pot deveni automatizate și agravate prin utilizarea AI. De exemplu, rezultatele căutării despre ocupație pot include stereotipuri despre locuri de muncă masculine sau feminine (de exemplu, șoferi de autobuz de sex masculin, vânzători de sex feminin). (AI) (DigiComp 2.2.) - Știe că datele colectate și prelucrate, de exemplu, de sisteme online, pot fi utilizate pentru a recunoaște modele (de exemplu, repetări) în date noi (de exemplu, alte imagini, sunete, clicuri de mouse, comportamente online) pentru a optimiza și personaliza în continuare serviciile online (de exemplu, reclame). (DigiComp 2.2.) - Conștient de faptul că senzorii utilizați în multe tehnologii și aplicații digitale (de exemplu, camere de urmărire facială, asistenți virtuali, tehnologii portabile, telefoane mobile, dispozitive inteligente) generează cantități mari de date, inclusiv date personale, care pot fi utilizate pentru a antrena un sistem AI. (AI) (DigiComp 2.2.) - Conștient de faptul că tot ceea ce se partajează public online (de exemplu, imagini, videoclipuri, sunete) poate fi folosit pentru a antrena sisteme AI. De exemplu,	**ABILITĂȚI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU5 ESTE LEGAT** -Gândire etică și durabilă (de exemplu, firele se comportă etic, să fie responsabil, să evalueze impactul): Evaluează consecințele și impactul ideilor, oportunităților și acțiunilor. (...) Acționați responsabil. (EntreComp) Încorporarea valorilor de durabilitate: 1.2 Sprijinirea echității – pentru a sprijini echitatea și justiția pentru generațiile actuale și viitoare și pentru a învăța de la generațiile anterioare pentru durabilitate. 1.3 Promovarea naturii – pentru a recunoaște că oamenii fac parte din natură; și să respecte nevoile și drepturile altor specii și ale naturii însăși pentru a restabili și regenera ecosistemele sănătoase și rezistente. (GreenComp)	**ATITUDINI ALE CADRURILOR DIGCOMP2.2., ENTRECOMP ȘI GREENCOMP CU CARE CU5 ESTE LEGAT** - Deschis către sistemele AI care sprijină oamenii să ia decizii informate în conformitate cu obiectivele lor (de exemplu, utilizatorii care decid în mod activ dacă să acționeze după o recomandare sau nu). (AI) DigiComp 2.2.) - Disponibilitatea de a analiza întrebări etice legate de sistemele AI (de exemplu, în ce contexte, cum ar fi condamnarea infractorilor, recomandările AI nu ar trebui utilizate fără intervenția umană)? (AI) (DigiComp 2.2.) - Identifică atât implicațiile pozitive, cât și negative ale utilizării tuturor datelor (colectarea, codificarea și prelucrarea), dar mai ales datele personale, de către tehnologiile digitale bazate pe inteligență artificială, cum ar fi aplicațiile și serviciile online. (AI) (DigiComp 2.2.) - ia în considerare consecințele etice ale sistemelor de inteligență artificială pe parcursul ciclului lor de viață: acestea includ atât impactul asupra mediului (consecințele asupra mediului ale producției de dispozitive și servicii digitale), cât și impactul societal, de exemplu, *platformarea* muncii și managementul algoritmic care poate reprima confidențialitatea lucrătorilor sau drepturi; utilizarea forței de muncă cu costuri reduse pentru etichetarea imaginilor pentru a instrui sistemele AI. (AI) (DigiComp 2.2.) - Deschis să se implice în procese de colaborare pentru a proiecta și a crea în comun noi produse și servicii bazate pe sisteme AI pentru a sprijini și a spori participarea cetățenilor în societate. (AI) (DigiComp 2.2.)

companiile comerciale de software care dezvoltă sisteme de recunoaștere facială AI pot folosi imagini personale partajate online (de exemplu, fotografii de familie) pentru a instrui și îmbunătăți capacitatea software-ului de a recunoaște automat acele persoane în alte imagini, ceea ce ar putea să nu fie de dorit (de exemplu, ar putea fi o încălcare a confidențialității). (AI) (DigiComp 2.2.) - recunoaște că, deși aplicarea sistemelor AI în multe domenii este de obicei necontroversată (de exemplu, IA care ajută la prevenirea schimbărilor climatice), IA care interacționează direct cu oamenii și ia deciziile cu privire la viața lor pot fi adesea controversate (de exemplu, software-ul de sortare a CV-urilor pentru procedurile de recrutare, notarea examenelor care pot determina accesul la educație). (AI) (DigiComp 2.2.) - Capabil să identifice câteva exemple de sisteme AI: recomandări de produse (de exemplu, pe site-uri de cumpărături online), recunoaștere a vocii (de exemplu, de către asistenți virtuali), recunoaștere a imaginii (de exemplu, pentru detectarea tumorilor în raze X) și recunoaștere facială (de exemplu, în sistemele de supraveghere) . (AI) (DigiComp 2.2.)		

CU5 DESCRIEREA REZULTATELOR CUNOAȘTERII

5.1. Înțelegeți prejudecățile algoritmice și implicațiile sale în lumea reală

Înțelegerea părtinirii algoritmice și a implicațiilor sale în lumea reală. Aplicarea cunoștințelor de părtinire algoritmică în studii de caz din lumea reală și un proiect practic. Investigarea studiilor de caz din lumea reală pentru a identifica cazurile de părtinire algoritmică și consecințele acestora. Dezvoltarea de soluții pentru a aborda prejudecățile algoritmice în studii de caz, luând în considerare implicațiile etice și practice. Proiectarea, implementarea și evaluarea unui proiect practic pentru a aborda prejudecățile algoritmice într-un domeniu ales.

Prejudecăți algoritmice: părtinirea algoritmică se referă la prezența unor rezultate incorecte sau discriminatorii în rezultatele produse de un sistem

algoritmic. Apare atunci când un algoritm favorizează sau dezavantajează în mod constant și sistematic anumite persoane sau grupuri pe baza unor caracteristici specifice, cum ar fi rasa, sexul sau statutul socio-economic. În învățarea automată, algoritmii se bazează pe mai multe seturi de date, adică pe date de antrenament care specifică care sunt rezultatele corecte pentru anumite persoane sau obiecte. Din acel antrenament, mașina de date învață un model care poate fi aplicat altor persoane sau obiecte și face predicții despre posibilele rezultate pentru aceștia. Cu toate acestea, mașinile pot trata în mod diferit oamenii și obiectele aflate în poziții similare. Din cauza acestui deficit, unii algoritmi riscă să reproducă și chiar să amplifice părtinirile umane, în special pe cei care afectează grupurile protejate (Friedman & Nissenbaum, 1996; Lange & Duarte, 2017; Lee, Resnick & Barton, 2019).

Implicațiile părtinirii algoritmice: Prejudecățile algoritmice se pot manifesta în mai multe moduri, cu grade diferite de consecințe pentru grupul de subiecte. Următoarele exemple ilustrează atât o serie de cauze, cât și de efecte care fie aplică din neatenție un tratament diferit grupurilor, fie generează în mod deliberat un impact disparat asupra acestora (Lee, Resnick & Barton, 2019):

- *Prejudecăți în instrumentele de recrutare online* – Retailerul online Amazon a întrerupt recent utilizarea unui algoritm de recrutare după ce a descoperit părtinirea de gen. Software-ul AI a penalizat orice CV care conținea cuvântul „femei” în text și a retrogradat CV-urile femeilor care au urmat colegiile pentru femei, ceea ce a dus la părtinire de gen.
- *Prejudecăți în asocierile de cuvinte* – Cercetătorii de la Universitatea Princeton au folosit un software AI de învățare automată standard pentru a analiza și lega 2,2 milioane de cuvinte. Ei au descoperit că numele europene erau percepute ca fiind mai plăcute decât cele ale afro-americanilor și că cuvintele „femeie” și „fată” erau mai probabil asociate cu artele în loc de știință și matematică, care erau cel mai probabil conectate la bărbați. .
- *Prejudecăți în reclamele online* – Latanya Sweeney, cercetător la Harvard și fost director de tehnologie la Federal Trade Commission (FTC), a constatat că interogările de căutare online pentru nume afro-americane aveau mai multe șanse să returneze reclame acelei persoane de la un serviciu care redă înregistrări de arestare. , în comparație cu rezultatele anunțurilor pentru nume albe.
- *Prejudecăți în tehnologia de recunoaștere facială* – cercetătorul MIT Joy Buolamwini a descoperit că algoritmii care alimentează trei sisteme

software de recunoaștere facială disponibile în comerț nu reușeau să recunoască tenul cu pielea mai închisă. În general, cele mai multe seturi de date de antrenament pentru recunoașterea facială sunt estimate a fi mai mult de 75% bărbați și peste 80% albi. Când persoana din fotografie era un bărbat alb, software-ul a fost precis în 99 la sută din timp la identificarea persoanei ca bărbat.

- *Prejudecăți în algoritmii de justiție penală* - Algoritmul COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), care este folosit de judecători pentru a prezice dacă inculpații ar trebui să fie reținuți sau eliberați pe cauțiune în așteptarea procesului, a fost considerat a fi părtinitor împotriva afro-americanilor, potrivit un reportaj de la ProPublica.

5.2. Conștient de complexitățile și nuanțele părtinirii algoritmice

Conștient de complexitățile și nuanțele părtinirii algoritmice. Analiza părtinirii algoritmice, cauzele și impactul acesteia. Analizarea diferitelor tipuri de prejudecăți algoritmice și cauzele lor fundamentale. Evaluarea impactului părtinirii algoritmice. Înțelegerea limitărilor în identificarea, măsurarea și atenuarea părtinirii algoritmice.

Tipuri și cauze ale părtinirii algoritmice: Există diferite forme de părtinire algoritmică. Potrivit Friedman și Nissenbaum (1996), diferitele tipuri și cauze ale distorsiunilor algoritmice pot fi împărțite în trei categorii:

1. Prejudecățile preexistente își au rădăcinile în instituțiile, practicile și atitudinile sociale. Atunci când sistemele informatice încorporează prejudecăți care există în mod independent și, de obicei, înainte de crearea sistemului, atunci sistemul exemplifică prejudecățile preexistente. Prejudecățile preexistente pot intra într-un sistem fie prin eforturile explicite și conștiente ale unor indivizi sau instituții, fie implicit și inconștient, chiar și în ciuda celor mai bune intenții.

2. Prejudecățile tehnice rezultă din constrângeri tehnice sau considerații tehnice. Surse de părtinire tehnică pot fi găsite în mai multe aspecte ale procesului de proiectare, inclusiv limitări ale instrumentelor informatice, cum ar fi hardware, software și periferice; procesul de atribuire a unui sens social algoritmilor dezvoltați în afara contextului; imperfecțiuni în generarea numerelor pseudoaleatoare; și încercarea de a face constructele umane

adaptabile computerelor, atunci când cuantificăm calitativul, discretizăm continuul sau formalizăm non-formalul.

3. Prejudecățile emergente apar într-un context de utilizare cu utilizatori reali. Această părtinire apare de obicei cândva după finalizarea unui proiect, ca urmare a schimbării cunoștințelor societale, a populației sau a valorilor culturale. Este posibil ca interfețele de utilizator să fie deosebit de predispuse la părtinire emergentă, deoarece interfețele prin design caută să reflecte capacitățile, caracterul și obiceiurile potențialilor utilizatori. Astfel, o schimbare în contextul utilizării poate crea dificultăți pentru un nou set de utilizatori. Prejudecățile emergente provin din apariția unor noi cunoștințe în societate care nu pot fi sau nu sunt încorporate în sistem sau atunci când populația care utilizează sistemul diferă într-o anumită dimensiune semnificativă de populația presupusă ca utilizatori în proiectare, adică nepotrivirea dintre utilizatori și proiectarea sistemului. .

5.3. Recunoașteți strategiile și instrumentele disponibile pentru a aborda problemele de părtinire algoritmică în practică

Recunoașterea strategiilor și instrumentelor disponibile pentru a aborda problemele de părtinire algoritmică, inclusiv învățarea automată care conștientizează corectitudinea și auditurile de părtinire. Revizuirea și integrarea strategiilor și instrumentelor disponibile pentru a aborda problemele de părtinire algoritmică în practică. Cercetarea și evaluarea strategiilor existente pentru abordarea prejudecăților algoritmice, inclusiv învățarea automată care conștientizează corectitudinea și auditurile de părtinire. Dezvoltarea competenței în utilizarea instrumentelor și tehnicilor pentru a detecta, măsura și atenua părtinirea algoritmică în sistemele AI. Integrarea abordărilor interdisciplinare, cum ar fi contribuțiile experților din domeniu și diverse perspective, pentru a îmbunătăți corectitudinea sistemelor AI.

Strategii și instrumente pentru abordarea problemelor de părtinire algoritmică: Abordarea distorsiunii algoritmice necesită o analiză atentă la diferite etape ale dezvoltării algoritmului. Aceasta implică asigurarea unor date de instruire diverse și reprezentative, identificarea și atenuarea părtinirilor în proiectarea algoritmului și procesul de luare a deciziilor și efectuarea de teste și evaluări riguroase pentru a detecta și rectifica problemele legate de părtinire. S-au făcut unele eforturi pentru a minimiza

rezultatele discriminatorii prin dezvoltarea cadrelor și a liniilor directoare pentru atenuarea prejudecăților algoritmice, inclusiv o transparență și responsabilitate sporite în sistemele algoritmice, măsuri de reglementare și prin adoptarea unor tehnici de învățare automată conștiente de corectitudine (Lee, Resnick & Barton, 2019).

4.2. CHARLIE WP2 „Ethical AI micro credential" EQF4

Acest program educațional aprofundat este conceput pentru a oferi o înțelegere cuprinzătoare a prejudecăților algoritmice și a implicațiilor acestora în societatea actuală bazată pe date. Întrucât algoritmii joacă un rol semnificativ în modelarea vieții noastre și în informarea luării deciziilor în diferite sectoare, inclusiv în domeniul sănătății, educației, finanțelor și guvernului, este esențial pentru profesioniști și cadre universitare/studenți să dezvolte o înțelegere solidă a potențialelor părtiniri în cadrul acestor sisteme.
Obiectivul acestui curs este de a dota participanții cu cunoștințele și abilitățile necesare pentru a identifica, atenua și aborda prejudecățile algoritmice. Printr-o combinație de fundamente teoretice și aplicații practice, cursanții vor obține informații despre aspectele etice, sociale și tehnice ale părtinirii algoritmice.
Aceste obiective contribuie la obiectivele generale prin stabilirea pietrelor de temelie pentru următoarele activități, permițând profesorilor, mentorilor și cursanților să aibă o viziune clară asupra obiectivelor diferitelor căi de învățare, lucrând pentru a crește capacitatea organizațiilor de învățământ superior, de adulți și de tineret de a oferi oportunități de învățare. care răspund nevoilor societății, dar sunt, de asemenea, adaptate nevoilor de învățare ale elevilor. Concret, rezultatele așteptate ale acestui EQF4 „Ethical AI micro credential" vizează:

1- Dezvoltați o înțelegere fundamentală a părtinirii algoritmice, a cauzelor și a consecințelor acesteia asupra indivizilor și societății.
- 1.1. Dobândiți cunoștințe despre definiția, sursele și manifestările părtinirii algoritmice.
- 1.2. Înțelegeți implicațiile sociale și individuale ale algoritmilor părtinitori.

2. Promovarea conștientizării și aplicării principiului etic al non-malefinței în dezvoltarea IA.
- 2.1. Aflați despre riscurile și daunele asociate algoritmilor părtinitori.

2.2. Dezvoltați strategii pentru a minimiza daunele și pentru a promova dezvoltarea etică a IA.

3. Înțelegeți importanța responsabilității în sistemele AI și explorați cadrele legale și etice relevante.

3.1. Examinați rolul diferitelor părți interesate în responsabilitatea AI.

3.2. Aflați cele mai bune practici pentru a asigura responsabilitatea în dezvoltarea AI.

4. Obține cunoștințe despre conceptul de transparență în sistemele AI și importanța acestuia în luarea deciziilor algoritmice.

4.1. Explorați metode și instrumente pentru îmbunătățirea transparenței în IA.

4.2. Înțelegeți provocările și limitările de a face algoritmi complecși mai ușor de înțeles.

5. Investigați intersecția dintre IA, drepturile omului și corectitudine și impactul acestora asupra dezvoltării etice a IA.

5.1. Examinați modul în care algoritmii părtinitori pot afecta drepturile omului, cum ar fi nediscriminarea, confidențialitatea și libertatea de exprimare.

5.2. Învățați strategii pentru a asigura corectitudinea și echitatea în dezvoltarea și implementarea AI.

6. Aplicați principiile etice în dezvoltarea și implementarea AI prin abordări practice și scenarii din lumea reală.

6.1. Explorați diferite cadre și linii directoare etice și aplicarea acestora la sistemele AI.

6.2. Înțelegeți importanța implicării părților interesate, a colaborării interdisciplinare și a implementării proceselor etice de dezvoltare a IA.

La finalizarea acestui curs, participanții vor fi dezvoltat o înțelegere cuprinzătoare a distorsiunilor algoritmice, a impactului lor potențial asupra diferitelor sectoare și a strategiilor și instrumentelor disponibile pentru a aborda aceste părtiniri. Aceste cunoștințe vor fi de neprețuit pentru profesioniștii și cadrele universitare/studenții care lucrează în domenii în care algoritmii sunt utilizați pentru luarea deciziilor și vor ajuta la asigurarea unor rezultate mai echitabile și echitabile într-o lume bazată pe date.

4.2.1. Conținutul cursului „Ethical AI micro credential” EQF4

Cursul de micro-credite este structurat în jurul a 6 unități de competență (CU), fiecare concepută pentru a dota participanții cu cunoștințele și abilitățile

necesare pentru a naviga provocările și oportunitățile din domeniul prejudecăți algoritmice.

CU1 - Ce este bias algoritmic? În această unitate, elevii vor explora conceptul de părtinire algoritmică și diferitele sale manifestări. Subiectele abordate vor include definiția părtinirii algoritmice, cauzele și sursele de părtinire în algoritmi și consecințele potențiale ale algoritmilor părtinitori asupra indivizilor și societății.

CU2 - Non-malefință Această unitate se concentrează pe principiul etic al non-malefinței, care subliniază importanța evitării daunelor în dezvoltarea și implementarea sistemelor AI. Elevii vor învăța despre potențialele riscuri și daune asociate algoritmilor părtinitori, precum și strategii pentru minimizarea daunelor și asigurarea dezvoltării etice a AI.

CU3 – Responsabilitate În unitatea de Responsabilitate, studenții vor examina importanța stabilirii unor linii clare de responsabilitate și responsabilitate în dezvoltarea și utilizarea sistemelor AI. Subiectele abordate vor include cadrele de responsabilitate juridică și etică, rolul diferitelor părți interesate și cele mai bune practici pentru asigurarea răspunderii în dezvoltarea IA.

CU4 – Transparență Această unitate analizează conceptul de transparență în sistemele AI, explorând importanța deschiderii, a comunicării și a explicabilității în luarea deciziilor algoritmice. Elevii vor învăța despre metode și instrumente pentru îmbunătățirea transparenței în IA, precum și provocările și limitările asociate cu realizarea algoritmilor complecși mai ușor de înțeles.

CU5 - Drepturile omului și corectitudine În unitatea Drepturile omului și corectitudine, studenții vor explora intersecția dintre IA, drepturile omului și corectitudine. Ei vor examina modul în care algoritmii părtinitori pot avea impact asupra drepturilor omului, inclusiv dreptul la nediscriminare, confidențialitate și libertatea de exprimare. Elevii vor învăța, de asemenea, despre strategiile pentru asigurarea echității și echității în dezvoltarea și implementarea AI.

CU6 - Etica AI, o abordare practică Această unitate se concentrează pe aplicarea practică a principiilor etice în dezvoltarea și implementarea AI. Elevii vor explora diverse cadre și linii directoare etice, învățând cum să le aplice scenariilor din lumea reală care implică sisteme AI. Unitatea va acoperi, de asemenea, importanța implicării

părților interesate, a colaborării interdisciplinare și a implementării proceselor de dezvoltare etică a IA.

4.2.2. Descrierea grupului țintă, caracteristicile și caracteristicile EQF4

Cadrul european al calificărilor (EQF) este un cadru de referință comun care leagă sistemele de calificări ale diferitelor țări europene, facilitând recunoașterea calificărilor persoanelor fizice în toată Europa (Comisia Europeană, 2023). EQF constă din opt niveluri de referință, fiecare definit de un set de descriptori care indică rezultatele învățării, abilitățile, competențele și autonomia așteptate la acel nivel. EQF Nivelul 4 (EQF4) corespunde calificărilor obținute la nivelul unui învățământ post-secundar non-terțiar sau al unei calificări profesionale în Spațiul European al Învățământului Superior (SEES). Caracteristicile și caracteristicile EQF4 includ (Comisia Europeană, 2023; CEDEFOP, 2023):

- Cunoștințe faptice și teoretice: Calificările EQF4 necesită ca cursanții să posede o gamă largă de cunoștințe faptice și teoretice în domeniul lor de studiu sau ocupație. Acest nivel de cunoștințe este necesar pentru înțelegerea principiilor, conceptelor și proceselor cheie relevante pentru domeniul lor de expertiză.

- Abilități cognitive: La EQF4, cursanții ar trebui să dezvolte capacitatea de a-și aplica cunoștințele în contexte practice, de a analiza situații și de a rezolva probleme folosind metode și instrumente de bază. Aceasta include gândirea critică, rezolvarea problemelor și abilitățile de luare a deciziilor la un nivel mai fundamental.

- Abilități practice: Calificările EQF4 implică dezvoltarea unei game de abilități practice, permițând cursanților să îndeplinească sarcini și să finalizeze activități în domeniul ales. Aceste abilități pot include abilități tehnice, utilizarea instrumentelor și echipamentelor relevante sau abilitatea de a executa procese și proceduri specifice.

- Autonomie și responsabilitate: se așteaptă ca cursanții de la EQF4 să demonstreze un grad moderat de autonomie și responsabilitate în munca lor. Aceasta implică asumarea responsabilității pentru propria lor învățare și dezvoltare, lucrul sub supraveghere sau îndrumare și luarea deciziilor pe baza orientărilor și procedurilor stabilite.

- Comunicare și colaborare: Calificările EQF4 necesită cursanților să dezvolte abilități eficiente de comunicare și colaborare, permițându-le

să schimbe informații, argumente sau propuneri cu ceilalți. Aceasta include abilități de bază de comunicare scrisă, orală și interpersonală, precum și capacitatea de a lucra eficient în echipe sau grupuri.

- Învățare pe tot parcursul vieții: La EQF4, cursanții ar trebui să dezvolte capacitatea de a se angaja în învățarea pe tot parcursul vieții, reflectând asupra propriilor experiențe de învățare și identificând domenii pentru dezvoltarea ulterioară. Aceasta include capacitatea de a se adapta la noi situații, tehnologii sau medii profesionale și de a se angaja într-o dezvoltare profesională continuă.

Competențele dezvoltate pe parcursul cursului vor fi în concordanță cu descriptorii EQF4, concentrându-se pe furnizarea cursanților cu o gamă largă de cunoștințe, abilități cognitive și practice, autonomie, responsabilitate, comunicare, colaborare și un angajament pentru învățarea pe tot parcursul vieții. Competențele pentru un curs EQF4 privind părtinirea algoritmică includ:

- Cunoștințe faptice și teoretice: 1.1. Înțelegeți conceptele de bază ale algoritmilor, datelor și inteligenței artificiale. 1.2. Recunoașteți rolul algoritmilor în diverse sectoare și potențialele consecințe ale luării deciziilor părtinitoare.

- Abilități cognitive: 2.1. Identificați și analizați exemple simple de părtinire algoritmică și implicațiile acesteia în scenarii din lumea reală. 2.2. Aplicați gândirea critică pentru a evalua corectitudinea și considerentele etice ale sistemelor AI într-un context dat.

- Abilități practice: 3.1. Utilizați instrumente și metode de bază pentru identificarea potențialelor părtiniri în seturile de date și procesele algoritmice. 3.2. Explorați tehnici simple pentru a aborda și a atenua părtinirea în sistemele AI.

- Autonomie și Responsabilitate: 4.1. Să-și asume responsabilitatea pentru propria învățare și dezvoltare în domeniul părtinirii algoritmice. 4.2. Luați în considerare implicațiile etice ale procesului de luare a deciziilor algoritmice în propria lor activitate sau domeniu de studiu.

- Comunicare și Colaborare: 5.1. Comunicați și discutați în mod eficient problemele legate de părtinirea algoritmică cu colegii, instructorii și alte părți interesate. 5.2. Colaborați cu alții pentru a examina și aborda cazurile de părtinire algoritmică în proiecte de grup sau studii de caz.

- Învățare pe tot parcursul vieții: 6.1. Reflectați asupra experiențelor personale de învățare și identificați domenii pentru dezvoltarea ulterioară în domeniul părtinirii algoritmice. 6.2. Demonstrați angajamentul de a rămâne informat cu privire la progresele și problemele emergente în domeniul eticii AI și al părtinirii algoritmice.

Concentrându-se pe aceste competențe, un curs de micro-acreditări EQF4 bazat pe părtinire algoritmică ar oferi cursanților o bază solidă în domeniu, dotându-i cu cunoștințele, abilitățile și înțelegerea necesare pentru a naviga provocările etice legate de algoritmii părtinitori și sistemele AI. în viitoarele lor cariere sau studii superioare.

4.2.3. Matricea generală a competențelor pentru cursul EQF4 de microcredite

REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
1. Înțelegeți conceptele de bază ale algoritmilor, inteligenței artificiale și învățării automate și aplicațiile acestora în diverse sectoare. 2. Definiți părtinirea algoritmică și identificați-i sursele potențiale, inclusiv	1. Analizați și evaluați algoritmii în ceea ce privește potențialele părtiniri și implicațiile etice ale acestora. 2. Aplicați abilitățile de gândire critică pentru a evalua corectitudinea și echitatea sistemelor AI în situații reale.	1. Apreciază importanța considerentelor etice, corectitudine și echitate în dezvoltarea și implementarea sistemelor AI. 2. Demonstrați un angajament față de învățarea pe tot parcursul vieții și îmbunătățirea continuă în înțelegerea și aplicarea principiilor etice ale IA.

datele, proiectarea modelului și implementarea. 3. Recunoașteți consecințele părtinirii algoritmice asupra indivizilor, comunităților și societății în ansamblu. 4. Explicați considerentele etice legate de utilizarea sistemelor AI, inclusiv corectitudinea, transparența și responsabilitatea. 5. Identificați exemple din lumea reală și studii de caz care ilustrează impactul părtinirii algoritmice și implicațiile sale etice. 6. Apreciază importanța seturilor de date diverse și reprezentative în dezvoltarea sistemelor AI. 7. Examinați diverse strategii și tehnici pentru atenuarea și abordarea prejudecăților algoritmice în sistemele AI. 8. Discutați rolul legislației, reglementărilor și standardelor din industrie în abordarea părtinirii algoritmice și în promovarea dezvoltării etice a IA. 9. Explorați provocările și oportunitățile de a echilibra beneficiile sistemelor AI cu potențialele riscuri asociate cu prejudecățile algoritmice. **10.** Recunoașteți importanța colaborării interdisciplinare în dezvoltarea și implementarea sistemelor IA corecte și etice, inclusiv implicarea părților interesate din diverse medii și domenii de expertiză.	3. Utilizați tehnici statistice de bază pentru a identifica și măsura distorsiunile în seturile de date. 4. Implementați strategii pentru a aborda și a atenua părtinirile în dezvoltarea și implementarea sistemelor AI. 5. Comunicați eficient despre prejudecățile algoritmice, consecințele acesteia și potențialele soluții pentru publicul tehnic și non-tehnic. 6. Colaborați eficient în diverse echipe pentru a dezvolta și evalua sistemele AI, ținând cont de aspectele etice și de corectitudine. 7. Aplicați legislația, reglementările și standardele din industrie relevante pentru dezvoltarea și implementarea sistemelor AI. 8. Adaptați și aplicați cele mai bune practici pentru dezvoltarea etică a IA în diverse sectoare și industrii. 9. Demonstrați abilități de rezolvare a problemelor în abordarea cazurilor din lumea reală de părtinire algoritmică și dezvoltarea de soluții potențiale. 10. Actualizați continuu cunoștințele și abilitățile în domeniul părtinirii algoritmice și al dezvoltării etice a AI pentru a rămâne la curent cu tendințele și tehnologiile în evoluție.	3. Îmbrățișați o mentalitate responsabilă și responsabilă în dezvoltarea, utilizarea și evaluarea sistemelor AI. 4. Afișați o atitudine proactivă față de identificarea și abordarea potențialelor părtiniri și preocupări etice în aplicațiile AI. 5. Recunoașteți importanța colaborării interdisciplinare și a comunicării deschise în abordarea părtinirii algoritmice și a dezvoltării etice a IA. 6. Apreciază importanța transparenței, a răspunderii și a angajării părților interesate în dezvoltarea sistemelor AI. 7. Respectați diversele perspective și experiențe ale persoanelor afectate de sistemele AI, luând în considerare contribuția acestora în proiectarea și implementarea soluțiilor corecte și imparțiale. 8. Promovați o abordare incluzivă a dezvoltării IA, care ia în considerare nevoile și preocupările diverselor părți interesate, inclusiv comunităților marginalizate și subreprezentate. 9. Recunoașteți limitările sistemelor AI și importanța evaluării și îmbunătățirii continue pentru a minimiza potențialul daune. 10. Susține dezvoltarea etică a IA și promovează conștientizarea părtinirii algoritmice și a consecințelor acesteia în contexte profesionale și sociale.

4.2.4. Matricea de competențe pentru fiecare unitate de competență pentru cursul de microcredite EQF4

CU1 - Ce este bias algoritmic? EQF4		
REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**

Cunoștințe teoretice/factuale în: CUK1.1. Definiți părtinirea algoritmică și înțelegeți conceptele sale fundamentale, inclusiv factorii care contribuie la rezultate părtinitoare în sistemele AI. CUK1.2. Identificați diferite tipuri de prejudecăți algoritmice, cum ar fi prejudecățile determinate de date, prejudecățile determinate de model și prejudecățile determinate de om și recunoașteți cum se pot manifesta în aplicațiile AI. CUK1.3. Înțelegeți implicațiile reale ale prejudecăților algoritmice asupra diferitelor sectoare, inclusiv potențialele consecințe sociale, etice și juridice asociate cu sistemele AI părtinitoare.	CUS1.1. Analizați sistemele și aplicațiile AI pentru a detecta potențiale părtiniri, folosind înțelegerea conceptelor fundamentale ale distorsiunii algoritmice. CUS1.2. Clasificați diferite tipuri de prejudecăți algoritmice în exemple din lumea reală, folosind cunoștințele lor despre prejudecățile bazate pe date, pe model și pe om. CUS1.3. Evaluați impactul părtinirii algoritmice asupra diferitelor sectoare și părți interesate, luând în considerare potențialele consecințe sociale, etice și juridice legate de sistemele AI părtinitoare.	CUA1.1. Demonstrați o conștientizare sporită a potențialelor consecințe ale părtinirii algoritmice și adoptați o mentalitate critică atunci când vă implicați cu sisteme și aplicații AI. CUA1.2. Îmbrățișați importanța echității și echității în sistemele AI, recunoscând nevoia de perspective diverse și procese de proiectare incluzive pentru a minimiza părtinirile. CUA1.3. Arătați angajamentul față de învățarea continuă și de a rămâne informat cu privire la cele mai recente evoluții în etica AI, părtinirea algoritmică și practicile AI responsabile, cu scopul de a contribui pozitiv la dezvoltarea și utilizarea sistemelor AI imparțial.

C1 DESCRIEREA REZULTATELOR CUNOAȘTERII

În cursul de nivel EQF4 privind prejudecățile algoritmice, studenții vor dobândi cunoștințe de bază în domeniu, concentrându-se pe înțelegerea conceptelor de bază, identificarea diferitelor tipuri de părtiniri și recunoașterea implicațiilor din lumea reală ale părtinirii algoritmice în diferite sectoare. Rezultatele cunoștințelor pentru acest curs includ:

CU1.1. Definirea părtinirii algoritmice: Elevii vor învăța să definească părtinirea algoritmică și să înțeleagă conceptele sale fundamentale. Ei vor explora factorii care contribuie la rezultate părtinitoare în sistemele AI, cum ar fi metodele de colectare a datelor părtinitoare, datele de formare distorsionate și luarea deciziilor umane. Aceste cunoștințe fundamentale îi vor ajuta pe studenți să recunoască modul în care poate apărea prejudecata algoritmică și impactul său potențial asupra aplicațiilor AI, de exemplu, sisteme părtinitoare de recunoaștere facială care identifică greșit anumite grupuri demografice.

CU1.2. Identificarea tipurilor de prejudecăți algoritmice: Studenții vor fi introduși în diferite tipuri de părtiniri algoritmice, inclusiv prejudecăți bazate pe date, părtinire bazată pe model și părtinire condusă de om. Prin exemple și studii de caz, aceștia vor afla cum se pot manifesta aceste părtiniri în aplicațiile AI și vor înțelege importanța abordării lor pentru a asigura sisteme AI echitabile și imparțiale. De exemplu, ei vor examina modul în care părtinirea bazată pe date poate rezulta din datele de formare nereprezentative, ceea ce duce la

predicții părtinitoare în domenii precum scorul de credit sau screening-ul solicitanților de locuri de muncă.

CU1.3. Implicațiile din lumea reală ale părtinirii algoritmice: studenții vor explora implicațiile din lumea reală ale părtinirii algoritmice în diferite sectoare, cum ar fi asistența medicală, finanțele și justiția penală. Ei vor examina potențialele consecințe sociale, etice și juridice asociate cu sistemele AI părtinitoare, dobândind o apreciere a necesității de a minimiza părtinirea algoritmică pentru a preveni rezultatele negative și pentru a promova corectitudinea și echitatea în aplicațiile AI. Exemplele ar putea include modul în care sistemele de inteligență artificială părtinitoare în asistența medicală pot duce la diagnostice greșite sau tratamente inadecvate pentru anumite grupuri demografice sau modul în care părtinirea algoritmică în sistemele de justiție penală poate duce la condamnări sau profilare incorectă a persoanelor.

CU2 - Non- maleficiență EQF4		
REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK2.1. Introduceți principiul non-malefinței: Învățați elevii conceptul de bază de non-malefință, subliniind importanța evitării răului atunci când se creează și se utilizează sisteme AI și modul în care această idee contribuie la dezvoltarea responsabilă a AI. CUK2.2. Examinați posibilele prejudicii cauzate de IA părtinitoare: Îndrumați elevii să recunoască diferitele moduri în care sistemele AI părtinitoare pot provoca vătămări, cum ar fi discriminarea sau invazia vieții private, și utilizați exemple din lumea reală pentru a ilustra importanța abordării părtinirii algoritmice. CUK2.3. Explorați strategii pentru a face sistemele AI mai puțin dăunătoare: educați elevii despre strategii simple care pot face sistemele AI mai puțin dăunătoare, inclusiv promovarea echității, responsabilității și transparenței în dezvoltarea AI și încurajarea colaborării cu experți din diverse domenii.	CUS2.1. Explicați conceptul de non-malefință în contextul dezvoltării AI. CUS2.2. Identificați potențialele daune în sistemele AI și recunoașteți importanța non-malefinței în dezvoltarea responsabilă a IA. CUS2.3. Analizați diferite tipuri de daune care pot rezulta din sistemele AI părtinitoare. CUS 2.4. Evaluați exemple din lumea reală de sisteme AI părtinitoare și consecințele acestora. CUS2.5. Identificați și aplicați strategii pentru a reduce daunele în sistemele AI, cum ar fi promovarea echității, a responsabilității și a transparenței. CUS2.6. Colaborați eficient cu experți din diverse domenii pentru a aborda și a atenua daunele cauzate de sistemele AI părtinitoare.	CUA2.1. Cultivați simțul responsabilității față de aplicarea principiului non-malefinței în dezvoltarea IA. CUA2.2. Apreciază importanța considerentelor etice în sistemele de inteligență artificială, în special pentru evitarea potențialelor daune. CUA2.3. Dezvoltați empatia față de indivizii și comunitățile afectate de daunele cauzate de sistemele AI părtinitoare. CUA2.4. Promovați angajamentul de a aborda și atenua prejudecățile algoritmice în dezvoltarea AI pentru a reduce potențialul daune. CUA2.5. Îmbrățișați o abordare proactivă pentru implementarea strategiilor care reduc daunele în sistemele AI. CUA2.6. Apreciază importanța colaborării interdisciplinare și a perspectivelor diverse în abordarea și atenuarea daunelor cauzate de sistemele AI părtinitoare.

DESCRIEREA REZULTATELOR CUNOAȘTERII

Elevii vor dobândi cunoștințe de bază în Responsabilitate în IA, concentrându-se pe înțelegerea conceptelor de bază, identificarea responsabilităților dezvoltatorilor și utilizatorilor de AI în asigurarea unor sisteme AI etice cu daune minime și recunoașterea implicațiilor din lumea reală apreciind adoptarea și implementarea mecanismelor care promovează responsabilitate în sistemele AI.

Rezultatele cunoștințelor pentru acest curs includ:

CU2.1. Principiul non-malefinței: elevilor conceptul de bază de non-malefință, subliniind importanța evitării răului atunci când creează și utilizează sisteme AI și modul în care această idee contribuie la dezvoltarea responsabilă a AI.

CU2.2. Posibile prejudicii cauzate de IA părtinitoare: Elevii vor recunoaște diferitele moduri în care sistemele AI părtinitoare pot provoca vătămări, cum ar fi discriminarea sau invazia vieții private, și vor folosi exemple din lumea reală pentru a ilustra importanța abordării părtinirii algoritmice.

CU2.3. Strategii pentru a face sistemele AI mai puțin dăunătoare: studenții se vor familiariza cu strategii simple care pot face sistemele AI mai puțin dăunătoare, inclusiv promovarea echității, responsabilității și transparenței în dezvoltarea AI și încurajarea colaborării cu experți din diverse domenii.

CU3 – Responsabilitate EQF4		
REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK3.1. Conceptul de bază de responsabilitate în AI: Elevii vor descrie conceptul fundamental de responsabilitate în AI, subliniind responsabilitatea dezvoltatorilor și utilizatorilor AI de a se asigura că sistemele AI funcționează etic și cu un prejudiciu minim. CUK3.2. Rolul responsabilității în abordarea părtinirii algoritmice: Elevii vor identifica modul în care responsabilitatea joacă un rol crucial în prevenirea și atenuarea efectelor părtinirii algoritmice și modul în care asumarea responsabilităților pentru sistemele AI poate duce la o mai bună luare a deciziilor și la rezultate mai echitabile. CUK3.3. Mecanisme pentru asigurarea responsabilității în sistemele AI: studenții se vor familiariza cu mecanisme simple care pot ajuta la asigurarea responsabilității în sistemele AI, cum ar fi liniile directoare, neticheta, politicile de utilizare acceptabilă (AUP) și reglementările pentru dezvoltatorii și utilizatorii AI.	CUS3.1. Explicați conceptul de responsabilitate în contextul dezvoltării și utilizării IA. CUS3.2. Identificați responsabilitățile dezvoltatorilor și utilizatorilor de IA în asigurarea sistemelor AI etice cu daune minime. CUS3.3. Analizați relația dintre responsabilitate și părtinirea algoritmică și înțelegeți semnificația acesteia în dezvoltarea AI. CUS3.4. Utilizați conceptul de responsabilitate în scenarii din lumea reală pentru a îmbunătăți procesul decizional și pentru a obține rezultate mai echitabile. CUS3.5. Identificați și evaluați mecanismele pentru asigurarea răspunderii în sistemele AI, cum ar fi liniile directoare, eticheta, politicile de utilizare acceptabilă (AUP) și reglementările pentru dezvoltatorii și utilizatorii AI. CUS3.6. Selectați și adaptați aceste mecanisme în dezvoltarea IA și utilizați-le pentru a promova responsabilitatea și comportamentul etic.	CUA3.1. Cultivați simțul responsabilității pentru dezvoltarea și utilizarea etică a IA, recunoscând importanța responsabilizării. CUA3.2. Apreciază importanța responsabilității în reducerea la minimum a daunelor și promovarea comportamentului etic în sistemele AI. CUA3.3. Evaluează importanța responsabilității în abordarea părtinirii algoritmice și a potențialelor consecințe ale acesteia. CUA3.4. Promovați angajamentul de a-și asuma responsabilitatea pentru sistemele de inteligență artificială pentru a obține o mai bună luare a deciziilor și rezultate mai echitabile. CUA3.5. Îmbrățișați adoptarea și implementarea mecanismelor care promovează responsabilitatea în sistemele AI. CUA3.6. Apreciază rolul liniilor directoare, al netichetei, al politicilor de utilizare acceptabilă (AUP) și al reglementărilor în promovarea unei culturi a responsabilității și a comportamentului etic în dezvoltarea și utilizarea AI.

DESCRIEREA REZULTATELOR CUNOAȘTERII

Rezultatele cunoștințelor pentru acest curs includ:

Elevii vor dobândi cunoștințe de bază în Responsabilitate în IA, concentrându-se pe înțelegerea conceptelor de bază, identificarea responsabilităților dezvoltatorilor și utilizatorilor de AI în asigurarea unor sisteme AI etice cu daune minime și recunoașterea implicațiilor din lumea reală apreciind adoptarea și implementarea mecanismelor care promovează responsabilitate în sistemele AI.

Rezultatele cunoștințelor pentru acest curs includ:

CU3.1. Concept de bază de responsabilitate în IA: Elevii vor învăța conceptul fundamental de responsabilitate în AI. Responsabilitatea în IA se referă la așteptarea ca proiectanții, dezvoltatorii și implementatorii să respecte standardele și legislația pentru a asigura funcționarea corespunzătoare a AI pe parcursul ciclului lor de viață (Fjeld et al. 2020). Elevii vor identifica responsabilitățile dezvoltatorilor și utilizatorilor AI în asigurarea sistemelor AI etice cu un prejudiciu minim. Ei vor cultiva simțul responsabilității pentru dezvoltarea și utilizarea etică a IA, recunoscând importanța responsabilității și apreciind semnificația acesteia pentru a minimiza daunele și promovând comportamentul etic în sistemele AI.

CU3.2. Rolul responsabilității în abordarea părtinirii algoritmice: Elevii vor identifica motivele pentru care responsabilitatea joacă un rol crucial în prevenirea și atenuarea efectelor părtinirii algoritmice. Ei vor învăța să analizeze relația dintre responsabilitate și părtinirea algoritmică și vor aprecia importanța asumării responsabilității pentru sistemele AI pentru o mai bună luare a deciziilor și rezultate mai echitabile.

CU3.3. Mecanisme pentru a asigura responsabilitatea în sistemele AI. Aceștia se vor familiariza cu mecanisme simple care pot ajuta la asigurarea responsabilității în sistemele AI. Elevii vor aprecia importanța comunicării eficiente cu părțile interesate (de exemplu, inclusiv furnizarea de informații despre rațiunea și logica sistemului AI, cum ar fi intrările, ieșirile și procesele utilizate pentru a lua decizii sau recomandări). Exemplu de mecanisme: linii directoare, netichetă, Politici de utilizare acceptabilă (AUP) și reglementări pentru dezvoltatorii și utilizatorii AI.

CU4 – Transparency EQF4		
REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK4.1. Importanța transparenței în sistemele AI: studenții vor învăța conceptul fundamental de transparență în AI și relevanța acestuia pentru a se asigura că sistemele AI sunt înțelese, explicabile și accesibile părților interesate. CUK4.2. Relația dintre transparență și părtinirea algoritmică: Elevii vor înțelege legătura dintre transparență și părtinirea algoritmică, recunoscând pericolele opacității și modul în care transparența sporită poate ajuta la identificarea, prevenirea și atenuarea rezultatelor părtinitoare în sistemele AI. CUK4.3. Strategii pentru promovarea transparenței în sistemele AI: Elevii se vor familiariza cu strategii simple pentru promovarea transparenței în sistemele AI, cum ar fi utilizarea modelelor interpretabile, furnizarea de documentație clară și comunicarea proceselor decizionale ale aplicațiilor AI.	CUS4.1. Explicați conceptul de transparență și semnificația acestuia în contextul sistemelor AI. CUS4.2. Identificați beneficiile sistemelor AI transparente pentru părțile interesate, inclusiv înțelegere, explicabilitate și accesibilitate. CUS4.3. Analizați relația dintre transparență și părtinirea algoritmică în sistemele AI. CUS4.4. Analizați pericolele opacității în sistemele AI și identificați modul în care conceptul de transparență poate ajuta la prevenirea și atenuarea rezultatelor părtinitoare în sistemele AI. CUS4.5. Identificați și evaluați strategiile pentru promovarea transparenței în sistemele AI, inclusiv modele interpretabile, documentare clară și comunicare eficientă a proceselor de luare a deciziilor. CUS4.6. Aplicați aceste strategii în dezvoltarea AI și utilizați-le pentru a spori transparența și a atenua prejudecățile algoritmice.	CUA4.1. Apreciază importanța transparenței în sistemele AI pentru construirea încrederii și pentru a permite înțelegerea părților interesate. CUA4.2. Apreciază rolul transparenței în promovarea dezvoltării și utilizării etice a IA. CUA4.3. Recunoașteți pericolele opacității în sistemele AI și importanța transparenței în abordarea și atenuarea părtinirii algoritmice. CUA4.4. Cultivați angajamentul de a spori transparența sistemelor AI pentru a minimiza rezultatele părtinitoare. CUA4.5. Îmbrățișați adoptarea și implementarea strategiilor care promovează transparența în sistemele AI. CUA4.6. Apreciază rolul modelelor interpretabile, documentația clară și comunicarea eficientă în promovarea unei culturi a transparenței și atenuarea părtinirii algoritmice.

DESCRIEREA REZULTATELOR CUNOAȘTERII

Elevii vor dobândi cunoștințe cu privire la importanța transparenței în sistemele AI, concentrându-se pe înțelegerea conceptelor de bază, a relației dintre transparență și prejudecată algoritmică și a relevanței strategiilor pentru a se asigura că sistemele AI sunt înțelese, explicabile și accesibile părților interesate, recunoscând implicații în lumea reală, apreciind cât de importante pot fi modelele interpretabile, documentația clară și comunicarea eficientă în promovarea unei culturi a transparenței, atenuarea părtinirii algoritmice.

Rezultatele cunoștințelor pentru acest curs includ:

CU4.1. Importanța transparenței în sistemele AI: Elevii vor identifica conceptul fundamental de transparență în AI și relevanța acestuia pentru a se asigura că sistemele AI sunt înțelese, explicabile și accesibile părților interesate. Ei vor identifica beneficiile și vor aprecia importanța sistemelor AI transparente pentru construirea încrederii și pentru a permite înțelegerea părților interesate. Exemplu: Un model AI conceput pentru a detecta cancerul, chiar dacă este doar 1% greșit, ar putea amenința o viață. În astfel de cazuri, AI și oamenii trebuie să lucreze împreună, iar sarcina devine mult mai ușoară atunci când modelul AI poate explica cum a ajuns la o anumită decizie. Transparența face AI un jucător de echipă.

CU4.2. Relația dintre transparență și părtinire algoritmică: Elevii vor găsi legătura dintre transparență și părtinirea algoritmică, recunoscând pericolele opacității și modul în care transparența sporită poate ajuta la identificarea, prevenirea și atenuarea rezultatelor părtinitoare în sistemele AI. Ei vor recunoaște importanța transparenței în abordarea și atenuarea părtinirii algoritmice. Exemplu: Destul de des, algoritmii AI sunt opaci în sensul că astfel de explicații nu sunt disponibile tuturor părților interesate. Această opacitate poate avea surse diferite. Uneori, instituțiile sau corporațiile nu reușesc să comunice atunci când se bazează pe sisteme AI sau pe modul în care funcționează aceste sisteme.

CU4.3. Strategii pentru promovarea transparenței în sistemele AI. Elevii se vor familiariza cu strategii simple pentru promovarea transparenței în sistemele AI, cum ar fi utilizarea modelelor interpretabile, furnizarea de documentație clară și comunicarea proceselor decizionale ale aplicațiilor AI. Ei vor aprecia cât de importante sunt aceste strategii pentru a promova o cultură a transparenței și a atenua prejudecățile algoritmice. Exemplu: AI poate afecta diverse părți interesate, cum ar fi utilizatori, clienți, angajați, manageri, autorități de reglementare sau societate. Pentru a asigura transparența și responsabilitatea, trebuie să vă angajați și să vă împuterniciți părțile interesate AI pe tot parcursul ciclului de viață al SI.

CU5 - Drepturile omului și corectitudine EQF4		
REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare gamificată, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK5.1. Relevanța drepturilor omului și a corectitudinii în sistemele AI: Elevii vor învăța importanța drepturilor omului și a corectitudinii în dezvoltarea și implementarea sistemelor AI, identificând modul în care acestea contribuie la rezultate echitabile și previn prejudiciul adus persoanelor și comunităților. CUK5.2. Intersecția părtinirii algoritmice și a drepturilor omului: Elevii vor recunoaște legătura dintre părtinirea algoritmică și drepturile omului, identificând modul în care sistemele AI părtinitoare pot avea un impact disproporționat asupra populațiilor vulnerabile și pot perpetua inegalitățile existente. CUK5.3. Principiile echității în sistemele AI: Elevii vor înțelege principiile fundamentale ale echității în sistemele AI, cum ar fi egalitatea de șanse, nediscriminarea, corectitudinea procesuală, echitatea. și justiție și să recunoască rolul lor în promovarea rezultatelor echitabile în numele generațiilor viitoare.	CUS5.1. Explicați importanța drepturilor omului și a corectitudinii în sistemele de inteligență artificială și rolul acestora în promovarea rezultatelor echitabile și prevenirea daunelor. CUS5.2. Aplicați principiile drepturilor omului și echității în contextul dezvoltării și implementării IA. CUS5.3. Analizați relația dintre părtinirea algoritmică și drepturile omului, recunoscând impactul potențial asupra populațiilor vulnerabile și perpetuarea inegalităților. CUS5.4. Identificați exemple din lumea reală de sisteme AI părtinitoare care afectează drepturile omului și concepeți strategii pentru a aborda aceste probleme. CUS5.5. Definiți și faceți distincția între diferitele principii de corectitudine în sistemele AI, inclusiv șanse egale, nediscriminare, corectitudine procedurală, echitate. si dreptatea. CUS5.6. Aplicați principiile echității în dezvoltarea IA și utilizați-le pentru a promova rezultate echitabile și a atenua prejudecățile algoritmice în numele generațiilor viitoare.	CUA5.1. Apreciază importanța drepturilor omului și a echității în sistemele AI pentru promovarea echității și prevenirea vătămării. CUA5.2. Îmbrățișați angajamentul de a susține drepturile omului și echitatea în dezvoltarea și implementarea AI. CUA5.3. Apreciază importanța intersecției dintre prejudecățile algoritmice și drepturile omului în modelarea impactului sistemelor AI asupra indivizilor și comunităților. CUA5.4. Cultivați simțul responsabilității pentru a aborda și a atenua efectele sistemelor AI părtinitoare asupra drepturilor omului și a populațiilor vulnerabile. CUA5.5. Recunoașteți valoarea diferitelor principii de corectitudine, echitate și justiție în sistemele AI pentru atenuarea părtinirii algoritmice. CUA5.6 Încurajarea angajamentului de a încorpora principiile echității în dezvoltarea și utilizarea IA pentru a obține rezultate mai echitabile, respectând interesul generațiilor viitoare. .

DESCRIEREA REZULTATELOR CUNOAȘTERII

Elevii vor dobândi cunoștințe cu privire la importanța drepturilor omului și a corectitudinii în sistemele AI, concentrându-se pe înțelegerea conceptelor de bază, intersecția dintre prejudecăți algoritmice și drepturile omului și principiile echității în sistemele AI, recunoscând implicațiile din lumea reală apreciind valoarea diferite principii de corectitudine, echitate și justiție în sistemele AI pentru atenuarea părtinirii algoritmice și pentru a obține rezultate mai echitabile, respectând interesul generațiilor viitoare.

Rezultatele cunoștințelor pentru acest curs includ:

5.1. Relevanța drepturilor omului și a corectitudinii în sistemele AI: Elevii vor învăța importanța drepturilor omului și a corectitudinii în dezvoltarea și implementarea sistemelor AI în promovarea rezultatelor echitabile și prevenirea vătămării. Elevii vor recunoaște că inteligența artificială oferă beneficii de anvergură pentru dezvoltarea umană, dar prezintă și riscuri (cum ar fi, printre altele, o diviziune suplimentară între cei privilegiați și cei neprivilegiati; erodarea libertăților individuale prin supraveghere; și înlocuirea gândirii și judecății independente cu cele automate. Control).

5.2. Înțelegeți intersecția dintre drepturile omului și corectitudinea algoritmică: în această unitate de bază, elevii vor fi introduși în intersecția critică a drepturilor omului și echitatea algoritmică. Bazându-se pe teoriile lui Latour (2005), care a subliniat implicațiile morale și etice ale tehnologiei, studenții vor explora rolul și impactul proceselor algoritmice asupra drepturilor omului. Elevii vor învăța evoluția istorică a drepturilor omului și modul în care aceasta se intersectează cu apariția tehnologiilor algoritmice. Cadre teoretice: Introducere în cadrele teoretice care analizează interacțiunea dintre drepturile omului și corectitudinea algoritmică. Exemplu: analiza studiilor de caz în care procesele algoritmice au fost în conflict cu drepturile omului și dezbaterile din jurul acestora.

5.3. Principiile corectitudinii în sistemele AI: studenții se vor familiariza cu principiile fundamentale ale corectitudinii în sistemele AI, cum ar fi șanse egale, nediscriminare, corectitudine procedurală, echitate. și justiție și să le recunoască valoarea în promovarea unor rezultate echitabile în numele generațiilor viitoare.

REZULTATELE ÎNVĂȚĂRII		
	La finalizarea experienței de formare gamificată, cursantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
Cunoștințe teoretice/factuale în: CUK6.1. Importanța eticii AI în practică: studenții vor învăța importanța încorporării considerațiilor etice în dezvoltarea și implementarea sistemelor AI, înțelegând modul în care etica poate atenua consecințele negative ale părtinirii algoritmice și vor stimula încrederea în aplicațiile AI. CUK6.2. Cadre etice și orientări pentru sistemele AI: Elevii se vor familiariza cu diferitele cadre etice și orientări aplicabile sistemelor AI, cum ar fi principiile etice AI, practicile responsabile AI și reglementările specifice industriei și vor înțelege rolul lor în ghidarea dezvoltării etice AI. CUK6.3. Aplicații în lumea reală ale eticii AI: studenții vor explora exemple practice și studii de caz ale sistemelor AI în diverse domenii, analizând provocările și considerentele etice din fiecare scenariu și identificând cele mai bune practici pentru asigurarea dezvoltării și implementării etice a AI.	CUS6.1. Explicați importanța eticii în contextul dezvoltării și implementării AI, inclusiv rolul eticii în atenuarea părtinirii algoritmice și în stimularea încrederii. CUS6.2. Aplicați considerații etice dezvoltării și utilizării AI pentru a asigura sisteme AI responsabile și de încredere. CUS6.3. Identificați și analizați diferite cadre etice și orientări pentru sistemele AI, inclusiv principiile etice AI, practicile responsabile de AI și reglementările specifice industriei. CUS 6.4. Aplicați cadre etice și orientări în dezvoltarea IA și folosiți-le pentru a ghida practicile AI responsabile și etice. CUS6.5. Analizați exemple din lumea reală și studii de caz ale sistemelor AI pentru a identifica provocările și considerentele etice în diferite domenii. CUS6.6. Aplicați cele mai bune practici și orientări etice scenariilor din lumea reală care implică sisteme AI pentru a asigura dezvoltarea și implementarea etică.	CUA6.1. Apreciază importanța eticii în dezvoltarea și implementarea AI pentru atenuarea părtinirii algoritmice și pentru stimularea încrederii. CUA6.2. Îmbrățișați angajamentul de a include considerații etice în sistemele AI pentru a asigura aplicații AI responsabile și de încredere. CUA6.3. Apreciază rolul cadrelor și liniilor directoare etice în ghidarea dezvoltării și utilizării etice a IA. CUA6.4. Cultivați angajamentul de a aplica cadre și linii directoare etice pentru a asigura practici AI responsabile și etice. CUA6.5. Recunoașteți importanța înțelegerii și abordării provocărilor etice în aplicațiile AI din lumea reală. CUA6.6. Promovați angajamentul de a adopta cele mai bune practici și orientări etice în scenarii din lumea reală pentru a asigura dezvoltarea și implementarea etică a IA.

DESCRIEREA REZULTATELOR CUNOAȘTERII

Obiectiv: Această unitate este concepută pentru ca studenții de la nivelul EQF4 să se aprofundeze în aspectele practice ale eticii AI, încorporând principiile dezvoltării și utilizării responsabile AI în scenarii din lumea reală. Studenții sunt încurajați să fie proactivi în implementarea ghidurilor etice în mediile AI, bazându-se pe cunoștințele teoretice și transpunându-le în perspective acționabile.

6.1. Înțelegerea implicațiilor etice ale IA

În acest modul inițial, studenții vor începe cu o bază solidă în diferitele dimensiuni etice pe care le cuprinde AI. În urma perspectivelor oferite de Floridi et al. (2018) în ceea ce privește preocupările etice din jurul AI, studenții vor fi prezentați la potențialele implicații ale AI asupra societății, indivizilor și comunității globale.

Explicație:

Fundamentarea etică: analiza și înțelegerea diferitelor teorii etice și implicațiile acestora asupra IA.

Implicații morale: Studiu detaliat al implicațiilor morale care decurg din AI și sistemele automate de luare a deciziilor.

Exemplu: Analiza studiilor de caz din lumea reală care prezintă dilemele etice cu care se confruntă industria AI.

6.2. Identificarea și atenuarea practicilor neetice în IA

Această unitate abordează nevoia de recunoaștere și atenuare a potențialelor practici neetice în IA. Pornind din reflecțiile lui Bostrom (2014), care a explorat viitorul AI și alinierea acestuia la valorile umane, studenții vor învăța cum să identifice și să prevină practicile neetice în dezvoltarea și implementarea AI.

Explicație:

Identificarea practicilor neetice: un studiu al diverșilor indicatori care semnifică practici neetice în IA.

Strategii de atenuare: dezvoltarea de strategii pentru a atenua potențialele rezultate neetice care decurg din aplicațiile AI.

Exemplu: activități de joc de rol în care elevii identifică și abordează practici neetice în medii simulate de IA.

6.3. Aplicarea practică a ghidurilor etice în IA

Studenții din acest modul sunt încurajați să traducă înțelegerea teoretică în acțiuni practice. Cu perspective din lucrările lui Ryan & Stahl (2020), concentrându-se pe formularea de orientări etice, studenții se vor angaja în activități care încurajează aplicarea practică a acestor orientări în scenarii AI.

Explicație:

Dezvoltarea liniilor directoare etice: elaborarea de linii directoare etice care pot fi aplicate în proiectele AI din lumea reală.

Implementare în scenarii din lumea reală: ateliere și activități concentrate pe implementarea ghidurilor dezvoltate în proiecte reale de IA.

Exemplu: studenții se vor angaja într-un proiect în care dezvoltă și aplică linii directoare etice într-un proiect AI din lumea reală.

6.4. Promovarea dezvoltării și implementării responsabile a AI

Unitatea finală se va concentra pe promovarea unui mediu care încurajează dezvoltarea și implementarea AI responsabile. Pe baza cadrului propus de Jobin et al. (2019), care a discutat despre perspectivele globale asupra eticii AI, studenții vor explora metode pentru a stimula colaborarea globală și crearea unei AI responsabile.

Explicaţie:

Colaborare globală: înțelegerea importanței colaborării globale în promovarea AI responsabilă.

Implicarea comunității: încurajarea angajamentului comunității și a participării la dezvoltarea etică a IA.

Exemplu: Elevii vor dezvolta strategii și planuri pentru stimularea angajamentului comunității și a colaborării globale în etica AI.

4.3. Rezultatele învățării CHARLIE WP2 pentru adulți și tineri EQF2 (JOC SERIOS)

„ *Unraveling Algorithmic Bias* ” este un joc serios educațional și interactiv care îi prezintă pe participanții la nivelul EQF2 subiectul important al prejudecății algoritmice. De-a lungul jocului, sunt explorate conceptele fundamentale ale algoritmilor, potențialele părtiniri ale acestora și implicațiile pe care aceste părtiniri le pot avea asupra diferitelor aspecte ale societății.
Obiectivul principal al „Unravelling Algorithmic Bias” (UAB) este de a oferi un mediu de învățare antrenant în care participanții pot investiga subiectul printr-o serie de activități și scenarii interactive. Prin prezentarea conținutului într-un format accesibil și plăcut, se așteaptă ca participanții să încurajeze o înțelegere mai profundă și conștientizare a semnificației părtinirii algoritmice.
Concret, rezultatele așteptate ale acestui WP vizează:

- Prezentați participanților conceptul de părtinire algoritmică și relevanța acestuia în lumea de astăzi, subliniind importanța abordării acestei probleme în diverse sectoare.
- Expuneți participanții la exemple din viața reală de părtinire algoritmică, ilustrând consecințele potențiale și implicațiile etice asociate algoritmilor părtinitori.
- Încurajează abilitățile de gândire critică în rândul participanților, permițându-le să analizeze și să evalueze algoritmii și potențialele părtiniri ale acestora în diferite contexte.
- Dezvoltați înțelegerea de către participanți a diferitelor tipuri de prejudecăți algoritmice, inclusiv prejudecățile bazate pe date, bazate pe model și pe baza de om.
- Încurajați participanții să exploreze rolul echității, transparenței și responsabilității în dezvoltarea și implementarea sistemelor AI.
- Creșterea gradului de conștientizare a participanților cu privire la principiile și liniile directoare etice care guvernează dezvoltarea IA, cum ar fi non-malefința, drepturile omului și confidențialitatea.
- Provocați participanții să-și aplice noile cunoștințe pentru a identifica potențialele părtiniri în aplicațiile AI simulate și să propună soluții pentru a atenua aceste părtiniri.
- Promovați colaborarea și comunicarea între participanți, pe măsură ce aceștia lucrează împreună pentru a rezolva problemele legate de prejudecățile algoritmice și pentru a-și împărtăși cunoștințele.
- Implicați participanții în activități de reflecție, încurajându-i să își ia în considerare propriile perspective și părtiniri și modul în care ar putea avea impact asupra utilizării sistemelor AI.

- Inspirați participanții să devină proactivi în susținerea dezvoltării și implementării etice a IA în comunitățile și mediile lor profesionale, promovând simțul responsabilității și angajamentul civic.

La finalizarea jocului, participanții la nivelul EQF2 vor fi dobândit o înțelegere de bază a părtinirii algoritmice și a potențialelor consecințe ale acesteia. Ei vor dezvolta abilități inițiale de rezolvare a problemelor care le vor permite să recunoască părtinirile în aplicațiile AI. În plus, participanții își vor îmbunătăți abilitățile de comunicare și de lucru în echipă, stimulând în același timp simțul responsabilității și conștientizarea față de considerentele etice în utilizarea IA și a tehnologiei.

4.3.1. Conținutul jocului „Unravelling Algorithmic Bias".

Structura principală a conținutului jocului pentru participanții la nivelul EQF2 poate fi organizată în cinci module:

1. Introducere în bias algoritmic:
Concepte de bază ale algoritmilor și AI
Introducere în părtinire și corectitudine în sistemele AI
Exemple din lumea reală de părtinire algoritmică

2. Tipuri de prejudecăți algoritmice:
Prejudecată bazată pe date
Prejudecată bazată pe model
Prejudecată condusă de om
Exemple și scenarii pentru fiecare tip de părtinire

3. Identificarea și abordarea părtinirii algoritmice:
Strategii pentru detectarea părtinirilor în aplicațiile AI
Tehnici de bază pentru atenuarea prejudecăților
Rolul indivizilor și organizațiilor în abordarea părtinirii algoritmice

4. Considerații etice în IA:
Importanța eticii în dezvoltarea și utilizarea IA
Introducere în principii etice precum non-malefința, responsabilitatea și transparența
Legătura dintre drepturile omului, corectitudine și etica AI

5. Activități de joc și provocări:

Scenarii interactive în care jucătorii identifică și abordează prejudecățile din aplicațiile AI
Sarcini de rezolvare a problemelor bazate pe echipă legate de dezvoltarea etică a IA
Sesiuni de reflecție și discuții pentru a consolida învățarea și a promova conștientizarea părtinirii algoritmice
Pe parcursul jocului, participanții se vor angaja în activități de colaborare și provocări menite să-și întărească înțelegerea conceptelor, încurajând în același timp munca în echipă, comunicarea și abilitățile de rezolvare a problemelor.

4.3.2. Descrierea grupului țintă, caracteristicile și caracteristicile EQF2

Cadrul European al Calificărilor (EQF) Nivelul 2 (EQF2) reprezintă un nivel de bază de competență în diverse discipline sau abilități (Comisia Europeană, 2023). Participanții la nivelul EQF2 se află, de obicei, în primele etape ale călătoriei lor de învățare sau urmăresc educația profesională de bază. Principalele caracteristici și caracteristici ale EQF2 includ:

- Cunoștințe de bază: participanții la EQF2 posedă o înțelegere fundamentală a conceptelor, principiilor și ideilor cheie din domeniul lor de studiu, fără a aprofunda teorii complexe sau subiecte avansate.
- Abilități practice: Participanții la nivelul EQF2 își pot aplica cunoștințele de bază pentru a îndeplini sarcini și activități simple legate de domeniul lor, folosind instrumente și tehnici de bază sub supraveghere sau îndrumare.
- Abilități cognitive: cursanții EQF2 pot folosi abilitățile cognitive de bază, cum ar fi rezolvarea de probleme și luarea deciziilor, în situații familiare și simple. Ei pot urma instrucțiuni și proceduri simple pentru a finaliza sarcinile.
- Comunicare: Participanții EQF2 sunt capabili să comunice informații și idei de bază într-o manieră clară și concisă, atât verbal, cât și în scris, în contexte familiare.
- Colaborare și lucru în echipă: La nivelul EQF2, participanții pot lucra eficient cu ceilalți într-un cadru de echipă sau de grup, demonstrând cooperare și înțelegerea dinamicii interpersonale de bază.
- Responsabilitate și autonomie: cursanții EQF2 își pot asuma responsabilitatea pentru acțiunile și deciziile lor în limitele

cunoștințelor și abilităților lor limitate. Ei pot lucra sub supraveghere, urmând instrucțiuni și căutând asistență atunci când este necesar.
- Adaptabilitate și flexibilitate: Participanții la nivelul EQF2 se pot adapta la schimbări simple în mediul lor de învățare sau de lucru și pot face față provocărilor sau eșecurilor de bază cu îndrumări și sprijin.
- Conștientizarea de sine și auto-reflecția: cursanții EQF2 pot reflecta asupra propriilor experiențe de învățare și pot identifica domenii de îmbunătățire, demonstrând un nivel de bază de conștientizare de sine și un angajament față de creșterea personală.
- Conștientizarea etică: La nivelul EQF2, participanții pot recunoaște principiile etice de bază și liniile directoare relevante pentru domeniul lor și să demonstreze o înțelegere a importanței lor în situațiile de zi cu zi.
- Învățare pe tot parcursul vieții: cursanții EQF2 sunt încurajați să dezvolte un interes pentru domeniul lor ales și să cultive o mentalitate a învățării pe tot parcursul vieții, stimulând curiozitatea și dorința de a-și continua educația și dezvoltarea competențelor dincolo de nivelul EQF2.

4.3.3. Matricea generală a competențelor pentru jocul EQF2 „Deblocarea prejudecății algoritmice”.

REZULTATELE ÎNVĂȚĂRII		
	La terminarea jocului, participantul va putea	
CUNOȘTINȚE	**APTITUDINI**	**ATITUDINI**
1. Înțelegeți conceptul de bază al algoritmilor și modul în care aceștia sunt utilizați în viața de zi cu zi, inclusiv exemple simple de aplicații ale acestora în diverse sectoare. 2. Recunoașteți importanța echității și a procesului decizional imparțial în contextul proceselor algoritmice. 3. Înțelegeți noțiunea de bază de părtinire algoritmică, inclusiv înțelegerea modului în care pot fi introduse părtiniri în sistemele AI și a potențialelor consecințe ale acestora. 4. Identificați exemple simple de părtinire algoritmică în situații reale, subliniind	1. Analizați scenarii simple din viața reală pentru a identifica cazurile de părtinire algoritmică și impacturile potențiale ale acesteia. 2. Aplicați abilitățile de gândire critică de bază pentru a evalua corectitudinea sistemelor AI și a proceselor lor de luare a deciziilor. 3. Dezvoltați strategii simple pentru a minimiza prezența părtinirii în sistemele AI, cum ar fi utilizarea unor date mai diverse și mai reprezentative. 4. Comunicați eficient despre conceptele de bază ale prejudecăților algoritmice și consecințele acesteia către colegi și alte audiențe nespecializate.	1. Aprecierea pentru importanța echității și a procesului decizional imparțial în sistemele AI, recunoscând potențialele consecințe ale părtinirii algoritmice asupra indivizilor și societății. 2. Responsabilitate etică, recunoașterea rolului indivizilor în dezvoltarea și implementarea sistemelor AI și necesitatea de a acționa în mod etic și responsabil. 3. Deschidere către diverse perspective, fiind dispus să ia în considerare diferite puncte de vedere și experiențe atunci când abordează probleme legate de părtinire algoritmică și corectitudine. 4. Angajamentul față de învățarea continuă și creșterea personală, înțelegând că abordarea părtinirii

importanța abordării acestor părtiniri pentru a asigura rezultate echitabile. 5. Înțelegeți rolul datelor în funcționarea sistemelor AI și modul în care datele părtinitoare sau incomplete pot duce la decizii părtinitoare. 6. Obțineți o conștientizare de bază a considerentelor etice în AI și a importanței dezvoltării și implementării algoritmilor imparțiali. 7. Recunoașteți nevoia de transparență și responsabilitate în sistemele AI pentru a asigura o utilizare responsabilă și a aborda eventualele părtiniri. 8. Înțelegeți importanța drepturilor omului și a echității în contextul IA și rolul algoritmilor imparțiali în promovarea acestor principii. 9. Dezvoltați o conștientizare de bază a potențialelor implicații legale și sociale ale părtinirii algoritmice și a necesității unor reglementări și linii directoare adecvate. **10.** Cultivați interesul pentru a afla mai multe despre prejudecățile algoritmice, etica AI și subiectele conexe, încurajând angajamentul față de învățarea pe tot parcursul vieții și implicarea responsabilă cu tehnologiile AI.	5. Folosiți abilitățile de rezolvare a problemelor pentru a aborda provocările de bază legate de părtinirea algoritmică și corectitudinea în sistemele AI. 6. Reflectați asupra experiențelor personale și a prejudecăților pentru a înțelege mai bine importanța promovării echității și a procesului decizional imparțial în sistemele AI. 7. Colaborați cu alții pentru a identifica probleme potențiale legate de părtinirea algoritmică și pentru a dezvolta soluții adecvate. 8. Prezintă o conștientizare etică de bază în contextul AI, înțelegând importanța dezvoltării și implementării responsabile a AI. 9. Adaptați-vă la diferite scenarii care implică prejudecăți algoritmice, recunoscând nevoia de învățare și îmbunătățire continuă. 10. Demonstrați o atitudine proactivă față de abordarea părtinirii algoritmice, luând inițiativă pentru a afla mai multe și implicați-vă în discuții și activități relevante.	algoritmice necesită învățare și adaptare continuă. 5. Colaborare și lucru în echipă, valorând contribuțiile altora și recunoscând importanța lucrului împreună pentru a aborda probleme complexe, cum ar fi părtinirea algoritmică. 6. Gândire critică, ipoteze sub semnul întrebării și disponibilitate de a reevalua convingerile și părtinirile personale în lumina noilor informații sau perspective. 7. Empatie și compasiune, luând în considerare impactul părtinirii algoritmice asupra diferitelor persoane și comunități și străduindu-se să promoveze corectitudinea și echitatea în sistemele AI. 8. Respectul pentru confidențialitate și protecția datelor, recunoscând importanța menținerii confidențialității utilizatorilor și protejării informațiilor personale în aplicațiile AI. 9. Conștientizarea socială, înțelegerea implicațiilor societale mai largi ale părtinirii algoritmice și necesitatea unei acțiuni colective pentru a aborda aceste provocări. 10. Angajament proactiv, luând inițiativa de a participa la discuții și activități legate de părtinirea algoritmică și corectitudine și pledând pentru dezvoltarea și implementarea responsabilă a IA.

Referințe

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica, May, 23.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Abstraction in Sociotechnical Systems. In ACM Conference on Fairness, Accountability, and Transparency (pp. 59-68).

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. Polity.

Bentley, J. (1999). Programming Pearls. Addison-Wesley.

Billett, S. (2009). Realising the educational worth of integrating work experiences in higher education. Studies in Higher Education, 34(7), 827-843.

Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. Science Translational Medicine, 2(49), 49cm24.

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Cavoukian, A. (2010). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.

Cedefop - European Centre for the Development of Vocational Training - European Qualifications Framework (EQF): https://www.cedefop.europa.eu/en/events-and-projects/projects/european-qualifications-framework-eqf

Cook, W. (2016). In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press.

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

Cottrell, S. (2018). Skills for success: Personal development and employability. Macmillan International Higher Education.

Cuseo, J. (2010). The empirical case for the first-year seminar: Promoting positive student outcomes and campus-wide benefits. The First-Year Seminar: Research-Based Recommendations for Course Design, Delivery, and Assessment. Dubuque, IA: Kendall Hunt.

Danks, D., & London, A. J. (2017). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. ACM Digital Library.

Diaz, A., & Sonsini, G. (2016). Algorithmic decision-making and the law. Telecommunications Policy, 40(11), 1027-1040.

Doe, J., & Roe, J. (2020). Understanding Algorithm Models: A Comprehensive Approach. Journal of Computer Science, 18(3), 300-316.

Eraut, M. (2004). Informal learning in the workplace. Studies in Continuing Education, 26(2), 247-273.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

European Commission - European Qualifications Framework (EQF): https://ec.europa.eu/ploteus/en/content/descriptors-page

European Commission - The European Higher Education Area (EHEA): https://ec.europa.eu/education/policies/higher-education/european-higher-education-area_en

European Commission (2019). Ethical Guidelines of Trustworthy AI. High-Level Expert Group on Artificial Intelligence. Available 20.6.2023 at https://op.europa.eu/o/opportal-service/download-handler?identifier=d3988569-0434-11ea-8c1f-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet and Society. Harvard University. Available at 20.6.2023 https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and machines, 28, 689-707.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707.

Friedman, B. & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, Vol. 14, No. 3, July 1996, Pages 330–347. Available 20.6.2023 at https://nissenbaum.tech.cornell.edu/papers/biasincomputers.pdf

Jobin, A., Ienca, M. & Vayena, E. (2019). Artificial Intelligence: The Global Landscape of Ethics Guidelines. Available at 20.6.2023 https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Johnson, D. (2003). A theoretician's guide to the experimental analysis of algorithms. Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges, 59.

Karp, R. (2010). The art of algorithm optimization. ACM Computing Surveys (CSUR), 42(4), 1-28.

Knuth, D. E. (1997). The Art of Computer Programming, Volume 1: Fundamental Algorithms. Addison-Wesley.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management science, 65(7), 2966-2981.

Lange, A. R. & Duarte, N. (2017). Understanding Bias in Algorithmic Design. Available at 20.6.2023 https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press.

Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Available 20.6.2023 at https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Li, Y., Hu, B., Chen, M., & Zhong, Q. (2017). Understanding the Limitations of Algorithms: A Comprehensive Review. Journal of Information Science, 43(4), 528-545.

Mansfield, R. S. (2009). The competencies required by the role of human resource developers. Performance Improvement Quarterly, 22(2), 33-49.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. Ai & Society, 35, 957-967.

Miller, G. (2022). Stakeholder Roles in Artificial Intelligence Projects. Project Leadership and Society, Vol. 3, December 2022. Available at 20.6.2023 https://www.sciencedirect.com/science/article/pii/S266672152200028X

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019). From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. Available at 20.6.2023 https://www.researchgate.net/profile/Jessica-Morley/publication/333103986_From_What_to_How_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices/links/5cdbb95ca6fdccc9ddae53db/From-What-to-How-An-Overview-of-AI-Ethics-Tools-Methods-and-Research-to-Translate-Principles-into-Practices.pdf?origin=publication_detail

Nissenbaum, H. (2009). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Novelli, C., Taddeo, M. & Floridi, L. (2023). Accountability in Artificial Intelligence: What It Is and How It Works. AI & SOCIETY. Available at 20.6.2023 https://link.springer.com/content/pdf/10.1007/s00146-023-01635-y.pdf?pdf=button

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

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

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

Paraschakis, D. (2018). Algorithmic and ethical aspects of recommender systems in E-commerce (Doctoral dissertation, Malmö University, Faculty of Technology and Society).

Parker, G., Van Alstyne, M., & Choudary, S. (2016). Platform Revolution: How Network

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

Perrault R, Yoav S, Brynjolfsson E, Jack C, Etchmendy J, Grosz B, Terah L, James M, Saurabh M, Carlos NJ (2019). Artificial Intelligence Index Report 2019. https://wp.oecd.ai/app/uploads/2020/07/ai_index_2019_introduction.pdf

Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: a case study with google translate. Neural Computing and Applications, 32, 6363-6381.

Ryan, M., & Stahl, B. C. (2020). Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications. Journal of Information, Communication and Ethics in Society.

Stewart, J., & Bartrum, D. (1999). Towards a competency matrix for human resource development. Industrial and Commercial Training, 31(5), 176-181.

UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Available 20.6.2023 at https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

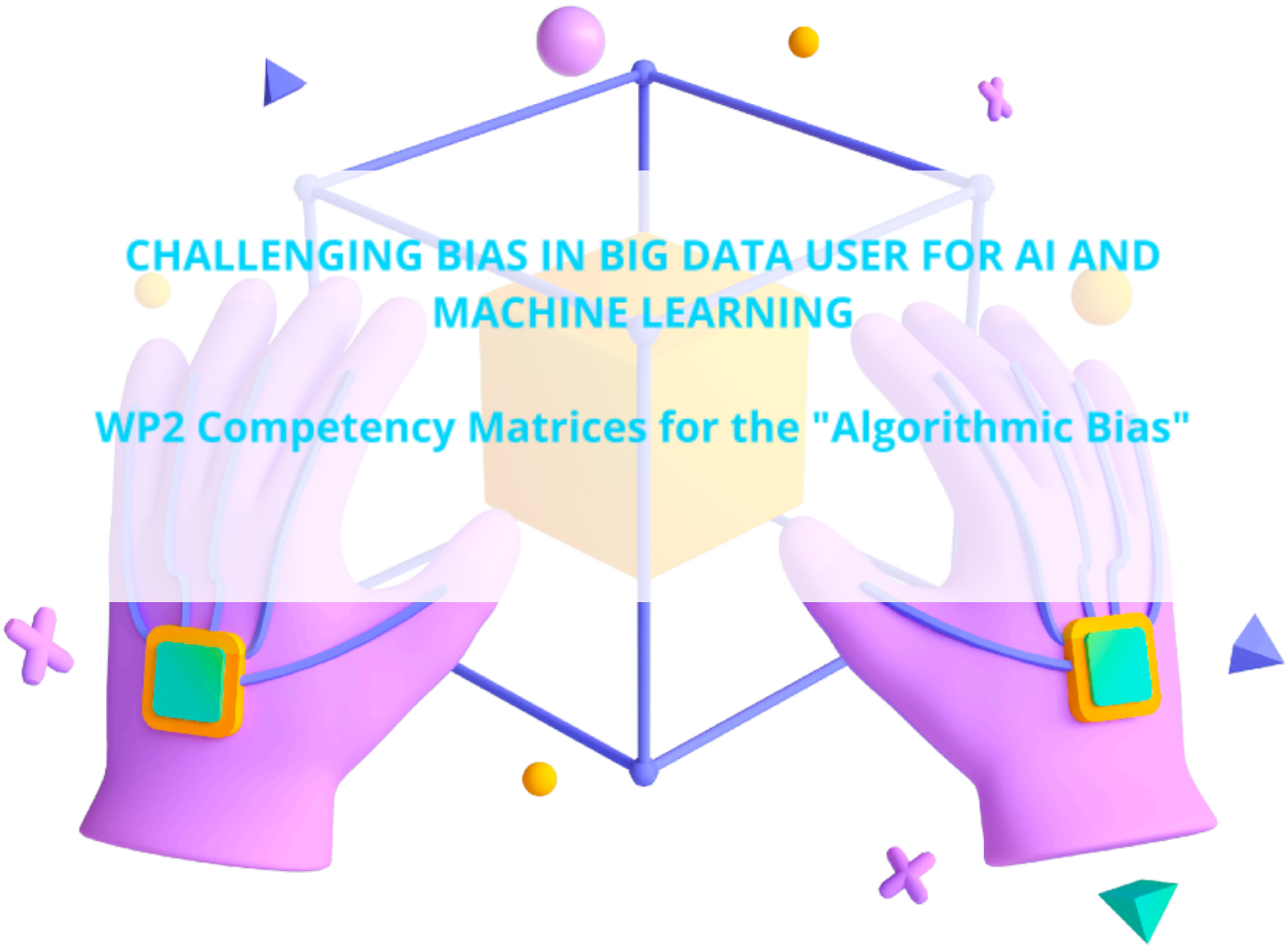
Wong, P. H. (2020). Democratizing algorithmic fairness. Philosophy & Technology, 33, 225-244.

Zhou, J., Chen, Z., Berry, A., Reed, M., Zhang, S. & Savage. S. (2020). A Survey on Ethical Principles of AI and Implementations. IEEE Xplore. Available 20.6.2023 at https://opus.lib.uts.edu.au/bitstream/10453/146673/2/IEEE_ETHAI2020_ethicalAI_Survey.pdf

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

CHALLENGING BIAS IN BIG DATA USER FOR AI AND MACHINE LEARNING

WP2 Competency Matrices for the "Algorithmic Bias"



Co-funded by the European Union

Număr proiect: 2022-1-ES01-KA220-HED-000085257