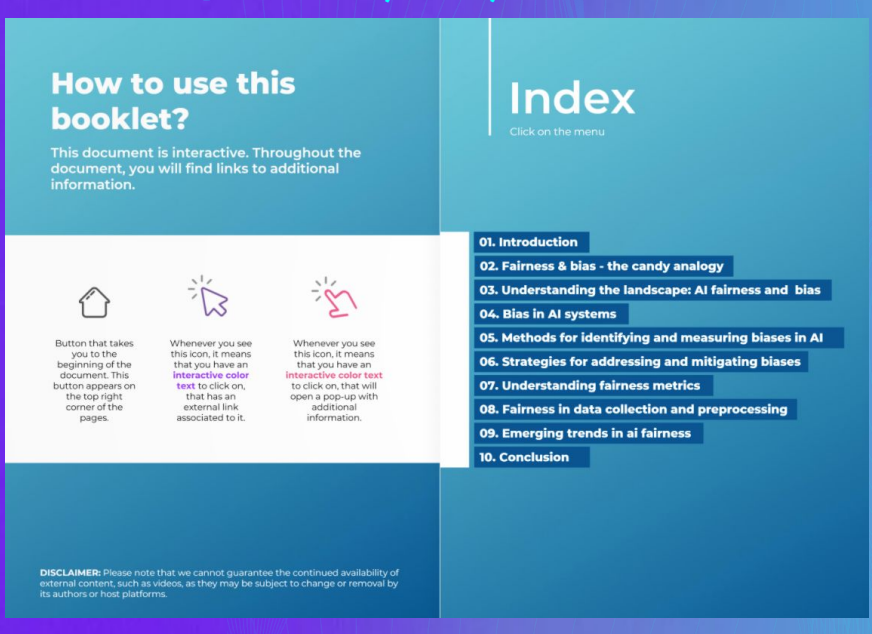
Buletin informativ 3
Octombrie 2024

Contestarea prejudecăților în utilizatorul de date mari pentru AI șiînvățare automată



Bine ați revenit la buletinul informativ al proiectului CHARLIE!

Bine ați venit la cea mai recentă ediție a buletinului informativ al proiectului CHARLIE! În această lună, punem accentul pe cursul EQF6 „Algorithmic Bias", conceput pentru instituțiile de învățământ superior care doresc să îmbunătățească înțelegerea eticii IA de către studenții lor. Având în vedere că algoritmii devin din ce în ce mai integrați în societate - în educație, sănătate, finanțe și multe altele - este esențial să abordăm prejudecățile ascunse din cadrul acestora.

Curs prezentat: Bias algoritmic

Cursul EQF6 „Algorithmic Bias" este un program educațional cuprinzător care abordează una dintre cele mai urgente provocări ale IA - prejudecățile din algoritmi. În lumea de astăzi, bazată pe date, algoritmii determină adesea cine are acces la locuri de muncă, împrumuturi, asistență medicală și chiar educație. Cu toate acestea, se pot strecura prejudecăți în aceste sisteme, fie că este vorba de un anunț de angajare care se adresează în mod disproporționat bărbaților sau de un algoritm de asistență medicală care favorizează anumite grupuri rasiale în detrimentul altora. Acest curs își propune să ofere participanților abilitățile și cunoștințele necesare pentru a identifica, atenua și gestiona astfel de prejudecăți.

De ce este important acest curs

Prejudecarea algoritmică nu este doar o problemă tehnică, ci are consecințe sociale și etice de mare amploare. În timp ce algoritmii sunt adesea considerați instrumente neutre, proiectarea, introducerea datelor și implementarea lor pot introduce discriminări neintenționate. De exemplu, anunțurile de locuri de muncă mai bine plătite sunt adesea afișate mai mult pentru bărbați decât pentru femei, iar sistemele de sănătate pot oferi o îngrijire mai bună anumitor grupuri rasiale. Aceste prejudecăți amenință nu numai corectitudinea, ci și încrederea publicului în sistemele de inteligență artificială, ceea ce face vital ca profesioniștii din domeniul tehnologiei și educatorii să înțeleagă aceste riscuri.

Cursul este structurat în jurul a cinci unități de competență (CU), fiecare fiind concepută pentru a oferi participanților instrumentele necesare pentru a naviga și a combate prejudecățile algoritmice în scenarii din lumea reală:

CU1: Algoritmi, modele și limitări
Învățați elementele de bază ale algoritmilor, limitările acestora și rolul pe care îl joacă în luarea deciziilor în toate sectoarele. Înțelegeți mecanismele care stau la baza algoritmilor și modul în care modelele acestora pot introduce în mod neintenționat prejudecăți.

CU2: Corectitudinea și părtinirea datelor în inteligența artificială
Analizați sursele de părtinire din sistemele de inteligență artificială, inclusiv datele utilizate pentru antrenarea acestor algoritmi. Participanții vor învăța tehnici de identificare, măsurare și reducere a prejudecăților din date, asigurând rezultate mai corecte.

CU3: Confidențialitatea și confortul IA
Explorați echilibrul delicat dintre păstrarea confidențialității utilizatorului și oferirea de confort în aplicațiile AI. Această unitate subliniază importanța protejării datelor utilizatorilor în timp ce se creează sisteme AI eficiente.

CU4: Etica IA - o abordare practică
Învățați să aplicați cadre etice la proiecte reale de inteligență artificială. Această unitate acoperă transparența, responsabilitatea și tehnicile de proiectare participativă pentru a asigura dezvoltarea etică a IA.

CU5: Studii de caz
În această unitate de vârf, participanții își vor aplica cunoștințele prin proiecte practice și studii de caz, dobândind experiență practică în abordarea prejudecăților algoritmice în diverse contexte.

Audiență țintă: Cursanți EQF6

Cursul „Algorithmic Bias" se adresează cursanților din învățământul superior (nivel licență) și profesioniștilor care doresc să își aprofundeze cunoștințele despre inteligența artificială etică. Cursul se aliniază la nivelul 6 al Cadrului european al calificărilor (EQF), care necesită cunoștințe avansate și abilități cognitive, precum și un grad ridicat de autonomie și responsabilitate.

Prin finalizarea acestui curs, studenții și profesioniștii vor dezvolta competențe critice, cum ar fi:

- **Cunoștințe avansate**

O înțelegere aprofundată a modului în care apar prejudecățile în sistemele AI și a impactului pe care acestea îl au în diverse sectoare. Cursanții vor explora măsurători ale corectitudinii, considerații etice și abordări tehnice pentru reducerea prejudecăților.

- **Competențe cognitive**

Capacitatea de a analiza cazuri reale de prejudecăți algoritmice, de a evalua corectitudinea sistemelor AI/ML și de a dezvolta strategii de atenuare a acestor prejudecăți în practică.

- **Competențe practice**

Implementarea tehnicilor de detectare și atenuare a prejudecăților, evaluarea eficacității acestora și integrarea principiilor etice în dezvoltarea sistemelor AI.

- **Autonomie și responsabilitate**

Să permită cursanților să ia decizii etice în cunoștință de cauză cu privire la sistemele AI/ML, luând în considerare atât impactul societal, cât și responsabilitățile individuale.

- **Comunicare și colaborare**

Dezvoltarea capacității de a explica probleme complexe legate de prejudecățile algoritmice atât părților interesate din domeniul tehnic, cât și celor din afara acestuia, încurajând colaborarea între discipline.

- **Învățarea pe tot parcursul vieții**

Angajamentul de a rămâne la curent cu cele mai recente evoluții în domeniul eticii IA și al atenuării prejudecăților, asigurând o dezvoltare profesională continuă.

Implică-te!

Sunteți entuziasmat de viitorul educației etice în domeniul IA? Consultați site-ul nostru web și urmăriți canalele noastre de social media pentru a fi la curent cu lansarea oficială a resurselor noastre și cu oportunitățile viitoare de implicare. De asemenea, vom împărtăși informații și știri din lumea eticii IA, așa că rămâneți pe recepție!
Împreună, să construim un viitor în care IA să servească binelui suprem!

https://charlie-project.uib.es/

2022-1-ES01-KA220-HED-C461966C

Parteneri

Co-funded by the European Union

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