



Abian da 'ebaluatoia', euskarazko txatbot irekia auzolanean hobetzeko bi asteko ekimena

ChatGPT eta antzeko eredu itxien pare edo hobeto dabilen euskarazko tresna irekia sortzea du helburu UPV/EHUko HiTZ zentroak

ALBISTEAK 2025/03/19



Euskal Herriko Unibertsitateko HiTZ zentroa Latxa euskarazko hizkuntza-eredua txatbot bihurtzeko lanean ari da, eta aldi berean txatboten kalitatea neurtzeko bide berriak irekitzen. Hizkuntza-ereduak adimen artifizialeko ereduak dira, giza hizkuntza ulertu eta sortzeko ikaskuntza automatikoko teknikak erabiltzen dituztenak. OpenAI edo Google bezalako enpresa handien ChatGPT eta antzeko eredu itxien pare edo hobeto dabilen euskarazko txatbot irekia sortzea da HiTZ zentroaren helburua. Enpresa horiek miloika dolarrekin eraiki dutenaren zati bat auzolanean egin nahi du HITZ-ek. Nola parte hartu? Oso erraza eta dibertigarria da: ebaluatoia.hitz.eus webgunean sartu, galdera edo eskaera egin, eta aukeratu zein den erantzun hoberena. Ebaluatoiaren lehen fase arrakastatsuan bezala, bigarren fase honetan ere lehiaketa eta sariak egongo dira. Martxoaren 19tik apirilaren 2ra egongo da ebaluatoian parte hartzeko aukera.

Hizkuntza-ereduen garapenak aurrerapauso nabariak egin ditu azken urteotan. Gaur egun, ChatGPT edo antzeko elkarrizketarako txatbotek erabiltzaileen galderei erantzuteko eta eduki berriak sortzeko gaitasun aurreratuak dituzte. Horiek eraikitzea enpresa handien esku dagoela dirudi, baina UPV/EHUko HiTZ Hizkuntza Teknologiaren Zentroan beste bide bat aukeratu da: existitzen den txatbot ireki on bat aukeratu eta horri euskaraz irakastea.

Pentsa daitekeen bezala proiektua anbiziosoa da, eta hainbat pauso egin behar dira euskarazko txatbot hoberena eraiki ahal izateko. Eneko Agirre HiTZ-eko zuzendariak dioenez, "euskarazko txatbota eraikitze bidean ikerkuntzako hainbat galdera berri planteatu zaizkigu. Gure helburua ez da bakarrik euskarazko txatbot hoberena eraikitzea, eta euskaraz gazteleraz bezain ondo ibiltzea. Hori lortzeko bidean ikerketa egiten ari gara, eta ikasten duguna argitalpen zientifikoetan kaleratuko dugu, jorratzen ari garen metodologia eta ikasten duguna beste hizkuntza askotarako ere baliogarria da eta".

Horregatik abiatzen da ebaluatoia.hitz.eus, ezinbestekoa baita erabiltzaile euskaldunek txatbotak erabili eta hoberena zein den aukeratzeko laguntzea. Ebaluatoian, erabiltzaileak eskaera bat egingo du webgunean, eta honek bi txatbot ezberdinen erantzunak emango dizkio bata bestearen ondoan, erabiltzaileak erantzunen artean onena aukeratu dezan. Alborapenak ekiditeko, erabiltzailearen bozka itsua da, hau da, erabiltzaileak ez daki zein bi txatbot ari den epaitzen. Txatboten sailkapena osatzeko xakean baliatzen den ELO ebaluazio-sistema erabiltzen da. Ereduak hasierako puntuazio arbitrario batetik abiatzen dira

(adibidez, 1.000 puntu bakoitzak), eta partidak irabazi, galdu edo berdindu ahala puntuazioa aldatuz joango da, aurkarien puntuazioaren arabera. Naiara Perez HiTZeko ikertzaileak dioenez “Ebaluazio mota honi arena deritzo, hizkuntza-ereduak elkarrekin lehiatzen direlako, eta txatbotak eszenatoki errealista eta praktikoa ebaluatzea ahalbidetzen du. Zentzu horretan ohikoa bihurtu da enpresa beraiek eredu hoberenak Ingeleserako Chatbot Arena delakoan lehiatzera bidaltzea. Guk nahiago izan dugu ebaluatoia deitzea”.

Ebaluatoian HiTZek eraikitako 18 txatbot probatuko dira, baita txatbot komertzial hoberenak ere. Denetarik dago, txatbot oso onak, ez hain onak, eta baita txarrak direnak ere, ikerketa zientifikoan ondorioak atera ahal izateko. Oscar Sainz HiTZeko ikertzailearen esanetan, "tartean egongo da estatuan inoiz eraiki den txatbot handiena ere, 70B parametroko Latxan oinarritutako txatbot erraldoia".

Duela gutxi ebaluatoiaren lehen fase bat antolatu zuen HiTZ zentroak. Oso arrakastatsua izan zen, bost milatik gora parte-hartze jaso dira, 250 parte-hartzailetik gora. Parte-hartzaile emankorrenak 470 ariketa egin ditu, eta horren arabera saria jaso du. Handik ikasitakoekin bigarren fase hau antolatu dugu, bi astez iraungo duena gaurtik aurrera, eta oraingoan helburua hogeimila ariketa jasotzea da, baietz lortu!

Informazio osagarria

Latxa hizkuntz eredu Eusko Jaurlaritzak finantzatutako IKER-GAITU proiektuaren esparruan garatu da. Eraldaketa Digitalerako eta Funtzio Publikorako Ministerioak eta Suspertze, Eraldatze eta Erresilientzia Planak, Europar Batasuneko NextGenerationEUk ere finantzatu du, ILENIA proiektuarekin lankidetzan, 2022/TL22/00215335 erreferentzia duena. HiTZek errendimendu handiko konputazio azpiegitura (HPC) propioa erabili du, eta azken ereduak CINECAren Leonardo superordenagailuan entrenatu dira, EuroHPC Joint Undertaking barruan (EHPC-EXT-2023E01-013 proiektua).

partekatzea