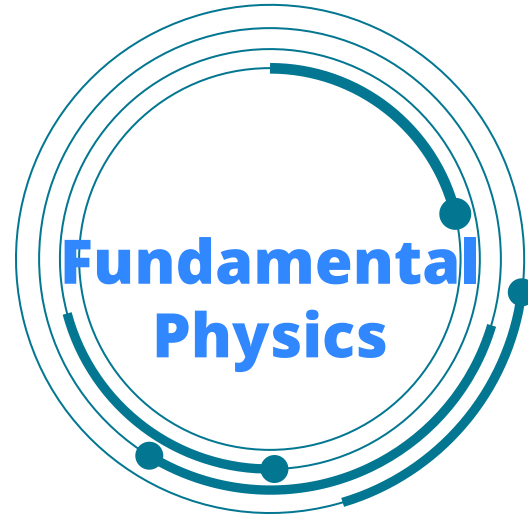


Data Science in



and the ridge to  
industry & society

**Paul Van Branteghem**

June 2024



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



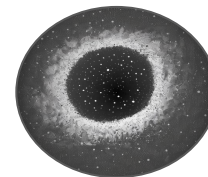
[@pauvanbr](https://twitter.com/pauvanbr)

[paul@spain-ai.com](mailto:paul@spain-ai.com)

# About me

## PAUL VAN BRANTEGHEM

- Development team for the national AI strategy at SEDIA (Ministry of Economy) contracted by INECO
- Co-founder of Spain AI, a leading association in AI dissemination
- Freelance and CTO co-founder of Big Onion
- Physicist with a master's degree in advanced AI, meteorology, and a higher program in Digital Analytics and CRM
- Professor in various Data Science / Big Data programs (ISDI, CUNEF, EOI, INTEF)
- Experience at Siemens-Gamesa, in the AI department at BBVA, Deloitte Digital, and StratioBD



BIG ONION





A mind map diagram with a central dark blue cloud containing the text "Spain AI Goals". Five lines radiate from the cloud to five different goal areas: "share AI knowledge", "Training", "Projects", "Networking & Collaboration", and "Non-profit association with the aim of democratizing Artificial Intelligence." Each goal area includes a brief description and a circular icon with a letter (D, T, P, N).

## Spain AI Goals

### share AI knowledge

AI events, podcasts and other initiatives with the aim of showcasing the capabilities of these technologies in real projects

D

### Training

Training for young people, administrations, and other entities with the aim of raising awareness about the potential of AI.

T

### Projects

Development of initiatives such as hackathons or development projects.

P

### Networking & Collaboration

Establish a network of contacts between start-ups, research centers and other companies that allows joint projects.

N

Non-profit association with the aim of democratizing Artificial Intelligence.  
We focus on makers, people who carry out these projects.  
We aim to be an important lever at the national level to help Spain grow.



SPAIN AI

# About Spain AI

## 2023 Report

### newsletter impacts 2023

**108.5K**

newsletter impact

### Events 2023

**16**

7 online and 9 hybrid events in different locations (Cádiz, Madrid, Zaragoza, Valladolid, Málaga,...)

**1.3K**

registered attendees at our events

**33.2K**

Visualizations on our Youtube Channel



### Our followers

**5.1k**

Subscribers on **Youtube**

**9.9k**

Followers on **LinkedIn**

**6.5k**

Followers on **Twitter**

**1.5k**

Subscribers on **MeetUp**

**2.5k**

Followers on **Eventbrite**

### Our Podcast 2023

**7.3k**

plays at 2023

**1.5k**

followers on Spotify

**30 regions in our network**

**80** active members in the network

# Index

- ❖ Data Science
- ❖ Roles in Data Science
- ❖ From academia/research to industry

# Index

- ❖ **Data Science**

- ❖ Roles in Data Science

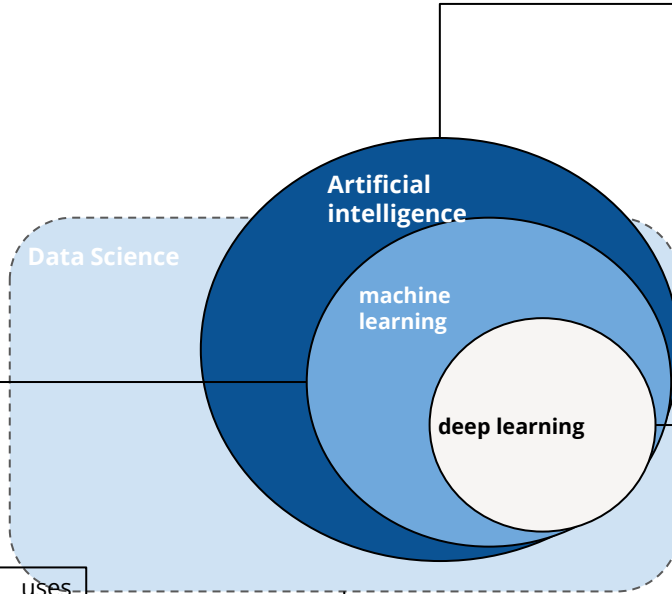
- ❖ From academia/research to industry

# Introduction to DS and AI

Artificial Intelligence, Machine Learning, Deep Learning and Data Science are different concepts.



A subset of AI that focuses on developing systems that learn, or improve performance, based on the data they consume.

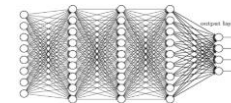
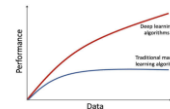


Set of techniques that tries to mimic the human intelligence processes by machines.



Area of study that uses scientific methods to obtain knowledge from existing data and support decision-making.

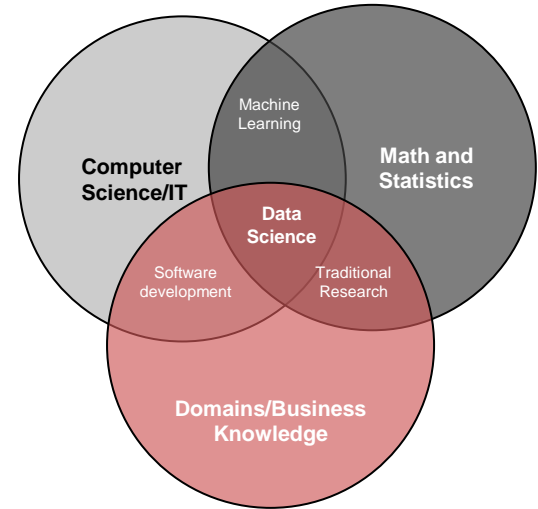
It is a set of algorithms based on multilayer neural networks, capable of capturing deep interactions.



# Data Science

When we talk about Data Science, we talk about a project typology that mixes 3 main domains: computer science, business knowledge and statistics.

- **Computer Science:** Use computer science techniques and tools to create algorithms, find patterns, ask questions, and launch experiments..
- **Business Knowledge:** Know what is the main objective, limitations, data and requirements of the project.
- **Statistics:** Define the best data transformations, cleaning, understand how the model works.





## Roles Data Science Projects



There is generally no Data Scientist who knows everything perfectly; Usually, these projects require multiple profiles.

# Index

- ❖ Data Science
- ❖ **Roles in Data Science**
- ❖ From academia/research to industry

# Roles Data Science Projects

Within a DS project we can find multiple roles.



**Data architect**

**Objective:** Data architects conceptualize and visualize data flow frameworks.

**Background:** Computer science, programming. It is usually a more experienced DE.

**Tools:** Scala, JAVA, MongoDB, Cloud Architectures.



**Machine Learning Engineer**

**Objective:** They take experimentation models and convert them into productive models to deploy.

**Background:** Computer Science with knowledge of ML

**Tools:** Python, Spark, Scala, Tensorflow,...



**Product owner/ Manager**

**Objective:** Define and manage business requirements.

**Background:** Business profile with high level analytical and development knowledge.

**Tools:** JIRA, Trello, Excels.



**Data engineer**

**Objective:** Data engineers develop and maintain data frameworks.

**Background:** Computer science, programming.

**Tools:** Scala, JAVA, MongoDB, Cloud Architectures.



**Data Scientist Research**

**Objective:** Experiment with the data to process and analyze the best models.

**Background:** Statistics, physics, mathematics, engineering.

**Tools:** Jupyter Notebooks, Python, R, Visualization Tools (Tableau,...)



**Data Analyst/Business Role**

**Objective:** Understand and convert business requirements into data.

**Background:** ADE, finance, actuarial,...

**Tools:** Excel, Oracle, Visualization Tools (Tableau, Microstrategy...)

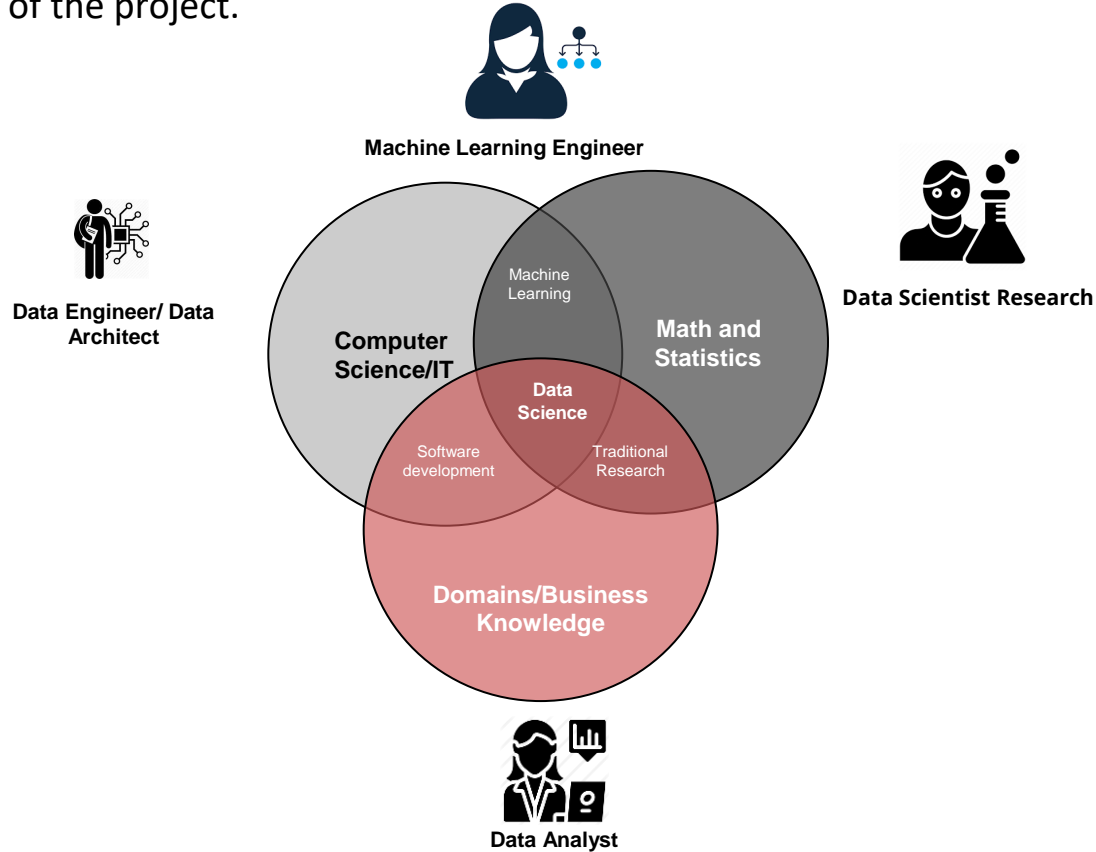
computer science

Statistics

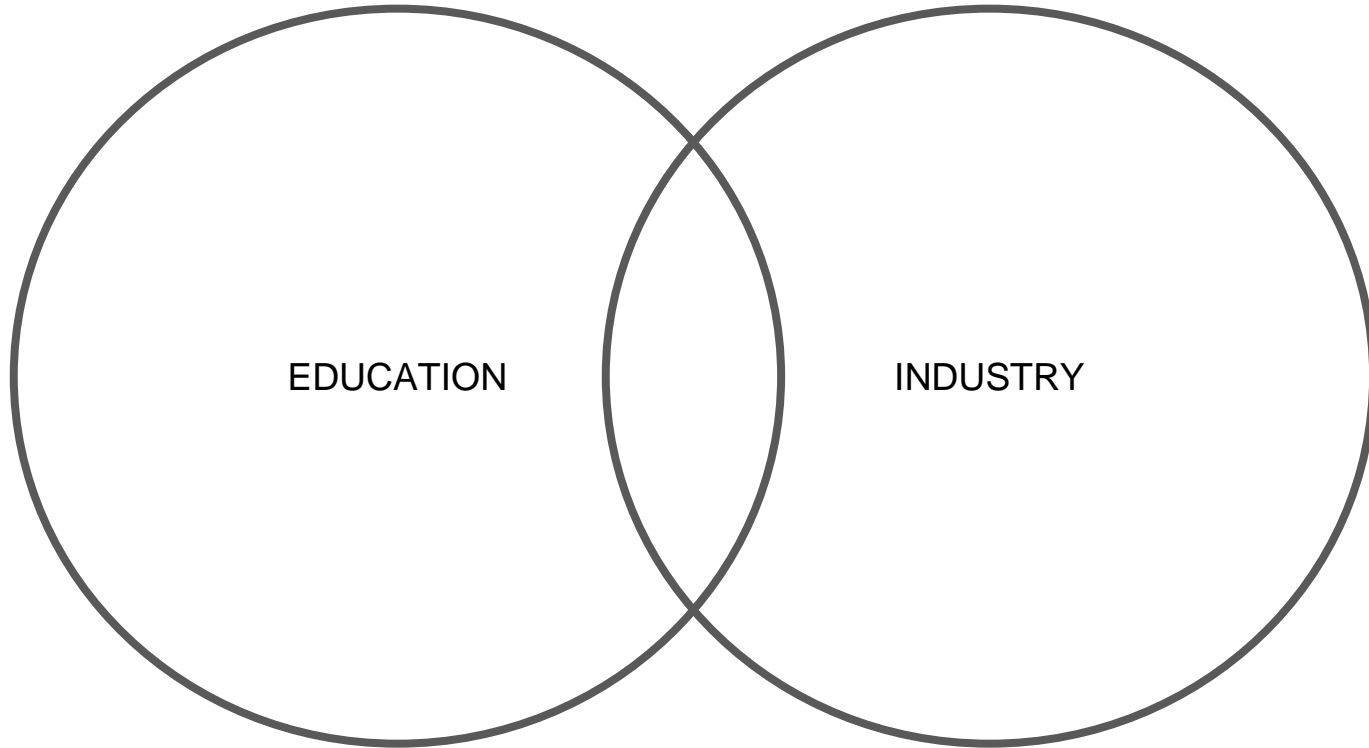
Business

# Roles Data Science Projects

Each of the roles may have a greater importance in the project depending on the scope and needs at each point of the project.



## My experience moving from university to industry



# My experience moving from university to industry as a physicist

Most profiles coming from physics, mathematics, or statistics are closer to the role of Data Scientist Research.

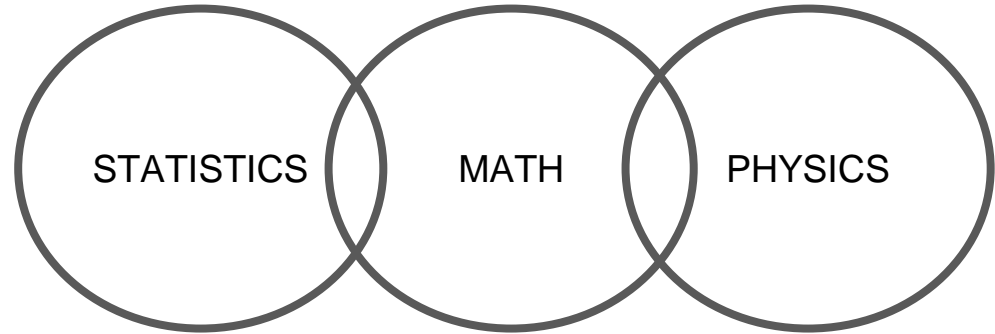


## Data Scientist Research

**Objective:** Experiment with the data to process and analyze the best models.

**Background:** Statistics, physics, mathematics, engineering.

**Tools:** Jupyter Notebooks, Python, R, Visualization Tools (Tableau,...)



- **Basic Programming Skills**
- **Strong Statistical Foundation**
- **Business Knowledge Gap**

# My experience moving from university to industry as a physicist/meteorologist

Most profiles coming from physics, mathematics, or statistics are closer to the role of Data Scientist Research.

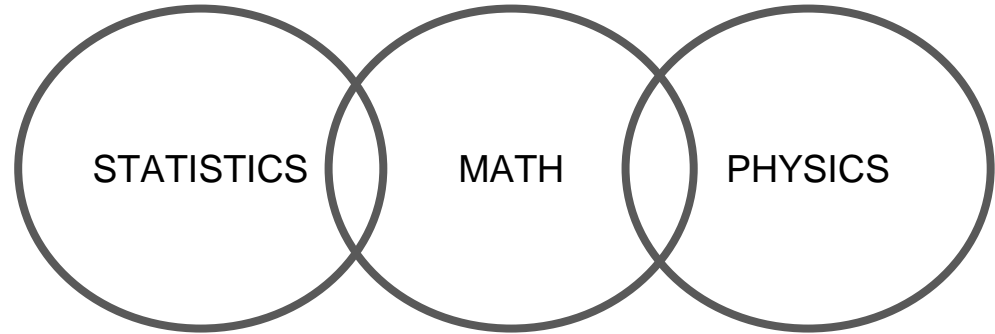


**Objective:** Experiment with the data to process and analyze the best models.

**Background:** Statistics, physics, mathematics, engineering.

## Data Scientist Research

**Tools:** Jupyter Notebooks, Python, R, Visualization Tools (Tableau,...)



In my case, I was lucky. I learned to program and understand computers better through the use of CESGA in developing weather prediction models. However, I still lacked many good coding practices and the use of collaborative tools.

A good foundation in time series modeling

- **Basic Programming Skills**
- **Strong Statistical Foundation**
- **Business Knowledge Gap**

# What skills did I find I needed to improve when I started working in the industry?

When I started working in the industry, I realized there were certain skills/goals that I needed to work on to improve my career path.

Model Metrics vs  
Performance Metrics



Clean Code &  
best practices



Soft Skills



focus on  
production



First, understand  
the business



Networking



Learn to adapt



## Example: Demand forecasting Model for a top retailer

Daily Demand Forecasting by SKU

+ 1000 shops

+ 1000 products per shop

**~1M models in production for  
daily demand forecasting**

- R didn't work
- ARIMA didn't work
- RMSE, MAE or other regression metrics are not the main metric that should be taken into account
- RMSE, MAE, or other regression metrics are not the main metrics that should be taken into account.
- First, consider the production requirements, and then try to find the best model.
- Huge team-> good and clean coding skills and collaborative tools

# Important initiatives that have helped me in my transition from being a student to my professional career



**Internship**



**Master's thesis  
enterprise  
collaboration**



**self-learning  
& continuous  
learning  
(Moocs)**

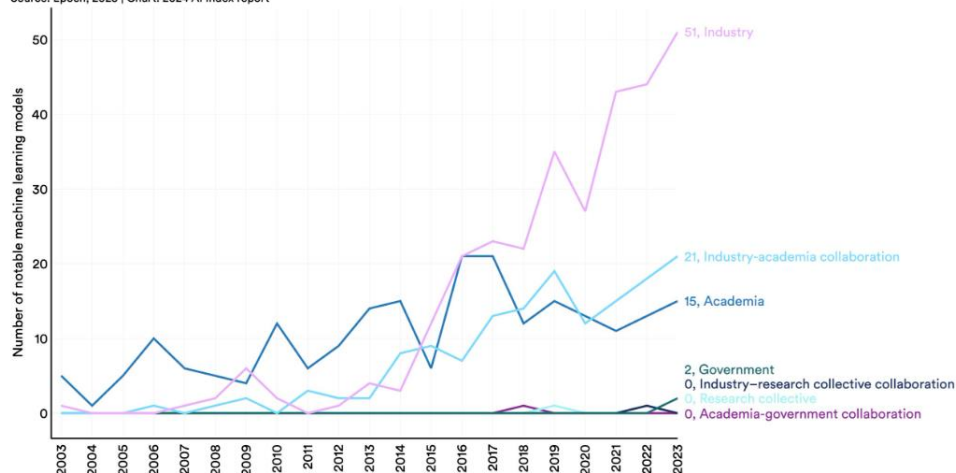


**Workshops/Even  
ts/networking/Ha  
ckathons**

# You can do research in Industry

Number of notable machine learning models by sector, 2003–23

Source: Epoch, 2023 | Chart: 2024 AI Index report



In 2023, industry produced 51 notable machine learning models, while academia contributed only 15. There were also 21 notable models resulting from industry-academia collaborations in 2023, a new high.

## AI Index Report 2024 By Stanford University



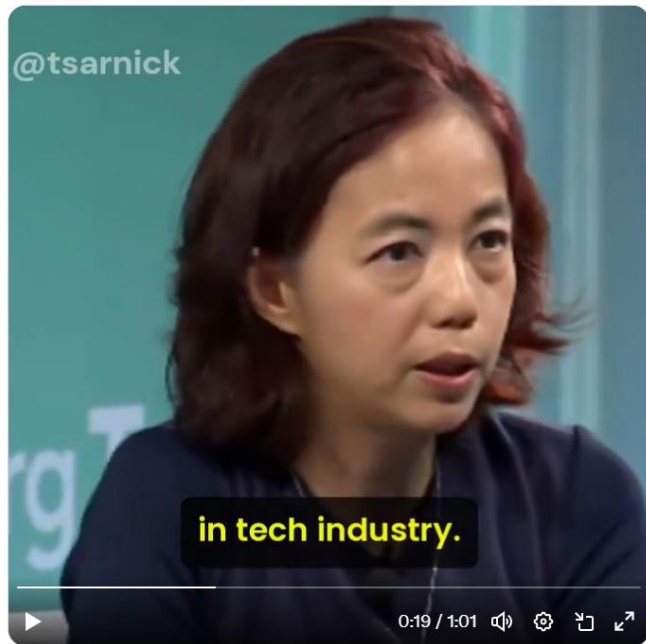
## Developing LLMs has a high cost that is affecting academic research in this field



Tsarathustra ✓

@tsarnick

Fei-Fei Li says Stanford's Natural Language computing lab has only 64 GPUs and academia is "falling off a cliff" relative to industry



""

*With President Biden one of the things we talked about is the moonshot mentality to invest in public sector AI ...*

*America's public sector academia is falling on the cliff very fast in terms of AI resources.""*

*May 2024, Fei Fei Li, Stanford professor*



**Thank you!**  
**Questions?**



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



[@pauvanbr](https://twitter.com/pauvanbr)



[paul@spain-ai.com](mailto:paul@spain-ai.com)