



La tecnología del lenguaje, la próxima frontera de la Inteligencia Artificial

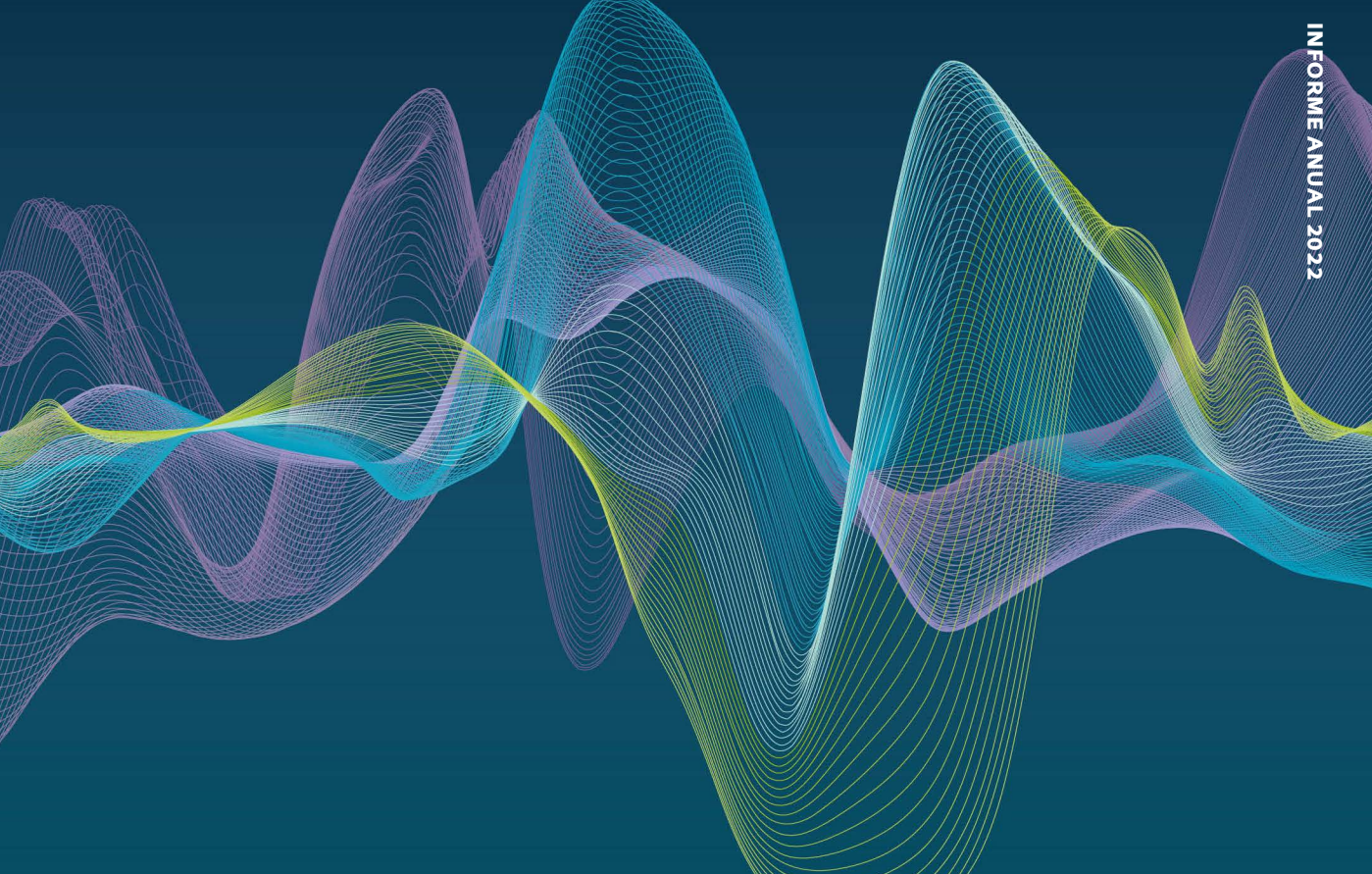
Informe anual
2022

INTRODUCCIÓN



La comprensión del lenguaje es uno de los principales objetivos de la inteligencia artificial. Los recientes e impresionantes avances en el uso de las técnicas de aprendizaje profundo combinadas con grandes cantidades de textos y capacidad de computación nos han llevado a una nueva frontera. De hecho, estamos siendo testigos de lo que, en opinión de algunos investigadores, es un nuevo renacimiento, en el que los grandes modelos del lenguaje facilitan el desempeño de tareas repetitivas, permitiendo a la gente centrarse en tareas más creativas y desafiantes. Al igual que sucede con todas las tecnologías disruptivas, también tiene sus riesgos, como la desinformación, los sesgos perjudiciales, las altas necesidades energéticas y el aumento de la brecha digital con las lenguas con pocos recursos.

Para hacer frente a los nuevos retos y oportunidades que generan la inteligencia artificial y la transformación digital, es



imprescindible fomentar la investigación abierta y pública en el área de la tecnología de la lengua. Es más, es fundamental que se investigue a nivel local, para evitar convertirnos en meros consumidores de la tecnología creada en otros países. Los beneficios de este planteamiento son claros: la investigación y los resultados se ponen a disposición de la economía y de la sociedad cercana, y permite auditar los riesgos fácilmente.

En HiTZ hacemos nuestra esta visión. 2022 ha sido un año fructífero en el que hemos visto crecer nuestro centro de forma considerable. Organizamos un congreso en torno al PERTE para la Nueva Economía de la Lengua que dió lugar a dos proyectos significativos financiados por el propio PERTE y el Gobierno Vasco: NEL-GAITU e IKER-GAITU. Estos proyectos trianuales, que se desarrollarán en estrecha colaboración con los principales agentes de Galicia, Cataluña y Comunidad

Valenciana, darán lugar a la próxima generación de tecnologías lingüísticas para el euskera, catalán y gallego. Además, HiTZ lideró, de la mano de la red de investigación INTELE, la integración de entidades investigadoras vascas y españolas en las infraestructuras de investigación europeas CLARIN y DARIAH.

Hace tres años creamos HiTZ con dos objetivos: ser un centro de referencia internacional para la investigación de la tecnología de las lenguas y serlo también para el procesamiento computacional del euskera. Nuestros logros nos animan a seguir con energía renovada y nos ayudarán a convertir nuestro país, con la colaboración de sus centros tecnológicos, empresas e instituciones, en un centro internacional para la tecnología de la lengua y la inteligencia artificial.

Eneko Agirre (Director del HiTZ) y German Rigau (Subdirector del HiTZ)

Investigación y transferencia

35

proyectos

5

contratos de
transferencia

47

publicaciones



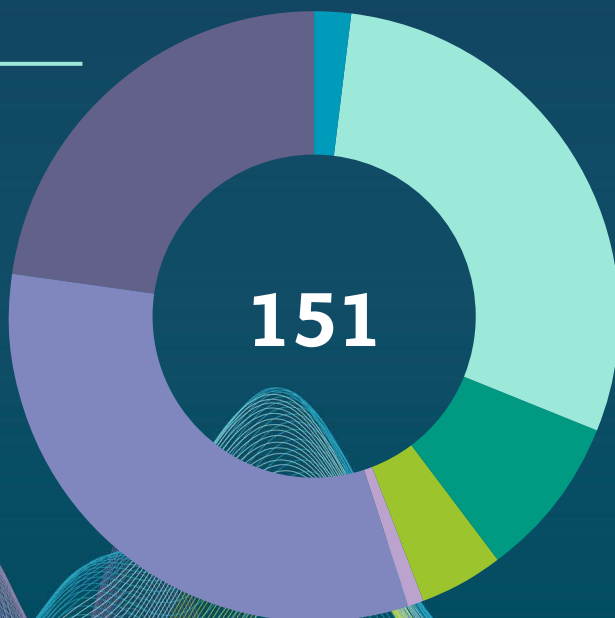
Personas

Miembros

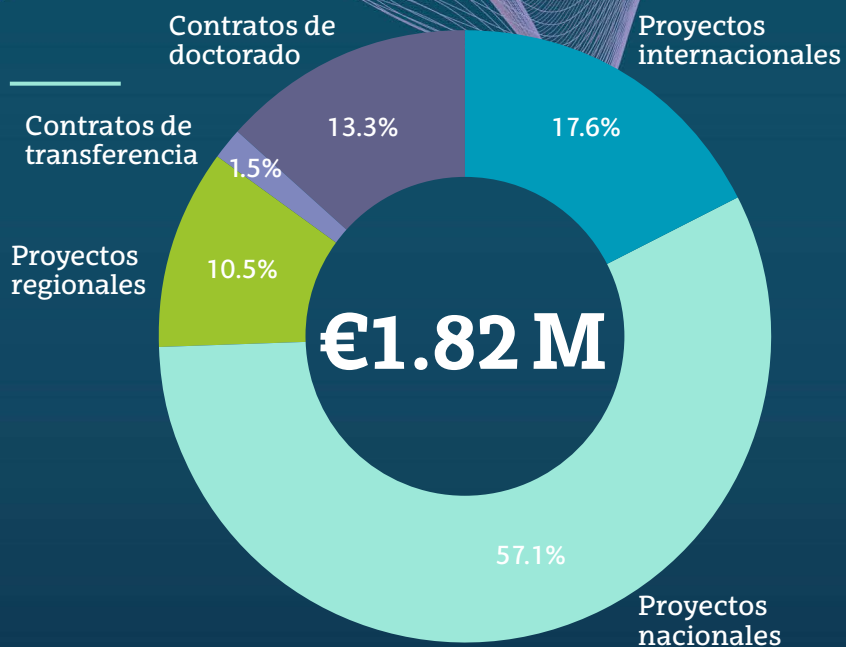
	Gestión y administración	3
	Profesores	44
	Becas de doctorado	13
	Investigadores postdoctorales	7
	Ramón y Cajal	1

Miembros Asociados

	Estudiantes de HAP/LAP	49
	Estudiantes de doctorado	34



Presupuesto



ORGANIZACIÓN

6

HiTZ es un centro de investigación multidisciplinar especializado en la **Inteligencia Artificial centrada en el lenguaje**, formado por miembros de 7 departamentos de la Universidad del País Vasco/Euskal Herriko Unibertsitatea. El objetivo del centro es **investigar** en el lenguaje y las tecnologías del habla, con un significativo esfuerzo en la **transferencia** de conocimiento y tecnología a las empresas. HiTZ está compuesto por dos grupos de investigación, Aholab e Ixa. Ambos poseen una amplia experiencia en investigación de base, creando recursos y herramientas lingüísticas y habiendo lanzado varios productos al mercado. HiTZ es **miembro fundador** de **CLARIN-K Center**, miembro de **CLAIRE** y miembro de pleno derecho de **BDVA** y **DAIRO**. A través de CLAIRE y BDVA, participamos en el European Partnership on Artificial Intelligence, Data and Robotics.

CLARIN
K CENTRE



CLAIRE



AIDA

ARTIFICIAL INTELLIGENCE
DOCTORAL ACADEMY



BDV BIG DATA VALUE
ASSOCIATION



Eneko
Agirre

Director



German
Rigau

Subdirector



Ander
Salaberria

Miembro de
la Junta



Maite
Oronoz

Miembro de
la Junta

Los miembros del centro son referentes internacionales en sus áreas científicas. En este momento HiTZ está formado por **más de 60 miembros** que incluyen informáticos, lingüistas y 3 técnicos de investigación. En el último lustro, los investigadores que forman el centro han publicado más de 200 **publicaciones**. El grupo es líder aplicando técnicas de aprendizaje profundo y, en los dos últimos años los trabajos realizados en éste área han sido **citados en más de 4000 ocasiones**. Los miembros del centro han realizado labores de asesoramiento en Plan Nacional de Tecnologías del Lenguaje y actualmente están asesorando al Gobierno Vasco en labores similares.

Tanto Ixa como Aholab han sido denominados



ahō LAB



**Itziar
Aldabe**

Miembro de
la Junta



**Inma
Hernaez**

Miembro de
la Junta



**Esther
Miranda**

Miembro de
la Junta



**Aitor
Soroa**

Miembro de
la Junta

grupos de alto rendimiento en la última evaluación realizada por el Consejo Científico del Gobierno Vasco. A lo largo de su historia, los grupos han participado en más de 200 **proyectos de investigación**, desde regionales hasta europeos. Además de participar en más 100 **contratos industriales** con el objetivo de transferir tecnología a la industria.

HiTZ es miembro de **Erasmus Mundus+ European Masters Program in Language and Communication Technologies (LCT)**, un programa diseñado para aunar las demandas tanto de la industria como de la investigación en un campo de rápido crecimiento como es el de la tecnología del lenguaje. HiTZ ofrece además

un **programa Doctoral** en Procesamiento del Lenguaje Natural, y es miembro de pleno derecho de la Academia Internacional de Doctorado en Inteligencia Artificial (AIDA).

La Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) es la principal entidad para la enseñanza e investigación en el País Vasco. La UPV/EHU está entre las 400 mejores universidades del mundo según el ranking Shanghai, y el Gobierno de España la ha reconocido como Campus de Excelencia. La Universidad del País Vasco es una institución de más de 30 años de antigüedad, con 45.000 alumnos, 5.000 académicos de nivel mundial e instalaciones de vanguardia, distribuidas en 20 centros en sus 3 campus.

ÁREAS DE INVESTIGACIÓN



Extracción y recuperación de información

Investigador/a principal:



Aitor Soroa



Traducción automática

Investigador/a principal:



Gorka Labaka



Interacción persona-computadora

Investigador/a principal:



Gorka Azkune



Recursos del habla y el lenguaje

Investigador/a principal:



Ainara Estarrona

INFRAESTRUCTURA

15

servidores multiprocesadores GNU/Linux

4

servidores de SPARC Solaris

1

HPC Cluster con 128 núcleos

1

sala aislada acústicamente con equipo de audio para grabaciones profesionales



Análisis de texto

Investigador/a principal:



**Rodrigo
Agerri**



Tecnologías de Voz

Investigador/a principal:



**Inma
Hernaez**



Dominios médico y legal

Investigador/a principal:



**Arantza
Casillas**



Humanidades digitales y educación

Investigador/a principal:



**Mikel
Iruskietta**

servidores con

45

GPUs

Más de

260^{TB}

de capacidad de
almacenamiento
bruto en red

1

Interfaz de
audio/MIDI
4x4 Behringer

1

Quiet PC
Sentinel
Fanless i10

INVESTIGACIÓN Y TRANSFERENCIA

35

proyectos de
investigación

5

proyectos de
transferencia de
conocimiento

5

tesis doctorales
defendidas
(3 internacionales)

16

artículos de
revistas
(7 Q1)

25

artículos en
congresos
(8 A o A+)

3

capítulos
de libro

10

CENTRO VASCO DE TECNOLOGÍA DEL LENGUAJE





49

estudiantes en
el Máster

34

estudiantes en
el programa de
doctorado

PRÁCTICAS

18

TFMs de
HAP/LAP
finalizadas

62

estudiantes en 2 cursos
complementarios de
Aprendizaje Profundo

8

Ikasiker

16

prácticas
internas
y externas

ACTIVIDADES

21

seminarios

8

webinarios

3

talleres

4

premios

Premio Gladys
(otorgado a Olatz Pérez de Viñaspre)

Aula Janssen UPV-EHU
(otorgado a Nuria Lebeña)

Finalista premio Institución del Año
Premios Archiletras de la Lengua
(otorgado a HiTZ)

Persona TIC
Asociación de Ingeniería de
Telecomunicación del País Vasco
(otorgado a Inma Hernández)