



Ghid pentru utilizarea responsabilă a instrumentelor de inteligență artificială generativă în INFLPR

Acest ghid a fost elaborat pe baza documentului “Living guidelines on the responsible use of generative AI in research” [1], European Commission, Directorate-General for Research and Innovation Directorate E-Prosperity, Unit E4 - AI in Science & Critical Technologies

1. Scop și domeniu de aplicare

Prezentul ghid stabilește un cadru practic pentru utilizarea responsabilă a instrumentelor de inteligență artificială (AI) generativă în activitățile de cercetare, dezvoltare, inovare, management de proiect, diseminare științifică și suport administrativ desfășurate în cadrul INFLPR.

Inteligența artificială generativă (AI generativă sau GenAI) reprezintă o categorie de inteligență artificială capabilă să creeze conținut nou, precum texte, imagini, materiale audio și video sau cod informatic. Această tehnologie analizează volume mari de date, identifică tipare și stiluri, iar apoi utilizează informațiile dobândite pentru a genera conținut pe baza unor prompturi — instrucțiuni sau cerințe formulate de utilizator. Printre cele mai cunoscute instrumente de GenAI se numără Microsoft Copilot, ChatGPT, Google Gemini, Claude, Midjourney, etc.

Ghidul urmărește valorificarea beneficiilor AI generativ fără afectarea integrității cercetării, a confidențialității, a protecției datelor, a drepturilor de proprietate intelectuală, a securității informației și a responsabilității umane asupra rezultatelor științifice.

Recomandările se aplică cercetătorilor, doctoranzilor, personalului tehnic, managerilor/directorilor/responsabililor de proiect, personalului suport pentru cercetare și, după caz, colaboratorilor și partenerilor care utilizează instrumente de AI generativ în relație cu activitățile INFLPR.

2. Introducere – AI-ul generativ și cercetarea științifică

AI-ul generativ oferă numeroase oportunități pentru toate sectoarele de activitate. Cu toate acestea, aduce cu sine și riscuri, precum generarea masivă de dezinformare, probleme legate de proprietatea intelectuală și protecția datelor, precum și alte utilizări neetice cu consecințe sociale și de mediu semnificative.

Cercetarea este unul dintre domeniile care ar putea fi profund transformate de AI-ul generativ. Această tehnologie are un potențial uriaș de a accelera descoperirile științifice, ducând la inovații majore și la creșteri semnificative ale productivității, îmbunătățind totodată eficiența și ritmul proceselor de cercetare și verificare.

Câteva exemple practice privind utilizarea pozitivă a acestor instrumente de către cercetători includ [2]:

- Sprijinirea vorbitorilor non-nativi în redactarea de texte în mai multe limbi;
- Crearea rapidă de rezumate din surse diverse, analizând volume masive de date;
- Extragerea și contextualizarea automată a unui volum vast de informații,
- Asistență în scrierea de cod pentru analiza datelor.

Cu toate acestea, tehnologia implică și un risc de abuz. Unele riscuri decurg din limitările tehnice ale instrumentului, în timp ce altele sunt legate de utilizarea sa (intenționată sau nu) în moduri care subminează bunele practici în cercetare.

Alte riscuri pentru mediul de cercetare ar putea proveni din natura proprietară (închisă) a unora dintre instrumentele AI - de exemplu: lipsa de transparență, taxele de acces la servicii sau modul în care sunt folosite datele introduse de utilizatori [3].

3. Principii cheie

Pornind de la punctele comune identificate în orientările elaborate de diverse părți interesate, principiile de bază pentru utilizarea responsabilă a AI-ului generativ în cercetare sunt:

- **Fiabilitate:** Asigurarea calității cercetării la nivel de concepție, metodologie, analiză și utilizare a resurselor. Aceasta presupune verificarea și reproducerea informațiilor generate de AI și conștientizarea riscurilor legate de egalitate și nediscriminare (prejudecăți, erori sau inexactități).
- **Onestitate:** Desfășurarea, revizuirea, raportarea și comunicarea cercetării într-un mod transparent, echitabil, riguros și imparțial. Acest principiu impune declararea explicită a utilizării instrumentelor de AI generativ.
- **Respect:** Considerație față de colegi, participanți și subiecți de cercetare, societate, patrimoniu cultural și mediu înconjurător. Utilizarea responsabilă a AI-ului generativ trebuie să țină cont de limitările tehnologiei, de amprenta sa ecologică și de efectele sociale (diversitate, echitate, prevenirea prejudiciilor). Acest principiu include gestionarea corectă a informațiilor, respectarea confidențialității, a protecției datelor cu caracter personal, a drepturilor de proprietate intelectuală și citarea corespunzătoare a surselor.
- **Responsabilitate:** Asumarea deplină a cercetării, de la idee până la publicare, precum și a managementului, formării, supravegherii și impactului societal general. Cercetătorul poartă răspunderea completă pentru toate rezultatele generate, pe baza principiului controlului și supravegherii umane.

4. Ghid privind utilizarea responsabilă a AI-ului generativ în cercetare

Recomandări pentru personalul CDI

Pentru a utiliza AI-ul generativ într-un mod responsabil, cercetătorii trebuie:

A. Să rămână pe deplin responsabili pentru rezultatele științifice obținute

- Cercetătorii sunt direct răspunzători pentru integritatea conținutului pe care îl produc, chiar dacă acesta este generat integral sau parțial cu ajutorul instrumentelor AI.
- Cercetătorii trebuie să păstreze o abordare critică față de rezultatele oferite de AI-ul generativ și să fie conștienți de limitările tehnologiei, precum erorile sistematice (*bias*), halucinațiile și inexactitățile.
- Sistemele AI nu sunt autori și nici coautori. Calitatea de autor implică asumarea responsabilității și capacitatea de acțiune conștientă, atribute care aparțin exclusiv cercetătorilor umani.

Modelele de AI generativ pot prezenta limitări care le afectează atât răspunsurile, cât și utilitatea practică. Cele mai frecvente tipuri sunt:

- **Erori sistematice din datele de antrenare (Training Data Bias):** Distorsiunile prezente în datele utilizate pentru antrenarea modelelor pot duce la răspunsuri părtinitoare, reflectând inexactitățile sau concepțiile sistematice din materialele sursă.
- **Influența instrucțiunii (Prompt Bias):** Modelele își pot alinia răspunsurile la convingerile sau preferințele percepute ale utilizatorului – un fenomen cunoscut sub numele de „comportament lingușitor/obedient” (*sycophantic behaviour*) – generând rezultate înșelător de concordante sau părtinitoare.
- **Citări inventate și rezumate incorecte:** Modelele de AI generativ pot inventa referințe bibliografice care par plauzibile, dar sunt complet false, inducând în eroare utilizatorii care se bazează pe aceste surse pentru cercetare sau verificare. Chiar și atunci când citările sunt corecte (lucrarea există, are titlul exact etc.), rezumatul generat de AI pentru acele articole poate fi eronat. Prin urmare, cercetătorii au responsabilitatea de a verifica toate referințele, conținutul citărilor și rezumatele obținute.
- **Interpretabilitate:** Modelele de AI generativ funcționează ca niște „cutii negre” (*black boxes*), ceea ce face dificilă înțelegerea modului în care este generat un anumit răspuns. Această opacitate subliniază importanța validării încrucișate (*cross-validation*), în special în analiza automată a datelor, unde răspunsurile generate de model pot influența semnificativ concluziile științifice.

B. Să utilizeze în mod transparent AI-ul generativ

- **Detalierea utilizării:** Pentru a asigura transparența, cercetătorii trebuie să specifice în detaliu ce instrumente de AI generativ au fost utilizate substanțial în procesele lor de cercetare. Atunci când AI-ul generativ influențează semnificativ rezultatele, utilizarea acestuia trebuie declarată în secțiunea de metodologie (sau echivalentă), conform cerințelor jurnalului sau standardelor din disciplina respectivă, evaluând în mod responsabil amploarea contribuției instrumentului. Acest principiu de transparență se aplică și atunci când AI-ul este folosit pentru detectarea sau semnalarea practicilor neconforme (de exemplu, detectarea halucinațiilor sau a plagiatului), menținând permanent supravegherea umană.
- **Caracterul stochastic și reproductibilitatea:** Cercetătorii trebuie să țină cont de natura stohastică (aleatorie) a instrumentelor de AI generativ, respectiv tendința acestora de a produce

rezultate diferite pentru aceeași instrucțiune de intrare (prompt). Cercetătorii trebuie să urmărească reproductibilitatea și robustețea rezultatelor și a concluziilor, declarând și analizând limitările instrumentelor utilizate (inclusiv eventualele erori sistematice din conținutul generat) și măsurile luate pentru atenuarea acestora.

C. Să acorde atenție sporită la confidențialitate, protecția vieții private și drepturile de proprietate intelectuală la partajarea informațiilor sensibile sau cu caracter protejat

- **Protecția datelor introduse:** Cercetătorii trebuie să fie conștienți de faptul că datele introduse sau încărcate (text, date brute, instrucțiuni, imagini etc.) pot fi utilizate în alte scopuri, cum ar fi antrenarea viitoarelor modele AI. Prin urmare, lucrările nepublicate sau sensibile (atât proprii, cât și ale altora) trebuie protejate prin evitarea încărcării lor în sisteme AI externe, cu excepția cazului în care există garanții explicite că datele nu vor fi refolosite (de exemplu, pentru antrenarea modelelor de limbaj) sau reutilizate într-un mod imposibil de urmărit și verificat.
- **Datele cu caracter personal ale terților:** Cercetătorii nu trebuie să introducă date cu caracter personal aparținând terților în sisteme externe de AI generativ, cu excepția cazului în care persoana vizată și-a dat consimțământul explicit și există un scop clar determinat, garantând astfel respectarea normelor europene privind protecția datelor (GDPR).
- **Înțelegerea riscurilor tehnice și juridice:** Cercetătorii trebuie să înțeleagă implicațiile tehnice, etice și de securitate privind viața privată, confidențialitatea și proprietatea intelectuală. În acest sens, este necesară verificarea ghidurilor instituționale, a setărilor de confidențialitate ale instrumentului, a entității care administrează serviciul (instituție publică, privată, companie etc.), a locului unde rulează instrumentul și implicațiile pentru orice informație încărcată. Mediul de rulare poate varia de la infrastructuri închise (locale) sau găzduire pe infrastructuri terțe cu confidențialitate garantată, până la platforme deschise, accesibile liber pe internet.
- **Utilizarea asistentilor/agentilor AI** (Comet, Atlas, OpenClaw, etc) care funcționează ca asistenți personali sau asistenți de sistem **nu este permisă în cadrul INFLPR conform regulamentelor interne**, deoarece acestea pot beneficia de un acces extins la date, aplicații și sisteme și pot genera riscuri semnificative privind securitatea și confidențialitatea informațiilor.

D. Să respecte legislația națională, internațională aplicabilă

- Rezultatele generate de AI pot fi deosebit de sensibile în raport cu drepturile de proprietate intelectuală și protecția datelor cu caracter personal:
- **Prevenirea plagiatului:** Cercetătorii trebuie să fie atenți la riscul de plagiat (text, cod, imagini etc.) atunci când utilizează conținut generat de AI. Dreptul de autor al terților trebuie respectat prin citarea corespunzătoare a lucrărilor originale. Conținutul generat de un model AI poate avea la bază rezultatele altor cercetători, necesitând recunoaștere explicită și citare adecvată.
- **Gestionarea datelor cu caracter personal din rezultate:** Rezultatele produse de AI pot conține date cu caracter personal. Dacă se identifică astfel de date, cercetătorii sunt direct responsabili să le trateze conform normelor europene de protecție a datelor (GDPR).

E. Să aibă un proces de învățare continuă și formare pentru utilizarea eficientă a instrumentelor AI

- **Actualizarea competențelor:** Instrumentele de AI generativ evoluează rapid, apărând constant noi modalități de utilizare. Cercetătorii trebuie să rămână la curent cu bunele practici și să le împărtășească colegilor și comunității academice.
- **Reducerea impactului ecologic:** Cercetătorii trebuie să urmărească minimizarea amprenteii de mediu a tehnologiei prin evaluarea critică a necesității de a folosi un instrument AI, selectarea soluției optime pentru sarcina respectivă și folosirea unor tehnici eficiente de formulare a cerințelor.

F. Să evite utilizarea substanțială a instrumentelor AI în activități sensibile de evaluare

- Se recomandă abținerea de la utilizarea substanțială a AI-ului generativ în activități critice cu impact direct asupra altor cercetători sau organizații (de exemplu: evaluarea semenilor/peer-review, evaluarea candidaților în concursuri de încadrare pe posturi de cercetare-dezvoltare, evaluarea propunerilor de proiecte de cercetare). Dacă totuși se folosesc aceste instrumente, cercetătorii evaluează și analizează propunerile pentru publicare, finanțare, numire în funcție, promovare sau premiere într-un mod transparent și justificat și declară utilizarea inteligenței artificiale și a instrumentelor automatizate [4].
- **Echitate în evaluare:** Evitarea acestor instrumente elimină riscul de evaluare incorectă cauzat de limitările AI (halucinații, erori sistematice).
- **Protecția confidențialității:** Se protejează lucrările originale și nepublicate ale colegilor împotriva expunerii neautorizate sau a integrării lor involuntare în seturile de antrenare ale modelelor AI.

G. Să ofere atenție la utilizarea AI-ului generativ de către terți în interacțiunile cu cercetătorii din INFLPR

- Cercetătorii din INFLPR trebuie să fie conștienți de faptul că terți pot utiliza instrumente de AI generativ – de exemplu, în cadrul ședințelor, pentru luarea de notițe, rezumarea discuțiilor sau realizarea de sinteze ale unor documente (care pot fi confidențiale). În consecință, aceștia se pot confrunta cu riscuri privind confidențialitatea, protecția datelor sau drepturile de proprietate intelectuală legate de informațiile protejate sau originale.
- Cercetătorii trebuie să informeze terți cu privire la utilizarea acestor tipuri de instrumente și la riscurile potențiale.
- În interacțiunile cu terți, dacă cercetătorul este cel care utilizează instrumente AI, trebuie să anunțe interlocutorii și să explice modul în care sunt protejate datele colectate.

5. Autoevaluarea importanței utilizării AI în generarea de conținut original

În continuare este prezentat un „checklist” care poate fi verificat înainte de utilizarea unui document generat cu un instrument de AI generativ, în vederea evitării unor posibile situații de încălcare a eticii:

- ☐ Conținutul pe care urmează să îl introduc este public sau poate fi transmis legal unui terț?
- ☐ Fișierul conține date personale, informații confidențiale, rezultate nepublicate, proprietate intelectuală sau informații ale partenerilor?

- ☐ Știu dacă furnizorul instrumentului reutilizează datele pentru antrenarea modelelor?
- ☐ Pot verifica independent răspunsul generat?
- ☐ Am verificat citările, cifrele, formulele și interpretările?
- ☐ Utilizarea AI are un impact substanțial asupra metodei, analizei sau concluziilor și trebuie declarată?
- ☐ Există o politică specifică a jurnalului, finanțatorului, proiectului sau partenerului care trebuie respectată?
- ☐ Este această activitate una sensibilă, de evaluare sau peer review, pentru care utilizarea AI ar trebui evitată?

Clasificarea practică a utilizărilor AI

Nivel	Exemple	Condiții minime	Declarație de utilizare AI
Nivel 1 – suport editorial	Corectură, reformulare, traducere, structurare, rezumat al unui text propriu	Verificare umană; fără date sensibile în servicii externe	De regulă nu este necesară, dacă editorul/finanțatorul nu solicită altfel
Nivel 2 – suport substanțial	Revizuire de literatură, ipoteze, cod pentru analiză, interpretarea datelor, identificarea lacunelor	Validare independentă; păstrarea trasabilității; verificarea surselor și a limitărilor	Recomandată / necesară conform cerințelor jurnalului, proiectului sau disciplinei
Nivel 3 – activitate sensibilă	Peer review, evaluare de proiecte, documente confidențiale, date personale, rezultate brevetabile nepublicate	A se evita în servicii externe; necesită mediu securizat și, după caz, aprobare instituțională	Obligatorie dacă utilizarea este permisă și relevantă

Îndrumări privind GenAI:

- *European Research Area Forum GenAI guidelines [1]*
- *GenAI curs online gratuit: <https://www.elementsofai.ro/>*

Referințe bibliografice

[1] “Living guidelines on the responsible use of generative AI in research”, European Commission, Directorate-General for Research and Innovation Directorate E-Prosperity, Unit E4 - AI in Science & Critical Technologies, https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/updated-era-living-guidelines-responsible-use-generative-ai-research-2026-05-08_en

[2] <https://doi.org/10.1038/d41586-023-02980-0>



[3] Recomandarea Consiliului Europei din 23 Mai 2024 asupra creșterii securității cercetării (ST/9097/2024/INIT)

[4] The European Code of Conduct for Research Integrity, <https://allea.org/code-of-conduct/>