Prietenie artificială - riscuri în interacțiunea cu chatboții AI

Această activitate este construită de Centrul pentru Jurnalism Independent în cadrul proiectului SNOOP-Educativ și face parte din colecția de resurse de educație media puse la dispoziție în cadrul Programului de Educație Media.

Concepte relevante din educație media:
inteligența artificială, siguranța online

Durata activității:
50 de minute

Recomandat pentru clasa/clasele:
a XI-a și a XII-a

Obiective de învățare:

La finalul activității, elevii vor putea să:
identifice riscuri ale utilizării aplicațiilor de inteligență artificială.

Pregătire (pentru profesor):

1. Citește articolul *„Sfaturi de la boți. Cum ajung adolescenții să-și caute răspunsurile în conversații cu inteligența artificială”*, publicat în cadrul proiectului SNOOP-Educativ, pentru a te familiariza cu tema de discuție. O parte dintre situațiile documentate în articol vor fi folosite în cadrul exercițiilor de la clasă.

2. Vizionează videoul (îl găsești în articol la același *link*), publicat în cadrul aceluiași proiect, care abordează aceleași teme. Secvențe din video vor fi folosite în cadrul exercițiilor de la clasă.

3. Citește Reperele teoretice (anexate mai jos), pentru a avea o privire de ansamblu asupra conceptelor pe care le vei discuta cu elevii.

4. Descarcă prezentarea *„Prietenie artificială - riscuri în interacțiunea cu chatboții AI”* pusă la dispoziție în pachetul de resurse. Aceasta conține repere teoretice explicate pe înțelesul elevilor și etapele activității transpuse pas cu pas.

5. Printează în cinci exemplare fișa de lucru pusă la dispoziție în Anexa 1. Elevii vor lucra în grupe în cadrul exercițiilor propuse.

6. Cere elevilor să vizioneze videoul ca temă pentru acasă înaintea orei în care vei implementa această activitate, pentru a-i familiariza cu tema de discuție.

Resurse necesare pentru implementarea activității:
proiector, acces la internet, fișe de lucru, sticky notes

Etapa 1: Valorificarea cunoștințelor elevilor privind tema abordată (prezentarea PDF, paginile 2-6)

Proiectează prezentarea PDF ***„Prietenie artificială - riscuri în interacțiunea cu chatboții AI”***
Durată: 5 minute

1. Prezintă tema discutată.
Ex.: Astăzi vom discuta despre inteligența artificială și cum ne poate afecta atunci când interacționăm cu ea zi de zi. Vă amintiți videoul pe care l-ați avut de vizionat ca temă pentru acasă? Vom discuta despre aceste lucruri pornind de la situațiile prezentate în video.

2. Proiectează activitatea online interactivă „Situații din viața de zi cu zi (AI)”. Obiectivul acestui pas este de a sonda grupul de elevi cu privire la interacțiunile lor cu aplicațiile de inteligență artificială și de a-i conecta cu tema de discuție.

Link activitate online interactivă: *https://wordwall.net/resource/101948593* (dacă nu ai posibilitatea de a lucra online, poți folosi paginile 3-6 din prezentarea PDF)

Activitatea online interactivă afișează patru cutii „surpriză”. Cere elevilor să aleagă ordinea în care deschideți cutiile (click pe cutia aleasă).

Fiecare cutie deschisă va dezvălui câte o situație de interacțiune cu aplicațiile de inteligență artificială. Cere elevilor să arate dacă au trecut prin situații similare ridicând mâna.

Situații incluse:

- Folosesc zilnic o aplicație AI sau un chatbot AI (ChatGPT, Gemini, Copilot, Claude etc.).
- I-am cerut sfaturi unui chatbot AI despre relațiile cu persoane din viața mea (familie, prieteni, colegi etc.).
- I-am cerut sfaturi unui chatbot AI despre sănătatea mintală și despre sănătate în general.
- Folosesc des chatboți AI pentru a obține informații de care am nevoie (pentru teme, despre interesele mele personale, știri etc.)

După fiecare moment de sondare, cere elevilor să ofere exemple personale (colectează 1-2 răspunsuri pentru fiecare situație).

Etapa 2: Fixarea reperelor teoretice și explorarea temei

Durată: 30 de minute

1. **Prezintă** riscurile asociate utilizării inteligenței artificiale. (10 minute, prezentarea PDF, paginile 7-11)

2. **Exercițiu** (20 de minute, prezentarea PDF, pagina 12):

- **Împarte elevii** în cinci grupe de lucru și oferă fiecărei grupe câte o fișă de lucru. Obiectivul acestui exercițiu este ca elevii să identifice riscurile utilizării inteligenței artificiale în niște situații date și să se gândească la modalități prin care ar putea gestiona expunerea la aceste riscuri.
- **Proiectează în plen videoul** (disponibil aici, selectează secvențele discutate în fișa de lucru). (Elevii pot viziona ulterior videoul accesând codul QR de pe fișa de lucru, dacă au nevoie.)

- **Cere elevilor să lucreze împreună pe grupe pentru a rezolva cerințele de pe fișa de lucru.** La finalul activității, fiecare grupă va desemna câte un raportor care va povesti rezultatele în plen.
  - **Timp de lucru:** 10 minute.
  - **Timp alocat procesării:** 10 minute.
  - **Repere** pentru rezolvarea cerințelor sunt disponibile în Anexa 2, mai jos.

- **Întrebări de sprijin pentru etapa de procesare a exercițiului:**
  - Vi se pare că tinerii umanizează chatbotul în situațiile prezentate, că îl consideră uman? Ce vă face să credeți acest lucru?
  - Cum își construiesc încrederea în chatbot tinerii din situațiile prezentate?
  - Dacă au încredere „oarbă" în chatbot, la ce riscuri se expun tinerii din situațiile prezentate?
  - Ce ați face voi pentru a vă proteja în fața riscurilor pe care le-ați identificat?

Pe parcursul procesării exercițiului, oferă **feedback corectiv** acolo unde este nevoie. Este esențială înțelegerea corectă a conceptelor și aplicarea lor pentru analiza situațiilor reale în care se pot găsi elevii.

Etapa 3: Reflecție

Durată: 10 minute

1. Concluzie (prezentarea PDF, paginile 15-16): Explică elevilor la final că oamenii ajung să acorde încredere sistemelor AI precum chatboții, ceea ce îi expune la o bună parte dintre riscurile discutate.

Odată ce le oferă încredere chatboților, multe persoane uită că ei nu sunt oameni, ci roboți, și că nu pot simți lucruri, nu pot gândi ca oamenii, că imită pur și simplu modul în care vorbesc oamenii.

Acest lucru înseamnă că pot greși informații, că pot transmite informații false utilizatorilor, sau că pot reflecta prejudecăți ale oamenilor, ori îi pot face să se izoleze de cei din jur.

De aceea, este important să verificăm informațiile pe care le primim de la un chatbot AI și să ne reamintim constant că nu este inteligent și că este doar un robot, o unealtă pe care o putem folosi, dar deciziile despre viața noastră trebuie să fie luate de noi.

2. Reflecție:

- Cere elevilor să răspundă la întrebarea „Ce vei face diferit de acum încolo în interacțiunea cu chatboți AI în viața ta de zi cu zi?" folosind câte un sticky note (2 minute).
- După ce au completat răspunsul, cere elevilor să vină pe rând în fața clasei, să citească răspunsul și să lipească sticky note-ul pe tablă.
- Pe măsură ce elevii citesc răspunsurile, le poți adresa întrebări suplimentare (Ex.: De ce este important pentru tine să schimbi acest aspect în comportamentul tău? Cum crezi că te va ajuta această schimbare de comportament? etc.)

Repere teoretice (pentru profesor):

Ce este inteligența artificială (AI)?

Inteligența artificială este un set de tehnologii care se referă la capacitatea unor sisteme informatice în a imita funcții umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea.

Sau, din punct de vedere juridic un sistem AI este considerat „un sistem bazat pe o mașină care poate, pentru un set dat de obiective definite de om, să facă predicții, recomandări sau decizii care influențează mediile reale sau virtuale".[1]

În ciuda termenului în sine, sistemele de inteligență artificială **nu au și nu dobândesc inteligență.** Ele funcționează de obicei pe baza unor algoritmi complecși (de învățare automată, învățare profundă prin rețele neuronale - vezi AI generativă etc.), adică sunt (auto-)programate să urmeze reguli, pași specifici pentru a rezolva o problemă sau a îndeplini o sarcină. Sistemele de inteligență artificială ajung la rezultate **analizând seturi foarte mari de date din care extrag tipare și concluzii după algoritmi matematici.** Inteligența artificială **generativă** reprezintă capacitatea sistemelor de a crea texte, imagini, muzică sau videouri noi, pornind de la elementele sau tiparele pe care le-au învățat din datele pe care le-a analizat.

Resurse suplimentare: https://apti.ro/inteligenta-artificiala-zeu-sau-masina

Ce este un chatbot AI?

Un chatbot AI este o aplicație care folosește inteligența artificială (de fapt, o anumită parte a acesteia numită model de limbaj, LLM - Large Language Model) pentru a genera răspunsuri la întrebările utilizatorilor sau pentru a încerca să îndeplinească sarcini primite de la aceștia. În funcție de model, chatboții AI pot efectua sarcini precum scrierea de texte, generarea de imagini sau video etc. Printre chatboții AI cei mai cunoscuți se regăsesc ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic).

Cum funcționează un chatbot AI?

Chatbotul AI este antrenat pe seturi foarte mari de date care provin din surse diverse, uneori selectate de oameni (cercetători, ingineri). Seturile de date pot conține texte, imagini, videouri - din cărți, pagini de web publice, articole, enciclopedii etc.

Antrenamentul presupune că sistemul AI analizează datele primite - prin transformarea lor în elemente pe care le poate procesa (tokeni)[2], identifică și extrage tipare. Acest lucru înseamnă că sistemul AI nu vede și nu dă un sens cuvintelor și textelor pe care le analizează, ci că poate doar identifica relații între tokeni, în formă de termeni, concepte sau elemente vizuale, combinații de cuvinte, structuri de construcție în diverse contexte care apar cel mai frecvent în datele cu care este antrenat. Antrenarea include și reguli și limite predefinite (guardrails[3]), care deși deseori sunt puse cu un scop anume (securitatea și confidențialitatea datelor, calitatea și relevanța răspunsurilor, filtrarea conținutului nepotrivit etc.), pot include forme de cenzură sau limitare.

Când primește o întrebare sau o sarcină, chatbot-ul AI aplică tiparele învățate, combinând cele mai des întâlnite elemente în datele analizate pentru a genera un răspuns statistic sau pentru a crea conținut nou care estimează că se potrivește cu ceea ce îi cere utilizatorul.
Astfel, un chatbot AI:

- nu dă răspunsuri adevărate sau false, pentru că criteriul de adevăr nu există legat de informațiile pe care le procesează;
- nu poate distinge în mod fiabil convingerile de cunoștințe și fapte;[4];
- nu gândește.

[1] *Definiția folosită în legislația din SUA privind reglementarea AI și, în linii mari, cea oferită de OECD. O definiție similară se va regăsi și în Regulamentul AI, proaspăt agreat de Uniunea Europeană.*
[2] *https://en.wikipedia.org/wiki/Large_language_model*
[3] *https://www.datacamp.com/blog/llm-guardrails*
[4] *https://www.nature.com/articles/s42256-025-01113-8*

Ce riscuri majore pot apărea în interacțiunea cu un chatbot AI?

Important!
Scopul pe termen lung al companiei care dezvoltă un chatbot AI generalist este să țină utilizatorii cât **mai mult timp conectați și interacționând cu aplicația.** Astfel, chatbotul AI poate **personaliza experiența utilizatorului** cât mai bine, ceea ce înseamnă profit pentru companie. Compania face bani, spre exemplu, prin afișarea reclamelor în conversații sau prin creșterea cotei de piață odată ce crește baza de utilizatori ai chatbotului.

Interacțiunea cu un chatbot AI implică o serie de riscuri, printre care:

1. Prejudecăți

Un chatbot AI poate **prelua prejudecățile** de orice tip (inclusiv culturale, sociale, rasiale sau de gen) din societate, indiferent că acestea sunt **urmare a datelor de antrenament sau lipsa ori inacurateța acestora, a modului de antrenare și a regulilor impuse sau a interacțiunii precedente cu unul sau mai mulți utilizatori.** Acest lucru se întâmplă pentru că, așa cum am explicat mai sus, sistemele AI nu înțeleg de fapt lumea, ele doar generează răspunsuri pe baza tiparelor identificate. De exemplu, dacă datele respective reflectă stereotipuri sau dezechilibre în reprezentarea anumitor grupuri, un chatbot le poate reproduce automat, fără să realizeze că ar fi prejudecăți. Dacă un chatbot ar fi antrenat cu seturi de date unde este descris mai frecvent că femeile spală hainele în familie, când va fi întrebat despre acest lucru, va susține că femeile spală hainele în familie.

În același timp, unele companii care dezvoltă chatboți AI încearcă să intervină prin **reguli de siguranță** pentru a diminua propagarea stereotipurilor și antrenează sistemele create să recunoască anumite combinații de cuvinte sau contexte care sunt asociate frecvent cu stereotipuri. Astfel, sistemul este învățat să semnaleze și să explice sau să evite aceste tipare atunci când generează răspunsuri, în funcție de cum este programat. Uneori, însă, chiar și cu aceste reguli, chatbotul tot poate să dea informații greșite.

Spre exemplu, un chatbot a fost instruit să folosească și persoane de culoare în răspunsurile sale - pentru a oferi o reprezentare incluzivă. Atunci când i s-a cerut să genereze imagini cu senatori ai Statelor Unite de la 1800, a inclus femei de culoare și indigene, ceea ce reprezintă incorect din punct de vedere istoric statutul femeilor în societatea americană și „șterge” o istorie a rasismului și a discriminării pe bază de gen. (Prima senatoare a fost o femeie albă abia în 1922.)[5]

Resurse suplimentare: https://apti.ro/inteligenta-artificiala-arbitru-corupt

2. Umanizare[6]

Chatboții sunt construiți pentru a fi „de ajutor”, utili, folosiți constant. Deseori sunt antrenați să răspundă politicos, să fie cooperativi și să mențină conversațiile într-un ton pozitiv, prietenos. Acest lucru poate aduce mai multe riscuri, cum ar fi:

- Chatboții **nu critică și nu atenționează întotdeauna utilizatorul** dacă greșește sau dacă spune ceva problematic, pentru a nu-l „supăra” și a-i face pe plac. Sau pentru că nu știe ce este o greșeală și ce este o problemă.
- Chatboții **sunt foarte des de acord cu utilizatorul,** confirmând uneori informații greșite sau false sau interpretări tendențioase sau dezechilibrate ale utilizatorului despre evenimentele din jurul său.
- Chatboții răspund mereu **prietenos folosind uneori tipare lingvistice care pot fi interpretate ca fiind forme de expresie ale empatiei, deși nu sunt de fapt decât o simulare și nu reflectă emoții reale.** Sistemele pot descrie sentimente și pot părea că le recunosc în poveștile pe care le spun utilizatorii, însă acest lucru se datorează antrenamentului, nu calității de a fi umani.

[5] *https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical*
[6] Înțelegem prin umanizare sau antropomorfizare „fenomenul psihologic prin care oamenii tind să interacționeze social cu entități non-umane (de exemplu, tehnologie, animale, plante, entități supranaturale, fenomene naturale sau sociale) ca și cum aceste entități ar fi umane.” (Festerling & Siraj, 2021)

- Chatboții își **personalizează răspunsurile** în funcție de tonul, stilul și interacțiunile cu utilizatorii, folosind deseori informații din conversații pentru a trage concluzii despre utilizator (spre exemplu, dacă utilizatorul cere des ajutor la teme despre concepte predate la istorie în clasa a X-a, chatbotul va infera că discută cu un adolescent și nu cu un adult), pe care le pot folosi apoi în răspunsurile pe care le generează. Acest lucru poate lăsa **impresia** utilizatorului că chatbotul este un bun ascultător, căruia îi pasă de el, care îl cunoaște, ceea ce crește șansele ca utilizatorul să dezvolte un atașament emoțional pentru chatbot.
- Interfața chatbotilor este creată si optimizată pentru a **părea cât mai umani**, inclusiv prin topologia de răspuns, crearea unei personalități a chatbotului sau introducerea de mici imperfecțiuni umane (de exemplu, răspunsuri precum „Lasă-mă să mă gândesc un pic… Hmm, uite o idee…”)

Tendința chatboților de a fi agreabili creează contextul potrivit pentru ca utilizatorii să îi umanizeze, crezând că discută cu cineva căruia îi pasă de ei, nu cu un sistem informatic. Acest lucru îi face pe utilizatori să acorde foarte ușor **un grad ridicat de încredere** chatbotului în legătură cu tot ceea ce spune, chiar și atunci când informațiile sunt incorecte sau incomplete.

Umanizând chatbotul, utilizatorii pot ajunge să se simtă **atașați emoțional** de acesta. Îl pot considera prieten, partener romantic sau substitut pentru părinți, profesori, doctori sau terapeuți, căruia îi pot cere sfaturi fără frica de a fi judecați sau criticați. În situațiile extreme, unii utilizatori ajung să se izoleze de persoanele din jurul lor, iar dependența de interacțiunea cu sistemul AI îi poate împiedica să ceară ajutor și sprijin în viața reală atunci când au nevoie.

Atenție! Interacțiunile dese cu chatboții predispun utilizatorii să **împărtășească date personale** cu aceștia. Informațiile respective nu sunt doar utilizate pentru a putea răspunde cât mai personalizat la întrebările utilizatorului, ele pot fi folosite și în moduri necunoscute acestuia (pentru antrenarea chatbotului, pentru crearea unui profil detaliat pentru scopuri comerciale, de țintire prin reclame - utilizările depind de politicile companiilor care dezvoltă chatboți AI). Acest lucru înseamnă că utilizatorul pierde controlul asupra lor sau că pot fi folosite împotriva dorinței lui.

Resurse suplimentare: https://apti.ro/inteligenta-artificiala-iti-poate-fi-chatbotul-prieten

3. Informații greșite/false/înșelătoare. Deepfake

Chatboții AI joacă un rol tot mai mare în **generarea și propagarea informațiilor greșite, false, sau înșelătoare**, atât neintenționat, ca urmare a limitărilor sistemelor AI, cât și intenționat, când sunt utilizați de oameni sau chiar de companiile care îi pun pe piață cu intenții malițioase.

Chatboții AI pot oferi răspunsuri care conțin informații greșite, false sau inventate, pentru că nu pot verifica context, informații sau surse așa cum o fac oamenii.

Așa cum am mai menționat, chatbotul AI funcționează pe baza unor modele statistice care analizează date pentru a identifica tipare. Ele calculează cât de des, în ce proporție, în ce contexte și relații apar combinații de cuvinte, imagini sau concepte. Aceste tipare sunt apoi folosite pentru a genera răspunsuri care par potrivite sau plauzibile, care **par corecte**. Sistemul AI nu știe dacă informațiile sunt corecte însă, el doar combină elementele întâlnite cel mai frecvent și care se potrivesc cel mai bine în contextul întrebării primite de la utilizator (și al celorlalte limitări impuse de dezvoltatori). Dacă în seturile de date analizate există informații incorecte, false sau înșelătoare, sistemele AI le pot **reproduce sau folosi în generarea răspunsurilor, fără să poată face distincția între ce este adevărat, corect și ce nu.**

Mai mult, chatboții AI pot **„halucina”**[7], adică pot genera fapte inventate, care nu există în realitate, dar care sunt prezentate având structura informațiilor corecte. Spre exemplu, au existat cazuri în care sisteme AI au inventat surse bibliografice care nu există (autorii existau, dar titlurile lucrărilor citate nu).

[7] *https://www.datacamp.com/blog/ai-hallucination*

Acest lucru se întâmplă pe de o parte pentru că chatbotul AI folosește tipare pentru a răspunde și nu are un „simț" al moralității - pentru sistem, dacă informația **arată corect conform tiparelor învățate,** atunci o folosește pentru a răspunde convingător utilizatorului fără să își pună problema că ar oferi o informație inventată. Pe de altă parte, această limitare provine și din faptul că fiind construite să fie mereu de ajutor, unele sisteme AI au dificultăți în a „recunoaște" că nu **dețin informațiile necesare** pentru a putea răspunde sau efectua o sarcină, pentru că ar însemna să nu mai fie utile. În consecință, **regurgitează informații** pentru a acoperi golurile pe care le au.

Chatboții AI sau aplicații similare pot fi utilizate pentru a crea texte, imagini, clipuri audio sau video care prezintă în mod **intenționat fals** un eveniment, cum ar fi **știrile false** sau inventate, videoclipuri **deepfake** care arată persoane în situații care nu au avut loc de fapt, sau teorii ale **conspirației** formulate convingător.

Un caz particular îl reprezintă **pornografia deepfake**, care reprezintă portretizarea unei persoane în ipostaze de natură sexuală (imagini sau video) fără consimțământul acesteia, folosind fotografii cu fața și, uneori, cu corpul acesteia. La momentul construirii acestei activități educaționale, victimele materialelor de pornografie deepfake erau covârșitor fetele și femeile, în proporție de 99%.[8]

Sfaturi pentru elevi privind interacțiunea cu chatboți AI

- Învață - cel puțin la nivel primar - cum funcționează un chatbot AI pentru a înțelege când și cum te poate ajuta și când nu.
- Testează întrebări cu un chatbot AI pe un domeniu pe care îl cunoști bine. Acest lucru te va ajuta să realizezi că nu este întotdeauna o sursă de informație corectă.
- Verifică **orice** informație oferită de un chatbot AI, inclusiv sursele sau interpretarea acestora.
- Nu povesti lucruri personale unui chatbot AI.
- Reține că un ton prietenos nu înseamnă că stai de vorbă cu un om. Chatbotul AI este un robot.
- Nu lua decizii doar pe baza sfaturilor unui chatbot AI. Verifică din mai multe surse. Cercetează și discută cu persoane de încredere din jurul tău atunci când ai nevoie.
- Nu înlocui interacțiunile reale cu discuțiile cu un chatbot AI. Există riscul să te atașezi emoțional de un chatbot și să te izolezi de cei din jur, ceea ce reprezintă o îngrijorare privind sănătatea mintală.
- Nu folosi tehnologia AI pentru a crea informații false sau înșelătoare, sau deepfake-uri prin care poți răni alte persoane. Și nu uita ca orice interacțiune cu un sistem informatic lasă urme.

[8] *https://www.scena9.ro/article/tehnologie-rusine-nud-deepfake-eleve-liceu-arad-exmatriculare*

Anexa 1: Fișă de lucru

Prietenie artificială - riscuri în interacțiunea cu chatboții AI

Vizionați videoul disponibil în articolul SNOOP scanând codul QR.

Discutați cu colegii de grupă și răspundeți la întrebările de mai jos.

La final, **alegeți un coleg** care să reprezinte grupa și să prezinte rezultatele voastre în fața clasei.

Întrebarea 1

Ce riscuri ale utilizării inteligenței artificiale considerați că apar în situațiile prezentate în video? Justificați răspunsurile voastre.

Situație	Riscuri
„Atunci când nu știu ce să gătesc, îi trimit lui ChatGPT o poză cu frigiderul meu, îi și scriu ce alimente mai am în frigider și îmi face un plan pentru toată săptămâna.”	
„Irina, studentă, a greșit temele de două ori din cauza ChatGPT-ului. O dată la desen, când a scris despre o pictură care de fapt nu exista. Și a trebuit să refacă tema. Altă dată la etică, unde a luat nota 1.”	
„Un adolescent din SUA l-a transformat pe ChatGPT în cel mai bun prieten și i-a mărturisit planul de-a se sinucide. ChatGPT nu i-a spus să caute sprijin, iar părinții au găsit conversațiile după ce băiatul nu mai era.”	

Întrebarea 2

Plecând de la informațiile oferite în video, explicați, pe scurt, de ce ajung unele persoane să aibă încredere „oarbă" în chatboți AI precum ChatGPT?

Întrebarea 3

Ce sfaturi ați oferi unui tânăr pentru a-l ajuta să gestioneze riscurile la care este expus prin utilizarea inteligenței artificiale?

Anexa 2: Fișă de lucru

Prietenie artificială - riscuri în interacțiunea cu chatboții AI (Cheie de răspuns - sugestii)

Vizionați videoul disponibil în articolul SNOOP scanând codul QR.

Discutați cu colegii de grupă și răspundeți la întrebările de mai jos.

La final, **alegeți un coleg** care să reprezinte grupa și să prezinte rezultatele voastre în fața clasei.

Întrebarea 1

Ce riscuri ale utilizării inteligenței artificiale considerați că apar în situațiile prezentate în video? Justificați răspunsurile voastre.

Situație	Riscuri
„Atunci când nu știu ce să gătesc, îi trimit lui ChatGPT o poză cu frigiderul meu, îi și scriu ce alimente mai am în frigider și îmi face un plan pentru toată săptămâna.”	***Sugestii de răspunsuri care pot fi corecte:*** • ***(Umanizare; control asupra vieții)*** *Apare riscul ca Alexia să considere chatbotul uman pentru că îi răspunde ca un prieten. De aceea, ea îi acceptă sfaturile, ceea ce înseamnă că nu mai ia ea decizii (despre mâncare, în acest exemplu), ci chatbotul;* • ***(Protecția datelor)*** *Apare riscul ca informațiile personale pe care i le dă Alexia chatbotului să fie folosite în moduri despre care ea nu știe;* • ***(Umanizare)*** *Tot pentru că îi oferă informații și poze personale, chatbotul poate să își adapteze foarte ușor răspunsurile ca să fie pe placul ei. Acest lucru o poate face pe Alexia să aibă foarte mare încredere în el și să uite că nu este decât un robot.*
„Irina, studentă, a greșit temele de două ori din cauza ChatGPT-ului. O dată la desen, când a scris despre o pictură care de fapt nu exista. Și a trebuit să refacă tema. Altă dată la etică, unde a luat nota 1.”	***Sugestii de răspunsuri care pot fi corecte:*** • ***(Umanizare; informații false/greșite)*** *Apare riscul ca Irina să considere chatbotul uman și să se bazeze pe el ca pe un prieten sau un expert. Acest lucru o poate face să nu verifice deloc informațiile pe care i le spune și să folosească mai departe informații false sau greșite. Chatbotul nu poate verifica dacă o informație e adevărată așa cum fac oamenii;* • ***(Dezvoltare personală)*** *Dacă se bazează doar pe ceea ce îi spune chatbotul, Irina nu mai învață nimic pe cont propriu și îi vor lipsi multe cunoștințe importante.*

„Un adolescent din SUA l-a transformat pe ChatGPT în cel mai bun prieten și i-a mărturisit planul de-a se sinucide. ChatGPT nu i-a spus să caute sprijin, iar părinții au găsit conversațiile după ce băiatul nu mai era.”	***Sugestii de răspunsuri care pot fi corecte:*** • ***(Umanizare; izolare)*** *Apare riscul ca tinerii să considere chatbotul uman, pentru că acesta se comportă ca un prieten. Acest lucru îi poate face să aibă încredere doar în chatbot și să se izoleze de persoanele din jurul lor. Uneori, vor simți că doar chatbotului îi pot împărtăși ce simt, prin ce trec și vor ascunde chiar și de părinții lor faptul că trec prin momente dificile;* • ***(Umanizare, confirmare constantă a opiniilor)*** *Chatbotul este creat să răspundă mereu într-un fel care îi dă dreptate constant utilizatorului. De aceea, este foarte ușor ca tinerii să îl considere uman și să aibă încredere în el că le oferă un spațiu sigur în care pot spune orice au nevoie. Această situație ne arată însă că chatbotul nu are capacitatea de a gândi ca un om și de a înțelege că sinuciderea este un pericol real și că ar fi trebuit să ofere sfatul de a căuta sprijin de la persoane reale.*

Întrebarea 2

Plecând de la informațiile oferite în video, explicați, pe scurt, de ce ajung unele persoane să aibă încredere „oarbă” în chatboți AI precum ChatGPT?

Sugestii de răspuns corect:
Unele persoane ajung să aibă încredere „oarbă” în chatboți AI pentru că sunt mereu disponibili, nu judecă, nu critică, sunt mereu de acord cu ceea ce li se spune, nu au cum să pedepsească utilizatorul în niciun fel. Iar aceste lucruri îi fac să pară umani, să pară că ar avea empatie. În realitate, nu sunt decât niște programe complexe care imită conversațiile umane, folosind tipare pe care le învață din cum comunică oamenii între ei.

Întrebarea 3

Ce sfaturi ați oferi unui tânăr pentru a-l ajuta să gestioneze riscurile la care este expus prin utilizarea inteligenței artificiale?

Sugestii de răspunsuri care pot fi corecte:

- *Să nu uite că un chatbot nu este decât un robot, nu este uman.*
- *Chatboții doar imită modul în care comunică oamenii, nu au empatie. Nu sunt prietenii noștri, nu sunt doctori, nu sunt profesori, nu sunt consilieri, nu sunt părinți. Dacă trecem prin momente dificile de sănătate mintală, ar trebui să căutăm sprijin la persoanele reale din jurul nostru.*
- *Chatboții pot să facă greșeli. De aceea e important să verifice informația înainte de a o folosi la teme sau pentru a lua decizii în viața de zi cu zi.*
- *Chatboții nu trebuie să fie cei care iau decizii pentru oameni. Pot să ne ajute să luăm decizii, dar nu ar trebui să ne bazăm doar pe un chatbot și să nu mai gândim deloc.*
- *E important să avem grijă ce informații personale dăm unui chatbot, pentru că le poate folosi în moduri despre care nu știm, dar și pentru a își adapta răspunsurile ca să ne răspundă mereu pe plac.*



CENTRUL PENTRU
JURNALISM INDEPENDENT





Această activitate este construită de Centrul pentru Jurnalism Independent în cadrul proiectului SNOOP Educativ și face parte din colecția de resurse de educație media puse la dispoziție în Programul de Educație Media, dezvoltat de CJI în parteneriat strategic cu Romanian-American Foundation.

SNOOP Educativ

PROIECT CO-FINANȚAT DE:
Administrația Fondului cultural național

„SNOOP Educativ" este un proiect co-finanțat de Administrația Fondului Cultural Național – AFCN.

Proiectul nu reprezintă în mod necesar poziția AFCN, care nu este responsabilă de conținutul sau de modul în care rezultatele pot fi folosite.