Leading Change
Inspiring Progress

Telefónica

Beyond the cloud: Edge Computing and the future of Distributed AI

Yolanda Bueno Morales

Infrastructure Marketing Manager,
Telefónica España

Telefónica

The development of hyper connectivity requires new data center infrastructures on which future use cases can be deployed

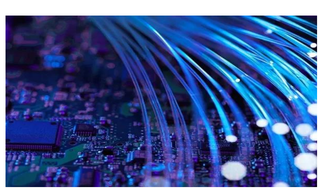
Telefónica's 5G and Fibre deployments are benchmarked at ...

5G

92%

- **Commercial rollout** of 5G coverage to populations
- High-performance 5G coverage

FIBRE



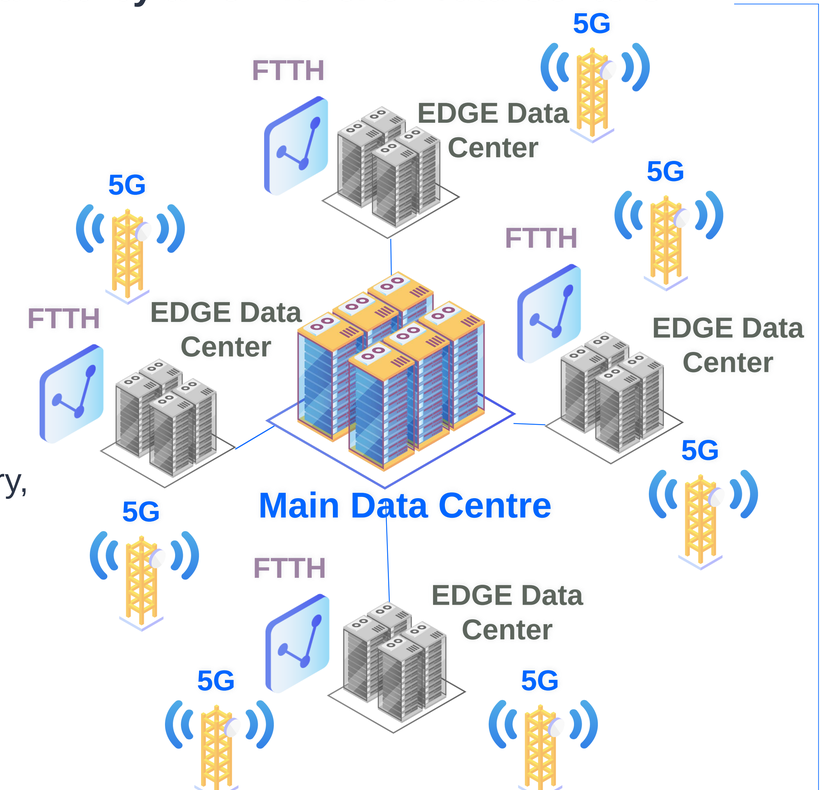
92% population coverage Spain *vs ~70% (EU 39)*

79% coverage rural households Spain *vs 64% Europe*

...to be accompanied by a new level of Data Centers

Large Data Centers for AI-based learning processes

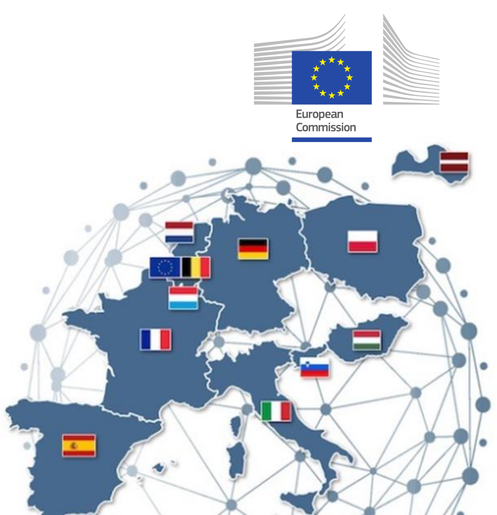
Edge Data Centers **smaller** and more capillary, where to run AI models close to the devices and act as technological innovation hubs



At Telefónica we are already working on preparing these infrastructures

IPCEI-CIS*

- Objective: To provide **interconnected Cloud Edge technologies and infrastructures** in Europe**.**
- **Nodes distributed** all over Spain
- **Interoperability between TELCOs** for EDGE services



Nodes Modernisation

- Reference architecture for **more efficient and sustainable new generation nodes**
- Enabling Edge Rooms to **bring Edge services closer to the customer** while minimising investment needs

Network Evolution and Cloud Capacity Deployment

- Best practices in **network integration** with **distributed computing capabilities** at the edge, offering **optimised latencies.**
- With guaranteed **sovereignty and residency** of data
- Facilitating the deployment of **use cases** for industry and services

Telefónica

** IPCEI-CIS: Important Project of Common European Interest for the Next Generation of Cloud Infrastructure and Services*

An Edge tailored to the Customer and their business

Integrating Telco network with Edge services

Distributed loads close to the user **to select the most suitable Edge for the user at all times.**

Low latencies

Apified

APIs defined in GSMA and integrated in OpenGateway

Sovereignty

Data residing in specific regions for business needs

An Edge tailored to the Customer and their business

Multiplatform and Cloud Continuum Orchestrator

A single entry point for the developer to deploy on nodes with different technologies. Integrated environments (Public Cloud, Private Cloud)

Federation

Between operators from different countries

AI and use cases as a service

Edge with AI as a Service and "ready for service" use cases to deploy on Edge nodes

Rafael Socas Gutiérrez

Cloud Architecture and Strategic Planning Manager,
Telefónica España

Telefónica

... with advanced functionalities

Complementing their high capacities with ....

.....Smart EDGE, Mobility/Roaming , Federation.....

.... in a **Multi-operator** environment

Smart EDGE

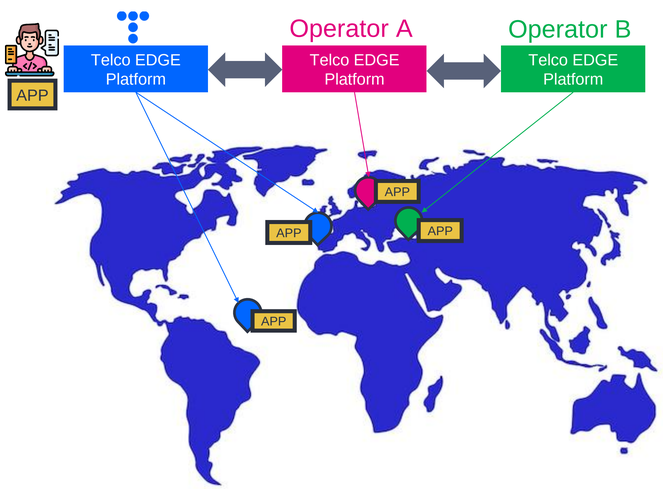
Both developers and the APPs themselves request the optimal EDGE node.

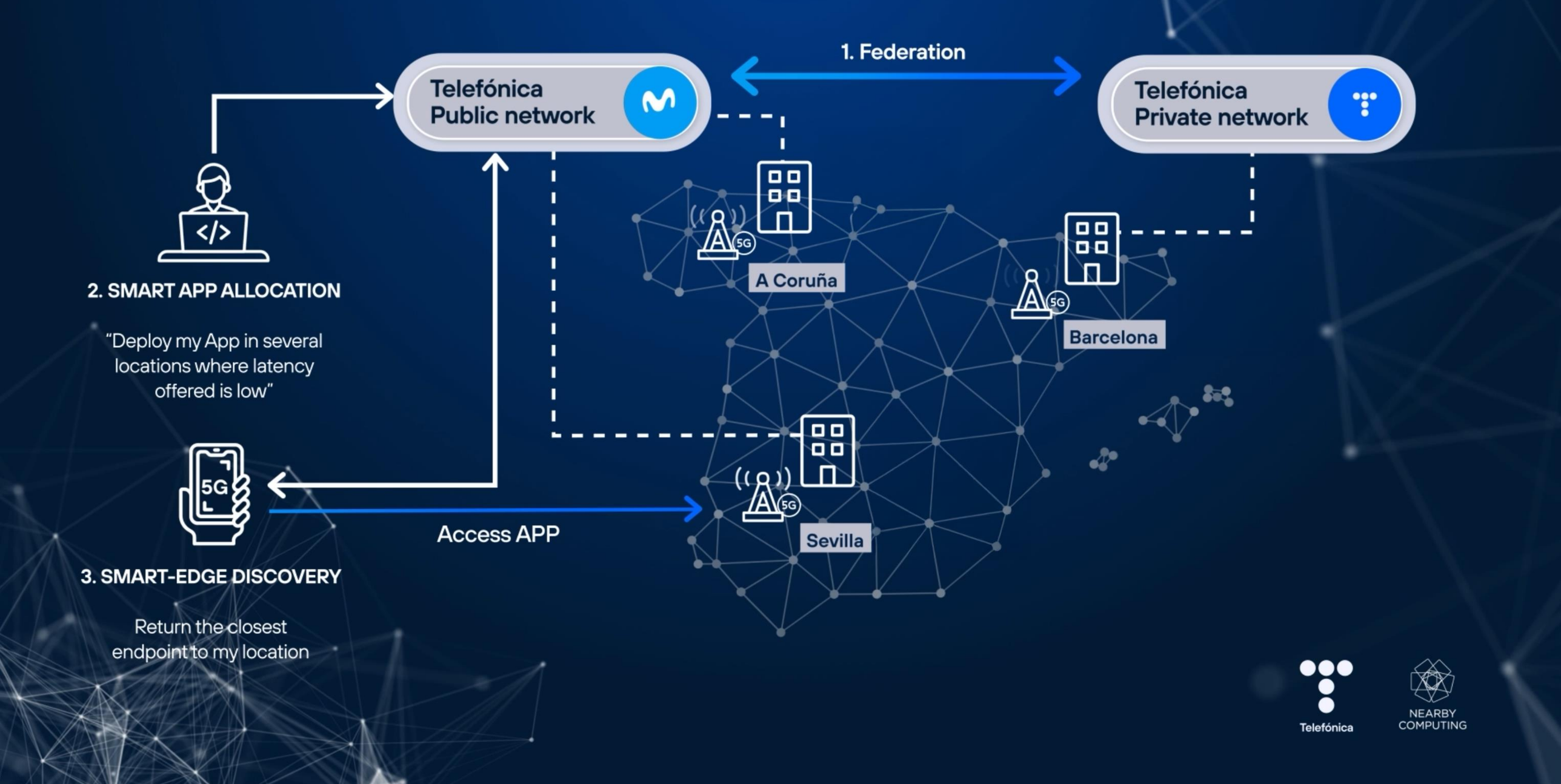
Latency, Localisation, Connectivity...

Resources, Network Capacity, Privacy...

Mobility/Roaming

Federation



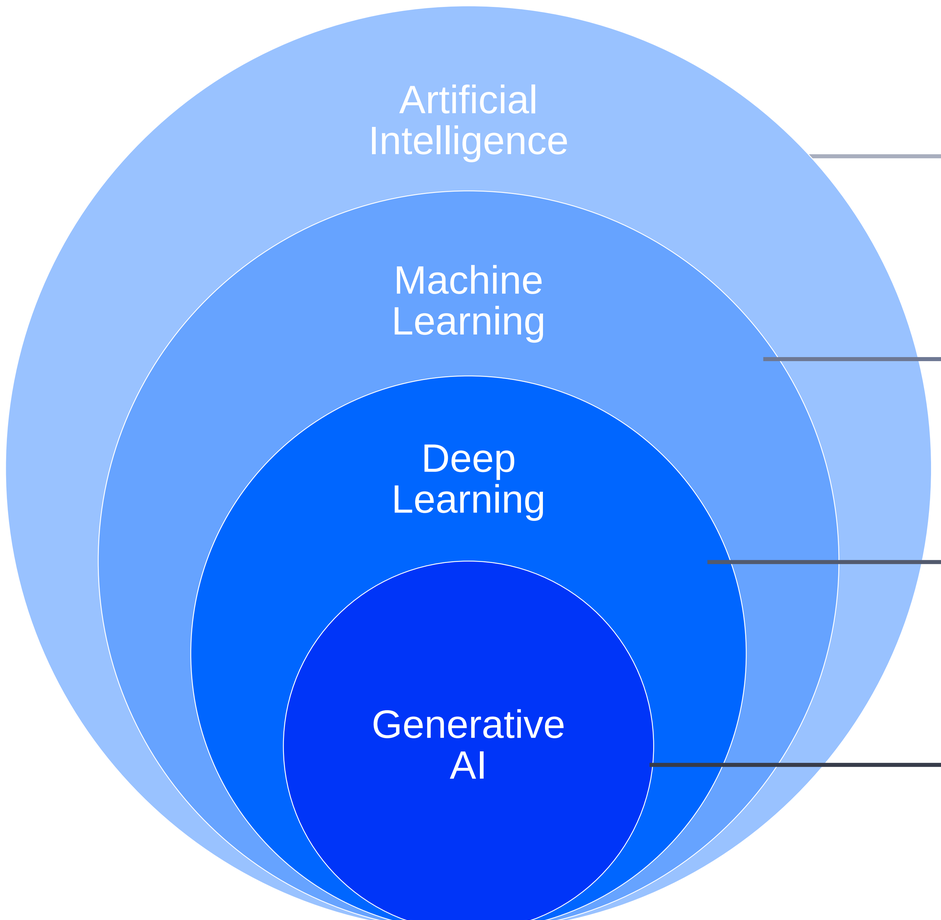

1. Federation
Telefónica
Public network
Telefónica
Private network
2. SMART APP ALLOCATION
"Deploy my App in several locations where latency offered is low"
A Coruña
Barcelona
Sevilla
5G
Access APP
3. SMART-EDGE DISCOVERY
Return the closest endpoint to my location
Telefónica
NEARBY COMPUTING

Daniel Ribaya González

Cloud Product & Service Management Director,
Telefónica Tech

Telefónica

What types of AI are there?



Field of Study. Programs capable of imitating human behavior

Branch of AI that focuses on the creation of machines/programs intelligent enough to learn from data

Subset of Machine Learning methods, based on Artificial Neural Networks that adapt and learn from large amounts of data

A type of artificial neural networks, capable of generating new data similar to the data with which they were trained

How has AI evolved over time?

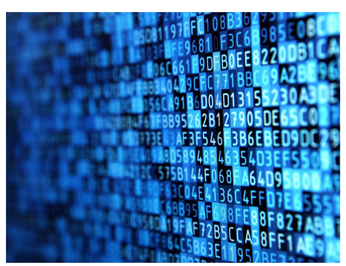
Compu… AI to simu… a human o… that has ex… particula…

Progra… neuron… identify ph… options a… conclus…

Branch… the cre… mac… learn fr…

Subse… methods, Neural Net… and learn… amounts …

A type of artificial neural networks, capable of generating new data similar to the data with which they were trained



Early Algorithms

Expert Systems

Neural Networks

Machine Learning

Deep Learning

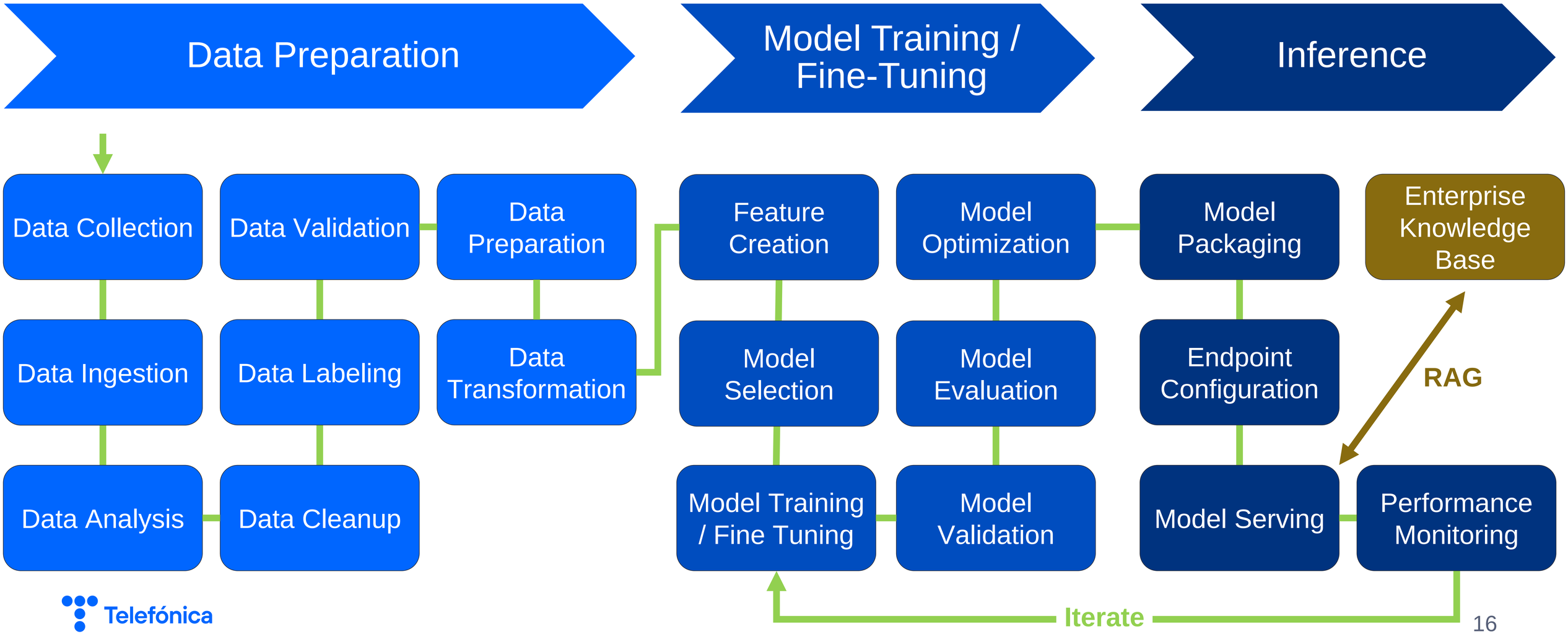
Gen AI

Late 60s – 70s

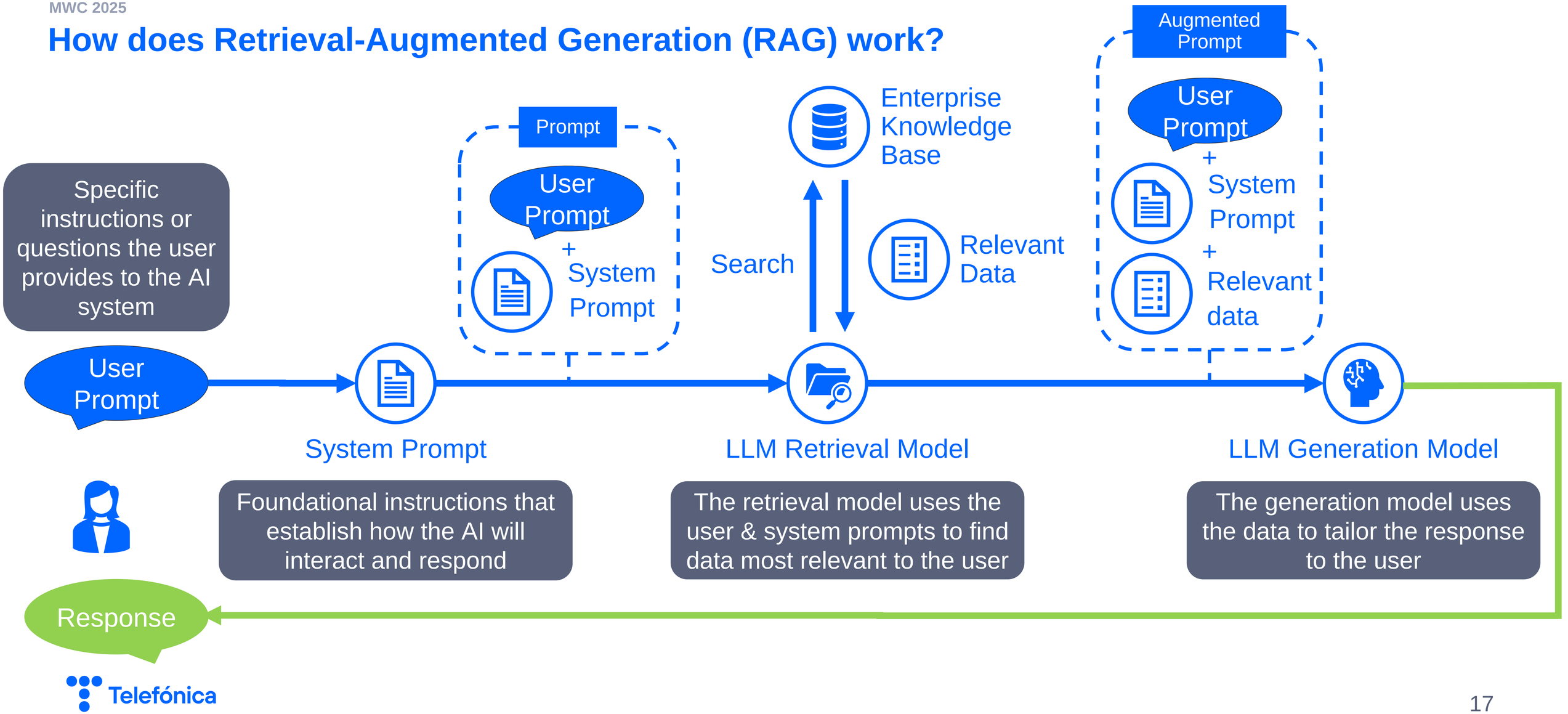
Late 80s – 90s

Modern AI

What are the Steps in building & maintaining a Gen AI agent?



How does Retrieval-Augmented Generation (RAG) work?



Where will my Gen AI workloads run?

Data Center

- Dataset preparation
- Model Training
- Model Fine-Tuning

Edge

- Basic inference
- RAG

Mobile Devices

- AI Client Applications

What type of infrastructure does each stage of the AI Lifecycle require?

	Training & Fine-Tuning	Advanced Inference	Basic Inference
Use Case	Training large AI models, Fine-Tuning existing ones	LLMs Gen AI inference & Deep Learning	Computer vision inference & Machine learning
Location	Centralized	Centralized / Multi-Access Edge	Multi-Access Edge
Node Type	AI Factory (bare metal) / Cloud AZ	Cloud AZ / Advanced Edge Location (IaaS)	Standard Edge Location (IaaS)
Required Investment	Several x $10M per site	Several x $Ms per site	Several x $100Ks per site
Multitenancy	Physical partitioning and time slicing of the node	IaaS: vGPUs PaaS: Tokens	IaaS: vGPUs PaaS: Tokens
Typical Software Stack	AI Foundry platforms e.g. NVIDIA NeMo, …	Libraries and microservices e.g. NVIDIA NIM, HUGS,...	Libraries and microservices e.g. NVIDIA NIM, HUGS,...

Telefónica

AI Models development → ← AI Models execution →

MWC 2025

What is the added-value of Telefónica in infrastructure for Sovereign AI?

Sovereignty

- Privacy, security, localization and jurisdictional safeguards for confidential business data
- Platforms and Data Centers with national operation, optimal for workload repatriation strategies
- Option of having isolated environments

Edge infrastructure

- Next generation networks (fibre, 5G) provide high bandwidth and lower latency and jitter, with intelligence provided by SmartEdge
- The Telco exchanges, provide space for computing & storage and they are at the first network hop
- LLM inference is an optimal use case for Multi-access Edge Computing because it requires very low latencies, high bandwidths and localization safeguards

Hybrid Cloud Experience

- Experience in setting-up & managing Infrastructure platforms (Compute + Comms + Security) adapted to customer needs
- Variety of choice between commercial and open-source AI solutions to avoid vendor lock-in.
- Experience in Storage infrastructure + Capabilities on Real-time private data inference: Retrieval Augmented Generation

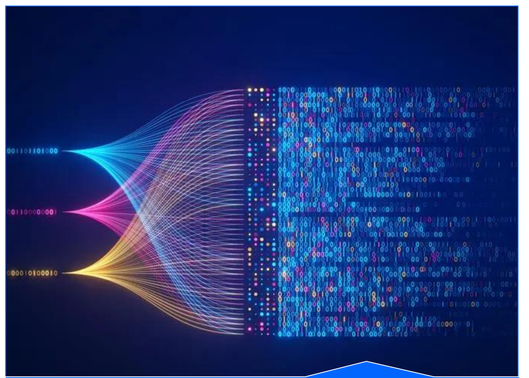
What are Edge AI most common Applied Use Cases?

Computer vision

- Retail (seamless checkout, image-based recommendation engines, theft prevention)
- Security (face and object recognition, threat detection)
- Manufacturing (quality control, optimization of manual processes)
- Diagnosis (pattern recognition in medical images)

On-site RAG-powered conversational agents

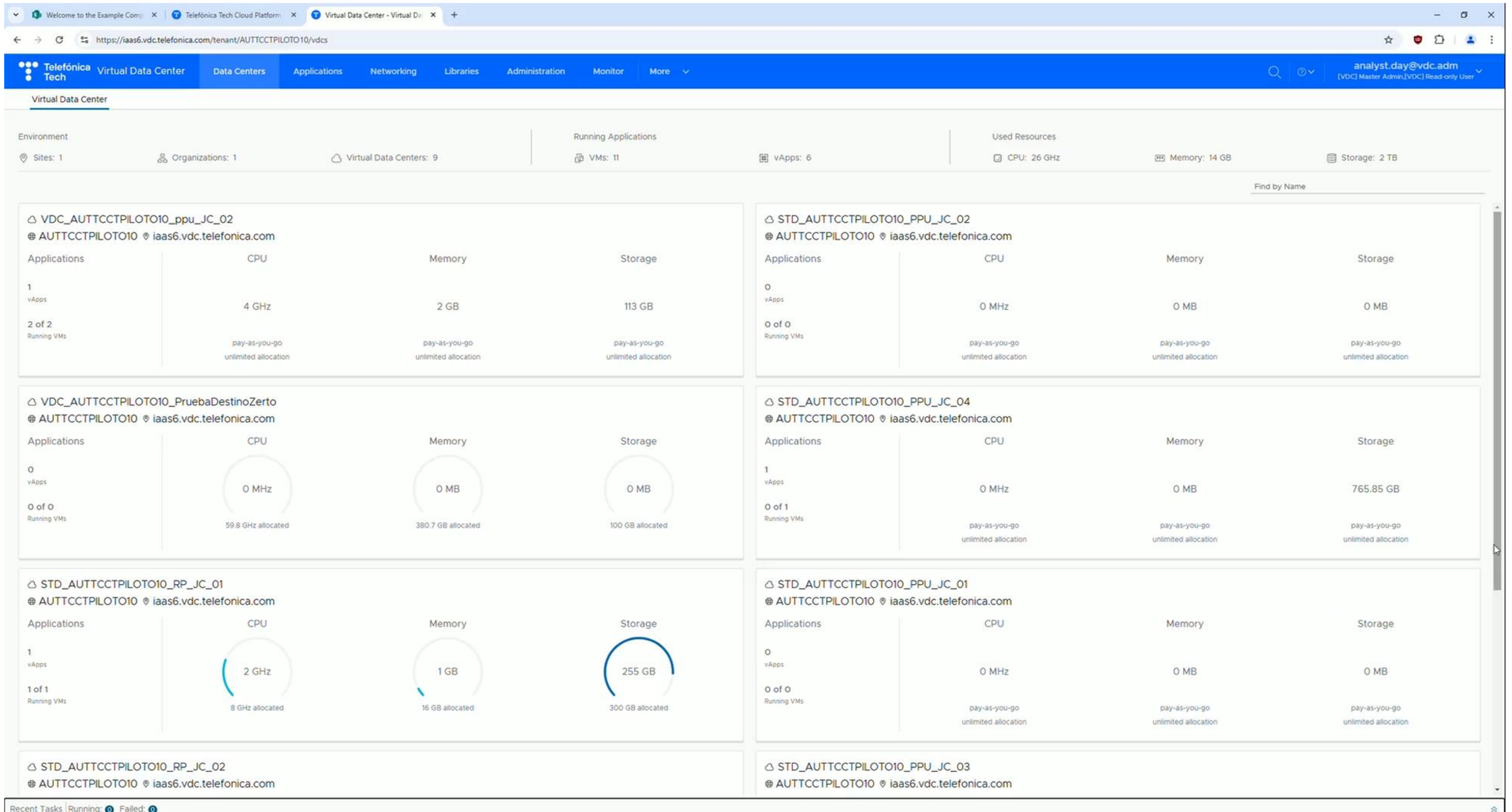
- Presales (requirements gathering, product recommendation)
- Sales (order taking or product/service contracting)
- Customer care (troubleshooting, incident reporting)



Big data & ML/DL models applied to industry

- Robotics (industry, agriculture, mining)
- Autonomous driving and drone control
- Logistics optimization (fleet, route and delivery/pick-up schedule optimization)
- Predictive maintenance

How does a Private Gen AI Agent generator work?




Telefónica