

ILENIA

IMPULSO DE LAS LENGUAS
EN LA INTELIGENCIA ARTIFICIAL

Hizkuntza Teknologien inpaktu handiko erabilera-kasuen azterketa

©2024 Hitz Zentroa

Aitortu-PartekatuBerdin 4.0 Nazioartekoa (CC-BY-SA 4.0)

<https://creativecommons.org/licenses/by-sa/4.0/legalcode.eu>

HiTZ Zentroa

EHU/UPV Informatika Fakultatea

Manuel Lardizabal Pasealekua, 1

20018 Donostia

hitz@ehu.eus

Egileak:

Hitz Zentroa, Langune eta Emun

Sustatzaileak:

Hitz Zentroa

1	Proiektuaren jatorria eta helburuak	1
2	Kontzeptualizazioa	5
2.1	<i>Adimen Artifiziala</i>	5
2.2	<i>Hizkuntza Teknologiak</i>	5
3	Diagnostikoa	7
3.1	<i>Proiektuaren faseak</i>	7
3.2	<i>Metodologia</i>	8
3.3	<i>Entzute fasearen emaitzak</i>	11
4	Irizpideen proposamena	19
4.1	<i>Irizpideen sailkapena eta ulermena</i>	19
4.2	<i>Irizpideen ponderazio-taulak</i>	20
5	Eragin handiko erakusleak	23
5.1	<i>Erakusleen fitxa teknikoak</i>	23
6	Ondorioak	43
	Eranskinak	47

Abreviaciones

AA	Adimen Artifiziala
LLM	Hizkuntza Eredu Handia (<i>Large Language Model</i> ingelesez)
HP	Hizkuntzaren Prozesamendua
STT	Ahotsaren ezagupena (<i>speech-to-text</i> ingelesez)
IA	Itzulpen Automatikoa
HT	Hizkuntza Teknologia
TTS	Ahots sintesia (<i>text-to-speech</i> ingelesez)
TRL	Teknologiaren heldutasun maila (<i>Technology Readyness Level</i> ingelesez)

Teknologiaren heldutasun mailak

- 1 Oinarrizko printzipioak hauteman eta jaso dira
- 2 Kontzeptu edota aplikazio teknologikoa adierazi da
- 3 Funtzio kritiko analitikoa eta esperimentala edota kontzeptu-proba bereizgarria
- 4 Osagaiak baliozkotu dira edota osagai horiek laborategiko ingurunean jarri dira
- 5 Osagaiak baliozkotu dira edota osagai horiek ingurune adierazgarri batean jarri dira
- 6 Sistema- edo azpisistema-eredua edo prototipoa ingurune adierazgarri batean probatu da
- 7 Sistema edo prototipoa ingurune erreal batean probatu da
- 8 Sistema osatua eta ziurtatu da, proba eta demostrazioen bidez
- 9 Ingurune errealean arrakastaz probaturiko sistema da

PROIEKTUAREN JATORRIA ETA HELBURUAK

Azken urte hauetan, ikasketa sakonak iraultza ekarri du Hizkuntzen Teknologien (HT) eta Adimen Artifizialaren (AA) alorretara. Teknologia horiek hizkuntza naturalaren prozesamenduan, itzulpen automatikoan, ahotsaren sintesian eta ezagutzan aplikatu daitezke, besteak beste, eta errotik aldatzen ari dira komunikatzeko, informazioa eskuratzeko eta gure ingurune sozial eta kulturalarekin elkarreragiteko modua.

ILENIA proiektu koordinatua¹ Hizkuntza Ekonomia Berriaren (HEB) Ekonomia Suspertzeko eta Eraldaketarako Proiektu Estrategikoaren (ESEPE) parte da. Proiektuaren helburu nagusia da Espainiako hizkuntza ofizialek adimen artifizialaren aplikazio aurreratuenetan txertatzeko beharrezkoak diren hizkuntza-baliabideak garatzea, hala nola corpusak eta eredu konputazionalak.



¹<https://www.proyectoilenia.es>

Hizkuntzen kudeaketa desegoki batek gure gizartearen hizkuntza-aniztasuna ikusezin bihur lezake. Testuinguru horretan, **ILENIA**k baliabide eta aplikazio eleaniztunen garapena du helburu. Berziki, testu-, ahots- eta itzulpen automatikoaren eredu eleaniztunak garatzea du xede, gizartearen beharrei erantzuteko eta gaur egungo teknologiaren eskakizunekin bat egiteko. Teknologia horretan, eleaniztasunak eta hizkuntzek funtsezko eginkizuna betetzen dute.

ILENIA modu koordinatuan garatzen ari dira hizkuntzaren teknologiaren arloko jardueretan puntakoak diren lau ikerketa-zentrok: Barcelona Supercomputer Center - Centro Nacional de Supercomputación (**BSC-CNS**²; AINA proiektua³), Hizkuntza Teknologia Zentroa (**HiTZ**⁴; IKER-GAITU proiektua⁵), Santiago de Compostelako Unibertsitatea (**USC**⁶; NÓS proiektua⁷) eta Centro de Inteligencia Digital de Alicante (**CENID**⁸; VIVES proiektua⁹).

Horrela, **ILENIA** proiektu koordinatuak gure gizartea beharrezko hizkuntza-azpiegituraz hornitzeko aukera paregabea eskaintzen du, gure hizkuntza ofizialek egungo eta etorkizuneko AAKo aplikazio aurreratuenetan txertatzea bermatzeko. **ILENIA**ren emaitzak honako puntu hauetan laburbil daitezke:



- 1.** Hizkuntzan oinarritutako AAKen sektorea bultzatzea, haren lehiakortasuna hobetzeko European eta mundu mailan.
- 2.** Gure hizkuntza ofizialak edozein AAKo aplikaziotan txertatzea errentagarri eta erakargarri egitea, bai multinazional handientzat, bai enpresa txiki eta ertainentzat.
- 3.** Mundu digitalean gure hizkuntza ofiziale-tako edozein erabiliz parte hartu ahal izatea, hizkuntza globalekin pareko mailan.
- 4.** Hizkuntzan zentratutako AAKen ikerketa behar bezala prestatzea, ezagutza oinarritutako gizarte digitalaren erronkei erantzuteko gai izan dadin.

²<https://www.bsc.es>

³<https://www.projecteaina.cat>

⁴<https://www.hitzeus>

⁵<https://www.hitzeus/iker-gaitu>

⁶<https://www.usc.gal>

⁷<https://www.nos.gal>

⁸<https://www.cenid.es>

⁹<https://www.vives.gplsi.es>

Txosten hau **ILENIA** proiektuko HiTZen **5.1 Entregagaia** da.

Txosten honen helburua da **inpaktu handiko erabilera-kasuak identifikatzea**, eta Euskadiko HT/AA arloko eragile nagusien aholkularitzarekin eta **ILENIA** proiektuaren helburuekin zuzenean lotutako industria-sektorearen interesarekin bat egitea. Zehazki, hizkuntzaren industriaren eragin handiko erabilera-kasu eta aplikazio berritzaileak identifikatu, hautatu eta garatzea proposatzen da, betiere **ILENIA** barruan garatu edo hobetu nahi diren teknologietan oinarrituta.

HiTZ zentroak **LANGUNE**¹⁰ eta **EMUN**¹¹ erakundeekin lankidetzan egindako proposamenetik abiatuta, eragin handiko balizko demostratzaileen hasierako zerrenda bat identifikatu ahal izan da, eta haien eragin ekonomikoa eta soziala neurtzeko adierazleen multzo bat definitu da.

¹⁰ <https://www.langune.eus>

¹¹ <https://www.emun.eus>

KONTZEPTUA- LIZAZIOA

Adimen Artifiziala

Adimen Artifiziala (AA) informatikaren adar bat da, zeinak tradizionalki giza adimena behar duten zereginak —esaterako, patroiak ezagutzea, erabakiak hartzea, hizkuntza ulertzea eta esperientziatik ikastea— burutzeko gai diren sistemak sortzea duen helburu. Sistema horiek datuak prozesatzen dituzten algoritmoetan oinarritzen dira erabakiak hartzeko. Gaur egun, eskuragarri dauden datu gehienak egituratu gabeko informazioz osatuta daude, idatzizko edo ahozko dokumentuetan, eta hainbat hizkuntzatan.

Horrenbestez, hizkuntza eta bere prozesamendu automatikoa ezinbestekoak dira AAren garapenean. Aldi berean, AA **Hizkuntza Teknologien** etengabeko garapen eta hobekuntza ahalbidetzen duen euskarri teknologikoa da. Biak batera, gizakien eta makinek arteko komunikazioa errazten duten hainbat aplikazio modernoren oinarria dira: itzulpen automatikoarena, bilaketa-motorrena eta laguntzaile birtualena, adibidez.

Hizkuntza ardatz duen AA teknologia giltzarri gisa finkatu da —AAren eremu berritzaile nagusietakoa da eta oso azkar hazten ari den inpaktu ekonomikoa du. Dagoeneko, hainbat enpresa teknologikok aurreikasitako hizkuntza-eredu handiak ezarri dituzte produkzioan, produktuen funtzionaltasunak eta erabiltzaile-esperientzia hobetzeko asmoz. Eredu horiek gizartearen hainbat sektore eraldatzeko potentziala dute, baina ez dira ahaztu behar planteatzen dituzten **erronka tekniko, sozial eta etikoak**.



Hizkuntza Teknologiak

Hizkuntza Teknologiek (HT) diziplinarteko alorra osatzen dute, zeinen helburua den makinek gizakion hizkuntza interpretatu eta sortzea ahalbidetzen duten sistemak eta aplikazioak ikertzea eta garatzea, idatzizko, ahotsezko edota keinu bidezko formetan, besteak beste. HTak modu desberdinetan sailka daitezke eta hizkuntzaren hainbat alor hartzen dituzte, bai idatzizkoa baita ahotsezkoa ere.

Azterketa honek HT eleaniztunak aztertzea du helburu, inpaktu handia izateko potentzial handia duten hiru eremu elkarlotutan:

- **Itzulpen automatikoa** (IA): Testua hizkuntza batetik bestera automatikoki itzultzen duten sistemak, hizkuntza desberdinen arteko komunikazioa errazteko.

- **Ahotsaren teknologiak:** Mintzamenaren transkripzioa edo ezagutza egiteko sistemak (STT, ingelesez *speech-to-text*), giza mintzaira testu idatzian bihurtzen dutenak; eta ahotsaren sintesia egiteko sistemak (TTS, ingelesez *text-to-speech*), testutik ahots artifizialak sortzen dituztenak.
- **Hizkuntza-eredu handiak** (LLM, ingelesez *large language model*): Testu-datu kopuru erraldoiekin trebatu diren AA eredu handiak, hizkuntza naturala modu aurreratuan ulertzeko eta sortzeko gai direnak.

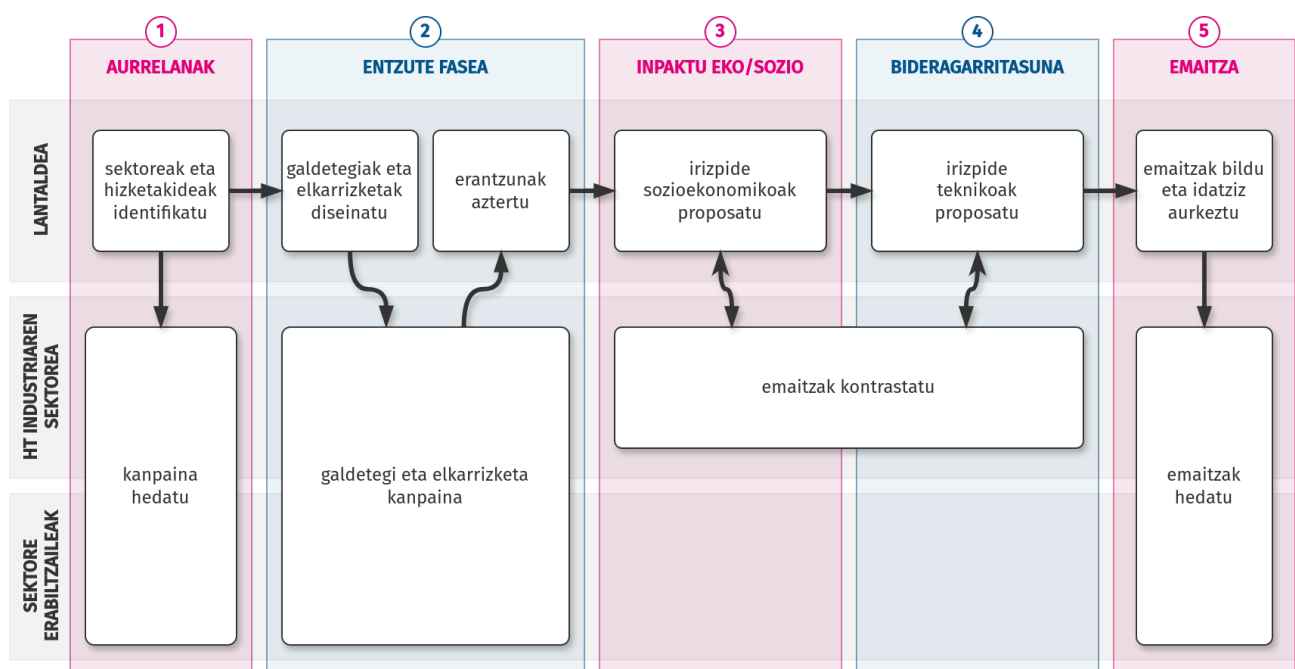
Hiru eremu horietako bakoitzak garrantzi handiko papera betetzen du giza-makina elkarrekinaren aurrerakuntzan, hizkuntza naturalaren ulermen eta sorkuntzarako ikasketa sakonaren teknologia aurreratuenaren egoeraren garapena bultzatuz.

Era berean, gero eta beharrezkoagoa da baliabide eleaniztunak sortzea, bereziki testu, ahots eta itzulpen automatikorako eredu eleaniztunak, gizartearen beharrak asetzeko eta gaur egungo teknologiaren mailara egokitzeko. Garapen horretan, eleaniztasunak eta hizkuntzen arteko transferentziak funtsezko eginkizuna dute.

DIAGNOS- TIKOA

Proiektuaren faseak

Lan honetan, **entzute faseak** berebiziko pisua izan du. Hartara, elkarriketak eta galdetegiak metodologiaren oinarri izan dira. Elkarriketen bidez, emaitza kualitatiboa eta sakonagoa jaso nahi izan dugu, eta horiek, galdetegien bitartez jasotako emaitza kuantitatiboekin osatu ditugu.



Metodologia

Lehenik eta behin, intereseko bi sektore nagusi identifikatu dira: batetik, Hizkuntza Teknologiak ikertzen eta garatzen diharduen sektorea —hau da, Hizkuntza Teknologien industria—; eta, bestetik, teknologia horiek erabiltzen edo aplikatzen ari diren sektoreen multzoa (adibidez, zerbitzu eta industria enpresak, hedabideak, hezkuntza-ren sektorea, administrazio publikoa, osasungintza eta justiziaren sektorea).

Bigarrenik, sektore horietako profil zehatzak aztertuta eta aukeratu dira, eta, jarraian, elkarrizketak zein inkestak gauzatu dira. Modu horretara, gaur egun Hizkuntza Teknologien inguruan dauden egoeraren ikuspegi zabalagoa lortu da, eta hainbat erronka, oztopo eta helburu identifikatu dira, baita teknologia mota bat edo bestea garatzeko edo aukeratzeko orduan kontuan hartu beharreko irizpideak ere.

Elkarrizketak

Elkarrizketak izan dira lan honetan erabili den metodologia nagusia. Kontuan hartu da modu aproposena dela teknika kualitatiboaren bidez informazio askotarikoa eta lanketa honetarako baliagarria den informazioa lortzeko.

Elkarrizketetan erabilitako gidoiak (ikus. A eranskina) hiru bloke nagusi izan ditu: lehenik, gaur egungo erabilera kasuak eta etorkizunean izan daitezkeenak ezagutzea; bigarrenik, gizarteak teknologiekin batera dituen erronkak, helburuak eta mugak aztertzea; eta, hirugarrenik, teknologia horien garapenean edo aplikazioan kontuan hartu beharreko irizpideak.

Elkarrizketa oro sarrera batekin hasi da. Sarrera horrek ikerketaren helburuak eta datuen erabileraren ingurukoak kokatu ditu, kasu honetan isilpekoak. Horrek lagundu du elkarrizketatuengan konfiantza sortzen, eta bizipen eta beharretatik irakurketa osoagoak egiteko baldintzak sortzen. Ondoren, gidoian aurreikusitako gaiei heldu zaie, elkarrizketatuaren interesetara egokitzekeo malgutasunez.

Guztira, **17** banakako edo talde-elkarrizketa egin dira. Hainbat profil eta sektore hartu dira kontuan, eta **30** erreferentziazko pertsonen testigantza jaso da denera.



- 7 HTen industriaren sektoreari.
- 2 industria eta zerbitzu enpresen sektoreari.
- 3 hezkuntzako sektoreari.
- 1 hedabideetako sektoreari.
- 4 sektore publikoko erakundeekin (administrazio publikoa, osasungintza eta justizia).

Elkarrizketetan eskuratutako informazioa biltzeko teknika ahozko grabaketak izan dira, eta elkarrizketetan oharrak ere hartu dira. Elkarrizketatu orori grabatzeko baimena sinatzea eskatu zaio (ikus. B eranskina).

Inkestak

Inkestak gauzatzeko, hainbat irizpide aintzat hartu dira; horien artean, galdetegiaren arintasuna eta erakargarritasuna.

Galdetegiaren hasieran, horien helburuak edota zertarakoa modu labur eta zehatzean azaldu da, eta guztiz anonimoak direla argi utzi da. Idazketaren zainketa ere ezinbestekotzat hartu da.

Profilen arabera 3 galdetegi prestatu dira:

- **G.1 galdetegia:** Hizkuntza Teknologien industriako kideentzat zuzendua, sektoreko joeren eta beharren ulermenean sakontzeko asmoz.
- **G.2 galdetegia:** zerbitzu eta industria enpresetako kideei zuzendua, teknologia berrien erabileraren eta horien inpaktuen gaineko ikuspegia jasotzeko.
- **G.3 galdetegia:** unibertsitateko ikasleen artean banatutakoa, etorkizuneko profesionalek teknologia horien inguruan duten jakintza eta jarrera aztertzeko.

Eginiko hiru inkesten galdera guztiak ez dira elkarren artean berdinak izan, bakoitza dagokion sektorerantz egokituta dagoelako; hala ere, hirurek galdera sekuentzia bera jarraitu dute (ikus. D eranskina):

Hasteko, galdera orokorrak edo identifikazio galderak egin dira. Horien bitartez, galdetegia bete-ko duten pertsonen inguruko segmentazio datu gutxi batzuk lortu dira, geroko datu-analisian ere aintzat hartu direnak.

Ondoren, aurre-lanketa sakona garatu ostean, galdera zehatzetara jo da, ikertu nahi den gaiaren inguruko galderak eginez, horiek baitira ikerketaren oinarri. Galdera itxien eta irekien arteko oreka mantendu da. Itxien arteko erantzun aukerak desberdinak dira, erantzun posible askotarikoak eta joera edo portaeren intentsitate maila neurtzekoak, besteak beste.

Galdera zehatzak bi bloketan banatuta daude: batetik, gaian kokatzea eta sintonizatzea xede dutenak, arazorik gabe ulertzen eta erantzuten direnak; bestetik, lan eremura eramandako gaiaren ingurukoak —hau da, azterlaneko input nagusiak izanen direnak. Elkarriketetan bezala,

galdetegiaren erabilitako galdera batzuk Hizkuntza Teknologien industriako hainbat erakundek 2023. urtean eginiko Europako hizkuntzen industriari buruzko ikerketan (ELIS survey, 2023) oinarrituta daude.

Azkenik, eta galdetegi bakoitzari amaiera emateko, galdera ireki bat dago, zeinetan inkestatuek, nahi izatekotan, iruzkin, ohar edota iritziak utzi ahal izan dituzten.

**Galdetegiak,
Hizkuntza Teknologia
industriako hainbat
erakundek 2023.
urtean eginiko
Europako hizkuntzen
industriari buruzko
ikerketan (ELIS
survey, 2023)
oinarrituta daude.**

Galdetegiak online plataforma batean egin dira eta ele bietan (euskaraz eta gaztelaniaz) jaso dute erabiltzaileek, posta elektronikoko bitartez, beraien hizkuntza hautua egiteko aukera izan dezaten.

Zabalkundea hainbat modutan egin da. Alde batetik, Languneraren sarea erabili da Hizkuntza Teknologien industriara gerturatzeko. Gainera, GAIA elkartearen parte denez, industria eta zerbitzu enpresetara zabaltzeko aukera egon da. Beste alde batetik, CLARIAH-ko mezu elektronikoen zerrenda erabili da zabalkunde masibo horrekin jarraitzeko, eta, azkenik, sare sozialetan ere argitaratu da publiko zabalgabato batera iristeko. Ondoren, profil zehatzen arabera hedapen zuzenagoa egin da, batez ere industria eta zerbitzu enpresetara; industria enpresak, banketxeak, zerbitzu informatikoak eskaintzen dituztenak... Azkenik, hezkuntzako zabalkundea egin da: irakasleei industria eta zerbitzu erakundeetako galde-

tegia betetzeko eskatu zaie eta ikasleei, berriz, hezkuntzara bideratutako galdetegia.

Galdetegietako informazioa kalkulu orri batean antolatu da, inkestaren galdera bakoitzaren datuak ordenatuta eta egituratuta edukitzeko eta, ondoren, horiek aztertu ahal izateko.

Dokumentu honen hurrengo ataletan metodologiari esker eskuratutako elkarrizketen eta galdetegiaren informazioa zein datuen emaitzak aurkeztuko dira.

Zenbakietara ekarrita, **1.000** galdetegi inguru bidali dira **3** kanal nagusi erabilita (GAIA eta Languneko bazkideen kanala, CLARIAH-ko kanala eta Mondragon Unibertsitateko kanala) eta horietatik **281** pertsonak erantzun dute, **% 28**ak.

G.1, Hizkuntza Teknologietan adituak:

20 pertsona

G.2, Zerbitzu eta industrietako profesionalak:

94 pertsona

G.3, Unibertsitateko ikasleak:

165 pertsona

Entzute fasearen emaitzak

Elkarrizketen emaitzak

Parte-hartzea

Guztira, 17 elkarrizketa egin dira, banaka edo taldeka. Profil desberdinen iritzia eta esperientzia kontuan hartu dira, eta honako sektore hauetako erreferente diren 30 pertsonen testigantzak bildu dira:

- Hizkuntza Teknologia
- Industria eta zerbitzuak
- Hezkuntza
- Komunikabideak
- Sektore publikoa: administrazio publikoa, osasungintza eta justizia

Erantzunen analisia

Elkarrizketetatik lortutako emaitzak txosten honen hiru azterketa arloetan banatu dira: itzulpen automatikoa (IA), ahots-teknologia (STT edo TTS), eta hizkuntza-eredu handiak (LLM). Zehazki, arlo bakoitzari dagozkion egungo adierazle, erronka, eta etorkizuneko adierazle eta erronkei buruz jasotako iritzia erregistratu dira.

Itzulpen automatikoari dagokionez, oinarriko beharra asetzeko nahikoa garapen maila lortu da elkarrizketatuen ustez, baita euskaraz ere. Mota horretako sistemak ohikoak dira jada administrazio publikoan, komunikabideetan eta hezkuntza-sektorean. Hortaz, lehendik dauden sistemak hobetzea da helburua gaur egun, hizkuntza-pare gehiago sartuz eta akatsak minimizatuz. Atzemandako erronka nagusietako bat medikuntza edo zuzenbidea bezalako arloetan espezializatutako itzultzaileak sortzea da.

Ahotsaren teknologia integratzen dituzten sistemak arlo garrantzitsu eta etorkizun handiene-koetakoak dira, aztertutako HT guztien artean. Elkarrizketa guztietan aipatu dira bai hizketa eza- gutza bai hizketa sintesia. Era berean, teknologia

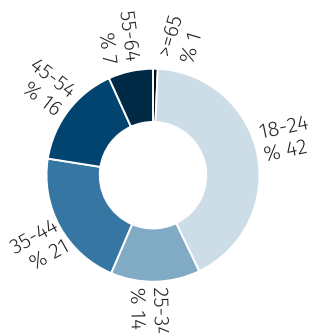
mota hori uste baino aurreratuago dagoela edo erabilera handiagoa izan beharko lukeela nabarmendu da, euskaraz ere bai. Ahots-teknologia hainbat erabileratarako baliatzen ari dira, hala nola robotekin eta bozgorailu adimendunekin elkarreragiteko, aktak eta bideoak transkribatzeko, ahotsak sortzeko, eta sistema horiek administrazio publikoan, justizian, osasunean eta hezkuntzan aplikatzeko proiektuetan.

LLMak produktu komertzialen parte gisa aipatu dira, eta elkarrizketatuek horietako batzuk erabilera pertsonalerako baliatzen dituztela adierazi dute. Elkarrizketetan argitu da bai hizkuntza-teknologiaren garatzaileek bai erabiltzaileek ondo ezagutzen dituztela egungo LLMen mugak eta erronkak, besteak beste, konputazio-gaitasuna, datuen segurtasunaren kudeaketa eta domeinu ezberdinetara egokitzeko gaitasuna. Oraingoz, ez da detektatu LLM propioak garatzeko jardura nabarmenik; hein batean, enpresen mesfidantzen ondorioz, beren datu pribatuak eredu horien garatzaileekin partekatzeko garaian. Horrek zaildu egiten du eredu horiek ingurune espezifikoetara egokitzea.

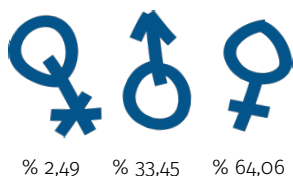
Azkenik, aztertutako arlo guztietan hainbat alderdi komun identifikatu dira. Lehenik eta behin, azpimarratu da garrantzitsua dela makinarekin elkarreragina ahalik eta naturalena izatea, erabiltzaileak bere erregistroa, dialektoa edo azentua moldatu beharrik izan gabe sistemarengandik emaitza onak lortzeko. Gainera, elkarrizketatuek azpimarratu dute funtsezkoa dela hizlariaren emozioak identifikatzea, ahoko zein idatzizko hizkuntzan, eta emozio horiek interpretatzeko gai diren teknologia garatzea. Azkenik, ezinbestekotzat jotzen da eskaintzen diren soluzioak enpresen eta azken erabiltzaileen beharretara egokitzea ez ezik, sistema propioetan modu eraginkorrean integratzea ere.

Inkesten emaitzak

3.1. Irudia: Adina



3.2. Irudia: Generoa



3.3. Irudia: Ama-hizkuntza



3.1. Taula: Sektorea (G.2)

Akademia	% 35.46
Zerbitzuak	% 29.17
Industria	% 13.54
Hedabideak	% 5.21
Ikerketa	% 5.21
Besterik	% 5.21
Administrazio publikoa	% 4.17
Ez du erantzun	% 1.04

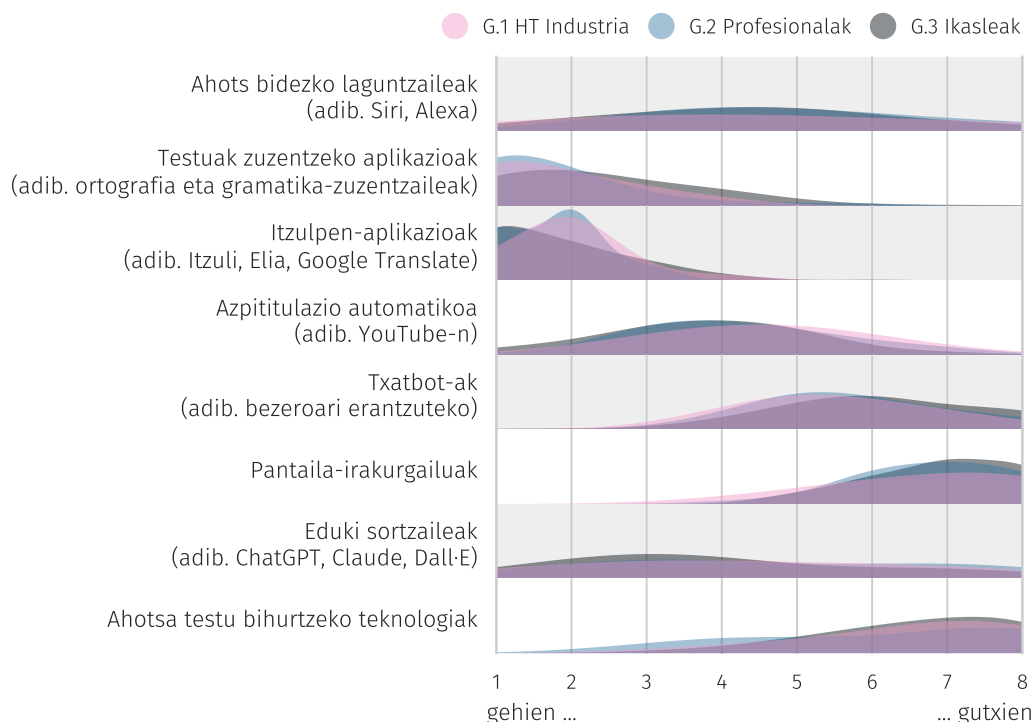
Parte-hartzea

Denera, 1.000 galdetegi inguru bidali dira 3 kanal nagusiren bidez: GAIA eta Langune elkartearen kanalak, CLARIAH sarearen kanala eta Mondragon Unibertsitatearen kanala. Horrez gain, inkesta LinkedIn eta X sare sozialen bidez ere zabaldu da. Azkenean, 281 pertsonak erantzun diote hiru galdetegietako bati: % 7,1ek beren burua Hizkuntza Teknologien aditutzat jo du (G.1 galdetegia); % 33,3k G.2 galdetegia erantzun du, zerbitzuetako eta industriako enpresetako kideentzat, hau da, teknologien erabiltzaile profesionalentzat zuzendutakoa; eta % 59,6a unibertsitateko ikasleei dagokie (G.3 galdetegia). G.2 profilarekin parte hartu dutenen sektorekako banaketa 3.1. taulan ikus daiteke. Parte-hartzaileen adinari dagokionez, % 42,0k 18 eta 24 urte artean ditu, % 50,2k 25 eta 55 urte artean, eta % 7,8k 56 urte baino gehiago. Genero-identitateari dagokionez, % 2,5ek beren burua ez-binariotzat jo du, % 33,4a gizon bezala identifikatu da, eta % 64,1 emakume bezala. Ama-hizkuntzari dagokionez, parte-hartzaileen % 71,4k euskara du ama-hizkuntza, % 25,4k gaztelania, % 1,4k euskara eta gaztelania batera, eta %1,8k beste hizkuntza bat.

Erantzunen analisia

Gaur egun, Hizkuntza Teknologian oinarritutako zein aplikazio erabiltzen dira? Nola eta zertarako erabiltzen dira? Inkestatutako talde guztiek batez ere testuen zuzenketarako eta itzulpenetarako aplikazioak erabiltzen dituzte, baita edukia sortzeko eta azpigitulu automatikoak egiteko tresnak ere (ikus 3.4. irudia). Ez da desberdintasun esanguratsurik aurkitu HTetako adituen (G.1) eta beste profesionalen (G.2) erabilera-ohituretan, ezta adituen eta ikasleen (G.3) artean ere. Hala ere, zerbitzuen eta industriaren arloko profesionalak zuzenketarako aplikazioei lehentasuna ematen dietela ikusi da, ikasleekin alderatuta [Mann-Whitney U, $p < .001$].

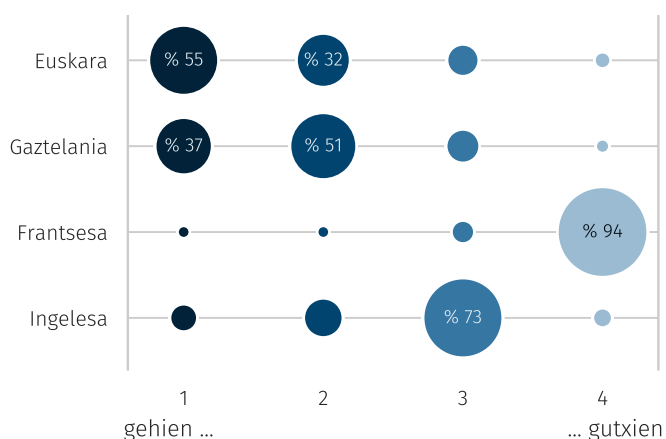
Nola eta zergatik erabiltzen dira? Teknologia horiek erabiltzea justifikatzen duten onurei dagokionez, talde guztiek gehien baloratzen dituzten alderdiak produktibitatea handitzea eta komunikazioaren eraginkortasuna hobetzea dira, eta, ondoren, errepikakorrak diren lanak automatizatzea. Hala ere, talde batzuen artean desberdintasun esanguratsuak daude. Batetik, unibertsitateko ikasleek komunikazioaren eraginkortasuna hobetzea eta testuen zuzenketa nabarmen gehiago baloratzen dituzte HTetako adituek baino [χ^2 , $p < .05$]. Bestetik, ahots bidezko ekintzak egiteko gaitasuna HT arloko profesionalentzat garrantzitsuagoa da beste erabiltzaile profesionalentzat baino [χ^2 , $p < .05$]. Azkenik, interakzioen pertsonalizazioaren onurak lehentasun txikia du talde guztietan. Emaiza horiek guztiak 3.5. irudian aurkezten dira. Teknologia horietan hizkuntzen erabilerari dagokionez, datuek erakusten dute gaztelania eta euskara direla



3.4. Irudia: Zein hizkuntza teknologia erabiltzen dituzu egunero-koan? Ordenatu gehienetik gutxienera.

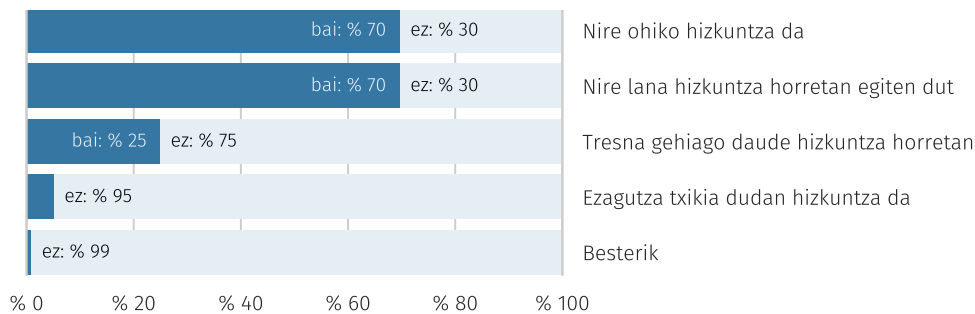
	G.1 HT Industria	G.2 Profesionalak	G.3 Ikasleak
Nire produktibitate maila handitzea	% 78	% 66	% 68
Testuen eta komunikazioaren kalitatea hobetzea	% 50	% 75	% 78
Ataza errepikakorrak automatizatzea	% 39	% 24	% 21
ataza desberdinak ahots bidez egin ahal izatea	% 17	% 1	% 6
Hizkuntza-zerbitzuen irismena handitzea	% 6	% 17	% 24
Interakzioak persontsonalizatu ahal izatea	% 0	% 3	% 4

3.5. Irudia: Onura hauetako zein ezaugarri lehenesten duzu gehien zure erabilera arrazoi-tzeko? Aukeratu nahi beste.

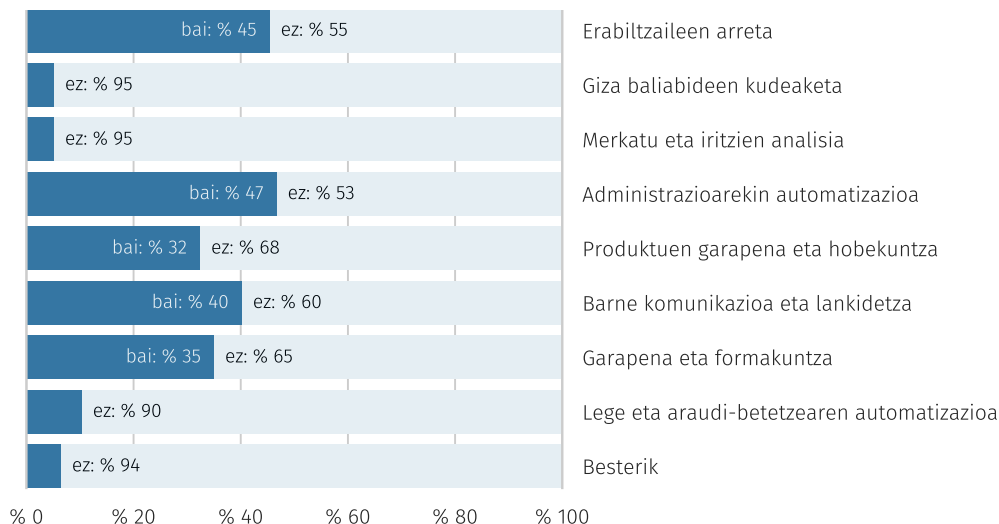


3.6. Irudia: Eguneroko zure erabileran, zein hizkuntzatan erabiltzen dituzu nagusiki? Ordenatu gehienetik gutxienera.

3.7. Irudia: Zergatik?



3.8. Irudia: Zein arlo zehatzetan erabiltzen dituzue erakundeetan HTak?



nagusi, eta haien atzetik datozela ingelesa eta, neurri txikiagoan, frantsesa (ikus 3.6. irudia). Aipatzekoa da profesionalak gaztelania askoz gehiago erabiltzen dutela ikasleek baino [Mann-Whitney U, $p < .001$]. Hizkuntza bakoitza zergatik erabiltzen den aztertuta (ikus 3.7. irudia), bi arrazoi nagusi nabarmentzen dira: erabiltzailearen ohiko hizkuntza izatea eta lan-jarduerak hizkuntza horretan egitea. Beste alde batetik, neurri txikiagoan, tresnen eskuragarritasunak ere eragiten du, eta, oso modu apalean, gutxi menderatzen diren hizkuntzak erabili beharrak ere izan dezake eragina.

Zertarako erabiltzen dira HTak lan-eremuan?

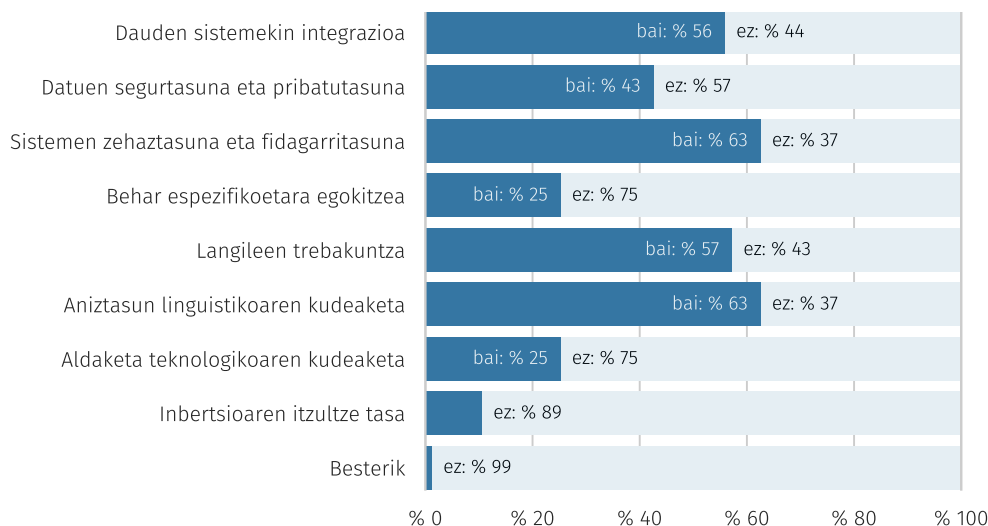
3.8. irudian erakusten den bezala, HTen adopzioak eredu desberdina aurkezten du eremu profesional ezberdinetan (G.2). Penetrazio handieneko aplikazioak bi eremu nagusitan kontzentratzen dira: administrazio eta dokumentu automatizazioan (% 47) eta bezeroaren arreta eta laguntza teknikoan (% 45). Hurbiletik jarraitzen dio barne komunikazio eta lankidetzaren hobekuntzak. Garapenak eta prestakuntzak, baita pro-

duktuen garapen eta hobekuntzak ere, adopzio maila moderatuagoak erakusten dituzte. Aitzitik, adopzio nabarmen baxua dago giza baliabideen kudeaketa eta hautaketa, merkatu eta iritzien analisia, eta lege eta araudien betetze automatizazioa bezalako eremuetan, non inkestatuen % 10ak baino gutxiagok haien erabilera aipatzen duten.

Zer zailtasun dituzte erabiltzaileek HTak lan-eremuan ezartzeko?

HTen lan-eremuetako erabileraren gainera inkestaren azterketak egoera konplexua erakutsi du: erronka teknikoak, antolaketakoak eta prestakuntza-beharrak uztartu behar dira. Alderdi hauek guztiek elkarrekin duten loturak agerian uzten du teknologia horiek ezartzeko ikuspegi osoa behar dela.

- **Antolaketa eta integrazio erronkak:** HTen integrazioa lehendik dauden sistema eta plataforma teknologikoekin erronka nagusietako bat da, inkestatuen %5 6k hala adierazi baitu (ikus 3.9. irudia). Datu kuantitatibo horrek isla du gaur egungo erakundearen egoera: oraindik hasierako



3.9. Irudia:
Zeintzuk dira
erronka na-
gusiak zurea
bezalako en-
presa batean
hizkuntza-
teknologiak
ezartzean?

fasean dago teknologien ezarpena, eta oso zaitkatuta dago. Inkestatuen esanetan, erakunde gehienak integrazioaren lehen urratsetan daude, eta gehienetan norbanakoen ekimen informalak dira, erakundearen estrategia bateraturik gabe. Hala ere, erakunde batzuk hasi dira integrazioa modu sistematikoagoan lantzen, eta teknologia horiek beren lan-fluxuetan txertatzen ari dira, batez ere itzulpen eta dokumentazio prozesuetan.

- **Erronka teknikoak:** Parte-hartzaileen % 63k aipatu du hizkuntza prozesatzeko sistemen fidagarritasuna eta zehaztasuna dela kezka nagusi. Erronka tekniko hau nabarmenagoa da euskararen kasuan, non muga handiak ikusten diren, bai itzulpenen kalitatean, bai euskalkien tratamenduan. Argi dago arlo honetan garapen teknologiko handiagoa behar dela, batez ere tokian tokiko hizkuntza-berezitasunetara egokitzeko. Fidagarritasunaren arazoak direla eta, etengabe egiaztatu behar dira emaitzak, bereziki testuinguru konplexuetan edo baldintza kaskarretan. Egoera are zailagoa da zeregin berezietan: hainbat parte-hartzailearen arteko elkarrizketak transkribatzean edo erakunde bakoitzaren testuinguru berezietara egokitzean. Azken hori inkestatuen % 25ek jo du erronkatzat.

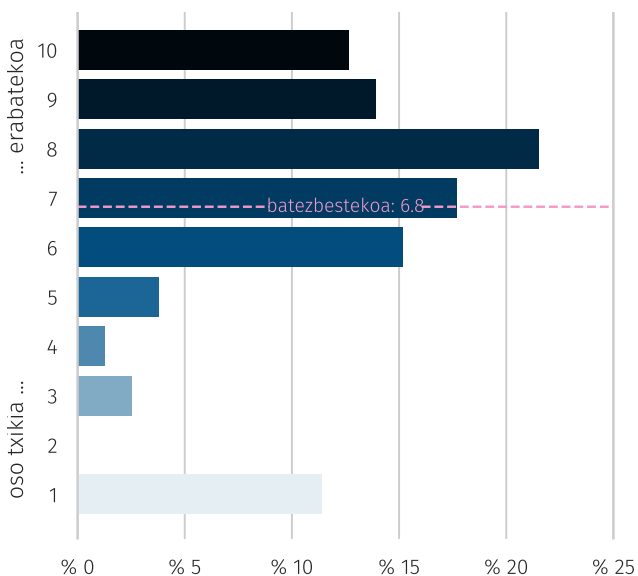
- **Gaitasun erronkak:** Teknologien ezarpenaren giza alderdia ere oso garrantzitsua da: parte-hartzaileen % 57k esan du langileen prestakuntza dela erronka nagusietako bat. Prestakuntzarik eza oztopo handia da, bai ezagutza teknikoaren faltagatik, bai alfabetatze teknologiko zabalagoa

behar delako. Arazo hau are larriagoa da motibazio falta dagoenean, eta kontuan hartu behar da askotan ez dagoela denborarik tresna berriei ikasi eta probak egiteko. Lan egiteko modu berrietara egokitzea ere erronka handia da, nahiz eta, bitxia bada ere, % 25ek bakarrik aipatu aldaketaren kudeaketa erronka gisa; horrek esan nahi dezake faktore horri behar baino garrantzi txikiagoa ematen zaiola. Bestalde, funtsezkoa da teknologia hauen erabileran irizpide kritikoa garatzea: HTak laguntzeko tresnak dira, eta beren erabileran giza gainbegiraketa eta irizpidea behar dira.

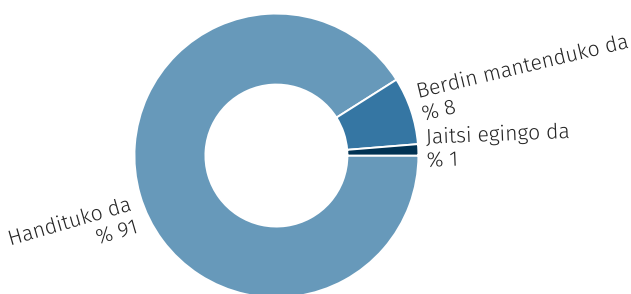
Zer zailtasun dituzte HTetako adituek teknologia garatzeko?

HTetako adituek (G.1) adierazten dute, teknologiak edozein eremu profesionale-tara potentzialki hedagarriak diren arren, haien ezarpen eraginkorra bi erronka handik baldintzatzen dutela, oztopo tekniko zein ekonomiko garrantzitsuak gainditzea eskatzen dutenak. Alde batetik, **datuen eskuragarritasuna eta kalitatea agertzen da oztopo tekniko nagusi gisa**. Adituek erabilera kasu espezifiko bakoitzerako datu nahiko eta egokiak lortzeko zailtasuna azpimarratzen dute, eta horrek azaltzen du neurri handi batean azken erabiltzaileek aipatzen dituzten zehaztasun eta egokitzapen arazoak. Erronka hori bereziki garrantzitsua da euskara bezalako baliabide gutxiagoko hizkuntzentzat, non kalitatezko datuen eskasiak soluzio sendoen garapena mugatzen duen. Bestetik, finantzaketa da HT proiektuak garatzeko beste erronka nagusia. Adituen %100ak dio irizpide ekonomikoak era-

3.10. Irudia: Zenbateko eragina dute gaur egun HTek zuen lan eremuan?



3.11. Irudia: Eta, etorkizunean, eragin maila...



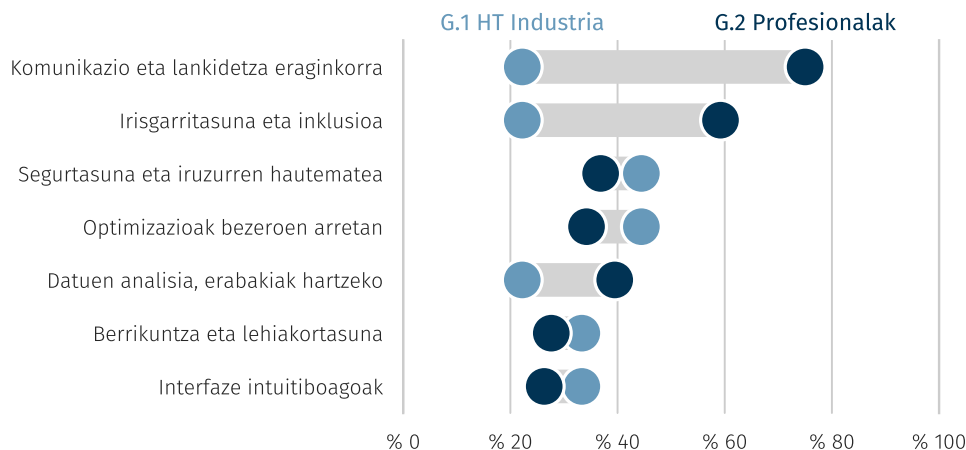
bakigarriak direla proiektu bat martxan jartzeko orduan; irizpide teknologiko-teknikoek (% 67) edo sozialek (% 22) baino pisu handiagoa dute. Ildo horretatik, proiektuak batik bat dirulaguntza publikoei eta funts propioei esker garatzen dira, eta inbertsio pribatuak zeregin apalagoa du.

Zer eragin dute HTek gaur egun eremu profesionalean? Zer eragin izango dute etorkizunean? Profesionalen erantzunen azterketak (G.2) askotariko iritziak erakusten ditu HTek beren lan-eremuan duten eraginari buruz, 1etik 10erako eskalan 6,8ko batezbestekoa lortu dutelarik. Erantzunak oso banatuta daude: talde esanguratsu batek (% 11,4) eragin txikia edo hutsa ikusten du,

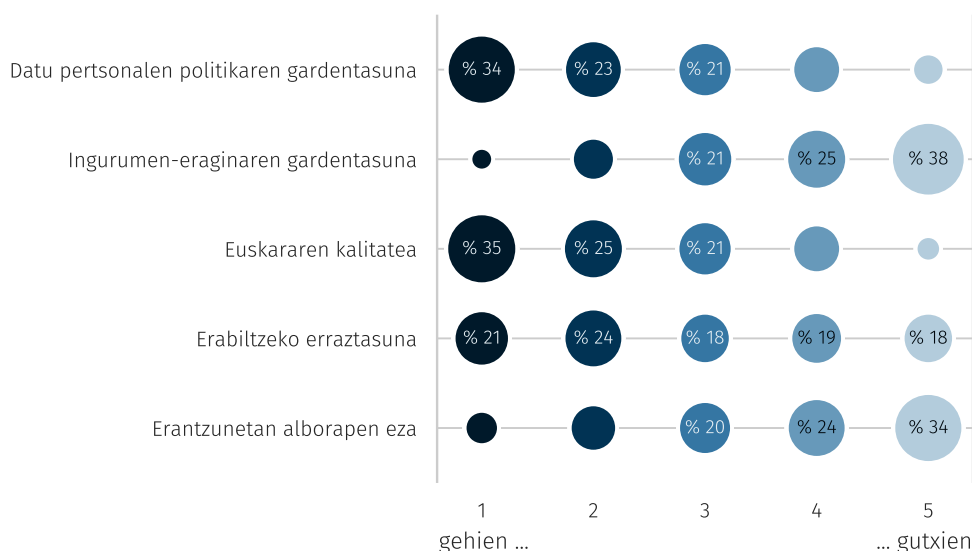
baina balorazio ertain eta altuak dira nagusi, eta asko dira 9 eta 10eko puntuazioak eman dituztenak. Gaur egungo eraginaren pertzepzioan dagoen aldakortasuna gorabehera, argi dago denek bat egiten dutela etorkizuneko bilakaeraren inguruan: profesionalen % 91ren ustez teknologia hauen eragina handituz joango da. % 8k soilik uste du berdin jarraituko duela, eta % 1ek baino gutxiagok aurreikusten du behera egingo duela (3.11. irudia).

Zein dira etorkizuneko erronka nagusiak HTei dagokienez? Komunikazioa eta lankidetzaz erraztea funtsezko erronka da erabiltzaile profesionalen ustez (G.2), eta garrantzi handiagoa ematen diote HT-etako adituek (G.1) baino [χ^2 , $p < .05$]. Era berean, erabiltzaileek diote erronka nagusietako bat dela irisgarritasuna eta inklusioa hobetzea, nahiz eta kasu honetan aldea ez den esanguratsua. Bi taldeek bat egiten dute lehentasunezko eremu hauek nabarmentzean: segurtasuna hobetzea iruzurrak detektatzeko teknologiaren eta hizkuntzan oinarritutako arrisku-analisiaren bidez, eta bezeroen arreta-prozesuak optimizatzea. Aldiz, bi taldeentzat lehentasun txikiagoko erronkak dira hauek: erabakiak hartzeko datu-bolumen handien automatizazioa, merkatuko berrikuntza sustatzea eta interfaze intuitiboagoak sortzea. Emaita horiek 3.12. irudian ikus daitezke. Horrez gain, inkestak berriz ere kezka berezia agertu dute hizkuntza gutxituen inguruko erronkekiko, bereziki euskararen kasuan. Erabiltzaileen arabera, oso garrantzitsua da euskara bezalako hizkuntzen eta ingelesa moduko hizkuntza nagusien artean zerbitzu-kalitate eta -kantitate parekoa lortzea, arreta berezia jarritz hizkuntza-naturaltasunean eta euskara batuaz gain euskalkien zaintzan ere.

Zein irizpidek gidatu beharko lukete HTen etorkizuneko garapena? Unibertsitateko ikasleen, hau da, sektoreko etorkizuneko profesionalen (G.3) ikuspegitik, **euskararen kalitatea da irizpiderik baloratuena**, inkestak % 35ek lehentasun goreneko posizioetan jarri baitu (ikus 3.13. irudia). Hizkuntza-kalitatearen inguruko kezka hau bat dator lehen aipatu ditugun hizkuntza gutxituen erronkekin. Era berean, datu pertsonalen



3.12. Irudia: Zeintzuk dira zure iritziz etorkizuneko helburu nagusiak hizkuntzaren teknologiei dago-kienez? Aukeratu nahi beste.



3.13. Irudia: Hurrengoetatik, zein irizpideri emango zenioke lehentasuna?

politikaren gardentasunak ere balorazio altua jaso du (lehentasun maila gorenetan %34), eta horrek erakusten du gero eta kontzientzia handiagoa dagoela segurtasun eta pribatutasunaren inguruan. Erabiltzeko erraztasunaren balorazioa orekatuago dago eskala osoan zehar, nahiz eta goiko mailetarako joera txiki bat dagoen (% 21 maila gorenean). Aitzitik, ingurumen-eraginaren gardentasuna eta emaitzen inpartzialtasuna ez dira hain garrantzitsutzat jotzen: % 38k eta % 34k, hurrenez hurren, eskalan behearen mailatan kokatu dituzte.

Zer HT aplikazio daude oraindik ikertzeko eta garatzeko? Erantzunek hainbat interes-esparru azaleratzen dituzte. Aipatutako funtsezko alderdietako bat da **elkarrerikintza modalitate desberdinak bateratzen dituzten sistemak** garatzeko beharra, eta horiek modu naturalean txertatu ahal izatea eguneroko aplikazioetan, bai etxean, bai lanean, bai kalean. Testuinguru honetan, nabarmendu egiten da ahots bidezko itzulpen zuzeneko eta elkarreragin adimenduneko sistemak garatzeko dagoen potentziala. Enpresa eta administrazioaren arloan, **dokumentu eleaniztunak kudeatzeko sistema eraginkorrak** garatu beharra ikusten da, eta horien barruan sartzen dira dokumentuak automatikoki sailkatzea, metadatuak erauztea eta hizkuntza naturaleko bilaketak ezartzea. Era berean, eskatzen da barne-komunikaziorako tresna sofistikatuagoak txertatzea eta **asistente**

birtual espezializatuak garatzea hainbat sektoretarako, hala nola hezkuntza eta osasunerako, datuen analisia, txostenen sorkuntza eta bezeroaren arreta bezalako zereginak errazteko. Bestalde, nabarmen agertzen da **gizarte-eragina duten aplikazioekiko** interesa, eta arreta berezia jartzen da irisgarritasun eta inklusioaren arloan. Esparru honetan, hainbat tresna garatzea proposatzen da aniztasun funtzionala duten pertsonentzat, hala nola ahots bidezko kontrol-sistemak eta testu-ahots bihurketa bi norabideetan. Era berean, adinekoen beharretara bereziki egokitutako teknologiak ere proposatzen dira, haien hizketa-ezaugarriak kontuan hartuta. Azpimarratzen da, halaber, administrazio-hizkuntza sinplifikatu eta ulergarriago egiteko sistemak sortzeko beharra, batez ere lege-terminologiarekin ohituta ez dauden pertsonentzat. Azkenik, beste bi garapen-eremu aipatzen dira: desinformazioa automatikoki detektatzeko tresnak eta hizkuntza ikasteko aplikazio berezituak, arreta berezia jarritz migratzaileen hizkuntza-integrazioa erraztean komunitate elebidunetan.

IRIZPIDEEN PROPOSAMENA

Irizpideen sailkapena eta ulermena

Teknologia/proiektu bat aukeratzeko orduan, garrantzitsua da lehenetsi beharreko irizpideak argi izatea. Entzute fasean egindako elkarriketetan, irizpideei buruz galdetu dugu eta erantzunak 3 irizpide multzo nagusitan jaso ditugu.

Garrantzirik handiena **irizpide ekonomikoari** ematen zaiola ikusi da entzute fasean zehar; oinarri ekonomiko sendorik gabe, teknologiarik oparoenak ere porrot egin dezake. Agerikoena da agian. Garrantzia eman diote hizkuntza teknologikoak industriakoek teknologia horren garapenerako behar den inbertsio ekonomikoari eta erabiltzaileek, berriz, teknologia horren aplikaziorako behar den inbertsio ekonomikoari. Enpresek, adibidez, teknologia bat etengabe erabiltzen jarrai dezaketen ebaluatu behar dute, hau da, eguneratzeak izango dituen eta beharko dituen euskarri tekniko berriak. Gainera, kostuak ez dira softwarea edo hardwarea eskuratzeari dagozkionak bakarrik, baita langileak trebatzeari eta lehendik dauden sistemekin integratzeari dagozkionak ere.

Irizpide sozialari dagokionez hainbat proposamen jaso ditugu. Garrantzia handia ematen zaion puntu bat dela esan dezakegu. Teknologia hauek gizartean ere inpaktu positibo eta negatiboa izan dezaketen heinean, garrantzitsuak dira eta kontuan izan behar ditugu, bereziki inklusioari, ekitateari eta etikari begirakoak. Horrek esan nahi du teknologiak hizkuntza komunitate guztientzat irisgarriak izatea eta alborapenak edo bazterketak ez iraunaraztea bermatu behar dela, besteak beste. Esango genuke profil guztietakoek aipatu dizkigutela irizpide sozialak.

Irizpide teknikoetan ere sakondu da elkarriketetan. “Bideragarria” hitza aipatu da hizkuntza teknologien industrian. Aurreko bi irizpideak garrantzitsuak dira, baina teknikoki gauzagarria izan behar du: gaur egun existitzen den teknologia aukerak erabilia, datuak erabilgarri izanda, maila konputazionala kontuan izanda... Gure gaitasunetara egokitzen dena izan behar dela aipatu da eta, batez ere, datuak izatearen gainean sakondu da. Erabiltzaileek erabilerara maiztasunari eta sistemen integrazioari eman diote garrantzia.

Ondorengo orrietan irizpide bloke bakoitzeko (irizpide ekonomikoa, soziala eta bideragarritasun teknikoa) proposamenetatik egindako hutsuketa bat proposatzen da. Azalpen labur bat jaso dugu eta garrantzia maila bat eman zaio.

Irizpideak neurtzeko moduari dagokionez, irizpide bakoitzaren proposamenarekin batera, entzute fasean jasotako garrantzia maila islatzen saiatu gara. Horretarako hiru maila ezberdindu ditugu: 1 maila, garrantzia baxuena izango da eta 3 maila, garrantzia maila altuena.

Irizpideen ponderazio-taulak

EZINBESTEKO irizpideak

PARAMETROAK	AZALPENA
Tresna euskaraz erabilgarri egotea	Tresnaren euskarazko erabilera bermatzea, lokalizaziorako bide ematen duen tresna izatea, hainbat hizkuntzatan erabilgarri jar daitekeena.
Aurrekontuaren egokitasuna	Dirulaguntza bidez edo "CALL"ean eskainitako diru kopuruarekin egingarria den planteamendua izatea.

Irizpide EKONOMIKOAK

PARAMETROAK	AZALPENA
Merkatu potentziala	- Merkatuaren behar bati erantzunez, diagnostiko edo azterketa batean oinarritzen den egiaztatzea. - Teknologiaren potentziala eta merkatuan duen hazkunde aukerak ebaluatzea. - Proiektua estandarizatzeko aukera ematen duen edo oso pertsonalizatua den zehaztea. - Negozio eredua definitu den egiaztatzea.
Balio ekonomikoa	- Garapenerako aurretiko iragazki gisa, ideia bideragarritasuna baloratzeko inbertsio maila ebaluatzea. - Soluzioa merkaturatzeko beharrezko inbertsio maila zehaztea. - Soluzioaren mantenuari bideratutako inbertsio edo gastua aztertzea. - Zaharkitzeko arriskua eta aurreikusitako epea ebaluatzea. - Soluzioaren fase bakoitzaren bideragarritasun ekonomikoa egiaztatzea.
Inbertsioaren itzulera	Teknologiak onura neurgarriak sortzeko duen gaitasuna ebaluatzea, hala nola produktibitatea handitzea, merkatua handitzea edo kostuak murriztea.
EMAITZA	3

Irizpide SOZIALAK

PARAMETROAK	AZALPENA
Irisgarritasuna eta erabilgarritasuna	• Teknologiarako sarbidea pertsona guztientzat dela egiaztatzea, ezgaitasunak dituzten pertsonak barne, diseinu inklusiboko garapenen bidez. • Teknologiaren eskuragarritasuna erabiltzaile kopuruaren eta erabilera maiztasunaren arabera sektore espezifikoetan zehaztea. • Teknologia erabiltzeko erraza izatea, erabiltzaileen gaitasun eta esperientzia maila desberdinen bidez interfaze intuitiboen bidez. • Teknologiaren prezioa eta sarbidea bideragarria eta bidezkoa izatea ahalik eta pertsona gehienentzat. • Lokalizazioa eta hizkuntza berrikustea, edukiak eta interfazea tokiko hizkuntza eta kulturara egokituz.
Etika adimen artifizialean	• Egungo legeria betetzea bermatzea. • AA sistemetan alborapenik ez izateko ikuspegia ebaluatzea. • Sistemaren funtzionamenduan eta erabaki automatizatueta gardentasuna bermatzea. • Diskriminazioa saihesteko ezarritako neurriak egiaztatzea. • Teknologiarekiko konfiantza handitzen duten ezaugarriak ebaluatzea.
Komunitatean duen eragina	• Tokiko komunitatea eta bere garapena sustatzen duen proposamena den zehaztea, eredu hegemonikoak saihestuz. • Komunitatearen testuinguruari erantzuten dion egiaztatzea. • Konponbideak komunitatean izan dezakeen eragin positibo zein negatiboa ebaluatzea.
Lankidetza	• Eragile publikoen, pribatuen eta herritarren arteko lankidetza planteatzen den egiaztatzea. • Proiektuak ekosistemaren garapenari laguntzen diola bermatzea.
Datuen pribatutasuna eta segurtasuna	• Erabiltzaileen datuen erabilera segurua eta pribatutasuna errespetatzen dela bermatzea, enkriptazioa eta pribatutasun-politika sendoak barne.
Ingurumena	• Konponbideak ingurumenean sortzen dituen eragin positibo edo negatiboak ebaluatzea. • Eraginak biderkatu edo murrizteko aukerak identifikatzea.
EMAITZA	1

Irizpide TEKNIKOAK

PARAMETROAK	AZALPENA
Teknologiaren zehaztasuna eta eraginkortasuna	• Erabiliko diren datu mota eta teknologiaren zehaztasuna eta eraginkortasuna identifikatzea • Dagoeneko existitzen diren teknologien erabilera eta horien maila ebaluatzea, helburua kanpoko teknologia bat egokitzea edo bertakoa erabiltzea den aztertzeke.
Eskalagarritasuna - integraziorako aukerak	• Teknologiak, errendimendua galdu gabe, eskaria handitzen denean hazteko eta egokitzeko duen gaitasuna egiaztatzea. • Teknologiak beste erakunde edo sektoreekin hedatzeko eta integratzeko duen gaitasuna ebaluatzea. • Teknologiaren eskalabilitatea bermatzea, gailu desberdinetan (mugikorrak, tabletak eta ordenagailuak) funtzionatzeko egokitzuz, diseinu optimizatuaren bidez.
Berrikuntza eta originaltasuna	• Ohikoak ez diren esplorazioak sustatu eta ikerketa-lerro berriak irekitzen dituzten proposamen berritzaileak txertatzea.
Teknologiaren irekitasuna	• Teknologiaren irekiera-maila aitortzea, bai programazioan, bai ereduetan eta datuetan, eta CC edo copyright bezalako estandar ezagunekin duen lerrokatzean.
EMAITZA	2

ERAGIN HANDIKO ERAKUSLEAK

Entzute fasean zehar gehien errepikatu diren erakusle zerrenda aurkeztuko da atal honetan, egungo behar, gabezi eta erronka nagusiei erantzuten dietenak, hain zuzen ere.

Erakusleen fitxa teknikoan modu argian azalduko dira:

- erakusle bakoitzaren erabilera edota zertarakoa
- erakusle horren erabiltzaile edo aplikazio sektorea
- erabilera-kasu adibide zehatzak
- teknologiaren heldutasun-maila (TRL)
- berau garatzeko beharrezkoak diren baliabideak
- erronka zientifiko-tekniko nagusiak

Erakusleen fitxa teknikoak

1. Erakuslea: Azpтитuluen sorrera automatikoa
2. Erakuslea: Akta jasotzailea
3. Erakuslea: Ahotsaren momentuko itzulpena
4. Erakuslea: Txatbot enpatikoa
5. Erakuslea: Ahotsaren klonazioa
6. Erakuslea: Erabiltzailean zentratutako itzultzailea
7. Erakuslea: Testu egokitzailea
8. Erakuslea: Bilatzaile propioa
9. Erakuslea: Prozesu bulegotikoen automatizazioa

Azpitituluen sorrera automatikoa

Teknologia nagusia: STT; aukeran, IA

Eragin handiko aplikazioa: Azpigituluak denbora errealeko itzulpenarekin bilera birtualetarako

Zein beharri erantzuten dio?	1. Ikus-entzunezko edukiak publiko askorentzat eskuragarriago eta ulergarriagoak egiteko beharrari erantzuten dio. Teknologia horrek edukia audiorik gabe kontsumitzea ahalbidetzen du eta, horrekin, hizkuntza-oztopoak gainditu eta entzumen urritasuna duten pertsonen lagundu diezaieke, adibidez. Gainera, azpigitulu automatikoek edukien hedadura globala areagotzen dute (adibidez, bilatzaileen indexazioa aberastu dezakete), eta eskuzko azpigitulazioaren ordezkotzat alternatiba azkar eta ekonomikoa eskaintzen dute, ikus-entzunezko material kopuru handiagoa denon eskura egon dadin. Azpimarratzekoa da teknologia hori bi ardatz nagusitan garatzen dela: alde batetik, azpigituluak egin daitezke denbora errealean (online) edo offline, behar eta testuinguru ezberdinetara egokituz; bestalde, itzulpen automatikoarekin edo gabe inplementatu daiteke. Horrela, hizkuntza beraren barruan irisgarritasuna eta hizkuntza ezberdinen arteko ulermena ahalbidetu ditzake.
Erabiltzaileak	2. Teknologia hori edozein sektoretan aplika daiteke. Erabiltzaile talde zabal bati zuzenduta dago, hala nola, eduki bisualak eskala handian ekoizten dituzten eta haien edukia irisgarritasun legeak betez eskuragarri jarri behar duten enpresa edo erakundeei, baita zuzeneko ekitaldiak, hitzaldiak edo webinar-ak antolatzen dituzten eta azpigituluak denbora errealean eskaini nahi dituztenei ere.
Aplikazioak	• Bilera birtualetan itzulitako azpigituluak denbora errealean eskaintzea, ingurune eleaniztunetan ulermena hobetzeko. • Zuzeneko emankizunetan azpigituluak eskaintzea, albisteetan edo kirolatan, adibidez. • Internet bidezko hezkuntzarako azpigituluak sortzea, hala nola online ikastaro eta hitzaldietan, entzumen-desgaitasuna duten edo bertakoa ez den hizkuntza batean ikasten duten ikasleentzako irisgarritasuna hobetzeko. • Museoetan, monumentuetan edo interes turistikoko guneetan azpigituluak ematea eta informazio-bideoak itzultzea.
Heldutasun Teknologikoko Maila (TRL)	4. Teknologia honen heldutasuna nabarmen alda daiteke bere aldaerak kontuan hartuta, batez ere euskara bezalako baliabide gutxiren testuinguruan: • Offline azpigituluak itzulpenik gabe: TRL 7-9 • Online azpigituluak itzulpenik gabe: TRL 6-7 • Offline azpigituluak itzulpenarekin: TRL 6-7 • Online azpigituluak itzulpenarekin: TRL 4-6 TRL hauek estimazioak dira eta faktore zehatz batzuen arabera alda daitezke, besteak beste, audioaren kalitatea eta edukiaren konplexutasuna (ikus behean Erronka zientifiko-teknikoak).

OINARRIZKOAK

- **Audio- eta testu-corpus lerrokatua:** Hizkuntza eta eremu esanguratsuetako audio kopuru handiak, dagozkien transkripzioekin, eredu akustiko bat edo ertzetik ertzerako STT eredu bat sortu edo egokitzeko.
- **Segmentazio softwarea:** Transkripzioa automatikoki azpтитulu egokietan banatzeko, hala nola luzera, gramatika-konposizioa eta pantailan egoteko denbora bezalako ezaugarriak kontuan hartuta.
- **Letra larrietarako eta puntuazio softwarea:** Transkripzioak aberasteko, irakurgarriagoak eta naturalagoak izan daitezen.

BESTE BATZUK

- **Testu corpora:** Hizkuntza eta eremu esanguratsuetako testu kopuru handiak, STT osagaiaren hizkuntza-eredua sortzeko edo egokitzeko.
- **Ahoskera-hiztegia:** Aplikazio-hizkuntzetako fonemak hitzekin lotzeko mapa.
- **Corpus paraleloa:** Lerrokatutako testu kopuru handiak hainbat hizkuntzatan, IA modulu bat trebatzeko edo egokitzeko.

5.
Zein baliabide
edo tresna
behar dira?

1. **Kalitatezko audioen bilketa:** Oinarrizko gailuekin eta ingurune zailetan egin-dako grabaketetan, audioaren kalitatea okertu ohi da. Horrek negatiboki eragiten dio hizketa-ezagutzaren zehaztasunari. Kalitate handiko banakako mikrofonoak erabiltzea konponbide tekniko eraginkorra da; hala ere, konponbide honek sistemaren erabilgarritasuna eta adopzioa asko mugatzen ditu.
2. **Hiztun anitzen hizkera espontaneo:** Errepikapenen, betegarrien eta ete-naldien gisako fenomenoak ohikoak dira eguneroko elkarrizketetan, bereziki hainbat hiztunekin, elkarri mozten diotenean. Fenomeno horiek STT sistemak nahas ditzakete, horiek behar bezala kudeatzeko diseinatuta ez badaude.
3. **Hiztegi tekniko eta espezializatua:** Testuinguru zehatzetan, hala nola eztabai-da mediko edo juridikoetan, egokitzapen-datu estandarretan ager ez daitez-keen termino tekniko eta espezializatuak erabiltzen dira. Funtsezkoa da siste-mek horiek zehaztasunez antzematea beren eremuetan erabilgarriak izateko.
4. **Ahozkoaren itzulpena:** Hizketa testu idatzia baino informalagoa izan ohi da. Gainera, esanahiaren interpretazioak testuinguruaren ulermen sakona eska-tzen du, eta hori automatizatzea zailagoa izan daiteke testu narratibo formal eta egituratuagoekin alderatuta.
5. **Itzulitako azpтитuluetan sinkronizazioa:** Itzulpenak jatorrizko testuaren luze-ra alda dezakete; hala ere, pantailan sartzeko moduan egokitu behar dira, irakurgarritasuna eta audioarekiko sinkronizazioa galdu gabe.
6. **Denbora erreala eta latentzia:** Denbora errealeko aplikazioetan, itzulitako transkripzioak eta ahots-sintesia azkar sortu behar dira ahalik eta latentzia txikienarekin. Horrek prozesamendu-katearen eta hura osatzen duten modu-lu ezberdinen integrazioa optimizatzea eskatzen du, prozesamendu-abiadura hobetzeko emaitzaren kalitatea kaltetu gabe.

6.
Erronka
zientifiko-
teknikoak

Akta jasotzailea

Teknologia nagusia: STT eta HP; aukeran, IA

Eragin handiko aplikazioak: Hizkuntza koofizialak dituzten lurraldeetan udalbatzetako akta eleaniztunak sortzea

1. Zein beharri erantzuten dio?	Teknologia honek modu eraginkorrean sortzen ditu hainbat bilera edo topaketatit eratorritako ideia nagusiak, erabaki garrantzitsuak eta ekintza-puntuak laburbiltzen dituzten erregistro formal edo aktak, gerora akordioak eta zereginak kontsultatu ahal izateko. Zehazki, parte-hartzaile ezberdinen esku-hartzeak modu egituratuan jasotzea ahalbidetzen du, bakoitzaren ekarpenak automatikoki esleituz, eta ideien trukea laburbiltzen du, elementu garrantzitsuak nabarmenduz eta bigarren mailakoak edo errepikakorrak baztertuz. Teknologia honek, gainera, parte-hartzaile guztiei bileraren helburuan erabat murgiltzea ahalbidetzen die, oharrak hartzearen distrakziorik gabe. Azkenik, itzulpen automatikoarekin konbinatuta, aktak hizkuntza ezberdinetako hitzunez ulertzea ahalbidetzen du nazioarteko edo testuinguru eleaniztunetan.
2. Erabiltzaileak	Teknologia hau edozein sektoretan aplika daiteke. Erabiltzaile talde zabal bati zuzenduta dago, bereziki bilerak, hitzaldiak eta elkarrizketak kudeatzen dituzten erakundeei. Bereziiki interesgarria da bilerak, bilera estrategikoak edo jarraipen-bilerak, besteak beste, azkar eta eraginkortasunez dokumentatu eta sintetizatu behar dituztenentzat. Zehazki, sektore publikoan eta gobernuan, akta paraleloak automatikoki hainbat hizkuntzatan sortzeak gardentasuna eta informaziorako sarbidea hobetzen lagun dezake.
3. Aplikazioak	• Udalbatzetako akta eleaniztunak sortzea, non solaskideek lurraldeko edozein hizkuntza ofizial erabiltzen duten. • Eskola-saioen aktak sortzea, zeintzuetan aurkeztutako funtsezko gaiak, ikasleen galderak eta barneratzeko azalpen gehiago beharko lituzketen kontzeptuak identifikatzen diren. • Proiektu teknikoan jarraipen-bileren akta xehatuak sortzea, zeintzuetan funtsezko erabakiak, kontingentzia-planak, zereginen esleipenak eta gauzatze-epeak jasotzen diren. • Talde-terapiako saioen edo laguntza-talde akta anonimizatuak sortzea, terapeutei edo bitartekariei aukera emanaz patroiak eta aurrerapenak aztertzeke, parte-hartzaileen pribatutasuna arriskuan jarri gabe.
4. Heldutasun Teknologikoko Maila (TRL)	Akta jasotzaile automatikoen heldutasun teknologiko-maila aldatu egiten da sistema osatzen duten osagai zehatzen arabera. Garrantzitsua da ulertzea teknologia honek hainbat I+G diziplina uztartzen dituela, bakoitza garapen-fase desberdinetan. Jarraian, osagai nagusien eta dagozkien TRL-en banaketa aurkezten da euskara bezalako baliabide gutxiko testuinguruetarako: • Transkripzio automatikoa: TRL 8-9 • Hizlarien identifikazioa: TRL 6-8 • Transkripzioen ulermena eta akten sorrera: TRL 2 • Itzulpen automatikoa: TRL 8-9

OINARRIZKOAK

- **Audio- eta testu-corpus lerrokatua, eta dagozkien aktak:** Eredu akustikoa edo ertzetik ertzerako STT ereduak, baita HP erabiliz transkripzioak ulertzeko eta aktak jasotzeko ereduak sortzeko edo egokitzeko hizkuntza eta eremu esanguratsuetako audio kopuru handiak, dagozkien transkripzioekin eta aktekin.
- **Identifikatu beharreko hizlarien laginak:** Parte-hartzaile bakoitzaren grabaketa labur baina askotarikoak (guztira 30-60 segundokoak izan ohi dira) hizlarien identifikazio-sisteman erregistratzeko, sistemak ahots ezberdinak ezagutu eta elkarriketaren txanda bakoitza dagokion hitzunari esleitu ahal izateko.

BESTE BATZUK

- **Hizkuntzak identifikatzeko softwarea:** Hizkuntza batean edo bestean ahozko edukiak bereizteko, eta dagozkien prozesamendu-tresnak aplikatzeko.
- **Testu corpora:** Hizkuntza eta eremu esanguratsuetako testu kopuru handiak, STT osagaiaren hizkuntza-eredua sortzeko edo egokitzeko.
- **Ahoskera-hiztegia:** Aplikazio-hizkuntzetako fonemak hitzekin lotzeko mapa.
- **Letra larrietarako eta puntuazio softwarea:** Transkripzioak aberasteko, irakurgarriagoak eta naturalagoak izan daitezen.
- **Corpus paraleloa:** Lerrokatutako testu kopuru handiak hainbat hizkuntzatan, IA modulu bat trebatzeko edo egokitzeko.

5. Zein baliabide edo tresna behar dira?

1. **Kalitatezko audioen bilketa:** Oinarrizko gailuekin eta ingurune zailetan egin-dako grabaketetan, audioaren kalitatea okertu ohi da. Horrek negatiboki eragiten dio hizketa-ezagutzaren zehaztasunari. Kalitate handiko banakako mikrofonoak erabiltzea konponbide tekniko eraginkorra da; hala ere, konponbide honek sistemaren erabilgarritasuna eta adopzioa asko mugatzen ditu.
2. **Hiztun anitzen hizkera espontaneoak:** Errepikapenen, betegarrien eta etenaldien gisako fenomenoak ohikoak dira eguneroko elkarriketetan, bereziki hainbat hitzunean, elkarri moztan diotenean. Fenomeno horiek STT sistemak nahas ditzakete, horiek behar bezala kudeatzeko diseinatuta ez badaude.
3. **Hiztegi tekniko eta espezializatua:** Testuinguru zehatzetan, hala nola eztabaida zientifiko, mediko edo juridikoetan, STT sistemen egokitzapen-datu estandarretan ager ez daitezkeen termino tekniko eta espezializatuak erabiltzen dira. Funtsezkoa da hiztegi espezializatuak zehaztasunez antzematea beren eremuetan erabilgarriak izateko.
4. **Zereginaren subjektibotasuna:** Aktak sortzeak bilera edo ekitaldi bateko informazio garrantzitsua interpretatu eta hautatzea dakar, subjektiboa izan daitekeena erabiltzaile ezberdinen ikuspuntuaren eta eztabaidaren asmoaren arabera. Zereginaren ezaugarri honek arazoa modelatzeko zehaztasunari eragiteaz gain, haren ebaluazioa zailtzen du.
5. **Testuinguru bisualaren falta:** Aurrez aurreko bileretan, parte-hartzaileek keinuak eta “zu” edo “hau” bezalako erreferentzia deiktikoak erabiltzen dituzte. Testuinguru bisualaren faltak sistema automatikoei erreferentzia horiek interpretatzea zailtzen du, eta horrek sortutako akten zehaztasunari eta koherentziari eragiten die.
6. **Emaitzen azalgarritasuna:** Aktak sortzeko zereginak berezko duen subjektibotasunak emaitzaren azalgarritasunaren garrantzia azpimarratzen du. Sistemak azaldu behar dute nola zehaztu eta hautatu diren ideia gakoak, gardenki aurkeztuz transkripzioaren zein zatik eragin duen sortutako laburpenean.

6. Erronka zientifiko-teknikoak

Ahotsaren momentuko itzulpena

Teknologia nagusia: STT eta IA; aukeran, TTS

Eragin handiko aplikazioak: Estatuko hizkuntza ofizialen eta migratzaile-komunitate nagusietako hizkuntzen arteko berehalako ahots-ahots itzulpena

1. Zein beharri erantzuten dio?	1. Teknologia honek hizkuntza desberdinak hitz egiten dituzten pertsonen arteko komunikazio arina izateko beharrari erantzuten dio, denbora errealean hizkuntza-oztopoak ezabatuz. Testuinguru eleaniztunetan edo nazioartekoe-tan elkar ulertzea errazten du, elkarreragin naturalagoak eta eraginkorragoak ahalbidetuz berehalako komunikazioa ezinbestekoa den egoeretan, enpresa arteko negoziatioetatik hasi eta atzerriko herrialdeetako larrialdietako mediku-laguntzaraino. Gainera, ahots-sintesi teknologia txertatzen bada, guztiz ahoko komunikazioa ahalbidetu daiteke. Aukera hau bereziki baliotsua da itzulitako testua irakurtzea deserosoa edo ezinezkoa den egoeretan, hala nola kontaktu bisuala mantentzea garrantzitsua den elkarreraginetan edo ikusmen-urritasuna duten pertsonentzat.
2. Erabiltzaileak	2. Teknologia hau hainbat sektoretan aplika daiteke, besteak beste, hezkuntzan, osasunean, administrazio publikoan, turismoan, nazioarteko negozioetan eta larrialdi zerbitzuetan. Onuradun nagusiak tokiko hizkeran mugak dituzten pertsonak dira, hala nola etorkinak, errefuxiatuak eta behin-behineko bisitariak (turistak, lanerako bidaiatzen duten pertsonak, etab.).
3. Aplikazioak	• Osasun zentro eta ospitaleetan, medikuen eta atzerriko pazienteen arteko komunikazioa erraztuz. • Hezkuntza-erakundeetan, ikasle etorkinen irakasle eta gurasoen arteko komunikazioa hobetzeko. • Familien baitan, guraso atzerritarrek beren seme-alaben hezkuntza-prozesuan aktiboki parte har dezaten erraztuz. • Gobernu eta administrazio zerbitzuetan, etorkinei eta errefuxiatuei ezinbesteko tramiteetan lagunduz. • Nazioarteko lantaldeetan. Orokorrean, ikusmen-urritasuna duten pertsonen inklusioa sustatzeko balio dezakeen teknologia da ingurune eleaniztunetan, hizkuntza ezberdinetako informazioa eta zerbitzuetara sarbidea erraztuz.
4. Heldutasun Teknologikoko Maila (TRL)	4. Heldutasun teknologikoaren mailan zuzenean eragiten duen aldagaietako bat sistemak ulertu eta itzultzeko gai izan behar duen hizkuntza-parea da. Zentzu honetan, Espainiako migratzaileen komunitate nagusien hizkuntzak zailtasun desberdinak aurkezten dituzte, adibidez, txinera estandarra (baliabideetan aberatsa den hizkuntza), urduera, errumaniera eta ukrainera (baliabide asko dituzten hizkuntzak) edo darija eta haren aldaerak (baliabide gutxikoak). Horri gehitu behar zaio hizkuntza horien eta gaztelania ez den estatuko hizkuntza ofizialen arteko itzulpenerako kalitatezko baliabide paraleloen eskasia. Ondorioz, teknologia hau TRL baxu eta ertainetan kokatzen da, 2tik 4ra. Hizkuntza hegemonikoen kasuan (ingeleza, frantsesa), TRL 8ra edo 9ra iritsi daiteke.

OINARRIZKOAK

- **Audio- eta testu-corpus lerrokatua:** Hizkuntza eta eremu esanguratsuetako audio kopuru handiak, dagozkien transkripzioekin, eredu akustiko bat edo ertzetik ertzerako STT eredu bat sortu edo egokitzeko. Corpus hau intereseko hizkuntzetan paraleloa izango balitz, ertzetik ertzerako transkripzio- eta itzulpen-eredu bat garatzea ahalbidetuko luke.
- **Corpus paraleloa:** Lerrokatutako testu kopuru handiak hainbat hizkuntzatan, IA modulu bat trebatzeko edo egokitzeko.
- **Letra larrietarako eta puntuazio softwarea:** Transkripzioak aberasteko, irakurgarriagoak eta naturalagoak izan daitezen.

BESTE BATZUK

- **Testu corpora:** Hizkuntza eta eremu esanguratsuetako testu kopuru handiak, STT osagaiaren hizkuntza-eredua sortzeko edo egokitzeko.
- **Ahoskera-hiztegia:** Aplikazio-hizkuntzetako fonemak hitzekin lotzeko mapa.

5.
Zein baliabide
edo tresna
behar dira?

1. **Kalitatezko audioen bilketa:** Oinarrizko gailuekin eta ingurune zailetan egindako grabaketetan, audioaren kalitatea okertu ohi da. Horrek negatiboki eragiten dio hizketa-ezagutzaren zehaztasunari. Kalitate handiko banakako mikrofonoak erabiltzea konponbide tekniko eraginkorra da; hala ere, konponbide honek sistemaren erabilgarritasuna eta adopzioa asko mugatzen ditu.
2. **Hiztun anitzen hizkera espontaneoak:** Errepikapenen, betegarrien eta etenaldien gisako fenomenoak ohikoak dira eguneroko elkarrizketetan, bereziki hainbat hiztunekin, elkarri moztan diotenean. Fenomeno horiek STT sistemak nahas ditzakete, horiek behar bezala kudeatzeko diseinatuta ez badaude.
3. **Hiztegi tekniko eta espezializatua:** Testuinguru zehatzetan, hala nola eztabaida mediko edo juridikoetan, egokitzen datu estandarretan ager ez daitezkeen termino tekniko eta espezializatuak erabiltzen dira. Funtsezkoa da sistemek horiek zehaztasunez antzematea beren eremuetan erabilgarriak izateko.
4. **Hizkuntza-aldaeren kudeaketa:** Teknologia honen aplikazio-eremuek hiztunen aniztasuna dute ezaugarri: dialekto, azentu eta hizkera desberdinak erabil ditzakete, eta horrek ahotsa automatikoki ezagutzea zein itzultzea zaildu egiten du. Erronka honi aurre egiteko, osagai desberdinak hizkuntza-aldaera horien adierazgarri diren datuekin trebatu behar dira, nahiz eta datu horiek biltzea eta etiketatzea garestia eta nekeza izan.
5. **Ahozkoaren itzulpena:** Hizketa testu idatzia baino informalagoa izan ohi da. Gainera, esanahiaren interpretazioak testuinguruaren ulermen sakona eskatzen du, eta hori automatizatzea zailagoa izan daiteke testu narratibo formal eta egituratuagoekin alderatuta.
6. **Norabide biko itzulpena:** Eszenatoki errealetan erabilgarria izateko, teknologiak zehaztasun onargarri bi noranzkoetan itzultzeko gaitasuna izan behar du, eta horrek bere garapenaren ahalegina eta zailtasuna biderkatzen ditu.
7. **Denbora erreala eta latentzia:** Denbora errealeko aplikazioetan, itzulitako transkripzioak eta ahots-sintesia azkar sortu behar dira ahalik eta latentzia txikienarekin. Horrek prozesamendu-katearen eta hura osatzen duten modulu ezberdinen integrazioa optimizatzea eskatzen du, prozesamendu-abiadura hobetzeko emaitzaren kalitatea kaltetu gabe.
8. **Hutsezintasunaren espektatiba:** Aplikazio-eremuetako askoren kritikotasuna dela eta, ia erabateko errendimendua espero da. Itzulpenean egindako akatsak ondorio larriak izan ditzakete, baina erabiltzaileek ezin dute sistemaren erantzuna berrikusi edo posteditatu. Horrek presio handia sortzen du zehaztasun oso altua lortzeko.

6.
Erronka
zientifiko-
teknikoak

Txatbot enpatikoa

Teknologia nagusia: STT eta HP; aukeran, TTS

Eragin handiko aplikazioa: Ongizate psikologikoa monitorizatzea eta buruko osasunarekin lotutako arazoak dituzten pertsonengandik informazio fidagarria biltzea

1. Zein beharri erantzuten dio?	Txatbot enpatikoek gizakien eta makinen arteko interakzioetan komunikazio gi-zatiarragoa eta ulerkorragoa izateko beharrari erantzuten diote. Ahotsaren bidez pertsonen emozioak automatikoki identifikatu eta aztertuz, txatbot horiek erabiltzailearen esperientzia hobetzen dute, laguntza emozionala eskainiz, frustrazioa murriztuz eta komunikazioaren eraginkortasuna areagotuz.
2. Erabiltzaileak	Txatbot enpatikoak bereziki erabilgarriak dira laguntza emozionala eta enpatia funtsezkoak diren testuinguruetan, hala nola bezeroarenganako arretan, osasun-arloan eta hezkuntzan.
3. Aplikazioak	• Isolamendu sozialeko testuinguruetarako txatbot enpatikoak, etengabeko laguntasuna eta babes emozionala eskaintzeko. • Osasun mentalaren jarraipena egiteko txatbot enpatikoak, buruko nahasmen-duak dituzten pazienteengandik datu errealak jasotzea errazteko. • Estres handiko egoeretarako txatbot enpatikoak, hala nola bezeroaren arretan, hezkuntza-prozesuetan edo industria-eremuetan, erabiltzailearen egoera emozionalari hobeto egokitutako erantzunak eskaintzeko eta, horrela, frustrazioa murrizteko. • Bideo-jokoetarako txatbot enpatikoak, interakzio naturalago eta zirrargarriagoen bidez jokalariaen esperientzia hobetzeko, jokoaren murgilketa eta errealismoa aberastuz.
4. Heldutasun Teknologikoko Maila (TRL)	Txatbot enpatikoen heldutasun teknologiko maila aldatu egiten da sistema osatzen duten osagai zehatzen arabera. Teknologia honek I+G diziplina anitz biltzen ditu, bakoitza garapen-fase desberdinetan. Gainera, erabilera-kasu batzuk errazagoak dira inplementatzeko, hala nola bezeroaren arretarako txatbotak; aldiz, edozein gairi buruzko elkarrizketa irekiak mantentzeko gai diren txatboten garapenak konplexutasun nabarmen handiagoa du. Jarraian, osagai nagusien eta dagozkien TRL-ak aurkezten dira: • Transkripzio automatikoa: TRL 8-9 • Ahotsean emozioak antzematea: TRL 5-7 • Elkarrizketa ulertzea eta erantzun enpatikoak sortzea: TRL 3-5 • Ahots-sintesi emozionala: TRL 4-6

OINARRIZKOAK

- **Audio- eta testu-corpus lerrokatua:** Hizkuntza eta eremu esanguratsuetako audio kopuru handiak, dagozkien transkripzioekin, eredu akustiko bat edo ertzetik ertzerako STT eredu bat sortu edo egokitzeke.
- **Ahots emozionalen datu-corpusa:** Emozio ezberdinekin etiketatutako ahotsen grabazioak, ahotsaren bitarteko ezagutza emozionalerako ereduak trebatzeko edo egokitzeke. Corpusa transkribatuta badago, testu-informazioa ere aprobe txatu daiteke helburu horretarako.

Txatbot motaren arabera, honako hauetako bat beharrezkoa izango da:

- **Arauetan oinarritutako elkarrizketa-sistema:** Egokiagoak dira elkarrizketa-fluxu mugatueterako, hala nola bezeroaren arretarako, non txatbotak zeregin kopuru jakin bat egiteko gai diren.
- **Elkarrizketen corpusa eta aurreikasitako LLMa:** Informazio emozionalarekin etiketatutako interakzio-multzoak, aurreikasitako LLM baten portaera doitzeke edo egokitzeke.

BESTE BATZUK

- **Testu-corpusa:** Hizkuntza eta eremu esanguratsuetako testu kantitate handiak, hizkuntza-eredu bat sortzeko edo egokitzeke.
- **Ahoskera-hiztegia:** Aplikazio-hizkuntzako fonemak hitzekin lotzeko mapa.
- **Letra larrietarako eta puntuazio softwarea:** Transkripzioak aberasteko, irakur-garriagoak eta naturalagoak izan daitezen.

**5.
Zein baliabide
edo tresna
behar dira?**

1. **Pribatutasuna eta segurtasuna:** Emozioen tratamendua oso gai sentsiblea da, emozioak erabiltzaileak abusatu, manipulatu edo modu desegokian esplotatzeko erabil daitezkeelako. Ondorioz, ezinbestekoa da tratatutako datuen babesa bermatzea eta dagokion araudia betetzea (adibidez, EB-ren AA Legea eta DBEO-a), nahiz eta horrek teknologiaren garapena kaltetu edo oztopatu.
2. **Kalitatezko audioen bilketa:** Oinarrizko gailuekin eta ingurune zailetan egindako grabaketetan, audioaren kalitatea okertu ohi da. Horrek negatiboki eragiten dio hizketa-ezagutzaren zehaztasunari. Kalitate handiko banakako mikrofonoak erabiltzea konponbide tekniko eraginkorra da; hala ere, konponbide honek sistemaren erabilgarritasuna eta adopzioa asko mugatzen ditu.
3. **Hizkera espontaneo:** Errepikapenen, betegarrien eta etenaldien gisako fenomenoak ohikoak dira eguneroko elkarrizketetan. Fenomeno horiek STT sistemak nahas ditzakete, horiek behar bezala kudeatzeko diseinatuta ez badaude.
4. **Hizkuntza-aldaeren kudeaketa:** Tresnaren erabiltzaileek dialekto, azentu eta hizkera desberdinak erabil ditzakete, eta horrek ahotsa automatikoki ezagutzea zein itzultzea zaildu egiten du. Erronka honi aurre egiteko, osagai desberdinak hizkuntza-aldaera horien adierazgarri diren datuekin trebatu behar dira, nahiz eta datu horiek biltzea eta etiketatzea garestia eta nekeza izan.
5. **Zereginaren subjektibotasuna:** subjektibotasuna emozioak interpretatzeari eta erantzun enpatikoak sortzeari bete-betean datxekio. Askotan, gizakientzat ere interpretatzeko zailak dira emozioak. Horrek ataza modelatzeko zehaztasunari eragiteaz gain, haren ebaluazioa zailtzen du.
6. **Denbora erreala eta latentzia:** Denbora errealeko aplikazioetan, erabiltzailearen txandaren transkripzioa eta azterketa, baita erantzunen sorrera eta sintesia, azkar sortu behar dira ahalik eta latentzia txikienarekin. Horrek prozesamendu-katearen eta hura osatzen duten modulu ezberdinen integrazioa optimizatzea eskatzen du, prozesamendu-abiadura hobetzeko emaitza-aren kalitatea kaltetu gabe.

**6.
Erronka
zientifiko-
teknikoak**

Ahotsaren klonazioa

Teknologia nagusia: TTS

Eragin handiko aplikazioa: Ahozko urritasuna duten pertsonen ahots sintetiko pertsonalizatuak eskaintzea

Zein beharri erantzuten dio?	1. Ahots-klonazioa ahots jakin baten ezaugarriak ahots sintetiko batean erreplikatzeko datza, testutik ahotserako (TTS) tekniken bidez lortutakoa. Teknologia honek ahots sintetikoaren aukerak zabaltzea eta ahots-multzo txiki bat erabili behar izatearen mugak gaitzartzea ahalbidetzen du. Behar hori bereziki garrantzitsua da komunikazio alternatibo eta handigarriaren gailuen erabiltzaileentzat, zeintzuek, normalean, ahots sintetikoaren katalogo oso murritzua izaten baitute.
Erabiltzaileak	2. Ahots sintetiko pertsonalizatuaren erabilera aplikazio ugari ditu hainbat eremutan. Gizarte-eragin handiko kasu bat da ELA diagnostikatu zaien pertsonen ahots pertsonalizatuak erabiltzea . Gaixotasun horretan, hitz egiteko gaitasuna pixkanaka galtzen da eta gaixoei komunikazio handigarri eta alternatiboko gailuak erabiliko dituzte beren eguneroko bizitzan. Ingurune sozialarekin komunikazioa mantentzea errazteko, ezaugarri egokietako ahots bat erabiltzea funtsezkoa da. Gaixoei diagnostikoa jasotzen dutenean, oraindik hitz egiteko gaitasuna izaten dute, eta, horri esker, teknologia honetaz baliatu daitezke garaiz grabazio multzo mugatu bat egiten badute.
3. Aplikazioak	• 3. erakuslearekin batera (ahotsaren momentuko itzulpena), pertsona baten hizketa beste hizkuntza batean itzuli eta erreproduzitzea ahalbidetzen du, jatorrizko ahotsaren ezaugarri bakarrak mantenduz. • Pertsonaia historikoen ahotsak berreskuratzea edo gordetzea, hezkuntza- eta museografia-erakusketetarako.
4. Heldutasun Teknologikoko Maila (TRL)	Ahots-klonazio sistema baten heldutasun teknologikoko maila aldatu egiten da sistema osatzen duten osagai zehatzen arabera. Hizkera neutroaren (emoziorik gabea) eta hizkera ez-dialektalaren sintesi pertsonalizatuaren alorrean, TRL 8 edo 9 mailaz hitz egin dezakegu. Pertsonalizazio aurreratuagoa duten sistemetarako, TRL 6 mailan gaude.

OINARRIZKOAK

- **Grabaketa-ekipamendua:** Grabatzeko gailuak eta klonatu beharreko ahotsaren laginak lortzeko tresnak.
- **Oinarritzko TTS ereduak:** Eskuragarri dauden grabazioekin entrenamendu eta egokitzen ahalbidetzen duten testu-ahots bihurketarako eredu irekiak.
- **Entrenamendu-corpora:** Kalitate handiko audio-grabazioak, eduki eta hiztun aniztasunarekin, ereduak egokitzeko.
- **Hedapenerako tresnak:** Eramangarrietan exekutatu daitezkeen ereduak lortzeko eta ohiko sistema eragileekin (Windows, Android) integratzeko softwarea.

5.
Zein baliabide
edo tresna
behar dira?

1. **“On-device” hedapena:** Teknologia gehienek ez dute tokiko gailuetan funtzionatzen, eta, hodeian badaude ere, ez dira bateragarriak ezgaitasuna duten pertsonen ohiko tresnekin.
2. **Laginen bilketa:** Kasu askotan ezin dira lortu klonatu nahi den ahotsaren laginak, hitz egiteko gaitasuna dagoeneko galduta dagoelako. Kasu hauetan, hainbat alternatiba azter daitezke, ahozko ezgaitasunaren graduaren eta motaren arabera:

- Erabateko galera kasuetan: aurretiko grabazioak berreskuratzea (bideoak, WhatsApp mezuak, elkarrizketak...). Kasu horretan, laginen kalitateak ahots sintetikoaren azken kalitatea muga dezake.
- Laginik gabeko erabateko galera kasuetan: antzeko ahotsak bilatu eta haien ezaugarriak aldatzen dituzten ahots-bihurketa teknikak aplikatzea.
- Disartriadun galera partzialeko kasuetan: norberaren ahotsaren ulergarritasuna hobetzeko bihurketa-teknikak aplikatzea.

Ahotsaren klonazioak, oro har, TTS sistema oinarritzko bat egokitzen du helburu-ahotsaren laginak erabiliz. Sistema horrek grabazio kopuru handiak behar izaten ditu, eta horiek ez dira hizkuntza guztietan eskuragarri egoten. Hori dela eta, beharrezkoa da bai oinarritzko sistemarako bai klonazio zehatz baterako eskakizunak minimizatuko dituzten teknikak garatzea.

3. **Hizkuntza gutxituak:** TTS sistemak batez ere hizkuntza nagusietan zentratzen dira, hala nola ingelesa, gaztelania eta txinera. Eskuragarritasuna txikiagoa da euskara edo galiziera bezalako hizkuntza gutxituetan, eskari txikiagoa eta baliabideak lortzeko zailtasuna direla eta. Euskalkien aniztasunak, ahotsaren nortasun eta izaerarako garrantzitsuak direnak, erronka gehigarria suposatzen dute.
4. **Ahotsaren adierazkortasuna:** Gaur egungo TTS sistemek mugak dituzte adierazkortasuna eta emozioa kontrolatzeko. Hizketaren naturaltasunak barre eta eztul bezalako elementuak barne hartzen ditu, oraindik behar bezala txertatu ez direnak.
5. **Pribatutasuna eta segurtasuna:** Ahotsa elementu biometrikoa denez, haren klonazioa iruzurrez erabil daiteke ahots-identifikazioko sistemetan edo telefono bidezko komunikazioetan. “Anti-spoofing” deritzen tekniken garapena, hau da, ahots sintetikoak eta giza ahotsak bereizteko teknikena, ikerketa-arlo aktibo bat da.

6.
Erronka
zientifiko-
teknikoak

Erabiltzailean zentratutako itzultzailea

Teknologia nagusia: IA

Eragin handiko aplikazioa: Justizia eta osasun arloetako dokumentazio espezializatuaren itzulpen automatikoa, terminologia eta estilo egokitzapen dinamikoarekin

Zein beharri erantzuten dio?	1. Itzulpen automatikoak (IA) aurrerapen nabarmena izan du azken urteetan, eta teknologia oso eskuragarria bilakatu da. Hala ere, gaur egungo sistemak gehienbat orokorrak dira eta mugak dituzte ingurune profesional espezializatuetan. Erabiltzailean oinarritutako itzultzaile automatikoak erronka hauetako bat edo gehiago lantzen ditu, domeinu espezifikoetara egokitzean, erabiltzaileen datu errealean bidezko etengabeko hobekuntzan, eta/edo itzulpenaren kalitatearen estimazio zehatzetan arreta jarritz . Hurbilketa honek IArein erabilgarritasuna eta fidagarritasuna handitu nahi ditu zehaztasuna eta terminologia zein estilo-egokitzapena funtsezkoak diren testuinguru profesionaletan.
Erabiltzaileak	2. Kalitate eta espezifikotasun handiko itzulpenak behar dituzten profesionalak eta erakundeak dira teknologia honen onuradun nagusiak. Osasuna eta justizia bezalako arloetan, non zehaztasuna funtsezkoa den, proposatutako hobekuntzen eragina bereziki nabarmena litzateke. Dokumentu espezializatu askoren itzulpenari aurre egiteko gaitasuna duen teknologia da, besteak beste, erregistro klinikoak, epai judizialak, artikulak zientifikoak, merkataritza-kontratuak, patenteak, finantza-txostenak eta gaurkotasuneko kazetaritza edukiak.
3. Aplikazioak	• Eredu generikoak eredu zehatzetara egokitzea terminologia espezializatuaren txertaketaren bitartez. • Erabiltzaileen feedbacketik denbora errealean ikastea ereduak arin eta etengabe hobetzeko. • Ereduak pertsonalizatzea, erabiltzaileen historia eta hobespenak erabiliz. • Kalitatearen estimazio automatikoa, orraztua behar duten segmentuak identifikatzeko.
4. Heldutasun Teknologikoko Maila (TRL)	Sistemak hainbat teknologia integratzen ditu, heldutasun-maila desberdinekin, nabarmen alda daitezkeenak baliabide ugari dituzten hizkuntzen eta baliabide mugatuak dituztenen artean: • Itzulpen automatikoaren nukleoa: TRL 8-9 • Domeinu espezifikoetara egokitzea eta terminologia txertatzea: TRL 5-7 • Erabiltzaileen feedbacketik denbora errealean ikastea: TRL 4-5 • Kalitatearen estimazio automatikoa: TRL 4-6

OINARRIZKOAK

- **Itzulpen automatikoaren oinarrizko eredua:** aurreikasitako itzulpen automatikoko eredu bat, gainerako egokitzenetarako oinarri gisa erabiliko dena (hala-berriak ezean, hizkuntza desberdinetan lerrokatutako testu kopuru handiak, eredu bat errentatzeko).
- **Corpus paralelo espezializatua:** Aplikazio-eremuko testu-kopuru handiak, hainbat hizkuntzatan lerrokatuta, oinarrizko eredu egokitzeko eta garapen desberdinen ebaluaziorako.
- **Glosario edo terminologia-base espezializatuak:** Erabiltzailearentzat interesgarriak diren terminoen bildumak eta haien itzulpenak hizkuntza-pare garrantzitsuetan, termino-zerrenda edo terminologia-biltegi egituratu gisa.
- **Posteditaturiko itzulpenen corpusa:** Gizakiek eskuz zuzendu dituzten itzulpen automatikoen bilduma. Baliabide hau kalitatea balioesteko sistemak zein online ikaskuntza-fluxuak garatzeko erabiltzen da.

**5.
Zein baliabide
edo tresna
behar dira?**

1. **Ahantura katastrofikorik gabeko egokitzena:** Ahantura katastrofiko gertatzen da ereduak, datu edo domeinu berrietara egokitzen, bat-batean galtzen duenean aurretik menderatzen zituen itzulpenak zuzen egiteko gaitasuna. Beraz, erroka da online ikaskuntza-teknikak garatzea, ereduari ezagutza berria txertatzea eta egoera berrietara egokitzea ahalbidetuko diotenak, aurretik gaitututako testuinguruetan bere errendimendua kaltetu gabe.
2. **Testuinguru luzeetan koherentzia modelatzea:** Egungo itzulpen automatikoko ereduak esaldi edo paragrafo laburren mailan ondo funtzionatzen dute, baina askotan ez dute elementu narratiboak testu luzeagoetan koherenteki atzemateko eta mantentzeko gaitasuna. Arazo hori bereziki kritikoa da aurretik aipatutako inguruneetan, non itzuli beharreko dokumentuak (adibidez, txosten klinikoak, patenteak) luzeak eta konplexuak izan daitezkeen.
3. **Hutsezintasunaren espektatiba:** Teknologia honen aplikazio eszenatoki askoren kritikotasuna dela eta (ikus Erabiltzaileak eta Aplikazioak), errendimendua ia perfektua espero da. Itzulpeneko akatsek ondorio larriak izan ditzakete osasun-arreta edo izapide juridikoak bezalako testuinguruetan eta, post-edizioa ezinezkoa denez, honek presio nabarmena sortzen du zehaztasun oso altua lortzeko.

**6.
Erronka
zientifiko-
teknikoak**

Testu egokitzailea

Teknologia nagusia: HP (LLMak)

Eragin handiko aplikazioa: Administrazio-dokumentuak eta lege-testuak automatikoki egokitzea irakurketa errazeko formatuetara, dibertsitate funtzional kognitiboa duten pertsonentzat

1. Zein beharri erantzuten dio?	Testuen egokitzapenak informazioa publiko desberdinentzat eskuragarriago eta ulergarriago egiteko beharrari erantzuten dio. Teknologia horrek eduki konplexuak bertsio sinpleagoetan eraldatzen ditu, laburtu edo hizkuntza argira egokitu, alfabetatze maila, espezializazio, gaitasun kognitibo edo behar espezifikoak dituzten pertsonen informazioa ulertzea eta eskuratzea errazteko.
2. Erabiltzaileak	Teknologia horrek gizartearen espektro zabal bati mesede egiten dio, besteak beste, ikasleei, adinekoei, urritasun kognitiboak dituzten pertsonen eta, orokorrean, testu konplexuei aurre egin behar dieten publiko orokorrari.
3. Aplikazioak	• Ikasmaterialen sinplifikazioa edota laburpena, gai konplexuen ulermena hobetzeko. • Osasun informazioaren egokitzapena, profesionalen eta pazienteen arteko komunikazioa errazteko. • Lege-dokumentuen sinplifikazioa, justiziarako sarbidea hobetzeko. • Albisteen eta gaurkotasun edukien bertsio egokituak sortzea, gaitasun kognitibo anitzak dituzten pertsonentzat. • Dokumentu luzeen laburpenak sortzea, ingurune profesionaletan eraginkortasuna hobetzeko. • Testu egokitzapena, belaunaldien arteko komunikazioa errazteko.
4. Heldutasun Teknologikoko Maila (TRL)	Heldutasun teknologikoaren maila aldatu egiten da modulu zehatzen arabera. Baldintza onenetan, maila hauek lortu dira: • Laburpena: TRL 7-8 • Sinplifikazioa: TRL 6-7 • Hizkuntza argira egokitzea: TRL 5-6 Baliabide gutxiko hizkuntzetarako, TRL orokorra 2ra edo 3ra jaitsi daiteke datu eta tresna espezifikoaren eskuragarritasun txikiagoagatik.

OINARRIZKOAK

- **Hizkuntza Eredu Handia (LLM):** Testuak egokitzeko lanetarako oinarria dira. Eredu hauek, xede-hizkuntzan aurreikasitakoak, testua modu aurreratuan ulertzen eta sortzen dute. Zeregin zehatzetarako egokitu daitezke, hala nola sinplifikazioa edo hizkuntza argira egokitzea, horien hizkuntzaren ezagutza orokorraz eta testu konplexuak ulertzeko gaitasunaz baliatzeko.
- **Erreferentzia corpora:** Xede-hizkuntzan dauden testu paraleloen (jatorrizkoak eta egokituak) bildumak dira. Gainera, corpus hauek xede-publikoaren adierazgarriak izan behar dute, erabiltzaile-talde ezberdinen beharrak islatzen dituzten konplexutasun, gai eta estilo aldaerak barne.

BESTE BATZUK

- **Hiztegiak eta tesauroak:** Hiztegiaren sinplifikazioa eta egokitzapena erraztu dezaketen baliabide lexikalak eskaintzen dituzte. Laburpena bezalako lanetarako guztiz beharrezkoak ez diren arren, sinonimo sinpleagoak aurkitzeko edo termino konplexuak azaltzeko erabilgarriak izan daitezke.

**5.
Zein baliabide
edo tresna
behar dira?**

1. **Esanahia gordetzea:** Hizkuntza ereduak jatorrizko testuan ez dagoen informazioa sor dezakete, eta horrek esanahia zaintzea arriskuan jartzen du. Erronka erabakigarria da “haluzinazio” horiek saihestea testu baten sinplifikazioan, egokitzapenean edo laburpenean jatorrizko mezuaren fideltasunari eusteko. Gainera, egungo ereduak askotan koherentzia eta kohesioa galtzen dute testu luzeak prozesatzen dituztenean, ereduak diseinuz dituzten mugak direla eta. Horrek dokumentu luzeetan ideien egitura logikoa eta harremanak mantentzea zailtzen du.
2. **Domeinu espezializatuak:** Testu teknikoak egokitzeak aplikazioaren domeinuaren ezagutza sakona eskatzen du, zehaztasuna mantentzeko, sinplifikatzen edo laburtzen duen bitartean. Erronka ereduak medikuntzaren, zuzenbidearen edo teknologiaren arlo zehatzetan adituak diren ezagutzak eskuratu eta aplikatzeko duten zailtasunetik dator.
3. **Hainbat xede-publikotara egokitzea:** Testu-egokitzapenaren eraginkortasuna nabarmen aldatzen da irakurlearen ezaugarrien arabera (adina, heziketa-maila, gaitasun kognitiboak, etab.). Egungo sistemek modu uniformearen sinplifikatzeko joera dute, xede-publiko ezberdinen behar espezifikoak kontuan hartu gabe, erabiltzaile batzuentzat sinpleegiak eta besteentzat are konplexuagoak izan daitezkeen egokitzapenak sortuz, egokitutako testuen horien erabilgarritasuna eta irisgarritasuna mugatuz.
4. **Ebaluazio eraikitzailea:** Testu egokitzapen “onaren” izaera subjektiboak zaildu egiten du informatiboak eta eraikitzaileak diren neurri automatiko hutsak sortzea. Erronka xede-publikoarekiko argitasuna, fideltasuna eta egokitasuna bezalako alderdi kualitatiboak jasotzen dituzten neurriak ezartzean datza.

**6.
Erronka
zientifiko-
teknikoak**

Bilatzaile propioa

Teknologia nagusia: HP (LLMak); aukeran, ordenagailu bidezko ikusmena

Eragin handiko aplikazioa: Jurisprudentzia eta aurrekarien bilatzaile eleaniztuna

1. Zein beharri erantzuten dio?	Hizkuntza-eredu handietan (LLM) oinarritutako bilatzaile pertsonalizatu batek erakundeetako barne-informazio bilduma heterogeneo handiak modu eragin-korreetan atzitzeko eta prozesatzeko beharrari erantzuten dio, bai testua bai multimedia edukia barne hartuz. Bilatzaile tradizionalak nabarmen hobetu badira ere, askotan ez dute testuinguruaren ulermenik ezta gaitasunik konplexuak, anbiguoak edota hizkuntza desberdinetan idatzitako kontsultak kudeatzeko. Teknologia honek LLM-en aurrerapenak aprobetxatzen ditu bilaketa zehatzagoak eta esanguratsuagoak eskaintzeko, kontsulten testuingurua eta asmoa ulertuz, eta hori dena hainbat hizkuntzatan.
2. Erabiltzaileak	Enpresak, ikerketa-erakundeak, gobernu-agentziak, lege- eta osasun-sektoreak izan daitezke erabiltzaile nagusiak, bereziki dokumentazioa eta datuen bilaketa eraginkorra behar duten langile, proiektu-kudeatzaile eta analistarentzat.
3. Aplikazioak	• Laguntzaile juridikoak, jurisprudentzia eta aurrekarietarako sarbide azkar eta zehatza errazten dutenak, zuzenbideko profesionalei beren argudioak oinarritzeko informazio juridiko bolumen handiak modu eraginkorreetan aztertzea ahalbidetzeko. • Diagnostikoan laguntzeko, osasun-profesionalek antzeko kasu klinikoak aurkitzeko eta alderatzeko, espediente mediko garrantzitsuak atzitzeko eta aurrekariak aztertzeko aukera emanez. • Hedabide eta kazetaritzan ezagutza kudeatzeko, multimedia- eta testu-artxiboetarako sarbidea bateratuz, kazetaritza-ikerketa eta albisteak testuinguruak jartzea erraztuz. • Informazio korporatiboa berreskuratzeko, denboran zehar pilatutako lan, proiektu eta barne-txostenen bilaketa eta sarbidea optimizatuz, taldeen arteko elkarlana hobetuz eta ahaleginen bikoizketak saihestuz. • Merkataritza-prozesuetan laguntzeko, aurrekontu eta negozio-proposamenen lanketa azkartuz aurreko eskaintza eta produktu edo zerbitzuen datu historikoetarako sarbide azkarraren bidez, horrela salmenta eta marketineko erabaki estrategikoak hartzea erraztuz.
4. Heldutasun Teknologikoko Maila (TRL)	Hizkuntza-eredu handietan (LLM) oinarritutako bilatzaile pertsonalizatu baten heldutasun teknologiko-maila nabarmen aldatzen da hizkuntza- eta aplikazio-testuinguruaren arabera. LLMen oinarri-teknologia nahiko finkatuta dago baliabide ugariako hizkuntzetarako, 7tik 8rako TRLa lortuz garapen baldintza onenetan. Hala ere, eredu hauek baliabide gutxiko hizkuntzetara egokitzean, hala nola euskara edo galizierara, edo arlo juridikoa edo medikoa bezalako domeinu oso espezializatueta, baita eremu hauetako ikus-analisi zeregin espezifikoetara ere, TRLa 2 edo 3ra jaitsi daiteke.

OINARRIZKOAK

- **Hizkuntza-eredu handia (LLM):** Testuen eta kontsulten errepresentazio bektorialak sortzen dituen osagai nagusia da. Multimodala bada, eduki bisuala ere prozesatzen du; eleaniztuna bada, hizkuntza desberdinen arteko bilaketak ahalbidetzen ditu.
- **Berezko datu-corpora:** Erakundearen berariazko dokumentu eta datuen bildumak, bilaketak egiteko ezagutza-basea sortzeko. Gainera, testu-bilduma honek hizkuntza-eredua egokitzeko balio du, xede-sektoreari dagokion hizkuntza tekniko eta espezializatuaren ezagutza hobetzeko.
- **Formatudun dokumentuak prozesatzeko tresnak:** PDF, DOCX edo PPT bezalako formatuetan dauden dokumentuak testu lau bihurtzen dituzte, hizkuntza-ereduek haien bektoretzea eta indexazioa egin ahal izateko.
- **Bektoreen indexazio eta berreskuratze motorea:** Dokumentuen irudikapen bektorialak indexatu eta berreskuratzen dituen sistema, datu-bolumen handietan bilaketa semantiko eraginkorrak ahalbidetuz.

BESTE BATZUK

- **Ordenagailu bidezko ikusmen eredua:** Multimodala ez den LLMen osagarria da, eduki bisualak aztertze eta prozesatzeko. Eredu hori irudien eta bideoen errepresentazio bektorialak kalkulatzeko arduratzen da, horiek indexatu eta bilatu ahal izateko.
- **Irudi eta bideoak aurreprozesatzeko tresnak:** Fotograma garrantzitsuak erauzteko, testudun irudietan OCR egiteko eta irudiak analisirako prestatzeko softwarea. Bereziki garrantzitsuak dira LLM ez-multimodalekin lan egingo bada.

5. Zein baliabide edo tresna behar dira?

1. **Eraginkortasuna eta eskalagarritasuna:** Denbora errealeko bilaketetan erantzun-denbora azkarrak mantentzea erronka bat da, LLMek baliabide asko kontsumitzen baitituzte. Sistemaren errendimendua optimizatu behar da aldi bereko kontsultak eta etengabe metatzen ari diren datuak kudeatzeko.
2. **Etengabeko eguneratzea:** Sistema birkonfigurazio osorik gabe eguneratu behar da, LLMa modu inkrementalean doituz datu eta testuinguru berriak ikasteko, baina era ntzunen garrantzia mantenduz datuetan ("data drift") edo kontzeptuetan ("concept drift") egon daitezkeen aldaketen aurrean.
3. **Dokumentu konplexuen kudeaketa:** Dokumentu luze eta askotarikoak semantikoki indexatzea funtsezkoa da bilatzailearen eraginkortasunerako. Pasaite eleaniztunak, taulak, zerrendak eta oharrak errepresentatzeko metodoak eskatzen ditu horrek, emaitza garrantzitsu eta esanguratsuak berreskuratze.
4. **Kalitate baxuko dokumentuak:** Sistemak kalitate eta kontserbazio-egoera desberdinetan dauden dokumentuak kudeatu behar ditu, mekanizatutako edo eskuz idatzitako dokumentuen argazki edo eskaneatuak barne. Funtsezkoa da baldintza eskasetan irudi-prozesamendu sendoa izatea.
5. **Integrazio eleaniztun eraginkorra:** Bilaketa-sistema eleaniztun baterako, hizkuntza guztietan ulermen eta prozesamendu homogeneoak bermatu behar dira, erantzun zehatz eta koherenteak emanez, kontsulten edo dokumentuen hizkuntza edozein dela ere.
6. **Modalitateen arteko lerrokatze semantikoa:** Funtsezkoa da testua eta irudiak erlazionatzeko errepresentazio semantiko bateratua sortzea, bilaketa gurutatuak ahalbidetzeko (hau da, testutik abiatuta irudiak aurkitzeko, edo alderantziz), ulermen multimodal eta espazio bektorial partekatuen bidez.

6. Erronka zientifiko-teknikoak

Prozesu bulegotikoen automatizazioa

Teknologia nagusia: HP eta ordenagailu bidezko ikusmena

Eragin handiko aplikazioa: Lizitazio publikoetako dokumentu-kudeaketa eleaniztunaren automatizazioa hizkuntza koofizialak dituzten erkidegoetan

1. Zein beharri erantzuten dio?	1. Bulegotika prozesuen automatizazioak negozio- eta erakunde-inguruneetako eraginkortasun- eta produktibitate-eskariari erantzuten dio. Teknologia horrek administrazio-zeregin errepikakorrak optimizatu eta erraztu nahi ditu, giza akatsak murriztuz eta balio erantsi handiagoa duten jardueretarako denbora askatuz. Haren ezarpenak informazioaren kudeaketa eraginkorragoa, erabakiak hartzeko hobekuntzak eta erakundearen beharrei erantzun azkarragoa ematea ahalbidetzen du.
2. Erabiltzaileak	2. Teknologia hau erabiltzaile talde zabal bati dago zuzenduta, tamaina guztietako enpresei, gobernu-erakundeei, hezkuntza-erakundeei eta datu eta dokumentazio-bolumen handiak maneiatzen dituzten edo logistika kudeaketan parte hartzen duten profesional independenteei barne. Prozesu bulegotikoen automatizazioak aplikazio oso desberdinak biltzen ditu, bi bloke teknologiko nagusitan taldeka daitezkeenak:
3. Aplikazioak	• Erabiltzaile edo aplikazio-sektoreari dagozkion dokumentu-moten sailkapen automatikoa (adibidez, fakturak, albaranak eta ordainagiriak, baina baita curriculumak, hezkuntza-tituluak, izena emateko inprimakiak, etab.) • Dokumentu horietan jasotako datu eta metadatu zehatzen erauzketa automatikoa, dagozkien datu-baseetan automatikoki isurtzeko eta dagozkien prozesuak automatikoki abiarazteko.
4. Heldutasun Teknologikoko Maila (TRL)	4. Hizkuntza-eredu handietan (LLMetan) oinarritutako aplikazio mota honetarako heldutasun-maila nabarmen aldatzen da hizkuntza- eta aplikazio-testuinguruaren arabera. LLMen oinarritzko teknologia nahiko finkatuta dago baliabide asko dituzten hizkuntzetarako, 7tik 8ra bitarteko TRL maila lortuz egotira mesedegarrienetan. Hala ere, eredu hauek baliabide gutxiko hizkuntzetara egokitzean, hala nola euskara edo galizierara, edo arlo juridikoa edo medikoa bezalako domeinu oso espezializatueta, baita eremu hauetako irudien analisi zeregin espezifikotara ere, TRLa 2 edo 3ra jaitsi daiteke.

OINARRIZKOAK

- **Hizkuntza Eredu Handia (LLM):** Hau da sistemari dokumentuen errepresentazio matematikoak prozesatzeko eta sortzeko gaitasuna ematen dion osagaia. Multimodala bada, irudi eta bideoen errepresentazioak ere prozesatu eta sortzake. Eleaniztuna bada, dokumentuak hainbat hizkuntzatan automatizatzeko aukera emango du.
- **Bulegotikako dokumentuen corpora:** Erakundearen dokumentu eta datu espezifikoek bildumak, automatizatu beharreko bulegotika-prozesuetan erabiltzen diren dokumentu-formatu eta moten aniztasuna irudikatzen dutenak (fakturak, inprimakiak, txostenak, etab.). Corpus hori bai hizkuntza-eredua domeinu espezifikora egokitzeko bai sistema probatu eta ebaluatzeko erabiltzen da.
- **Formatudun dokumentuak prozesatzeko tresnak:** Sarri, erakundeek beren eza-gutza gizakiek irakurtzeko dokumentuetan gordetzen dute, hala nola PDF, DOCX edo PPT formatuetan. Formatu horiek ez dira zuzenean prozesagarriak testuan oinarritutako hizkuntza-ereduentzat; beraz, beharrezkoa da testu laua erauzten duten tresnekin prozesatzea, gero gehiago landu ahal izateko.

BESTE BATZUK

- **Ordenagailu bidezko ikusmen eredia:** Multimodala ez den LLMen osagarria da, eduki bisualak aztertzeko eta prozesatzeko. Eredu hori irudien eta bideoen errepresentazio bektorialak kalkulatzeko arduratzen da, horiek sailkapen edo informazioa erauzteko algoritmoen bidez automatikoki prozesatu ahal izateko.
- **Irudi eta bideoak aurreprozesatzeko tresnak:** Fotograma garrantzitsuak erauzteko, testudun irudietan OCR egiteko eta irudiak analisirako prestatzeko softwarea. Bereziki garrantzitsuak dira LLM ez-multimodalekin lan egingo bada.

**5.
Zein baliabide
edo tresna
behar dira?**

1. **Eraginkortasuna eta eskalagarritasuna:** Denbora errealeko bilaketetan erantzun-denbora azkarrak mantentzea erronka bat da, LLMek baliabide asko kontsumitzen baitituzte. Sistemaren errendimendua optimizatu behar da aldi bereko kontsultak eta etengabe metatzen ari diren datuak kudeatzeko.
2. **Pribatutasuna eta segurtasuna:** Izen, identifikazio-zenbaki, datu mediko edo enpresen informazio konfidentziala bezalako informazio sentsiblea dagoenez, entrenamendurako beharrezkoak diren datu-corpus errealean sarbidea mugatua edo ezinezkoa izan daiteke. Erronka honek pribatutasuna arriskuan jarri gabe datu adierazgarriak lortu eta erabiltzeko estrategiak garatzea dakar, agian datu sintetikoak sortuz edo datu errealak anonimizatuz.
3. **Formatu eta edukien aldakortasuna:** Sistemak gai izan behar dute kategoria bereko dokumentuetatik informazio garrantzitsua prozesatu eta erauzteko, nahiz eta egitura eta eduki oso desberdinak izan (adibidez, fakturen formatu desberdinak konposizio, kolore, letra-tamaina eta abarren arabera).
4. **Kalitate baxuko dokumentuak:** Sistemak kalitate eta kontserbazio-egoera desberdinetan dauden dokumentuak kudeatu behar ditu, mekanizatutako edo eskuz idatzitako dokumentuen argazki edo eskaneatuak barne. Funtsezkoa da baldintza eskasetan irudi-prozesamendu sendoa izatea.

**6.
Erronka
zientifiko-
teknikoak**

ONDORIOAK

Atal honetan, ikerketaren ondorioak aurkezten dira, elkarrizketa eta galdetegi bidez jasotako datu eta iritzien laburpen gisa. Azterketak agerian uzten du nola erabiltzen eta baloratzen diren gaur egun Hizkuntza Teknologietan oinarritutako tresnak. Gure ikerketak bi ardatz nagusi izan ditu: batetik, tresna hauen egungo eta izan litezkeen erabilerak aztertzea, eta bestetik, erabiltzaileek dituzten behar eta itxaropenak identifikatzea.

Hona hemen aztertutako hiru atal teknologikoen ondorio nagusiak:

Itzulpen automatikoa

- Euskarazko itzulpen automatikoak asko aurreratu du, oinarritzko beharrak betetzeko moduko mailara iritsi arte. Horregatik, elkarrizketetan batik bat garapen berrietan eta eremu espezializatuertarako alderdi garrantzitsuetan jarri da arreta. Parte-hartzaileen esanetan, gaur egungo erronka nagusia da zerbitzuaren doitasuna eta kalitatea hobetzea, ahal den neurrian akatsak murriztuz.

Ahotsaren prozesamendua

- Teknologia honek protagonismo nabarmena izan du elkarrizketa gehienetan. Gizakion berezko gaitasuna den heinean, etorkizuneko garapenertarako funtsezko elementu gisa aitortu da. Emaitez erakusten dutenez, teknologia aurreratuago dago jendeak uste edo erabiltzen duena baino, baita euskararen kasuan ere. Aipagarria da arlo honetan bai erabiltzaileek bai Hizkuntza Teknologietako adituek gaur egungo prototipo eta dagoeneko konponduta dauden beharretan jarri dutela arreta. Erabiltzaile eta garatzaileen arteko lankidetzaproiektu ugariak bi taldeen ikuspegiaren arteko aldeak leundu dituzte.

Hizkuntza eredu handiak (LLM)

- Erabiltzaileen alorrean, oraindik ere urria da teknologia hauen berezko inplementazioa, hala enpresa eta zerbitzu-erakundeetan nola sektore publiko eta akademikoan. Gaur egun, norberaren erabileran zentratzen da batik bat, Copilot eta ChatGPT bezalako tresnak nagusi direla. Hala ere, argi dago denek ikusten dutela etorkizunerako teknologia giltzarri modura. Parte-hartzaileek erabilera esanguratsu ugari ikusten dizkiote: hainbat tresnatan hardwarea automatizatu eta txertatzeko aukeretatik hasi, eta askotariko erabileretarako irtenbideak ezartzera arte. Nabarmendu dute, halaber, erakundeei balio erantsia emateko duen ahalmena, bai komunikazioa hobetzeko, bai produktuen diseinuan bertan, baita helburu estrategiko askotarikoetan erabiltzeko ere. Garapen-ahalmen handiko eremua dela uste dute, nahiz eta onartu badirela oraindik erronka handiak gaingidzeko.

Hona hemen ondorio orokor batzuk:

1. Itzulpen automatikoak eta testuak zuzentzeko teknologiak dira gaur egun gehien erabiltzen direnak.

Tresna hauek, nagusiki, produktibitatea eta komunikazioa hobetzeko baliatzen dira. Inkesta erantzun duten pertsonen iritziz, Hizkuntza Teknologien garrantzia eta eragina areagotuz joango da, harik eta eguneroko bizitzako alor askotan ezinbesteko bihurtu arte.

Inkesta erantzun duten pertsonen iritziz, Hizkuntza Teknologiek gero eta garrantzi eta eragin handiagoa izango dute

2. Sektore bakoitzak berariazko behar eta erronkak ditu.

Azterketan bi tresna erabili dira: batetik, galdetegi orokorrak, eta bestetik, aurrez aurreko elkar-errikteta sektorizatuak. Hizkuntza-industrietatik abiatu da ikerketa, arlo honetako erreferente direlako, eta gero industria, justizia, osasun, administrazio publiko, komunikabide eta hezkuntza arloetara hedatu da. Sektore bakoitzak berezko beharrak eta errealitate propioa ditu.

- **Justizia eta osasungintzan** datuen beharra nabarmentzen dute, baina informazio sentisibla izanik, garapen teknologikoa bereziki arretaz egin behar da. Sektore publikoan ere lehenetsia ematen diote datuen babesari.
- **Industrian**, datuen babesa garrantzitsua bada ere, lan-segurtasunaren eskakizunek dute pisurik handiena teknologia ezartzeko orduan, zibersegurtasuna eta prebentzio-neurriak barne.
- **Hezkuntzan** askotariko beharrak daude. Teknologia berriak onuragarritzat jotzen dituzte ikasuntzarako, baina bi erronka nagusi dituzte: iturrien fidagarritasuna eta ziurgabetasunaren kudeaketa.

3. Gizartearen ezagutza eta kontzientziazio maila handitzea beharrezkoa da, eta horrek ohitura aldaketarako prozesu baten beharra eskatzen du.

Erakundeetan, digitalizazioari buruzko ezagutza-maila apala dela antzeman da, eta funtsezkoa da haren erabilgarritasuna eta aplikazioa hobeto ulertaraztea. Horrek hainbat esparrutan finkatutako ohiturak berraztertzea dakar. Euskararen kasuan zehazki, hizkuntza-teknologiak ezinbestekoak dira normalizaziorako, eta, gainera, beharrezkoa da hizkuntza honetan lan egiteko konpromisoa hartzea eta erakundeetan modu eraginkorrean erabiltzeko berariazko prestakuntza ematea.

4. Euskarak berebiziko garrantzia izan du teknologia hauen erabilera zein hedapenean, bai erronketan bai etorkizuneko erabilera-kasuetan ere.

Ikerketak arreta berezia jarri dio euskarari, parte-hartzaile gehienek lehen hizkuntza delako eta Hizkuntza Teknologien erabilera tresna nagusia dutelako. Emaiztekin erakusten dute gaztelaniaren pareko erabilera-maila duela, eta oinarritzko beharrak asetzen dituen garapena lortu dela. Hala ere, gaitasunak indartzeko beharra identifikatu da, eremu espezializatuetan eta garapen berrietan erabilera egokia bermatzeko.

5. Etorkizunari begirako teknologia hauen helburuek, kezkek eta erronkek hartu dute protagonismo handia:

- Sistemen zehaztasuna eta fidagarritasuna bermatzeko beharra.
- Hizkuntza-aniztasuna egoki kudeatzeko premia, erronka hau areagotu egingo baita hizkuntza eta erabilera-maila berriak gehitu ahala.
- Ingurune informatiko apalgotara egokitu daitezkeen sistemak garatzea, beste sistema batzuekin integratzea erraztuko dutenak, bereziki erakundeetan lehendik daudenekin.
- Erabiltzaileen esperientzia hobetzeko prozesuak optimizatzea.
- Behar beste datu-corpus lortzea. Erronka hau are garrantzitsuagoa da, kontuan hartuta gero eta joera handiagoa dagoela pertsonalizaziorako, eta horrek eremu bakoitzerako datu-multzo berezituak eskatzen dituela.

6. Irizpide ekonomikoari garrantzia nagusia esleitu zaio, irizpide soziala gutxien aintzat hartzen den irizpidea izanik.

Emaitza horiek haztaperen-taulan ikus daitezke. Kontuan izan behar da elkarriketa eta galdaketetan parte hartu duten profil nagusiak merkatuan lehiatzen diren enpresa edo erakundeak direla. Horrez gain, haien lan-eremuan egin dugu ikerketa, bai elkarriketak bai galdetegiak laneko posta elektronikoz edo lantokian bertan egin baititugu eta, beraz, ulertzekoa da alderdi ekonomikoa gailentzea. Produktibitatea buruan zuten unean egin ditugu elkarriketak eta, ondorioz, eraginkorrak izateko irtenbideak bilatu dira, eta askotan gure elkarriketak ere bide horretatik bideratu ditugu.

7. Hala eta guztiz ere, ikuspegi sozialaren ezinbestekotasuna aipatu beharra dago, handia izan baitaiteke Hizkuntza Teknologiek gizarteari egin diezaiokeen ekarpena.

Funtsean, Hizkuntza Teknologia gure eguneroko bizitzaren giltzarri bilakatu dira eta bilakatzen jarraituko dute; horregatik, garrantzitsua da ikuspegi soziala gogoan izatea. Hizkuntza teknologian ari direnen elkarriketatik, eta irakurritako hainbat txostenen ondorioetatik, ikusi dezakegu gizarte irizpideak ere oso garrantzitsuak direla teknologien ikerketa eta garapenean. Sektoreko adituekin egindako elkarriketatik eta hainbat txostenen azterketatik ondorioztatzen da irizpide sozialak funtsezkoak direla teknologia hauen ikerketa eta garapenean. Egungo prototipoak aztertzean, inklusioari begirako teknologia identifikatu dira. Adibidez, sektoreko erakunde batek mugimen-zailtasunak dituzten ikasleek idazlanak diktatu bidez egiteko aukera ematen duen irtenbidea garatu du. Ildo beretik, hainbat talderen bizikaltatea hobetzen duten irtenbideak garatzen ari dira: adibidez, pertsonen laguntzeko ahots-sistemak, istripu edo gaitasun baten ondorioz hizketa galdu duten pertsonentzako ahots-klonazio teknologia, eta komunikazioa errazten duten sistemak.

**Hizkuntza
Teknologiek gure
bizitzaren
praktikotasunean
giltzarri bilakatu dira
eta bilakatzen
jarraituko dute;
horregatik,
garrantzitsua da
ikuspegi soziala
gogoan izatea**

Irizpide sozialek presentzia nabarmenik izan ez badute ere elkarrizketa eta galdeketetan, garrantzi handia dute azter-tutako txostenetan, eta ez da gutxietsi beharreko irizpidea. Izan ere, adimen artifizialaz ari garenean eguneroko testuin-guruetan, alderdirik garrantzitsuenetako bat da, hain zuzen ere, gizarteari onura ekartzeko duen ahalmena.

Horregatik guztiagatik, txosten honek, eragin handiko erabilera-kasuen erakusleak aurkezteaz gain, digitalizazioa-ren eta ikuspegi sozialaren arteko harremanaren garrantzia azpimarratu nahi du.

Eranskinak

A Eranskina: Elkarrizketen gidoia

B Eranskina: Elkarrizketak grabatzeko baimen orria

D Eranskina: Galdetegiak

A eranskina: Elkarriketen gidoia

Helburua

- Demostradoreen menua osatzea – gaur egungo mapa – zer dago
- Inpaktuak neurtzeko irizpideak jaso – irizpide komunak aurkitzeko
- Erabilera kasu aurreratuak aurkitu

Gaur egun dagoena

1. Zein teknologia garatzen/lantzen ari zarete?
2. Zer dago gaur egun erabilgarri? Merkatuan?
3. Zenbateko inplementazioa maila izan dute, enpresa edo erakundeetara transferitu da?
4. Zein sektoretara edo esparrura dira hedagarriak aplikazio horiek?

Gaur egungo erronkak eta mugak

5. Zeintzuk dira hizkuntza naturalaren prozesamenduaren arloan egungo erronka nagusiak?
6. Zeintzuk dira zuen garapen proiektuetan aurkitzen dituzuen oztopo edo erronka nagusiak?

Hizkuntza gutxituen sustapena

7. Nola ari dira teknologiak hizkuntzen aniztasuna kudeatzen?
8. Hizkuntza minorizatuak dagokienez (euskara, adib.), zein dira hizkuntza teknologiek duten abantailarik nagusienak?
9. Zein aplikazio sor zitekeen euskal komunitatean inpaktua eduki liratekeena?
10. Nola lagun dezakete hizkuntza teknologiek hizkuntza gutxituen biziraupenean eta sustapenean?
11. Zein dira hizkuntza hauek digitalizatzeko orduan aurkitzen diren erronkak?

Interdiziplinaritatea

12. Zein beste diziplinarekin uste duzu lankidetzat handiagoa izan beharko litzatekeela hizkuntza teknologien garapenean?

Etorkizuneko joerak

13. Zer norabide hartuko du hizkuntza teknologien ikerketak hurrengo hamarkadetan?
14. Zein dira hurrengo urteetarako hizkuntza teknologien joera nagusiak?
15. Zein aplikazio berri edo hobekuntza espero daitezke?

Erabilera Kasuak eta Aplikazio Praktikoak

16. Zein dira, zure ustez, hizkuntza teknologiak izan duen eraginik handieneko erabilera kasuak gaur egun? Zergatik?

17. Nola imajinatzen duzu hizkuntza teknologiak gizartean eta teknologian duen lekua 10 urte barru? Zein aldaketa edo aplikazio espero dituzu?
18. Zein soluzio/aplikazio garatuko zenuke dauden hizkuntza teknologiak erabiliz beharrezko baliabide guztiak bezenitu? betiere, jendartearen egunerokotasuna errazteko?
19. Zer baliabide lehenetsiko zenituzke “beharrezko baliabide guztiak bazenitu” multzo horretan?

Irizpideak

20. Hizkuntza teknologiak aukeratzeko orduan, zein irizpide hartzen dituzue kontuan?
21. Nola aukeratuko zenuke proiektu bat aukeratu edo beste bat?
 - (a) Zein irizpideren arabera? Arlo ekonomiko, sozial eta teknikoaren artean zer lehenetsiko zenuke? Zergatik?
 - (b) Zein dira gaur egun sektorean edota zure erakundean garapen proiektua martxan jarri edo ez erabakitzeko irizpide nagusiak?

B eranskina: Elkarrizketak grabatzeko baimen orria

 GAIA TALDEA	Datu Pertsonalen Babesa Irudia/ahotsa tratatzeko baimena	Edizioa: 2024/04/01 Erref.: IRU-003 Orrialdea: 1 / 2
--	---	--

DATU PERTSONALEN BABESA – IRUDI ETA AHOTSAREN TRATAMENDUA

Datu Pertsonalak Babesteko legedia betez, jakinarazi nahi dizugu:

1. Elkarrizketa honetan jasotzen den informazio guztia “Zerbitzu hornitzaileak” Hizkuntza Teknologien aplikazio kasuen inguruko azterlanerako eta gai horrekin bat egiten duen GBL proiektuaren helburuetarako bakarrik erabiliko da. Informazioa konfidentziala izango da, euskarria edozein dela ere. Informazioa ez da publikoa izango eta ez da zabalduko elkarrizketatutako pertsonen alde zureko baimenik gabe.
2. Datu pertsonalak babesteko araudiaren arabera, Europako Parlamentuaren eta Kontseiluaren apirilaren 27ko 2016/679 (EB) Erregelamendua (“Datuak Babesteko Erregelamendu Orokorra” edo “DBEO”) eta Datu Pertsonalak Babesteko Lege Organikoa. Datu pertsonalak biltzen dituen informazio guztia ahal bezain laster anonimotuko da. Halaber, datu pertsonalak eta informazio konfidentziala dituen euskarri oro suntsitu egingo da proiektuaren azken txostena idatzi ondoren.
3. Proiektuan bildutako edo landutako datu pertsonalak ezin izango zaizkie hirugarrenei laga.
4. Konfidentzialtasun-konpromiso honek “Zerbitzu hornitzailearen” eta elkarrizketatutako pertsonen arteko harremana amaitutakoan ere iraungo du.
5. Dokumentu hau onartzeak esan nahi du deskribatu diren tratamenduak eta komunikazioak egiteko **baimen zehatza** ematen diozula “zerbitzu hornitzaileari”. Zure datuak tratatzeko legezko oinarria zure baimena da.

Arrasaten,(ko)(ren) (ean)

Interesduna

Zerbitzu hornitzailea

Izen-Abizenak:
NAN:

Izen abizenak:
NAN:

D eranskina: Galdetegiak

Sarrera

Adimen artifizialetik eta hizkuntza naturalaren prozesamendutik abiatutako hizkuntza teknologiak azken hamarkadetan nabarmen garatu dira hainbat aplikazio ahalbidetuz: itzulpen automatikoa, ahotsaren ezagutza eta sintesia, ChatGPT bezalako hizkuntza eredu handiek dakarten testuen analisia eta sorkuntza, eta beste hainbat. Teknologia horiek jada erabiltzen ditugu gure egunerokoan, eremu pertsonalean zein profesionalean. Hizkuntzen kudeaketa desegoki batek gizarte honetako hizkuntza aniztasuna ikusezin egiten du, eta horregatik, **ILENIA proiektuak** Espainiako estatuko hizkuntzetan aplikazio eleanitzak garatzeko helburua du. **LANGUNETik**, hizkuntza industrien elkartetik, galdetegi honen bidez hausnarketarako gune bat ireki nahi dugu.

- **Zertarako:** Hizkuntza teknologia horien bidez, gizarteak, sektore industrialak edo beste sektoreek izan ditzaketen inpaktu sozioekonomiko handiko erabilera-kasuak identifikatzeko. Gaur egun jada egiten dena jasotzeko, eta, horietatik abiatuta, hedagarri diren eremu edo erabilera berriak antzemateko.
- **Nola:** Galdera itxi eta irekien bidez. Guztiz anonimoa da eta azterlan honetarako baliatuko da soilik. Egun baliatzen ditugun erabileratik abiatuta, etorkizuneko erabilera esparru berriak irudikatzea da asmoa.
- **Zer:** Azterlan bat abiatu da eta galdetegi hau hainbat sektoretako eragileak entzuteko kanal bat izango da. Azterlanaren emaitza publikoa izango den txosten batean jasoko da.
- **Galdetegia betetzeko epea:** maiatzak 27

Informazio orokorra

[G-1] Adina:	[G-2] Generoa:	[G-3] Ama-hizkuntza:	[G-4] Lan eremua: *G.2 bakarrik
☐ 18-24	☐ Emakumea	☐ Euskara	☐ Industria
☐ 25-34	☐ Gizona	☐ Gaztelania	☐ Hedabideak
☐ 35-44	☐ Ez bitarra	☐ Frantsesa	☐ Akademikoa
☐ 45-54	☐ Besterik	☐ Besterik	☐ Administrazioa
☐ 55-64	(zehaztu)	(zehaztu)	☐ Besterik (zehaztu)
☐ >=65			

Gaiarekin sintonizatzen...

[G-5] Zein hizkuntza teknologia erabiltzen dituzu egunerokoan?

Ordenatu gehienetik gutxienera.

- ≡ Laguntza pertsonaleko aplikazioak, ahots bidezko aginduak (adib. Siri, Alexa)
- ≡ Testuak zuzentzeko aplikazioak (adib. ortografia eta gramatika-zuzentzaileak)
- ≡ Itzulpen-aplikazioak (adib. Elia, Itzuli, Google Translate, DeepL)
- ≡ Azpititulazio automatikoa (adib. albiste-erreportajea, YouTube)
- ≡ Txatbot-ak
- ≡ Eduki sortzaileak (adib. ChatGPT, Claude, Dall-E, Gemini)
- ≡ Ahotsa testu bihurtzeko teknologiak
- ≡ Pantaila-irakurgailuak

[G-6] Onura hauetako zein ezaugarri lehenesten duzu gehien zure erabilera arrazoitzeko?

Aukeratu nahi beste.

- ☐ Komunikazioaren eraginkortasuna hobetzeko, testuaren zuzentasuna zaintzeko
- ☐ Nire produktibitate maila handitzeko
- ☐ Ahots bidez erosoago egiten zaidalako mota ezberdinetako ekintzak bideratzea
- ☐ Erabiltzaileekiko interakzioak pertsonalizatzea
- ☐ Ataza errepikakorren automatizazioa
- ☐ Hizkuntza-zerbitzuen irismena handitzea
- ☐ Besterik (zehaztu mesedez)

[G-7] Eguneroko zure erabileran, zein hizkuntzatan erabiltzen dituzu nagusiki?

Ordenatu gehienetik gutxienera.

- ≡ Euskara
- ≡ Gaztelania
- ≡ Frantsesa
- ≡ Ingelesa

[G-8] Zergatik?

Aukeratu nahi beste.

- ☐ Nire ohiko hizkuntza delako
- ☐ Nire lan-jarduerak hizkuntza horretan egiten ditudalako
- ☐ Tresna teknologiko gehiago daudelako hizkuntza horretan
- ☐ Ezagutza txikia dudan hizkuntza delako
- ☐ Besterik (zehaztu mesedez)

Lan eremura etorri...

G.1 Hizkuntza teknologien industria

[G.1-9] Zein hizkuntza teknologia nagusi garatu edo aplikatu dituzue eta zein erabilera eman diozue? *Azaldu laburki.*

[G.1-10] Zeintzuk dira zuen garapen proiektuetan aurkitzen dituzuen oztopo edo erronka nagusiak? *Azaldu laburki.*

[G.1-11] Zein sektoretara edo esparrura dira hedagarriak aplikazio horiek? *Azaldu laburki.*

[G.1-12] Zeintzuk dira etorkizuneko helburu nagusiak hizkuntzaren teknologiei dagokienez?

Aukeratu 3.

- ☐ Irisgarritasuna eta inklusioa hobetzea, bizitzak erraztea
- ☐ Komunikazioa eta lankidetzaren erraztea
- ☐ Testu eta ahots datuen bolumen handien sailkapena eta analisisia automatizatzea, erabakiak hartzeko
- ☐ Erabiltzaile-interfaze intuitiboagoak eta naturalagoak garatzea
- ☐ Bezeroari zerbitzua emateko prozesuak optimizatzea, adibidez, kontsulta konplexuak ulertzeko eta ebazteko
- ☐ Merkaturaren berrikuntza eta lehiakortasuna sustatzea
- ☐ Datuen segurtasuna hobetzea, iruzurrak hautemateko teknologiak eta hizkuntzan oinarritutako arrisku-analisiak erabiliz (biometria, ahots identifikazioa, ...)
- ☐ Besterik (zehaztu mesedez)

- [G.1-13] Zein dira hizkuntza teknologien esparruan esploratu gabeko aplikazio aukerarik handienak?
Azaldu laburki.
- [G.1-14] Zein soluzio/aplikazio irudikatzen duzu beharrezko baliabide guztiak bazenitu?
Azaldu laburki.
- [G.1-15] Zein dira gaur egun sektorean edota zure erakundean garapen proiektua martxan jarri edo ez erabakitzeko irizpide nagusiak? *Aukeratu nahi beste.*
- ☐ Irizpide ekonomikoak
 - ☐ Irizpide sozialak
 - ☐ Irizpide teknologiko-teknikoak
- [G.1-16] Zuen erakundeak beste zein erakunderekin lankidetzan dihardu hizkuntza teknologien garapenean? *Aukeratu nahi beste.*
- ☐ Akademikoa
 - ☐ Industria - Zerbitzu enpresak
 - ☐ Administrazioa
 - ☐ Besterik (zehaztu mesedez)
- [G.1-17] Zeintzuk dira zure erakundeak hizkuntza teknologia berriak garatzeko finantzatzeko dituen iturri nagusiak? *Ordenatu handienetik txikienera.*
- ≡ Dirulaguntza publikoak (Europa, Espainia, Euskal Herria...)
 - ≡ Fondo propioak (salmentak, autofinantzazioa)
 - ≡ Inbertitzaileen ekarpena
- [G.1-18] Gaiaren inguruan besterik adierazi nahiko bazenu beheko espazioan egin dezakezu.

G.2 Industria eta zerbitzu enpresak

- [G.2-9] Nolakoa/zenbateko eragina dute gaur egun hizkuntza teknologiek zuen lan eremuan?
1 = oso txikia, 10 = erabatekoa.
- [G.2-10] Eta etorkizunean eragin maila...
Aukeratu 1.
- ☐ Handituko da
 - ☐ Berdin mantenduko da
 - ☐ Jaitsi egingo da
- [G.2-11] Zein arlo zehatzetan erabiltzen dituzue erakundean hizkuntzaren teknologiak?
Aukeratu nahi beste.
- ☐ Bezeroen/ikasleen arretarako eta laguntza teknikorako
 - ☐ Enplegu/Praktikaldi baliabideen kudeaketarako eta hautapenerako
 - ☐ Merkatuaren analisirako eta iritzien ikuspegirako
 - ☐ Administrazioa eta dokumentazioa automatizatzeko prozesuetarako
 - ☐ Produktuen garapena eta hobekuntza
 - ☐ Barne komunikazioa eta lankidetzaren hobetzeko
 - ☐ Garapena eta prestakuntza errazteko
 - ☐ Lege eta araudi-betetzearen automatizaziorako
 - ☐ Besterik (zehaztu mesedez)
- [G.2-12] Zuen erakundean beste teknologia edo sistema batzuekin hizkuntza teknologiak nola integratu dituzue? *Azaldu laburki.*

[G.2-13] Zeintzuk dira aplikatzerako orduan aurkitzen dituzuen oztopo edo erronka nagusiak? *Azaldu laburki.*

[G.2-14] Zeintzuk dira zure iritziz etorkizuneko helburu nagusiak hizkuntzaren teknologiei dagokienez? *Aukeratu 3.*

- ☐ Irigarritasuna eta inklusioa hobetzea, bizitzak erraztea
- ☐ Komunikazioa eta lankidetzaren erraztea
- ☐ Testu eta ahots datuen bolumen handien sailkapena eta analisisia automatizatzea, erabakiak hartzeko
- ☐ Erabiltzaile-interfaze intuitiboagoak eta naturalagoak garatzea
- ☐ Bezeroari zerbitzua emateko prozesuak optimizatzea, adibidez, kontsulta konplexuak ulertzeko eta ebazteko
- ☐ Merkaturaren berrikuntza eta lehiakortasuna sustatzea
- ☐ Datuen segurtasuna hobetzea, iruzurrak hautemateko teknologiak eta hizkuntzan oinarritutako arrisku-analisiak erabiliz (biometria, ahots identifikazioa,...)
- ☐ Besterik (zehaztu mesedez)

[G.2-15] Aurreko galderan aukeratu dituzun 3 helburu nagusiak kontuan hartuta, zein erabilera/aplikazio posible bururatzen zaizkizu? *Azaldu laburki.*

[G.2-16] Zure ustez, zeintzuk dira erronka nagusiak zurea bezalako enpresa batean hizkuntza-teknologiak ezartzean? *Aukeratu gehienez 4.*

- ☐ Dauden sistema eta plataforma teknologikoekin integratzea
- ☐ Datuen segurtasuna eta pribatutasuna bermatzea
- ☐ Hizkuntza prozesatzeko sistemen zehaztasuna eta fidagarritasuna bermatzea.
- ☐ Enpresaren eta bezeroen behar espezifikoetara egokitzea
- ☐ Langileak tresna eta teknologia berriak eraginkortasunez erabiltzeko gai izatea
- ☐ Aniztasun linguistikoa kudeatzea
- ☐ Teknologia berrien onarpena eta aldaketa kudeaketa
- ☐ Inbertsioaren itzultze tasa neurtzea eta justifikatzea
- ☐ Besterik (zehaztu mesedez)

[G.2-17] Zein dira, zure ustez, hizkuntza teknologien esparruan (eta zuen sektorean / gizartean) esploratu gabeko aplikazio aukerarik handienak? *Azaldu laburki.*

[G.2-18] Zein soluzio/aplikazio irudikatzen duzu beharrezko baliabide guztiak bazenitu? *Azaldu laburki.*

[G.2-19] Gaiaren inguruan besterik adierazi nahiko bazenu beheko espazio honetan egin dezakezu.

G.3 Hezkuntza sektorea

[G.3-9] Nola uste duzu hizkuntzaren teknologiek hezkuntza-prozesuari mesede egin diezaioketela? *Aukeratu nahi beste.*

- ☐ Trebetasun-maila eta premia desberdinak dituzten ikasleentzako hezkuntza-materialen irisgarritasuna hobetuz
- ☐ Nazioarteko ikasleentzako hezkuntza-baliabideen itzulpen automatikoa erraztuz
- ☐ Idatzizko ebaluazioetan atzeraelikadura automatizatuz, hala nola saiakeretan
- ☐ Tutoretza birtualeko sistemen bidez ikaskuntza pertsonalizatuz
- ☐ Ahotsaren ezagutza eta bihurketa teknologiak erabiliz irakaskuntza interaktiboagoak eskainiz
- ☐ Hizkuntza-desberdinetako ikasleen inklusioa bultzatuz
- ☐ Kolaborazioa eta komunikazioa erraztuz tele-hezkuntzan
- ☐ Gaitasun linguistikoak modu ludikoan garatuz
- ☐ Besterik (zehaztu mesedez)

[G.3-10] Zure ustez, zein da hizkuntza-teknologiek hezkuntza arloan edo gizartean duten erronkarik handiena? *Aukeratu nahi beste.*

- ☐ Hezkuntza-erakundeetan azpiegitura teknologiko egokirik ez izatea
- ☐ Irakasleek eta hezkuntza-administratzaileek aldaketari aurre egitea
- ☐ Teknologia horiek eraginkortasunez erabiltzeko trebakuntza espezifiko behar izatea
- ☐ Sistema automatizatuaren erabileran, osotasun akademikoarekiko eta plagioarekiko kezka
- ☐ Aniztasun linguistikoa eta kulturartekotasuna onartzea
- ☐ Pribatutasuna eta datuen segurtasuna
- ☐ Teknologia horien kostua eta irisgarritasuna
- ☐ Besterik (zehaztu mesedez)

[G.3-11] Hurrengoetatik, zein irizpideri emango zenioke lehentasuna?

Ordenatu gehienetik gutxienera.

- ≡ Aplikazioak datu pertsonalekin zer egiten duen azaltzeari
- ≡ Aplikazioaren karbono-inpaktua modu argi batean azaltzeari
- ≡ Aplikazioaren euskararen mailari eta kalitateari
- ≡ Aplikazioaren “sinpletasunari” ahalik eta jende gehienara iristeko
- ≡ Aplikazioaren emaitzetan alborapenik ez emateari (adib. genero alborapena)

[G.3-12] Gizarteko erronka bati aurre egiteko zein hizkuntza teknologia asmatuko edo erabiliko zenuke? *Azaldu laburki.*

[G.3-13] Zein beste aplikazio berritzaile sortu daitezke hizkuntza teknologiek? *Azaldu laburki.*

[G.3-14] Gaiaren inguruan besterik adierazi nahiko bazenu beheko espazio honetan egin dezakezu.



IMPULSO DE LAS LENGUAS
EN LA INTELIGENCIA ARTIFICIAL

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology