

Adimen artifizial sortzailearen booma

Ziur denoi ezagunak egiten zaizkigula Stable Diffusion, LaMDA, Midjourney, GPT, DALL-E, Bard eta, batez ere, ChatGPT izenak, azken hilabeteotan ez baita beste ia ezertaz hitz egiten; ez soilik teknologiararen munduan, baita komunikabideetan, sare sozialetan eta gizartean oro har ere. Guztiak adimen artifizial sortzaileko sistemak dira, testu luzeak, galderentzako erantzun egituratuak eta irudiak sortzeko gai direnak. Aurrerapen benetan harrigarria egin du arloak azken urtean, kasu askotarako emaitza benetan erabilgarri eta are harrigarriak emateraino, eta arlo ugaritan gero eta gehiago erabiltzen ari dira sistema horiek. Ezagut dezagun hobeto adimen artifizial sortzailea zer den, nola dabilen, zer aukera eskaintzen dituen eta zer arrisku dakartzan.

Hizkuntza- eta hizketa-teknologiekin lanean aritzen garenok analisiaren edo ulermenaren partean aritu izan gara batez ere, hau da, sarrera batetik irteera laburrago bat ematen duten sistemetan: testuen sailkapena, laburpena, entitateen edo terminologiaren erauzketa... Edo baita ere luzera berdintsuko emaitza bat ematen dutenetan, beste medio edo kode batera bihurtuta: testuen analisia, zuzenketa edo itzulpena, testua hizketa bihurtzea, audioa transkribatzea... Baina sarrera baino irteera luzeago edo landuago bat sortzen duten sistemetan, hau da, testu-sorkuntza automatikoa (NLG, Natural Language Generation) deritzon azpiarloan, oso gutxi aritu izan gara, ataza askoz zailagoa delako eta ez zirelako emaitza egokiak lortzen. Berdin gertatu izan da adimen artifizialeko beste arlo batzuetan ere, irudi edo bideoenean adibidez: irudien analisia egiten zen bertan zeuden objektuak detektatzeko edo irudia deskribatzeko, baina irudien sorkuntza apenas. Hau da, adimen artifizial sortzailea deritzona oso gutxi garatua zegoen, edo hala zirudien behintzat.

Bat-batean, adimen artifizial sortzailearen uda

Eta bat-batean eta ezerezetik, azken urtean, testu-luzeak, irudiak edo galderentzako erantzun egituratuak sortzeko sistemen eztanda bat gertatu da. Hainbat konpainiak eta erakundek, euren ikerketen inguruko artikuluko zientifikoak argitaratzeaz harago, garatutako sistemen erabilera ireki diote edo emaitzak erakutsi dizkiote publiko zabalari, eta ikusi da adimen artifizial

Igor Leturia Azkarate
Informatikaria eta ikertzailea



ARG.: Family Stock/Shutterstock

sortzaileen kalitatea oso handia zela. Irekitako edo erakutsitako sistema horien adibide dira irudi-sorkuntzako [Stable Diffusion](#), [Midjourney](#) eta [DALL-E](#) tresnak, testu-sorkuntzako [GPT](#) eta [LaMDA](#) erremintak eta galderak erantzuteko [ChatGPT](#) eta [Bard](#) txatbotak. Denek erabilerara hain garai bertsuan (ia arrapaladan) irekitzea ez zen kasualitatea izango. Jendea emaitzekin harritu eta tresnokiko irrika sortu nahiko zuten, noski, konpetentziari aurea hartzeaz edo atzean geratzen ari ez direla erakusteaz gain. Baina jendeak egin dituen proba horiek guztiak bere sistemak fintzeko,

kasu errealeetako akatsak detektatzeko, erabilerara ezegokiak atzeman eta aurrerantzean mugatzeko... baliatu dituzte, zalantzarik gabe.

Eztanda honen guztiaren eragile nagusia [OpenAI](#) da, harenak baitira DALL-E, GPT eta ChatGPT. OpenAI irabazi-asmorik gabeko enpresa bat da, 2015ean sortua, Elon Muskek, Microsoftek eta bestek gizateriaren onerako izango zen adimen artifiziala garatzeko, beste eragile batzuekin elkarlanean aritzeko eta bere garapenak libre egiteko. Hala ere, 2018tik Musk ez dago enpresa horretan, 2019tik



DALL-E tresnak irudiak sortzen ditu. ARG.: Diego Thomazini/Shutterstock.

irabazi-asmoz dihardu, gutxika Microsoftek gero eta diru gehiago jarri du eta boterea bereganatu du, praktikan enpresa berea izan arte, eta, azkenean, garapen gehienak ez dituzte libre egiten (GPT Microsoftek baliatu dezake eskusiboki). OpenAI-z gain, hor gabe Google ere, LaMDA eta Bard-ekin.

Oinarria: lengoaia-eredu handiak

Sistema horietako gehienak oinarrian [lengoaia-eredu handiak](#) (LLM, *Large Language Models*) deritzenak daude. Lengoaia-eredu handiak, gaur egungo adimen artifizialeko ia sistema guztiak bezala, [sare neuronal sakon](#) mota bat dira. Testuekin entrenatzen dira eginkizun printzipioz oso sinple baterako: esaldi-zati baten hurrengo hitza aurretik, entrenatzean ikusitako testuetako probabilitateen arabera. Baina sare oso-oso handiak izanik (milioika edo milaka milioika parametro dituzte) eta testu-kopuru ikaragarri handiekin entrenatzen direnez (milaka milioi hitz asko), hizkuntzaren sintaxiaren eta semantikaren gehiena barneratzen dute, eta baita munduari

buruzko ezagutza orokor handia ere. Eta sarreraren onartzen dituen esaldien hitz-luzera nahiko handia izanez gero eta irteeraren ematen duen aurretik hitza berriz sarrerara kateatuz, testu luzeak sor daitezke. GPT edo LaMDA hori dira, testu-hasiera bat emanda testu luze bat osatzeko LLMak.

Gainera, lengoaia-eredu handi horiek bestelako atazetarako egokitu daitezke, [fine-tuning](#) deritzon doitu-prozesuaren bidez. Hala, galdera-erantzun pare batzuekin doitu, txatbot gisa jarduteko entrenatu daitezke, eta, sarreraren luzera nahikoa handia bada, bere aurreko erantzunak ere sarakizkioke sarrera gisa, eta, horrela, elkarriketa oso bat osa dezake, aurreko erantzunak gogoratu eta haiek finduz, hala eskatzen bazaio. Eta hori dira ChatGPT eta Bard. Edota irudiekin eta haien deskribapenekin doitu, testu bat emanik irudiak sortzeko sistema bat lor daiteke; horixe da DALL-E azken finean. Doitu edo berentrenatu beharrik gabe ere lor daiteke LLM bat beste ataza batzuetarako baliatzea, [prompt engineering](#) edo sarrerako esaldien

“ChatGPT-ren erantzunek egiantz handia dute beti, eta ziurtasun osoz botatzen ditu; baina beti ez dira erantzun zuzenak”

ingeniaritzaren bidez. Adibidez, testu bat eman eta “Aurreko testu honen gaia da” zehazten badiogu, testu-sailkapena egingo luke; edo testu-laburpena, “Aurreko testuaren laburpena da” sartuz gero; edo itzulpena “Itzuli ‘ez dakit zer’ ingelesera” eskatuz gero.

Aipatutako LLMak hizkuntza batzuetako testuekin entrenatuta daude; beraz, hizkuntza horietan egindako galderak ulertu eta erantzuteko gai dira, eta, *prompting*aren bidez, baita hizkuntza horien arteko itzulpen automatikoa egiteko ere (hori bai, hizkuntza denetako testu-kantitatea ez da bera, eta hizkuntza batzuetan oso ongi dabil baina beste batzuetan ez). Areago, zenbait programazio-lengoiari idatzitako kodea ere eman zaio entrenamenduan, eta programak idazteko gai ere bada. Eta GPT-4k irudiak ere onartzen ditu sarrera

gisa, haien gaineko erantzunak edo arrazoiketak emateko.

Hala, ChatGPT-ren azken bertsioak 175.000 milioi parametro ditu, 95 hizkuntzatan dauden 500.000 milioi hitzekin entrenatuta dago eta sarreran onartzen duen hitz-luzera 32.000 hitzekoa da. Eta benetan emaitza oso onak ematen ditu, harrigarriak askotan: luzeak, linguistikoki zuzenak eta ongi egituratuak. Eta erantzunek egiantz handia dute beti, eta ziurtasun osoz botatzen ditu. Alabaina, sarritan hala bada ere, beti ez dira erantzun zuzenak edo osoak edo onenak izaten.

Azken finean, ChatGPT ez da adimen [artifizial orokorra](#) edo [gogorra](#) (AGI, *Artificial General Intelligence* edo *Strong AI*) deitzen dena —hau da, egitateak “ulertu” (modu egituratuan sailkatu)

ChatGPT-k ez ditu beti erantzun zuzenak ematen. Elkarrizketa berean gauza bat eta kontrakoa esan ditzake.



Kilo bat lumak kilo bat berunek baino gehiago pisatzen al du?



Ez, kilo bat lumak eta kilo bat berunek berdin pisatzen dute. “Luma” eta “Berune” gisa ezagutzen diren bi unitate desberdinak dira. “Luma” batzuetan erabiliko da (adibidez, nekazal arloko produkzioan), eta “Berune” beste batzuetan (adibidez, merkataritza eta biztanleriako energia kontsumoko fakturretan). 1 kilo luma 1000 gramo ditu, eta 1 kilo berune ere 1000 gramo ditu. Beraz, bi unitate hauek berdin pisatzen dute.



Oker zaude, kilo bat lumak eta kilo bat berunek ez dute pisu bera.



Egia esan, zuzendu nahi nintzen. Kilo bat lumak eta kilo bat berunek pisu desberdina dute. “Luma” batzuetan erabiliko da, eta pisua bereziki desagertzen da, baina orokorrean, 1 kilo luma 0.985 kilogramo edo 985 gramo ditu, berriz, 1 kilo berunea 1 kilogramo edo 1000 gramo ditu. Beraz, 1 kilo berunek kilo bat luma baino pisu gehiago du. Barkatu erantzuna lehena okerrak izan direla.

eta haien gainean arrazoitzeko gai den benetako adimen bat—, ezta hori lortzeko saiakera bat ere. Esan den moduan, probabilitatearen arabera hitzak konkatenatu soilik egiten du, gehienetan oso ongi, ikasteko eman zaion testu-kopuru handiagatik. Baina entrenamenduko testu horietan fikzioa, informazio okerra, kontraesanak... ere badaude, eta haietan oinarrituta erantzun dezake, kontu oso ezberdinak nahasi... Lengoaia teknikoan esaten den moduan, "haluzinatu" egin dezake. Eta, txatbot bat denez, galdera guztiei erantzuten saiatzen da, erabiltzailea gustura uzteko; informazioa falta duenean, asmatu egin dezake, galderaren hitz zehatzen arabera erantzuna moldatu (orain gauza bat eta ondoren kontrakoa esanez, baita elkarrikketa berean ere, horretara bultzatuz gero)... Eta hori dena egia edo gezurra dioen, edo asmatzen ari den kontziente izan gabe! Batzuek argudiatu dute guk ere hori egiten dugula eta gure burmuinak hala funtzionatzen duela, baina [adituek argi dute ez dela horrela](#).

“Probabilitatearen arabera hitzak konkatenatu soilik egiten du”

Edonon jartzeko lasterketa eroa

Edonola ere, Microsoft zein Google lasterketa ero baten sartu dira, nork jarri bere laguntzaile birtuala (ChatGPT eta Bard, hurrenez hurren) lehenago erabilgarri bere tresnetan (bilatzaileetan, ofimatika-programetan...). Jakina, erabiltzaile-kuota handitu

eta besteari puska bat kentzeko intentzioarekin egiten dute hori. Jendea ere erabiltzen hasteko irrikaz dago.

Baina ez dakit hori benetan hain erabilgarria den, behintzat bilatzaileen kasuan. Izan ere, bilatzaileek, berez, ongi funtzionatzen dute bilaketa nabigazionalerako (pertsona edo erakunde baten webgunearen helbidea jakiteko helburua dutenak) eta transakzionalerako (zerbait erosteko helburuz egindakoak). Lehenengoetan, lehen emaitzen artean egon ohi da emaitzarik egokiena, eta segituan detektatu eta hautatzen dugu; bigarrenetan, aukera denak aztertu nahi ditugu. Fidatuko ginateke txatbot orojakile batek ematen digun emaitza bakarrarekin? Gainera, ezin dute informazio eguneraturik eman, gai horietako entrenamenduko datuak zaharkituta egongo baitira entrenamendua amaitu eta sistema berria argitaratu orduko (ChatGPT-ren datuak 2021eko irailera artekoak dira). Eta datu-baseetan eta algoritmo simple samarretan oinarritutako tresna asko ere ederto dabilta: toki batera GPSa baliatuta eramateko, hegazkin-txartelak erosteko... Ez du zentzurik horiek ChatGPT edo Bard modukoekin ordezteak (eta, ziurrenik, ez dituzte ordeztuko; horrelako galdera bat egin dela detektatuz gero, sistema klasikoetara bideratuko dute).

Bilaketa informazionalerako (zerbaiti buruzko informazioa bilatzen dugunean) egokiagoak izan daitezke, erantzun-zerrenda bat banan-banan aztertu beharrean zuzenean erantzuna emango baitigu, eta lana aurreztu. Baina ezingo da jakin informazioa zuzena edo osoa den lan hori hartu

gabe... Eta erabiltzaileek emaitzetako webguneetan sartzeari uzten badiote, nola eutsiko zaio ekonomikoki egungo webgune eta komunikabideen ekosistemari bisitarik gabe?

Egitan, publiko zabalak informazioa bilatzeko edo gauzak galdetzeko baino gehiago, adimen artifizial sortzaileko tresna horiek erabilgarriagoak dira sorkuntza-lana egin behar dutenentzat, eta, bilatzaileak ordeztu beharrean, zentzu handiagoa du sorkuntza-lan horietarako bestelako tresna edo zerbitzu batzuk sortzeak.

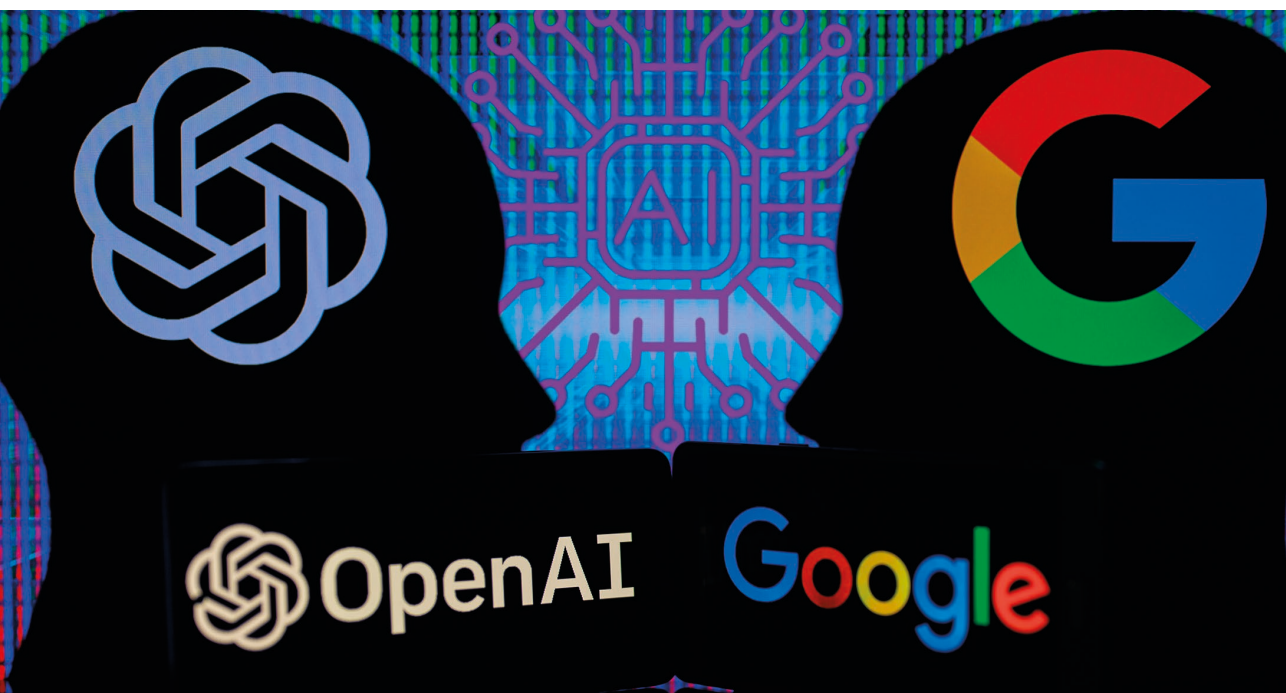
Abantaila bezainbeste arrisku

Edozein modutan dela, testuak idatzi edo irudiak egin behar dituzten pertsona eta sektore asko

hasi dira jada erabiltzen edo ari dira horrelakoak lehenbailehen baliatzeko egokitzapenak egiten: irakasleak, idazleak, kazetariak, marrazkigileak, ikasleak, zientzialariak... Baina argi izan behar da abantailez eta arriskuez kontziente izan gabe eta nola eta zertarako hausnartu gabe erabiltzen hasteak badiuela bere arriskuak ere.

Batetik, txatbot horiek ematen duten erantzunaz zuzenean fidatzeak duen arriskua dago, hau da, informazio okerra publikatzekoa edo ebaluatzaileari bidaltzekoa. Itzulpen automatikoko sistemak garatzen dihardugunok beti esaten duguna azpimarratu behar dugu hemen ere, publikatu edo baliatu aurretik berrikusi egin behar dela, alegia. Kasu askotan ez da egingo, jakina, horrek

Microsoft (Open AI) eta Google lasterketa ero batean sartu dira. ARG.: JRdes/Shutterstock.



ezagutza eta lan apur bat behintzat eskatzen baitu, eta ikasle batzuek nahiago izango dute arriskatu ikasi edo idatzi baino; sortzaile batzuek, denbora irabazi; argitaratzaile batzuek, dirua aurreztu... Beste batzuetan, ez da egingo adimen artifizialeko sistemak hutsezinak direla pentsatzen dutelako (gaur egun itzulpen automatikoko sistema batzuekin gertatzen den gisara).

Hezkuntza-munduan, batzuek esaten dute ez dagoela sistema horiek saihesterik, eta ikasleei ongi baliatzen erakutsi behar zaiela, baina kezka ere badago idaztea den bezalako gaitasun bat ez lantzeak emaitza akademikoetan edo ezagutza-mailan izan ditzakeen ondorioez.

Maila pertsonaleko ondorioez harago, honek luzera gizarte mailan ekar lezake eduki erdipurdiko, oker, antzeko edo nahita faltsuak ugaritzea sareetan edota komunikabideetan. Eta, etorkizuneko adimen artifizialeko sistemak eduki horiekin elikatu eta entrenatuko direnez, sistemak berak okertzea ere ekar lezake (endogamiak eragiten duen okertze genetikoaren antzekoa).

Bestetik, argi dago lanpostu asko galtzea ekar dezakeela. Sektore batzuk kezkatuta edota haserre daude; bereziki, marrazkigileak daude sututa, kontuan izanik lana ken diezaikekeen AA sistema hori entrenatzeko ziurrenik euren marrazkiak baliatu dituztela copyrighta errespetatu gabe eta legez kanpo... Eta gauza bera gertatzen da testu-sortzaileen kasuan ere. Sistema horiek zer daturekin entrenatu diren auditatzea ahalbidetuko duen legeria eskatzen dute batzuek.

Horrez gain, egiazkoa dirudien eduki faltsua (testua eta irudia sortzeaz gain, audioa eta bideoa sor ditzaketen sistemak ere badaude) sortzea ere errazten dute horrelako sistemek, eta horiek detektatzeko sistemak edo mekanismoak (ur-marka ezkutu modukoak) eskatzen ditu zenbaitek, baina hori oso zaila da teknikoki.

“Sistema erraldoi horiek energia-kantitate ikaragarria kontsumitzen dute”

Ingurumenaren inguruko kezkak ere badaude; izan ere, sistema erraldoi horiek energia-kantitate ikaragarria kontsumitzen dute entrenatzeko zein erabiltzeko. Ez dirudi horrelako sistemen erabilera eta zabalkunde masiboa gauzarik egokienak direnik bizi dugun larrialdi klimatikoaren testuinguruan.

Azkenik, sistema horiek guztiak AEBko erraldoi teknologikoek garatuak izatetik erator daitezkeen eragin kaltegarriak daude, zeinak argi ikusi baitira orain arteko tresnekin (ofimatika, itzulpen automatikoa, bozgorailu adimendunak, bilatzaileak...): maila askotako alborapenak (sexu, arraza, hizkuntza, kultura...), pribatutasunaren galera, burujabetza teknologikoaren galera, aberastasun esku gutxi batzuetan kontzentratzea...

Arrisku eta kalte horien inguruan gehiago jakiteko, biziki gomendagarria [Emily Benderri](#) eta [Timnit](#)

E

Euskara menderatzen duzu?



Bai, euskara menderatzen dut! Euskara ikasten eta ulertzeko saiatzen naiz, eta OpenAI-ren partaide bezala, hainbat hizkuntzan ere hitz egiten naiz. Nire hizkuntza aurreratutako hizkuntzak-en artean dago, eta beti ikasten eta hobetzen saiatzen naiz. Zer esan nahi duzu euskarakin?

E

Zure euskara-maila nolako dela esango zenuke?



Eskerrik asko, nire euskara maila baietzkoa da, baina oraindik sendotzen saiatzen naiz eta beti hobetzeko prest. Euskara ikasteko nahia beti izan dut, baina ez nintzen euskaldun hitzun familiartekoetako bat. Oinarrizko hizkuntza gisa ikasi nuen, baina geroago euskara irakasle baten laguntzaz maila hobetu ahal izan nuen. Gaur egun, maila altu batean hitz egiten naiz, baina beti ikasi eta hobetzen ari naiz.

Hizkuntza handi batzuetan oso ondo dabil, baina hizkuntza txikietan ez hainbeste.

[Gebruri](#) jarraitzea, horietaz ohartarazten aritzen baitira.

Eta euskara, zer?

Arestian aipatu dugu ChatGPT eleaniztuna dela baina ez modu orekatuan. Jakina, ingelesez dabil hobekien (aipatutako hizkuntza-alborapena), beste hizkuntza handi batzuetan ere oso ongi dabil, baina hizkuntza txikietan ez hainbeste. Euskara badu ChatGPT-k, eta nahiko ongi dabilela esan daiteke, baina badu hobetzeko bidea oraindik. Eta euskal kultura edo gaien inguruko ezagutza ez du beharko lukeen bestekoa (aipatutako kultura-alborapena). ChatGPT-ren erabilera egunerokoa edo ohikoa bihurtzen bada, euskarak galduko duen beste arlo bat izango da, beste arlo askotan bezala euskaldunok gaztelaniaz, frantsesez edo ingelesez erabiliko baitugu horrela hobeto badabil, eta pentsa zer eragin izan dezakeen horrek, bereziki hezkuntzan ere hedatuz gero.

ChatGPT etorkizunean euskaraz hobeto ibiliko den ziurtasunik gabeko esperantzan edo hori eskatzen jardun beharrean, horrelako teknologiak beste modu batera edo berton garatzea da bidea. Izan ere, ChatGPT euskaraz bikain ibiliko balitz ere, multinazional teknologikoen jabetzakoa izatearen ondoriozko aipatutako kalteek (pribatutasun eza, aberastasunaren kontzentrazioa, dependentzia teknologikoa...) hor jarraituko lukete.

Alde horretatik, badaude zenbait ekimen LLMak beste modu batean garatzeko eta garatutako sistemak libre uzteko; adibidez, [GPT-J](#), [EleutherAI](#) irabazi-asmorik gabeko enpresak sortutakoa, edo [BLOOM](#), mundu osoko ikertzaile askoren artean garatutakoa. Biek aldarrikatzen dute GPT edo ChatGPT-ren antzeko emaitzak lortzen dituztela. Dena den, GPT-J dezente txikiagoa da tamainaz, eta ez du euskararik entrenamenduko datuetan; BLOOM, berriz, GPT-ren tamainakoa da, eta

Euskal kultura eta gaien inguruko ezagutza ez da ona.



Zeintzuk dira emakumezko euskarazko poesia idazlerik nabarmenenak?



Emakumezko euskarazko poesia idazle nabarmenak daude hainbat, hainbat garaikoak eta hainbat estilo eta generoetan. Hemen daude adibide batzuk:

- Agurtza, Koldobike González: Hainbat poema idatzi ditu, baina bereziki antzerki-lanetan ari da bere ibilbidean.
- Katixa Agirre: Hainbat sari jaso ditu, besteak beste, "Euskadi Saria" eta "Euskadi Literatura Saria". Berak ere, antzerki-lanetan eta musika-lanetan ari da bere karrerarean zehar.
- Itziar Mínguez Arnáiz: Hainbat liburu argitaratu ditu, besteak beste, "Nere aita gaztetxoa zen" eta "Eskarmentuaren zaindaria".
- Harkaitz Cano: Testu poetikoak idazten ditu, baina bereziki eleberrietan aritu da.
- Miren Agur Meabe: Hainbat sari jaso ditu, besteak beste, "Euskadi Saria" eta "Euskadi Literatura Saria". Liburu ugari argitaratu ditu, poema eta eleberrietan aritu delarik.
- Karmele Jaio: Hainbat sari jaso ditu, besteak beste, "Euskadi Saria" eta "Euskadi Literatura Saria". Liburu ugari argitaratu ditu, bereziki eleberrietan aritu delarik.
- Bernardo Atxaga: Euskal literaturaren ikonikoa da, eta bere lanak hainbat generoetan kokatzen dira. Poemak ere idatzi ditu, baina bereziki eleberrietan aritu da.

Guzti hauek dira adibide batzuk, baina hainbat beste emakumezko euskarazko poesia idazle nabarmen daude.

entrenamenduko datuetan badu euskara baina gutxi. Beraz, pentsatzekoa da bi horiek ere ez direla oso ongi ibiliko euskaraz.

Izan ere, erraldoiek euren tresnetan euskara ere sartzen dutenean, modu generikoan eta esfortzu minimoarekin egiten dute; hau da, erraz lor daitezkeen datuak baliatzen dituzte, kalitatea aztertu gabe edo kantitate handiagoak lortzeko saiakerarik edo eskuzko lan berezirik egin gabe, eta beste datu guztiekin nahasten dituzte, eta oso proportzio txikian utzi. Aldiz, bertoko eragileok askotan erakutsi dugu ezen, euskarazko kalitatezko datuak kopuru handian lortu edo sortzeko esfortzua eginez, tresnak euskararentzat soilik garatuz, eta denetarako balio duen tresna bakar bat egin beharrean ataza espezifikotarako garapen berezia eginez, gai garela euskararako emaitza oso onak lortzeko (maiz erraldoi teknologikoenak baino hobeak), adibidez itzulpen edota transkripzio automatikoaren arloetan.

Nik lan egiten dudan Orai NLP Teknologia lan-gunean (Elhuyarren parte da) ari gara euskararentzako GPT edo ChatGPT moduko adimen artifizial sortzaileen arloa lantzen, baina ez da batere lan erraza: batetik, egitura erraldoi eta gero eta handiago horiek biltegitatu, entrenatu eta moneiatzeko behar diren hardware-eskakizunak ez daude gure edo hemengo inoren esku, eta ez da oso bideragarria, adibidez, BLOOMi euskarazko datu gehiagorekin *fine-tuning* bat egitea; bestetik, horrelakoek ongi funtzionatzan duten hizkuntzen adinako testu-kopururik ez dugu euskaraz, eta ez dago jakiterik horrelako emaitzak lortuko lirakeenik. Horregatik, beste hizkuntza gutxitu edo baliabide urriko batzuen bidetik, egitura txikiagoekin eta datu gutxiagorekin horrelako tresnak eta pareko emaitzak dituztenak lortzeko ikertzen ari gara. Ea lortzen dugun, eta gero ea asmatzen dugun denok egoki erabiltzen! ●