

# Language technology, the next frontier of Artificial Intelligence

Annual Report  
2022



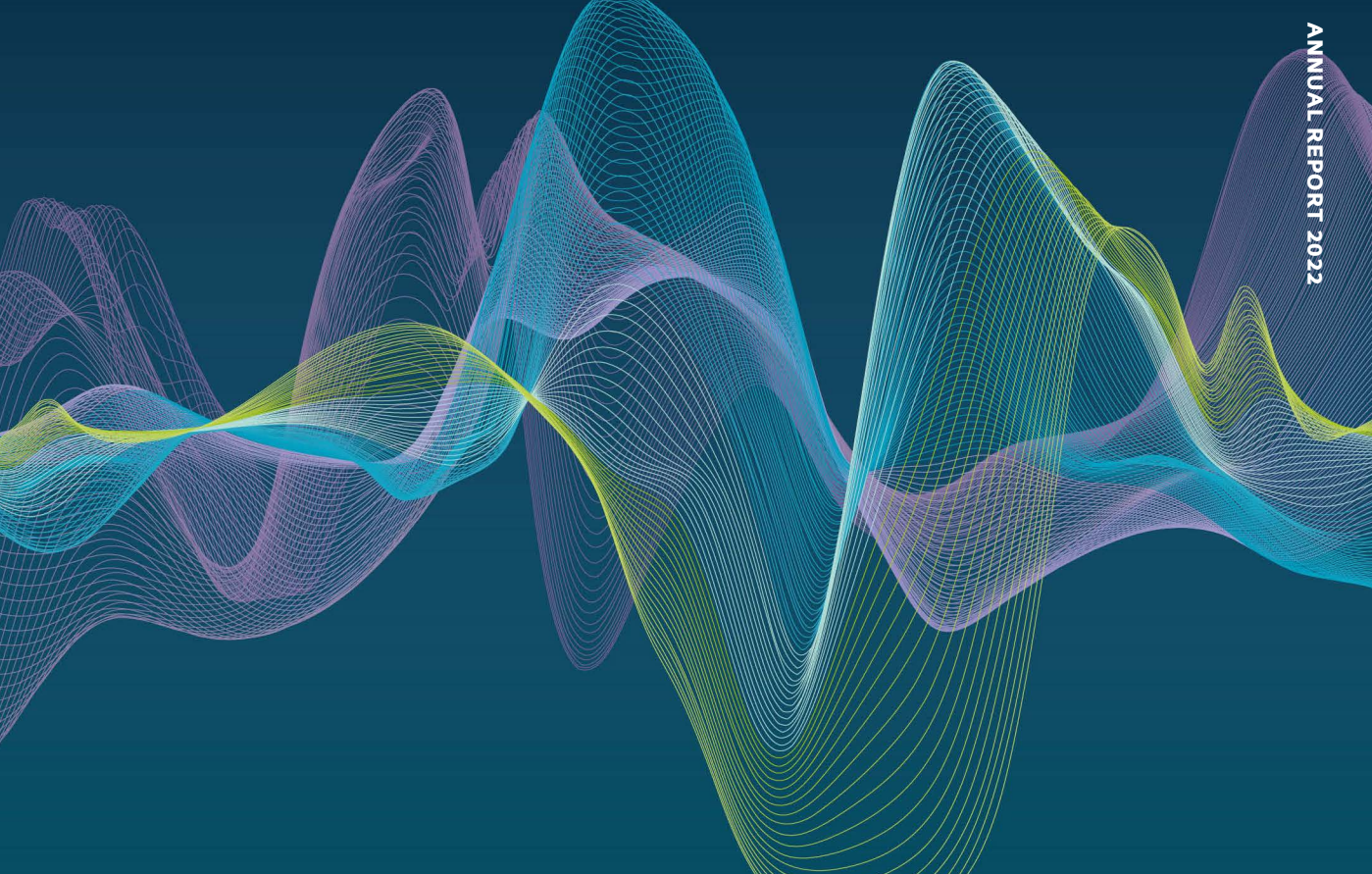
# INTRODUCTION



Language understanding is a main goal of artificial intelligence. Recent impressive advances using deep learning techniques that harness massive amounts of text and computing power have brought us to a new frontier. Indeed, we are witnessing what some researchers believe to be a renaissance, where large language models (LLM) aid people to accomplish repetitive tasks much more easily, allowing them to focus on more creative and challenging tasks. As with all disruptive technologies, there are also risks, including disinformation, harmful biases, high energy needs, and a widening of the digital divide for low-resource languages.

In order to face the new challenges and opportunities engendered by artificial intelligence and the digital transformation, it is essential to foster





open and public research in language technologies. Moreover, it is equally crucial to conduct research domestically, for without homegrown research we become mere consumers of technology created elsewhere. The benefits of such a stance are clear: research and results are made accessible to the local economy and society, while any associated risks may be easily audited.

We take this vision to heart at HiTZ. 2022 was a productive year that saw our center grow substantially. We organized a meeting for the PERTE on the New Economy of Language, which led to two significant projects funded by the PERTE and the Basque Government: NEL-GAITU and IKER-GAITU. These three-year projects will produce the next generation of language technologies for Basque, Catalan and Galician in close

collaboration with the main players in Galicia, Catalonia and the Valencia region. In addition, HiTZ led the incorporation of Basque and Spanish research institutions into the CLARIN and DARIAH European research infrastructures via the INTELE research network.

Three years ago we created HiTZ with a dual purpose: to be an international center of reference for language technology research and the computational processing of Basque. Our most recent achievements allow us to continue with renewed energy and, in cooperation with technology centers, companies and institutions, will help turn our country into an international hub for language technologies and artificial intelligence.

Eneko Agirre (Director of HiTZ) and German Rigau (Deputy Director of HiTZ)



# HiTZ IN NUMBERS

## Research & Transfer

35

Projects

5

Transfer Contracts

47

Publications



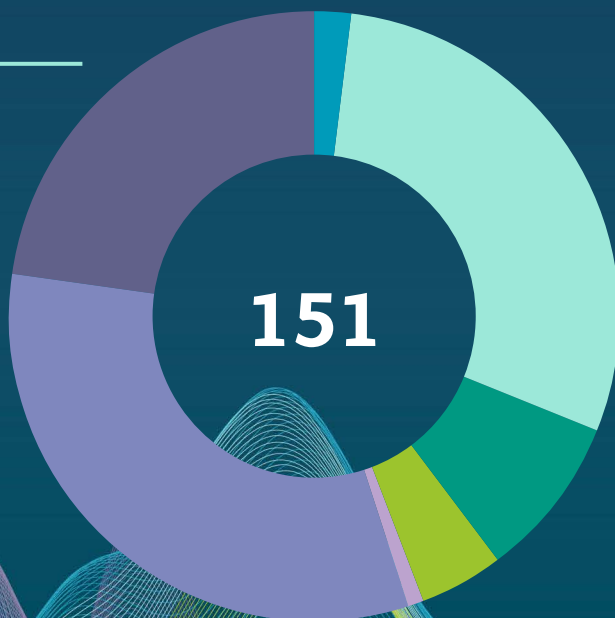
# People

## Members

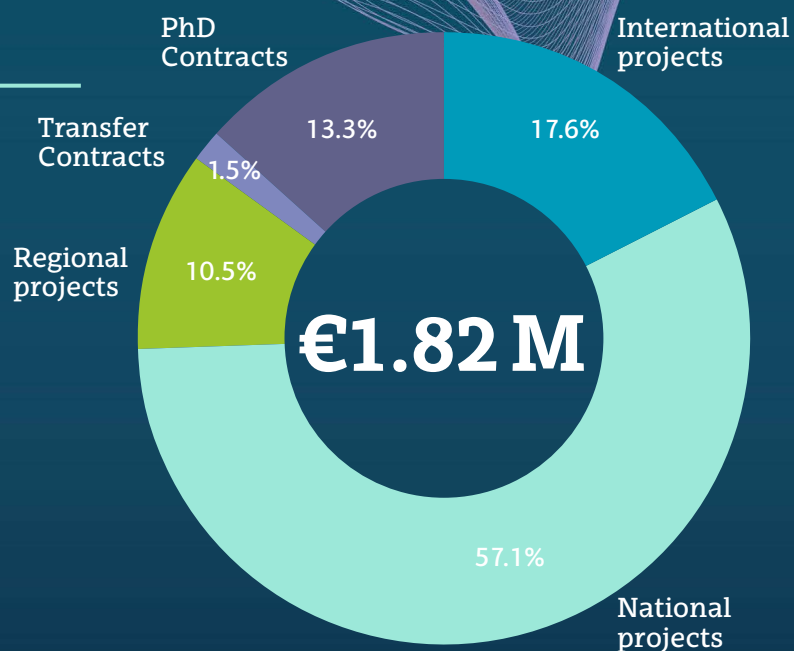
	Management and administration	3
	Lecturers	44
	Funded predoctoral researchers	13
	Postdoctoral researchers	7
	Ramón y Cajal	1

## Associated Members

	HAP/LAP students	49
	PhD students	34



# Budget





# ORGANIZATION

HiTZ is a multidisciplinary research center on **Language-centric 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, performing basic research, creating linguistic resources and tools and launching different commercial products on the market. HiTZ is also a **founding member** of **CLARIN-K Center**, member of **CLAIRE** and full member of **BDVA** and **DAIRO**. Through CLAIRE and BDVA, we also participate in the European Partnership on Artificial Intelligence, Data and Robotics.



**Eneko Agirre**

Director



**German Rigau**

Subdirector



**Ander Salaberria**

Vocal



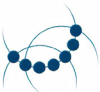
**Maite Oronoz**

Vocal

The members of the center are international referents in their scientific areas. At the moment, it is formed by **more than 60 members**, including computer scientists, linguists and 3 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 techniques to language processing and its recent work in the area has been **cited more than 4,000 times** in the last two years.

The members of the center have been **advisors** in the creation of the National Plan for Spanish Language Technologies and are currently advising the Basque Government's equivalent counterpart.

**CLARIN  
K CENTRE**



**CLAIRE**



**AIDA**

ARTIFICIAL INTELLIGENCE  
DOCTORAL ACADEMY



**BDV** BIG DATA VALUE  
ASSOCIATION



# ahōLAB



**Itziar  
Aldabe**

Vocal



**Inma  
Hernaez**

Vocal



**Esther  
Miranda**

Vocal



**Aitor  
Soroa**

Vocal

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** with the aim of transferring technology into the industry.

HiTZ is also a member of **Erasmus Mundus+ European Masters Program** in Language and Communication Technologies (LCT) **program**. It is designed to meet the demands of industry and research in the rapidly growing field of language technology. HiTZ also offers a **Doctoral**

**Programme** in Language Analysis and Processing, and is a full member of the International Artificial Intelligence Doctoral Academy (AIDA).

**The University of the Basque Country (UPV/EHU)** is the leading teaching and research institution in the Basque Country, a prosperous region stretching along the Atlantic coast of northern Spain. The UPV/EHU is among the best 400 universities in the world according to the Shanghai ranking, and has been recognized as an International Excellence Campus by the Spanish Government. The University of the Basque Country, a vibrant 30-year-old institution with 45,000 students, 5,000 world-class academic staff and state-of-the-art facilities distributed throughout 20 centers in its three campuses.



# 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

15

multiprocessor  
GNU/Linux  
servers

4

SPARC  
Solaris  
servers

1

HPC Cluster  
with 128  
cores

1

acoustically isolated room  
with audio equipment for  
professional recordings





## 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  
Iruskietia**

10 servers with  
**45**  
GPUs

Over  
**260**<sub>TB</sub>  
of raw Network  
storage capacity

**1**

Behringer 4x4  
audio/MIDI  
interface

**1**

Quiet PC  
Sentinel  
Fanless i10

# RESEARCH & TRANSFER

35

Research  
projects

5

Knowledge  
transference  
projects

5

Doctoral  
theses  
defended  
(3 International)

16

Journal  
papers  
(7 Q1)

25

Conference  
papers  
(8 A or A+)

3

Book  
chapters

10





49

Students  
in masters

34

Students  
in doctoral  
program

**TRAINING**

18

HAP/LAP  
Master Thesis  
Finalized

62

Students in 2 Deep  
Learning complementary  
courses

8

Ikasiker

16

Internal and  
external  
internships



# ACTIVITIES

# 21

Seminars

# 8

Webinars

# 3

Workshops

# 4

Awards

Gladys Award  
(awarded to Olatz Pérez de Viñaspre)

Aula Janssen UPV-EHU  
(awarded to Nuria Lebeña)

Institución del Año Award finalist  
(awarded to HiTZ)

Persona TIC Asociación de Ingeniería de  
Telecomunicación del País Vasco  
(awarded to Inma Hernández)