



# Microcredențial de IA etică

CARTEA

**CU3 | Responsabilitate**

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



# Cum să utilizați acest Flipbook?

Acest document este interactiv. De-a lungul documentului, veți găsi linkuri către informații suplimentare.



Buton care vă duce la începutul documentului. Această pictogramă apare în colțul din dreapta sus al paginilor.



Ori de câte ori vedeți această săgeată, înseamnă că aveți un **text color interactiv** pe care trebuie să faceți clic, care are asociat un link extern.

**DECLINARE DE RESPONSABILITATE:** Vă rugăm să rețineți că nu putem garanta disponibilitatea continuă a conținutului extern, cum ar fi videoclipurile, deoarece acestea pot fi modificate sau eliminate de către autorii sau platformele gazdă.

# Index

Faceți clic pe meniu

**01. Introducere**

**02. Definirea responsabilității și cultivarea  
responsabilității în IA**

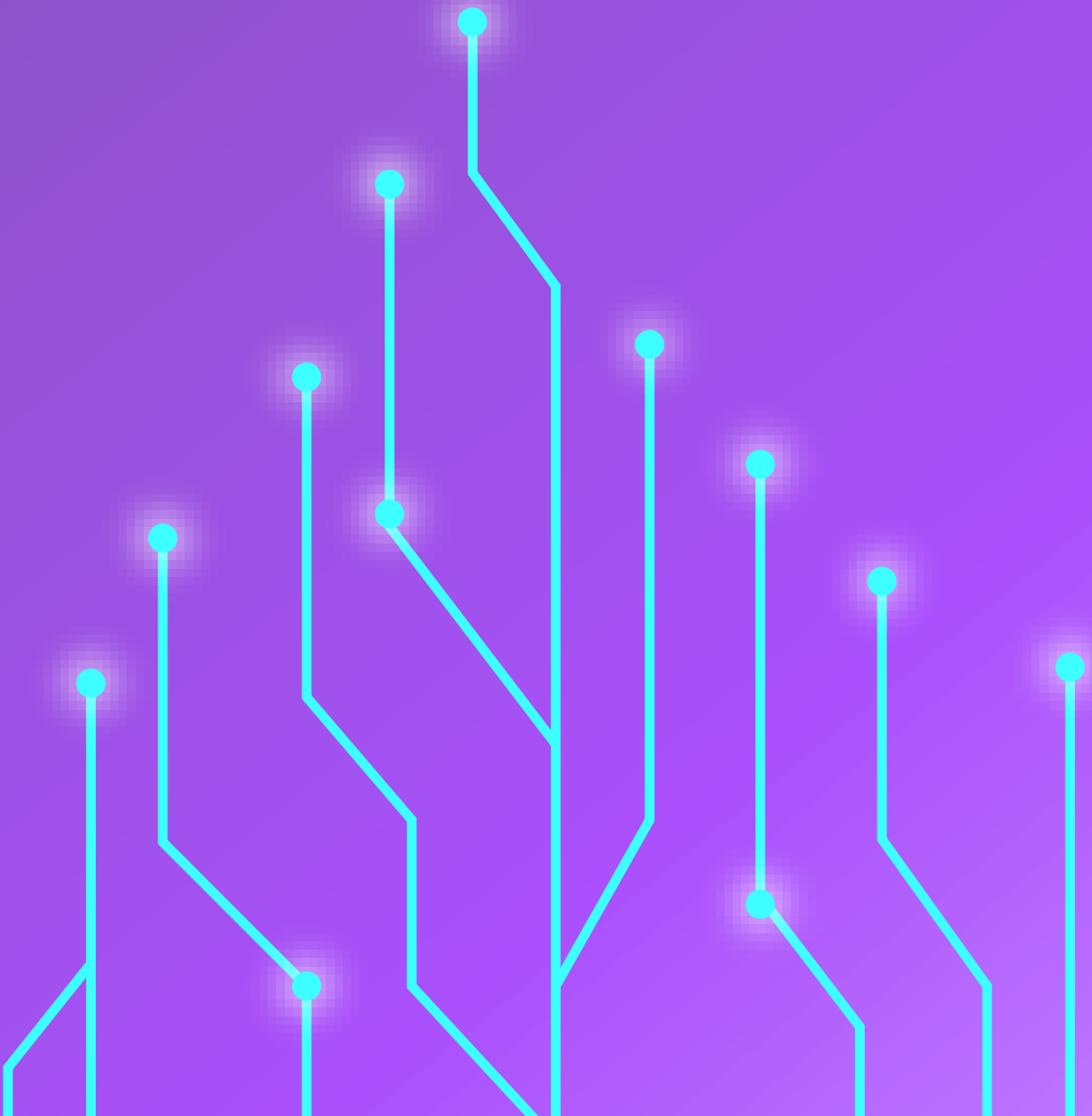
**03. Rolul responsabilității în abordarea  
Bias algoritmic**

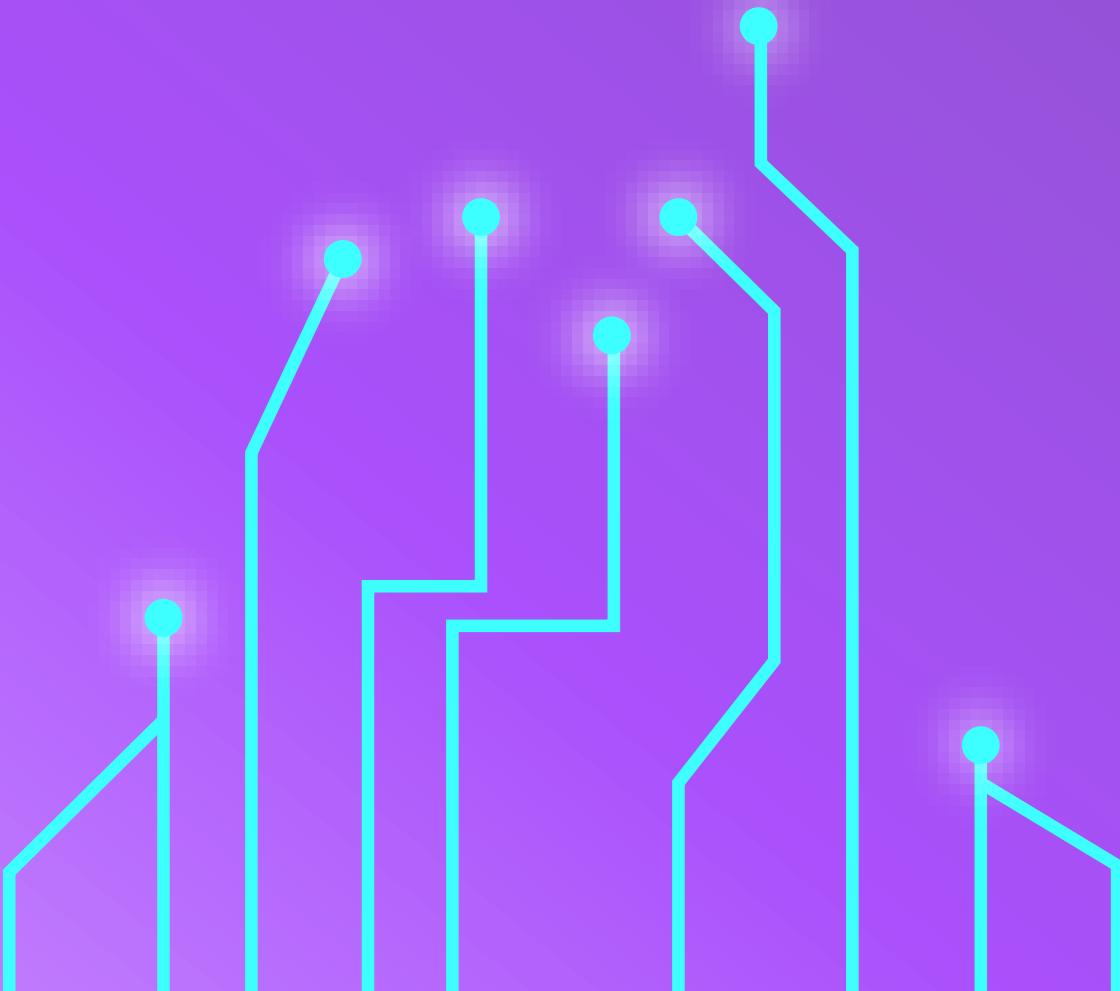
**04. Mecanisme pentru asigurarea  
responsabilității în sistemele AI**

**05. Concluzii**

# 01. Introdúcere

CU3 | Responsabilitate





## 01. Introducere

În această unitate de competență, cursanții vor dobândi cunoștințe fundamentale în domeniul responsabilității în inteligența artificială, concentrându-se pe înțelegerea conceptelor de bază, identificarea responsabilităților dezvoltatorilor și utilizatorilor de inteligență artificială în asigurarea unor sisteme etice de inteligență artificială cu prejudicii minime și recunoașterea implicațiilor în lumea reală ale adoptării și implementării mecanismelor care promovează responsabilitatea în sistemele de inteligență artificială.

Rezultatele cunoștințelor pentru această unitate cuprind:

- **Definirea responsabilității și cultivarea responsabilității în IA:** această unitate se concentrează asupra conceptului fundamental de responsabilitate în IA, care presupune așteptarea ca proiectanții, dezvoltatorii și implementatorii să adere la standarde și legislație pentru a asigura funcționarea corespunzătoare a IA de-a lungul ciclului lor de viață (Fjeld et al., 2020). Vor fi prezentate responsabilitățile dezvoltatorilor și utilizatorilor de inteligență artificială în promovarea sistemelor etice de inteligență artificială cu daune minime, subliniindu-se importanța cultivării unui simț al responsabilității pentru dezvoltarea și utilizarea etică a inteligenței artificiale și a aprecierii semnificației responsabilității în minimizarea daunelor și promovarea comportamentului etic în sistemele de inteligență artificială.

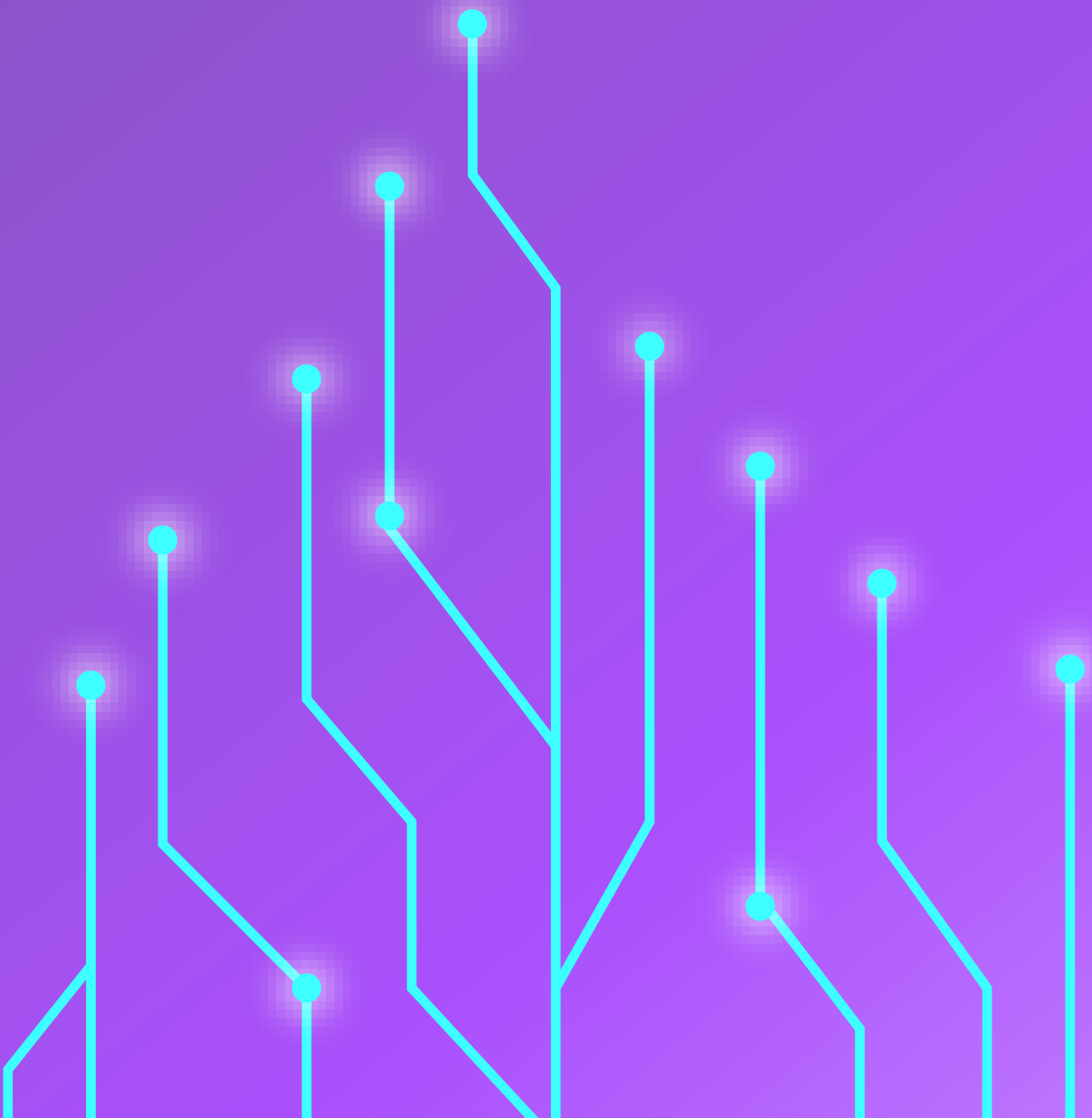


- **Rolul responsabilității în abordarea prejudecăților algoritmice**, recunoscând rolul esențial al responsabilității în prevenirea și atenuarea impactului prejudecăților algoritmice. Vom explora relația dintre responsabilitate și prejudecățile algoritmice, apreciind importanța asumării responsabilității pentru sistemele AI în vederea îmbunătățirii procesului decizional și a promovării unor rezultate mai echitabile.
- **Implementarea mecanismelor de responsabilizare în sistemele de inteligență artificială** pentru a fi familiarizați cu diverse mecanisme menite să asigure responsabilitatea în sistemele de inteligență artificială. Vom explora importanța comunicării eficiente cu părțile interesate, cum ar fi furnizarea de informații cu privire la raționamentul și logica sistemelor AI, inclusiv intrările, ieșirile și procesele decizionale. Exemple de astfel de mecanisme includ orientări, netiquette, politici de utilizare acceptabilă (AUP) și reglementări pentru dezvoltatorii și utilizatorii de IA.

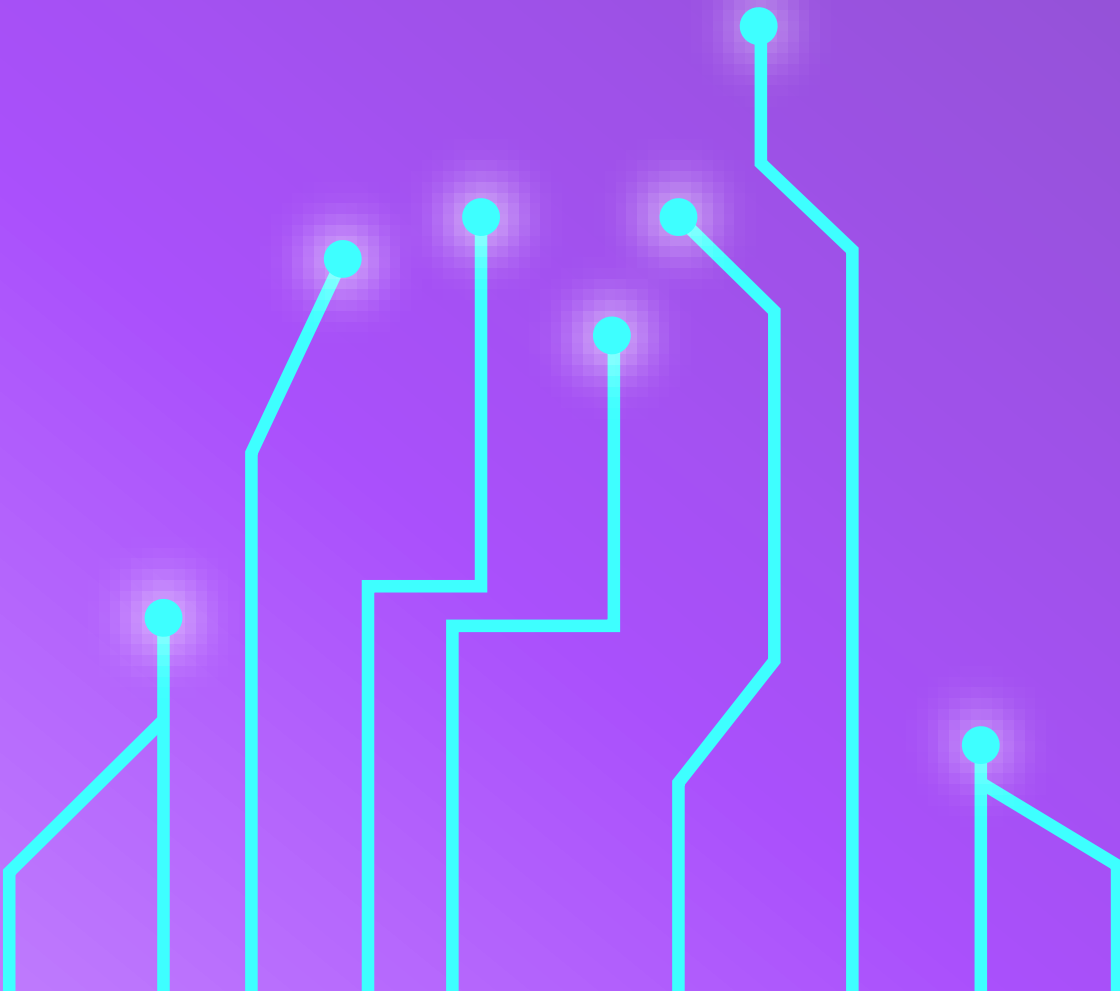


# 02. Definirea responsabilității și cultivarea responsabilității în IA

CU3 | Responsabilitate







## 02. Definirea responsabilității și cultivarea responsabilitate în IA

Responsabilitatea este fundamentală pentru dezvoltarea și implementarea etică a sistemelor de inteligență artificială (AI).

Acesta servește drept garanție împotriva practicilor imorale, asigurând că tehnologiile IA funcționează în conformitate cu standardele, reglementările și cadrele etice stabilite. Această secțiune analizează în profunzime conceptul de responsabilitate în IA, explorând principiile, mecanismele și implicațiile sale în peisajele tehnologice contemporane.

În esență, responsabilitatea în domeniul IA implică responsabilitatea proiectanților, dezvoltatorilor și implementatorilor de a respecta standardele etice pe tot parcursul ciclului de viață al sistemelor IA. Aceasta implică transparența în luarea deciziilor, trasabilitatea rezultatelor algoritmice și capacitatea de reacție la feedback-ul părților interesate. În plus, responsabilitatea presupune identificarea și atenuarea proactivă a riscurilor pentru a promova dezvoltarea și utilizarea etică a IA.

Cadrele juridice, standardele industriale și orientările etice conturează peisajul responsabilității IA. Responsabilitatea juridică se intersectează cu cerințele de reglementare și considerațiile privind răspunderea, impunând obligații de a respecta legile privind protecția datelor și reglementările privind protecția consumatorilor.



Standardele industriale și orientările etice oferă cadre pentru navigarea prin dileme etice și promovarea unor practici responsabile în domeniul IA.

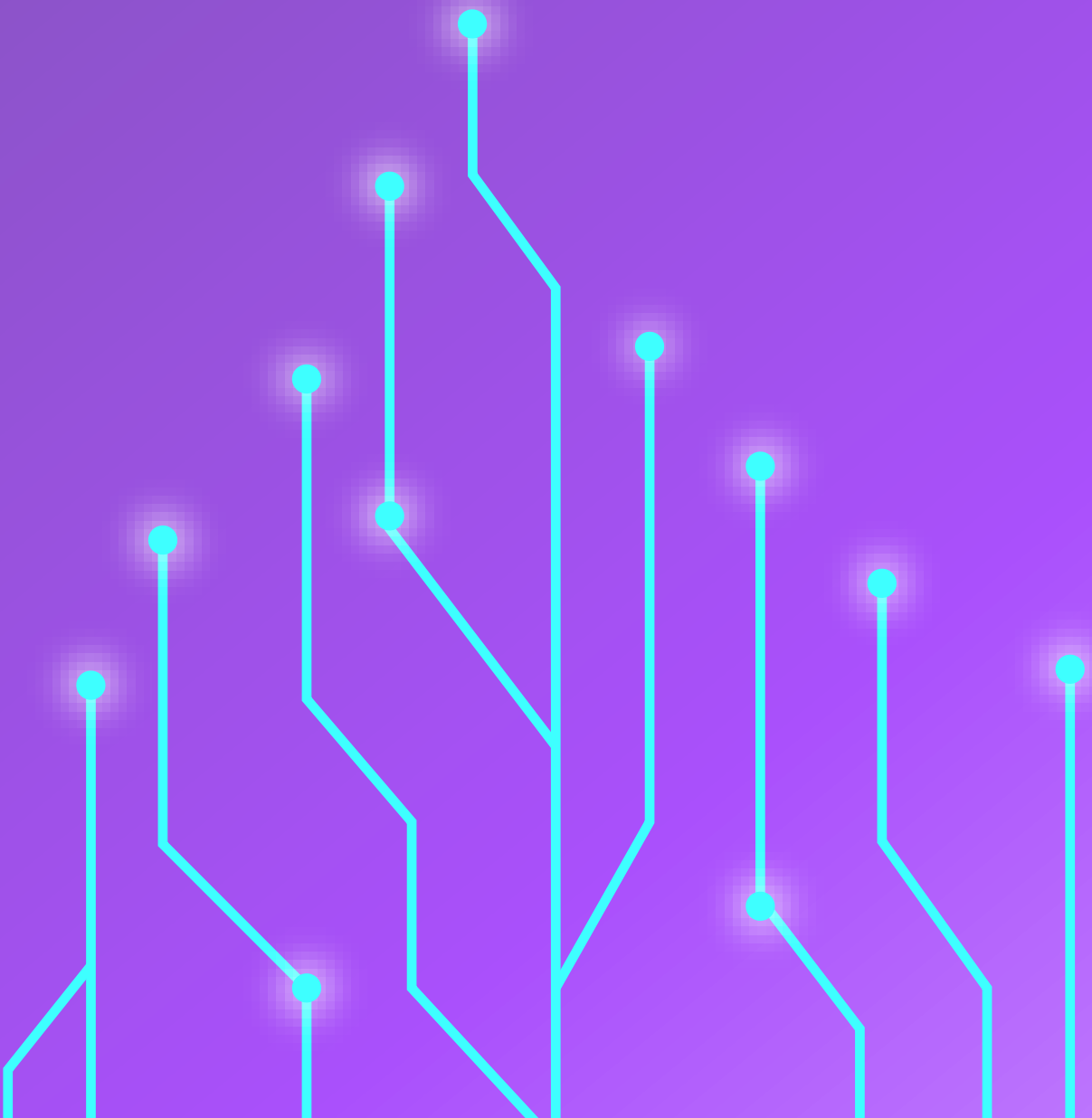
Responsabilitatea algoritmică implică evaluarea acurateței tehnice a algoritmilor AI și înțelegerea impactului socio-economic și a implicațiilor etice ale acestora. Prejudecățile din algoritmi AI pot conduce la rezultate disparate, perpetuând inegalitățile și subminând încrederea. Astfel, responsabilitatea necesită vigilență în detectarea și atenuarea părtinirilor, alături de mecanisme de transparență și explicabilitate.

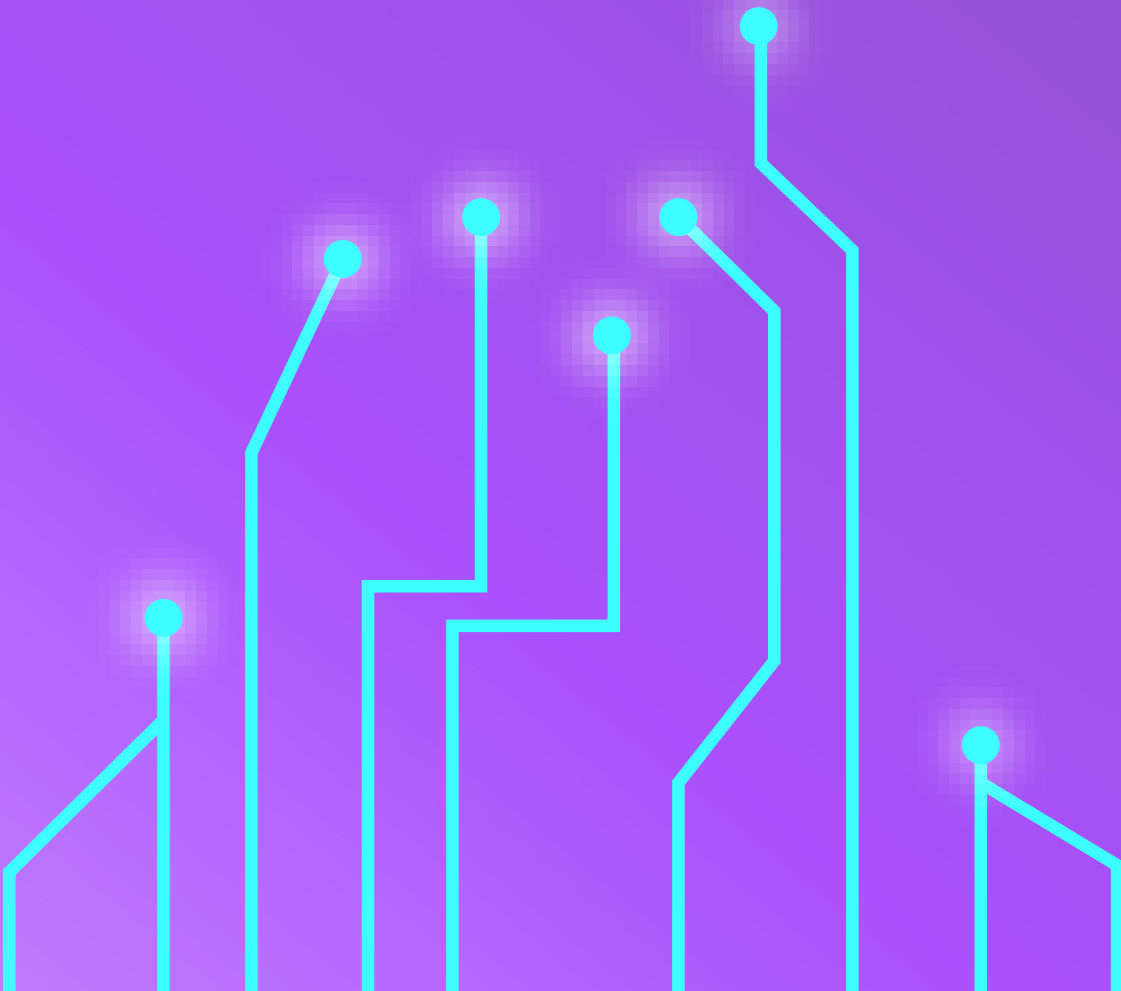
Transparența și explicabilitatea sporesc încrederea și responsabilitatea în sistemele AI, permițând părților interesate să înțeleagă procesele decizionale și rezultatele algoritmice. Responsabilitatea etică necesită alinierea dezvoltării AI la principiile etice, drepturile omului și valorile societale. Responsabilitatea societală presupune implicarea părților interesate pentru a se asigura că tehnologiile IA respectă demnitatea umană, autonomia și justiția.



# 03. Rolul responsabilității în abordarea Bias algoritmic

CU3 | Responsabilitate





### 03. Rolul responsabilității în abordarea Bias algoritmic

Responsabilitatea joacă un rol esențial în abordarea prejudecăților algoritmice din cadrul sistemelor de inteligență artificială (AI).

Pe măsură ce tehnologiile IA devin din ce în ce mai integrate în diverse aspecte ale societății, preocupările privind prejudecățile și echitatea au ajuns în prim-plan. Prejudecățile algoritmice se referă la erorile sistematice sau nedreptatea din algoritmi AI, care conduc la rezultate discriminatorii pentru anumite persoane sau grupuri. Aceste prejudecăți pot perpetua inegalitățile existente, pot consolida stereotipurile și pot submina încrederea în sistemele AI. În această secțiune, vom analiza modul în care responsabilitatea poate contribui la identificarea, atenuarea și prevenirea prejudecăților algoritmice, asigurând că tehnologiile IA sunt corecte, etice și echitabile.

Unul dintre aspectele-cheie ale responsabilității în abordarea prejudecăților algoritmice este **transparența**. Transparența se referă la deschiderea și accesibilitatea sistemelor AI, permițând părților interesate să înțeleagă cum funcționează algoritmi, de ce sunt luate anumite decizii și ce factori influențează rezultatele acestora. Sistemele AI transparente permit utilizatorilor să analizeze și să conteste rezultatele algoritmice, promovând încrederea și responsabilitatea.





De exemplu, dacă un sistem de inteligență artificială este utilizat pentru luarea automată a deciziilor privind aprobarea împrumuturilor, transparența ar implica divulgarea criteriilor și a factorilor luați în considerare în procesul decizional, cum ar fi istoricul de credit, venitul și informațiile demografice. Această transparență permite persoanelor să înțeleagă de ce cererea lor de împrumut a fost aprobată sau respinsă și le oferă posibilitatea de a recurge la o cale de atac în cazul în care consideră că decizia a fost părtinitoare sau nedreaptă.

Un alt aspect al responsabilității în abordarea prejudecăților algoritmice este **explicabilitatea**. Explicabilitatea se referă la capacitatea sistemelor AI de a oferi explicații interpretabile pentru deciziile și acțiunile lor. Inteligența artificială explicabilă permite utilizatorilor să înțeleagă raționamentul din spatele rezultatelor algoritmice și să identifice și să corecteze prejudecățile sau erorile. De exemplu, în cazul în care un instrument de angajare bazat pe inteligență artificială respinge automat cererile de angajare ale persoanelor cu anumite caracteristici demografice, explicabilitatea ar implica furnizarea de explicații privind motivul pentru care aceste cereri au fost respinse, cum ar fi evidențierea criteriilor sau caracteristicilor specifice care au condus la decizie. Această explicabilitate permite utilizatorilor să conteste deciziile părtinitoare și să ia măsuri corective pentru a atenua părtinirea din sistemele AI.

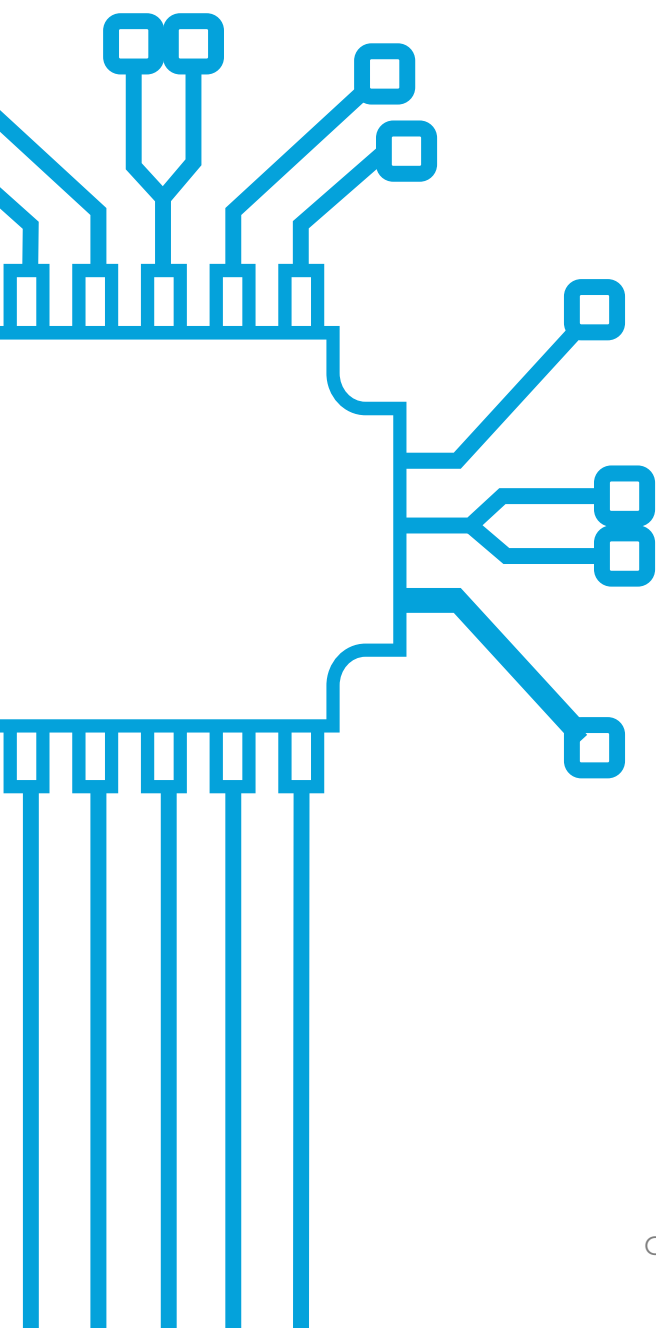
Responsabilitatea implică, de asemenea, **responsabilitatea pentru proiectarea, dezvoltarea și implementarea sistemelor de inteligență artificială**. Dezvoltatorii și practicienii IA au responsabilitatea de a se asigura că sistemele lor sunt corecte, imparțiale și echitabile. Această responsabilitate include efectuarea de teste și validări amănunțite pentru a identifica și atenua prejudecățile din algoritmi AI, precum și monitorizarea și auditarea sistemelor AI pentru a asigura respectarea permanentă a standardelor etice și juridice. De exemplu, dezvoltatorii de inteligență artificială pot utiliza tehnici precum algoritmi de detectare a părtinirilor, metrici de echitate și teste contradictorii pentru a identifica și aborda părtinirea din modelele lor. Aceștia pot, de asemenea, să pună în aplicare procese de audit și revizuire periodică a sistemelor de inteligență artificială pentru a identifica și aborda prejudecățile care pot apărea în timp.

În plus, responsabilitatea în abordarea prejudecăților algoritmice necesită **colaborarea și implicarea cu diverse părți interesate**. Aceasta include implicarea comunităților afectate, a grupurilor de advocacy, a factorilor de decizie politică și a organismelor de reglementare în proiectarea, dezvoltarea și supravegherea sistemelor AI. Prin colaborarea cu părțile interesate, profesioniștii din domeniul IA pot obține informații valoroase cu privire la impactul potențial al tehnologiilor lor și se pot asigura că sistemele lor sunt concepute și implementate într-un mod care respectă drepturile și interesele tuturor persoanelor și grupurilor.



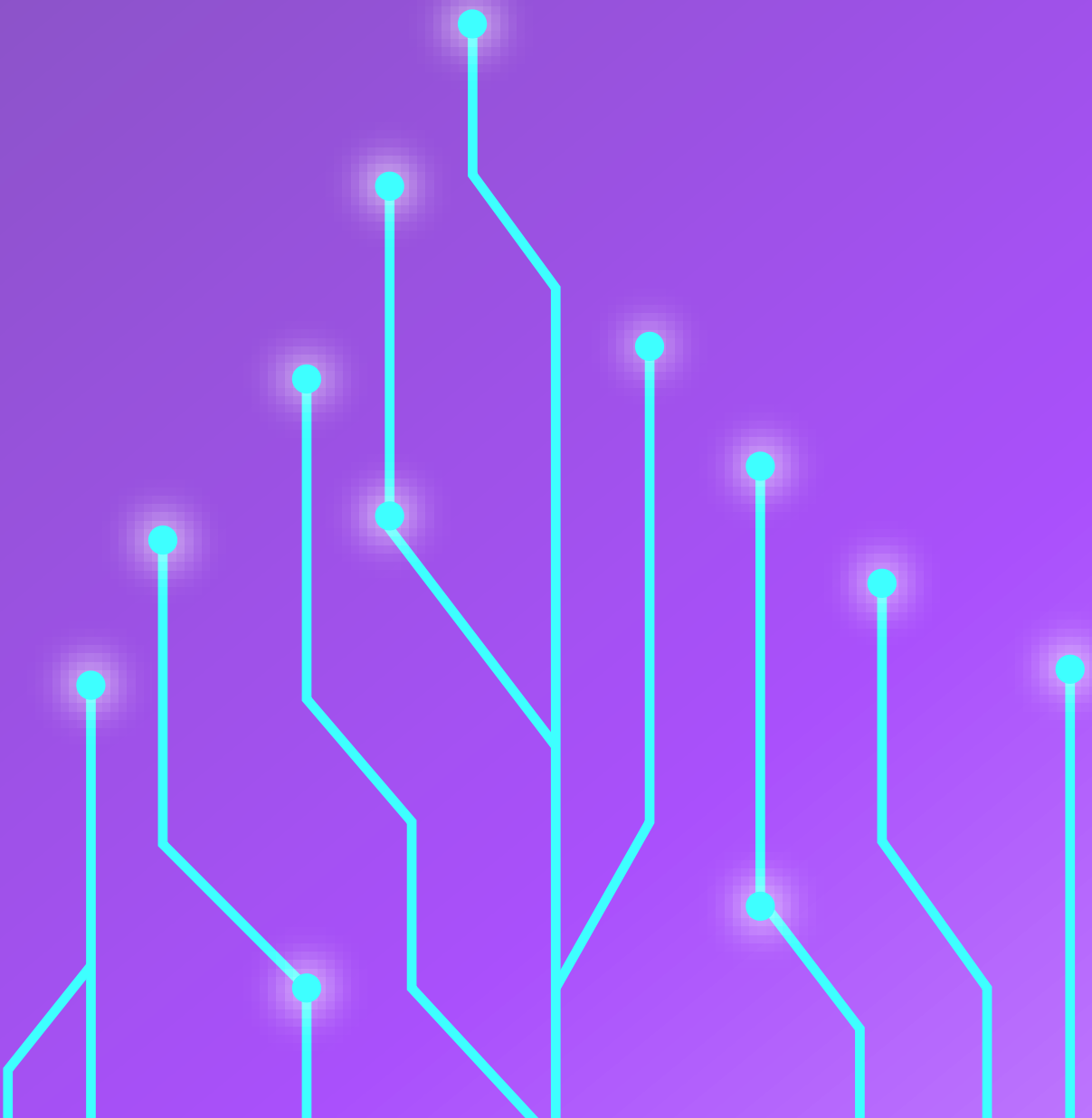


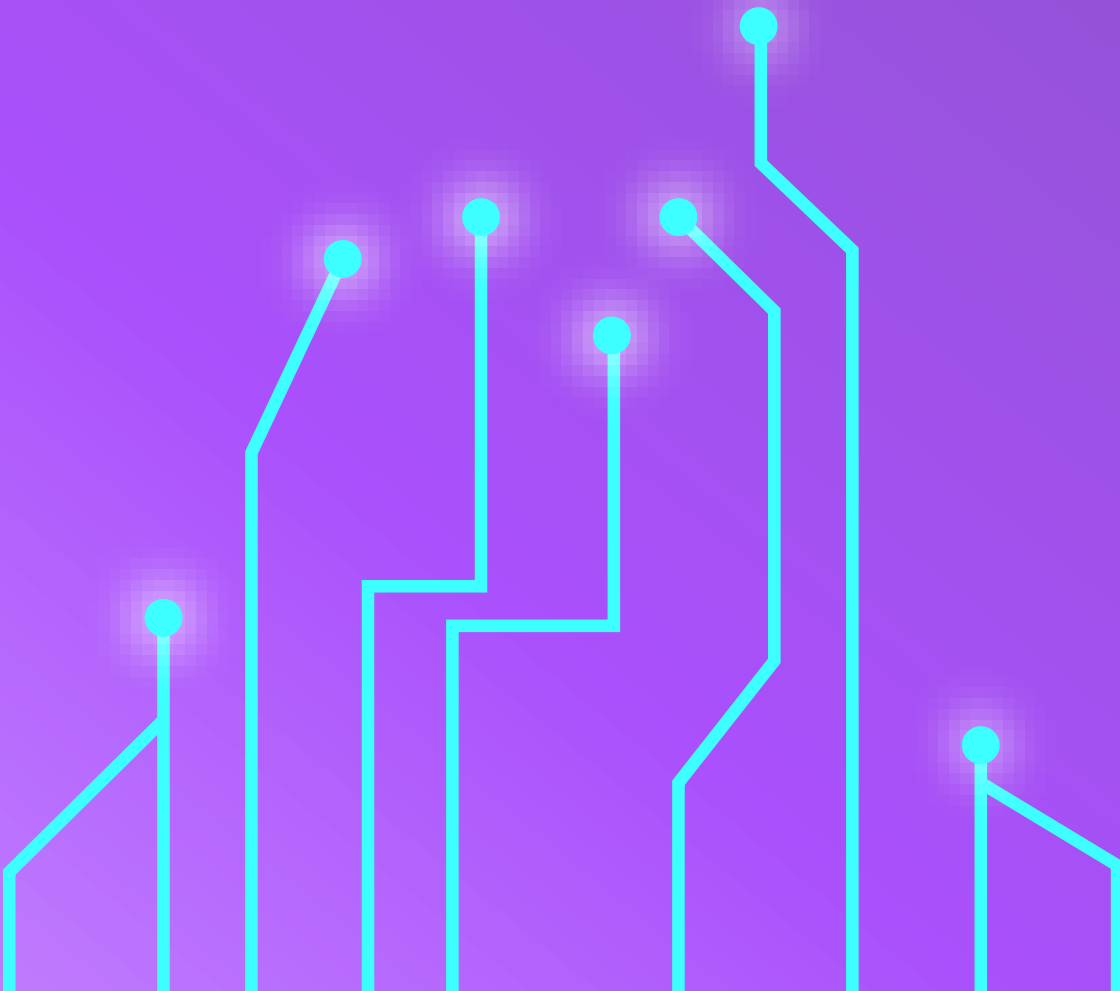
De exemplu, dacă un sistem de inteligență artificială este utilizat pentru poliția predictivă, responsabilitatea ar implica consultarea membrilor comunității și a organizațiilor pentru drepturile civile pentru a înțelege preocupările și perspectivele acestora cu privire la utilizarea inteligenței artificiale în aplicarea legii și pentru a aborda orice prejudecăți sau prejudicii potențiale.



# 04. Mecanisme pentru asigurarea responsabilității în sistemele AI

CU3 | Responsabilitate





## 04. Mecanisme pentru asigurarea responsabilității în sistemele AI

În peisajul tehnologic actual, care evoluează rapid, în care sistemele de inteligență artificială (AI) exercită o influență semnificativă asupra proceselor decizionale, asigurarea responsabilității este esențială. Mecanismele de asigurare a responsabilității în sistemele AI cuprind o serie de instrumente și cadre menite să promoveze comportamentul etic, transparența și corectitudinea în dezvoltarea și utilizarea tehnologiilor AI:

- **Orientările** servesc drept principii generale și bune practici pentru dezvoltatorii și utilizatorii de inteligență artificială, oferind îndrumări privind considerațiile etice, gestionarea responsabilă a datelor și strategiile de reducere a riscurilor.
- Principiile **Netiquette**, sau eticheta internetului, pot fi adaptate pentru a promova un comportament etic în sistemele AI, punând accentul pe respect, civilizație și comunicare responsabilă între utilizatori.
- **Politicile de utilizare acceptabilă (AUP)** descriu standardele de conduită și orientările de utilizare pentru sistemele AI, delimitând comportamentele acceptabile și consecințele utilizării necorespunzătoare sau abuzive.





- **Cadrelle de reglementare** stabilesc cerințe și standarde juridice pentru dezvoltarea și implementarea inteligenței artificiale, inclusiv dispoziții privind responsabilitatea, transparența și supravegherea. În Uniunea Europeană (UE), **cadrele de reglementare precum Regulamentul general privind protecția datelor (GDPR) și propunerea de lege privind inteligența artificială** stabilesc cerințe și standarde juridice pentru dezvoltarea și implementarea inteligenței artificiale. Aceste reglementări includ dispoziții privind responsabilitatea, cum ar fi cerințele de transparență, corectitudine și supraveghere în dezvoltarea și utilizarea sistemelor de inteligență artificială. De exemplu, în conformitate cu GDPR, dezvoltatorii trebuie să asigure transparența modului în care sistemele AI prelucrează datele cu caracter personal și să furnizeze persoanelor informații cu privire la logica implicată în luarea automată a deciziilor.

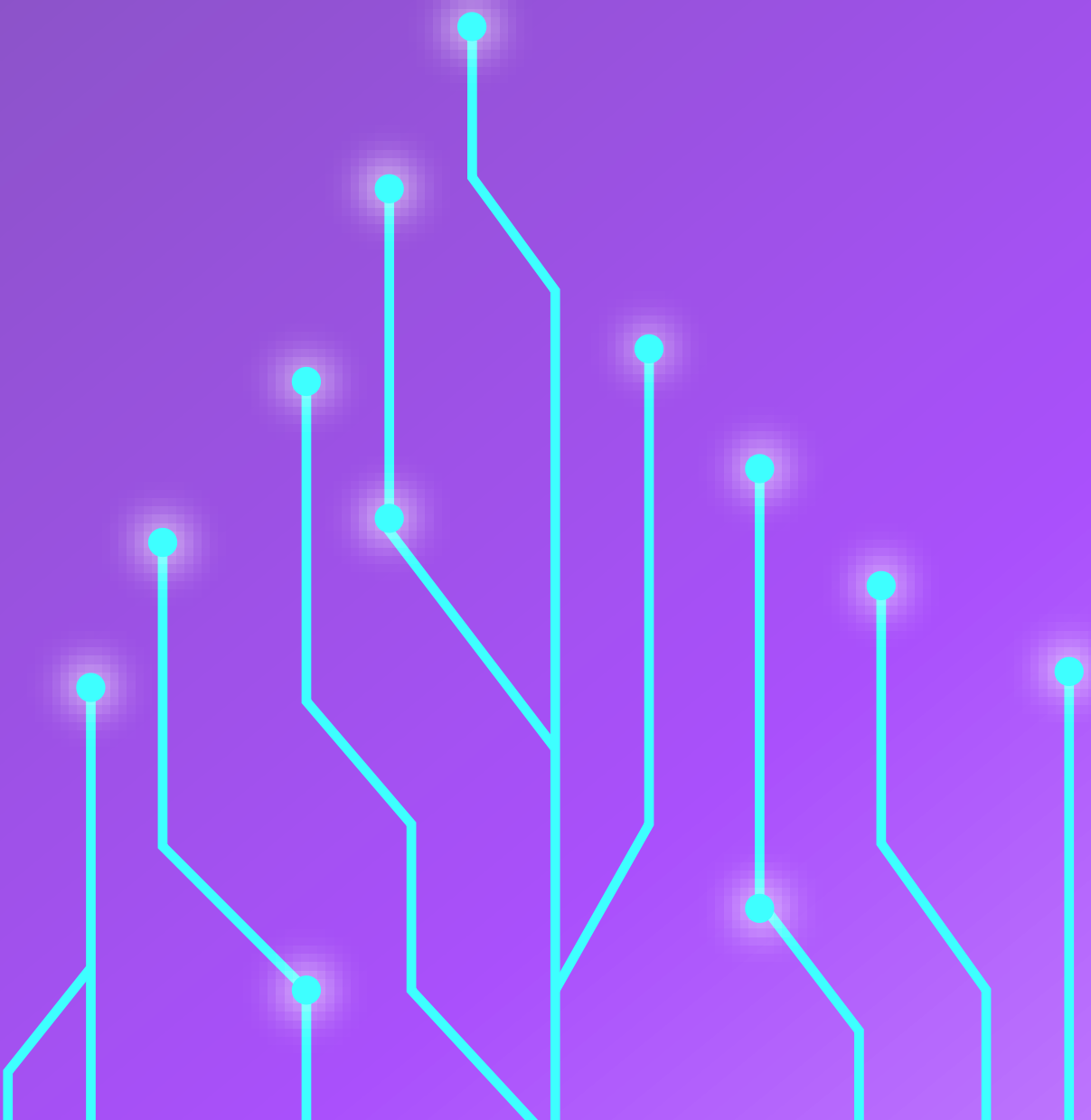


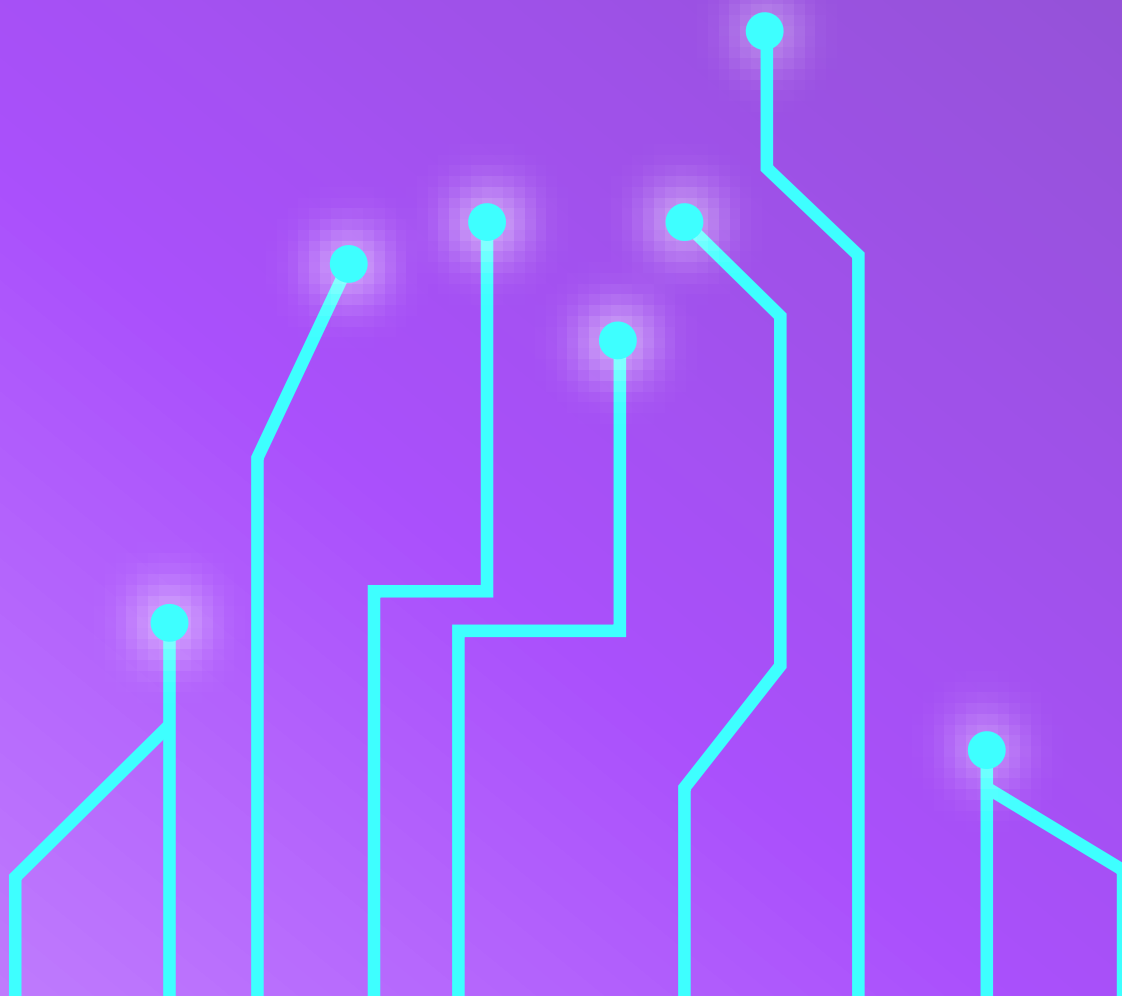
Prin valorificarea acestor mecanisme, părțile interesate pot promova o cultură a responsabilității și a comportamentului etic în dezvoltarea și implementarea IA, asigurându-se că tehnologiile IA sunt utilizate în mod responsabil și etic în beneficiul societății.



# 05. Concluzii

CU3 | Responsabilitate





## 05. Concluzii

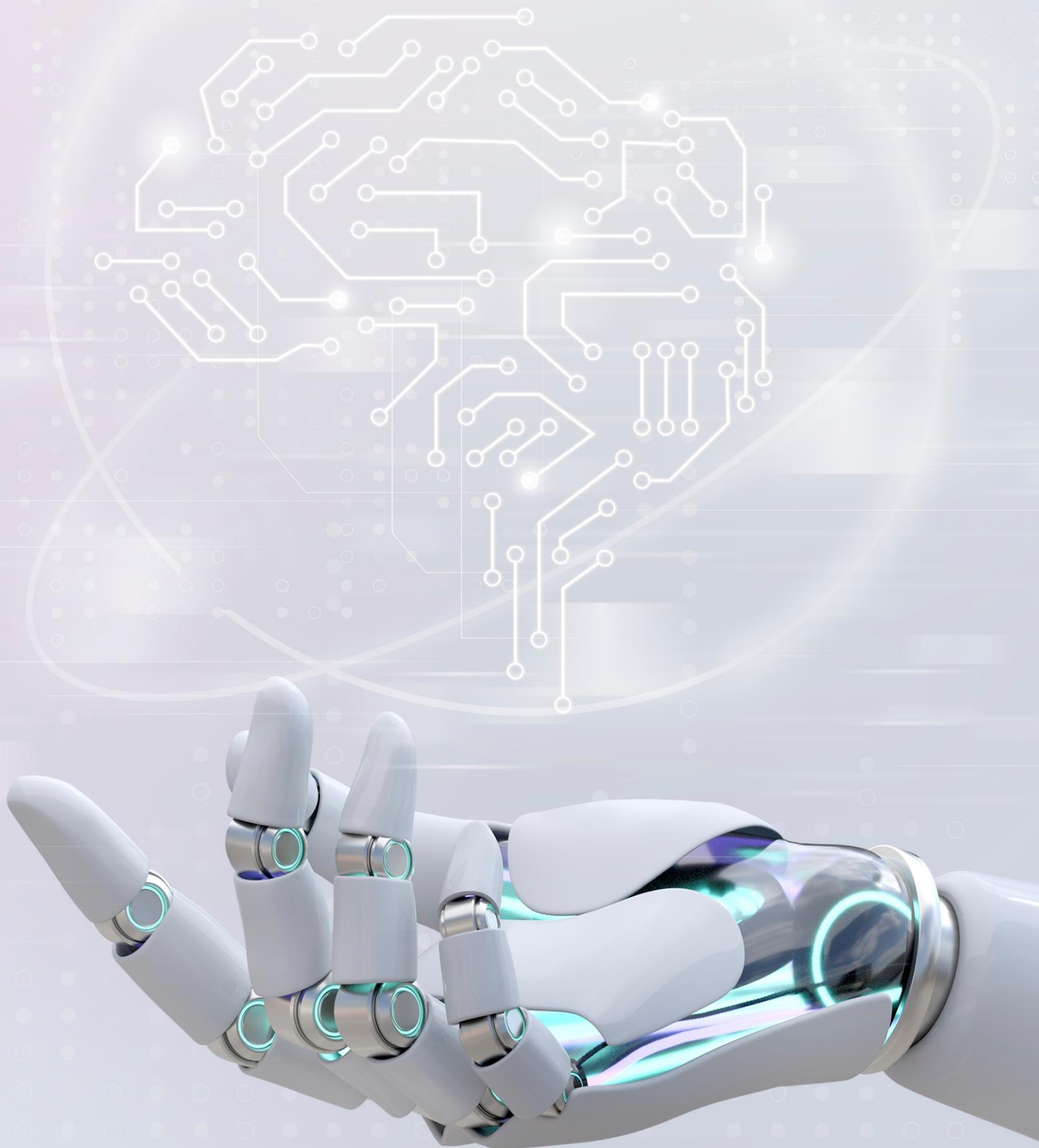
În concluzie, responsabilitatea reprezintă o piatră de temelie în dezvoltarea și implementarea etică a sistemelor de inteligență artificială (IA). Pe parcursul acestui CU, am explorat natura multidimensională a responsabilității în IA, aprofundând principiile sale fundamentale, mecanismele subiacente și implicațiile practice. De la transparență și explicabilitate la implicarea părților interesate și respectarea reglementărilor, responsabilitatea servește drept principiu director pentru a garanta că tehnologiile IA sunt dezvoltate și utilizate într-un mod care respectă standardele etice, încurajează transparența și promovează corectitudinea și echitatea.

Prin asumarea responsabilității, părțile interesate pot naviga prin complexitatea tehnologiilor IA cu integritate, responsabilitate și angajament față de bunăstarea societății.

Pe măsură ce IA continuă să evolueze și să pătrundă în diferite aspecte ale vieții umane, responsabilitatea rămâne esențială pentru conturarea unui viitor în care tehnologiile IA să servească binele comun și să susțină valorile eticii, justiției și demnității umane.









# CharTe



Cofinanțat de  
Uniunea Europeană

Finanțat de Uniunea Europeană. Punctele de vedere și opiniile exprimate aparțin, însă, exclusiv autorului (autorilor) și nu reflectă neapărat punctele de vedere și opiniile Uniunii Europene sau ale Agenției Executive Europene pentru Educație și Cultură (EACEA). Nici Uniunea Europeană și nici EACEA nu pot fi considerate răspunzătoare pentru acestea.



Universitat  
de les Illes Balears



helixconnect



2022-1-ES01-KA220-HED-000085257