



**HITZ**

Hizkuntza Teknologiako  
Zentroa  
Basque Center  
for Language Technology



**LATXA**

Basque  
open  
chatbot

[latxa.hitz.eus](http://latxa.hitz.eus)

**Frontier  
research  
on Generative  
Artificial  
Intelligence**

Annual Report  
2025

# ORGANIZATION

HiTZ is a multidisciplinary research center on Language Technology and Generative Artificial Intelligence with members from seven departments of the University of the Basque Country. The objective of the center is to investigate language and speech technologies, with a significant effort towards the transfer of knowledge and technology to companies. It comprises two research groups Aholab and Ixa, both with extensive experience since 1993. Both IXA and Aholab have been evaluated as high-performance research groups in the last research evaluation exercise by the science agency of the Basque Government. During their history, the groups have participated in more than 200 research projects ranging from regional to European projects. It has also participated in more than 100 industrial contracts. HiTZ is a member of CLAIRE and a full member of BDVA and DAIRO. Through CLAIRE and BDVA, we participate in the European Partnership on Artificial Intelligence, Data and Robotics.

The members of the center are international referents in their scientific areas. It is formed by more than 100 members, including computer scientists, linguists and 5 research technicians. In the last five years, the researchers now in the center have published more than 200 scientific publications. The group is a leader in applying deep learning and generative AI techniques to text, speech and visual-language models. Its recent work in the area has been cited more than 4,000 times in the last two years.

HiTZ coordinates the CLARIAH-ES distributed research infrastructure, part of the European Strategy Forum on Research Infrastructures (ESFRI). Our deputy director, German Rigau, is the national coordinator of both infrastructures and leads the CLARIAH-ES consortium.

The impact of the center on society is boosted by the Chair of Artificial Intelligence and Language Technology (AI&LT), and the two masters coordinated by the center: the Erasmus Mundus+ European Master Program in Language and Communication Technologies (LCT) and the local program. The master, together with the bachelor degree on AI and the four courses for lifelong learning of professionals, are designed to meet the demands of industry and research in this rapidly growing field. HiTZ also offers a Doctoral Programme in Language Analysis and Processing.



**Eneko Agirre**

Director



**German Rigau**

Deputy Director



**Suna Şeyma Uçar**

Board Member



**Aritz Farwell**

Board Member



**Maite Oronoz**

Board Member



**Inma Hernaez**

Board Member



**Esther Miranda**

Board Member



**Aitor Soroa**

Board Member

# WORDS

# FROM THE DIRECTOR



**Eneko Agirre**  
Director of HiTZ

Generative Artificial Intelligence (GAI, or GenAI) has made remarkable progress in recent years, especially in Language Technology (LT). Its rapid rise is exemplified by Large Language Models (LLMs), as shown by the yearly advances of systems such as GPT or Gemini. Once their ability to communicate with humans was demonstrated, last year they leapt beyond language into programming and mathematics. Key to this progress have been new techniques such as Test-Time Compute and Reinforcement Learning with Verifiable Rewards (RLVR), which relies on verifiable values. At the same time, LLMs have been endowed with the ability to use tools and, through autonomy and coordination with other LLMs, this has led to an explosion of agents. The expansion of LLM capabilities has gone beyond text, incorporating seamless work with images, video, and speech, thereby extending into new applications, such as interacting with what we see on a screen. It is hard to believe that all of this has happened in less than a year.

In this context, tools based on GAI have a profound impact on many aspects of technology and society. This transformative era offers new opportunities to those capable of harnessing state-of-the-art GAI technology. However, the sudden emergence of GAI has also sparked important debates about responsible development and deployment, including the transformation of many professions, disinformation, harmful biases, high energy demands, and the widening digital divide between minority languages and dominant languages.

To face this new era and take advantage of the opportunities it offers, it is essential to promote open and public research in Language Technology and GAI: without local research, our society risks becoming a mere consumer of technologies developed elsewhere. Beyond LT, GAI is spreading to other technologies, including computer vision and robotics, and is driving a revolution in other scientific disciplines as well. For example, the 2024 Nobel Prize in Chemistry was awarded to the creators of AlphaFold, a technology based

on GAI, and GAI-based models have also become dominant in weather forecasting.

At HiTZ we fully embrace this new perspective and have structured our activity along two main lines: GAI and LT on the one hand, and GAI for the humanities and sciences on the other. Many of Spain's leading experts in GAI and LT are members of the HiTZ center and we are at the forefront of research in this field. At the same time, our multidisciplinary center continues to grow steadily. As a result, we have published a significant number of papers in top-tier conferences and journals, and we currently host more doctoral students than ever before. We lead Spain-wide artificial intelligence research networks and multi-year transformative projects. A clear example of this is Latxa, which since the release of its first version in 2024 has become the leading open Basque-language model. This achievement received the Best Paper Award at ACL, the most prestigious conference in the field. In 2025, we adapted Latxa to be able to converse with humans, reaching GPT-level performance in public human evaluations. Latxa is now available for testing, together with state-of-the-art open models for speech processing, under open licenses.

Finally, the HiTZ Chair in Artificial Intelligence and Language Technology enables us to train professionals and to transfer knowledge and technology to society. Likewise, the CLARIAH-ES research infrastructure coordinated by HiTZ has strengthened our national and international collaboration networks by providing digital tools and resources for the social sciences and humanities. In closing, by actively working to extend the benefits of GAI to new domains, we have developed pioneering models for visual languages, and together with neuroscientists we are working on an ambitious project to explore the interface between speech and the brain. All these successful initiatives provide HiTZ with a solid foundation to face future challenges with enthusiasm and, at the same time, position our country as a reference in research on Language Technology and Generative Artificial Intelligence.



# CLARIAH

HiTZ leads CLARIAH-ES, the distributed digital research infrastructure that steers Spain's participation in the European Research Infrastructure Consortia CLARIN (Common Language Resources and Technology Infrastructure) and DARIAH (Digital Research Infrastructure for the Arts and Humanities). CLARIAH-ES strives to enhance digitally-enabled research and provide researchers with access to state-of-the-art digital resources, tools, and services. Its multi- and interdisciplinary approach is designed to advance Spanish research in the social sciences and humanities, foster Spain's digital humanities community, and help strategically position Spain's researchers in international projects and programs, particularly within the European Research Area. German Rigau, Deputy Director of HiTZ, is Spain's National Coordinator for CLARIAH-ES, CLARIN, and DARIAH.

CLARIAH-ES currently includes partners from twelve leading institutions and research groups across Spain that are dedicated to advancing digital capabilities within the social sciences and humanities. In 2025 the infrastructure further consolidated its network of experts and increased its outreach towards the academic community and general public. In recognition of its work in this area, CLARIAH-ES received the best infrastructure award at the 7th Congress of the International Society for Hispanic Digital Humanities.

HiTZ also leads the CLARIAH-EUS node, which encompasses several institutions and groups dedicated to research in and for the Basque language in the social sciences and humanities. CLARIAH-EUS receives funding and support from the Basque Government, the Provincial Council of Gipuzkoa, the University of the Basque Country (EHU), and HiTZ.

More information available at: [www.clariah.es](http://www.clariah.es) and [www.clariah.eus](http://www.clariah.eus)



**German Rigau**

Deputy Director of HiTZ  
Spain's National Coordinator for  
CLARIAH-ES, CLARIN, and DARIAH



**Xabier Arregi**

Coordinator of CLARIAH-EUS

 **CLARIAH-ES**

 **CLARIAH-EUS**



In 2025 we consolidated the HiTZ Chair of Artificial Intelligence and Language Technology (AI&LT), launched in 2024 as part of the National Artificial Intelligence Strategy (ENIA). The Chair is the only funded chair in the Basque Country and the only chair in Spain dedicated to Language Technology (more information at [www.hitz.eus/katedra](http://www.hitz.eus/katedra)).

In this consolidation phase, the Chair has deepened its three pillars (education, research, and knowledge transfer) to maximize AI&LT's positive impact on society and the productive sector, strengthen research activities, and solidify the knowledge base that supports academic and professional programs. This work is carried out with our affiliated companies: Avature SL, Elhuyar Fundazioa, Ikerlan SCL, Euskaltel-MasOrange, Multiverse Computing, and Tecnalía Foundation for Research and Innovation. The relevance, number, and excellence of these organizations underline the productive sector's confidence in the Chair's potential to benefit society.

The Chair continues to offer four specialization courses in Generative AI, Deep Learning, and Language Technology—updated with the latest advances and designed for professionals, researchers, and students—taught by leading experts ([hitz.eus/training](http://hitz.eus/training)).

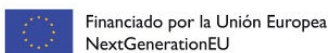


**Aitor Soroa**

Director of the HiTZ Chair of Artificial Intelligence and Language Technology



### Funded by:



### Partner companies:



# HiTZ IN NUMBERS

## Research & Transfer

31

Research projects

9

Knowledge transference projects

4

Other projects

7

Doctoral theses defended  
(3 International)

23

Journal papers  
(13 Q1)

41

Conference papers  
(15 A or A+)

3

Book chapters

1

Book







## KNOWLEDGE TRANSFER

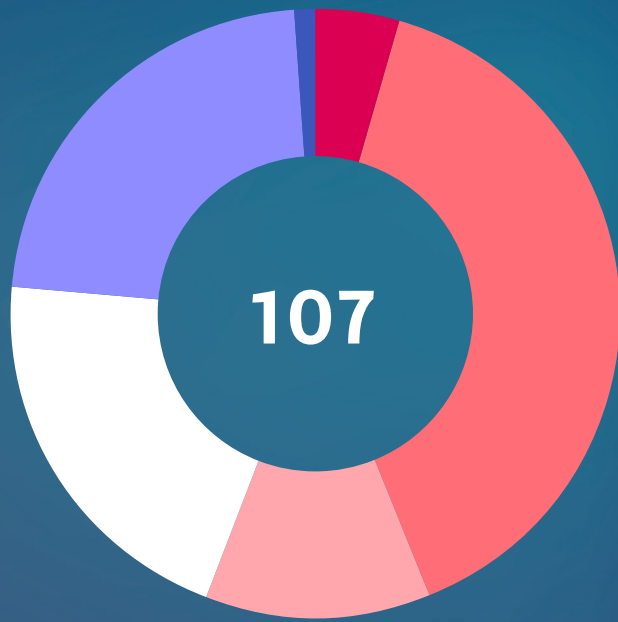
All the models, code, and data we have developed for Basque at the HiTZ center are publicly available. These include the best open chatbots, speech recognition models, and speech synthesis models for Basque, all of which allow commercial use.

[hitz.eus/results](https://hitz.eus/results)



# People

## Members

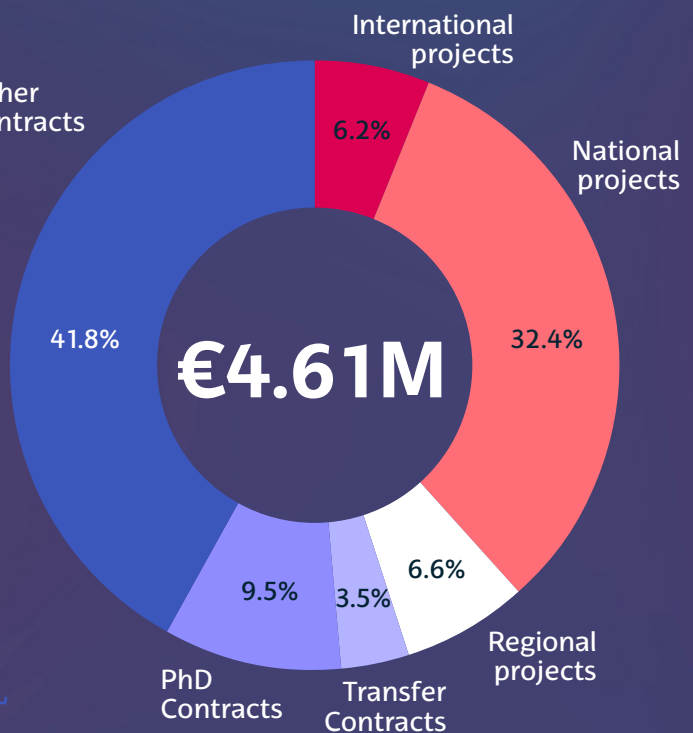
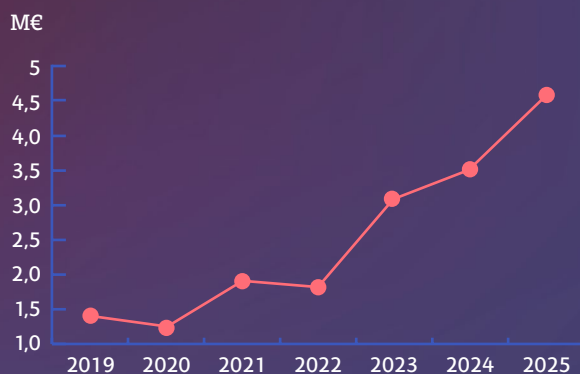
	Administrative and technical staff	5
	Lecturers	42
	Postdoctoral researchers	13
	Other researchers	22
	Funded predoctoral researchers	24
	Academic collaborator	1



## Students

	Master students	68
	PhD students	46

# Budget



# RESEARCH AREAS



Information  
Extraction  
and  
Information  
Retrieval

Main Researcher:



Aitor  
Soroa



Machine  
Translation

Main Researcher:



Gorka  
Labaka



Human-  
Computer  
Interaction

Main Researcher:



Gorka  
Azkune



Speech and  
Language  
Resources

Main Researcher:



Ainara  
Estarrona

# INFRASTRUCTURE

18

Multiprocessor  
GNU/Linux

1

SPARC  
Solaris  
server

8

H200 141GB  
vram GPU  
server

36

A100 80GB  
vram GPU  
cluster

56

A100 80GB vram  
GPU cluster  
(shared with DIPC)

14

L40 GPUs  
with 48GB  
vram

26

GPU  
various  
models

1.2

of raw Network  
storage capacity

PB



### Text Analysis

Main Researcher:



Rodrigo Agerri



### Speech Technologies

Main Researcher:



Inma Hernaez



### Medical and Legal domains

Main Researcher:



Arantza Casillas



### Digital humanities and education

Main Researcher:



Mikel Irukieta

# 1

acoustically isolated room

# 1

audio interface Midi 4x4 Behringer

# 1

Quiet PC Sentinel Fanless i10

# 1

Delsys Trigno™ Wireless EMG System

Access to 900,000 H100 GPU hours (valued at 2,200,000 euros) at the EuroHPC SuperComputer to research large models for European languages with few resources.



# TRAINING

68

Students  
in masters

46

Students  
in doctoral  
program

12

EMLTC  
Master Thesis  
finalized

21

HAP/LAP  
Master Thesis  
Finalized

95

Students in 4 LLM  
complementary  
courses

3

Ikasiker  
internships

7

HiTZ-funded  
internships



# ACTIVITIES

31  
Seminars



7  
Webinars

3  
Workshops

4  
Awards

Winning project in the 3rd Seminar Competition of the Duques de Soria Foundation for International Hispanism: "Humanidades digitales en acción: herramientas para el análisis de prensa histórica en español" (organized by CLARIAH-ES)

CLARIAH-ES awarded "Best Developed Tool, Resource, and Infrastructure" at the 7th Edition of the International Society of Hispanic Digital Humanities (HDH) Awards

Award for Best Article or Presentation in the area of Engineering and Architecture at the 6th edition of the IkerGazte Conference: Jaione Bengoetxea Azurmendi for the paper "Euskara eta gaztelaniazko kontratarratiben sorkuntza: datuen sorrera eta ebaluazioa"

Eika Prize for the best project in the Basque language at the 2025 TFG-TFM MONDRAGON Awards: Irune Palacios Orueta for the project "Euskarazko Medikuntzaren Domeinuko Lehenengo Corpus Sintetikoaren Sorrera: Metodologia eta Hastapenak"



# www.hitz.eus

✉ [hitz@ehu.eus](mailto:hitz@ehu.eus)

🦋 [hitz-zentroa.bsky.social](https://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