



HITZ

Hizkuntza Teknologiako
Zentroa
Basque Center
for Language Technology



LATXA

Euskarazko
txatbot
irekia

latxa.hitz.eus

**Adimen
Artifizial
Generatiboari
buruzko ikerketa
aurreratua**

Urteko Txostena
2025

ANTOLAKUNTZA

HiTZ diziplina anitzeko ikerketa-zentroa da. Hizkuntza Teknologian eta Adimen Artifizial Sortzailean goi-mailako zentroa da, eta Euskal Herriko Unibertsitateko zazpi sailetako adituak biltzen ditu. Zentroaren egiteko nagusia hizkuntza- eta hizketa-teknologiak ikertzea da, enpresa-sektorearekiko ezagutzaren eta teknologiaren transferentzian enfasi berezia jarritz. HiTZek 1993tik arlo horietan esperientzia zabala duten bi ikerketa-talde biltzen ditu, IXA eta Aholab. Bi taldeak errendimendu handiko ikerketa-talde gisa aitortu zituen Eusko Jaurlaritzaren zientzia-agentziak bere azken ebaluazioan. Haien historian zehar, bi taldeek 200 ikerketa-proiektu baino gehiagotan parte hartu dute, eskualdeko ekimenetatik Europako ekimenetara, eta 100 kontratu baino gehiago egin dituzte enpresa desberdinekin. HiTZ BDVAko kide da, eta elkarte horren bidez, zentroak Adimen Artifizialaren, Datuen eta Robotikaren Europako Partzuergoan parte hartzen du.

HiTZek 100 kide baino gehiago ditu, informatikariak, hizkuntzalariak eta bost ikerketa-teknikari. Nazioartean aitorten handia duten ikertzaileak dira, eta horren erakusgarri, azken bost urteetan zentroko ikertzaileek 200 argitalpen zientifiko baino gehiago argitaratu izana. Zentroa liderra da ikasketa sakonean eta adimen artifizial sortzailearen teknikak testuari, hizketari eta ikusmen-hizkuntza ereduari aplikatzen. Arlo horretan egindako lan berriek 4.000 aipu baino gehiago izan dituzte azken bi urteetan.

HiTZek CLARIAH-ES ikerketa-azpiegitura banatuaren bulego teknikoa kudeatzen du. Azpiegitura horrek Espainiaren parte-hartzea kudeatzen du ESFRiko (*European Strategy Forum on Research Infrastructures*) parte diren CLARIN eta DARIAH ikerketa-azpiegituretan. HiTZeko zuzendariordeak, German Rigauk, koordinatzen du CLARIAH-ES partzuergoa, eta era berean, CLARINeko eta DARIAHko Espainiako koordinatzaile nazionala da.

Zentroaren gizarte-eragina Adimen Artifizialaren eta Hizkuntza Teknologien Katedraren (AA&HT) eta zentroak koordinatutako bi masterren bitartez indartzen da: Hizkuntza eta Komunikazio Teknologien Europako Erasmus Mundus+ Master Programa (LCT), eta Hizkuntzaren Analisia eta Prozesamendua Master Programa. Master programa horiek, graduondoko ikastaroa eta profesionalentzako lau etengabeko prestakuntza-ikastaro industriaren eta ikerketaren eskakizunak asetzeko diseinatuta daude. Gainera, HiTZek Hizkuntzaren Analisia eta Prozesamendua Doktorego Programa ere eskaintzen du.



**Eneko
Agirre**

Zuzendaria



**German
Rigau**

Zuzendariordea



**Suna Şeyma
Uçar**

Batzordeko
kidea



**Aritz
Farwell**

Batzordeko
kidea



**Maite
Oronoz**

Batzordeko
kidea



**Inma
Hernaez**

Batzordeko
kidea



**Esther
Miranda**

Batzordeko
kidea



**Aitor
Soroa**

Batzordeko
kidea

ZUZENDARIAREN MEZUA



Eneko Agirre
HITZeko zuzendaria

Adimen Artifizial Sortzaileak (AAS edo GenAI) aurrerapen nabarmenak izan ditu azken urteotan, batez ere Hizkuntzaren Teknologian (HT). Haren gorakada azkarra Hizkuntza Eredu Handiek (HEH edo LLM) erakusten dute, GPT edo Geminik urtero emandako aurrerapausoek erakutsi bezala. Behin pertsonekin komunikatzeko duten gaitasuna erakutsi ondoren, joan den urtean hizkuntzaz haratago salto egin dute programaziora eta matematiketara. Horretarako gakoa izan dira teknika berriak, hala nola Egiatzatu daitezkeen Balioen bidezko Reinforcement Learning (RLVR) deritzona egiaztatu daitezkeen balioekin ezkonduz eta Inferentzia Garaiko Konputazioa (Test-Time Compute). Era berean, HEHei tresnak erabiltzeko gaitasuna eman zaie, eta aldi berean autonomia eta beste HEHeekin koordinatzeko gaitasunaren bidez, Agenteen leherketa ekarri dute.

HEHen gaitasunen hedatzea testuaz haratago joan da, irudi, bideo eta ahotsarekin modu gardenean lan egiteko gaitasuna garatuz, eta horren bidez, aplikazio berrietara hedatuz, horien artean, pantailan ikusten dugunarekin lan egiteko aukera. Gezurra dirudi hori dena urtebete eskasean gertatu izana. Gauzak horrela, AASrekin egindako tresnek eragin sakona dute teknologiaren eta gizartearen alderdi askotan. Aro eraldatzaile horrek aukera berriak eskaintzen dizkie punta-puntako AAS teknologia aprobeatzeko gai izango direnei. Hala ere, AASren bat-bateko agerpenak eztabaida garrantzitsuak eragin ditu garapenaren eta hedapen arduratsuen inguruan, hala nola, lanbide askoren eraldaketa, desinformazioa, alborapen kaltegarriak, energia-behar handiak eta hizkuntza gutxituen eta hizkuntza handien arteko arrakala digitala zabaltzea.

Aro berri horri aurre egiteko, eta eskaintzen dituen aukerak aprobeatzeko, funtsezkoa da Hizkuntza Teknologiarenekin eta AASren ikerketa irekia eta publikoa sustatzea: bertako ikerketarik gabe, gure herria beste nonbait garatutako teknologiaren kontsumitzaile huts bihurtu baitaiteke. HTatik haratago, AAS beste teknologiatara hedatzen ari da, ikusmen artifiziala edo robotika barne, eta horrekin beste diziplina zientifiko batzuetara ere ekarri du iraultza, eta horren lekuko esaterako,

2024ko Kimikako Nobel saria teknologia hori oinarri duen AlphaFold-en sortzaileei eman izana, edo eguraldi iragarpenean AASan oinarritutako ereduak gailendu izana.

HiTZen erabat geure egiten dugu ikuspegi berri hori, eta bi lerro nagusitan banatu dugu gure jarduna: AAS eta HT alde batetik, eta, bestetik, AASa humanitate eta zientzatarako. Espainiako AAS eta HTko aditu nagusienetako asko HiTZ zentroko kideak dira, eta ikerketaren abangoardian gaude arlo horretan. Aldi berean, gure diziplina anitzeko zentroa etengabe hazten ari da. Horren ondorioz, goi-mailako konferentzia eta aldizkarietan artikulu kopuru esanguratsua argitaratu dugu, eta inoiz baino doktorego-ikasle gehiago ditugu. Espainia mailako Adimen Artifizialeko ikerkuntza sarea eta urte anitzeko proiektu eraldatzaileak lideratzen ditugu. Horren erakusle da Latxa. 2024an lehen bertsioa kaleratu zenetik euskarazko eredu ireki nagusia da. Lorpen horrek ACLn, arloko konferentziarik entzutetsuenean, artikulu onenaren saria lortu zuen. 2025ean gizakiokin hitz egiteko gai izan zedin egokitu genuen, GPTren mailara iritsiz gizakiok egindako ebaluazio publikoetan. Latxa orain probatarako erabilgarri dago lizentzia irekiekin, ahotsarekin lan egiteko eredu ireki hoberenerekin batera.

Azkenik, Adimen Artifizialaren eta Hizkuntza Teknologiarenekin HiTZ Katedrak profesionalak prestatzeko eta ezagutza eta teknologia gizarteari transferitzeko aukera ematen digu. Era berean, HiTZek koordinatzen duen CLARIAH-ES ikerketa-azpiegiturak gure lankidetzaren sare nazionalak eta nazioartekoak sendotu ditu gizarte-zientzietarako eta humanitateetarako tresna eta baliabide digitalak eskainiz. Bukatzeko, AASen onurak domeinu berrietara zabaltzeko aktiboki lan eginez, hizkuntza bisualen eredu aitzindariak sortu ditugu, eta neurozientzialariek batera asmo handiko proiektu batean lanean ari gara, hizketaren eta garunaren arteko elkargunea aztertuz. Egitasmo arrakastatsu horiek guztiak etorkizuneko erronkei ilusioz aurre egiteko oinarri sendoa dira HiTZentzat, eta era berean, gure herria Hizkuntza Teknologiarenekin eta Adimen Artifizial Sortzailearen ikerketan erreferente gisa kokatzen dute.



CLARIAH

HiTZek CLARIAH-ES izeneko ikerketa-azpiegitura digital banatua zuzentzen du, eta azken horrek gidatzen du Espainiaren partaidetza CLARIN (Hizkuntza Baliabide eta Teknologia Azpiegitura Komuna) eta DARIAH (Arte eta Humanitateetarako Ikerketa Azpiegitura Digitala) Europako Ikerketa Azpiegituren Partzuergoetan. CLARIAH-ESen helburua da baliabide digitalen bidezko ikerketa indartzea eta ikertzaileei punta-puntako baliabide, tresna eta zerbitzu digitaletarako sarbidea ematea. Bere diziplina anitzeko eta diziplinarteko ikuspegia diseinatua dago gizarte-zientzietan eta humanitateetan Espainiako ikerketa aurrera eramateko, Espainiako humanitate digitalen komunitatea sustatzeko eta Espainiako ikertzaileak nazioarteko proiektu eta programetan —bereziki Europako Ikerketa Esparruaren barruan— estrategikoki kokatzen laguntzeko. HiTZeko zuzendariorde German Rigau da Espainiako koordinatzaile nazionala CLARIAH-ES, CLARIN eta DARIAH erakundeetarako.

Gauregun, CLARIAH-ESek Espainia osoko hamabi erakunde eta ikerketa-talde nagusitako kideak biltzen ditu, gizarte-zientzien eta humanitateen esparruan gaitasun digitalak aurrera eramateko lan egiten dutenak. 2025ean, azpiegiturak bere adituen sarea are gehiago sendotu du, eta komunitate akademikoari zein publiko orokorrari zuzendutako zabalkundea areagotu du. Arlo honetan egindako lanaren aitortza gisa, CLARIAH-ESek azpiegiturarik onenaren saria jaso zuen Humanitate Digital Hispanikoen Nazioarteko Elkartearen 7. Kongresuan.

HiTZek CLARIAH-EUS nodoa ere zuzentzen du; bertan biltzen dira gizarte-zientzien eta humanitateen esparruan euskararen eta euskararako ikerketan diharduten hainbat erakunde eta ikerketa-talde. CLARIAH-EUSEk Eusko Jaurlaritzaren, Gipuzkoako Foru Aldundiaren, Euskal Herriko Unibertsitatearen (EHU) eta HiTZen finantzaketa eta babesa jasotzen ditu.

Informazio gehiago hemen: www.clariah.es eta www.clariah.eus



German Rigau

HiTZeko zuzendariordea
CLARIAH-ES, CLARIN eta DARIAHen
Espainiako koordinatzaile nazionala



Xabier Arregi

CLARIAH-EUS-en koordinatzailea

 **CLARIAH-ES**

 **CLARIAH-EUS**



KATEDRA



HiTZ

Adimen Artifizial eta Hizkuntza
Teknologiako Katedra
Cátedra de Inteligencia Artificial
y Tecnología de la Lengua

5
URTEKO TXOSTENA 2025

2025ean, Adimen Artifiziala eta Hizkuntza Teknologien (AA&HT) HiTZ Katedra egonkortu dugu. Katedra hori 2024an jarri zen abian, Adimen Artifizialaren Estrategia Nazionalaren (ENIA) barruan. Katedra hori Euskal Autonomia Erkidegoan finantzatutako katedra bakarra da, eta Espainian Hizkuntzaren Teknologiarri eskainitako katedra bakarra (informazio gehiago: www.hitz.eus/katedra).

Egonkortze-fase honetan, Katedrak bere hiru ardatzetan (irakaskuntza, ikerketa eta ezagutzaren transferentzia) sakondu du, AA&HTak gizartean eta sektore produktiboan duen eragin positiboa ahalik eta handiena izan dadin, ikerketa-jarduerak indartzeko eta programa akademiko eta profesionalak sostengatzen dituen ezagutzainarria finkatzeko. Lan hori gure enpresa laguntzaileekin batera egiten da: Avature SL, Elhuyar fundazioa, Ikerlan SCL, Euskaltel-MasOrange, Multiverse Computing eta Tecnalía Research and Innovation fundazioa. Erakunde horien garrantziak, kopuruak eta bikaintasunak agerian uzten dute sektore produktiboak baduela Katedrak gizartearentzako onurak sortzeko ahalmenean konfiantza.

Katedrak Adimen Artifizial Sortzailea, Ikasketa Sakona eta Hizkuntzaren Teknologia arloetako aditunagusiak ematen dituzten lau espezializazio-ikastaro eskaintzen jarraitzen du. Ikastaro horiek azken aurrerapenekin eguneratuak daude, eta bereziki profesionalentzat, ikertzaileentzat eta ikasleentzat diseinatu dira. hitz.eus/training

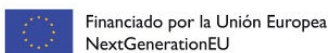


Aitor Soroa

Adimen Artifiziala eta Hizkuntza Teknologien (AA&HT) HiTZ Katedraren zuzendaria



Funded by:



Partner companies:



HiTZ

ZENBAKITAN

Ikerkuntza eta transferentzia

31

ikerkuntza proiektuak

9

ezagutza transferitzeko kontratuak

4

bestelako proiektuak

7

doktoretza-tesiak defendaturik (3 nazioarteko)

23

aldizkari-artikuluak (13 Q1)

41

artikuluak biltzarretan (15 A edo A+)

3

liburu-kapituluak

1

liburua







TRANSFERENTZIA

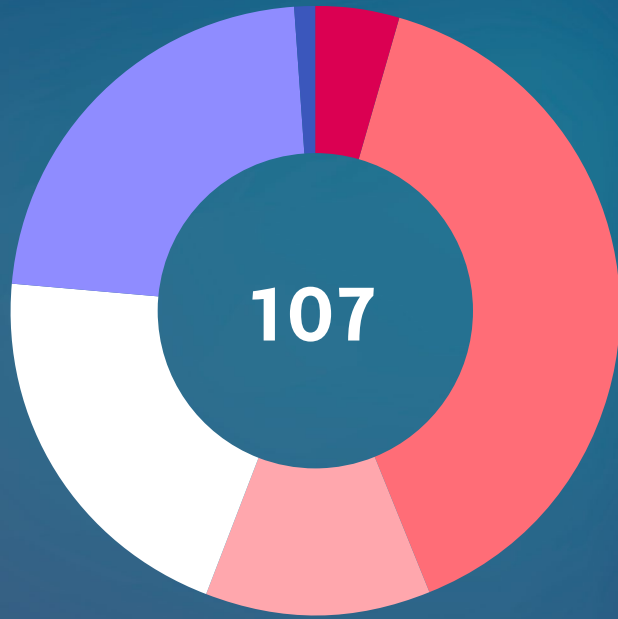
HiTZ zentroan euskararako sortu ditugun eredu, kode eta datu guztiak daude eskuragarri. Tartean daude euskararako txatbot, ahots ezaugatzaille eta ahost sortzaile ireki hoberenak, erabilera komertziala baimentzen dutenak.

hitz.eus/eu/emaitzak

Pertsonak

Kideak

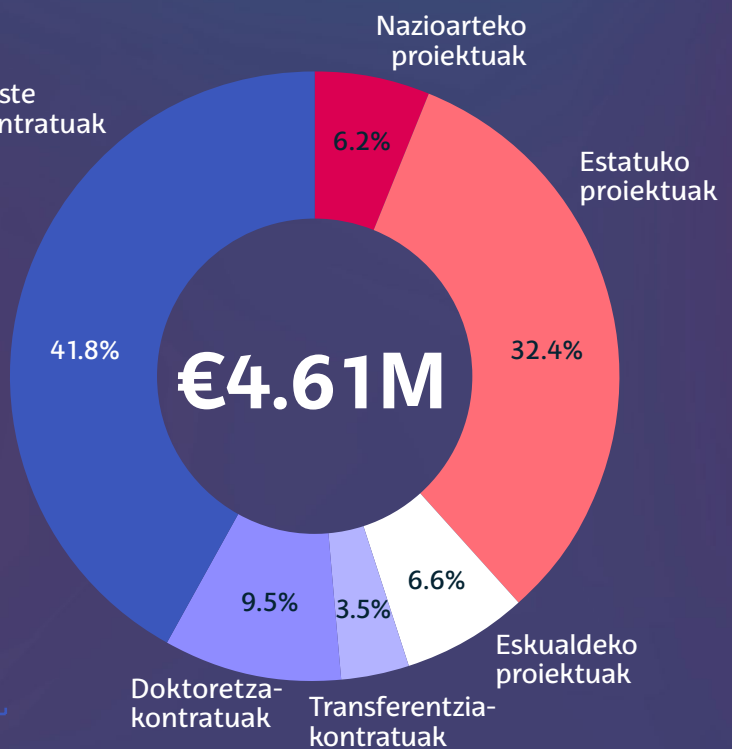
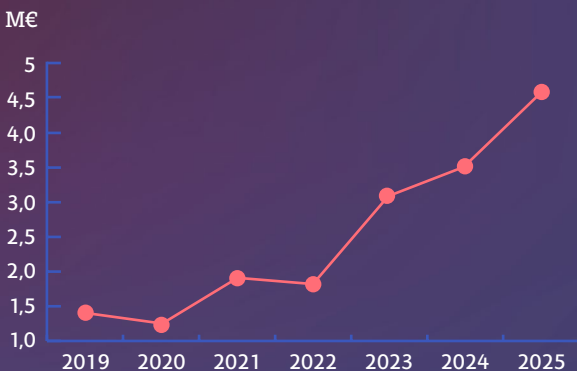
	Kudeaketa eta administrazioa	5
	Irakasleak	42
	Doktoratu ondoko ikertzaileak	13
	Beste ikertzaileak	22
	Finantzatutako aurredoktoregoko ikertzaileak	24
	Kolaboratzaile akademikoa	1



Ikasleak

	Master ikasleak	68
	Doktoretza ikasleak	46

Aurrekontua



IKERKETA-ARLOAK



Informazioaren Erauzketa eta Berreskurapena

Ikertzaile nagusia:



Aitor Soroa



Itzulpen Automatikoa

Ikertzaile nagusia:



Gorka Labaka



Gizakiaren eta Ordenagailuaren arteko elkarreragina

Ikertzaile nagusia:



Gorka Azkune



Hizketa- eta Hizkuntza-Baliabideak

Ikertzaile nagusia:



Ainara Estarrona

AZPIEGITURAK

18

Multiprocessor GNU/Linux

1

SPARC Solaris zerbitzaria

8

H200 141GB vram GPU zerbitzaria

36

A100 80GB vram GPU cluster

56

GPU cluster A100 80 GB vram (partekatua / DIPC)

14

L40 GPU (48 GBko vram)

26

GPU modelo desberdinak

1.2 PB

sareko biltegitarte ahalmen gordina



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
Irukieta

1

Akustikoki
isolaturiko
gela

1

Behringer 4x4
audio/MIDI
interfazea

1

Quiet PC
Sentinel
Fanless i10

1

Delsys Trigno™
Wireless EMG
System

**EuroHPC SuperComputer-en
900.000 GPU orduko sarbidea (2,2
milioi euroko baliokoa), baliabide
gutxiko Europako hizkuntzetarako
eredu handiak ikertzeko.**



PRAKTIKAK

68

masterretako
ikasle

46

doktorego
programako
ikasle

12

MALak
amaiturik
EMLCTn

21

MALak
amaiturik
HAP/LAPen

95

ikasle HEHrekin lotutako
4 etengabeko hezkuntza
ikastarotan

3

Ikasikerrak

7

barne- eta
kanpo-praktikak



JARDUERAK

31
mintegiak



11
URTEKO TXOSTENA 2025

7
web-
mintegiak

3
tailer
antolatuak

4
sariak

III Certamen de Seminarios de la Fundación Duques de Soria de Hispanismo Internacional. Proiektu irabazlea: "Humanidades digitales en acción: herramientas para el análisis de prensa histórica en español" (CLARIAH-Esek garatua)

Premios de la Sociedad Internacional de Humanidades Digitales Hispánicas (HDH). VII. Edizioa. "Mejor herramienta, recurso e infraestructura desarrollada" izeneko saria. Saritua: CLARIAH-ES

IkerGazte kongresua. VI. Edizioa. Ingeniaritza eta Arkitektura alorreko artikulua eta aurkezpen onena. Saritua: Jaione Bengoetxea Azurmendi, "Euskara eta gaztelaniazko kontra-narratiben sorkuntza: datuen sorrera eta ebaluazioa" lanarengatik

Premios TFG-TFM MONDRAGON Sariak 2025. Euskarazko proiektu onenaren Eika saria: Irune Palacios Orueta "Euskarazko Medikuntzaren Domeinuko Lehenengo Corpus Sintetikoaren Sorrera: Metodologia eta Hastapenak" lanarengatik



www.hitz.eus

✉ hitz@ehu.eus

🦋 hitz-zentroa.bsky.social

📺 [hitz_zentroa](#) | [hitz-zentroa](#)

✂ [@hitz_zentroa](#)

in [hitz-zentroa](#)

📺 [@Hitz_zentroa](#) | [@hitz.zentroa](#)



HiTZ

Hizkuntza Teknologiako
Zentroa
Basque Center
for Language Technology