

Hizkuntzaren teknologia, adimen artifizialaren hurren go jomuga

Urteko Txostena
2023



Adimen Artifizial Sortzailea (AA) izugarri aurreratu da azken urteotan. Alren igoera azkarra, zeinen adierazlerik nagusienak GPT bezalako Hizkuntza Eredu Handiak (LLM) diren, gizagaingidiko adimenaren inguruko itxaropen desegokiak zein etorkizunari buruzko iragarpen katastrofikoak sortu ditu, erronka sotil eta mundutarragoak ezkutatuz. Zorionez, teknologia berri horren onura edo arrisku izugarritzkoen inguruan hasiera batean sortutako ekaitza baretzen hasi da: LLMk urrun daude gizakien mailako adimenaren jabe izatetik, eta ez dira inolako mehatxu existentziala; aitzitik, aparteko tresnak dira, askotariko betebeharrak profesionalak dituzten erabiltzaile askori eginkizunak hobeto eta azkarrago egiten laguntzeko. Teknologia disruptibo guztiekin gertatzen den bezala, arriskuak egon badaude, hala nola, desinformazioa, alborapen kaltegarriak, energia-premia handiak, eta baliabide handiko eta gutxiko hizkuntzen arteko arrakala digitala handitzea.

Adimen artifizialak eta eraldaketa digitalak sortzen dituzten erronka eta aukera berriei aurre egiteko, ezinbestekoa da hizkuntza teknologien ikerketa irekia eta publikoa sustatzea. Are gehiago, bertako ikerketa egitea funtsezkoa da, bertako ikerketarik gabe beste leku batzuetan sortutako teknologiaren

language

hitzak

kontsumitzaile huts bihurtzen baikara. Jarrera horren onurak argiak dira: ikerketa eta emaitzak tokiko ekonomiaren eta gizartearen eskura jartzen dira, eta horiekin lotutako edozein arrisku erraz ikuska daiteke.

HiTZen oinarrietan sendo errotuta dago Ikuspegi hori. 2023 urtea oso emankorra izan da, gure zentroa funtsezko lau proiekturen inguruan nabarmen hazi baita. Eusko Jaurlaritzak IKER-GAITU proiektua hiru urterako finantzatzea erabaki zuen euskararen arrakala digitala txikitzeko bideak ikertzeko asmoz. Proiektu horri esker, euskararentzako orain arte sortutako Latxa izeneko LLMrik handiena eta onena eraiki eta ezagutzera eman dugu. Eredu erraldoi hori sortzeko beharrezkoak izan diren kalkuluak gure zerbitzari propioek eta EuroHPCK emandako dirulaguntza lehiakor bati esker lortutakoek eskaini dizkigute. Horretaz guztiaz gain, Espainiako Gobernuak, Galizia, Katalunia eta Valentziako eragile nagusiekin lankidetza estuan, hiru urterako ILENIA proiektua finantzatu zuen, euskararako, katalanerako eta galizierarako hizkuntza teknologia irekien hurrengo belaunaldia ekoizteko. Bi ekimen horiei CLARIAH-ES ikerketa-azpiegitura banatuaren sorrera ofiziala gehitu behar zaie. HiTZek koordinatuta, CLARIAH-ESek aukera ematen die Euskal Herriko eta Espainiako

ikertzaile eta ikerketa-erakundeei CLARINen eta DARIAHen parte hartzeko, hau da, Europako azpiegitura digital garrantzitsuenetan, humanitateak eta gizarte-zientziak ikertzeko. Azkenik, Espainiako Gobernuko Digitalizazio eta Adimen Artifizialeko Estatu Idazkaritzak eta Avature software enpresak lau urterako asmo handiko Adimen Artifizial eta Hizkuntzaren Teknologiako Katedra finantzatzea erabaki zuten. Katedra horrek indartu egingo ditu industriarako transferentzia teknologikoa, hizkuntza teknologiako etengabeko hezkuntza eta dibulgazio-jarduerak, eta era berean, gizartean eragin handia duten ikerketaren funtsezko bi alderditan zentratuko da: LLM berdeak eta LLM justuak.

Helburu bikoitzaz sortu genuen HiTZ. Alde batetik, hizkuntzen teknologien ikerketarako nazioarteko erreferentzia-zentro bihurtzea, eta bestetik, euskararen prozesamendu konputazionalerako erreferentzia izatea. Gure azken lorpenei esker, ikerketa- eta teknologia- zentro, enpresa eta erakunde nagusiekin lankidetzan jarraitzeaz gain, gure herrialdeak gero eta presentzia handiagoa izan dezan laguntzen jarraitu ahal izango dugu, hizkuntza teknologien eta adimen artifizialaren nazioarteko gune gisa.

Eneko Agirre (HiTZeko zuzendaria) eta German Rigau (HiTZeko zuzendariordea)

Ikerkuntza eta transferentzia

39

proiektu

2

transferentzia-
kontratu

hitzak

70

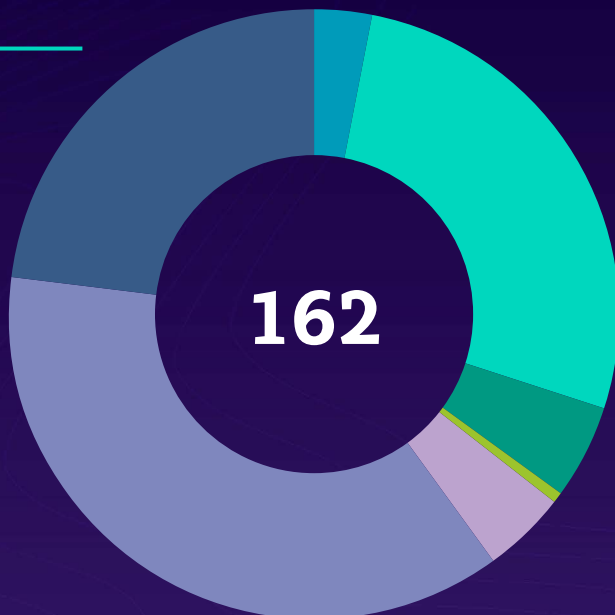
argitalpen



Pertsonak

Kideak

	Kudeaketa eta administrazioa	5
	Irakasleak	44
	Doktoratu ondoko ikertzaileak	8
	Ramón y Cajal	1
	Beste ikertzaileak	7
	Doktoretza-bekak	18

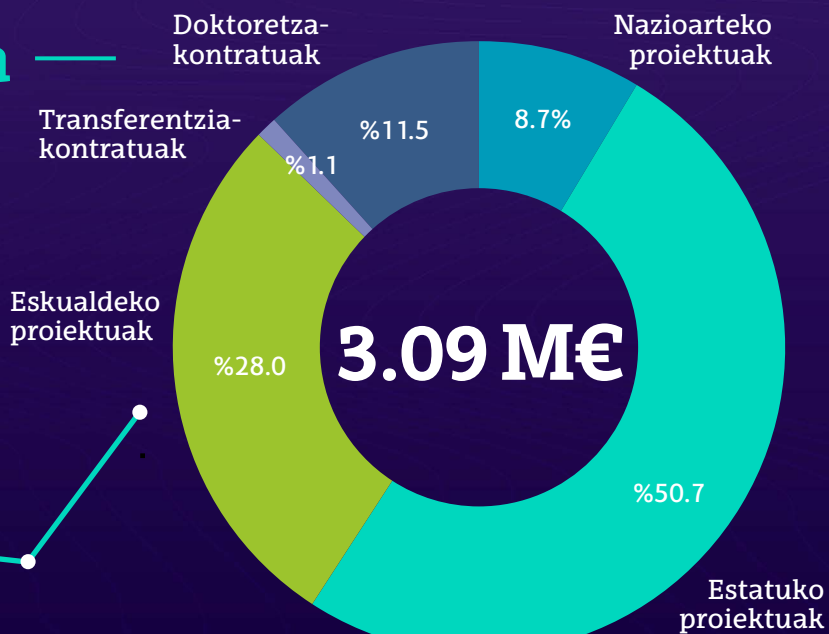


Ikasleak

	Master ikasleak	60
	Doktoretza ikasleak	37



Aurrekontua





Hizkuntza ardatz duen adimen **artifizialean** diharduen ikerketa-zentroa da HiTZ. Euskal Herriko Unibertsitateko zazpi sailetakoko kideek osatzen dutelarik, diziplina anitzeko taldea da. Hizkuntzaren eta hizketaren teknologiak ikertzea da zentroaren helburua, enpresei ezagutza eta teknologia transferitzeko ahalegin nabarmena eginez. Bi ikerketa taldek osatzen dute: **Aholabek** eta **Ixak**. Bi taldeok esperientzia handia dute, 1993az geroztik oinarritzko ikerketa egin, hizkuntza-baliabideak eta -tresnak sortu eta hainbat produktu merkaturatu dituztela. HiTZ zentroa CLAIREko kide da eta baita BDVAko eta DAIROko eskubide osoko kide ere. Claire eta BDVAren bidez, European Partnership on Artificial Intelligence, Data and Robotics-en parte hartzen du. Era berean, Spanish CLARIN K zentroko kide sortzaileetako bat gara.

2023ko irailaz geroztik, HiTZek CLARIAH-ES ikerketa-azpiegitura banatua koordinatzen du. Azpiegitura horrek Humanitateetako eta Gizarte Zientzietako Europako azpiegitura zientifiko handienetako bi biltzen ditu Espainian: CLARIN (Common Language Resources and Technology Infrastructure), Gizarte Zientzien eta Humanitateen arloko ikerketa errazteko hizkuntza teknologiako datuak, tresnak eta zerbitzuak eskaintzen dituen, eta DARIAH (Digital Research Infrastructure for the Arts and Humanities), Arteetan eta Humanitateetan baliabide digitaletan oinarritutako ikerketa eta irakaskuntza bultzatzen dituen. Bi azpiegitura horiek Ikerketa Azpiegituretarako Foro

CLARIN



DARIAH-EU



CLARIAH-EUS



Estrategikoaren (ESFRI, ingelesezko siglak) parte dira. Foro hori EBko estatu kideek eta Europako Batzordeak osatzen dute, eta 2002an eratu zen, Europa osorako instalazio zientifikoaren eta ikerketarako azpiegituren arloan estrategia komun bat koordinatzeko helburuarekin. Gure zuzendariordea, German Rigau, bi azpiegituren koordinatzaile nazionala da, eta CLARIAH-ES partzuergoaren buru da. Partzuergo horrek Gizarte Zientzien, Arteen eta Humanitateen, bibliotekonomiaren, hizkuntzalaritzaren, adimen artifizialaren, hizkuntza teknologien, informatikaren eta errendimendu handiko konputazioaren arloetako ikerkuntzan lider diren hamar erakundetako diziplina anitzeko aditu talde bat biltzen du.



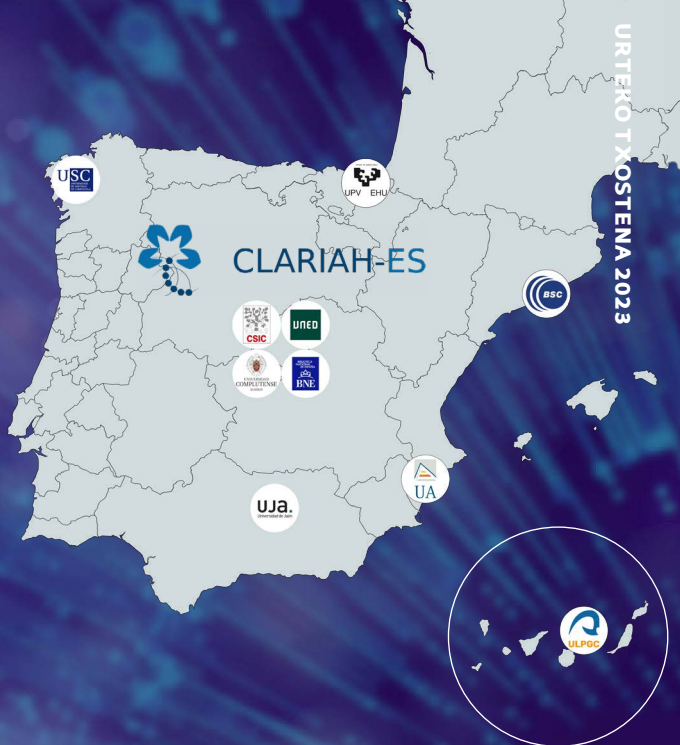
**Eneko
Agirre**

Zuzendaria



**German
Rigau**

Zuzendariordea



**Suna Seyma
Uçar**

Batzordeko
kidea



**Maite
Oronoz**

Batzordeko
kidea



**Itziar
Aldabe**

Batzordeko
kidea



**Inma
Hernaez**

Batzordeko
kidea



**Esther
Miranda**

Batzordeko
kidea



**Aitor
Soroa**

Batzordeko
kidea

Zentroko kideak nazioarteko erreferenteak dira beren jakintza-arloetan. Gaur egun 80 kidek baino gehiagok osatzen dute zentroa: informatikariak, hizkuntzalariek eta 5 ikerketa-teknikaririk. Azken bost urteetan, 200 argitalpen zientifiko baino gehiago argitaratu dituzte zentroa osatzen duten ikertzaileek. Ikaskuntza sakoneko teknikak hizkuntzaren prozesamenduan aplikatzen liderra da taldea, eta azken bi urteetan, 4.000 aipu baino gehiago izan ditu arlo horretan egin duen lanak. Zentroko kideek aholkulari gisa lagundu dute Espainiako Hizkuntzaren Teknologien Plan Nazionala sortzen, eta, gaur egun, Eusko Jaurlaritzak aholkatzen ari dira antzeko lanetan.

Bai lxa bai Aholab errendimendu handiko ikerketa-talde izendatu ditu Eusko Jaurlaritzaren Zientzia Kontseiluak, ikerketa ebaluatzeko azken ekitaldian. Historian zehar, taldeek 200 ikerketa-proiektutan baino gehiagotan parte hartu dute, eskualde-proiektuetatik hasi eta Europakoetaraino. Halaber, teknologia industriara transferitzeko helburuarekin, 100 industria-kontratutan baino gehiagotan parte hartu dute.

Erasmus Mundus+ European Masters Program **Hizkuntzaren eta Komunikazioaren Teknologien Europako Masterreko (LCT)** kide ere bada HiTZ. Industriak eta ikerketak hizkuntzaren teknologiaren alorrean dituzten beharrak asetzeko diseinatuta dago Masterra, azkar hazten ari den esparrua baita. HiTZek, halaber, Doktorego Programa bat eskaintzen du Hizkuntzaren Analisi eta Prozesamenduz, eta Adimen Artifizialeko Nazioarteko Doktoretza Akademiako (AIDA) kide osoa da.

Euskal Herriko Unibertsitatea (UPV/EHU) Euskal Autonomia Erkidegoko irakaskuntza-eta ikerketa-erakunde nagusia da. UPV/EHU munduko 400 unibertsitate onenen artean dago, Shanghaiko rankingaren arabera, eta Espainiako Gobernuak Nazioarteko Bikaintasun Campus izendatu du. Euskal Herriko Unibertsitateak, 30 urteko erakunde biziak, 45.000 ikasle ditu, mundu mailako 5.000 langile akademiko eta abangoardiako instalazioak, bere hiru campusetako 20 zentrotan banaturik.

IKERKETA-ARLOAK



**Informa-
zioaren
Erauzketa
eta Berres-
kurapena**

Ikertzaile nagusia:



**Aitor
Soroa**



**Itzulpen
Automatikoa**

Ikertzaile nagusia:



**Gorka
Labaka**



**Gizakiaren eta
Ordenagailua-
ren arteko
elkarreragina**

Ikertzaile nagusia:



**Gorka
Azkune**



**Hizketa- eta
Hizkuntza-
Baliabideak**

Ikertzaile nagusia:



**Ainara
Estarrona**

AZPIEGITURA

36

GPU
cluster
A100 80GB
vram

48

GPU cluster
A100 80 GB vram
(partekatua /
DIPC)

26

GPU
modelo
desberdinak

1

HPC
Clusterra,
128
nukleoduna

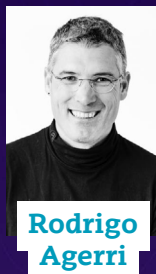
EuroHPC SuperComputer-en 1,5 milioi GPU orduko sarbidea (4,2 milioi euroko baliokoa), baliabide gutxiko Europako hizkuntzetarako eredu handiak ikertzeko





Testuen Analisia

Ikertzaile nagusia:



Rodrigo
Agerri



Hizketa- teknologiak

Ikertzaile nagusia:



Inma
Hernaez



Hizkuntzaren Prozesamen- dua Medi- kuntzan eta Zuzenbidean

Ikertzaile nagusia:

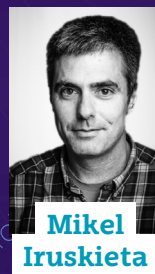


Arantza
Casillas



Humanitate Digitalak eta Hezkuntza

Ikertzaile nagusia:



Mikel
Iruskietta

450

TBtik
gora

sareko biltegiatze
ahalmen gordina

1

Behringer
4x4 audio/
MIDI
interfazea

1

Quiet PC
Sentinel
Fanless
i10

1

Akustikoki isolaturiko
gela, grabazio
profesionalerako audio-
ekipo batez hornitua

IKERKUNTZA ETA TRANSFERENTZIA

39

ikerkuntza
proiektuak

2

ezagutza
transferitzeko
kontratuak

4

doktoretza-
tesiak
defendaturik
(2 nazioarteko)

23

aldizkari-
artikuluak
(14 Q1)

36

artikuluak
biltzarretan
(9 A edo A+)

11

liburu-
kapitulua





PRAKTIKAK

60

ikasleak
masterrean

37

ikasleak
doktorego
programan

4

MALak
amaiturik
EMLCTn

21

MALak
amaiturik
HAP/LAPen

46

ikasleak, sakoneko
ikaskuntza bi ikastaro
osagarritan

6

Ikasikerrak

17

barne- eta
kanpo-praktikak

JARDUERAK

24
mintegiak

5
web-mintegiak

2
tailer
antolatuak

2
sariak

Carlos J. Finlay
Ordena (2022)

Radio Bilbao irratiaren
Bikaintasun Saria
Euskara kategorian

