Logotipo

Descripción generada automáticamente

LOGO CLIENTE

(Tamaño máximo 4,68x1,23cm)

Strategic Vision  
  
AI Generic

2026 - 2029

Global RFI  
  
10 de April de 2025

Index

[1. Introduction 3](#_Toc191286566)

[1.1 The RFI 3](#_Toc191286567)

[1.2 Vendor answer 3](#_Toc191286568)

[2. Generic architecture 4](#_Toc191286569)

[2.1 Architectural schema 4](#_Toc191286570)

[2.2 Requirements 4](#_Toc191286571)

[3. Specific AI evolution 6](#_Toc191286572)

[3.1 AI Technology 6](#_Toc191286573)

[3.2 AI use cases 6](#_Toc191286574)

# Introduction

## The RFI

This document will try to collect all the supplier information regarding AI generic support and roadmap plans. Generically, the RFI is composed by several sections:

* Generic architecture
* Specific AI evolution

## Vendor answer

1. Please state your full roadmap timelines for the various capabilities, assuming there will always be phasing, including any limitations in early phases:

SUPPLIER: Radisys solutions currently include, and support AI powered capabilities in various domains and applications, including network operations and real-time conversational intelligence, as well as AI technologies for computer vision. The AI capabilities leverage public providers of AI LLM (such as OpenAI), but also an on-premise solution is also available where the data content and local LLMs are deployed for security and privacy reasons. The ability to create custom datasets for LLM with generative AI capabilities are available in various industry vertical applications, including live virtual assistants and AI powered technical network support. Capabilities leveraging RAG and various Agentic AI capabilities provide enable a wide range of applications and solution to be designed, customized and deployed for diverse range of industries as well as for personal and enterprise solutions.

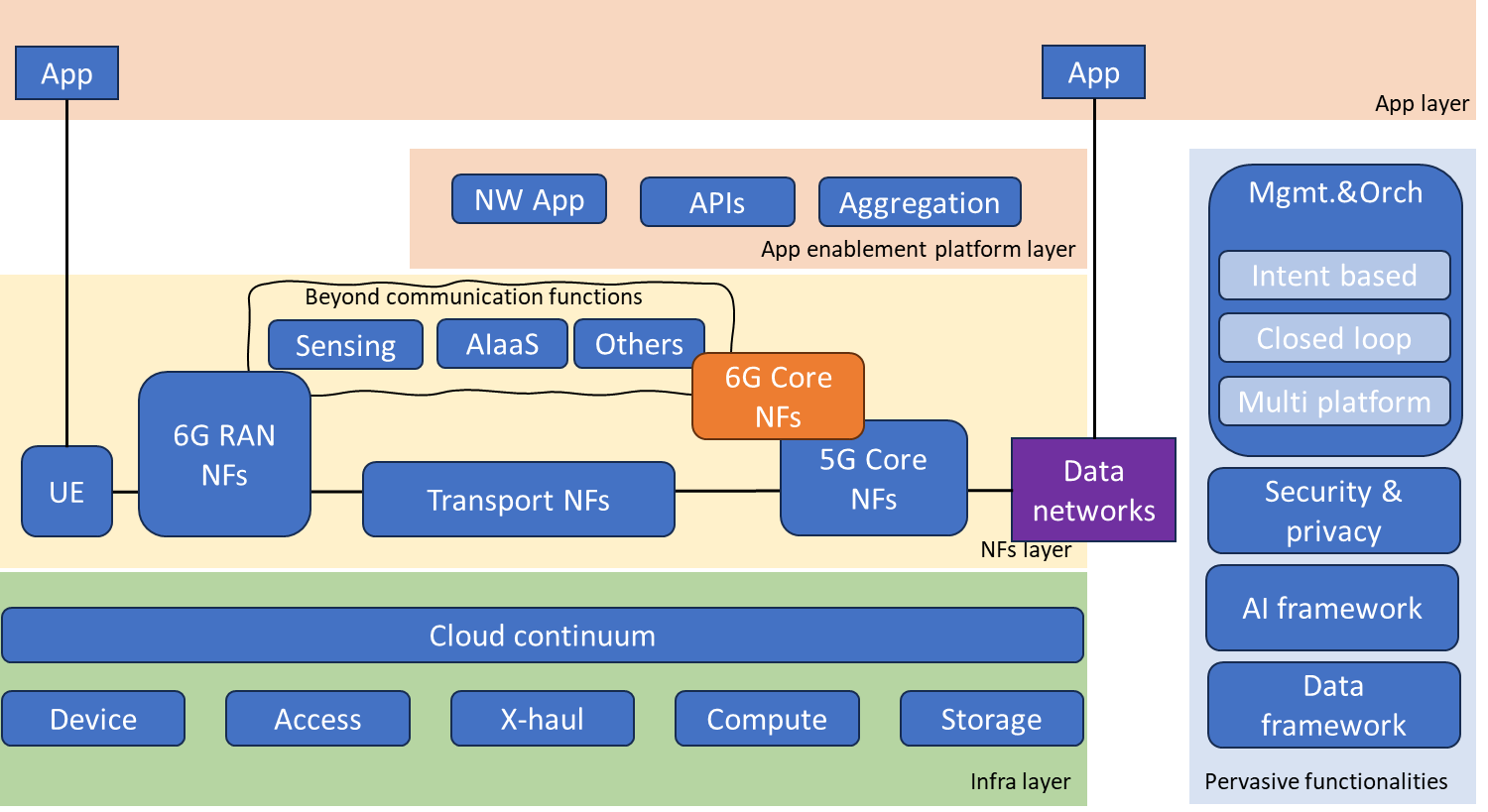
1. Please feel free to include here any other aspect that you may consider as relevant to share with Telefónica and has not been requested in any concrete requirement:

SUPPLIER: Radisys solutions provide a unique set of AI powered capabilities and applications which are deployed in-network, i.e.: as part of the operators voice and data networks, including live mobile telephony calls and the open API approach facilitates ease of integration with external services such as CRM databases or custom interaction with enterprise private servers, eg: Salesforce or other enterprise services. Agentic AI capabilities of the Radisys EDP platform provide a powerful and unique set of applications for operator’s voice subscribers and enterprise B2B and B2C customers.

# Generic architecture

## Architectural schema

This is the generic architectural diagram that will be referenced in each of the requirements included in this RFI.



## Requirements

1. Please describe how you foresee the 5G architecture evolution with the introduction of AI:

SUPPLIER: AI technology is being deployed to enhance and optimize 5G networks (evolving to 6G) for network optimization and automation for efficient network resource utilization as well as network operational management with predictive analytics. In additional to network operational management, the real-time media (audio, video, messaging) traversing these networks opens a wide range of AI powered applications, such as digital virtual assistants, fraud detection, AI powered call screening, and generative AI powered enterprise and consumer services. The wide range of these AI powered digital applications are applicable to numerous industry verticals, including healthcare, entertainment, e-commerce, SMB markets, IoT, industrial factory and warehouse automation as just a few examples.

1. Please describe under your view what are the main consumers of AI in 6G networks across the different layers (infra, NFs, App enablement, Pervasive functionalities…) and domains:

SUPPLIER: AI technology leverage into the following main categories:

* Network optimization and operations (eg: predictive maintenance, fault resolution, security or threat detection)
* Network resource management and predictive analytics (eg: slicing, on-demand resource allocation, RAN and network resource allocations)
* Digital applications utilizing voice, video, and messaging with AI powered capabilities to provide a unique set of services to consumers and enterprise, such as AI powered digital assistants, auto-generated call transcripts, live language translation, AI powered in-call fraud detection as a few examples.

1. Please describe the different operational use cases based on AI introduction that could provide benefits to network platforms:

SUPPLIER: AI powered solutions can analyse network performance metrics in real-time to predict and prevent service degradation before customers is impacted. Virtual AI assistants and chatbots provide automated troubleshooting, network diagnostics, and personalized recommendations for customers. As well, AI-driven analytics can help operators optimize service quality by proactively resolving issues affecting user experience.

1. Please describe the monetization use cases exploiting IA that should be considered by Telecom operators:

SUPPLIER: Numerous AI powered digital applications can be offered by telecom operators to the customers as well as enterprise market segments.

Additionally, AI powered solutions can analyse customer behaviour, usage patterns, and preferences to offer personalized data plans, content bundles, and upgrades.

AI-powered chatbots and virtual assistants proactively recommend premium services (e.g., 5G upgrades, unlimited data, cloud storage, security add-ons).

Dynamic pricing models ensure optimal ARPU (Average Revenue Per User) growth by offering the right plans at the right time.

AI powered personal digital assistants can be offered to consumers and enterprise which are leverage AI LLM, RAG, and Agentic capabilities to serve a broad range of applications and solutions leveraging generative AI capabilities.

Additional AI powered applications listed below:

Speech Recognition and Analytics:

Radisys Engage Digital Platform offers programmable speech recognition and analytics capabilities that allow for the development of applications that can interpret and analyze spoken language.

Video Analytics:

The platform includes programmable video analytics, enabling advanced video processing and analysis for various applications.

AI-Powered Customer Care Services:

Engage Digital Platform enables the creation of intelligent AI-powered customer care services, improving customer interactions and support efficiency.

Natural Language-Based Virtual Agents:

The platform supports eCommerce and customer care applications using AI-driven natural language processing to create virtual agents that can interact with users in a conversational manner.

Real-Time IoT Services:

With support for 5G networks, the platform powers real-time IoT services that leverage AI capabilities such as computer vision and speech analytics.

Biometric Authentication:

Radisys Engage Digital Platform includes advanced media features like biometric authentication, utilizing AI to enhance security measures.

1. Please describe the architectural components and associated products that should be taken into account in order to sustain monetization use cases exploiting IA capabilities:

SUPPLIER: Radisys solution encompasses Programmable Communications stack (APIs and SDKs with reference applications) and AI capabilities covering speech analytics, video analytics (computer vision), LLMs, generative AI, RAG framework, and Agentic AI application builders. These capabilities can be deployed on public or private cloud (on-prem datacenter) infrastructure.

1. Please describe the technology and protocols involved on UE data collection for AI/ML:

SUPPLIER: All media (audio, video, message) content from UE is accessible by the Radisys EDP platform over 3GPP/IETF compliance standards (e.g.: RTP, RTCP, SRTP) for consumption and use by AI/ML model, e.g.: real time audio can be transcribed to text and used by AI model to determine user intent based on NLU (natural language understanding). UE device status information (including location status, SIM status, etc) can be obtained via a Programmable Networks APIs (as defined by the Camara project).

# Specific AI evolution

## AI Technology

1. Please describe how AI is embedded in your products:

SUPPLIER: For on-prem solutions AI LLMs, RAG framework, and generative AI capabilities can be deployed in isolated environments (for privacy and security), as well third-party external solutions can be accessed via well define REST APIs. AI model training and updates can be applied to the platform on an as needed basis without affecting the run-time services.

1. Please describe the LLM technology used to leverage your products

SUPPLIER: Third party (external) LLMs can be utilized by Radisys EDP applications (eg: OpenAI), as well as integrated LLAMA 3.x is provided.

1. Please describe the SLM technology used to leverage your product

SUPPLIER: Currently LLAMA 3 is supported, however smaller LLM can also be provided, specifically LLAMA 2 (7B), LLAMA 2 (3B), and TinyLLAMA, however, these smaller models are targeted for embedded devices and UEs versus network-based deployments.

1. Please describe how you foresee AI and NF decoupling from an architectural perspective

SUPPLIER: Radisys approach has been to architecturally decouple “Programmable Communications with APIs/SDK similar to CPaaS” from AI Analytics capabilities (audio, video, and messaging) such that these capabilities can be enhanced independently and multiple combinations of solutions can be designed more efficiently

1. In case of AI and NF decoupling, please describe the interface and protocols used in the communication between both entities.

SUPPLIER: For leveraging AI capabilities for real-time media, typically web-socket or gRPC interfaces are used (direct RTP based media interfaces are also available and supported), along with HTTP REST for control interfaces to the AI subsystems.

1. Please describe infrastructure requirements needed for the introduction of AI against the different use cases identified

SUPPLIER: Radisys EDP components are fully containerized with a set of micro-services which can be invoked, managed, and scaled independently. The Kubernetes based CNF infrastructure allows maximum flexibility and optimizations on general purpose Intel or AMD based server infrastructure. Additionally, for higher densities and performance, GPU based acceleration is also supported via NVIDIA. As such, individual applications and use-cases can be optimized and scaled independently.

## AI use cases

1. Please describe from a timeline perspective what are the key evolution steps to be considered by an operator from an AI perspective

SUPPLIER: Realtime in-network conversational applications with AI technologies (eg: live transcriptions, live language translation, live AI virtual assistants, AI powered real-time fraud detection, and generative AI based Agentic Agents are currently in process of lab trials and early operator deployments for consumer and enterprise use cases. We expect these applications leveraging AI technologies to be pervasive in all operator networks and application use-cases for over the next 1-3 yrs, while additional capabilities and performance optimized solutions will continue to grow as additional advancements in AI technology become available. AI for computer vision, generative AI capabilities, and network operations with predictive analytics are expected areas of additional advancements over the next 1-3 years.

1. Please describe operational processes that could be proposed or improved by introducing GenAI

SUPPLIER: GenAI technology is currently available in Radisys solutions for a number of applications and use cases, covering network ops, technical assistance, consumer based AI assistants, enterprise use-cases with GenAI capabilities for sales and customer support, as well as contact-center use-cases. GenAI capabilities are supported alongside custom datasets which can be easily uploaded (eg: documents, files, PPTs, XL sheets, websites, data logs, product literature, and external databases such as CRM repositories). Additionally, live voice call transcripts can also be easily uploaded into the application or user’s custom dataset available for GenAI based applications.

1. Please describe how GenAI will be introduced and considered in your products

SUPPLIER: Customized dataset management along with localized LLMs (along with RAG framework and Agentic AI capabilities) are currently supported for a variety of real-time applications as described earlier.

1. Please describe the use cases you foresee in 6G networks around pure analytics, prediction and classification capabilities

SUPPLIER: There are a number of AI driven advancements which can be leveraged in 6G networks, namely a few itemized below:

AI-Powered Network Intelligence & Self-Learning Networks

Predictive and Proactive Fault Prediction & Self-Healing Networks

AI-Powered Cybersecurity & Anomaly Detection AI-Driven Smart Cities & Mobility Analytics Real-Time User Behaviour Analytics & Service Personalization

AI-Optimized Energy & Sustainability Management

1. Please describe the use cases you foresee in 6G networks around service automation powered by AI

SUPPLIER: A few use-cases itemized below to be targeted for AI powered 6G networks.

Autonomous Network Management & Zero-Touch Provisioning

AI-Driven Dynamic Network Slicing & Monetization

AI-Powered Autonomous Service Delivery & Edge Automation

AI-Driven Customer Service Automation & Digital Assistants

AI-Powered Predictive Maintenance & Self-Healing Networks

AI-Enabled Autonomous Security & Threat Mitigation

AI-Powered Smart Supply Chain & Logistics Automation

1. Please describe the use cases you foresee in 6G networks around closed loop networks powered by AI

SUPPLIER: 6G networks to be fully autonomous, self-optimizing, and self-healing through AI-driven closed-loop automation. Leveraging AI technology, networks can be continuously monitored, analyzed, and adapted in real-time with minimal human intervention. AI can power closed-loop decision-making across network management, service assurance, security, and optimization. A few areas where these can be beneficial are for self-organizing and self-optimizing and self-healing networks as well as AI driven resource optimizations by managing network slices in closed loops, including AI driven dynamic pricing of network resources.

1. Please describe the use cases you foresee in 6G networks around intent-based management (zero-touch) networks powered by AI

SUPPLIER: A few use-case highlights can be expected in 6G networks around intent-based AI powered management networks:

AI-Powered Service Assurance & QoE Optimization: Use Case: Real-time intent-based user experience management

Autonomous Network Configuration & Self-Optimization: Use Case: Zero-touch deployment & AI-driven optimization

Intent-Based AI-Driven Security & Cyber Threat Mitigation: Use Case: Zero-touch AI-driven security enforcement

AI-Powered Dynamic Network Slicing with Zero-Touch Control: Use Case: Real-time, intent-driven slice creation & management

Self-Healing Networks & Automated Fault Management: Use Case: Intent-based network self-recovery

1. Please describe what kind of tools you foresee will be necessary to manage the lifecycle of AI execution environments everywhere in the 6G network (e.g. UE, RAN, transport, core…)

SUPPLIER: AI-powered 6G networks are expected to be deploy embedded AI capabilities, from User Equipment (UE) to Radio Access Network (RAN), transport, and core network. Managing AI across this distributed landscape will require specialized lifecycle management tools to ensure efficiency, security, and adaptability.

A few lifecycle management capabilities powered with AI execution are expected to be deployed in 6G networks:

AI Model Orchestration Platforms. Deploy, update, and scale AI models across UE, RAN, Transport, and Core

AI Workload Scheduling & Resource Optimization Tools. Dynamically allocate compute, memory, and power resources for AI execution

AI powered security and trust management

AI powered Lifecycle Monitoring & Observability Tools

1. Data quality management will be key for a success AIaaS implementation. Please describe the ingestion and refinement models you foresee in multivendor 6G networks.

SUPPLIER: Automated AI powered tools will be essential to continuously monitor, enhance, and refine models using a set of AI tools as this activity will be unmanageable via human intervention. Interoperability and consistency will be key in multivendor 6G networks and dynamic adaption shall require AI powered capabilities which can be automated. Synthetic data sources may be utilized in data quality management to create AI generated scenarios for accuracy and consistency verification, in an automated and continuous manner.

1. Please describe what are the main enhancements needed to support AI/ML services exposure.

SUPPLIER: Service exposure shall be based on an easy to consume set of APIs (REST and gRPC are commonly used) and SDKs for model consumption, but also the ability to manage and automate model updates. Intent based service exposure shall be provided to control and take certain actions based on the analytics generated by the applied AI models during inferencing.

1. Please describe how AI/ML will help to improve energy efficiency.

SUPPLIER: RAN resources can benefit drastically by AI models for radio resources as well as energy consumption. Similarly, edge and core network functions can be optimized for energy consumption by intelligently managing, based on AI predictive analytics, the compute resources as needed and colling and thermal management. Energy efficiency shall be a key characteristic across the entire network for AI model to learn from run time based on specific traffic load and apply predictive analytics to reduce energy consumption.

For example: Traffic-aware network scaling, where AI models dynamically adjust bandwidth and processing resources based on real-time demand. AI-powered predictive maintenance can prevent equipment failures and inefficiencies that increase energy consumption, and dynamic spectrum sharing allows AI models which reallocates unused spectrum dynamically to avoid energy waste.

1. Please describe what kind of services of interest for a telco operator can be implemented by using a GenAI as a service framework (audio bots, personal assistants…).

SUPPLIER: GenAI based virtual assistants are currently supported for any real-time telephony services. These AI based assistants utilized localized and customizable datasets, in private repositories, to provide immersive user experience in real-time audio and video responses to user, rather than chat-bot only interactions. AI driven personal virtual assistants can be extremely beneficial for consumer, enterprise business, as well as contact-centre use-cases. As well, Radisys EDP platform provide a wide range of languages (100+), with NLU (natural language understanding), prompt engineering and tuning responses based on custom datasets. Audio and Video based AI generated virtual assistants can leverage AI technologies to create dynamic content in speech and video matching the customer use cases, including live audio and text-based language translation services.

1. Please describe how such GenAI as a service framework should be implemented, detailing the functional blocks needed in terms of data acquisition and processing, multimodal models, interface towards application level, model training, infra layer... and the tools for the operator and their partners to self-develop GenAI based applications like bots (i.e. a GenAI studio).

SUPPLIER: As many GenAI service agents and applications can serve a large diverse application market segments and use-cases, the solution shall provide a easy to use (in many cases no-code or low-code) capabilities for personalization and customization capabilities. The capabilities span a wide range of multi-model access, including real-time audio, video, messaging and real-time connectors to external servers for live data fetch and private data access, e.g.: with CRM and customer databases. The Radisys EDP platform provides all these capabilities with Agentic AI capabilities for rapid application design, verification, and deployment capabilities via set of easy-to-use REST APIs, SDKs, reference applications and creation of application flow and design logic. Model creation and training are provided via design portals to abstract the underlying complexities, such that non-experts are able to design and deploy GenAI based Agentic applications directly by the operator or can be exposed to third-party developers or enterprises.

|  |
| --- |
|  |
| www.telefonica.com |