



HITZ

Hizkuntza Teknologiko
Zentroa
Basque Center
for Language Technology



LATXA

Chatbot
ouvert
en basque

latxa.hitz.eus

Recherche de Pointe sur l'Intelligence Artificielle Générative

Rapport annuel
2025

ORGANISATION

HiTZ est un centre de recherche multidisciplinaire de premier plan en technologies du langage et en intelligence artificielle générative (GenAI), rassemblant des experts de sept départements de l'Université du Pays Basque. La mission principale du centre est d'étudier les technologies du langage et de la parole, avec une forte orientation vers le transfert de savoir et de technologie vers le secteur industriel. HiTZ regroupe deux groupes de recherche, IXA et Aholab, qui possèdent une vaste expérience depuis 1993. Ces deux groupes ont été reconnus comme groupes de recherche de haut niveau par l'agence scientifique du Gouvernement Basque lors de sa dernière évaluation. Au fil des ans, ils ont participé à plus de 200 projets de recherche à l'échelle régionale, nationale et européenne, et réalisé plus de 100 contrats avec diverses entreprises. HiTZ est membre de la BDVA, et via cette association, il participe au Partenariat européen pour l'intelligence artificielle, les données et la robotique.

HiTZ rassemble plus de 100 membres, dont des informaticiens, des linguistes et cinq techniciens de recherche. Ils sont internationalement reconnus dans leurs domaines scientifiques. Au cours des cinq dernières années, les chercheurs du centre ont publié plus de 200 articles scientifiques. Le groupe est à la pointe de l'application des techniques d'apprentissage profond et de GenAI au texte, à la parole et aux modèles de langage visuel. Leurs travaux récents dans ce domaine ont été cités plus de 4.000 fois au cours des deux dernières années.

HiTZ coordonne également l'infrastructure de recherche distribuée CLARIAH-ES, le consortium qui supervise la participation de l'Espagne aux infrastructures de recherche européennes CLARIN et DARIAH, toutes les deux inscrites sur la feuille de route de l'ESFRI (European Strategy Forum on Research Infrastructures). Notre directeur adjoint, German Rigau, dirige le consortium CLARIAH-ES et est le coordinateur national de l'Espagne pour CLARIN et DARIAH.

L'impact du centre sur la société est renforcé par la Chaire en Intelligence Artificielle et Technologies du Langage (IA&TL), ainsi que par deux masters qu'il coordonne : le Master Erasmus Mundus+ européen en Technologies du Langage et de la Communication (LCT), et le Master en Analyse et Traitement du Langage. Ces programmes, accompagnés d'une licence en IA et de quatre formations continues à destination des professionnels, visent à répondre aux besoins de l'industrie et de la recherche dans ce domaine en pleine croissance. HiTZ propose également un programme doctoral en Analyse et Traitement du Langage.



**Eneko
Agirre**

Directeur



**German
Rigau**

Sous-directeur



**Suna Şeyma
Uçar**

Membre du
conseil



**Aritz
Farwell**

Membre du
conseil



**Maite
Oronoz**

Membre du
conseil



**Inma
Hernaez**

Membre du
conseil



**Esther
Miranda**

Membre du
conseil



**Aitor
Soroa**

Membre du
conseil

MESSAGE DU DIRECTEUR

Eneko Agirre
Directeur de HiTZ



L'intelligence artificielle générative (IAG ou GenAI) a connu des avancées remarquables ces dernières années, en particulier dans le domaine des technologies du langage (TL). Son essor rapide est illustré par les grands modèles de langage (GML ou LLM), comme en témoignent les progrès annuels de systèmes tels que GPT ou Gemini. Une fois leur capacité à communiquer avec les humains démontrée, ils ont franchi l'an dernier une nouvelle étape en s'étendant au-delà du langage vers la programmation et les mathématiques. Des techniques nouvelles ont été déterminantes dans cette évolution, notamment le Reinforcement Learning with Verifiable Rewards (RLVR), fondé sur des valeurs vérifiables, ainsi que le Test-Time Compute. Parallèlement, les GML ont été dotés de la capacité d'utiliser des outils et, grâce à leur autonomie et à leur coordination avec d'autres GML, cela a entraîné une véritable explosion d'agents. L'extension des capacités des GML est allée au-delà du texte, intégrant de manière transparente le traitement des images, de la vidéo et de la parole, ce qui a permis leur déploiement dans de nouvelles applications, comme l'interaction avec ce que nous voyons à l'écran. Il est difficile de croire que tout cela se soit produit en moins d'un an.

Dans ce contexte, les outils fondés sur l'IAG ont un impact profond sur de nombreux aspects de la technologie et de la société. Cette ère de transformation offre de nouvelles opportunités à ceux qui sauront tirer parti des technologies d'IAG les plus avancées. Toutefois, l'émergence soudaine de l'IAG a également suscité d'importants débats sur le développement et le déploiement responsables, notamment en ce qui concerne la transformation de nombreuses professions, la désinformation, les biais préjudiciables, les besoins énergétiques élevés et l'élargissement de la fracture numérique entre langues minoritaires et langues dominantes.

Pour faire face à cette nouvelle ère et exploiter les opportunités qu'elle offre, il est essentiel de promouvoir une recherche ouverte et publique en technologies du langage et en IAG : sans recherche locale, notre société risque de devenir un simple consommateur de technologies développées ailleurs. Au-delà des TL, l'IAG s'étend à d'autres technologies, telles que la vision artificielle ou la robotique, et provoque une révolution dans d'autres disciplines scientifiques. À titre d'exemple, le prix Nobel de chimie 2024 a été attribué aux créateurs

d'AlphaFold, une technologie fondée sur l'IAG, et des modèles basés sur l'IAG se sont également imposés dans la prévision météorologique.

À HiTZ, nous adoptons pleinement cette nouvelle approche et avons structuré notre activité autour de deux grands axes: l'IAG et les TL d'une part, et l'IAG au service des humanités et des sciences de l'autre. De nombreux experts espagnols de premier plan en IAG et en TL sont membres du centre HiTZ, et nous nous situons à l'avant-garde de la recherche dans ce domaine. Parallèlement, notre centre pluridisciplinaire est en croissance constante. En conséquence, nous avons publié un nombre significatif d'articles dans des conférences et revues de haut niveau, et nous accueillons aujourd'hui plus de doctorants que jamais. Nous dirigeons des réseaux de recherche nationaux en intelligence artificielle ainsi que des projets transformateurs pluriannuels. Un exemple emblématique est Latxa, qui, depuis la publication de sa première version en 2024, est devenu le principal modèle ouvert en langue basque. Cette réalisation a reçu le prix du meilleur article à ACL, la conférence la plus prestigieuse du domaine. En 2025, nous avons adapté Latxa afin qu'il puisse dialoguer avec des humains, atteignant un niveau comparable à celui de GPT dans des évaluations publiques humaines. Latxa est désormais disponible en version de test, aux côtés des meilleurs modèles ouverts de traitement de la parole, sous licences ouvertes.

Enfin, la Chaire HiTZ en intelligence artificielle et technologies du langage nous permet de former des professionnels et de transférer connaissances et technologies à la société. De même, l'infrastructure de recherche CLARIAH-ES, coordonnée par HiTZ, a renforcé nos réseaux de collaboration nationaux et internationaux en fournissant des outils et des ressources numériques pour les sciences sociales et les humanités. Pour conclure, en travaillant activement à étendre les bénéfices de l'IAG à de nouveaux domaines, nous avons développé des modèles pionniers de langages visuels et collaborons avec des neuroscientifiques dans un projet ambitieux visant à explorer l'interface entre la parole et le cerveau. L'ensemble de ces initiatives réussies constitue une base solide permettant à HiTZ d'aborder avec enthousiasme les défis futurs et positionne, en parallèle, notre pays comme une référence en matière de recherche en technologies du langage et en intelligence artificielle générative.

CLARIAH

HiTZ dirige CLARIAH-ES, l'infrastructure de recherche numérique distribuée qui coordonne la participation de l'Espagne aux consortiums européens d'infrastructures de recherche CLARIN (Infrastructure commune pour les ressources et les technologies linguistiques) et DARIAH (Infrastructure numérique de recherche pour les arts et les humanités). CLARIAH-ES vise à renforcer la recherche soutenue par le numérique et à offrir aux chercheurs un accès à des ressources, outils et services numériques de pointe. Son approche pluri- et interdisciplinaire est conçue pour faire progresser la recherche espagnole en sciences sociales et en humanités, favoriser la communauté espagnole des humanités numériques et contribuer au positionnement stratégique des chercheurs espagnols dans les projets et programmes internationaux, en particulier dans le cadre de l'Espace européen de la recherche. German Rigau, directeur adjoint de HiTZ, est le coordinateur national de l'Espagne pour CLARIAH-ES, CLARIN et DARIAH.

CLARIAH-ES réunit actuellement des partenaires issus de douze institutions et groupes de recherche de premier plan à travers l'Espagne, consacrés au développement des capacités numériques dans le domaine des sciences sociales et des humanités. En 2025, l'infrastructure a encore consolidé son réseau d'experts et renforcé sa visibilité auprès de la communauté universitaire et du grand public. En reconnaissance de son travail dans ce domaine, CLARIAH-ES a reçu le prix de la meilleure infrastructure lors du 7e Congrès de la Société Internationale des Humanités Numériques Hispaniques.

HiTZ dirige également le nœud CLARIAH-EUS, qui regroupe plusieurs institutions et équipes dédiées à la recherche sur la langue basque et pour la langue basque dans le domaine des sciences sociales et des humanités. CLARIAH-EUS bénéficie d'un financement et d'un soutien du Gouvernement basque, du Conseil provincial de Gipuzkoa, de l'Université du Pays basque (EHU) et de HiTZ.

Pour en savoir plus, cliquez ici:

www.clariah.es et www.clariah.eus



German Rigau

Directeur adjoint de HiTZ
Coordinateur national à l'Espagne
de CLARIAH-ES, CLARIN et DARIAH



Xabier Arregi

Coordinateur de CLARIAH-EUS

 **CLARIAH-ES**

 **CLARIAH-EUS**

CHAIRE

En 2025, nous avons consolidé la Chaire HiTZ d'intelligence artificielle et de technologie du langage (AI<), lancée en 2024 dans le cadre de la Stratégie nationale pour l'intelligence artificielle (ENIA). Cette chaire est la seule chaire financée du Pays basque et la seule en Espagne consacrée à la technologie du langage (plus d'informations sur www.hitz.eus/katedra).

Au cours de cette phase de consolidation, la Chaire a approfondi ses trois piliers (formation, recherche et transfert de connaissances) afin de maximiser l'impact positif d'AI< sur la société et le secteur productif, de renforcer les activités de recherche et de consolider la base de connaissances qui soutient les programmes académiques et professionnels. Ce travail est mené avec nos entreprises partenaires : Avature SL, Elhuyar Fundazioa, Ikerlan SCL, Euskaltel-MasOrange, Multiverse Computing et Tecnalia Foundation for Research and Innovation. L'importance, le nombre et l'excellence de ces organisations soulignent la confiance du secteur productif dans le potentiel de la Chaire à bénéficier à la société.

La Chaire continue de proposer quatre cours de spécialisation en IA générative, apprentissage profond et technologie du langage — mis à jour avec les dernières avancées et conçus pour les professionnels, les chercheurs et les étudiants —, dispensés par des experts de premier plan (hitz.eus/training).



HiTZ

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

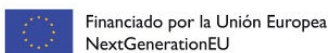


Aitor Soroa

Directeur de la Chaire HiTZ d'Intelligence Artificielle et de Technologie Linguistique



Funded by:



Partner companies:



HiTZ EN CHIFFRES

Recherche et transfert

31

Projets de
recherche

9

Contrats de
transfert de
connaissances

4

Autres
projets

7

Thèses de doctorat
soutenues
(3 internationales)

23

Articles de
revue
(13 Q1)

41

Communications
de congrés
(15 A ou A+)

3

Chapitres
de livre

1

Livre







TRANSFERT

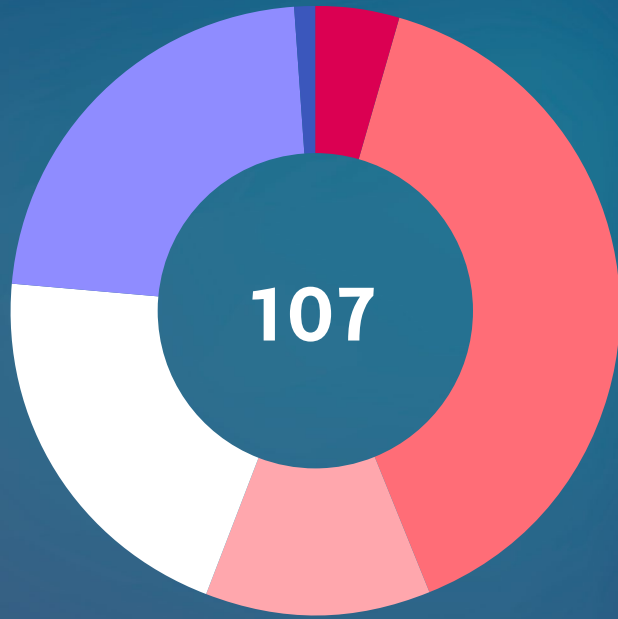
Tous les modèles, codes et données que nous avons développés pour le basque au centre HiTZ sont accessibles au public. Ils comprennent les meilleurs chatbots ouverts, modèles de reconnaissance vocale et modèles de synthèse vocale pour le basque, autorisant un usage commercial.

hitz.eus/fr/resultats



Personnes

Membres

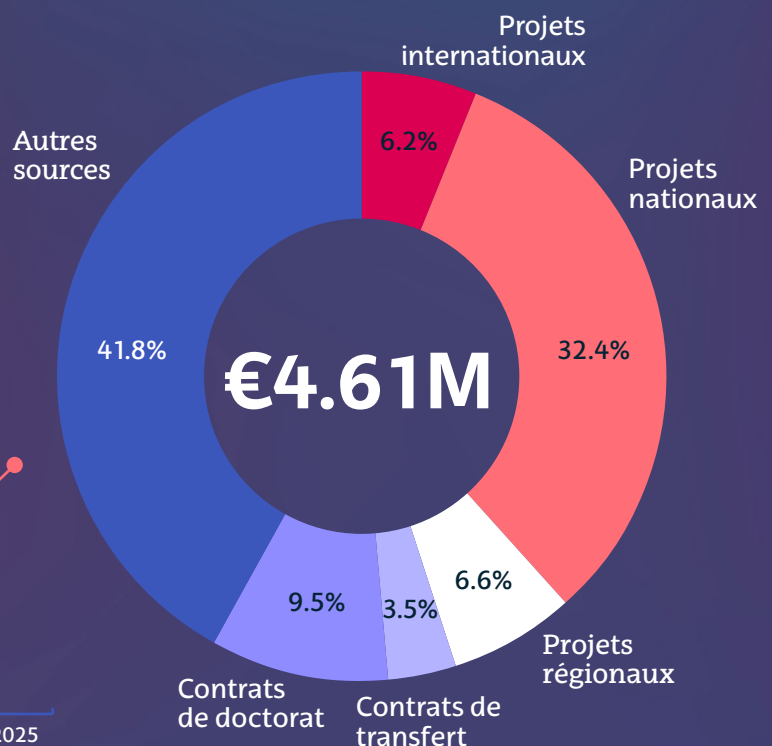
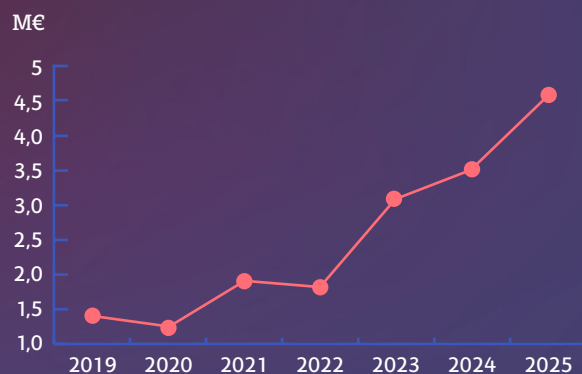
	Gestion et administration	5
	Enseignants	42
	Chercheurs postdoctoraux	13
	Autres chercheurs	22
	Chercheurs titulaires d'une bourse prédoctorale	24
	Collaborateur académique	1



Étudiants

	Étudiants en master	68
	Étudiants en doctorat	46

Budget



DOMAINES DE RECHERCHE



Chercheur principal :

**Extraction
et recherche
d'information**



**Aitor
Soroa**



Chercheur principal :

**Traduction
automatique**



**Gorka
Labaka**



Chercheur principal :

**Interaction
homme-
machine**



**Gorka
Azkune**



Chercheur principal :

**Ressources
vocales et
langagières**



**Ainara
Estarrona**

INFRASTRUCTURE

18

multiprocesseur
GNU/Linux

1

serveur
SPARC
Solaris

8

serveurs
H200 141GB
vram GPU

36

GPU cluster
A100 80GB
vram

56

GPU cluster
A100 80 GB vram
(partagé / DIPC)

14

L40 GPUs
avec 48GB
vram

26

GPU
divers
modèles

1.2

PB
de capacité de
stockage brute
en réseau



Analyse de textes

Chercheur principal :



Rodrigo Agerri



Technologies de la parole

Chercheur principal :



Inma Hernaez



Domaines médical et juridique

Chercheur principal :



Arantza Casillas



Sciences humaines numériques et éducation

Chercheur principal :



Mikel Iruskietia

1

salle acoustiquement isolée

1

interface Behringer 4x4 audio/MIDI

1

Quiet PC Sentinel Fanless i10

1

Delsys Trigno™ Wireless EMG System

Accès à 900.000 heures de GPU (d'une valeur de 2,2 millions d'euros) au Superordinateur EuroHPC pour la recherche de grands modèles pour les langues européennes avec peu de ressources



FORMATION

68

Étudiants
en masters

46

Étudiants en
doctorat

12

Thèses du
master EMLTC
soutenues

21

Thèses du
master HAP/LAP
soutenues

95

Étudiants de 4
cours de formation
continue liés au LLM

3

Ikasiker

7

Stages internes
et externes



ACTIVITÉS

31

Séminaires



7

Webinaires

3

Workshops

4

Prix

Projet lauréat du III^e Concours de Séminaires de la Fondation Duques de Soria d'Hispanisme International: "Humanidades digitales en acción: herramientas para el análisis de prensa histórica en español" (développé par CLARIAH-ES)

Prix de la "Meilleure outil, ressource ou infrastructure développée" décerné à CLARIAH-ES lors de la VII^e édition des Prix de la Société Internationale d'Humanités Numériques Hispanophones (HDH)

Prix du Meilleur Article ou Présentation dans le domaine de l'ingénierie et de l'architecture, lors de la VI^e édition du congrès IkerGazte: Jaione Bengoetxea Azurmendi pour le travail "Euskara eta gaztelaniazko kontra-narratiben sorkuntza: datuen sorrera eta ebaluazioa"

Prix TFG-TFM MONDRAGON Saria 2025. Prix Eika au meilleur projet en langue basque: Irune Palacios Orueta pour le travail "Euskarazko Medikuntzaren Domeinuko Lehenengo Corpus Sintetikoaren Sorrera: Metodologia eta Hastapenak"



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