

# Erik Sarriegui

Madrid, España

[linkedin.com/in/eriksarriegui](https://www.linkedin.com/in/eriksarriegui)

+34 679 75 85 68

[erik@sarriegui.es](mailto:erik@sarriegui.es)

[eriksarriegui.github.io](https://eriksarriegui.github.io)



## SOBRE MI

---

*Científico de Datos en formación con una sólida base autodidacta en Inteligencia Artificial y desarrollo de software. Mi fascinación por la capacidad de los modelos de IA para interpretar información compleja me ha llevado a especializarme en este campo a través de proyectos personales y formación continua. Actualmente estoy cursando un Máster en Ciencia de Datos, buscando combinar mi entusiasmo por el aprendizaje con una oportunidad profesional que me permita aplicar y expandir mis conocimientos en el análisis de datos y machine learning. Mi objetivo es contribuir al desarrollo de soluciones innovadoras en el campo de la IA mientras continúo creciendo profesionalmente.*

## EXPERIENCIA

---

### **Vicomtech**

Desarrollador de Negocio en prácticas

Donostia y Bilbao, España

Mayo 2024 – Octubre 2024

Como parte del departamento de Salud Digital y Tecnologías Biomédicas, investigué la aplicación de Grandes Modelos de Lenguaje aplicados a Salud Mental en un ámbito clínico. En este contexto:

- Realicé ajustes a modelos de lenguaje utilizado Transformers y PyTorch. Uno de estos modelos, cuyo código es accesible, es MediFlow ([Modelo](#), [Artículo](#)) que realiza triajes automáticos con una precisión del 90%.
- Construí una demo con la que testear el interés de potenciales clientes y colaboradores en la solución propuesta.

### **Experiencia Internacional**

Estudio de Tendencias de Digitalización del Sector Salud Surcoreano

Seúl, Corea del Sur

Enero 2023 – Abril 2023

Se trató de un estudio respaldado por dos hospitales del País Vasco en el que se analizaron tanto las tendencias digitales del sector sanitario surcoreano como herramientas y proyectos tecnológicos específicos.

## FORMACIÓN

---

### **Máster en Ciencia de Datos**

Universidad Complutense de Madrid

Madrid, España

2025 - 2026

### **Grado en Liderazgo Emprendimiento e Innovación**

Universidad de Mondragón

Irún, España

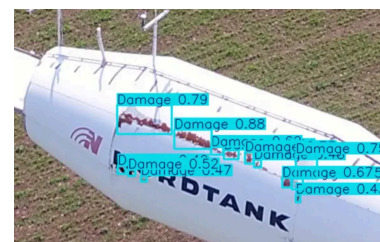
2020 – 2024

## PROYECTOS & HITOS

---

### **DeepWind**

Inspirado por proyectos de drones autónomos que se utilizan para la inspección de palas de aerogeneradores en movimiento. Desarrollé DeepWind, un sistema de Visión Artificial basado en técnicas de Aprendizaje Profundo capaz de detectar defectos en Aerogeneradores. He publicado todo el proyecto en GitHub ([link](#)) con el objetivo de que esté accesible para todos.



GPT-Bi

Se trata de un modelo de lenguaje abierto basado en Transformers (arquitectura de GPT-2) pre entrenado desde cero para experimentar con la generación de texto en euskera. El proyecto tiene como objetivo principal avanzar en el campo del procesamiento del lenguaje natural (PLN) para lenguas minoritarias, como el euskera, que históricamente han tenido menos recursos y modelos. El modelo está accesible a través de Hugging Face ([link](#)).

Contradictory My Dear Watson

Obtuve el primer puesto en la competición "Contradictory, My Dear Watson" de Kaggle, centrada en Inferencia de Lenguaje Natural (NLI). El desafío consistía en desarrollar un modelo capaz de determinar la relación semántica entre pares de frases en quince idiomas diferentes, lo que añadió un alto nivel de complejidad al problema. Para alcanzar este resultado, implementé y adapté una versión del modelo RoBERTa utilizando la librería Transformers de Hugging Face. Optimicé el proceso de entrenamiento para lograr una precisión superior al 89%, demostrando un gran control sobre LLMs.

#	Team	Members	Score	Entries	Last	Join
1	Erik Perez		0.89894	3	20h	
2	Minaash Ahmad		0.80630	2	23h	

IDIOMAS

- **Español:** Nativo
- **Inglés:** B2. Cuento con plena capacidad para trabajar en inglés, como lo demuestra un estudio de mercado realizado en Corea, desarrollado íntegramente en este idioma. Además, leo regularmente artículos científicos en inglés.
- **Euskera:** B2

SKILLS

- **Aprendizaje autónomo:** Desarrollo autodidacta en Ciencia de Datos demostrado a través de proyectos competitivos y contribuciones en GitHub, con capacidad para adquirir y aplicar nuevos conocimientos de forma independiente.
- **Adaptabilidad:** Alta capacidad para adaptarme a nuevos entornos y metodologías de trabajo, evidenciada tanto en proyectos internacionales como en colaboraciones con diferentes empresas y sectores.
- **Orientado a resultados:** Historial probado de éxitos en competiciones técnicas y desarrollo de soluciones empresariales, manteniendo altos estándares de calidad y cumplimiento de objetivos.
- **Comunicación técnica:** Experiencia presentando conceptos complejos a diferentes audiencias, desde demostraciones técnicas a clientes hasta documentación de proyectos en código abierto.

VOLUNTARIADOS

<b>Orden de las Misioneras de la Caridad   Madre Teresa de Calcuta</b> Orfanato Daya Dan, para niños con discapacidad	Calcuta, India Abril 2023
<b>Fundación Parque La Libertad</b> Monitor	San José, Costa Rica Febrero 2022
<b>Asociación Gipuzkoana de Ilusionismo</b> Asociado y, desde el 2024, Tesorero	Donostia, España 2018 - actualidad

# Erik Sarriegui

Madrid, Spain

[linkedin.com/in/eriksarriegui](https://www.linkedin.com/in/eriksarriegui)

+34 679 75 85 68

[erik@sarriegui.es](mailto:erik@sarriegui.es)

[eriksarriegui.github.io](https://eriksarriegui.github.io)



## ABOUT ME

---

*Aspiring Data Scientist with a strong self-taught foundation in Artificial Intelligence and software development. My fascination with AI models' ability to interpret complex information has driven me to specialize in this field through personal projects and continuous learning. I am currently pursuing a Master's in Data Science, seeking to combine my passion for learning with a professional opportunity that allows me to apply and expand my knowledge in data analysis and machine learning. My goal is to contribute to the development of innovative AI solutions while continuing to grow professionally.*

## EXPERIENCE

---

### **Vicomtech**

Business Developer Intern

Donostia and Bilbao, Spain

May 2024 – Oct 2024

As part of the Digital Health and Biomedical Technologies department, I researched the application of Large Language Models in Mental Health within a clinical setting. In this context:

- I fine-tuned language models using Transformers and PyTorch. One of these models, whose code is publicly accessible, is MediFlow ([Model](#), [Article](#)), which performs automatic triage with 90% accuracy.
- I built a demo to test the interest of potential clients and collaborators in the proposed solution.

### **International Experience**

Study on Digitalization Trends in the South Korean Healthcare Sector

Seoul, South Korea

Jan 2023 – Apr 2023

It was a study supported by two hospitals in the Basque Country, which analyzed both the digital trends in the South Korean healthcare sector and specific technological tools and projects.

## EDUCATION

---

### **Master in Data Science**

Universidad Complutense de Madrid

Madrid, Spain

2025 - 2026

### **Bachelor's Degree in Leadership, Entrepreneurship, and Innovation**

Universidad de Mondragón

Irún, Spain

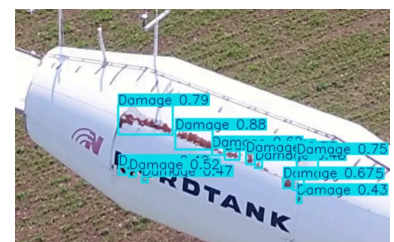
2020 – 2024

## PROJECTS & MILESTONES

---

### **DeepWind**

Inspired by projects involving autonomous drones used for inspecting wind turbine blades in motion, I developed DeepWind, a Computer Vision system based on Deep Learning techniques capable of detecting defects in wind turbines. I have published the entire project on GitHub ([link](#)) to make it accessible to everyone.



GPT-Bi

It is an open-language model based on Transformers (GPT-2 architecture) pre-trained from scratch to experiment with text generation in Basque. The project's main goal is to advance the field of natural language processing (NLP) for minority languages, such as Basque, which have historically had fewer resources and models. The model is accessible through Hugging Face (link).

Contradictory My Dear Watson

I achieved first place in Kaggle's "Contradictory, My Dear Watson" competition, which focused on Natural Language Inference (NLI). The challenge involved developing a model capable of determining the semantic relationship between pairs of sentences in fifteen different languages, adding a high level of complexity to the problem. To achieve this result, I implemented and adapted a version of the RoBERTa model using the Hugging Face Transformers library. I optimized the training process to achieve an accuracy above 89%, demonstrating strong expertise in LLMs.

#	Team	Members	Score	Entries	Last	Join
1	Erik Perez		0.89894	3	20h	
2	Alimash Zhondot		0.80670	5	25h	

LANGUAGES

- **Spanish:** Native
- **English:** B2. I am fully capable of working in English, as demonstrated by a market study conducted in Korea, developed entirely in this language. Additionally, I regularly read scientific articles in English.
- **Basque:** B2

SKILLS

- **Autonomous Learning:** Self-taught development in Data Science demonstrated through competitive projects and contributions on GitHub, with the ability to acquire and apply new knowledge independently.
- **Adaptability:** High ability to adapt to new environments and work methodologies, demonstrated in both international projects and collaborations with different companies and sectors.
- **Results-oriented:** Proven track record of success in technical competitions and business solution development, maintaining high standards of quality and meeting objectives.
- **Technical Communication:** Experience presenting complex concepts to diverse audiences, from technical demonstrations to clients to project documentation.

VOLUNTEERING

<b>Order of the Missionaries of Charity   Mother Teresa of Kolkata</b> Daya Dan Orphanage, for Children with Disabilities	Kolkata, India Apr 2023
<b>Fundación Parque La Libertad</b> Instructor	San José, Costa Rica Feb 2022
<b>Gipuzkoa's Association of Illusionism</b> Member and, since 2024, Treasurer	Donostia, España 2018 - present