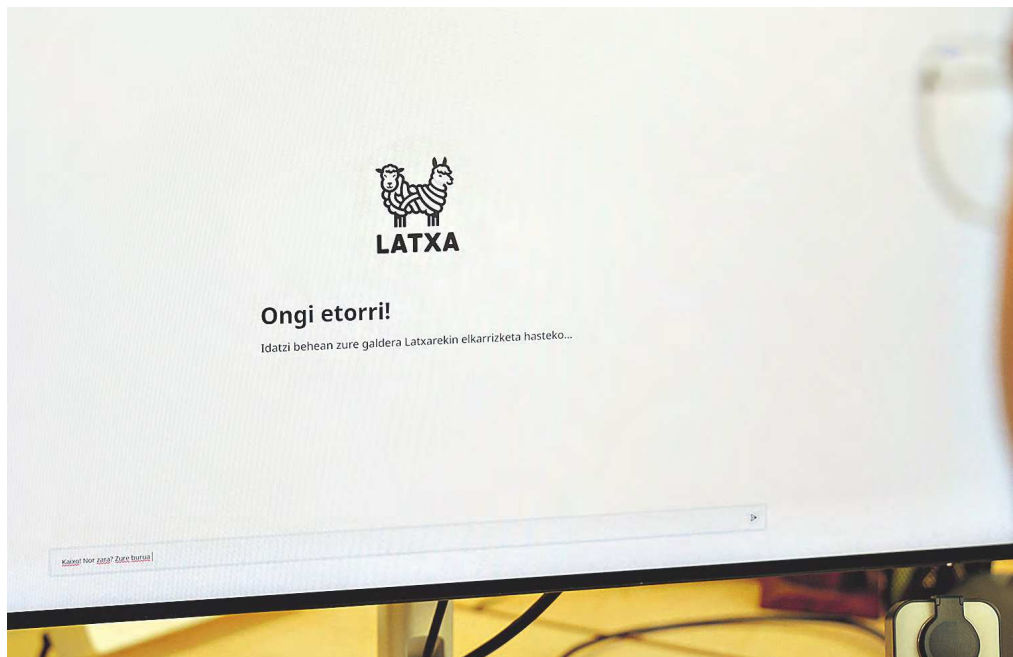




**ADIMEN
ARTIFIZIALA
EUSKARAZ.**
Latxa txatbota,
HiTZ zentroak
sortutakoa.
MAIALEN ANDRES
/ FOKU



«Kaixo, Latxa naiz. Zer jakin nahi duzu gaur?»

- Euskarazko txatbota sortu du EHUko HiTZ ikerketa zentroak.
- Oraindik ez dute jendaurrean zabaldu, baina garatzaileek eta enpresek eskuratzeko aukera dute.

Mikel Garcia Martikorena

DONOSTIA—«Kaixo! Ni Latxa naiz, adimen artifizial laguntzaile bat. Nire helburua da erabiltzaileei informazioa ematea eta haiekin elkarrekin aritzea, galderak erantzunez eta gai ezberdinei buruzko ezagutza partekatuz. Zer jakin nahi duzu gaur?». Hitz horiekin aurkezten du bere burua Latxak, EHUko HiTZ ikerketa zentroak sortu duen euskarazko txatbotak. Egunerokoa txertaturik dauden ChatGPT eta Gemini tresnen gisakoa da, baina baditu bi ezberdintasun nabarmen: alde batetik, ez dago enpresa multinazional baten esku; bestetik, entrenamendua euskarazko testuekin egin dute.

Testuak laburtu, haien ideia nagusiak zeintzuk diren adierazi, mezu bati erregistroa aldatu, Baionan asteburu-pasa egiteko plana egin eta bakailaoa pil-pilean kozinatzeak urratsak zehaztu. Latxak egin ditzakeen lanak askotarikoak dira; gaur-gaurkoz, ordea, ez dago jendearen esku. Dena dela, zenbait enpresa eta garatzaile erabiltzen hasi dira: astebetean, ehundik gora aritu dira. HiTZ zentroko kideek nabarmendu dutenez, erantzuna «ona» da. Horrenbestez, jendaurrean noiz eta nola zabaldu ari dira az-

tertzen, Eusko Jaurlaritzarekin batera. «Txatbot batek azpiegitura erraldoiak behar ditu ondo funtzionatzeko; beraz, haiek nola lortu ahal ditugun aztertzen ari gara», gaineratu du Eneko Agirre Bengoak, HiTZ ikerketa zentroko zuzendariak.

Latxa mugarri bat da informatikarako eta hizkuntza gutxituetarako. Izan ere, gai izan dira metodo berri bat garatu eta hizkuntza gutxitu batean adimen artifizialean oinarritutako txatbota sortzeko.

Agirrek azaldu duenez, iker-tzaileek egun dituzten galderetako bat da ea hizkuntza gutxituek duten testu kopuruarekin maila berera heldu daitekeen —ingelesezko testuak, esaterako, euskaraz baino mila aldiz gehiago dira Interneten—. Eta emaitzak «oso onak» direla adierazi du. Are gehiago, ChatGPT4.0 tresnak adinako ezagutza du Latxak, eta emaitza «paregabeak» ditu euskarazko probetan: EGA euskararen gaitasun agirik azterketak gaitzen ditu, adibidez.

EHUko ikertzaileen metodo berriak aurretik uste zenaren aurka egiten du. Izan ere, orain arte, testuak eskuz sartu behar ziren txatbotak entrenatzeko: milioika euro eskatzen dituen lan eredu bat. Latxa, baina, beste eredu ba-

ten gainean eraikita dago: *Llama* izeneko hizkuntza eredu irekia euskarazko testu masekin entrenatzen dute, ahanztura katastrofikoari —hau da, aurretik zekiena ahazteari— aurre eginez. Horrek bidea emango du multinazionalatik aldentzeko eta eredu ireki baten alde egiteko, ikertzaileen iritziz. Horrek euskara eta euskal kultura errespetatzeko modua emango duela iritzi dio Agirrek, eta egun «gero eta hedatuago» dauden tresnetan presentzia handiagoa izateko modua ere bai.

Proiektua aurrera eramateko, BERRIATEN testuak «ezinbestekoak» izan direla nabarmendu du Agirrek. Izan ere, askotariko gaien inguruan irakasteko aukera eman die egunkari honek. BERRIAKoez gain, *Hitzetako*, Bilbao Hiria Irratiko eta Euskarazko Wikipediako testuak erabili dituzte, besteak beste.

Erakutsitakoez gain, bestelako gaitasunak «bere kabuz» ikasi dituela adierazi du Naiara Perez Miguel lantaldeko kideak: testu bat erdaratik euskarara itzultzen, adibidez. Itzultzeko orduan, gainera, erregistroa aldatzeko gaitasuna ere baduela esan du Perezek. Etengabe probak egin behar zaizkiola gaineratu du. Ildo horretan, 1.600 pertsonak ebaluazio bat egin berri diote Latxari. Ebaluazio horren ondorio garbiena zera izan da, ikerketa zentroaren ustez: «HiTZek asmatutako metodoa da euskarazko txatbotak garatzeko metodorik hoberena».

Erabilerak eta erronkak

«Laguntzaile bat da, idazkari bat», esan du Oscar Sainz Gimenez HiTZ zentroko kideak, izango dituen erabileren inguruan galdetuta. Lanerako tresna bat dela gaineratu du, «gauzak azkarrago eta hobeto» egiteko. Hala ere, nabarmendu du ez duela ordezkatuko pertsonen lana.

Agirrek argi du zein den azken helburua, «utopikoa» dela onartu duen arren: «Wikipedia bezalako eredu erraldoi bat eratzea, zeinak kultura eta hizkuntza denak errespetatuko dituen».

Erronkak ere askotarikoak dira. Alde batetik, Latxa txatbotak hobetzeko dituen hainbat alderdi aipatu ditu Perezek: audioak transkribatzea, dokumentuak igo eta haiek ulertu ahal izatea, eta irudiak sortzea, besteak beste. Gainera, sarean era irekian dauden testu gehienak euskara batuan daudenez, euskalkiak ulertzen arazoak dituela nabarmendu du Agirrek. Perezek esan du ezen, bestelako adimen artifizialean bezala, gizartearen izan ditzakeen ondorioak eta erabilerak aztertzea behar direla. ●