



Se lansează primul model dedicat limbii române pentru inteligența artificială de tip ChatGPT

Comunicat de presă

București, 14 mai

O echipă de cercetători români a creat un model de limbaj (LLM – „large language model”) dedicat limbii române, ce poate fi folosit pentru dezvoltarea unor unelte și platforme A.I. Modelul este „open source”, deci va putea fi accesat și folosit de oricine dorește să construiască instrumente bazate pe inteligența artificială. Odată cu publicarea acestui LLM, inițiatorii proiectului lansează și comunitatea OpenLLM-Ro, care dorește să-i aducă împreună pe toți cei care vor să contribuie în diverse forme la dezvoltarea tehnologiilor A.I. pentru limba română. Ambele proiecte sunt inițiate și derulate de către POLITEHNICA București, Universitatea din București și Institutul de Logica și Știința Datelor, cu sprijinul BRD Groupe Société Générale.

Deși este o tehnologie la care avem acces la scară largă de doar câțiva ani, mulți dintre noi am interacționat deja de foarte multe ori cu roboți conversaționali, precum ChatGPT (produs de OpenAI), Copilot (dezvoltat de Microsoft), Gemini (dezvoltat de Google) etc. Însă, pentru limba română rezultatele sunt uneori imprecise, pentru că modelele pe care sunt instruite nu au fost expuse multor surse românești de date. Totodată, aceste genuri de unelte nu pot fi folosite în companii, deoarece accesul direct poate fi restricționat din rațiuni de securitate și confidențialitate. O soluție în aceste situații este implementarea unui model local aflat în infrastructura companiei. Aceste modele publice care pot fi folosite local sunt în general, însă, antrenate pe limba engleză sau folosind un număr redus de documente în limbile cu circulație mai scăzută.

Modelul românesc lansat astăzi este adaptarea unui LLM public dezvoltat în principal pentru limba engleză. Însă acesta a fost expus la câteva milioane de documente în limba română, pentru a înțelege mai bine semnificația cuvintelor. Acest lucru e esențial pentru performanța unor astfel de modele în situații unde exprimarea cerinței sau a întrebării introduse de utilizator, cât și a răspunsului trebuie să fie în limba română. Începând din a doua parte a anului 2023, la dezvoltarea și instruirea acestui LLM a lucrat o echipă de cercetători de la POLITEHNICA București, Universitatea din București și Institutul de Logica și Știința Datelor. Partenerii academici au contribuit cu cercetători care au lucrat pro-bono și, în plus, POLITEHNICA București a pus la dispoziție și puterea de calcul necesară antrenării modelului. Partenerul principal al proiectului este BRD Groupe Société Générale, care susține inovația și tehnologiile viitorului în România în toate formele lor.

„Pentru ca mediul economic și/sau instituțional din România să poată folosi această nouă tehnologie atât de promițătoare, este nevoie de modele specializate care să fi întâlnit o mulțime de conversații și documente în limba română. Motivul este simplu: pentru a ne putea oferi informațiile de care avem

nevoie. În BRD lucrăm în mod constant la soluții care să ne îmbunătățească procesele de lucru, folosind cele mai noi tehnologii care pot aduce valoare adăugată în primul rând clienților noștri. Însă înțelegem totodată că nevoile noastre sunt comune cu ale multor alți actori instituționali și ne-am asumat misiunea de a sprijini timpuriu inovația în domeniul inteligenței artificiale. Implicându-ne în peisajul extrem de animat al acesteia, putem ajuta ca cele mai noi tehnologii să aibă un impact pozitiv în societatea românească aproape în același ritm cu evoluțiile din domeniu la nivel internațional”, a declarat **Horia Velicu, Head of Innovation Lab în cadrul BRD Groupe Société Générale.**

„Câteva dintre exemplele de utilizare ale modelului românesc sunt: căutarea de informații în baza de cunoștințe a unei organizații, cu ghiduri și proceduri de lucru, sau roboți conversaționali pentru clienții companiilor sau ai instituțiilor care să îi ghideze în parcurgerea pașilor necesari pentru utilizarea unui produs sau serviciu. În ambele cazuri, angajații și/sau clienții economisesc timp în accesarea informației, beneficiind în multe situații și de îmbunătățirea calității acesteia”, a spus **Alin Ștefănescu, directorul Departamentului de Informatică din cadrul Universității din București și vicepreședintele Institutului de Logică și Știința Datelor.**

Efortul de specializare a unui model dedicat unei limbi este frecvent coordonat de către comunitatea academică asociată limbii respective, exemple recente fiind din țări precum Franța, Germania, Spania, Finlanda, Bulgaria. Resursele necesare sunt însă considerabile atât din punct de vedere al infrastructurii tehnice necesare (de exemplu, hardware dedicat, precum plăci grafice de mare putere) cât și în ce privește cercetătorii și programatorii experimentați. Este deci nevoie de o susținere pe scară largă și pe termen mediu și lung din partea multor actori-cheie ai societății: mediul economic, cel academic și nu în ultimul rând mediul guvernamental, prin programele dedicate dezvoltării tehnologiilor de Inteligență Artificială.

De aceea, dezvoltatorii acestui model lansează în același timp și [comunitatea OpenLLM.ro](https://openllm.ro). Aceasta își propune să încurajeze interacțiunea între diverși actori sau facilitatori ce doresc să contribuie la dezvoltarea acestei tehnologii pentru limba română și lansarea unor modele specializate pentru anumite domenii. Inițierea acestui dialog într-un mediu „open source” va accelera crearea unor modele mai performante, implementate în companiile sau instituțiile românești, ceea ce va rezulta într-un plus de productivitate per ansamblu a întregii societăți.

„Sperăm ca lansarea acestui model să fie doar începutul unui efort de lungă durată din care să rezulte LLM-uri mai bune pentru limba română. Deja am descoperit o metodă pe care vrem să o aplicăm altor modele lansate recent (Llama-3 și Mistral) și care au performanțe în general mai bune ca cel de la care am pornit (Llama-2). Totuși, pentru a avea modele performante pentru limba română avem nevoie de 2 tipuri de resurse: colecții de date mari și curatoriate, de calitate bună, precum și resurse hardware (în special, GPU-uri pentru antrenarea modelelor). Sperăm ca atât entitățile private, cât și cele publice să înțeleagă importanța dezvoltării modelelor mari de limbaj și multimodale (text-imagini) pentru limba română. Pe toți îi așteptăm să ni se alăture în cadrul inițiativei OpenLLM-Ro și a proiectelor de cercetare care o vor sprijini”, a declarat **Traian Rebedea, conferențiar la POLITEHNICA București și cercetător principal la NVIDIA, unul dintre coordonatorii tehnici ai inițiativei OpenLLM-Ro.**

Raportul tehnic poate fi găsit aici: <https://arxiv.org/abs/2405.07703>

Modelul LLM poate fi descărcat de pe platforma Hugging Face: <https://huggingface.co/OpenLLM-Ro>

Codul asociat modelului poate fi descărcat de pe GitHub: <https://github.com/OpenLLM-Ro>

Alte detalii despre proiect: <https://ilds.ro/llm-for-romanian>

BRD pentru educație, tehnologie & inovație

BRD susține pregătirea viitoarelor generații de specialiști și antreprenori în tehnologie. Proiectele pe care BRD le are în vedere merg în direcția educației în disciplinele STEM: **FIRST Tech Challenge Romania**, **laboratoarele de robotică din cadrul universității POLITEHNICA București**, **programul Innovation Labs**, **parteneriatul cu Applied Data Science Centre din Universitatea din București**, **programul Innovators for Children** sau **Conferința How to Web**. Toate acestea reunesc elevi, studenți, profesori, tineri antreprenori, experți și au impactat până în prezent peste 50.000 de beneficiari.