

# NOLA ERABILI DA ADIMEN ARTIFIZIALA EUSKARA ZABALTZEKO?

Asko dira teknologia euskaratik lantzen ari diren ikerlariak; horien artean, Naiara Perez EHUko HiTZ Hizkuntza Teknologiako Zentroko ikerlaria, Itziar Cortes Elhuyarreko ikerlaria eta Eli Pombo lametzako kudeatzailea. Euskarak adimen artifizialean daukan lekuaz, aukerez, erronkez eta zailtasunez mintzatu dira, besteak beste.

 OLALIA L. GARAIALDE

**D**enen ahotan dago eta uste baino gehiago erabiltzen da adimen artifiziala egunerokoan. Baina zer da? “Definizio simple bat ematearren esanen nuke matematikaren eta informatikaren arteko arlo bat dela, zeinetan ordenagailuei gizakien gaitasunak ematen zaizkien; hala nola hizkuntza gaitasuna, ikusmena, mugimendua...”, azaldu du Naiara Perezek. Adimena zehatza edo orokorra izan daiteke. Hau da, zehatzak ekintza batzuetarako balio du; adibidez, tunel batera sartu eta autoko argiak automatikoki pizteko edota ahots laguntzaileei gauzak galdetzeko. Orokorra aldiz, gizakiaren pareko zerbait da eta ekintza bat baino gehiago egiteko gai da; horren adibide da, Open AI enpresaren ChatGPT solaskide artifiziala.

Sare neuronalen bidez funtzionatzen du adimen artifizialak. Konputazio tres-

nekin gizakien nerbio sistema eta ikas-teko modua imitatzen saiatzen dira sare neuronalak. Horretarako, informazioa eman eta ikasketa algoritmoak diseinatu tresnak trebatu behar dira patroiak ikas ditzaten. “Oraindik ez gara heldu, baina hasi gara benetan planteatzen ea datozen urteetan adimen artifizialak gizaki baten adimena ordezkatu dezakeen”, esan du Perezek.

Hala ere, ohar bat egin du Pombok, eta gogorarazi du adimen artifiziala tresna bat dela: “Ez nuke gure burua liluratzen utziko. Teknologia zentzuz, benetan baliagarriak diren erabileretarako eta modu etikoan erabili behar dugu”.

Adimen artifiziala sektore askotan aplikatu daiteke; horien artean hizkuntzaren prozesamenduan. Arlo horretan sartzen dira; besteak beste, hizkuntza idatzia ezagutu eta bat-bateko itzulpenak egiten dituzten sistemak; ahozko

hizkuntza jaso eta idazten dutenak eta testuak ahots bihurtzen dituzten sistemak. Perezen arabera, azken urteetan Euskal Herrian “asko” dira hizkuntza teknologian eta garapenean ikertzen ari direnak, eta sektore “indartsu bat” dago: “Guretik sortzen ari gara”. Hortaz, egon badago euskaraz ikertzen duen sektore bat, baina teknologiaren zurrumbilo horretan guztian zein da euskarak daukan lekua? Zein onura eta traba ditu? Zein zailtasun eta aukera ditu euskarak?

“Adimen artifiziala inguratzen gaituen errealitatea da, eta euskarak hor egon behar du. Bestela, aukerak galtzen ditu presentzia digitalean”, nabarmendu du Pombok. Gaineratu du euskara beste lekuetara zabaltzeko aukera bat izan daitekeela. Bat egin du horrekin Perezek, eta gogorarazi du proiektio handia duten tresnak, ChatGPT kasu, enpresa handien esku egon ohi direla: “Ezin gara







enpresa handiek eginen dutenaren zain gelditu. Google eta Microsoft bezalako enpresen lehentasuna ez da munduko hizkuntza guztiak kontuan hartzea. Horrelako enpresak bezero gehien izanen dituzten hizkuntzetan jartzen dute arreta; hau da, hizkuntza hegemonikoetan". Hala ere, horrelako enpresetan euskara ere txertatzen hasi direla dio.

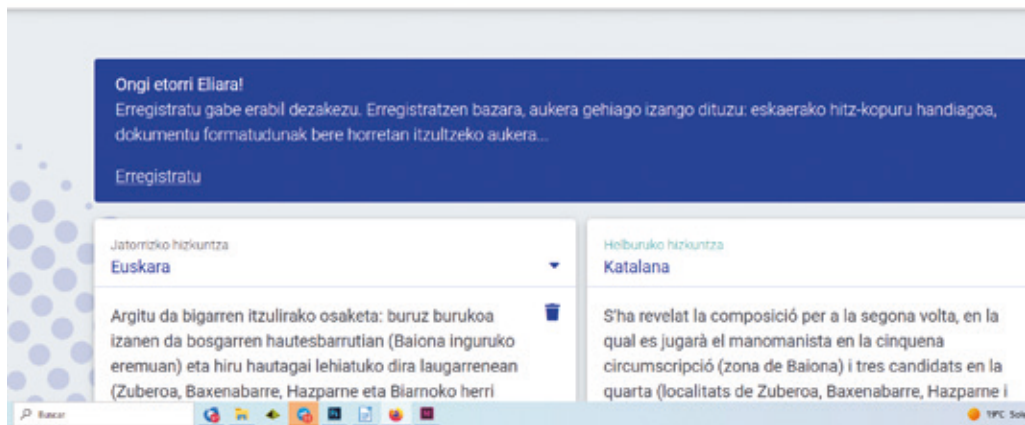
Zentzu horretan euskaraz lanean ari diren ikerlariak "borrokan" ari direla sentitzen du Pombok: "Trena galdu nahi ezean eta lekua egin nahian gabiltzala sentitzen dut, baina ez naiz ezkorra. Zailtasunak aparte, uste dut aldeko haizea ere badugula". Erakunde publikoen eta herritarren partetik gauzak egiteko "borondate handia" ere badagoela dio. Gainera, Pomboren arabera, enpresa handiek egiten dituzten aurrerapausoen kontra lehiatzea "ezinezkoa" da: "Gurean jarraitu behar dugu, behar diren gauzak zentzuz egiten eta frustratu gabe".

## BURUJABETZA TEKNOLOGIKOA

Teknologia, oro har, datuetan oinarritzen denez, Pomboren ustez, informazioa enpresa handien esku ez uzteko "modu burujabe" kudeatu behar da: "Software libreak aukera ematen dizu zuk egindakoa besteek hobetzeko, eta tokiko ekonomia indartzen du". Datu horien erabileraz mintzatu, eta "zentzuz" erabili direla ohartarazi du Cortesek. Horregatik, Elhuyarren ez dituzte bezeroen datuak erabiltzen tresnak entrenatzeko: "Adi egon behar gara, ez dugu letra txikia irakurtzen edo ez digute abisatzen eta nahi gabe sistema horiek elikatzen ari gara". Datuen jatorrian ere azpimarra egin du Perezek, eta eredu libreetan alderdi etikoa "garbiagoa" dela esan du.

Tokian tokiko beharretatik eta nahietatik lan egiteak duen inpaktua ere azpimarratu dute. "Guk egiten badugu, gure intereseko gaiak duten edukia sortuko dugu", esan du Perezek. Gaineratu du "punta-puntako teknologia" sortzea Euskal Herriko sektore teknologikoa eta ikerketa sektorea "elikatzeko" lagungarria dela: "Modu irekian lan egiten badugu hemengo ikerketa zentroen arteko elkarlana sustatu dezakegu".

Asko dira teknologia garatzeko tresnak besteen eskura jartzen dituztenak. Horien adibidea da EHUko HiTZ Hizkuntza Teknologiako Zentroak sortutako Latxa. Hizkuntza eredu handi bat da, eta horrelako ereduak testu segida bat ematen zaienean, hitz probableena



“Trena galdu nahi ezean eta lekua egin nahian gabiltzala sentitzen dut, baina ez naiz ezkorra. Zailtasunak aparte, uste dut aldeko haizea ere badugula”

**Eli Pombo, lametza**

ematen dute. "Laxak bere horretan ez du deus egiten, beste aplikazioak sortzeko motor bat da", azaldu du Perezek. Adibidez, zuzentzaile ortografikoak, galderak erantzuteko aplikazioak eta irakaskuntzan ariketa automatikoak sortzeko balio du: "Guk, adibidez, Latxa erabiltzaile moduan sartu dugu *Egunean Behin* jokoan erantzunak emateko". Sarean eskuragarri dago eta edonork deskargatu dezake.

## INFORMAZIO FALTA, ZAILTASUN

Datuak dira adimen artifizialaren altzorra. Izan ere, tresnak trebatzeko kalitatezko ahalik eta datu gehien behar dira.

Adibidez, ahotsa entzun eta zuzenean transkribatzen duen sistema bat sortzeko grabazio eta transkripzio asko behar dira. Elkarrizketatutako ikerlariak nabarmendu dute datu kopuru handiak lortzea dela euskarak daukan zailtasun handietako bat. "Hizkuntza hegemonikoekin alderatuta, euskarak ez dauka hainbeste eduki, eta orduan zailagoa da emaitzak lortzea", esan du Perezek. Bat egin du horrekin Cortesek; hala ere, uste du beste hizkuntza gutxitu batzuekin alderatuta euskaraz eduki "gehiago" dagoela.

Egungo tresna gehienak euskara batuan trebatuta daude; hala ere, badira euskalkiak lantzen dituzten tresnak ere. Batua itzulpen zerbitzua, esate baterako, adimen artifizialean eta sare neuronaletan oinarrituta dago eta Vicomtech teknologia zentroak egindako eta Euskaltelek, Mondragonlinguak eta EITBk sustatutako proiektua da. Itzulpen zerbitzu horrek euskara batua, frantsesa, gaztelania, ingelesa eta bizkaiera ezagutzen ditu. Horrez gain, euskararen eta hizkuntza horien guztien artean itzulpenak egiteko gai da. "Euskalkien kasuan askoz zailagoa da datuak lortzea; euskara batuan kostatzen bada, pentsa bizkaieran edo lapurteran", esan du Perezek. Euskalkie dagokienez, arau faltak eta zonalde bakoitzean dauden aldaerek prozesua zailtzen dute: "Ez dago bizkaiera baturik, orduan Getxoko, Gernikako edota Ondarroako euskalkiekin elikatzen badugu makina, oso zaila da pratoi bat sortzea".

Zailtasunak zailtasun, horrek ez du



Eskerrik asko, eta ez da ni zoriontzeko, baizik eta fakultateaz ari.

Elhuyarrek, besteak beste, Elia eta Aditu tresnak ditu. Elia itzultzaile neuronalak testu soilak eta dokumentu formatudunak itzultzen ditu segundo gutxitan. Aditu plataformak aurrez grabatutako audio eta bideo fitxategiak prozesatu eta haien transkripzioak eta azpigituluak sortzen ditu.

esan nahi Euskal Herrian kalitatezko edukia sortzen ez denik. Izan ere, Cortesek nabarmendu du “asko” zaintzen dela euskaraz sortzen den edukia: “Dauzkagun datuekin oso emaitza txukunak lortzen ari gara”. Gaineratu du adimen artifizialak beste ate batzuk ireki dizkiola euskarari, sistema zaharrek “mugatuta” zeudelako. Azaldu du lehenengo sistemekin ez zirela iristen “benetan baliagarriak” diren sistemak sortzera eta eskuragarri jartzera.

Praktikara eramanda itzulpen automatikoarekin egindako aurrerapausoak azaldu ditu: “Lehen ez zuen ongi funtzionatzen, baina orain bai”. 2007an Elhuyarrek eta EHUK Matxin –lehenengo itzultzaile automatiko librea– sortu zuten. Garai horretako sistema ez zegoen adimen artifizialean oinarritua: “Euskararen kasuan, horrek ematen zituen emaitzak ez ziren hurbiltzen gaur egun ditugun emaitzetara. 2016ra arte ez genuen lortu euskararen eta gaztelaniaren arteko itzulpenak egiteko kalitatezko sistema bat”. Egun, Elia itzultzaile automatikoa izenez ezaguna da.

Aldiz, beste hizkuntza batzuen kasuan, esate baterako, gaztelaniaren eta galegoaren arteko itzulpenak egiteko, gaur egun oraindik adimen artifizialean oinarritzen ez diren sistema zaharrak erabiltzen dira: “Gertuago dauden edo antzekotasunak dituzten hizkuntzetan oso emaitza onak lortzen ziren. Euska-

“Ezin gara enpresa handiek zer eginen duten zain gelditu. Google eta Microsoft bezalako enpresen lehentasuna ez da munduko hizkuntza guztiak kontuan hartzea”

**Naiara Perez, HiTZ zentroa**

raren kasuan deklinabideak daude, aditzak desberdinak dira eta hitzen ordena librea da. Horrek zailtasunak sortzen ditu hizkuntza batetik bestera pasatzeko erregelak sortzeko. Adimen artifizialari esker gaur egun ez dago ia mugarik”.

### **KALITATEA ZAINTEKO TRESNAK**

Teknologia geroz eta gauza gehiago

egiteko kapaza denez, “zentzuz” jokatu behar dela uste du Cortesek. Bestela, euskara zabaltzeko tresna izan beharrean, kalterako izan daitekeela dio: “Euskaraz sortutako edukia itzultzeko erabiltzen badugu, nolabait euskaraz sortutako edukiari zabalkundea ematen ari gara. Baina itzultzaile automatikoa erabiltzen badugu euskaraz ez dagoena euskarara pasatzeko eta ez badugu begiratzen itzulpena ongi dagoen, ez diogu mesederik egiten euskarari”. Gaineratu du horrek epe ertainera “kalitate gutxiko” testuak izatea eragingo duela: “Testu horiek erabiltzen baditugu etorkizuneko sistemak entrenatzeko, euskararen kalitatea baxua izango da”.

Sortzen den edukia berrikustez gain, zer egin daiteke kalitatea bermatzeko? HiTZ Hizkuntza Teknologiako Zentroak zenbait bide hartu ditu horretarako. Lehenik, Creative Commons lizentzia daukaten euskal hedabideetako edukiak erabili dituzte. Datu horiek nahikoak ez direnez, artxibo handietara ere jo dute eta bertako edukia filtro batetik pasa. “4.000.000 dokumentu lortu ditugu, baina ChatGPT bezalako tresna bat sortzeko ez da nahikoa; hala ere, emaitza onak lortu ditugu behintzat”.

Beste adibide bat Elhuyarrek dituen Elia, Aditu eta TTS tresnak dira. Adimen artifizialean oinarritzen dira hirurak. Elia itzultzaile neuronalak testu soilak eta dokumentu formatudunak itzultzen ditu segundo gutxitan. Aditu plataformak aurrez grabatutako audio eta bideo fitxategiak prozesatu eta haien transkripzioak eta azpigituluak sortzen ditu. TTS neuronalak testua ahots bihurtzen du. Elhuyarren teknologiek euskara, gaztelania, frantsesa, ingelesa, katalana eta galegoa ezagutzen dituzte. Horrek esan nahi du Elia, Aditu eta TTS sei hizkuntza horietan erabili daitezkeela. Elhuyarren azken berrikuntza da: “Ikusi dugu bezeroak eleaniztunak direla, eta gure teknologia erabili nahi dutela euskara ez den beste hizkuntzetan. Euskara ardatz izanik tresna bakarra hizkuntza gehiagotan erabili dezakete”.

Tresna horiek guztiak euskara “gehiago” zabaltzeko lagungarriak izan daitezkeela uste du Cortesek: “Ez dugu euskaraz sortzeko beldurrik izan behar. Jende gehiagorengana heldu nahi badugu, itzultzaile profesionalz gain, tresna ugari ditugu gaur egun”. Elhuyarren, adibidez, gehienetan euskara hutsean



Elhuyarrek zientzia.eus webgunean itzultzaile neuronal integratuta dauka. Oharra jarria du: "Euskaraz idatzitako testua eta Eliaren bidez automatikoki itzultitakoa, ondorengo gainbegiraturik gabe". Adibide honetan testua katanalez ageri da.

sortzen dituzte gauzak, baina Euskal Herritik kanporako gauza batzuk itzultzaile automatikoarekin itzultzen dituzte. Beti ere erabiltzaileari ohartaraziz eta jatorrizko bertsiara jotzeko aukera eskainiz.

Adibidez, zientzia.eus atarian itzultzaile automatikoa integratuta daukate. Horrek esan nahi du pertsona batek Interneten jartzen badu *Como xurdiu o sistema solar?* –Nola sortu zen eguzki-sistema? galegoz–, zientzia.eus atarira iritsi daitekeela. Webgunean, albo batean, beste hizkuntzetan ere irakurtzeko aukera eskaintzen du, baita jatorrizko bertsiara ikusteko ere. "Argi dugu irakurleak jakin behar duela itzultzaile automatikoak egindako testu bat dela, eta ez benetan pertsona batek egindakoa", esan du Cortesek.

ARGIAk duela gutxi Itzulinguru proiektuan hartu du parte, eta zientzia.eus webguneak bezala, itzultzaile automatikoa integratu du Soziolinguistika Klusterrak eta EHUKo Innoklab ikerketa taldeak proposatutako esperimentuan. Proiektu horretan ondorengoek ere hartu dute parte: Elhuyar, AEK, Orai

“Erabiltzen badugu itzultzaile automatikoa euskaraz ez dagoena euskarara pasatzeko eta ez badugu begiratzen itzulpena ongi dagoen, ez diogu mesederik egiten euskarari”

**Itziar Cortes, Elhuyar**

adimen artifizialeko zentroa, Osakidetza, Eusko Jaurlaritzako Hezkuntza Saila eta Hekimen.

Itzultzaile automatikoa erabili arren, ez dute dena tresnen esku uzten. Elhuyarreko kideek edukia birpasatzen

dute kalitatea bermatzearren. Batzuetan ikerlariak egiten dute lan hori, eta beste batzuetan itzultzaile profesionalak behar dira automatikoarekin ez duzulako lortzen kalitatearen %100. Gainera, ezin duzu edozeinen esku utzi testu bat jakiteko ona den edo ez, denok ez ditugulako irizpide berak”.

Gaur egun dauden tresna eta baliabide horiek guztiak hizkuntzak ikasteko lagungarriak izan daitezke. Halere, zer gertatzen da ikasteko gogoa piztu beharrean, hizkuntza ikasteko nahirik ez izatea eragiten badute? Puzen arabera, hizkuntza bat ikasteko nahia edo beharra komunikatzeko beharrari lotuta dagoen arren, ez da soilik horretara mugatzen: “Euskarari dagokionez, ez dut uste inork soilik komunikatzeko ikasten duenik. Hautu bat da eta beste gauza asko daude inguruan. Benetan, euskara, frantsesa, arabiera edo edozein hizkuntza ikasi nahi baduzu horrelako tresnek bidea erraztuko dizute, baina itzultzaile batek ez dizu emango euskaraz zuzenean irakurtzeak ematen duen plazera”. ●

BABESLEA: