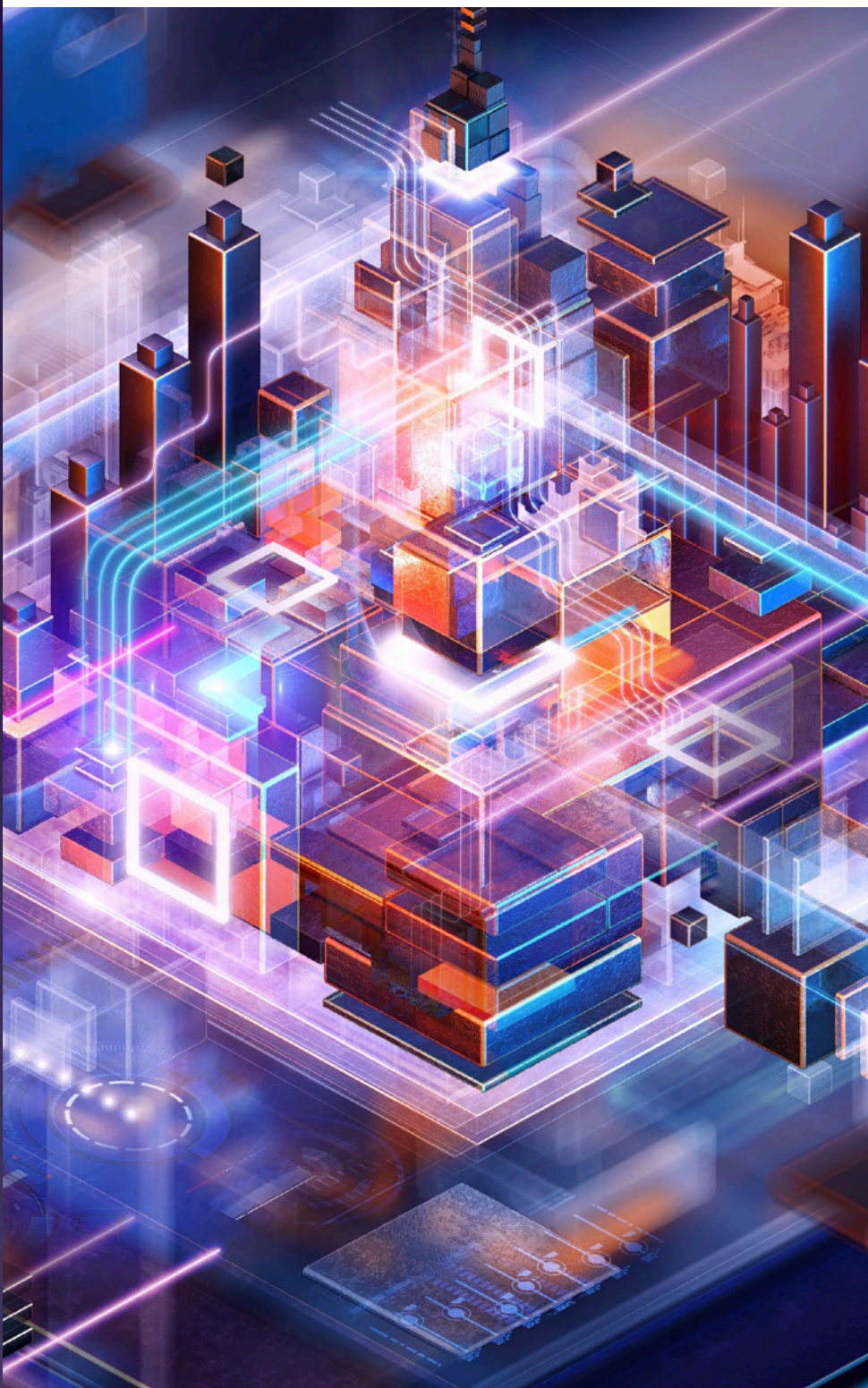


INNOVATION¹⁰

december 2025



INNOVATION 10

december 2025

Editors-in-Chief

João Paulo Firmeza
Ana Margarida Almeida
Daniela Rei
Pedro Carvalho

Pedro Miguel Neves
Raquel Pinheiro
Ricardo Cadime
Rui Calé
Sara Furão
Sérgio Ramalho
Victor Marques

Editorial Board

Ana Margarida Almeida
Cláudio Rodrigues
Daniela Rei
Fausto de Carvalho
Filipe Cabral Pinto
Francisco Fontes
Luís Cortesão
Nuno Silva
Pedro Carvalho
Pedro Neves
Vítor Marques

Title

Innovaction
Altice Labs Journal

Art Coordinator and Graphic Edition

Cátia Pinto
Carlos Figueiredo

Edition

2025

Manuscript editing and proofreading

Miguel Neto

Design

Altice Labs

Legal deposit

512167/23

Reviewers

Alexandre Laranjeira
André Gomes
Bruno Vítor Cabral
Carlos Guilherme Araújo
Catarina Dionísio
Daniel Duarte
Daniela Rei
Filipe Cabral Pinto
Francisco Fontes
João Bastos
Jorge Gonçalves
José Salgado
Luís Azevedo
Luís Castro
Luís Cortesão
Luís Reis
Miguel Freitas
Nuno Silva
Paulo Jesus
Pedro Carvalho

ISSN | DOI

2975-8459 | 10.57650/
INNOVATION





Opening note

Welcome to a new edition of *InnovAction*. It is with great enthusiasm that I introduce this year's magazine at a moment when Altice Labs is deeply engaged in redefining the future of connectivity, intelligence, and digital trust.

We enter this new edition at a time of profound transformation. The global technological landscape is evolving at unprecedented speed, shaped by the convergence of Artificial Intelligence, the virtualization of processes, the decentralization of business models, and the increasing demand for sustainable solutions. This context challenges organizations to rethink their strategies, adapt with agility, and remain firmly committed to creating value through innovation. AI, automation, next-generation networks, edge computing, and data sovereignty are reshaping the way we design, operate, and secure digital infrastructures. This context challenges us, but also reinforces our purpose: to create technology that is smarter, more sustainable, more resilient, and more human-centric.

At Altice Labs, we embrace this transformation with determination and a clear sense of purpose. The innovation showcased in the articles of this edition reflects our ability to anticipate trends, develop cutting-edge technologies, and deliver solutions that strengthen our partners, operators, and communities. From next-generation networks to AI-driven platforms, from cybersecurity to sustainable architectures, each project demonstrates not only our technical expertise but also our readiness to face the challenges and opportunities presented by today's world.

This is also a moment to transform the legacy that shaped Altice Labs into a catalyst for the future, reinforcing our commitment to change, collective innovation, and a bold, forward-looking vision. As we navigate this new chapter, the collective talent, creativity, and dedication of our teams remain our greatest strengths. We continue to push boundaries, explore new paradigms, and anticipate what comes next – from intelligent automation to advanced network architectures, from quantum-safe approaches to open, collaborative innovation models.

I invite you to explore this new edition of *InnovAction*, certain that it mirrors the energy of transformation that we are living and the ambition with which we build the future.

João Paulo Firmeza

General Manager of Altice Labs

PFirmeza@alticelabs.com



Editorial note

Welcome to this year's edition of our technical magazine, where we delve into the forefront of innovation in telecommunications, artificial intelligence, and digital connectivity. This collection of fifteen articles showcases the remarkable breadth and depth of research and development underway at Altice Labs and its partners, offering a compelling glimpse into the future of intelligent automation, connected systems, and data sovereignty.

We start with **"Open Innovation at Altice Labs –The Road Ahead"**, which outlines the shift from isolated innovation to collaborative ecosystems, emphasizing co-creation as a driver of technological progress.

The next set of articles explores how AI is transforming telecommunications, encompassing customer experience, enterprise automation, network design and configuration. These contributions share a vision of smarter, more adaptive, and inclusive systems:

- **"Towards Agentic AI: The Next Evolution in Enterprise Automation"** introduces autonomous systems capable of decision-making and contextual adaptation. The work of Altice Labs' BOTSchool reveals both the potential and complexity of integrating agentic flows into enterprise environments.
- **"AURA: Bridging Human Voice and Network Reality through AI-Powered Diagnostics"** presents a system that converts voice complaints into structured insights, enabling early detection of service issues and improving customer satisfaction.
- **"Towards MCP Server on NOSSIS-Based Autonomous Operations"** describes a modular protocol for secure, context-rich interactions between AI models and operational systems, paving the way for intelligent OSS platforms.
- **"Agentic AI Meets Connected Home Ecosystem: A Case Study of MyButler"** brings agentic intelligence into smart homes. MyButler autonomously manages devices, optimizes energy use, and supports assisted living through multimodal interaction and scripting.
- **"Autonomous Network Design: Leveraging AI in Reducing CAPEX and Planning Time"** showcases an AI-driven framework using genetic algorithms to optimize GPON topologies, reducing costs and improving scalability.
- **"Prompting the Shift: Democratizing Telco Configuration with Low-Code and AI"** explores how low-code platforms and AI interfaces empower non-experts to manage telco systems, emphasizing trust and human-centered design.

- **“A SHAP-Driven Framework for Misclassification Analysis and Feature Optimization in ML Models”** tackles AI interpretability by using SHAP values to analyze misclassifications and optimize model performance.
- **“Network Configuration and Operation via Intent-Based Networking”** introduces a paradigm where operators express high-level intents, automatically translated into network actions. The article outlines Altice Labs’ plans under the ALBATROZ project.

The next set of articles focuses on network transformation, where fiber, 5G, and automation converge to shape next-gen connectivity. These contributions reflect Altice Labs’ commitment to scalable and intelligent connectivity:

- **“Powering the Future of Connectivity with Integrated Edge and AI in 5G Small Cells”** presents a modular 5G solution with embedded edge computing, enabling ultra-low latency and localized intelligence for verticals like healthcare.
- **“Driving the Expansion of the 5G Ecosystem through Impactful Success Stories”** highlights Altice Labs’ 5G lab in Aveiro and its role in national and European research, fostering innovation and collaboration.
- **“The Next Step for NG-PON2: 50G TWDM-PON”** explores advancements in fiber access networks, positioning 50G TWDM-PON as a foundation for future-proof broadband infrastructure.
- **“VHSP Technology Assessment: IMDD vs. Coherent PON”** compares two high-speed PON technologies. IMDD offers simplicity and cost-efficiency, while Coherent PON provides superior performance for complex deployments.

Finally, the last two articles address trust and control in the digital age. As data and computation shape our world, securing communications and ethical governance become essential:

- **“Securing Today Against Tomorrow’s Quantum Threats”** discusses quantum computing’s threat to cryptography and advocates for Post-Quantum Cryptography (PQC) as a safeguard, highlighting telecom’s role in quantum-safe communications.
- **“Shaping the Future of Data: Why Gaia-X and IDSA are Crucial for a Stable and Innovative Digital Europe”** calls for data sovereignty through federated frameworks that promote ethical governance and empower European organizations.

Together, these fifteen articles reflect a bold and thoughtful approach to innovation. They demonstrate how Altice Labs and its partners are not only responding to technological trends but actively shaping them, building systems that are intelligent, inclusive, and resilient.

We invite you to explore our insightful perspectives and imagine what comes next!

Ana Margarida Almeida, ana-margarida-almeida@alticelabs.com

Pedro Carvalho, PCarv@alticelabs.com

Nuno Silva, NunoS@alticelabs.com

Contents

01

Open Innovation at
Altice Labs – The
Road Ahead

(p. 8)

02

Towards Agentic AI:
The Next Evolution in
Enterprise Automation

(p. 18)

03

AURA: Bridging Human
Voice and Network
Reality through AI-
Powered Diagnostics

(p. 32)

04

Towards MCP
Server on NOSSIS-
Based Autonomous
Operations

(p. 44)

05

Agentic AI Meets
Connected Home
Ecosystem: A Case
Study of MyButler

(p. 56)

06

Autonomous Network
Design: Leveraging AI
in Reducing CAPEX
and Planning Time

(p. 72)

07

Prompting the Shift:
Democratizing Telco
Configuration with
Low-Code and AI

(p. 88)

08

A SHAP-Driven
Framework for
Misclassification
Analysis and Feature
Optimization in ML
Models

(p. 106)

09

Network Configuration
and Operation
via Intent-Based
Networking

(p. 120)

10

Powering the Future
of Connectivity with
Integrated Edge and
AI in 5G Small Cells

(p. 138)

11

Driving the Expansion
of the 5G Ecosystem
through Impactful
Success Stories

(p. 152)

12

The Next Step for
NG-PON2, 50G
TWDM-PON

(p. 168)

13

VHSP Technology
Assessment: IMDD vs.
Coherent PON

(p. 182)

14

Securing Today
Against Tomorrow's
Quantum Threats

(p. 198)

15

Shaping the Future of
Data: Why Gaia-X and
IDSA are Crucial for a
Stable and Innovative
Digital Europe

(p. 218)



01

Open Innovation at Altice Labs – The Road Ahead

Abstract

Altice Labs has been a pioneer in electronics and telecommunications for nearly 80 years, evolving from a mindset where everything had to be made in-house to a dynamic and collaborative ecosystem that engages universities, research centres, startups, and technology partners worldwide. Fostering strategic partnerships with a broad range of entities enables Altice Labs to attract and retain top talent and knowledge through participation in projects supported by both national and European funding, as well as by developing internal initiatives. A notable example is the recently created “New Products Incubation Model”, which transforms ideas into prototypes and demonstrators while promoting knowledge transfer, experimentation, and collaboration with academia to accelerate innovation.

In parallel, Altice Labs and MEO have adopted a structured approach to corporate-startup collaboration, combining venture clienting, strategic partnerships, challenge prizes, and open calls, implemented through programs such as ENTER. Looking ahead, Altice Labs aims to amplify the impact of these collaborations, responding to a rapidly evolving telecommunications landscape in which operators are transitioning into Technology Companies (or “TechCos”). At the same time, emerging challenges (including AI, geopolitical fragmentation, and rising internal resistance to innovation) demand renewed strategic focus.

This article presents Altice Labs’ comprehensive approach to Open Innovation and outlines the path forward in a world where things are rapidly changing.

Authors

Fausto de Carvalho, Altice Labs, cfausto@alticelabs.com

Herlander Jorge Santos, Altice Labs, herlander-j-santos@alticelabs.com

João Ramos, MEO/GPE, joao.m.ramos@meo.pt

Lúis Manuel Reis, Altice Labs, luis-m-reis@alticelabs.com

Pedro Carvalho, Altice Labs, pcarv@alticelabs.com

Raquel Ariana Pinheiro, Altice Labs, raquel-a-pinheiro@alticelabs.com

Keywords

Open Innovation, Internal Incubation, New Products Incubation Model, Startups, Collaboration

DOI

10.57650/INNOVATION/2025.01

Introduction

Altice Labs was created almost 80 years ago, following the surge of landline telephone installation and utilization all over the world, under the original designation of GECA (the Portuguese acronym for “Special Group for Automatic Switching”). Founded at the same time as electronic voice communications were being established across Europe, the company has, from the beginning, followed and often shaped the evolution of state-of-the-art communication technologies.

Being at the forefront of a recent technology wave early in our history, we understood the value of innovation as a fundamental success factor, both for our customers and our business. However, the ability to innovate relies on one simple yet hard-to-obtain ingredient: knowledge. More specifically, in our line of business, knowledge of electronics and telecommunications technologies was scarce at the time. Hence, we also had to innovate in the creation of knowledge.

Recognizing the importance of education and research, and because of a partnership between our shareholders and the Portuguese Government, the first university to teach Electronics and Telecommunications Engineering in Portugal was established on our current campus location. Thus began a journey that continues to this day, always guided by our commitment to create better products for our customers, empowering them to lead their businesses.

Over the last decades, our innovation capacity, methods, and tools have evolved significantly. We transitioned from a mindset and business model where everything could be created, designed, and manufactured in-house – leading to the well-known “not invented here” syndrome – to a state of technological development where knowledge is widely available to everyone and most innovation emerges through collaboration.

Today, innovation involves a network of partners, from academia to Research & Development (R&D) companies, technology producers and end customers. Furthermore, innovation cycles, from the idea to the viable product, evolved from years to months, sometimes weeks, making it effectively impossible to innovate alone.

In the following sections, we will briefly describe our Open Innovation (OI) methodology, involving universities and research centers, internal incubation of ideas, and collaboration with startups. We will conclude with some considerations on the Road Ahead for Open Innovation at Altice Labs.

Working with universities: getting the right talent

Raw knowledge is the primary innovation fuel at Altice Labs, and it is created and nurtured at universities and research centers worldwide. Therefore, proximity to the entities producing the most relevant knowledge for Altice Labs is essential for us. Over the years, we have established collaborative relationships with most of Portugal’s leading universities and polytechnic schools, formalized through signed protocols that define the objectives and expected results of these partnerships.

Altice Labs is fueling tomorrow's breakthroughs with today's brightest minds

Our research and innovation work with universities and research centers follows different approaches, depending on the objectives pursued. When categorizing our interactions with those entities, we identify the following lines of action:

- **Talent identification and hiring** – We engage with our partner universities (and are often approached by them) regularly, following the academic agendas to propose engineering projects or master’s thesis topics to graduating students. Selected students join the company part-time while developing their work, and the top graduates are invited to join our team permanently. More recently, we have extended this approach to include PhD students, both internal and external to our institution.
- **Contract research** – When we face a complex product development challenge that requires time-consuming internal research or for which we do not possess the necessary knowledge, we approach selected university partners, discuss the challenge with them, and eventually set up a research and development contract with the best proposer. Our internal teams closely monitor the research, and once completed, the results are transferred to Altice Labs and integrated into our product innovation process.
- **Product development projects** – Our preferred approach for developing new product lines and evolving existing ones is through collaborative research projects in partnership with specialized research centers. In these projects, they perform the industrial research tasks while we focus on the product development and marketing tasks. This model also facilitates access to public research and innovation funding.
- **Collaborative research & innovation** – We regularly prepare project proposals in exploratory and cutting-edge technological areas in partnership with European universities and research centers. This is an extremely interesting approach to acquiring the necessary knowledge for future technologies and products, often supported by European Union (EU) funding that helps mitigate the risks associated with exploration and experimentation.

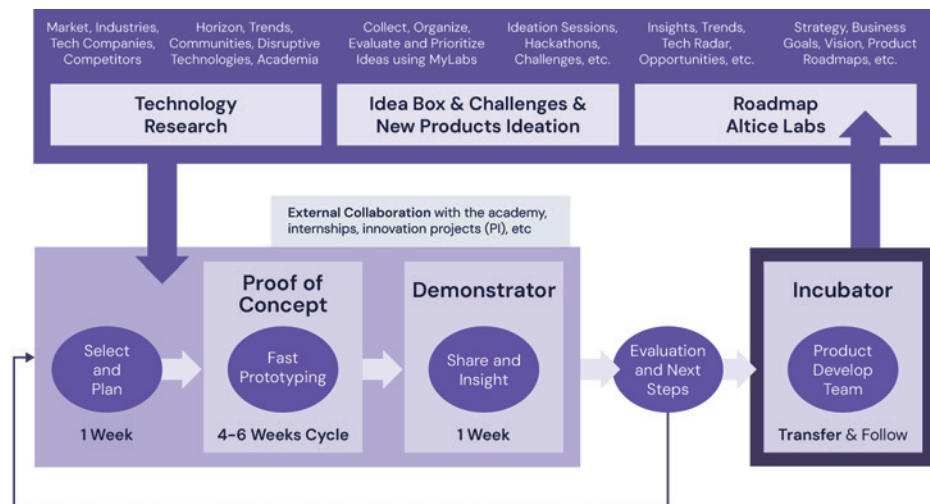


Internal incubation: getting the best out of our talent

One particularly interesting way to leverage internal talent, knowledge, and entrepreneurial spirit is the activity of internal incubation, which we are exploring at Altice Labs as an integral part of our innovation process. This process involves the internal incubation of new products, which is grounded in the creation of demonstrators and prototypes that explore emerging technologies with the potential to become part of the company's product portfolio.

The "New Products Incubation Model", represented in **Figure 1**, is supported by several cross-organizational pillars, maintaining a close relationship with product areas and placing special emphasis on the outcomes of the Innovation Committee and the Altice Labs roadmap. Our technology research encompasses a range of activities, from market analysis to identifying technological trends across relevant industries to monitoring the technology strategy established by the innovation management team.

Figure 1: New Products Incubation Model



From brainstorming to prototyping, ideas evolve into innovations through rapid experimentation and collaboration

Idea generation is fueled by brainstorming and ideation sessions, hackathons, and other informal activities where new concepts, features, and products can be explored. Proposals are then evaluated and prioritized based on criteria such as market and demonstration potential, technical feasibility, and alignment with the company's strategy. The most promising incubation lines are selected and may be developed through rapid experimentation, proof-of-concept (PoC), or prototyping.

Rapid experimentation involves limited effort over a brief period, assessing the idea's potential, technological limitations, and feasibility. While a PoC entails preliminary development or integration to validate the technical and functional viability of the idea, a prototype represents a more advanced stage of the concept, incorporating feedback gathered through internal interactions and interim presentations, and may extend over a longer period (three to four months). A close relationship with academia enables short internships (such as the summer camp program), which enhance and amplify the internal capacity for experimenting and prototyping of emerging technologies.

These primarily technological activities are complemented by internal dissemination initiatives, including article and blog writing, as well as interactive presentations in hybrid formats (in-room or video conference), which are highly practical and often feature live demonstrations of developed prototypes. Some of these demonstrations are also showcased at Future Labs.

Once a functional prototype is achieved, the work produced by the incubation team – including prototype hardware, code, and documentation – is transferred to the relevant product team through joint sessions. This ensures effective engagement, knowledge transfer, and enables the prototype results to be incorporated and aligned with the product roadmap.

Working with startups: leveraging focus and speed

Innovation in partnership with other companies, particularly recently created technology corporations, is generally considered the most complex and advanced form of Open Innovation. This involves two or more entrepreneurial entities at different maturity stages, contractual and investment decisions, shared intellectual property rights (IPR), and, last but not least, several different cultures, objectives, and personalities.

In the following sections, we describe the main models for this type of cooperative innovation and explain how Altice Labs implements them in direct collaboration with MEO.

Corporate–Startup collaboration: best practices and models

In the context of corporate–startup collaboration, a robust set of best practices and proven collaboration models is increasingly being adopted worldwide. For corporates seeking growth and competitiveness through innovation, partnering with startups has evolved into a strategic core activity, rather than a peripheral initiative [1].

Startups bring agility, rapid prototyping capabilities, and shorter time-to-market. However, these advantages come with inherent risks, including resource scarcity, frequent pivots, or even irreversible failures in achieving product–market fit.

To mitigate these challenges, companies should adopt a comprehensive framework of best practices, regardless of the specific collaboration model selected. According to the European Innovation Council, four fundamental pillars underpin successful partnerships [2]:

1. Strategic Alignment

Clearly define objectives, expected outcomes, and mutual value propositions from the outset. Aligning goals ensures that startup solutions and capabilities directly support the corporates' strategic focus areas.

Corporate–startup collaboration has shifted from peripheral experiment to strategic engine of innovation

2. Mutual Commitment and Value Creation

Securing buy-in from corporate stakeholders is essential from top management to operational teams. Transparent feedback and iterative goal setting help close the gap in culture, pace, and expectations between corporates and startups.

3. Skills and Capabilities

Corporates should develop cross-functional teams capable of conducting comprehensive due diligence that encompasses legal, technical, and business aspects, while also providing mentorship, resources, and market insights to startups.

4. Experimentation and Iteration

While some partnerships may remain purely financial, the most impactful collaborations incorporate a hands-on experimentation component. Pilot projects, PoC, and rapid prototyping are critical to validating solutions and refining the collaborative approach.

Collaboration models

In addition to embracing best practices, corporates must select one or more collaboration models that align with their strategic objectives and the maturity stage of their startup partners. As stated in [1], **Table 1** discriminates the identified models for startup and enterprise collaboration:

Table 1: Corporate and startup collaboration models

Model	Description	Typical Application
Venture Clienting	The corporate acts as the startup's first client for its product	Rapid operational/process innovation
Accelerator/Incubator	Corporates provide tailored resources and mentorship to early-stage startups	Culture shift; incremental innovation
Challenge Prizes	Companies issue public calls or hackathons to source solutions	Ideation; broad solution scouting
Corporate Venture Capital	Equity investments in high-potential startups	Access to new markets; early signals of disruption
Strategic Partnerships/Joint Ventures	Deep, sustained co-development relationships	Full-scale integration; transformative change
Mergers & Acquisitions	Direct acquisition of startup assets or teams	Long-term capability and technology absorption

A notable example of mixed/diversified collaboration models is the Microsoft–OpenAI partnership, which has combined multiple approaches [3], [4], [5], [6]:

- 1. Venture Clienting** – Microsoft was an early adopter of OpenAI's GPT models, integrating them into Microsoft products and services ahead of widespread adoption.

2. **Corporate Venture Capital** – Microsoft made substantial equity investments in OpenAI.
3. **Strategic Partnership / Joint Venture** – A long-term collaboration involving co-development of AI technologies, integration into Microsoft products, and shared infrastructure.

Our approach at Altice Labs and MEO

At Altice Labs and MEO, we have implemented a hybrid strategy that combines Venture Clienting and Strategic Partnerships/Joint Ventures with Prize Challenges and Open Calls. These initiatives operate under the umbrella of the Open Innovation and Corporate Venture Program “ENTER”, developed in collaboration with technology startups and key stakeholders in the innovation and entrepreneurial ecosystem.

Programs like AIIA – “Altice International Innovation Award” – and projects as NextGen Mobility and Aveiro Smart Connected Spaces Test Beds (powered by Altice Labs, funded by the Portuguese Republic PRR Programme), MEO 5G Challenge, MEO 5G API Sprint or MEO Intern Hackathon form part of the recurring Open Innovation calls, which over the past several years have attracted a large and diverse pool of startups and individual talents.

Future Focus

Looking ahead, considering Open Innovation with startups, Altice Labs and MEO aim to strengthen and accelerate the outcomes of its startup collaborations, focusing on consistently improving performance against established metrics. Efforts will target:

- Boosting conversion rates from initial evaluation meetings to signed PoC or pilots;
- Increasing pilot-to-scale success, ensuring more initiatives progress to commercial contracts or large-scale rollouts;
- Maximizing strategic impact, driving stronger revenue growth, cost savings, and regulatory compliance achievements.

The road ahead for Open Innovation

Historically, Altice Labs’ customers have been telecom operators or related companies, operating under a B2B business model. The company has traditionally sold its products directly to these customers, while also providing installation, support, and operational services.

However, in a world where digital marketing and sales dominate, new opportunities arise for B2B companies to evolve into B2C players, selling directly to end customers. Our vision is to embrace this transformation, extending beyond

our current business-to-business model to reach individuals globally. By leveraging our expertise in the telecom sector, we aim to position ourselves in a worldwide addressable market, delivering innovative solutions that connect people everywhere.

Considering our traditional market, one fact needs particular attention: Telecom Operators are pivoting their businesses at an amazingly fast pace. Under pressure from demanding shareholders, fierce competition from other operators, OTT (Over-the-Top) market substitution (promoted by unregulated digital giants), and rapidly changing customer habits and demands, telecom operators are trying to survive and grow. As a result, many are choosing between becoming low-cost infrastructure providers or, more frequently, transforming into “TechCos”, increasingly operating like the giant software providers which are menacing them.

Innovation ecosystems must evolve fast, as AI disruption, geopolitical shifts, and cultural barriers reshape collaboration

A TechCo is characterized by innovation, agility, openness, diversity of digital technologies supported and the facility of integration with other digital and cloud businesses. To align with this new market positioning and the agility required of future TechCos, Altice Labs will need to adapt its innovation ecosystem, models, processes, and tools to fit in.

Nevertheless, due to other factors, Open Innovation scenarios are evolving at an amazing rate. Even though the traditional operational difficulties of OI – such as IPR management, venture investment, cultural resistance, and misalignment of partner objectives – still exist and must continue to be addressed, new high-impact challenges are emerging. These will require a structured response from Altice Labs, namely:

- **Artificial Intelligence (AI) and Data Governance** – Never before has an innovative technology been deployed with the capacity to disrupt so many lines of business, so fast, like AI promises to. Its pervasiveness introduces several new challenges with data governance, such as ownership and confidentiality, among others. In addition, managing IPR in a multi-partner environment, governed by AI Agents, may prove to be extremely difficult.
- **Geopolitical Fragmentation** – It is widely recognized that the global order established and generally accepted after the end of World War II is changing fast. This leads to the need to choose Open Innovation partners and their talent sources with extreme care, confirming their venture capital origins and reorganizing supply and partner chains to make them more resilient.
- **Increased Organization Resistance to Open Innovation** – With business uncertainty on the horizon and the need to prepare for the impact of AI, organizations are likely to focus on faster Return on Investment (ROI) projects and to avoid longer-term, more disruptive and uncertain initiatives with external partners. At the same time, AI replacement fear may revive a strong “not invented here” syndrome, reinforcing internal resistance to incorporating external, more agile partners into the innovation process.

We would like to close by recalling the famous words of Bob Dylan, Nobel Prize-winning songwriter and singer, written back in 1964 but remarkably fitting for the times we are living in:

*The line it is drawn
The curse it is cast
The slow one now
Will later be fast
As the present now
Will later be past
The order is rapidly fadin'
And the first one now
Will later be last
For the times they are a-changin'*

References

- [1] C. Stephan, M. Mordecai, and D. Sun, "Accelerate Enterprise Growth Through Startup Collaboration," Gartner, Doc. No. 6724434, 2025. Accessed: Aug. 12, 2025. [Online]. Available: <https://www.gartner.com/en/documents/6724434>
- [2] European Innovation Council and SMEs Executive Agency (EISMEA), "Unlocking Innovation through Corporate–Startup Collaboration. The EIC Corporate Partnership Programme," Publications Office of the European Union, Luxembourg, Luxembourg, 2025.
- [3] Microsoft, "Microsoft and OpenAI evolve partnership to drive the next phase of AI," blogs.microsoft.com, Jan. 21, 2025. Accessed: Aug. 12, 2025. [Online]. Available: <https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/>
- [4] OpenAI, "OpenAI and Microsoft extend partnership," openai.com, Accessed: Aug. 12, 2025. [Online]. Available: <https://openai.com/index/openai-and-microsoft-extend-partnership/>
- [5] CDO Magazine Bureau, "Microsoft and OpenAI Strengthen Strategic Partnership Through 2030," cdomagazine.tech, Accessed: Aug. 12, 2025. [Online]. Available: <https://www.cdomagazine.tech/aiml/microsoft-and-openai-strengthen-strategic-partnership-through-2030>
- [6] Aranca, "The Microsoft – OpenAI Partnership," aranca.com, Accessed: Aug. 12, 2025. [Online]. Available: https://www.aranca.com/assets/docs/Special%20Report_Microsoft-OpenAIPartnership.pdf



02

Towards Agentic AI: The Next Evolution in Enterprise Automation

Abstract

The field of generative Artificial Intelligence (AI) is undergoing a significant transformation. What began as rule-based chatbots and intent-driven assistants has evolved into a new generation of intelligent agents – systems capable of generating natural, human-like conversations while also acting with clear purpose, autonomy, and contextual awareness. Agentic AI marks this shift, introducing systems that can make decisions, plan, and dynamically adapt to user needs across multiple domains.

At BOTSchool, Altice Labs' conversational AI platform, we have been exploring how this approach can improve automation and deliver richer user experiences in business environments. Building agent-based capabilities has revealed both promising opportunities and practical challenges, particularly in areas like memory persistence, orchestration flows, security, and seamless integration with external tools. In this article, we share some insights, use cases, and offer our perspective on how we see agentic AI reshaping the way enterprises interact with digital systems.

Authors

Sara Furão, Altice Labs, sara-s-furao@alticelabs.com

Ricardo Ângelo Filipe, Altice Labs, ricardo-a-filipe@alticelabs.com

Jorge Miguel Sousa, Altice Labs, jorge-m-sousa@alticelabs.com

José Pereira, Altice Labs, jose-m-pereira@alticelabs.com

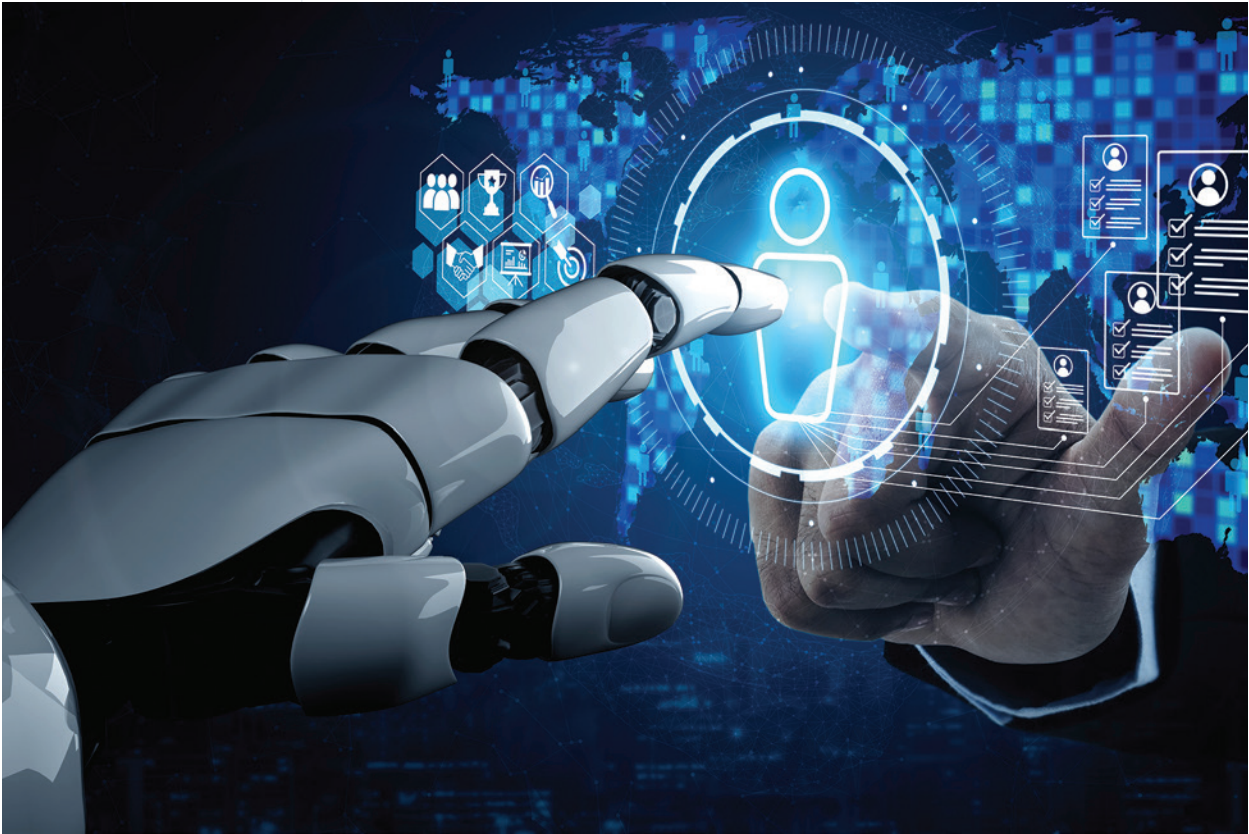
Rodrigo Almeida, MEO, rodrigo-l-almeida@telecom.pt

Keywords

Agentic AI, BOTSchool, LLM, MCP

DOI

10.57650/INNOVATION/2025.02



Introduction

Over the last decade, language models have advanced far beyond simple, static text generation, giving rise to intelligent agents – AI systems capable of reasoning, making decisions, and interacting with the world in adaptive, goal-oriented ways. This shift was catalyzed by a series of influential research breakthroughs that expanded the capabilities and ambitions of AI, leading to what is currently known as Agentic AI – systems able to have autonomy and agency.

A pivotal moment in this evolution came with the work of Christiano et al. [1], which enabled models to learn desired behaviors directly from human feedback rather than relying on inflexible, hand-crafted reward functions. Additionally, Wei et al. [2] revealed that models trained for specific tasks could generalize entirely new, unseen challenges without additional retraining. Other works enabled language models to interleave intermediate reasoning steps with action [3] or the integration of external tools, such as calculators, search engines, or databases, empowered language models to carry out complex tasks well beyond their pre-trained knowledge [4].

These developments collectively transformed language models from passive generators of text into agentic AI: active, autonomous entities capable of setting and pursuing goals, adapting to new situations, and engaging constructively with the world around them. AI has entered a new era, one defined by meaningful, responsible, and versatile agency. By leveraging existing product Application Programming Interfaces (API), we can integrate them into interconnect agents designed to process and execute user requests.

It is within this context that BOTSchool, a conversational AI platform, has been exploring how agentic AI can be applied in enterprise environments. Beyond the academic breakthroughs, our focus has been on the translation of these ideas into business automation and user experiences. In this article, we share insights, highlight real-world use cases, and offer our perspective on how agentic AI is re-shaping the way organizations interact with digital systems.

State of the art

Up until a few years ago, Natural Language Understanding (NLU) techniques used to define user Intent were the main core of virtual assistants' platforms. Intent-based platforms relied on predefined rules to match user input to known actions. They provided predictable, controlled responses but require manual updates for new cases, making conversations rigid if users deviate from the original topic. With the advent of generative systems, especially Large Language Models (LLM), new capabilities arose, with the LLM being able to answer a wider range of questions, including unexpected ones, with more natural, flexible, and human-like interactions. They can summarize, paraphrase, or generate creative solutions, offering a richer experience.

In this section, we present some of the emerging technologies and the current state of the art, focusing on AI Agents, their ability to use tools, the interaction between multiple Agents, and their integration into business workflows.

AI Agents

From prompts to performance: AI agents evolve, learn, and pursue your goals with precision and clarity

AI agents are advanced AI solutions built on LLM, designed to do much more than generate text. They combine sophisticated reasoning, persistent memory, domain-specific expertise, and the ability to interact with external tools, allowing them to handle complex, multi-step processes in rapidly changing business environments. Unlike conventional generative AI, which responds only to individual prompts, AI agents maintain context across interactions, adapt to evolving situations, and actively pursue defined business goals.

These agents retain knowledge from past interactions and accumulated data, giving them a more comprehensive understanding of context and the capacity to continuously improve performance. Many are customized for specific industries or business functions, embedding specialized expertise that enables them to manage complex or regulated scenarios – such as financial compliance, engineering design, or logistics – with precision and reliability.

AI Agents have some key aspects that allow them to interact with the world and improve reliability. First, with Tools or Model Context Protocol (MCP) servers [5], we enhance Agent's abilities to connect in real time with external applications, data sources, or even execute code. This connectivity allows agents to access current information, automate workflows, and execute tasks that translate insights into tangible business results. Secondly, to ensure delegation, agentic frameworks include the concept of handoffs. Handoffs enable routing of tasks or decision-making to specialized

sub-agents or human overseers when the agent reaches limits in confidence or authority. Finally, to ensure reliability, guardrails define boundaries for safe and compliant actions, allowing agents to validate inputs and outputs and halt execution early if rules are violated, preventing propagation of errors or undesired behavior.

Tools and MCP

LLM-based agents are strong in reasoning and general knowledge, but are limited to the training data. To overcome this, tools serve as external, specialized functions or systems that can be integrated with LLM-based agents. Tools enhance Agentic AI robustness and capabilities, making agents more reliable and suitable for real-world applications [10].

As organizations integrate diverse tools and services, MCP [5] emerge as a universal standard to unify connections between AI agents and business systems. MCP simplify integration by providing a consistent interface that breaks down data silos, lowers development overhead, and ensures secure and controlled access across the enterprise.

Multi-agent architectures

Agentic design patterns are the blueprints that transform specialized agents into autonomous systems capable of mastering complex, dynamic tasks beyond the reach of generalist AI

Generalist agents can struggle with complex tasks because of their broad scope and limited depth in specific domains. This limitation motivates the use of multi-agent architectures, where multiple specialized agents collaborate, each focusing on a distinct topic or expertise. When developing a multi-agent solution, we can follow a set of architectural and behavioral frameworks, commonly referred to as agentic design patterns, to enhance system robustness and efficiency. These patterns act as blueprints that guide the creation of autonomous AI systems capable of performing complex, dynamic tasks with structured workflows and adaptive decision-making. Although depending on the source, we might have distinct architectures and terminology [7], we can group them into two areas: sequential and parallel flows.

A sequential multi-agent system is adopted when a single AI agent can no longer handle all tasks effectively. Instead of concentrating responsibilities, an orchestrator distributes subtasks to specialized agents arranged in sequence. Each agent performs only its designated function, for example, querying data, analyzing results, generating visualizations, or drafting reports, and then passes the output to the next agent until the final answer is produced. This approach offers several advantages, such as greater specialization, modularity, scalability, and improved transparency in the process. However, it also comes with disadvantages: higher costs due to multiple model calls, increased coordination complexity, difficulties in debugging, and the need for rigorous maintenance to prevent overlapping responsibilities [8], [9].

A parallel multi-agent system follows an orchestrator-worker model. The lead agent receives the query, defines the strategy, and spawns specialized sub-agents that operate in parallel, each focusing on a specific aspect of the problem. The results are returned to the lead agent, which determines whether to initiate further branches or synthesize the findings. The final output might be

refined and completed by a dedicated citation agent. The advantages include higher efficiency in addressing open-ended problems, simultaneous exploration of multiple solution paths, more comprehensive information synthesis, and reduced risk of a single-agent bottleneck. The challenges, however, involve significant computational cost (high token usage), coordination complexity among agents, and potential redundancies or irrelevant explorations.

From these two core architectures, more complex ones were created, as discussed in [10].

Agentic frameworks

The evolution of agentic AI frameworks began with Microsoft’s AutoGen in September 2023, designed to build scalable multi-agent systems capable of collaborating through conversational interfaces. Its flexibility allowed developers to configure and compose agents dynamically, making it suitable for workflows requiring coordination among multiple agents.

In early 2024, CrewAI launched as a Python framework for orchestrating teams of agents, enabling specialized roles and coordinated workflows that balance high-level simplicity with low-level control.

In March 2025, OpenAI introduced the Agents Software Development Kit (SDK), a toolkit for rapid deployment of multi-step workflows with memory, tool integration, multi-agent orchestration, and production-ready support for tasks like web search, file handling, and code execution. Shortly after, Google unveiled the Agent Development Kit (ADK), providing a modular framework for enterprise-ready, conversational, multimodal agents. With support for Python and Java, seamless system integration, and session-based memory, it enables agents to handle complex workflows while maintaining context across interactions.

These frameworks trace a clear chronological progression, from scalable multi-agent orchestration with AutoGen, to team-like coordination in CrewAI, and finally to production-ready, enterprise-integrated solutions with OpenAI SDK and Google ADK [11]. **Table 1** shows an overall summary of our experiment and research done through some of the most known frameworks.

Table 1: Agentic frameworks overall compare

Attribute	Open AI Agents SDK	Google ADK	AutoGen	CrewAI	Lan Chain	Lang Graph	n8n
Scalability (throughput)	✓	✓	✓	⚠	✓	⚠	⚠
Open Source	✓	⚠	✓	✓	✓	✓	✓
Production ready	✓	✓	⚠	⚠	✓	✓	✓
Integration with Altice Labs	✓	⚠	⚠	⚠	⚠	⚠	✓

Conversational AI platform: evolution of customer journeys

Early conversational AI solutions were limited to rule-based or statistical dialogue systems [12]. The advent of LLM marked a breakthrough, enabling greater fluency, adaptability, and (in some contexts) open-ended reasoning capabilities [13].

Google Dialogflow is a platform for developing conversational interfaces, providing robust natural language understanding and seamless integration across diverse communication channels. More recently, n8n, an open-source workflow automation platform, has introduced LLM-driven agents to coordinate communication flows across messaging, APIs, and—with external integrations—voice channels, enabling low-code design of sophisticated interaction models. In parallel, CrewAI has gained traction with its multi-agent paradigm, where specialized agents collaborate within orchestrated workflows, highlighting how communication AI can evolve beyond single-agent architectures.

Although these platforms differ in their architectural approaches – Dialogflow emphasizes scalable NLU frameworks, n8n focuses on extensible low-code orchestration, and CrewAI pioneers multi-agent collaboration – they converge on the business objective to deliver intelligent orchestration of digital communication. This convergence highlights how conversational AI is transitioning from isolated dialogue management toward ecosystem-level orchestration, integrating human interaction, workflow automation, and agentic collaboration to enable next-generation communication platforms.



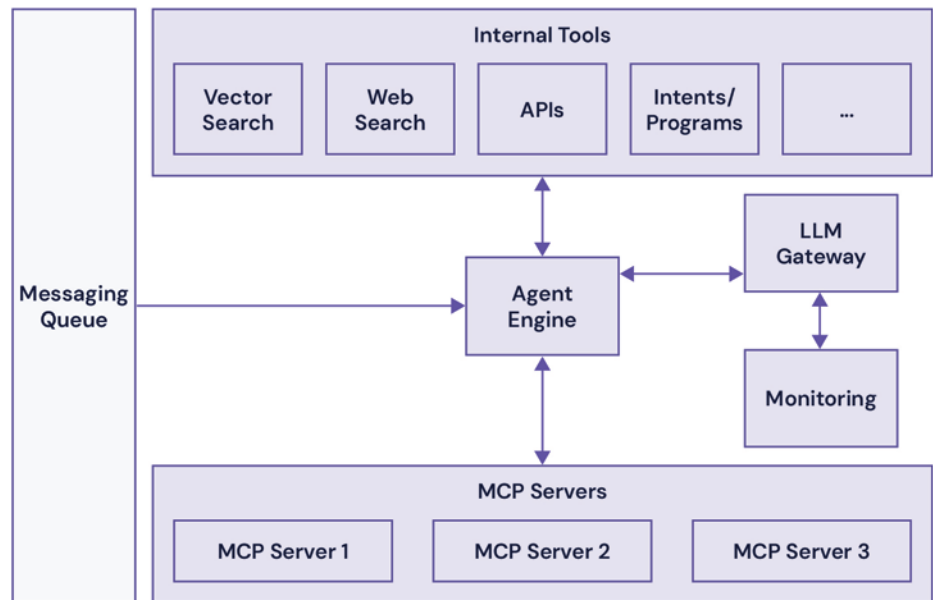
Development of agents – BOTSchool

Altiace Labs' BOTSchool has evolved from earlier NLU/NLP approaches to the more recent paradigm of Agentic AI. Despite this evolution, the goal remains unchanged: enhancing customer experience and overall satisfaction. In this section, we explore how Agentic AI was implemented in BOTSchool and integrated into workflow journeys, fostering a continuously evolving ecosystem rather than a disruptive shift.

Agent architecture

Our agentic solution was built using several state-of-the-art technologies. We took into consideration our own review. Additionally, the focus was to build an extensible architecture that integrates agent functions seamlessly. The system ensures resilience and scalability through a robust microservices architecture and comprises the core components shown in **Figure 1**.

Figure 1:
High-level architecture
of the Agent Engine
ecosystem



The main components are:

- **Agent Engine:** responsible for orchestrating the conversation flow, using the OpenAI Agent SDK, receives messages via Messaging Queue, interacts with the LLM Gateway service to perform reasoning and plan actions, and decides when to invoke the Tool Engine or access memory.
- **LLM Gateway:** provides a canonical interface to connect and manage different LLM providers.
- **MCP Servers:** enable AI agents to go beyond inherent knowledge by connecting them to external prompts, tools, or resources. BOTSchool agentic architecture was built with MCP in mind, understanding that many internal systems, namely Sigo [14], NetQ, MyButtler [15], already rely on their own pre-established toolsets and want to extend their use-cases with an Agentic AI layer. By supporting these tools through MCP interfaces, the system enables seamless integration, eliminating the need for custom connectors tailored to each client's existing infrastructure.
- **Internal Tools:** provide a plugin architecture that allows new tools to be added seamlessly. There are already some internal tools available for BOTSchool's clients, such as: API Tool, which allows agents to seamlessly interact with external API; Vector Search Tool, which retrieves information from vectorized databases, improving contextual awareness; Web Search Tool, which enables agents to conduct real-time searches across the web. Additionally, it is relevant to mention that a key strength of Internal Tools is the ability to bridge the pre- and post-Agentic worlds. Existing programs (models based on NLP) can be seamlessly integrated with Agentic AI through Internal Tools, ensuring more deterministic workflows when they are better suited to client needs.
- **Monitoring:** tracks detailed traces of every agent interaction, including LLM calls, decisions, and tool usage. The framework delivers comprehensive observability by tracing every step within agent workflows, capturing inputs, outputs, latency, token usage, and errors in detail. This enables precise

detection of hallucinations, bugs, and implementation defects, which are critical for maintaining system accuracy. It also allows to do some continuous improvement techniques such as human-in-the-loop, annotations, or evals.

Whenever a new message arrives in the Messaging Queue, the Agent Engine listening for incoming messages consumes it and begins processing. The first interaction with the LLM Gateway interprets the content of the message and understands the user's intent and necessary actions. Based on the classification returned by the LLM Gateway, the tools are identified as either internal or external (MCP-type). Once all required tools have been invoked and their results obtained, the agent-engine performs a second interaction with the LLM Gateway to compose the complete response, using the appropriate prompt instructions and contextual information. The final response is then sent back to the user through the messagingQueue. Throughout this entire process – from message reception to response delivery – all operations, events, and execution details are continuously tracked and recorded by the Monitoring component, ensuring full observability, traceability, and system transparency.

Journeys architecture

BOTSchool's communication platform is inherently omnichannel, allowing flows to remain independent of the interaction channel – whether voice, webhook, or any other touchpoint – while maintaining the same orchestration mechanism. This ensures a uniform user experience, enhanced scalability, and robustness throughout the entire journey.

Agents emerge as a new shape of intelligence in BOTSchool's communication journeys. They introduce advanced reasoning, adaptability, and execution capabilities, strengthening the existing architecture and significantly expanding the scope and impact of the journeys.

BOTSchool's communication journeys, illustrated in **Figure 2** and **Figure 3**, are designed to simplify and automate complex processes through a low-code

Figure 2:
High-level architecture
of BOTSchool's
communication
journeys

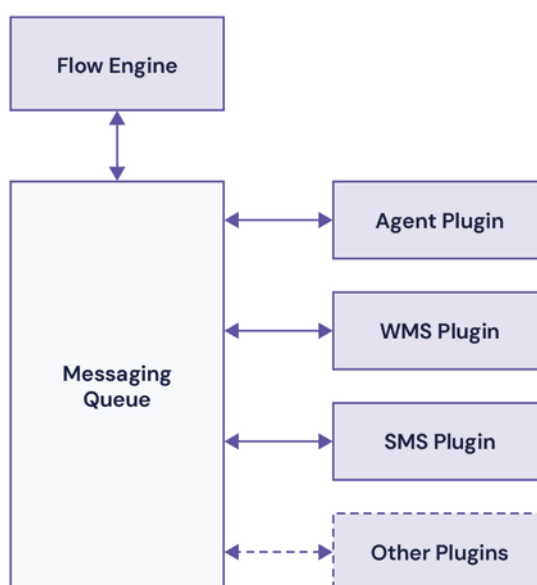
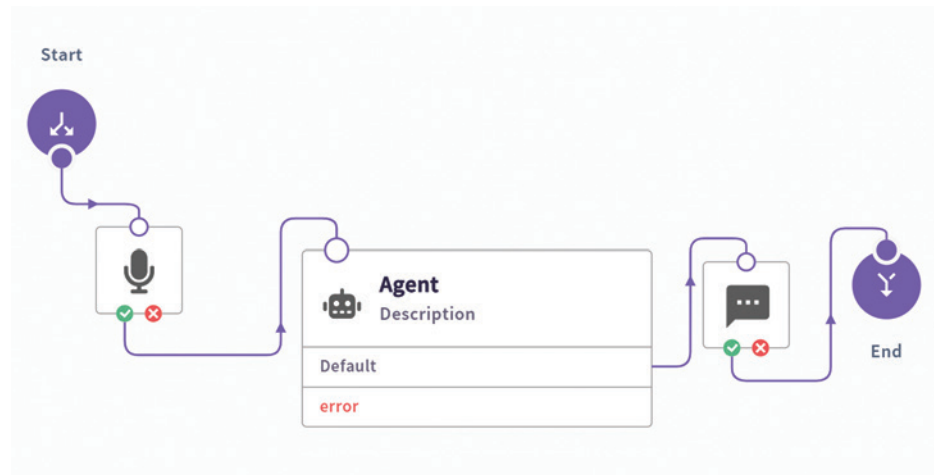


Figure 3: Frontend interface of BOTSchool's communication platform



approach. These journeys are defined as structured flows, built from interconnected nodes that represent actions, decisions, or integrations. Organized in a Directed Acyclic Graph (DAG) structure, the flows describe how information moves and decisions are taken step by step. For example, a journey can start by detecting a voice interaction through the Windless Media Server (WMS), followed by handling the interaction with an agent, and concluding by triggering a Short Message Service (SMS) confirmation. This approach enables users to design end-to-end communication experiences in a streamlined way, without requiring deep technical expertise.

At the core of this architecture, as shown in **Figure 2**, lies the Flow Engine, the orchestrator responsible for receiving events from multiple sources and routing them through the defined nodes. Each event traverses the journey's DAG, ensuring that actions are executed in the correct order. This orchestration is supported by an extensibility layer, materialized through plugins that expand and specialize the system's capabilities. Existing plugins, such as the SMS Plugin for direct communication with users, and the WMS Plugin with integrated Interactive Voice Response (IVR), Text-to-Speech (TTS), and Automatic Speech Recognition (ASR) capabilities, illustrate how journeys can integrate external services and engage through multiple channels.

Among these, the Agent Plugin stands out as a key component, acting as the intermediary between the Flow Engine and the Agent Engine. This integration ensures seamless alignment between the journey logic orchestrated by the Flow Engine and the decision mechanisms powered by agents, enabling more coherent and scalable interactions across all channels.

BOTSchool empowers communication journeys with intelligent agents, low-code flows, and scalable channels, transforming complex interactions into simple experiences

Use cases

Automatic product catalog registry

The use case is straightforward: when a new product enters the market, it must be registered in the web store, ensuring customers can properly compare it with similar products. Until now, this process has been entirely manual, taking roughly

1 hour per product. With Agentic AI, the workflow becomes much faster and more efficient – the AI automates data collection and entry, while the human agent shifts to a quality assurance role, ensuring accuracy and consistency.

The adoption of an Agentic AI solution, combining different/multiple tools such as Web Search and Vector Store, for example, enables intelligent optimization of product sheet creation and completion. Through Web Search, the agent automatically collects accurate and up-to-date information from multiple reliable external sources, while the Vector Store ensures efficient organization and retrieval of the company's internal knowledge, and from ground-truth sources. This approach allows the agent to operate autonomously – searching, analyzing, and consolidating relevant data – to generate complete and precise product sheets with minimal manual intervention. Beyond process automation and productivity gains, the economic impact is substantial, with the time required per sheet reduced by nearly 90%.

Increase productivity in the physical store

Agentic AI empowers retail staff with instant product insights, real-time comparisons, and promotions, driving faster decisions, stronger trust, and higher sales

In physical retail environments, customers often ask store staff about product features, ongoing campaigns, or comparisons between different models in the same category. Traditionally, this requires employees to rely on memory, printed materials, or manual lookups, which can slow down service and reduce the overall customer experience.

With an Agentic AI-powered assistant, employees gain instant access to consolidated product data, promotional details, and side-by-side comparisons. Through vector store knowledge, real-time web search, or integration with internal systems, the AI assistant can provide accurate, context-aware responses on the spot. This not only enhances the speed and quality of customer interactions but also boosts staff confidence in delivering reliable information.

The result is a more productive sales floor, where human assistants can focus on building trust and guiding purchase decisions while the AI handles the heavy lifting of information retrieval. This hybrid approach increases conversion rates, shortens decision cycles, and creates a more engaging customer experience, ultimately translating into higher sales and customer loyalty.

Agentic AI in Call Centers

Call centers, whether technical or non-technical, face growing challenges: increasing complexity in diagnostics, rising customer expectations, and escalating operational costs that make traditional models less scalable. Although call centers have already leveraged AI (particularly traditional approaches such as NLP) to successfully support the handover of self-service operations to virtual assistants, Agentic AI introduces several new capabilities that significantly enhance the self-service rate of operations. Autonomous, goal-driven agents that can handle repetitive tasks and structured customer interactions, allowing human agents to focus on the cases where empathy, creativity, or complex decision-making are essential.

In technical call centers, Agentic AI can evolve platforms into AI-native diagnostic systems, becoming the backbone for network and platform intelligence. These AI agents not only accelerate resolution times but also reduce the workload on technical staff, freeing them to concentrate on high-value problems that require domain expertise or nuanced judgment. In non-technical call centers, Agentic AI empowers customer service with seamless automation of inquiries such as account management, order tracking, or basic troubleshooting. Cognitive agents can manage the majority of routine interactions end-to-end, while human operators step in for more challenging conversations where emotional intelligence or negotiation is critical.

AI agents can have access to vector stores, web search, other tools, and even MCP servers. This enables them to act, for example, by creating support tickets, analyzing real-time network status, or enriching customer profiles with contextual insights from both internal and external sources. This integration transforms call centers into intelligent ecosystems where AI agents do not just respond but proactively manage and optimize operations.

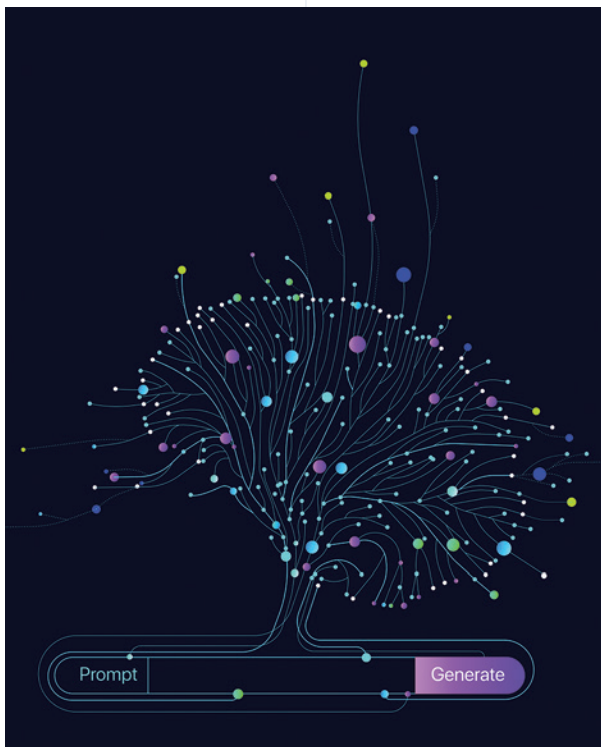
Conclusion and future work

In the industry, we can find a plethora of agentic AI projects reshaping Enterprise Operations and Products. Studies show that where agents are in place, the impact is clear – 66% see productivity gains, 57% cost savings, and finance teams report up to 90%-time savings and 40% better forecasting [16]. JPMorgan's Coach AI [17] helps advisers access research 95% faster, contributing to 20% sales growth, and Gartner expects Generative AI to hit \$644B by 2025 [18].

Although there is a great opportunity ahead, several challenges arise for Agentic AI projects. There are two main barriers to a swifter transformation. First, technical challenges such as escalation cost or unclear business. Secondly, organizational problems: resistance to change, slow workforce adoption, and workflows designed for human interactions.

In this paper, we explored the evolution, capabilities, and frontiers of Agentic AI. As a practical step, we integrated Agentic AI into our low-code conversational AI platform, BOTSchool, to enrich client interactions and pursue goals on the client's behalf. By leveraging BOTSchool's features and seamless integration with existing IT systems through Tools and MCP servers, enterprises can deploy agents that resolve requests, access information, and complete tasks across channels, reducing costs and delivering better customer experience.

For future work, there are three main directions worth exploring. First, Small Language Models (SLM), which – although they do not match LLM in open-ended



conversations – may be better suited for agents due to lower cost, faster execution, and the ability to be fine-tuned for domain-specific tasks, reducing unexpected behaviors such as hallucinations. Secondly, the Agent Communication Protocol (ACP) [19], which focuses on messaging and interaction between agents. Finally, the Agent-to-Agent Protocol (A2A) [20] aims to improve coordination, task delegation, and collaboration among multiple agents.

As seen in this paper, agentic AI is already delivering measurable benefits. Yet its full potential is likely far greater than current metrics suggest. This moment in time echoes Amara's Law: "We tend to overestimate the effect of a new technology in the short term and underestimate its long-term impact". The cumulative impact of agentic AI on enterprises adopting platforms such as BOTSchool may ultimately prove far greater than we currently imagine.

References

- [1] P. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," 2017, arXiv:1706.03741.
- [2] J. Wei et al., "Finetuned language models are zero-shot learners," 2022, arXiv:2109.01652.
- [3] S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," 2023, arXiv:2210.03629.
- [4] A. Parisi, Y. Zhao, and N. Fiedel, "TALM: Tool augmented language models," 2022, arXiv:2205.12255.
- [5] Anthropic, "Introducing the model context protocol," Anthropic. Accessed: Sep. 25, 2025. [Online]. Available: <https://www.anthropic.com/news/model-context-protocol>
- [6] Z. Xi et al., "The rise and potential of large language model based agents: A survey," 2023, arXiv:2309.07864.
- [7] LangChain, "Multi-Agent architectures," LangGraph. Accessed: Oct. 6, 2025. [Online]. Available: https://langchain-ai.github.io/langgraph/concepts/multi_agent/#multi-agent-architectures
- [8] Seahorse, "Sequential agentic workflow in LangGraph," Medium. Accessed: Oct. 3, 2025. [Online]. Available: <https://medium.com/@seahorse.technologies.sl/sequential-agentic-workflow-in-langgraph-0c17d7e0d51e>
- [9] T. Chen and C. Rickman, "Semantickernel: MultiAgent orchestration." Microsoft DevBlogs. Accessed: Oct. 3, 2025. [Online]. Available: <https://devblogs.microsoft.com/semantic-kernel/semantic-kernel-multi-agent-orchestration/>
- [10] Anthropic. "How we built our multi-agent research system." Anthropic. Accessed: Oct. 3, 2025. [Online]. Available: <https://www.anthropic.com/engineering/multi-agent-research-system>

- [11] Harsh. "OpenAI Agents SDK vs LangGraph vs AutoGen vs CrewAI." Composio Blog. Accessed: Oct. 3, 2025. [Online]. Available: <https://composio.dev/blog/openai-agents-sdk-vs-langgraph-vs-autogen-vs-crewai>
- [12] M. M. Mariani, N. Hashemi, and J. Wirtz, "Artificial intelligence empowered conversational agents: A systematic literature review and research agenda," *J. Bus. Res.*, vol. 161, p. 113838, 2023, doi: 10.1016/j.jbusres.2023.113838.
- [13] T. Adewumi, F. Liwicki, and M. Liwicki, "State-of-the-Art in open-domain conversational AI: A survey," *Information*, vol. 13, no. 6, Art. no. 298, Jun. 2022, doi: 10.3390/info13060298.
- [14] M. M. Castro, R. Â. Filipe, C. G. Araújo, and L. Cortesão, "Towards the MCP server on NOSSIS-based Autonomous Operations," *InnovAction*, no. 10, pp. 44-55, 2025.
- [15] P. Chainho, R. Â. Filipe, V. Santos, J. L. Silvestre, and J. Sousa, "Agentic AI meets Connected Home ecosystem: A case study of MyButler," *InnovAction*, no. 10, pp. 56-71, 2025.
- [16] PwC. "AI Agent Survey." PwC. Accessed Sep. 24, 2025. [Online]. Available: <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-agent-survey.html>
- [17] N. Anand. "JPMorgan says AI helped boost sales, add clients in market turmoil." Reuters, Accessed: Sep. 24, 2025. [Online]. Available: <https://www.reuters.com/business/finance/jpmorgan-says-ai-helped-boost-sales-add-clients-market-turmoil-2025-05-05/>
- [18] Gartner, Inc. "Gartner forecasts worldwide GenAI spending to reach \$644 billion in 2025." Gartner. Accessed: Sep. 24, 2025. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/2025-03-31-gartner-forecasts-worldwide-genai-spending-to-reach-644-billion-in-2025>
- [19] R. Surapaneni, M. Jha, M. Vakoc, and T. Segal. "Announcing the Agent2Agent Protocol (A2A)." Google Developers Blog. Accessed: Sep. 24, 2025. [Online]. Available: <https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>
- [20] K. Martineau. "An open-source protocol for AI agents to interact." IBM. Accessed: Sep. 24, 2025. [Online]. Available: <https://research.ibm.com/blog/agent-communication-protocol-ai>



03

AURA: Bridging Human Voice and Network Reality through AI-Powered Diagnostics

Abstract In today's telecommunications landscape, customer complaints often serve as an early signal of potential service degradations. However, they tend to be unstructured and subjective, making it challenging to correlate them with technical data. AI for Understanding and Resolving Anomalies (AURA) addresses this challenge by structuring customer feedback and combining it with real-time diagnostic test results. Using a continuous analysis process, AURA trains machine learning models to detect anomalies early across multiple features. This approach uncovers emerging issues – whether caused by equipment faults, misconfigurations, or customer-specific conditions – so operators can act proactively. Operational results show AURA detects degradations sooner, reduces false alerts, speeds up case resolution, and improves customer satisfaction. The paper presents AURA's methodology, key deployment results, and future applications in mobile and Internet of Things (IoT) networks.

Authors **Cátia Felizardo**, Altice Labs, catia-s-felizardo@alticelabs.com
João Almeida, Altice Labs, joao-f-almeida@alticelabs.com
Márcio Sousa, Altice Labs, marcio-m-sousa@alticelabs.com
Rita Oliveira, Altice Labs, rita-a-oliveira@alticelabs.com
Daniel Teixeira, Altice Labs, daniel-o-teixeira@alticelabs.com
Ricardo Ângelo Filipe, Altice Labs, ricardo-a-filipe@alticelabs.com

Keywords NOSSIS Genius, NOSSIS Assurance, Agentic AI Studio, AURA, Customer Experience, Call Center

DOI 10.57650/INNOVATION/2025.03

Introduction

Telecommunication operators face a growing challenge: ensuring consistent, high-quality service delivery while understanding, in real time, what customers are truly experiencing. The digital transformation of networks has brought agility, but also complexity, decentralization, and reduced visibility over the customers' subjective perception of quality of experience. Despite the increasing sophistication of monitoring systems, these tools often miss the nuances that only human feedback can reveal, such as intermittent performance, degraded app usability, or confusing service behavior [1].

Customers report these issues through a range of natural language interactions including interactive voice response (IVR) calls, chatbot exchanges, and app-based complaints, using intuitive and emotional descriptions, for example, "The internet is working but too slow to watch Netflix," "I cannot find a channel I used to have," or "Calls keep dropping every evening." Such feedback is invaluable but notoriously difficult to process. Conventional systems often disregard or treat it as noise, resulting in recurring challenges:

- Problems are detected reactively, only after many complaints are accumulated;
- Agents follow generic troubleshooting scripts disconnected from actual technical states;
- Valuable patterns go unnoticed due to a lack of structured correlation between user expression and network telemetry [2].

Another critical challenge arises when multiple customers with similar issues call within the same time window but are queued and handled independently. Without early detection of a shared root cause, these cases are misclassified as isolated, leading to overloaded call queues and prolonged resolution times. Meanwhile, unrelated issues are handled in the same flows, creating inefficiency and growing dissatisfaction. AURA helps identify and group such converging cases earlier, enabling better triage, more targeted communications, and reduced queue congestion.

AURA was created to close these gaps. It acts as a correlation layer between what the customer says and what the network reports, turning qualitative complaints into structured, explainable diagnostic signals. It does this by representing each interaction as a feature vector and evaluating it against known patterns of service behavior, allowing operators to detect emerging problems before they escalate.

AURA integrates three pillars:

- **Voice of the customer**, captured via NOSSIS Genius AI studio Natural Language Understanding (NLU) interface;
- **Real-time network insights**, provided by NOSSIS Assurance Diagnostics;
- **Correlation and pattern recognition**, powered by itemset mining and outlier detection.

This article presents the rationale behind AURA, its architecture, methods, and operational outcomes. It argues about the strategic importance of aligning customer-side point-of-view and engineering, making customer voice an actionable signal for proactive network care.



State of the art

Modern telecommunication environments rely heavily on a wide array of monitoring tools to detect and resolve network and service disruptions. These include Simple Network Management Protocol (SNMP) based network management systems, passive telemetry collectors, flow analyzers, and topology-aware fault detectors. While these tools offer deep technical insight into specific infrastructure components, they tend to operate independently of the systems that interface directly with the customer, namely IVR flows, chatbots, and self-care applications [3].

If a customer calls the contact center to report a service anomaly, technical diagnosis typically starts with the customer context and device fingerprinting, then triggers remote tests (e.g., TR-O69/USP customer premises equipment (CPE) checks, line/optical power, synthetic probes) and cross-correlates the results with outage maps, maintenance windows, and Operations Support System/Network Management System (OSS/NMS) alarms to triage faults across home Wi-Fi, access, transport/core, or Content Delivery Network (CDN).

Most existing anomaly detection frameworks are designed around predefined thresholds, static topologies, or signature-based alarms, with intensive instrumentation (i.e., white-box approach). While this works well for known fault patterns, it fails to capture degradations that manifest primarily through user-experience signals, such as video freezing, inconsistent app behavior, or availability issues for specific channels/segments. This challenge is further compounded by the limitations of black-box observability, where only external symptoms are visible, making it difficult to pinpoint root causes without deeper system insights [4]. In recent years, advances in natural language processing and pattern mining

have significantly expanded the ability to process and understand unstructured data. Techniques such as intent classification, sentiment analysis, summarization, and keyword extraction enable the interpretation of customer messages. Frequent pattern mining algorithms such as Apriori and FP-Growth reveal a group of items (i.e., itemsets) that occur together, highlighting associations and behavioral trends [5], [6].

Recent developments in generative AI have enhanced the capability to analyze unstructured customer interactions beyond traditional intent classification and keyword extraction. Large Language Models (LLM) can function directly with call transcripts and chat records to:

- Generate abstractive summaries of customer complaints, making them concise and machine-actionable;
- Extract themes and categories that go beyond surface keywords, enabling robust classification of issues (e.g., IPTV vs. broadband vs. device);
- Determine key terms and entities that encapsulate the core of the customer's experience.

These capabilities position LLM as powerful complements to traditional anomaly detection, enabling a more faithful representation of how customers describe their problems.

Despite these developments, few operational platforms integrate NLU with service assurance in a unified anomaly detection strategy. Most telecom AI initiatives continue to treat these domains separately, with one side focused on infrastructure performance and the other on customer interaction.

AURA seeks to bridge this divide by integrating both black-box and white-box observability paradigms. Its approach is grounded in the idea that actionable intelligence emerges when customer experience signals (black-box) and network telemetry (white-box) are evaluated simultaneously rather than in isolation.

This positions AURA within a new generation of intent-aware, behavior-centric diagnostic platforms, capable of transforming free-text and voice complaints into structured, explainable, and actionable insights that support proactive operations and more empathetic customer care.

High-level architecture

NOSSIS is Altice Labs' modular, API-first OSS [11] suite that underpins inventory, fulfillment, and assurance for digital telcos. Within this suite, NOSSIS Assurance Diagnosis provides the end-to-end service assurance and diagnostics layer exposing real-time service-health indicators, topology/provisioning context, and troubleshooting signals across access, aggregation, and home domains. Complementing it on the customer-interaction side, NOSSIS Genius AI Studio (powered by BOTSchool, Altice Labs' conversational AI platform) delivers AI capabilities for conversation understanding, including summarization and keyword extraction over chat inputs, yielding normalized, intent-centric features. AURA consumes both streams to correlate customer-side signals with network/service

NOSSIS demonstrates how converging conversational AI with network diagnostics enables a measurable leap in service-health understanding

telemetry. With these roles defined, **Figure 1** details the high-level architecture: from data ingestion and time-aligned window processing, through baseline/drift analytics and itemset mining, to the generation of explainable common-fault clusters that feed operations and call-center tools.

1. Data Ingestion

- a. Customer side stream: The Customer Issue Feed collects voice and digital interactions; voice calls are processed by automatic speech recognition (ASR), enriched with NLU and NOSSIS Genius AI Studio for summarization and keyword extraction;
- b. Network side stream: NOSSIS Assurance Diagnostic provides service health indicators, diagnostics, and topology/provisioning data.

2. Processing Window

- a. Both streams are synchronized within a rolling observation window of four hours, refreshed every five minutes;
- b. This strategy balances responsiveness with statistical robustness.

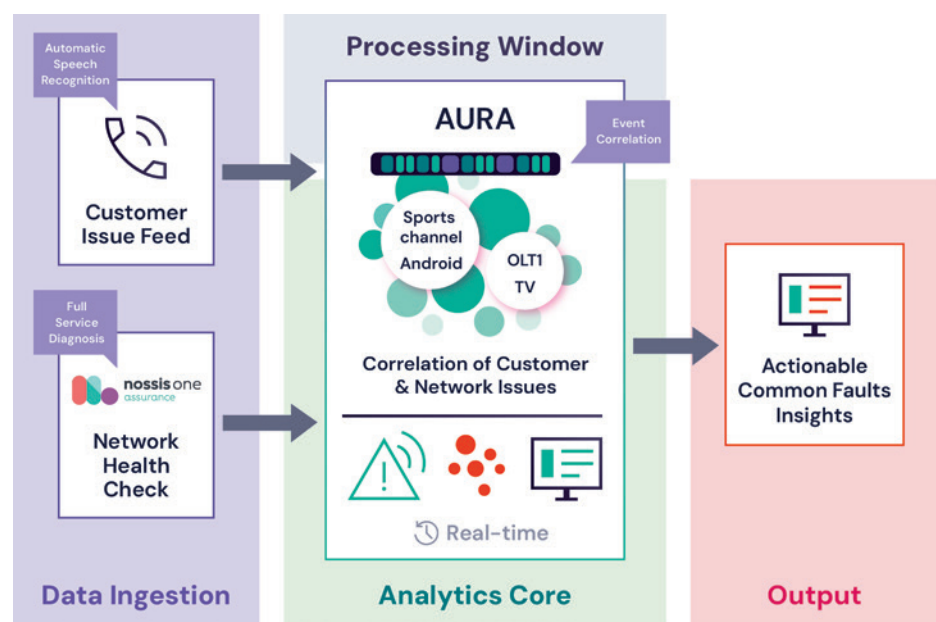
3. Analytics Core

- a. Computes empirical feature distributions and monitors drift;
- b. Applies frequent itemset mining to surface recurring co-occurrences;
- c. Performs clustering and outlier detection to organize anomalies into operationally meaningful groups.

4. Outputs

- a. Generates interpretable “common-fault clusters” with label, impacted scope, and confidence score;
- b. Results are published in real time to network operations center (NOC) dashboards and customer operations tools, enabling triage and proactive messaging, opening the possibility to trigger adaptive IVR/chatbot flows.

Figure 1: AURA high-level architecture

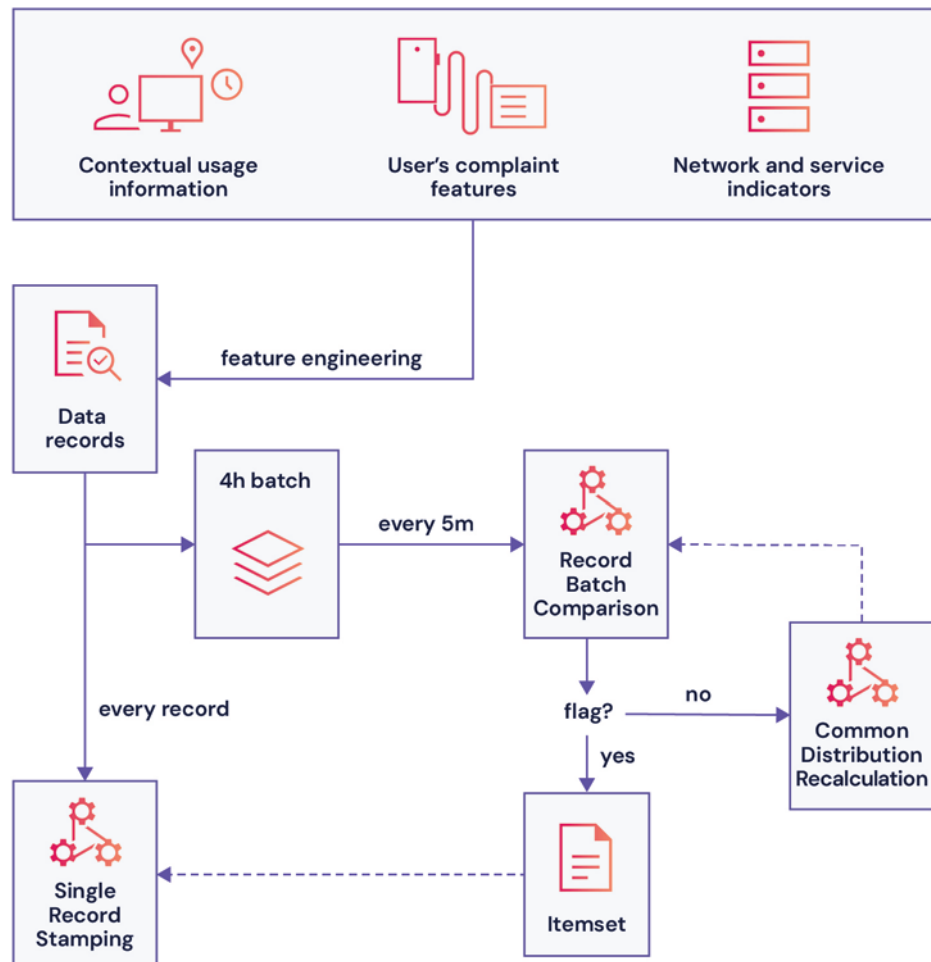


Methodology

AURA operationalizes customer and network-side evidence into explainable diagnostics through a structured, real-time pipeline. The design follows three principles: signal parity (customer interaction signals and network/service indicators contribute with equal weight), temporal alignment (all features are evaluated on a synchronized rolling window), and operational explainability (outputs must be interpretable and actionable for care and operations teams). Inputs are gathered from voice/chat interactions, contextual usage of metadata, and live assurance telemetry. These heterogeneous signals are normalized, quality-scored, and fused into records suitable for statistical comparison and pattern discovery.

As summarized in **Figure 2**, the process unfolds in four stages:

Figure 2: AURA methodological pipeline



1. **Feature modeling:** each interaction voice, chat, or app is converted into a structured feature vector that may include:
 - **User's complaint features:** To process spoken input, NOSSIS Genius AI Studio integrates ASR to transcribe calls and NLU to extract intent and entities directly from speech transcripts as described [7] and [8]. Both extractive and abstractive summarization are applied to generate concise representations of call content, while keyword extraction highlights terms

most indicative of customer-perceived issues [9], [10]. These processes ensure that spoken input is translated into structured, explainable features suitable for downstream correlation and anomaly detection. The classification process is based on a small LLM that analyzes customer conversations and assigns them to well-defined macro-categories. All outputs are normalized, ensuring consistency across synonyms, phonetic variations, and service-related terminology. The main categories include: technical issues (covering all types of technical issues and service disruptions), contracts and negotiation (such as new contracts, renewals, cancellations, portability, and service adhesion), billing and tariffs (invoices, payments, and tariff plans), service (specific telecom services like Internet, Box, Cloud, etc.), assistant and customer support (requests for human or technical assistance), and other for non-standard cases. Additionally, the system extracts keywords from the customer's input, prioritizing service names, faults, and identifiers, while filtering out irrelevant words. This structured classification enables consistent, accurate categorization of customer interactions while maintaining a simplified taxonomy suitable for automation.

- **Contextual information**, such as device type/model, firmware/software version, Wi-Fi capabilities, subscribed services (e.g., product bundle, channel packages/premiums), access technology (e.g., Fiber to the Home (FTTH), Digital Subscriber Line (DSL), Hybrid Fiber-Coaxial (HFC), 4G, 5G), usage time, and geographical/segment attribution;
- **Network and service indicators** from NOSSIS Assurance in real time encompassing telemetry and assurance checks, such as trends of performance KPIs (e.g., latency, jitter, packet loss), associated alarm and health states (e.g., Loss of Signal (LOS), Loss of Frame (LOF), interface down), service flags (e.g., authentication/authorization failures, IPTV session errors), diagnostic flags, and provisioning status (e.g., profile mismatches, CPE/set-top box (STB) configuration drifts).

2. Windowed baseline estimation and drift detection: Rather than evaluating interactions individually, AURA aggregates them in a rolling observation window (4 hours, updated every 5 minutes). At each refresh, the observed window is compared with a learned baseline: if drift exceeds stability thresholds, the window is flagged as anomalous.

3. Itemset mining: The same data are then processed through the Apriori algorithm [5] to mine frequent itemsets, yielding interpretable co-occurrence patterns that characterize and explain anomalies by highlighting their most informative feature combinations. If no anomaly is detected, the empirical feature distribution is recalculated to update the baseline. This unsupervised, self-labeling method offers several advantages:

- discovery of previously unknown or evolving issues;
- continuous adaptation as more interactions are processed;
- reduced reliance on static inventories, taxonomies, or rigid troubleshooting flows.

4. Clustering and visualization: To support explainability and operational confidence, all detected anomalies are grouped into clusters of similar

cases. These clusters are displayed in custom dashboards for NOC and customer care teams, showing the following metrics:

- representative feature composition;
- most affected services or devices;
- diagnostic signals from the network side;
- common keywords and intents from user complaints.

Instead of isolated tickets or opaque alerts, operators see consolidated, interpretable patterns that help them prioritize responses, communicate clearly, and escalate with precision. All outputs surface in live dashboards that expose common failures in real time with technical explanations and geographic maps. Analysts can pivot to history to replay events, compare against baselines, and trace resolution.

Results

AURA was validated on both historical datasets and real field operations. The results demonstrate its strength in capturing weak signals from customer complaints as they reach the call center, transforming them into actionable insights that conventional assurance systems systematically overlook.

Early detection of hidden issues

One of AURA's main strengths is its ability to recognize emerging patterns in customer complaints, long before they accumulate enough volume to trigger conventional dashboards or alarms.

For example, during Android TV disruptions or clusters of "TV not working" calls, AURA aggregated signals from just a handful of early interactions and correlated them with network telemetry. This provided operators with early awareness of customer pain points at a stage where traditional Fault Management/Customer Feedback (FM/CF) systems still showed no anomaly.

This capability allows call center and NOC teams to act faster, adjusting IVR flows, prioritizing triage, and preparing proactive communication before the incident escalates.

Real cases discovered in the field

In live operations, AURA uncovered several issues that had gone undetected by standard monitoring:

- **Misdiagnosed STB replacements:** In a given locality, multiple customer tickets were closed by replacing set-top boxes after reports of intermittent video issues. With AURA, the concentration of similar symptoms across subscribers would have indicated that the problem was unlikely to

In the noise of early customer complaints, AURA identifies structure, revealing service issues before conventional systems register any anomaly

be isolated to customer equipment, avoiding unnecessary STB replacements and associated costs.

- **Low-Density Impact:** Failures in sparsely populated regions generated too few complaints to be visible globally. AURA aggregated these weak signals, surfacing degradations that dashboards missed.
- **Automatic Recording Failures:** A group of customers unable to access on-demand or recorded content were all using the same STB model, a correlation revealed only through clustering.
- **Channel-Specific Outage:** A regional outage of a single TV channel went unnoticed in aggregated KPIs. AURA detected and classified the issue in near real-time, enabling targeted communication to affected customers.

These cases illustrate how AURA converts isolated voices into clear, actionable clusters, turning noise into insight.

Operational gains

AURA is delivering significant improvements across support operations, including:

- Faster detection of multi-customer problems by identifying scattered complaints are part of the same incident;
- Reduced duplicate call volume, supported by proactive IVR/care messaging driven by AURA's real-time signals;
- IVR deflection, enabled by real-time signals and messaging templates that allow the IVR to inform affected callers and offer alternatives, such as self-service, notifications, and callback, when an incident is confirmed. This reduces the average speed of answer (ASA) and abandonment rates, freeing agents to focus on complex cases;
- Call center process optimization, shifting from ticket-by-ticket handling to resolution at the cluster level;
- Focus on true root causes, preventing wasted troubleshooting and unnecessary CPE replacements by surfacing the actual underlying fault.

Conclusion

AURA introduces a new-centric approach to telecom service assurance. By combining the voice of the customer with real-time network telemetry, it closes a structural gap: operators no longer need to choose between what customers say and what the network reports. Both perspectives are reconciled into explainable, actionable insights.

In recognition of its impact, AURA powered by NOSSIS Genius received the "Most Innovative AI Application to Enhance Customer Experience" award at FutureNet World 2025. This external validation highlights the uniqueness and maturity of the solution, reinforcing its potential as a strategic layer in telecom assurance.

AURA establishes a data-centric model where emerging issues become detectable before they exist at scale

Field validation further demonstrates that the approach is not only feasible but also transformative: beyond detecting anomalies, AURA helps operators anticipate impacts and root causes, adapt processes, and communicate with customers in a more transparent and proactive manner.

In practice, AURA enables:

- **Better customer experience**, enabled through faster resolution and proactive communication.
- **Lower operational costs** by reducing repeated diagnostics and avoidable equipment swaps, and by optimizing workforce allocation.
- **Scalable automation enablement**, as AURA supplies real-time signals and templates that IVR and care flows can use to self-adjust to the current service state.
- **Faster time-to-detect (TTD)**, achieved through windowed aggregation, drift detection against a learned baseline, and rapid surfacing of common-fault clusters.

Crucially, AURA is designed to fit into existing IVR systems, care workflows, and NOC dashboards, enhancing what operators already do while unlocking new levels of efficiency. Looking ahead, AURA positions itself as a strategic layer for digital telcos, one that transforms scattered interactions into knowledge, knowledge into action, and action into measurable improvements in both customer experience and operational performance.

Future work

AURA's journey is only beginning. Future development will focus on bold directions that can redefine how operators connect customer voice with network reality:

- **Customer-aware diagnostic memory**: Imagine a system that remembers every past issue a customer experienced, learns from the fixes applied, and uses that memory to deliver faster, more empathetic, and more efficient support the next time they reach out.
- **Expansion to mobile and IoT**: AURA will tackle anomalies in 4G/5G networks (dropped calls, handover failures, app degradations) and the highly diverse IoT ecosystem, where problems are device-specific, context-dependent, and often scattered over time.
- **Closed-loop self-healing**: Taking anomaly detection to the next level by connecting diagnostics directly to automated repair workflows. When AURA recognizes a recurring fault pattern, it won't just raise an alert; it will trigger the fix itself, where permitted, creating the foundation for true self-healing networks.
- **Model Context Protocol (MCP)**: Building a new interface layer where autonomous agents can access AURA's enriched datasets via MCP. This will enable AI-driven collaboration, scenario simulation, and even human-AI co-piloting for network assurance.

AURA will continuously grow and adapt to listen, anticipate, and heal networks as events unfold

These initiatives will take AURA from today's research-proven framework to tomorrow's operational intelligence engine — always-on, adaptive, and pushing the boundaries of what it means to listen to customers and heal networks in real-time.

References

- [1] T. G. Dietterich, "Machine learning for sequential data: A review," in *Structural, Syntactic, and Statistical Pattern Recognition*, T. Caelli, A. Amin, R. P. W. Duin, D. de Ridder, and M. Kamel, Eds., vol. 2396, *Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2002, pp. 15–30.
- [2] R. Grishman, "Information extraction: Techniques and challenges," in *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, M. T. Pazienza, Ed. Berlin, Germany: Springer, 1997, vol. 1299, *Lecture Notes in Computer Science*, pp. 10–27.
- [3] ITU-T, *Architectural framework for machine learning in future networks including IMT-2020*, Rec. ITU-T Y.3172, International Telecommunications Union (ITU-T), Geneva, Switzerland, Jun. 2019.
- [4] R. Filipe, "Client-side monitoring of distributed systems," Ph.D. dissertation, Univ. of Coimbra, Coimbra, Portugal, 2020.
- [5] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, Santiago, Chile, 1994, pp. 487–499.
- [6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.
- [7] G. Hinton, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [8] B. Liu and I. Lane, "Joint online spoken language understanding and language modeling with recurrent neural networks," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist. (NAACL)*, San Diego, CA, USA, 2016, pp. 22–30.
- [9] Y. L. et al., "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. 58th Annu. Meeting Assoc. Comput. Ling.*, Jul. 2020, pp. 7871–7880.
- [10] R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," in *Proc. Conf. Empir. Methods Natur. Lang. Process. (EMNLP)*, Barcelona, Spain, Jul. 2004, pp. 404–411.
- [11] OpenAPI Initiative, *API Design Guidelines*, OpenAPI Initiative, Accessed: May, 2025. [Online]. Available: <https://www.openapis.org>



04

Towards MCP Server on NOSSIS-Based Autonomous Operations

Abstract

The Model Context Protocol (MCP) servers are modular services that let Large Language Models (LLM) securely access tools and data through a standardized protocol. They connect to local and remote resources to retrieve or act on structured and unstructured data. MCP servers enrich prompts with dynamic context, execute external actions, and support multiple parallel connections for scalability. By standardizing and unifying the way applications provide context to LLM (including persistent memory and long-term user context), MCP servers simplify development and integration.

For NOSSIS and Altice Labs' OSS suite, MCP could drive more intelligent and autonomous network operations. It provides a unified, context-aware framework that enhances interpretation between models, systems, and data. This enables adaptive decision-making, automated workflows, and interoperability across complex environments. By supporting agentic AI, systems that plan and act autonomously, MCP helps evolve OSS from static, rule-based automation to adaptive, self-governing platforms, advancing toward scalable and intelligent autonomous operations.

Authors

Maria Manuel Castro, Altice Labs, maria-m-castro@alticelabs.com

Ricardo Ângelo Filipe, Altice Labs, ricardo-a-filipe@alticelabs.com

Carlos Guilherme Araújo, Altice Labs, carlos-guilherme-araujo@alticelabs.com

Luís Cortesão, Altice Labs, luis-m-cortesao@alticelabs.com

Keywords

MCP, Agentic AI, Autonomous Operations, AI Orchestration

DOI

10.57650/INNOVATION/2025.04

By minimizing repetitive adaptation work, MCP servers expand team capacity and accelerate value-driven development

Introduction

Over the past years, the way intelligent systems connect with external resources has evolved considerably. At first, integrations were built case by case, requiring specialized development to connect each model to an API, database, or tool. This approach, although functional, created silos, increased maintenance costs, and limited scalability.

The next stage brought the rise of standardized interfaces such as OpenAPI, which simplified access to services but still required significant technical expertise to make models work seamlessly across different environments.

Afterwards, Agentic AI introduced new possibilities for orchestration and automation, but also new challenges. While AI Agents could chain tasks and call tools dynamically, they often incurred high tolling in terms of complexity and integration between vendors. Many organizations began to experience vendor lock-in, as proprietary platforms offered powerful integrations but constrained flexibility, making it difficult to migrate or adapt solutions across different LLM [1].

It is in this context that the MCP emerges as a promising solution. MCP enables models such as LLM to interact with APIs, databases, and software tools in a truly standardized and interoperable manner.

For NOSSIS products, this shift is strategic. MCP servers not only streamline integration but also expand team capacity, as developers spend less time on repetitive adaptation tasks and more on value creation. In addition, the protocol opens the door for business professionals without deep technical expertise to leverage these connections, using domain knowledge to directly influence how systems are applied in practice. The rise of 5G, virtualized infrastructures, and real-time service expectations has pushed operators to seek platforms that enable automation, agility, and intelligent decision-making. The shift toward autonomous operations – where systems can sense, analyze, decide, and act without human intervention – is becoming essential to ensure scalability, resilience, and operational efficiency.

This article investigates how we can elevate traditional automation into adaptive, self-governing operations. The goal is to understand how MCP servers can become a key enabler of an autonomous OSS, unlocking new capabilities. We also present the practical results of their implementation over an OpenAPI interface, including the technical challenges faced and the lessons learned.

What are MCP servers?

Paraphrasing the creators of MCP: “Think of MCP like a USB-C port for AI applications. Just as USB-C provides a standardized way to connect electronic devices, MCP provides a standardized way to connect AI applications to external systems” [2].

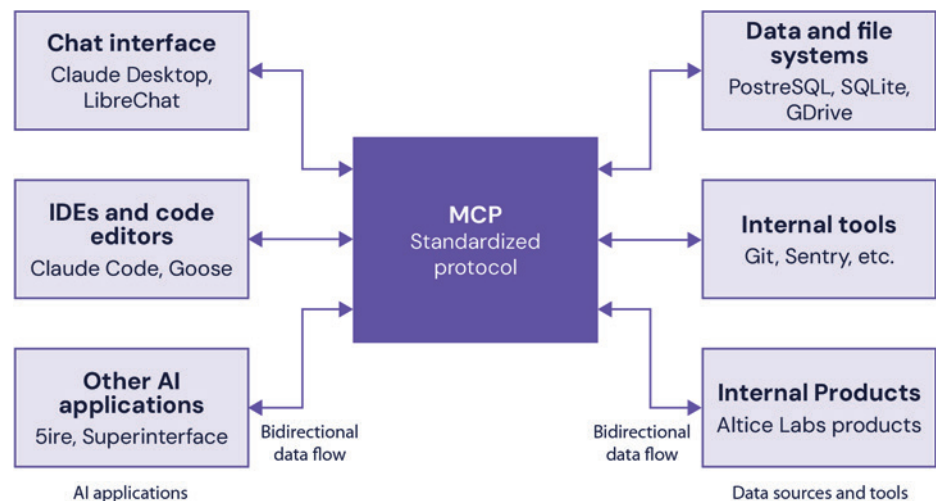
MCP standardizes how LLM agents access and utilize external resources. By providing a unified framework for communication with data and tools, it expands LLM’s knowledge and enables it to address complex, real-world challenges more effectively. In addition, developers no longer need to maintain multiple integrations across vendor-specific tools, simplifying development and reducing overhead.

The MCP has the potential to transform the industry by guiding how AI systems connect with external tools, data, and other products. Despite these advantages, its implementation presents a significant challenge. MCP illustrates the relevance of Conway's Law [3], often paraphrased as: "the structure of a software system mirrors the communication structure of the organization that builds it." Organizations with aligned structures can leverage MCP to develop AI systems that are more coherent, efficient, and maintainable, whereas misaligned structures may struggle to realize their full benefits. Ultimately, MCP positions the industry to create more interoperable and sustainable AI ecosystems.

How to implement MCP servers?

MCP goes with an MCP server/client logic. MCP servers can be used by any kind of client that supports the protocol, as illustrated in **Figure 1**. On the left, the MCP Clients with LLM capabilities invoke the MCP Server (represented by the blue shape). The bidirectional arrows are because of some protocol capabilities (i.e. sampling) that allow the MCP server to request completions to the LLM client. On the right, we have the MCP capabilities, which can go from direct access to databases or filesystem to internal products or modules.

Figure 1: MCP Protocol [2]



The core concept is that MCP servers can expose resources to AI agents, enabling seamless integration and enhancing the agents' capabilities.

The protocol allows the following features for the client:

- **Roots:** allows the client to expose filesystem areas to servers, defining accessible areas;
- **Sampling:** servers can request LLM completions through clients, ensuring client control over access or model use;
- **Elicitation:** Lets clients control user interactions and data sharing while enabling servers to collect needed information dynamically. Useful when you want to ensure that the MCP server has all the information to invoke a specific downstream resource.

Additionally, it also provides several capabilities from the server-side point of view:

- **Prompts:** instructions that guide model interactions;
- **Resources:** content that provides context to the model;
- **Tools:** functions that allow models to perform actions or retrieve information from the system.

The first three functionalities are client-side, while the remaining three are server-side. Although new capabilities have been introduced over time – most recently “Elicitation” in the latest version of the protocol – the most widely adopted functionality continues to be Tools. Nevertheless, the protocol’s overall functionalities are designed to enable future use cases that may not yet be feasible today.

MCP servers are invoked by clients, such as code editors or autonomous agentic AI, which operate with the autonomy and agency required to execute a designated task by leveraging MCP capabilities alongside other resources available to the client.

Additionally, since MCP are API capable of executing actions on a given system, implementing authentication is strongly recommended. Servers may employ standard authentication mechanisms, including OAuth 2.0 or bearer tokens.



MCP servers strategically reshape how organizations design, operate, and evolve their systems

Why use MCP servers?

The adoption of MCP servers is not merely a technical improvement; it is a strategic enabler for transforming how organizations design, operate, and evolve their systems. In increasingly complex environments, where agility, automation, and intelligence are essential, MCP servers offer a unifying layer that bridges the gap between AI capabilities and operational execution.

By introducing a standardized protocol for interaction between models and systems, MCP servers reduce fragmentation and vendor dependency, allowing organizations to build solutions that are modular, portable, and future-proof. This standardization is particularly relevant in ecosystems like NOSSIS, where multiple components must interoperate seamlessly across domains such as fault management, diagnostics, and orchestration.

Beyond technical efficiency, MCP servers empower teams. Developers benefit from reduced integration overhead and can focus on delivering value rather than maintaining custom connectors. At the same time, business professionals gain the ability to influence system behavior through natural language and domain knowledge, without requiring deep technical expertise.

This convergence of roles fosters collaboration and accelerates innovation.

In the context of autonomous operations, MCP servers play a foundational role. They enable systems to interpret context, make informed decisions, and execute actions dynamically – moving beyond static automation toward adaptive, self-governing platforms.

Ultimately, MCP servers are a catalyst for operational intelligence. They enable organizations to scale AI adoption responsibly, align technology with business goals, and unlock new levels of autonomy, resilience, and efficiency.

Integrating MCP servers into the NOSSIS ecosystem

The initial practical exploration focused on the areas of Problem Management and Test & Diagnostic, which are critical components of the Assurance layer within NOSSIS One. These areas were selected because they offer clear and measurable use cases for validating the MCP approach, and they align closely with the automation and proactive response goals of the NOSSIS Genius platform.

By transforming API calls into MCP tools, the system gains flexibility, reduces integration complexity, and enables adaptive, goal-oriented workflows.



Implementation methodology: MCP server over OpenAPI

The implementation methodology followed an iterative and exploratory approach, focusing on technical validation and value demonstration. The foundation of the implementation was the OpenAPI specifications from the Problem Management and Test & Diagnostic areas of NOSSIS One.

Tools and development environment:

- Visual Studio Code (VSCode) was used as the main Integrated Development Environment (IDE);
- The programming language chosen was Python, due to its flexibility and rich ecosystem;
- Integration with the agent was achieved via GitHub Copilot, configured as an MCP agent with access to the exposed tools;
- The model used was o4-mini, suitable for quick testing and contextual interactions.

MCP server structure:

- Two distinct MCP servers were created: one for Problem Management and another for Test & Diagnostic;
- Each server exposed a set of tools derived from the OpenAPI endpoints, such as `create_ttk`, `search_ttk`, `get_ttk`, `diagnose_gpon_network`, `diagnose_access_network`, `execute_netq_order`, `get_netq_order_status`, ...
- The tools were enriched with clear semantic descriptions, making it easier for the agent to understand their purpose.

Integration process:

1. **Selection of relevant endpoints:** Endpoints with the highest potential for automation and operational impact were chosen.
2. **Conversion into MCP tools:** Each endpoint was transformed into a tool with descriptive metadata.
3. **Unit testing:** Each tool was tested individually to validate its functionality and response.
4. **Composite testing:** Scenarios requiring the use of multiple tools in sequence were simulated.
5. **Limitation testing:** Cases where MCP usage was not ideal were explored to identify the boundaries of the approach.

Practical experience: MCP calls over OpenAPI

During the testing phase, the MCP agent (GitHub Copilot configured with the o4-mini model) was able to interact directly with the NOSSIS One APIs via the MCP server, executing real-time queries and actions. The tools exposed through the MCP server allowed the agent to retrieve and manage tickets (including

Table 1: Representative interactions of the MCP agent with NOSSIS One APIs

creation, status update, and data aggregation) and to conduct diagnostics across multiple services such as Gigabit Passive Optical Network (GPON) fiber, Internet Protocol Television (IPTV), Internet, and Voice over Internet Protocol (VoIP). The agent interpreted technical metrics, identified anomalies, and proposed service-specific corrective actions.

In **Table 1**, there are some representative examples of interactions:

Services	Prompt	Response
Querying Ticket Details	Give me details of the TTK TT-2025191-000172420	The agent returned a full set of attributes, including ID, type, priority, severity, origin, status (CANCELLED), assigned team, timestamps, and SLA metrics.
Status Check and Assignment	What is the current status of TTK TT-2025025-000343474, and which team is it assigned to?	Status was reported as PENDING, assigned to the Field Force team.
Ticket Creation via MCP Tool	Create a new TTK that reports the same as TT-2025191-000172420 in the state ACKNOWLEDGED.	A new ticket was successfully created with identical attributes, but with status set to ACKNOWLEDGED.
Status Update Attempt with Permission Handling	Change the status of the TTK TT-2025191-000172420 to In Progress	The agent attempted the update but received a 403 Forbidden error, correctly identifying a lack of permissions and suggesting administrative action.
Aggregated Queries	Could you provide the total number of TTKs per status under the Infrastructure category?	The agent returned a breakdown: <ul style="list-style-type: none"> • ACKNOWLEDGED: 33 • IN_PROGRESS: 4 • CANCELLED: 5 • CLOSED: 1
Listing Specific Tickets	List me the Cancelled TTKs IDs	A list of five ticket IDs was returned, all under the Infrastructure category.
Time-Series Analysis Attempt	What is the average of TTKs by day in the last year. By month.	The agent explained the limitations of the API (lack of aggregation support), and proposed alternative approaches using data export and external analysis tools.
Entity Inspection	What can you tell me about the entity 16015410?	The agent returned a complete overview of the entity, including access type (FIBER), location, commercial and technical profiles, active services (IPTV, HSI, VoIP), and transmission rates.
Full Service Diagnostics	Make a diagnostic over the services that this entity has.	The agent performed diagnostics on GPON fiber, IPTV, Internet, and VoIP. Issues such as excessive optical attenuation, NAT profile inconsistencies, and neighboring ONT failures were identified. Specific corrective actions were also suggested for each service.

These examples demonstrate the agent's ability to interpret prompts accurately, execute API calls via MCP tools, and handle errors gracefully. They also highlight the importance of semantic clarity in tool descriptions and the need for robust permission management.

Implementation insights: design considerations and best practices

The implementation of MCP servers over OpenAPI within the NOSSIS ecosystem offered valuable lessons that go beyond technical execution. These insights, drawn from real-world testing and agent interactions, are intended to guide future MCP server designers toward more effective, scalable, and agent-friendly solutions.



Design considerations for agentic interactions

Several design-level observations emerged during the project that are essential for enabling smooth and intelligent interactions between agents and MCP tools:

- **Semantic Precision in Tool Descriptions:** Tools must be described in a way that is both machine-readable and semantically meaningful. Vague or overly technical descriptions led to misinterpretation and incorrect tool selection by the agent.
- **Managing Context Load:** APIs with a large number of endpoints introduced latency and, in some cases, timeouts. Strategies such as tool grouping, dynamic exposure, or context filtering are essential to maintain agent responsiveness.
- **Bridging the Semantic Gap:** Agents operate in natural language, while APIs are inherently structured. Ensuring that tools can handle abstract queries or reformulations improves usability and reduces friction in agent interactions.
- **Handling Silent Failures:** In some cases, the agent failed without returning an error or result. This highlighted the need for better observability, error propagation, and fallback mechanisms.

These considerations reflect the shift from traditional API integration to dynamic, agent-driven orchestration – a paradigm where context, semantics, and adaptability are as important as code.

Lessons learned and recommendations

The experience also revealed a set of best practices that can serve as guiding principles for future MCP server projects:

- **Use clear and concise tool descriptions** to improve agent comprehension and reduce ambiguity;
- **Validate inputs and outputs rigorously** to prevent misuse and ensure data integrity;
- **Apply usage limits and sandboxing** to protect systems from overload and maintain operational safety;
- **Include human-in-the-loop mechanisms** for critical operations where full automation may not be appropriate;
- **Implement continuous monitoring and audit integration** to ensure transparency, traceability, and compliance.

Together, these insights form a practical foundation for designing and operating MCP servers in production environments. They highlight the importance of treating MCP servers not just as technical components, but as cognitive enablers – interfaces that must serve both human and machine reasoning.

Strategic business and technical positioning of NOSSIS in the MCP server context

By cutting integration friction, MCP servers accelerate innovation while strengthening service quality and system resilience

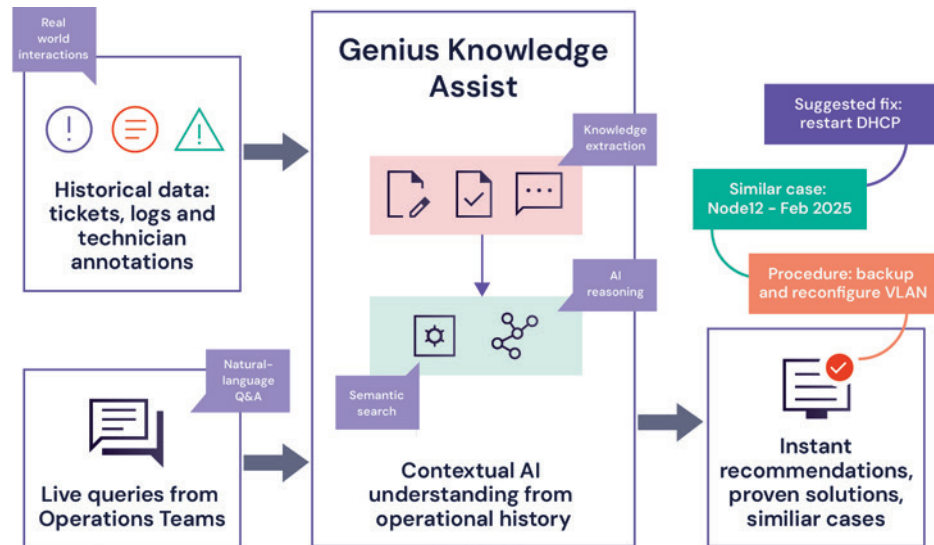
The adoption of MCP servers in the NOSSIS ecosystem represents a decisive step that is both technical and strategic. On the technical side, MCP provides a standardized interface that transforms fragmented API connections into a modular and interoperable layer. On the business side, MCP reduces integration overhead, accelerates time-to-market for new workflows, and increases service quality and resilience. Together, these benefits position NOSSIS not only as a future-proof OSS platform, but also as a pioneer of AI-native operations.

NOSSIS intends to evolve beyond static dashboards and manual workflows toward a pervasive model of contextual intelligence. MCP servers enable this transformation by exposing tools and data sources that can be orchestrated by AI agents across Assurance, Fulfillment, and Inventory. The next strategic step is to distribute prompt-based interfaces across the suite, embedding conversational assistants into each application. Instead of navigating multiple menus and clicks, users will interact with the system through natural language, asking questions and receiving contextual responses. These embedded assistants will guide operators through troubleshooting, provisioning, diagnostics, or reporting, combining business knowledge with technical execution in a seamless way.

Beyond embedded assistants, NOSSIS also envisions a global assistant capable of interacting with the entire ecosystem through a single conversational interface. In this model, the user does not need to know where a particular capability

resides, since the assistant dynamically orchestrates MCP tools across multiple domains. For example, an operator may ask for details about a specific Optical Line Terminal (OLT). The assistant retrieves the technical and topological information from the Inventory module and then correlates it with Assurance data to verify whether this OLT has experienced recent alarms or service degradation. Similarly, a user may request an overview of open trouble tickets related to that OLT or its connected subscribers. The assistant responds by aggregating ticket data, highlighting active incidents, and linking them with network and service diagnostics. What previously required multiple systems and workflows becomes a single contextual interaction, accessible through natural language.

Figure 2: Diagram of Intelligent Knowledge Assistance for Operations Teams



This approach is consistent with the principles of the TM Forum Open Digital Architecture (ODA) [4]. NOSSIS keeps all system boundaries exposed through TM Forum Open APIs and structures its modules as reusable components. MCP servers complement this by providing a standardized way for AI agents to access tools and data internally, which aligns with ODA goals of modularity, interoperability, and plug-and-play integration. Furthermore, the conversational and intent-driven paradigm enabled by embedded and global assistants resonates with the TM Forum Autonomous Networks vision, where interactions are guided by intents rather than manual commands.

This dual strategy creates value for all stakeholders. Technical teams benefit from a reduction in custom integration work, allowing them to focus on innovation. Business teams gain direct influence over system behavior through natural-language interactions, lowering the barrier between domain expertise and operational automation. For customers, the result is faster problem resolution, reduced OPEX, and improved customer experience. For NOSSIS, it strengthens its market positioning as one of the first OSS platforms to embrace MCP not only as a technical connector, but as a foundation for a new way of working where conversational AI becomes the default interface for operations.

By combining MCP standardization with embedded assistants, a global assistant, and prompt-driven interactions, NOSSIS moves closer to its long-term vision of autonomous, AI-native OSS platforms. This vision places the MCP

Through standardized MCP frameworks and AI-driven assistants, NOSSIS edges closer to a fully autonomous, intelligent OSS future

server not as an isolated component, but as the backbone of a strategy where intelligence is distributed, accessible, and aligned with both business goals and operational realities.

Conclusions

The implementation of MCP servers over OpenAPI interfaces within the NOSSIS ecosystem has proven to be a strategic and transformative step toward AI-based autonomous operations. By standardizing how AI agents interact with tools and data, MCP servers reduce integration complexity, promote modularity, and enable goal-oriented workflows. This approach not only enhances technical efficiency but also democratizes access to automation, allowing business professionals to directly influence system behavior through natural language.

The conducted tests demonstrated that agents configured with MCP access can perform diagnostics, manage trouble tickets, and interpret technical metrics accurately, showcasing the potential of intelligent orchestration in OSS environments. The lessons learned, from the importance of semantic clarity in tool descriptions to context management and silent failure handling, provide a solid foundation for future implementations.

NOSSIS's vision for an AI-native OSS platform, with embedded conversational assistants and a global assistant capable of interacting with the entire ecosystem, aligns with the principles of TM Forum ODA and the evolution toward Autonomous Networks. MCP is not just a technical connector, but the backbone of a new operational paradigm where intelligence is distributed, accessible, and aligned with business goals.

As next steps, the focus will be on expanding integration across other platform domains, exploring new use cases, and reinforcing security, auditing, and continuous monitoring mechanisms. With these developments, NOSSIS can further establish itself as a reference in autonomous, resilient, and user-centric operations.

References

- [1] S. Furão, R. Â. Filipe, J. M. Sousa, J. Pereira, and R. Almeida, "Towards Agentic AI: The next evolution in enterprise flows", *Innovaction*, no. 10, pp. 18–31. 2025.
- [2] Model Context Protocol. "Getting Started: Model Context Protocol (MCP)." Model Context Protocol. Accessed: Oct. 1, 2025. [Online]. Available: <https://modelcontextprotocol.io/docs/getting-started/intro>
- [3] M. E. Conway, "How do committees invent," *Datamation*, vol. 14, no. 4, pp. 28–31, 1968.
- [4] TM Forum. "About Open Digital Architecture (ODA)." TM Forum. Accessed: Oct. 1, 2025. [Online]. Available: <https://www.tmforum.org/oda/about/>



05

Agentic AI Meets Connected Home Ecosystem: A Case Study of MyButler

Abstract

This paper presents the design and application of Agentic AI technologies in the development of MyButler, a Virtual Intelligent Agent integrated into the Altice Labs Connected Home platform. MyButler continuously learns user routines, preferences, and contextual cues (such as time of day, calendar events, weather conditions, and human behaviors) to proactively manage smart home devices. It enhances daily living by autonomously adjusting lighting, blinds, environmental controls, enforcing security and assisted living, and optimizing energy/water efficiency based on user behavior and inferred intent, such as detecting moods through voice or interpreting calendar events.

MyButler supports multimodal interaction, including voice, text, and visual input, allowing users to engage naturally. It interprets commands like “What’s that light blinking?” by analyzing camera feeds or provides visual tutorials in response to queries like “Show me how to reset the router.” It recognizes implicit triggers such as “I left home” or “It’s cold” to take autonomous action.

A built-in scripting environment enables users to define or adjust routines using natural language, e.g., “Turn off all lights and set the alarm when I say, ‘I’m out,’” or “Change my morning routine to include the weather forecast.” MyButler aims to reduce utility bills through intelligent energy management, prevent appliance failures, and streamline repetitive tasks, creating a more efficient, personalized, and proactive smart home experience.

Authors

Paulo Chainho, Altice Labs, paulo-g-chainho@alticelabs.com

Ricardo Ângelo Filipe, Altice Labs, ricardo-a-filipe@alticelabs.com

Vitor Santos, Altice Labs, vsantos@alticelabs.com

Jorge Miguel Sousa, Altice Labs, Jorge-m-sousa@alticelabs.com

Keywords

AI, Multi-Agent Societies, MyButler, AI Studio, Connected Home

DOI

10.57650/INNOVATION/2025.05



Introduction

The evolution of Artificial Intelligence (AI) in daily life has moved rapidly from novelty to necessity. What began with simple voice chat assistants that could understand intents and answer questions has grown into systems that seamlessly integrate into our routines. AI is no longer just a tool we call upon; it is increasingly a partner that learns from our habits, adapts to our preferences, and reduces friction in everyday tasks. This shift is transforming homes from collections of devices into intelligent environments that actively support well-being, efficiency, and comfort [1].

As technology matures, customers' demands are shifting from reactive assistants to proactive, agent-based ecosystems. Instead of waiting for commands, customers want anticipation: dimming the lights before bedtime, adjusting temperature based on occupancy, securing the home when it detects everyone has left, or even reminding me to go back if I forgot my wallet.

MyButler differentiates itself in this market by fully embracing this vision, building on the existing strategy of Altice Labs in the AI area. The reuse of technological components, infrastructures, and accumulated knowledge makes it possible to accelerate development while ensuring robust and interoperable solutions [2]. MyButler learns not just from individual device usage but from the whole household's routines. By positioning itself as a trusted, proactive partner rather than just another app or device, MyButler creates a smarter, more human-centric home that stands apart in the smart home marketplace.



In this paper, we will explore how agentic AI has the potential to transform the connected home ecosystem. By having context-aware agents, we can redefine the role of smart home technology. The promise of agentic AI is a home that not only responds, but understands, and in doing so, reshapes the very way we experience living.

From the history of AI to multi-agent societies

The history of AI is marked by evolving paradigms that reflect different conceptions of intelligence and computation. Early developments in the 1950s and 1960s were dominated by symbolic AI, which assumed that reasoning could be represented through formal rules and logic. Expert systems of the 1970s and 1980s built upon this paradigm, encoding human knowledge in rule-based engines to support decision-making in domains such as medicine or engineering. While effective in narrow contexts, their brittleness and lack of adaptability limited broader applications.

In parallel, the rise of connectionist approaches, such as artificial neural networks, reintroduced learning and pattern recognition as central capabilities of intelligent systems. These foundations eventually enabled the breakthroughs of Machine Learning (ML) and deep learning in the 21st century, which dominate much of today's AI landscape. Importantly, throughout these shifts, researchers explored not only intelligence in isolation, but also coordination among multiple intelligent entities.

The concept of Multi-Agent Societies (MAS) emerged to address distributed problem solving, autonomy, and collaboration [3]. Rather than a single monolithic AI, MAS research proposed societies of agents, autonomous, proactive, and socially capable entities that are able to negotiate, share knowledge, and act in dynamic environments. This paradigm resonates with early visions of distributed computing and telecommunications architectures. For example, the Mirror Agent Society model, proposed in the 1990s within the Telecommunications Information Networking Architecture (TINA) framework, highlighted how distributed agents could mirror stakeholders, negotiate on their behalf, and enforce business rules across heterogeneous systems. Such efforts anticipated many principles later applied in digital ecosystems, from negotiation protocols to ontology-based knowledge sharing.

Historically, MAS approaches found application in domains where distributed control was essential. Robotics research explored cooperative robots ("multi-robot systems"), while transportation leveraged agents for traffic management and resource allocation. Early smart home prototypes, although limited by available hardware, also experimented with agents for energy management

and task automation. These pioneering use cases revealed both the opportunities and complexities of coordinating decentralized intelligence.

Today, the relevance of multi-agent societies has grown substantially with the proliferation of Internet of Things (IoT) ecosystems. A modern connected home integrates dozens of devices – sensors, appliances, energy systems, and digital assistants – that must operate in concert. Achieving this requires precisely the type of negotiation, coordination, and distributed reasoning envisioned decades ago in MAS research. Contemporary platforms, such as MyButler, embody this evolution by orchestrating multimodal interaction and autonomous routines across diverse devices. The continuity from symbolic AI to agent-based systems and, more recently, to generative and agentic AI underscores that societies of agents are not a new vision but a longstanding trajectory toward collective intelligence.

MyButler in the Connected Home ecosystem

MyButler embodies the transition from traditional personal assistants toward a household-level agent, continuously learning from contextual cues such as time, occupancy, schedule, or weather. This continuous adaptation goes beyond simple automation: it reflects lifestyle patterns, anticipating needs, and mediating potentially conflicting user demands.

From personal assistant to household agent

Unlike single-user assistants, MyButler represents the entire household. It manages preferences across family members and guests, arbitrating conflicts (e.g., balancing temperature preferences or media choices) while optimizing resource consumption. This positions MyButler as the digital steward of the home, ensuring both harmony and efficiency.

MyButler as a broker of specialized agents

MyButler is not a monolithic system, but rather an orchestrator of domain-specific agents running within the AI Engine, as seen in **Figure 1**. Each sub-agent is responsible for a well-defined scope, while MyButler coordinates their interplay:

- **MyConnect** – Connectivity & Technology: manages Wi-Fi, IoT interoperability, diagnostics, cybersecurity, and the lifecycle of smart devices.
- **MyGuard** – Security: integrates alarms, surveillance, and access control; coordinates emergency response.
- **MySustain** – Sustainability: optimizes energy, water, and gas; manages renewables and negotiates with utilities.
- **MyFun** – Entertainment: organizes multi-room experiences, content recommendations, and family media.

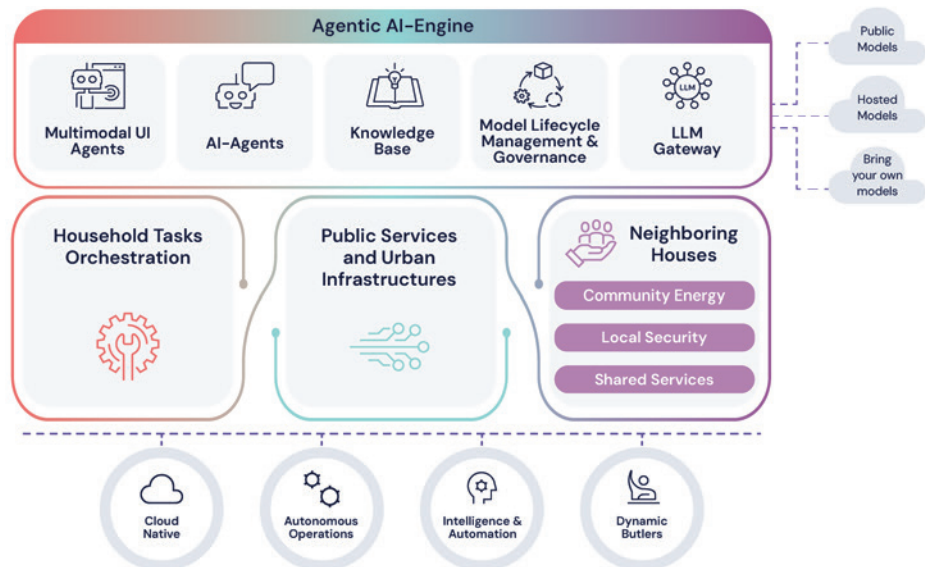
- **MyCare** – Health & Wellbeing: provides assisted living support, monitors health, and supports elderly care.
- **MyChef** – Culinary Expert: plans meals, manages groceries, and supports multimodal cooking guidance.
- **HouseKeeper** – Home Organization: coordinates cleaning, inventory, and household readiness.

This multi-agent orchestration model reflects established approaches in distributed systems and business-oriented agent societies, enabling modularity, resilience, and flexibility.

The household as a node in a larger society of agents

In the long term, each household can act as a digital node in a broader agent society, connected to municipalities and urban infrastructures [5], represented externally by MyButler. Such a framework would connect homes with municipalities (for public services or emergencies), with utility providers (for tariffs and resource optimization), and with neighboring houses (for community energy sharing, local security, or shared services). This vision expands MyButler from a domestic orchestrator into a participant in smart city ecosystems, where households collectively contribute to efficiency, sustainability, and resilience.

Figure 1: MyButtler as orchestrator of internal agents and interacting with external Society agents

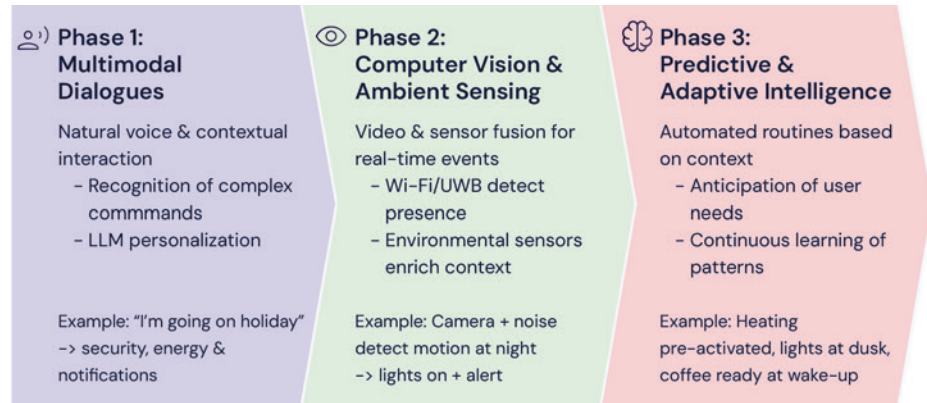


Differentiating use cases: a roadmap perspective

The connected home does not become intelligent in a single leap; it evolves through a series of milestones that progressively expand how artificial intelligence interacts with, perceives, and anticipates human behavior. MyButler differentiates itself by following a clear roadmap where AI acts as the engine of transformation,

progressively introducing new layers of capability. Each stage represents a new level of maturity in the household’s intelligence: from understanding natural dialogue, to perceiving ambient context, to learning and adapting proactively. Together, these stages form a clear trajectory – from interaction to anticipation – illustrating how MyButler integrates AI evolution into everyday living (see **Figure 2**).

Figure 2: MyButler Use cases evolution



Multimodal dialogues

Beyond commands and menus, interaction evolves into conversations powered by context and intent

The first stage of MyButler’s roadmap is about transforming interaction itself. Rather than relying on rigid commands or predefined menus, MyButler enables natural, multimodal dialogues. Residents can use voice, text, or even visual cues (like pointing a camera at a blinking device) to engage in conversations that feel intuitive and human.

Powered by Large Language Models (LLM), MyButler goes beyond simple command recognition: it understands context, infers intent, and adapts its responses to each household. Instead of being reactive, it orchestrates complex sequences of actions from a single request.

Examples include:

- A resident says, "Good evening, I'm watching a movie" → MyButler dims the lights, lowers blinds, silences notifications, and configures the home theater setup.
- A child points to a blinking purifier and asks, "What is that?" → MyButler explains the device’s function and shows indoor air quality levels.
- Saying "I'm going on holiday" → triggers a full routine of presence simulation, energy optimization, door locking, and tailored notification rules.

Through these dialogues, MyButler evolves into a household companion that interprets language in all its forms (spoken, written, or visual) and translates it into proactive orchestration.

Computer vision and ambient sensing

The second stage strengthens contextual awareness, where the home becomes capable of perceiving its environment and reacting intelligently. By combining video analysis, Wi-Fi sensing, Ultra-Wideband (UWB) localization, and

*No longer reactive,
intelligent systems
now anticipate, adapt,
and evolve seamlessly
with your lifestyle*

environmental sensors, MyButler builds a real-time understanding of household dynamics without requiring explicit user input.

Examples include:

- An exterior camera detects unexpected motion late at night, in low light, combined with unusual noise → MyButler activates lights, records video, and sends an alert.
- Motion and acoustic sensors detect a fall in the hallway → a caregiver is automatically notified in real time, ensuring rapid intervention.
- Rising humidity in the basement triggers early flood prevention → MyButler shuts off the main water valve autonomously.

Beyond perception, this stage also empowers residents through a built-in scripting environment. Instead of coding rules, users can create or adapt routines in natural language:

- “Turn off all lights and set the alarm when I say, ‘I’m out.’”
- “Change my morning routine to include the weather forecast.”

This dual capability – autonomous sensing and user-driven customization – positions MyButler as both an intelligent observer and a flexible orchestrator of the connected home. By interpreting ambient signals such as temperature, luminosity, and occupancy, it not only enhances security but also adapts comfort and energy usage in proactive, human-centric ways. Stage 2 marks the transition from passive detection to adaptive environmental intelligence, a key milestone on the roadmap.

Predictive & adaptive intelligence

The third stage represents the leap from reaction to anticipation. At this level, MyButler leverages predictive models and continuous machine learning to adapt routines automatically and to foresee user needs before they are expressed.

Instead of waiting for commands, MyButler identifies patterns of behavior and context: what time family members arrive, when lights are usually needed, or which appliances show signs of wear. Over time, it builds a living model of the household that evolves seamlessly with lifestyle changes.

Examples include:

- On a hot summer afternoon, MyButler lowers blinds on the sunny side of the house to reduce cooling costs without being asked;
- Before the family arrives home on a rainy evening, it preheats the living room, adjusts the lighting to a cozy ambiance, and queues up a playlist that matches their usual mood;
- It detects that the dishwasher is drawing unusual power and proactively suggests maintenance before a breakdown occurs.

In this phase, MyButler becomes a self-adapting agent, shaping household routines around efficiency, comfort, and resilience. By turning data into foresight, it closes the loop of the roadmap: from intuitive interaction to intelligent perception, to predictive anticipation.

From interaction to anticipation

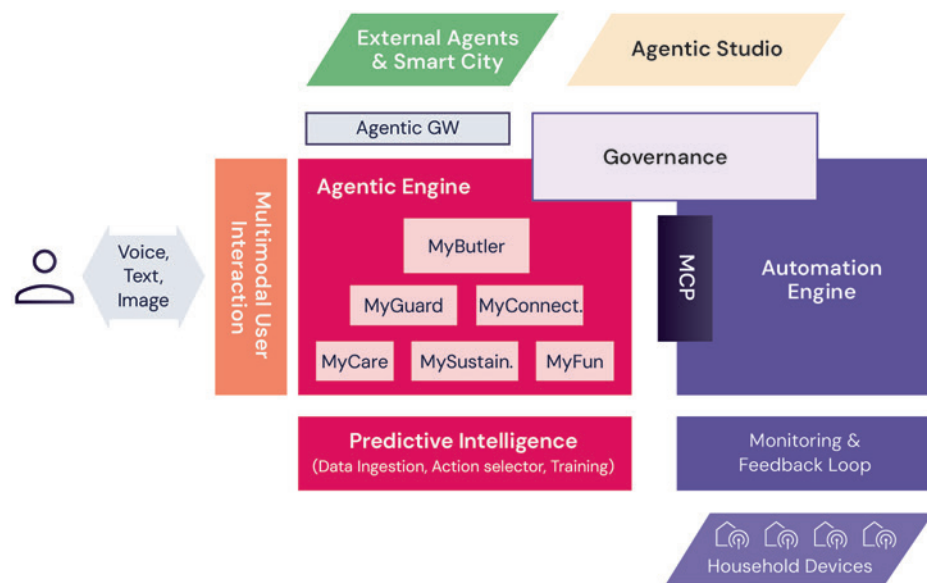
This evolutionary roadmap – from multimodal dialogues to ambient sensing to predictive learning – demonstrates how MyButler differentiates itself in the connected home landscape. Each stage adds a layer of value: first making communication natural, then embedding intelligence into the environment, and finally anticipating needs before they arise.

Through this journey, MyButler moves from being a useful assistant to becoming a trusted household agent, orchestrating convenience, efficiency, and safety in ways that feel effortless.

Architecture and technology

The architecture of MyButler integrates agentic AI, multimodal interaction, and predictive intelligence into a coherent technological framework designed for scalability, adaptability, and trust. As shown in **Figure 3**, the system is composed of interdependent layers that collectively enable perception, reasoning, and proactive orchestration within the connected home ecosystem.

Figure 3: MyButler Architecture



At the center lies the **Agentic Engine**, powered by BotSchool, Altice Labs' conversational AI platform. Here, the main orchestrator, MyButler, coordinates domain-specific agents such as MyGuard (security), MyCare (wellbeing), and MySustain (sustainability). Each agent operates autonomously using a specific agentic framework, while MyButler ensures the handover through shared ontologies and context models. These mechanisms enable consistent semantic exchanges, negotiation, and explainable dialogues across the MAS ecosystem while remaining compatible with cloud and edge environments.

Interaction between users and the system occurs through the Multimodal Interface, which supports voice, text, and visual inputs. Voice recognition and

synthesis rely on neural models such as Whisper or some Text-to-Speech (TTS) model, while natural language understanding is supported by LLM like GPT or open-source alternatives (LLaMA, Mistral). Visual inputs (e.g., identifying devices or interpreting gestures) leverage lightweight vision transformers or multimodal fusion models. The multimodal interface bridges human communication with agentic reasoning, ensuring that intents expressed in natural language are converted into structured agent-level actions. This interface layer allows for natural, conversational interaction while maintaining local privacy by running partial inference at the private cloud.

Beneath this orchestration layer, the **Predictive Intelligence Module** drives the shift from reactive automation to adaptive intelligence. It not only integrates core machine learning capabilities – such as data ingestion pipelines to collect sensor and contextual data, model training, and serving frameworks – to predict occupancy, energy usage, and daily routines, but also an Action Selector that applies rule-based logic and reinforcement learning to choose the most effective proactive actions.

The **Monitoring & Feedback Loop** complements this intelligence layer by capturing performance metrics and user overrides. It provides feedback, closing the loop of continuous learning.

At the foundation, **Knowledge Bases** and **LLM** collaborate to ground reasoning. Long-term memory stores maintain semantic consistency between agents, while LLM provide natural language reasoning, summarization, and explanation capabilities. Together, they form the cognitive layer that bridges structured knowledge with conversational intelligence and interoperable agent communication.

The **Automation Engine**, interfaced through the **Model Context Protocol (MCP)**, executes validated actions over the Connected Home IoT infrastructures using standard protocols such as Matter, Zigbee, and Z-Wave, ensuring interoperability across vendor ecosystems and compatibility with existing home platforms (Alexa, Google Home, or Apple HomeKit).

Externally, the **Agentic Gateway** allows seamless integration with Smart City and service ecosystems, enabling interaction with energy providers, health platforms, and municipal infrastructures. This federated layer extends MyButler far beyond the home, transforming it into a vital node within a wider society of agents.

Overarching all layers, the **Governance framework** enforces comprehensive lifecycle management, policy compliance, and transparency. It combines an AI policy engine (e.g., guardrails) with an audit and explainability subsystem that traces every single decision, from model input to executed action. Governance ensures secure data flows, ethical constraints, and adherence to regulations such as the General Data Protection Regulation (GDPR) and the EU AI Act, which guarantees consistent accountability and robust oversight

Through this layered architecture, MyButler evolves into an agentic, predictive, and trustworthy digital steward that is capable of understanding multimodal cues, learning from daily life, anticipating needs, and acting responsibly within both the household and the broader connected ecosystem.

Evolving into an agentic digital steward, MyButler understands multimodal cues, adapts routines, and anticipates tomorrow's needs



Benchmarking and market comparison

The evolution of smart home platforms has been marked by the search for more natural, efficient, and integrated human–technology interaction. Today’s market is shaped by large consumer ecosystems, specialized premium solutions, and open–source communities. Examining these approaches highlights the distinctive value proposition of MyButler.

Mainstream voice assistants: Google Home, Alexa, Siri [5], [6], [7]

These platforms dominate the mass market, excelling as reactive, voice–first assistants. Their strengths lie in executing simple commands, such as “turn on the lights” or “play music”, within a polished user experience. Yet they remain constrained by limited contextual reasoning and their dependence on closed ecosystems, where interoperability is possible but often optimized around proprietary devices and services. This strategy ensures consistency at scale but sacrifices flexibility, personalization, and true autonomy.

Josh.ai [8]

Josh.ai provides a premium alternative, prioritizing advanced natural voice control, privacy, and local data processing. Its ability to understand nuanced voice

commands has positioned it as a preferred solution for high-end smart home installations. However, Josh.ai's model is still primarily command-driven, with limited proactivity and no broader orchestration role within a heterogeneous ecosystem. Its value lies in the quality of interaction, but it does not venture into the domain of adaptive, multi-modal, or multi-agent intelligence.

Home Assistant, IFTTT, and ChatGPT integrations [9], [10]

At the other extreme, open-source and modular solutions such as Home Assistant, often extended with ChatGPT or IFTTT, offer unparalleled flexibility. They allow technically skilled users to script complex automations and integrate a vast array of devices and services. The trade-off is a fragmented, highly technical experience: configuration and maintenance require effort, and usability remains a barrier for mainstream adoption. These solutions illustrate the possibilities of openness and customization, but they fall short of delivering a seamless, user-friendly household agent.

MyButler: market differentiation

Against this backdrop, MyButler positions itself as a next-generation household agent, not confined to voice reactivity or technical scripting, but designed as an orchestrator of intelligence. Its differentiation emerges on several dimensions:

- **Context-driven proactivity:** MyButler interprets implicit triggers, such as "I left home" or "It's cold", to act autonomously, anticipating needs rather than waiting for explicit commands.
- **Multimodal interaction:** Voice, text, visual, and gestural inputs are supported, ensuring interaction that is natural, contextual, and not limited to speech.
- **Deep ecosystem integration:** MyButler connects with the broader Connected Home environment, transcending brand silos and coordinating across protocols and devices.
- **Personalized continuous learning:** It adapts dynamically to routines, schedules, and preferences, evolving with the household over time, and allows a continuous feedback that enriches the agent and improves the customer experience.
- **Multi-agent orchestration:** Unlike competitors, MyButler is designed as the orchestrator of a society of specialized agents. Sub-agents (e.g., MyConnect for connectivity, energy, or security agents) can operate autonomously within their domains, while MyButler coordinates them to deliver consistent, household-level intelligence.
- **Openness to external agents:** Looking ahead, MyButler's architecture supports interaction with external agentic ecosystems – from energy providers to health services –, enabling cooperation across domains and extending the smart home's intelligence into the wider digital society.

This shift – from a reactive assistant to a context-aware, multi-agent orchestrator – defines MyButler’s competitive advantage (see **Table 1**). While other solutions excel in single dimensions (voice quality, technical flexibility, or ecosystem control), none combine proactivity, multimodality, orchestration, and openness in the way MyButler does. It represents the transition from smart devices controlled by commands to a true household agent society, where intelligence is distributed, cooperative, and continuously evolving.

Table 1: MyButler Compared with other products

Dimension	Google Home / Alexa / Siri	Josh.ai	Home Assistant + IFTTT / ChatGPT	MyButler
Interaction Mode	Voice-first, reactive	Advanced voice control	Highly configurable (scripts, APIs)	Multimodal (voice, text, vision, gestures)
Contextual Proactivity	Limited (explicit commands)	Low (mainly reactive)	Possible but manual/technical	High: anticipate needs, implicit triggers
Ecosystem Model	Closed, brand-centric	Premium, installation-focused	Open but fragmented	Deep integration across Connected Home
Personalization	Basic routines	Some adaptation	Manual configuration	Continuous learning, lifestyle adaptation
Mass-Market Readiness	Mass market, simple UX	Niche, high-end	Tech-savvy user base	Mainstream-ready, user-friendly orchestration
Agentic Architecture	Single assistant	Voice-driven agent	Independent scripts and flows	Multi-agent society, orchestrated by MyButler
Future Openness	Limited to brand ecosystem	Closed premium	Open source, DIY integrations	Open to external agents and services

Benefits and risks

The deployment of intelligent agents in the connected home introduces a clear duality of tangible benefits and critical risks. MyButler, as a multi-agent orchestrator, exemplifies both the opportunities and the challenges of this new paradigm.

Benefits

Smart home intelligence offers measurable improvements in everyday life.

- **Comfort & Convenience:** Automating repetitive routines – from adjusting lighting to preparing morning setups – creates a seamless and personalized living experience. Multimodal interaction (voice, text, visual) further reduces friction, allowing users to control their home naturally and intuitively.

- **Energy Efficiency & Sustainability:** By learning household patterns and optimizing device usage against tariff schedules and weather forecasts, MyButler contributes directly to cost reduction and environmental sustainability.
- **Predictive Maintenance:** Early detection of anomalies in appliances prevents costly breakdowns, extending the life of household investments while reducing unexpected disruptions.
- **Enhanced Security & Assisted Living:** Context-aware monitoring increases safety, from detecting unusual presence at night to supporting elderly residents with proactive alerts and guidance.

Risks

Yet, these benefits come with risks that must be addressed systematically.

- **Privacy & Data Security:** Multi-agent systems rely on the continuous collection of personal and contextual data. Whether stored locally in the Connected Home infrastructure or exchanged through orchestration platforms, this information is highly sensitive. The use of LLM introduces further concerns around conversation history and inference leakage.
- **Unexpected Behavior in Simple Tasks:** Unlike deterministic, rule-based systems, LLM are probabilistic. Even straightforward commands like “Turn off all lights” may lead to clarification requests or unintended actions, undermining reliability in mission-critical tasks.
- **Over-Dependence on Automation:** Excessive reliance on autonomous routines may erode user skills, increase frustration when systems fail, or reduce willingness to act manually.
- **Interoperability Gaps:** Despite progress in standards, differences between device protocols and vendor ecosystems can cause fragmented experiences, particularly when introducing new devices.
- **AI Biases & False Detections:** As with any AI system, biased training data or imperfect sensor fusion may generate false positives / negatives (e.g., misclassifying benign movement as a security threat).

Mitigation Strategies

MyButler addresses these risks through a layered approach:

- **Transparency & Explainability:** Providing users with understandable rationales for decisions builds trust.
- **User Override & Control:** Manual intervention options ensure the human remains in charge, particularly in safety-critical scenarios.
- **Security-by-Design & Compliance:** Adherence to GDPR and equivalent privacy frameworks is mandatory. Explicit consent management, data minimization, and secure persistence mechanisms are integrated from the outset [11].

- **Safety Nets:** MyButler's is complemented with guardrails that prevent unsafe automation and allow rapid rollback of undesired behaviors.

Conclusion

MyButler stands as a living example of agentic AI applied to the home, showcasing how autonomous, context-aware agents can orchestrate daily life in ways that extend well beyond reactive automation. By continuously learning from user routines, preferences, and contextual signals, it transforms the household into an adaptive ecosystem rather than a collection of isolated devices.

The tangible benefits of this approach are already visible. Efficiency emerges through intelligent energy management and predictive maintenance, reducing costs and resource waste. Convenience is achieved by eliminating repetitive tasks, enabling seamless multimodal interaction, and anticipating user needs. Safety and assisted living capabilities provide reassurance, from proactive security measures to support for elderly residents. Sustainability becomes embedded in daily routines, as the system optimizes energy and water consumption in response to environmental conditions.

Beyond individual households, MyButler's vision extends much further. As a broker of specialized agents, it exemplifies how local intelligence can federate into broader networks. The same orchestration mechanisms that align appliances, sensors, and services in the home can scale to neighborhoods, smart communities, and even smart cities, where multiple agentic ecosystems collaborate to share resources, coordinate infrastructures, and enhance collective well-being.

In this sense, MyButler is not only a technical achievement but also a pathfinder for the future of digital society. It demonstrates how agent-based systems can balance autonomy and coordination, individual benefit and collective good. Home is the starting point; the horizon lies in interconnected, collaborative agent societies that will define the next generation of sustainable, human-centered living environments.

The journey begins at home and extends to agent societies that will shape tomorrow's sustainable, human-centered world

References

- [1] P. Chainho and F. C. Pinto, "Connected life: the path to a smarter ecosystem," *InnovAction*, no. 7, pp. 94–105, 2022.
- [2] M. M. Castro, R. Â. Filipe, C. G. Araújo, and L. Cortesão, "Towards the MCP server on NOSSIS-based Autonomous Operations," *InnovAction*, no. 10, pp. 44–55, 2025.
- [3] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice," *The Knowledge Engineering Review*, vol. 10, pp. 115–152, 1995, doi: 10.1017/S0269888900008122.
- [4] Lee, L. V. do Nascimento and J. P. M. de Oliveira, "A multi-agent architecture for context sources integration in smart cities," *Future Gener. Comput. Syst.*, vol. 172, Art. no. 107862, Nov. 2025, doi: 10.1016/j.future.2025.107862.

- [5] Google Home Developer Center. "Google Home Developer Center." Google Home Developer Center. Accessed: Sep. 18, 2025. [Online]. Available: <https://developers.home.google.com>
- [6] Amazon Developer. "Alexa Developer Documentation – Home." *Amazon Developer Documentation*. Accessed: Sep. 18, 2025. [Online]. Available: <https://developer.amazon.com/en-US/docs/alexa/documentation-home.html>
- [7] Apple Developer. "HomeKit." Apple Developer. Accessed: Sep. 18, 2025. [Online]. Available: <https://developer.apple.com/design/human-interface-guidelines/homekit>
- [8] Josh.ai. "Control with Voice." Josh.ai. Accessed: Sep. 19, 2025. [Online.] Available: <https://josh.ai/control-voice/>
- [9] Home Assistant. "Home Assistant." Home Assistant. Accessed: Sep. 19, 2025. [Online.] Available: <https://www.home-assistant.io>
- [10] OpenAI. "Hello GPT-4o." OpenAI.com. Accessed: Sep. 19, 2025. [Online.] Available: <https://openai.com/index/hello-gpt-4o>
- [11] European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)," *Official Journal of the European Union*, L 202, 12 Jul. 2024. [Online]. Available: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>



06

Autonomous Network Design: Leveraging AI in Reducing CAPEX and Planning Time

Abstract

The use of Artificial Intelligence (AI) is becoming increasingly widespread, providing significant advantages in developing robust, efficient, and innovative solutions across multiple sectors of the global market. AI demonstrates strong potential in overcoming challenges related to time, resources, and constraints typically found in human-designed network topologies. Notably, AI can autonomously design entire networks from the ground up, offering substantial benefits to organizations. This autonomous approach can lead to major improvements by reducing capital expenditure (CAPEX) through optimized design and lowering operational expenditure (OPEX) by minimizing the time required for network planning and deployment.

This article presents an AI-based framework for autonomous network planning using Genetic Algorithms to create complete GPON topologies. From customer building locations, it automatically defines optimal central office sites, cell boundaries, equipment, and cable dimensions, achieving coherent, cost-efficient designs that lower CAPEX and OPEX.

The framework combines Machine Learning clustering and graph partitioning to improve automation, scalability, and efficiency in complex network design and deployment. Tests showed that clustering supported effective deployment management, while graph partitioning efficiently divided large networks into manageable sub-networks. While these methods demonstrate promising outcomes, further refinement may enhance the results across diverse, large-scale, and real-world network scenarios.

Authors

Xavier Silva, Altice Labs, xavier-s-silva@alticelabs.com

Manuel Magalhães, Altice Labs, manuel-f-magalhaes@alticelabs.com

João Guerrinha, Altice Labs, joao-t-guerrinha@alticelabs.com

Luís Cortesão, Altice Labs, luis-m-cortesao@alticelabs.com

Pedro Antero Carvalhido, Altice Labs, pedro-a-carvalhido@alticelabs.com

Keywords

Autonomous Network Design, AI, FTTH, /GPON, Network Partitioning

DOI

10.57650/INNOVATION/2025.06

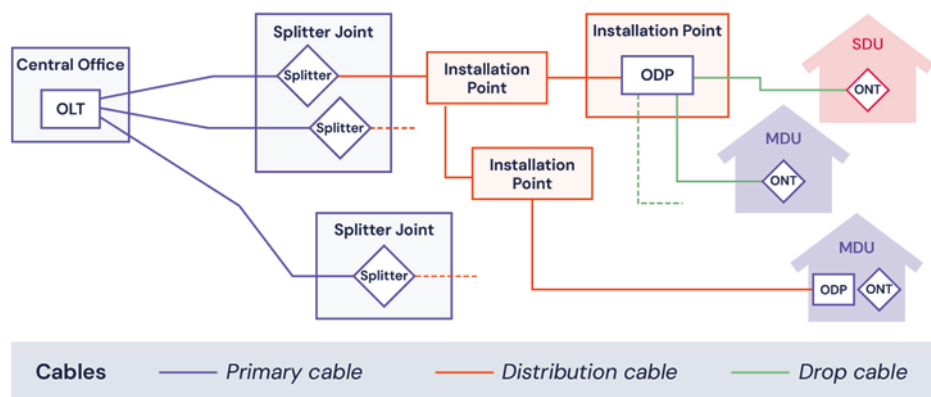
Introduction

The rapid expansion of fiber-to-the-home (FTTH) GPON deployments has significantly increased the complexity of telecommunication network design. Traditional manual engineering processes, which typically require up to 1.5 days per design, cannot scale efficiently to meet the accelerating demands for deployment [1]. Autonomous network design emerges as both a technical and economic necessity, leveraging advanced optimization techniques to streamline the planning of fiber access networks.

Understanding the deployment context is fundamental to design strategy. Greenfield deployments involve planning networks without pre-existing infrastructure, allowing optimization to focus purely on efficiency and cost minimization. By contrast, brownfield scenarios require careful integration with existing assets, navigating deployed infrastructure, and established routing patterns while still pursuing cost optimization [2].

Access networks follow a two-tier hierarchical structure. The feeder (primary) network encompasses the Central Office and its connections to Splitter Joints, branching nodes that serve as starting points for distribution cells. High-capacity primary cables connect the Central Office with Splitter Joints, covering longer distances and carrying aggregated traffic. The distribution (secondary) network extends from Splitter Joints to customer premises via Optical Distribution Points (ODP), serving Multi-Dwelling Units (MDU) or Single-Dwelling Units (SDU), with drop cables completing the final connection to homes. ODP may be located externally or inside MDU, depending on customer density. This structure is illustrated in **Figure 1**.

Figure 1:
Representation of a
GPON FTTH network



Optical splitters can be placed at multiple levels, including Central Office, Splitter Joints, or ODP, thereby multiplying a single fiber into several to optimize utilization and reduce costs. However, each splitting stage introduces insertion losses requiring careful power budget management. Physical routing also impacts design – for instance, buried routes bring the need to manage duct occupancy and respect physical constraints.

Network design encompasses interdependent criteria for balancing technical feasibility and cost-effectiveness. Core decisions include component placement, splitting ratios, and cable routing. These choices directly impact

deployment costs, operational efficiency, and the network's long-term scalability. Furthermore, the design must respect optical constraints, including power budget limitations and latency thresholds, while simultaneously optimizing resource utilization and construction efficiency.

In brownfield scenarios, coordinating physical work – such as trenching, pole installation, pavement repairs, and access to maintenance holes – presents critical challenges. Clustering network deployments addresses this by grouping construction activities into shared tasks, minimizing redundant operations, reducing urban disruption, and improving deployment timelines.

Building upon our earlier work using GAs to automate greenfield network design, this paper extends the methodology toward more complex and realistic deployment conditions, delivering two principal contributions [1].

First, an optimization framework capable of handling brownfield distribution networks, enabling cost-effective integration with pre-existing infrastructure while respecting operational constraints and leveraging existing assets. Second, a hierarchical cell-based decomposition strategy that partitions large-scale networks into manageable feeder and distribution units, systematically scaling autonomous design from individual cells to metropolitan deployments.

Together, these advancements represent a significant step toward fully autonomous, end-to-end network design processes that can address the diverse challenges of next-generation broadband deployments.

Background

Previous work on autonomous FTTH network design addressed a critical industry challenge: reducing the typically up to 1.5 days required to design an average network manually [1]. This labor-intensive process involves positioning equipment, planning cable pathways, defining network topology, and configuring connections, making rapid fiber network expansion difficult and expensive.

To address this issue, our work proposed a GA-based approach tailored explicitly for GPON distribution network design. Unlike traditional optimization methods, which become computationally impractical with large solution spaces, the proposed solution explored multiple network configurations simultaneously through an evolutionary process. The algorithm encoded designs as chromosomes representing ODP placement and used genetic operations – mutation, selection, and crossover – to iteratively refine solutions while balancing conflicting objectives: minimizing cable length, optimizing equipment placement, and ensuring complete coverage [1].

The fitness function evaluates solutions based on five cost factors: drop and distribution cable costs, ODP, splitters, and penalties for unallocated homes. Pre-optimization techniques, using relevant graph theory algorithms, reduced the initial graph to its critical connections, thereby improving computational efficiency.

The results were compelling: cost reductions between 3% and 78% were achieved across evaluated networks, varying according to their complexity, with all optimizations completed in less than one minute. The automated approach

consistently delivered networks with fewer ODP, more efficient cable routing, and better resource utilization while maintaining 100% coverage. This reduced design time from days to minutes and established AI as a practical accelerator for fiber deployment, though the scope remained focused on distribution networks, leaving feeder network optimization as a clear next step.

Evolving the distribution network

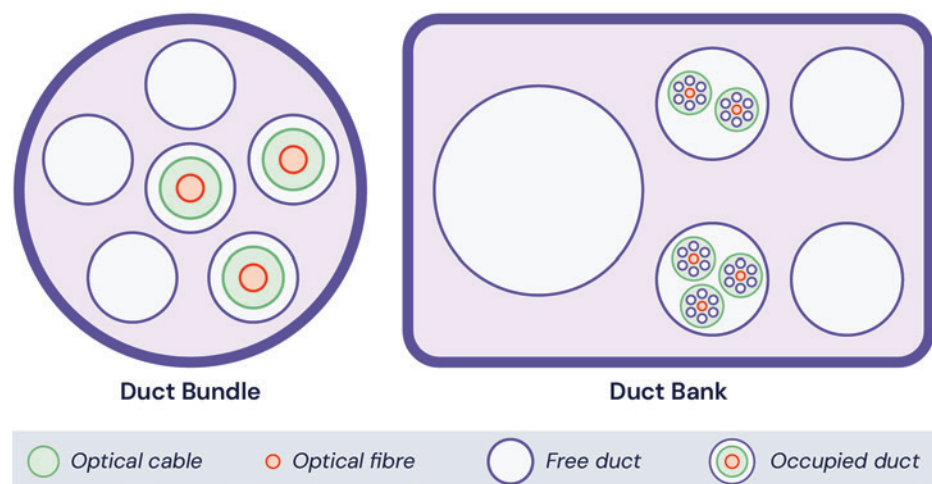
One of the key vectors of evolution has been the distribution network itself, shifting from greenfield scenarios, focused on new deployments, to brownfield contexts, where existing infrastructure is expanded and optimized. This section examines how duct structures support this transition, explaining the technical and operational mechanisms that enable brownfield deployments and the rationale behind them. It also explores the practical aspects of physical implementation, concluding with a real-world case that illustrates these concepts in action.

Managing duct bank and duct bundle routes

Duct infrastructures are essential for preserving the longevity and reliability of fiber optic cables, protecting them from corrosion, weather conditions, and temperature fluctuations. Commonly organized as duct bundles or duct banks, these structures form a foundational layer of network resilience. Duct banks and duct bundles are elements represented in the solution as infrastructures that can hold a set of ducts. The ducts were represented realistically, considering physical concepts such as their diameter, occupancy, and obstruction.

Representing the notion of available space within duct infrastructures is crucial, as ducts may be partially or fully occupied, directly affecting how new cables are allocated and how routes are chosen. Assessing duct capacity is, therefore, a critical step for preventing improper or infeasible cable assignments. Structurally, duct bundles accommodate only one cable per duct, while duct banks can contain multiple cables within a single duct. A visual comparison of these two configurations is provided in **Figure 2**.

Figure 2: Duct infrastructures representation



The move from greenfield to brownfield challenges fiber providers to reconcile theoretical designs with existing network realities

The automated network design system integrates duct infrastructures by assigning distribution cables based on available capacity along predefined routes. During this process, ducts are evaluated for their ability to accommodate new cables within defined spatial constraints, ensuring efficient use of existing infrastructure.

Two main criteria govern this evaluation: occupancy and obstruction. Occupancy checks ensure that adding a new cable does not exceed a duct's maximum allowable fill rate, particularly relevant in duct banks where multiple cables may share a single duct. Obstruction status, on the other hand, marks ducts as unavailable when blocked or already in use. Both characteristics are updated during the design flow.

Expanding beyond greenfield

The transitioning process from greenfield to brownfield is a well-known challenge that largely depends on the scale of the change [3]. The evolution of distribution network design reflects this shift, requiring a deeper understanding of network topologies.

In greenfield scenarios, the system designs networks for new areas without existing infrastructure, creating all connections from the Central Office to customer homes. This is ideal for emerging areas and new fiber deployments.

Brownfield scenarios arise when fiber providers need to expand existing networks to support additional customers using deployed infrastructure. New customer connections leverage existing distribution points through two modes: automatic selection, where the solution finds the optimal distribution point by minimizing distance, reducing both cost and optical budget; or manual selection, where network planners choose equipment locations based on operational requirements.

This transition expands the system from theoretical planning to practical deployment, aligning the design process with real-world network topologies.



Autonomous physical deployments

The transition from network design to physical deployment marks the point where planning meets real-world execution, especially in brownfield scenarios where new connections rely on existing infrastructure. In these contexts, understanding which installations can share deployments is key to improving efficiency and minimizing civil disruptions. Although automated tools optimize fiber layouts, construction still entails costly manual tasks, such as accessing maintenance holes, trenching, cable routing, and repaving. This makes it crucial to identify overlapping routes and shared infrastructure access points to reduce redundant labor through coordinated activity grouping.

From a problem to an algorithm

This problem brings unique challenges for algorithmic selection. Unlike supervised learning scenarios, which require labeled data and predefined outputs, this task involves uncovering latent structure within project interdependencies, driven by cost-sharing opportunities. The absence of ground truth and the combinatorial nature of minimizing shared costs make graph-based clustering and unsupervised optimization methods provide a more natural and interpretable framework for addressing the problem.

Several algorithmic families were considered. Evolutionary approaches, such as Ant Colony Optimization, excel at sequential pathfinding but struggle with combinatorial grouping, where value emerges from the overlap of infrastructure. Graph-based heuristics can identify shared segments, yet they lack principled mechanisms for balancing aggregate savings against practical constraints, such as manageable work package sizes.

Hierarchical clustering proved uniquely well-suited to this domain. It operates directly on pairwise economic relationships, without requiring training data, and encodes optimal groupings at every granularity level. This hierarchy enables planners to navigate trade-offs between maximum savings and operational feasibility in real-time. By transforming infrastructure overlap into mathematical distances, the approach delivers both provable cost optimization and the practical flexibility essential for real-world deployment.

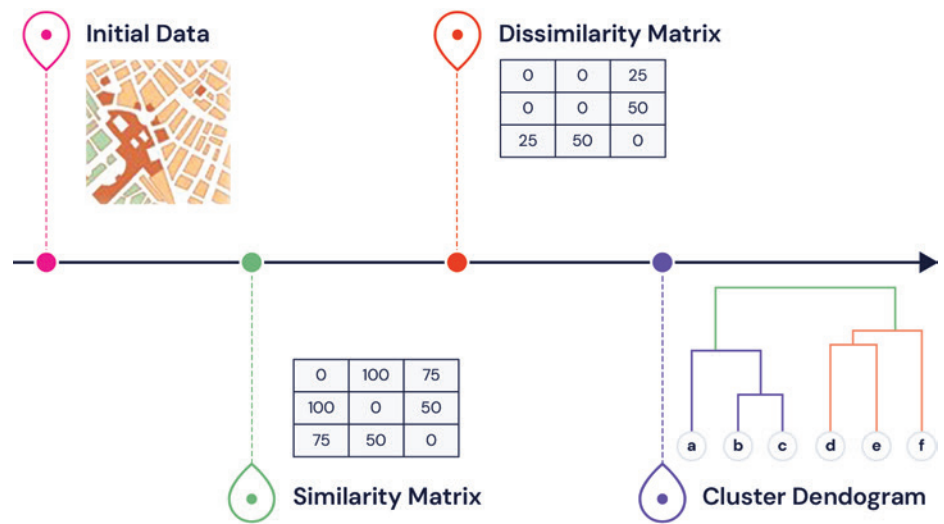
From savings to network clusters

A hierarchical clustering approach was developed to automatically group construction projects based on their shared work potential. The algorithm analyzes project paths alongside available network infrastructure to identify common segments that translate directly into shared construction activities (e.g., opening a distribution point once for multiple connections, or routing several cables through the same conduit in a single operation).

To measure these gains, a savings similarity matrix is calculated, quantifying the cost benefits of executing any two projects together. This matrix undergoes a mathematical transformation into a dissimilarity matrix for a distance-based representation, as shown in **Figure 3**.

Transforming infrastructure overlap into measurable distances, hierarchical clustering balances cost savings with operational feasibility

Figure 3: Hierarchical clustering pipeline



The final cluster selection applies a maximal savings criterion, prioritizing configurations that yield the most significant cost reduction while favoring higher granularity when multiple options achieve equivalent savings. This ensures operational flexibility by allowing construction activities to be divided into smaller, more manageable work packages without compromising overall efficiency.

The underlying savings model is formulated as an extensible matrix-based framework that can integrate additional cost factors, such as shared traffic management, coordinated street closures, or joint deployment of security resources, as operational requirements evolve. This adaptability maintains the model's analytical validity and optimization potential over time. Empirical results demonstrate that this approach achieves significant deployment cost reductions by systematically identifying shared construction opportunities that would otherwise be difficult or infeasible to detect manually across large-scale project portfolios.

A NOSSIS Inventory use case

The practical impact of autonomous deployment clustering can be demonstrated through a real-world case on the NOSSIS Inventory environment. Three customers in a common neighborhood requested fiber connections, resulting in three brownfield network installations (Networks A, B, and C) from the existing infrastructure to their homes. While one client requires an isolated connection, the other two share a portion of their path and connect through the same distribution point. **Figure 4** illustrates the three networks individually and their physical overlap, revealing the latent opportunity for construction consolidation.

Figure 4: Representation of the physical networks in the use case



Table 1 presents the deployment costs for each network connection when built independently, encompassing civil works, cable infrastructure, and management. These costs reflect actual values from operational deployments, providing a realistic baseline for evaluating the potential for optimization. Building each network separately could require independent site mobilizations, redundant access to shared infrastructure elements, and discrete civil work operations – even where physical paths overlap.

Table 1: Network deployment costs

Network	Civil work cost	Cable cost	Management cost	Total cost
A	€3,473.89	€481.35	€1,256.63	€5,211.86
B	€2,830.62	€528.68	€1,256.63	€4,615.93
C	€2,472.73	€801.05	€1,256.63	€4,530.40

The proposed solution automatically identified the optimal grouping strategy for this scenario: combining Networks A and B into a single deployment cluster, while keeping Network C independent. **Figure 5** displays the deployment clusters and the hierarchical dendrogram, which clearly shows substantial savings between Networks A and B at lower merge heights. At the same time, Network C maintains high distance values, indicating no cost-sharing benefit. The corresponding map visualization highlights how the algorithm correctly identifies the infrastructure overlap between A and B, as well as the physical isolation of Network C.

Figure 5: Hierarchical clustering dendrogram and clusters generated

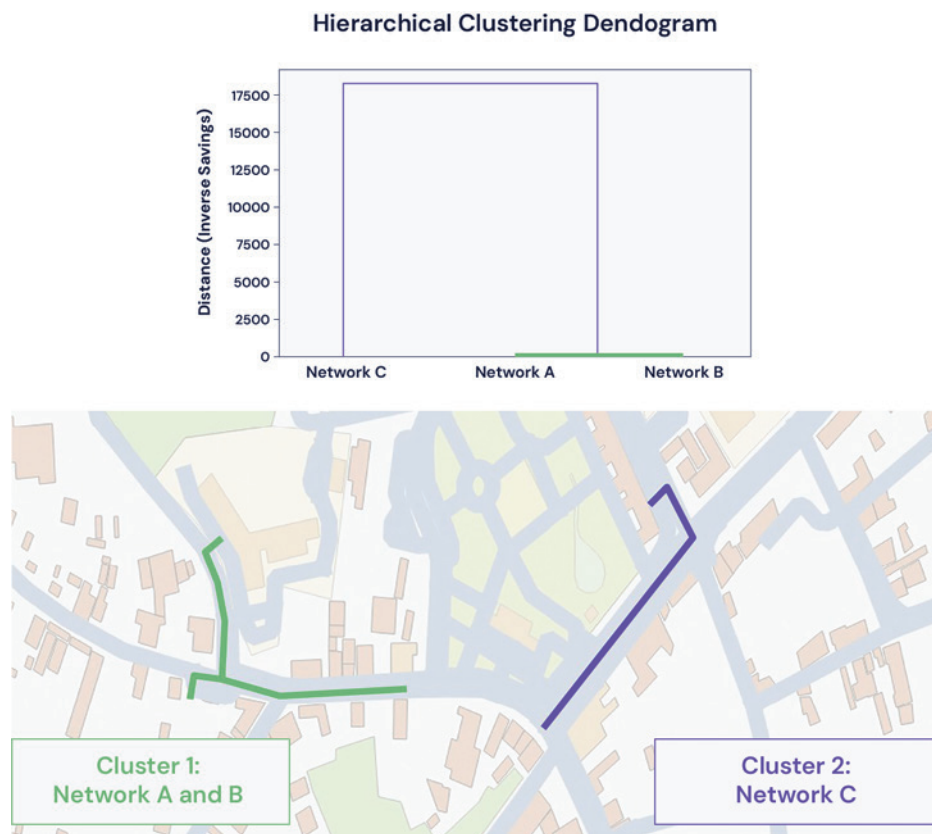


Table 2: Cluster deployment costs

Cluster	Civil work cost	Cable cost	Management cost	Total cost
NC	€2,472.73	€801.05	€1,256.63	€4,530.40
NA+NB	€5,479.82	€1,010.03	€1,256.63	€7,889.10

Table 2 presents the cost breakdown for the two identified clusters. The consolidated deployment of Networks A and B achieves approximately a 15% reduction in total costs, compared to independent construction. These savings stem from several factors: shared maintenance holes require only one access event rather than multiple interventions, and common route segments are excavated and repaved once. Beyond direct cost savings, this consolidation reduces deployment time by eliminating redundant site visits, thereby streamlining the process. It also minimizes disruptions to existing infrastructure, lowering both the risk of construction errors and the impact on neighborhood traffic and services. As a result, the combined deployment approach enhances both cost-effectiveness and implementation reliability.

This case demonstrates how intelligent clustering transforms abstract cost models into tangible operational benefits. The algorithm successfully navigated the trade-off between maximum consolidation and practical deployment constraints, recognizing that forcing Network C into the cluster would provide no economic advantage while potentially complicating construction logistics. The 15% cost reduction, combined with accelerated deployment timelines and reduced infrastructure stress, validates the approach's value in real-world brownfield scenarios where every optimization compounds across hundreds or thousands of customer connections.

Breaking down the feeder network

The cell division problem addresses the transformation of large-scale network datasets into manageable feeder cells that align with operational and business requirements – essentially a graph-partitioning challenge. Starting with a graph representation of a primary network that includes thousands of installation point nodes (such as poles, cabinets, and ducts) and numerous home nodes distributed across a geographic area, the objective is to generate individual sub-networks, each connected to a designated central office node.

Effective cell division must satisfy multiple constraints simultaneously: adherence to the maximum client capacity per cell, maintenance of relative uniformity in cell size and composition, and preservation of connectivity as single sub-graphs reachable via a single cable route from the central office. This section examines the rationale behind the design of the distribution cells, the dynamic placement of their key components, and the various approaches developed to address this partitioning challenge, concluding with an analysis against a real-world network scenario.



Designing the distribution cells

The design of distribution cells within the feeder network follows a structured process to ensure balanced and connected topology. The system loads network data, determines central office placement, and constructs a graph representation of the topology. The network is then divided into multiple distribution cells. Each cell is designed to connect Splitter Joints, ODP, and homes through optimized cable routing. Finally, cells are integrated into the feeder network, creating a unified, non-overlapping network structure.

Dynamic positioning of key components

The placements of Central Office and Splitter Joints are essential for constructing the feeder network and ensuring efficient data distribution. It is optimized to balance connection costs and maintain uniform distances between network elements, minimizing unnecessary cable usage.

During the feeder design phase, the system calculates the ideal central office placement based on input data and user preferences, identifying the network's weighted center through a graph-based analysis. The nearest node to this center becomes the central office, unless a user specifies a preferred area or exact location. Similarly, Splitter Joints are positioned to achieve optimal coverage of all homes, with the algorithm selecting locations that minimize cable lengths and ensure cost-effective deployment.

Exploring different strategies

To address the cell partitioning challenge, three distinct algorithmic approaches were investigated. Several traditional approaches to the graph-partitioning problem were experimented with, but most cannot map the design criteria and graph properties. Three approaches stood out: Genetic Algorithms, Integer Linear Programming, and Spectral Clustering. The following subsections detail the implementation and issues associated with each approach in handling the partitioning constraints.

Integer Linear Programming

Integer Linear Programming is a mathematical optimization method in which variables are restricted to integer values, aiming to optimize an objective function while satisfying predefined constraints [4]. When some variables remain continuous, the approach becomes Mixed Integer Linear Programming (MILP). For the cell division problem, MILP was selected because cell connectivity cannot be represented purely in binary form – continuity requires continuous values to model the connection paths between nodes properly.

The formulation encodes the problem through variables representing node-to-cell assignments, inter-node connections within cells, and flow-based values that establish continuous connectivity. The objective function maximizes internal cell connectivity to ensure cohesive, connected partitions. Constraints enforce node exclusivity (each node belongs to one cell), client capacity limits per cell, and connectivity requirements. The connectivity constraint represents the most complex aspect: a flow-based approach designates a root node in each cell and propagates a connectivity unit through the nodes, verifying uninterrupted paths to confirm that each cell forms a single connected subgraph.

Genetic Algorithm

Genetic Algorithms, inspired by principles of natural selection and evolution [5], are a powerful and flexible tool for solving optimization problems. The problems are modeled as chromosomes that evolve through iterations, similar to the real world, with diversity operations such as mutation and crossover.

The chromosome directly encodes cell assignments across all installation points in the network. Each gene in the chromosome contains an integer value ranging from one to the total number of cells, creating a straightforward mapping. The fitness function evaluates each partition based on compliance with capacity constraints, cell similarity metrics, and connectivity rules.

To ensure an effective starting population, k-means clustering is employed to generate initial partitions, thereby promoting diversity and providing a feasible baseline solution. These initial assignments are then refined through iterative genetic operations – crossover and mutation – that explore alternative configurations and guide the population toward more optimal and well-balanced network divisions.

Spectral Clustering

Spectral Clustering is a graph-based partitioning technique that leverages the eigenstructure of matrices [6], derived from the network topology, to identify natural divisions within a graph. Unlike traditional clustering methods that rely on distance metrics in feature space, spectral approaches exploit the connectivity patterns encoded in the graph's adjacency or Laplacian matrix [6], making them particularly suitable for network partitioning problems where the topological structure is essential.

For the cell partitioning problem, the network is represented as a Laplacian matrix, capturing connectivity relationships and spatial proximity among installation points. The algorithm computes the eigenvectors associated with the smallest eigenvalues of this matrix, embedding the graph nodes into a lower-dimensional spectral space where clusters become more distinct. The resulting embeddings are then clustered using k-means, assigning each installation point to one of the k-cells based on its position in the spectral space. This procedure groups nodes that are strongly interconnected while separating regions with weaker connectivity.

The method effectively identifies partitions aligned with the graph's intrinsic community structure, producing cells with high internal cohesion and minimal inter-cell links. However, spectral clustering does not inherently account for domain-specific constraints, such as capacity limits or connectivity requirements, which are instead managed through post-processing heuristic validations to ensure the practical feasibility of the resulting partitions.

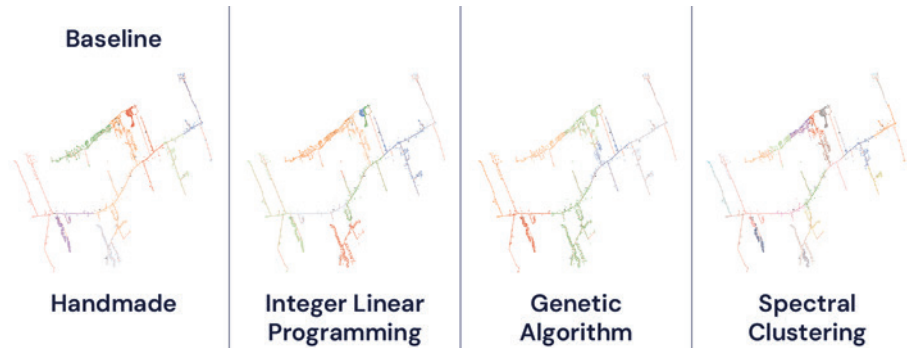
A NOSSIS Inventory use case

To anchor the analysis in real-world conditions, each approach was applied to a network dataset from the NOSSIS Inventory platform, which contained a manually engineered cell partition as a practical baseline. The handmade division enforces a maximum of 128 client units per cell – a threshold that maximizes component reusability in secondary network design. This same constraint was applied uniformly across all three approaches to ensure comparable evaluation of their efficiency and quality, against both each other and the baseline design.

Table 3 and **Figure 6** present the comparative results across all approaches, measuring duration; number of cells generated; average, standard deviation, and dispersion of client units per cell; and average absolute distance from the maximum clients per cell. The baseline, although not directly comparable in duration due to its manual nature (requiring several minutes to hours as part of a broader planning process), comprises nine cells that exhibit relatively high dispersion and absolute distance from the maximum capacity threshold. This indicates a design philosophy that accepts smaller cells for improved overall network consistency.

The MILP approach required 298 seconds to generate seven cells, achieving the most statistically coherent partition with the lowest dispersion and minimal deviation from the 128-unit target. However, while cells respect capacity constraints and demonstrate mathematical optimality, the design reveals significant spatial fragmentation, as nodes are frequently assigned to cells outside their geographic

Figure 6: Cell division by each methodology



boundaries. The approach’s scalability remains severely limited, with computational time growing for larger networks, despite delivering well-balanced cell sizes.

The GA completed execution in 49 seconds, producing seven cells with decent average client occupancy but exhibiting the highest dispersion, and a high average absolute distance from target capacity. Despite extensive hyperparameter tuning and experimentation with diverse crossover and mutation operators, premature convergence consistently limited performance – optimal solutions typically emerged from the initial k-means population generation with minimal subsequent improvement. While offering superior flexibility in encoding business criteria, the design shows frequent capacity violations and poor topological coherence, with cells containing nodes from geographically proximate but disconnected network segments.

Spectral Clustering demonstrated the best computational efficiency, completing in approximately 21 seconds, but generating 22 cells – significantly more than other approaches. The results reveal extremely low average client units per cell, with the highest absolute distance from the target threshold, confirming the method’s tendency toward conservative and unbalanced partitioning. Unable to directly incorporate capacity constraints, the algorithm creates excessively small cells to avoid violations.

Table 3: Cell division quality metrics analysis

Approach	Duration (s)	Num. cells	Avg. num. client units	Std. num. client units	Dispersion	Avg. abs. dist. num. clients
Baseline	n/a	9	91.56	35.70	94.00	36.44
MILP	298	7	117.71	17.80	50.00	10.29
GA	49	7	117.71	67.89	188.00	54.57
Spectral Cl.	21	22	37.45	24.94	86.00	90.55

To evaluate the practical quality of each cell partitioning approach, the designed cells were used as input for generating the complete network. Each cell division approach – MILP, GA, Spectral Clustering, and the human baseline – underwent the same distribution network design process described earlier. The distribution cell design logic was applied individually to each cell, generating ODP placement, cable routing, and equipment allocation, which were then combined to form a complete end-to-end network topology.

The resulting networks were evaluated using three key metrics: combined fitness value (aggregating individual cell fitness scores based on cable costs, equipment usage, splitters, connectors, and penalties for design violations) [1], penetration rate (percentage of homes successfully allocated to the network), and number of unallocated homes (customers that could not be served under the current cell division). These metrics provide a comprehensive view of both cost-effectiveness and coverage quality for each partitioning strategy, as detailed in **Table 4**.

Table 4: Network evaluation metrics analysis

Approach	Fitness value	Penetration rate	Unallocated homes
Baseline	282,306.17	99%	9
MILP	879,510.91	88%	92
GA	1,004,238.04	85%	113
Spectral Cl.	282,913.62	99%	9

The baseline human-designed cell division achieved the best overall performance with a 99% penetration rate, only nine unallocated homes, and the lowest fitness value. Spectral Clustering delivered the second-best result despite generating more cells, matching the baseline penetration rate and unallocated homes count, but with a slightly higher fitness value. MILP scored third with an 88% penetration rate and 92 unallocated homes. At the same time, GA performed worse, with a fitness value more than three times higher than the baseline and Spectral Clustering, achieving only an 85% penetration rate and leaving 113 homes unallocated.

Conclusions

This work represents a significant evolution from distribution-only network design toward fully autonomous, end-to-end FTTH network planning. By extending the GA approach to encompass both feeder and distribution networks, while integrating real-world duct infrastructure and brownfield deployment scenarios, the solution transitions from theoretical optimization to practical industry applications.

Brownfield integration and deployment clustering emerged as critical advancements for reducing both CAPEX and OPEX. The duct infrastructure management system facilitates accurate cable routing by taking into account all current assets and infrastructure. Clustering orders after the design phase demonstrated tangible value in the NOSSIS case study, achieving 15% cost reduction and accelerated deployment timelines. Beyond direct CAPEX savings, this approach reduces OPEX by minimizing time, urban disruption, and streamlining site operations. These factors compound significantly across large-scale rollouts, aligning directly with the dual objectives of cost efficiency and time reduction.

Cell partitioning results, while modest, validate the practical viability of automated approaches. Despite none of the three methods delivering exceptional

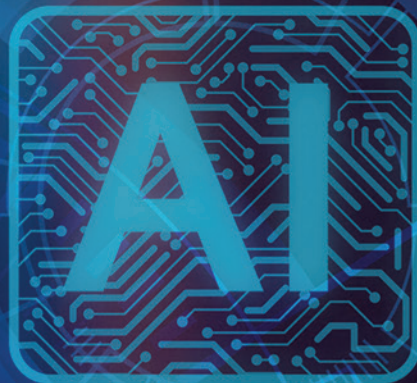
Extending optimization across feeder, distribution, and brownfield realities moves FTTH planning closer to practical, scalable automation

performance individually, the baseline results were nearly matched in a fraction of the time. This dramatic time compression makes these tools highly valuable for network designers, significantly boosting productivity by automating the labor-intensive partitioning phase, while maintaining design quality equivalent to manual engineering.

Future work should focus on refining cell partitioning approaches, particularly MILP and Spectral Clustering, which show clear potential for improvement through enhanced constraint modeling and geographic cohesion mechanisms. In addition, the order clustering approach could be further developed to better reflect more realistic deployment scenarios and adapt to emerging requirements, as network conditions and design objectives evolve. Beyond algorithmic refinement, the natural progression extends to infrastructure network design, completing the autonomous design cycle. This involves moving from planning networks over existing infrastructure to designing the infrastructure itself, representing one step further toward truly comprehensive autonomous AI-based network deployment.

References

- [1] X. Silva et al., "Automatic FTTH network design," *InnovAction*, no. 9, pp. 106–117, 2024.
- [2] M. Tahon et al., "Improving the FTTH business case—A joint telco-utility network rollout model," *Telecommun. Policy*, vol. 38, no. 5–6, pp. 426–437, Jun. 2014, doi: 10.1016/j.telpol.2013.01.004.
- [3] J. W. Axehill, E. Herzog, J. Tingström, and M. Bengtsson, "From Brownfield to Greenfield Development – Understanding and Managing the Transition," *INCOSE Int. Symp.*, vol. 31, no. 1, pp. 832–847, Jul. 2021, doi: 10.1002/j.2334-5837.2021.00873.x.
- [4] F. Clautiaux and I. Ljubić, "Last fifty years of integer linear programming: A focus on recent practical advances," *Eur. J. Oper. Res.*, vol. 324, no. 3, pp. 707–731, Aug. 2025, doi: 10.1016/j.ejor.2024.11.018.
- [5] R. Malhotra, N. Singh, and Y. Singh, "Genetic Algorithms: Concepts, Design for Optimization of Process Controllers," *Comput. Inf. Sci.*, vol. 4, no. 2, Feb. 2011, doi: 10.5539/cis.v4n2p39.
- [6] H. Jia, S. Ding, X. Xu, and R. Nie, "The latest research progress on spectral clustering," *Neural Comput. Appl.*, vol. 24, no. 7–8, pp. 1477–1486, Jun. 2014, doi: 10.1007/s00521-013-1439-2.



11001
01010
10011



07

Prompting the Shift: Democratizing Telco Configuration with Low-Code and AI

Abstract

The telecommunications industry is undergoing a shift toward more inclusive configuration workflows, driven by the rise of low-code/no-code (LC/NC) platforms and AI-powered interfaces. Traditionally, telco software required deep technical expertise, limiting usability and increasing reliance on support teams. This paper explores how LC/NC design paradigms and prompt-based AI can democratize configuration tasks, enabling broader participation without compromising precision or control. We examine the evolving profile of system administrators, the usability challenges of complex configuration interfaces, and the role of predictive UX and intent-based networking in simplifying user interaction. Central to this transformation is the issue of trust in AI systems, particularly concerns around transparency, reliability, and control. Drawing on frameworks like Human-Centered AI (HCAI), Trustworthy AI (TAI), and Closed-Loop Automation, we propose design strategies that foster user confidence and support scalable, resilient telco ecosystems. Through case studies and industry data, we highlight how thoughtful interface design can bridge the gap between technical depth and user autonomy, paving the way for more accessible and intelligent configuration experiences.

Author

Leonor Flores, Altice Labs, maria-l-igreja@alticelabs.com

Keywords

Low-Code/No-Code, Prompt-Based AI, Predictive UX, Human-Centered AI, Trustworthy AI, User Trust, Democratized Software Tools

DOI

10.57650/INNOVATION/2025.07



Introduction

At Altice Labs, we continuously design and develop new features and enhancements to improve the adoption, usability, and capabilities of our software products, which are deployed by telecommunications companies worldwide. These products, such as NOSSIS and NGIN, enable users to execute highly complex configurations across network operations, charging systems, and policy management. However, the sophistication of these tasks often necessitates a high level of technical expertise, resulting in a reliance on specialized personnel and frequent support requests. This dependency highlights a critical challenge: the usability of our tools is tightly coupled with the complexity of the operations they support, creating barriers for less technically inclined users.

In response to this challenge, recent years have seen a strategic shift toward more accessible design paradigms, particularly through the adoption of Low-Code/No-Code (LC/NC) platforms. These approaches aim to democratize configuration tasks by abstracting technical complexity and enabling users to perform advanced operations through intuitive, visual interfaces. While LC/NC solutions represent a

meaningful step forward, they are not a panacea. Significant gaps remain across other products and features, and the evolving expectations of users – driven by the rapid proliferation of Artificial Intelligence (AI) – demand further innovation.

AI technologies, especially those centered around prompt-based and predictive interactions, offer promising avenues for enhancing user autonomy and simplifying engagement with complex systems. Yet, the integration of AI into existing products introduces new challenges, particularly around user trust and adoption. Studies indicate that the primary barrier to AI adoption is not technical capability but a lack of trust, stemming from concerns about transparency, reliability, and control. This issue is especially pronounced when AI features are introduced into familiar workflows, where users may resist changes that feel opaque or unpredictable.

This article explores the intersection of product design, usability, and AI integration within the telco industry. It examines how LC/NC platforms and AI-driven interfaces can transform user experiences, reduce support dependency, and foster greater autonomy. Central to this exploration is the role of trust in human-AI interaction, and how design principles – rooted in frameworks such as Human-Centered AI (HCAI), Trustworthy AI (TAI), and Closed-Loop Automation – can guide the development of intuitive, inclusive, and ethically sound solutions. By aligning technological innovation with user needs and expectations, we aim to chart a path toward more accessible and resilient telco software ecosystems.

Telco industry complex configuration paradigm

As providers of software solutions for telecommunications companies all around the world, Altice Labs is aware of the complexity of the use cases covered in our products. This is a result of answering our customers' needs, who require solutions for their highly specific edge-cases, allowing them to provide the best service possible to their respective customers. One significant problem has come from this: the usability of our products ends up reflecting the intricacy of the use cases they respond to. In turn, this contributes to a high volume of requests to our support teams, as our clients are not completely autonomous as users of our products. Our tools can support the complex configurations they need to implement, but our support teams need to assist them in performing them on the interface.

This is also connected with a paradigm shift that has been noticed in the telco industry. There has been a change in what is called the "system administrator group" [1]. If before these were users who had a deep understanding of the systems and coding skills, as well as other technical knowledge that would allow them to perform configurations in highly technical interfaces, nowadays this group of users has expanded to include less technical-savvy administrators [1]. This has lowered the level of expertise expected from these professionals, but our products are only now taking the first steps to accommodate this new set of users, burdening support teams with a growing number of help requests.

From a product design perspective, we must address the established difficulties encountered by administrators. These challenges, identified in research on configuration interfaces, include [1]:

1. **Lack of guidance and information:** Administrators must find parameters related to expected system behavior and set correct values from hundreds of available options.
2. **Inconsistency and ambiguity:** This applies to the interface's appearance (such as parameter naming and data formats), configuration requirements (e.g., case sensitivity), and program behavior related to configuration.
3. **Complexity beyond capability:** Configuring systems is inherently complicated, and this difficulty prevents end-user administrators – who often lack the time or capability – from dealing with such complexity.
4. **Lack of environmental awareness:** Unlike end-user applications, system software involves multiple interacting components. When functionality fails, it is hard to know which component is improperly configured.
5. **Lack of technical support:** Online technical support for system admins is often slow, incomplete, or inefficient due to missing diagnostic details.

The key shortcoming of today's configuration interfaces is their inability to help administrators understand the configuration knobs (e.g., constraints, correlated environment information) and their impact on the system [1]. In this context, product designers should be particularly aware of the factors presented in

Table 1: Findings on configuration problems and their implications for configuration interface design [1]

Table 1, namely the design implications of the configurations problems and the cognitive obstacles to configuring systems. Addressing these obstacles, especially with the evolved system administrator group, necessitates interfaces that are more informative, instructive, and user-friendly. This context validates the necessity of introducing solutions based on the Low-Code/No-Code (LC/NC) paradigm.

Characteristics of Configuration Problems (Section 2, 3)	Design Implications
<p>1. The shift of administrators. The system administrator group has expanded greatly to include non- and semi-professional administrators.</p>	<p>New interface for system configuration is desired to accommodate the shift of the administrator group.</p>
<p>2. Configuration and Programming are anti-correlated. Administrators and programmers (including scripting) from different communities and have different skill sets.</p>	<p>System configuration should be studied as a separate problem from programming. The principles of building (end-user) programming interface might not be applicable to configuration.</p>
<p>3. The separation of understanding and manipulation. The separation of user manuals and configuration files causes administrators' cognitive difficulties and errors.</p>	<p>System vendors should not assume that manuals can help users solve their configuration problems. Instead, the configuration interfaces should integrate the information in user manuals.</p>
<p>4. Difficulties rather than errors. In more than 65% of cases, the configuration problems are administrators' difficulties (e.g., finding related parameters, setting values) rather than committing errors.</p>	<p>The configuration interface should try to guide and educate administrators rather than directly asking them to input values of configuration parameters.</p>
Cognitive Obstacles to Configuring Systems (Section 4)	Design Implications
<p>1. Lack of guidance and information. This is the major cause of today's administrators' configuration problems.</p> <p>This results in two challenges towards configuration:</p> <ul style="list-style-type: none"> • Finding the right parameters relevant to tasks from the entire parameter set; • Setting the parameters' values to achieve the intended system behavior. 	<p>Configuration interfaces should be more informative to help administrators address the two challenges.</p> <p>To address the first challenge, configuration interfaces should provide administrators with dependency, correlation, and association information regarding their settings.</p> <p>To address the second challenges, constraints, potential impact, and working examples should be provided by the interfaces.</p>
<p>2. Inconsistency and ambiguity. Inconsistency of interface appearance, correctness rules, and system behavior are one major cause of configuration problems, including both difficulties and errors.</p>	<p>Conceptual integrity should be carefully maintained for configuration interfaces, between interfaces and user manuals, and between interfaces and system behavior.</p>

Cognitive Obstacles to Configuring Systems (Section 4)	Design Implications
<p>3. System and control complexity. A significant portion of users' configuration difficulties are caused by their incapability in dealing with system and control complexity. Our hunch is that non-professional administrators have less performance and security concerns as professional administrators.</p>	<p>We should decouple the configuration interface for "dummies" and professional administrators, in a similar way as [8], [17]. Configuration parameters with different necessity and skill prerequisite should be separated in the different interfaces.</p>
<p>4. Lack of environment awareness. This is one common difficulty of diagnosing and resolving configuration-related system anomalies. The environment of a running system includes its underlying stacks (e.g., OS) and co-running software.</p>	<p>Configuration interfaces should help administrators recognize the environment information correlated to the configuration settings, for example, constraints, entities, and resources.</p>
<p>5. Lack of technical support. Administrators have difficulties in using Internet as technical support. Many questions of configuration problems are not answered or with unsatisfied answers.</p>	<p>Internet-based technical support services should try to reduce the response time and improve the efficiency of diagnosing/solving configuration problems.</p>

There is a specific trend in the design industry that has been known to be highly effective in situations parallel to the one we face with this new group of administrator users. This type of interaction, known as LC or NC, allows non-technical users to perform highly technical tasks by abstracting that complexity and using interface design to hide it behind the curtain.

LC/NC as a democratizing factor

LC/NC is defined as "a software development approach that allows for the creation of applications using visual drag-and-drop tools, rather than traditional coding" [2]. Its main benefits relate to faster time-to-market, due to faster development and deployment, editing capacity by people without technical expertise, fewer errors and bugs, and overall significant savings in costs, mainly related with development [3], [6]. Gartner and Forrester have been predicting for a few years the increasing adoption of LC/NC tools across several industries, amongst which are telcos [4]. Although most articles focus on LC/NC platforms used by telcos to develop their products, the present discussion takes a different perspective. It explores how Altice Labs can design and develop LC/NC solutions for the produced software. Going back to the previous section, this is a step forward in democratizing the usability of our products, increasing user autonomy, and decreasing errors and bugs that so often occur in these complex configuration tasks [7].

According to a Mendix study, four in ten businesses in Europe and the U.S. were already using LC/NC software for mission-critical operations back in 2021 [2].

LC/NC is reshaping software development across industries, and in telcos, it is becoming a catalyst for autonomy, agility, and fewer configuration errors

In the meantime, the expectation for more capabilities in this format of development has increased, as these platforms are no longer viewed as a crisis technology, but a core one [3], [6]. The highlighted benefits in this study include business agility in the form of rapid change, closer business/Information Technology (IT) collaboration, legacy renewal, and facilitated technology adoption [2]. In another study published in 2024, Gartner projects that the LC/NC market will grow around 16% from 2022 to 2027, with a forecast of over 16 billion dollars in that year, expecting good opportunities for both current vendors and new players [5].

The User Experience/User Interface (UX/UI) aspect of these platforms is crucial, as these are spaces where a user interacts with a graphic interface to develop programs, applications, configurations, or whatever tasks their use case requires. They allow users to build their desired outcome mainly through drag-and-drop interaction and fine-tuning specific characteristics, which also promotes collaboration across multiple users working in the same build, as well as future editing and a clear and intuitive understanding of the final result [4]. Crucially, the increase in user autonomy and the decrease in configuration errors directly translate into reduced operational expenditures (OPEX) by minimizing reliance on specialized support teams.

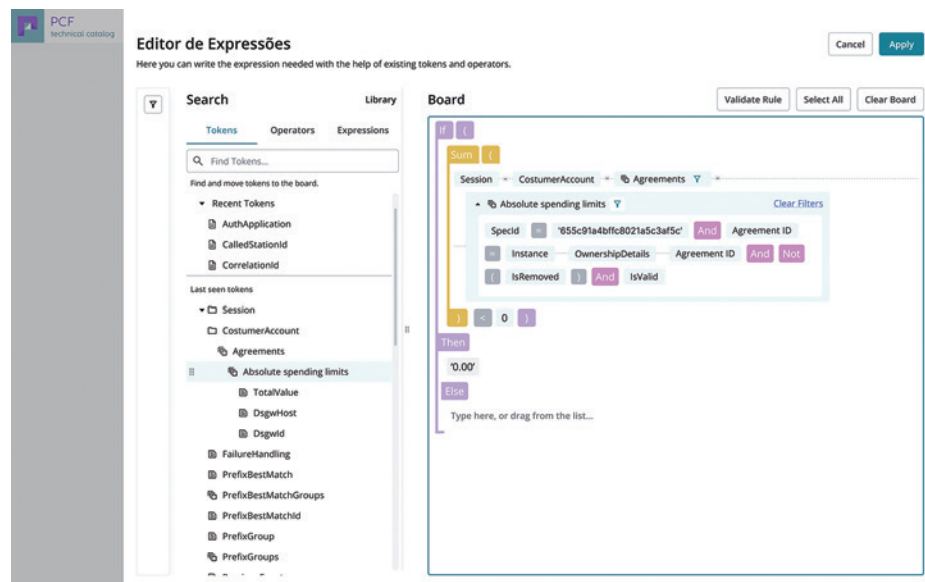
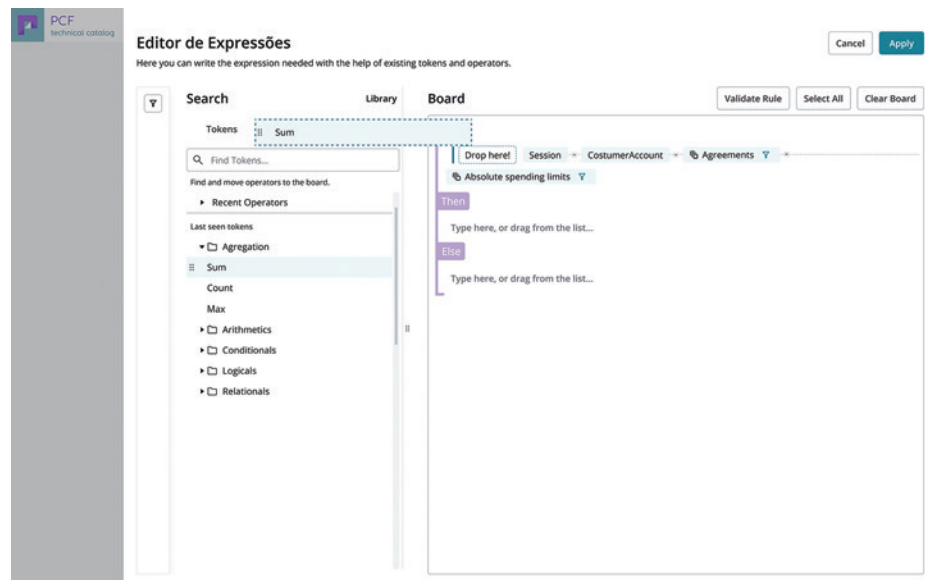
The main identified down-sides or limitations of LC/NC platforms as of today are the difficulty of access to source code, possible slowness in loading and publishing, and potential loss of personalization for unique corner-cases [4]. These are important things to consider when designing and implementing these types of platforms or solutions.

As a case study, let us consider the Expression Editor feature that was redesigned back in 2022 for the technical catalogs of the NGIN product suite. This is a low-code approach to perform the task of creating rules for charging configurations, which, up to now, could only be written in code. Even though this new feature still requires coding knowledge, it is at a much more basic level, as the logic functioning of the expression is quite intuitive.

In this area of the product, the user can drag and drop items from a library of tokens, operators, and pre-made expressions that can serve as a base for a new one and use these components to build the rule they envisioned. It is also possible to validate the expression before committing to it, which provides a big sense of safety and trust, as well as turning an expression into code, allowing the more technical users to have more control over their editing preferences.

This feature has since been developed and will go live for clients at the end of this year. It will represent an enormous step forward in leveling the ground for less technical-savvy users to be able to create these rules, allowing more people across the client's operation to perform and/or collaborate on these tasks, as well as relieving our support teams of some requests. However, as it was only possible to dedicate development effort to it this year, we should expect that the client's expectations have changed since this was designed. In just a bit over two years, AI has grown exponentially, and as we all feel its impact on our day-to-day life, so do our clients expect to have their work optimized and simplified with this technology. This does not diminish the fact that the new Expression Editor, represented in **Figure 1**, was and will be a step in the right direction; it just means we need to take the next one quickly, as our users are expecting and demanding us to.

Figure 1: New Expression Editor



In particular, it is important to determine how we can take the LC/NC mindset and keep moving towards an accessible and democratized usage of our configuration tools, whether in Business Support System (BSS) or Operations Support System (OSS) products. As generative-AI and prompt-based interaction increasingly infiltrate our daily life, it seems as if there is a big shiny arrow pointing in that direction. The introduction of AI-powered automation solutions in the telco industry is already a reality, and there have been seen significant reductions in operational costs and improvements in service delivery efficiency. Research points to two fields as the ones with the most proven benefits of AI introduction: Automated Operations and Maintenance, and Advanced Analytics and Decision Support [8].

The first shows over 50% reductions in manual interventions, as well as in unplanned downtime and overall maintenance costs [8]. The latter, with the processing of real-time data and advanced pattern recognition algorithms, can suggest actions and identify errors with over 90% accuracy [8]. There is, however, much to account for regarding the usability challenges that come with this

The journey from low-code interfaces to AI-driven intent marks a natural evolution: moving from manual configuration toward strategic, system-level control

approach, and that should be taken into consideration, so that telco companies can have a positive impact with introducing AI solutions in their workflows.

This continued pursuit of abstraction, by moving complexity “behind the curtain”, naturally leads to Intent-Based Networking (IBN) [9], which is already accelerating across the telecommunications industry. IBN serves as the evolution of the LC/NC mindset; it is an automated, software-driven approach that translates high-level business objectives, or intents, into network configurations and policies [9]. This methodology allows organizations to focus on strategic goals rather than manual device configuration, essentially taking the complexity abstraction pioneered by LC/NC to an AI-driven, systemic level [9]. The core challenge, however, shifts from dealing with configuration parameter obscurity to managing user trust in the automated intent translation.

AI-Driven configuration: beyond the prompt and towards predictive UX

A few years ago, prompt-based interaction started being a part of our lives. It was introduced mainly in the context of generative-AI, most famously with ChatGPT, but then with many others like Perplexity, DeepSeek, Claude, Gemini, Copilot, etc. These have quickly become important tools for our work as well as our personal lives, and users around the world rely on these platforms to help them with an enormous variety of subjects and issues, so much so that this type of interaction has quickly spread to tools that have been on the market for years, in the form of new features. The conversational aspect of this type of interaction has proven to be key in increasing ease of use, having come a long way from traditional chatbot interaction [11].

However, the surge of new roles like prompt-engineer or specialist, the hundreds of prompt hacks found throughout articles and social media, and our own personal experience point out some frustration in carefully crafting a prompt, just to have the AI tool produce a result that is nothing of what the user intended. “The smart defaults are either too broad or users must articulate their full intent upfront and then fall back on iterative chat-based back-and-forth to refine context.” [10].

Writing effective prompts has become a skill, but this frustration has been picked up by user experience specialists and started conversations around predictive AI [11]. This is when, instead of users having to articulate every detail of what they need, the system can anticipate it based on common behaviors and patterns. Instead of the user having to spend so much time crafting the perfect prompt, the system provides options for them to select, basically building the prompt together by narrowing down on the available options. This is not applicable in all contexts, but it is a good principle to keep in mind, as personalization, understood here as the need for well written prompts that will generate the expected result, has been identified as the second biggest blocker of adoption of AI tools for work-related tasks [12].

The trust barrier: opacity, aversion, and the need for transparency

The introduction of AI-powered automation solutions in the telco sector is not merely theoretical; it is already delivering tangible benefits, with research pointing to a 31% reduction in operational costs and a 42% improvement in service delivery efficiency among tier-1 providers [8]. These proven gains in Automated Operations and Maintenance, particularly through over 50% reductions in manual intervention, firmly establish AI as the necessary path forward for efficiency and agility [8].

Table 2: Results of a Survey of AI-Related Concerns Across Global and U.S. Populations [16]

However, multiple studies and articles point to trust as the number one blocker to AI adoption [11], [8], [14], [15], [16]. According to the Edelman Trust Barometer’s 2024 study, trust in AI companies has been declining since 2019, mainly due to privacy concerns and social bias, which are almost two times more widespread than concerns over job impacts [16], as shown in **Table 2**.

GLOBAL 28				
I am less than enthusiastic about AI because it...		Developed	Developing	U.S.
could compromise my privacy	39	41	35	52
may devalue what it means to be human	36	35	38	38
could be harmful to people	35	38	31	57
could be harmful to society	35	40	29	61
is not adequately tested and evaluated	35	40	30	54
is happening too fast	32	36	28	35
will worsen societal inequalities	27	28	26	33
could threaten my job security	27	18	26	19

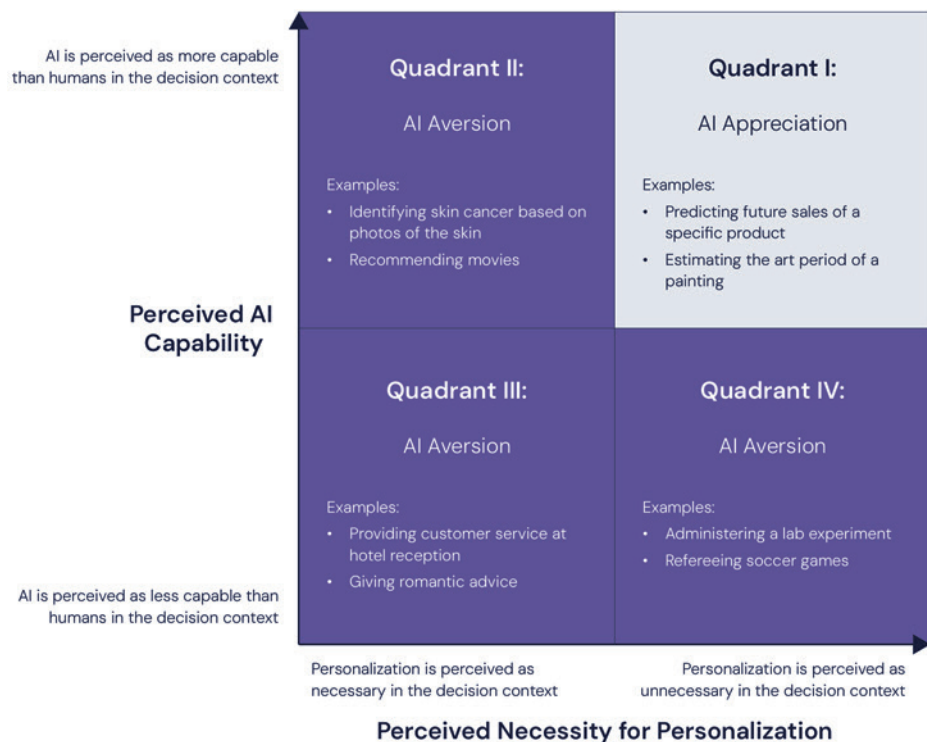
U.S. most concerned over AI's potential harm to society and people

Threats to job security less of a concern

Other studies point to AI characteristics that increase user distrust, such as being opaque (also referred to as black box), difficult to interpret, and biased – all of which impose risks to using AI-generated work in mission critical tasks [14], [15]. The black box issue is one of the most concerning, as it hides the logic that generated those results, making it harder to find the origin of errors and decisions, to predict AI behaviors, to be externally verified or questioned, and, therefore, to trust the generated results, as they may lead to unforeseen risks and biased decisions. “AI refers to machines capable of performing cognitive functions commonly attributed to the human mind, such as learning, reasoning, and decision making” [13]. So, for users, it is the same as if a human provided them with an answer without explanation or justification – they are immediately less inclined to trust it. Users need visibility of the logical path that led to that answer, so they can judge whether it makes sense and has no bias, and only then conclude if the result is correct and fair [13], [15].

Some studies go in a slightly different direction and refer to this phenomenon as “AI aversion”, as opposed to “AI appreciation”, as it has been shown that people tend to behave more negatively toward AI than they do toward humans [13]. This phenomenon, however, is not observed in every case. It depends heavily on the tasks or information that the user is trying to achieve. If their perception points to AI being more capable than humans in that context, they are more likely to trust and use it, for example, in computation tasks and chess-related games [13]. However, if the subject involves morality, emotion, ethics, or creativity, people have shown to be more averse to resorting to AI [13]. As mentioned previously, this also ties with the need for personalization – studies point us to the conclusion that, as of today, users are more likely to resort to AI when two conditions co-exist: they believe it will be more capable than humans in that context and that it will not require a lot of personalization of their request to obtain the desired result (see **Figure 2**) [13].

Figure 2: Four quadrants based on the capability-personalization framework of AI aversion versus AI appreciation [13]



Designing for trust: the Human-AI collaboration (HAX and Closed-Loop frameworks)

This leads us to a different question: if having to personalize the interaction too much is a blocker to adoption, what is the optimal balance for human-AI interaction? Research points to a framework of “closed-loop” collaboration as a good methodology, specifically in the telco industry. In this framework, human expertise comes into play by ensuring strategic oversight and decision-making, and by training and refining the systems [8]. Significantly better outcomes have come out of companies that have humans overseeing the automated results, especially in cases with complex regulations, with Network Operations Centers (NOC)

reporting over 50% reductions in critical incidents. Humans remain more competent than AI when it comes to “interpreting and implementing regulatory requirements, particularly in areas requiring contextual understanding and ethical considerations”, as well as in identifying emerging edge cases with accuracy [8].

Table 3: Performance improvements through closed-loop automation implementation [8]

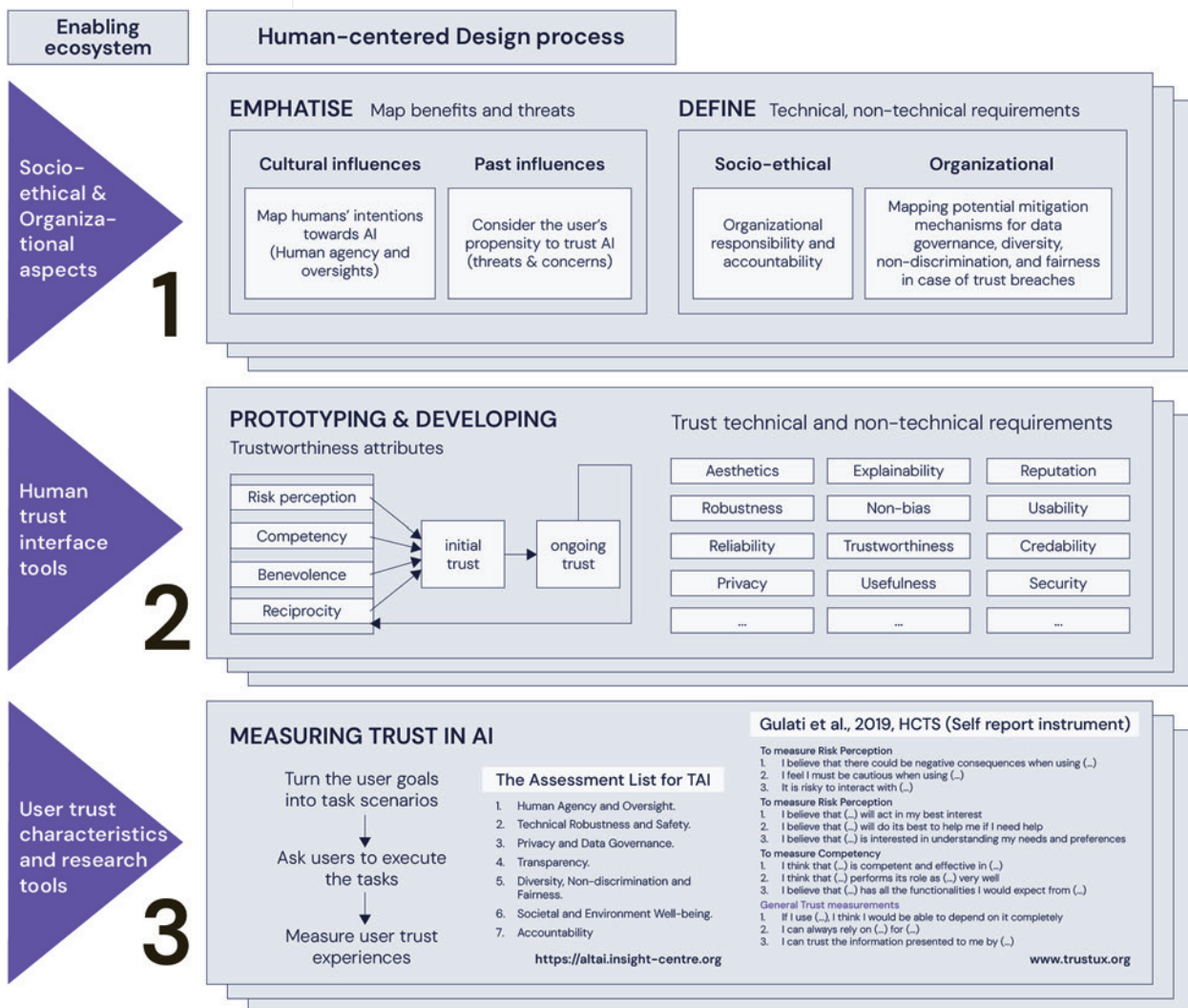
Performance Category	Metric	Improvement (%)
System Performance	Manual intervention reduction	80
	Network performance enhancement	35
	Data accuracy rate	99.5
	Fault prediction accuracy	94
Operational Efficiency	OPEX reduction	25
	MTRR reduction	55
	Network optimization	45
	Energy consumption reduction	30
Human Interface & Management	Decision-making time reduction	48
	Operator satisfaction rate	95
	Configuration error reduction	65
	Manual task automation	85
Service Quality	Service incident reduction	72
	Change success rate improvement	68
	Throughput improvement	58
	Latency reduction	45

As seen in **Table 3**, the implementation of the “closed-loop” automation framework can lead to 80% reductions in manual interventions as well as improvements in network performance up to 35% [8]. The learning aspect of “closed-loop” systems reinforces the importance of the human validation process, as mentioned previously, simultaneously bringing down the number of recurring issues and improving prediction algorithms and even proactive resolutions [8]. However, all these improvements are highly dependent on the interfaces through which users consume the information necessary to make decisions, spot errors, or predict anomalies. Research shows, as does empirical evidence from our history as a company, that well designed interfaces have a significant impact in optimizing the efficiency of the telco administrator or operator, namely by implementing advanced visualization techniques that make it clearer to the user where the errors and incidents are, reducing decision-making time, and make configurations intuitive and clear so as to avoid errors and support requests [8], [14].

In the experience design industry, some paradigms are slightly shifting to embrace new usability concerns that come up with the introduction of AI features in products. Methodologies such as HCAI [14] and Human–AI interaction (HAX) [17], evolutions of Human–Centered interaction (HCI), and Trustworthy AI (TAI) [14] have been introduced in relation to topics of “user trust and socioethical characteristics to overcome potential AI threats and vulnerabilities” [14]. Studies point to a misalignment between AI development and user trust in AI, due to reasons mentioned before, such as transparency and reliability, but they also emphasize social and ethical considerations as fundamental for building TAI. This is a very common consideration in the user experience work, as it is the base for the first step in the Design Thinking methodology – empathize, where designers deep-dive in the context of their users’ lives, their problems and needs, understand their motivations, map their journeys, etc. In HCAI, the empathize phase should consider both cultural influences, by understanding humans’ intentions towards AI, and their past influences, namely the user’s propensity to trust AI, what threats and concerns can turn into adoption blockers [14].

Figure 3: The interconnections among the four layers of the HCTframe analysis [14]

The scheme presented in **Figure 3**, taken from a recent study in HCAI and TAI, illustrates a framework of building and designing for user trust throughout the entire design process [14].



For AI to succeed, it must reflect users' expectations, perceptions, and the realities of their daily work

Designing for AI features and products must consider the human element – including users' attitudes, expectations, perceptions – as well as their social and organizational context, such as responsibilities in the company/organization, fairness, and non-discrimination. Alongside these, the technical aspects of the system are equally important. In developing TAI, as with almost everything else, the first interaction is critical and can determine adoption success. However, we cannot neglect the fact that trust is also built, maintained, and enhanced over time [14].

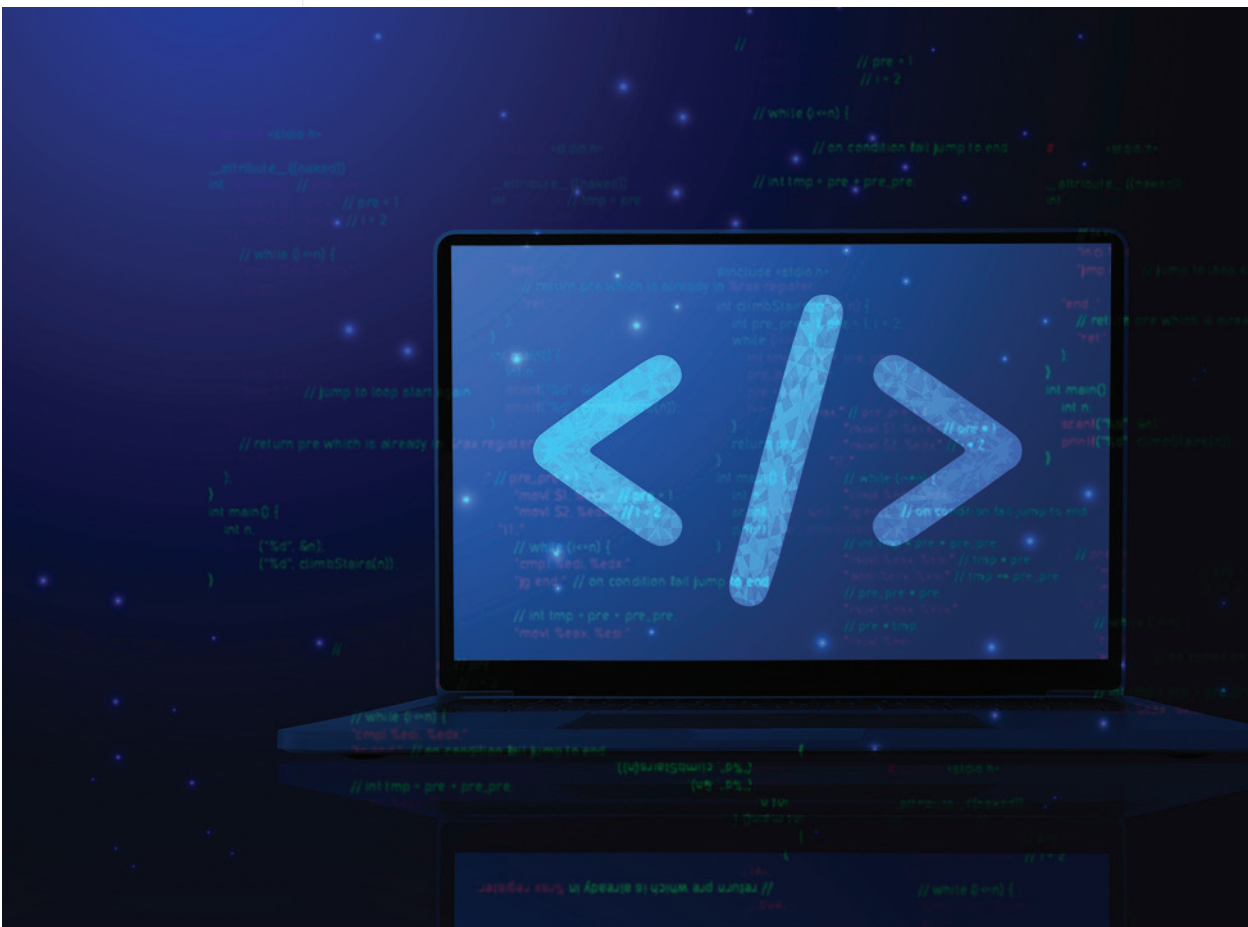
Microsoft has released a list of principles that organizations using AI in their products should consider to fill this blank space left by the rapid AI developments [18]. These principles are as follows [18]:

1. Privacy and security
2. Reliability and safety
3. Accountability
4. Inclusiveness
5. Transparency
6. Fairness

It is clear and interesting how these principles match the user concerns that reduce AI adoption that we have explored throughout this article, which clearly provides designers and product developers with focus points to create great user experiences in their products. Following these principles, Microsoft went a step further and developed them into guidelines for designing Human-AI interaction (HAX), as follows [17]:

1. Make clear what the system can do
2. Make clear how well the system can do what it can do
3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did
12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions
17. Provide global controls
18. Notify users about changes

If we analyze these principles attentively, we can clearly see that they are born from Jakob Nielsen's ten usability heuristics. These are universally followed principles for designing interactions, published in 1994 but still as relevant today, as they were thought of to be "broad rules of thumb, and not specific usability guidelines" [19]. Microsoft has used them as a base to specify these principles that can help designers create positive user experiences with AI products and features [17], [18]. The most important heuristics to remember, as they are the ones most reflected in these new considerations, are #1 Visibility of System Status and #3 User Control and Freedom – with this meaning no disregard for the eight remaining heuristics, as they are all considered in these principles in some way [19]. However, these two will be crucial to building TAI. The first (Visibility of System Status) is focused on communicating what is happening with the system, its status, task progress, reasoning, history, and any other relevant information. This heuristic states that "when we understand the system's state, we feel in control – we can rely on the system to act as expected in all circumstances", meaning that communication and predictability create trust [19]. The third heuristic (User Control and Freedom) is all about allowing the user to quickly get out of a wrong choice or action, which allows experimentation [19]. This is obviously crucial in new AI features and products, as these are new technologies and new types of interactions with an expected learning curve, so users must feel confident in knowing they can undo or redo an action, go back a step, backtrack on mistakes, so that they can explore the interface freely and in an enjoyable way.



AI offers immense potential for automation, smarter decisions, and streamlined workflows, but only if supported by trust, ethics, and human-AI synergy

Conclusion

The telecommunications industry is undergoing a profound transformation, driven by the increasing complexity of network operations and the evolving profile of system administrators. In this context, the adoption of LC/NC platforms represents a pivotal advancement in making configuration tools more accessible, intuitive, and inclusive. By abstracting technical complexity and enabling users with varying levels of expertise to perform sophisticated tasks, LC/NC solutions not only enhance user autonomy but also reduce operational bottlenecks and support overhead. The Expression Editor case study exemplifies how thoughtful design can bridge the gap between technical depth and usability, empowering a broader spectrum of users to contribute meaningfully to mission-critical operations.

However, as user expectations continue to evolve – particularly with the rapid proliferation of AI technologies – the next imperative lies in integrating AI-driven capabilities that are not only powerful but also trustworthy and transparent. The promise of AI in automating operations, enhancing decision-making, and streamlining workflows is immense, yet its success hinges on user trust, ethical design, and seamless human-AI collaboration.

Frameworks such as HCAI, Trustworthy AI, and Closed-Loop Automation offer valuable guidance in this endeavor, emphasizing the importance of transparency, user control, and contextual understanding. By embedding these principles into the design and development of our products, we can ensure that AI features are not perceived as opaque or unpredictable, but rather as reliable partners that augment human capabilities.

At Altice Labs, our commitment is to continue advancing the usability and intelligence of our solutions, fostering a future where complex configurations are not a barrier but an opportunity for innovation, inclusivity, and operational excellence.

References

- [1] T. Xu, V. Pandey, and S. R. Klemmer, “An HCI view of configuration problems,” 2016, arXiv:1601.01747.
- [2] I. Scales. “How low can telcos go?,” TelecomTV, Oct. 31, 2022. Accessed: Sep. 5, 2025. [Online]. Available: <https://www.telecomtv.com/content/digital-platforms-services/how-low-can-telcos-go-45830/>
- [3] Mendix, “Low-code software development: From crisis technology to core technology, says new Mendix-Sponsored survey,” Mendix Press Release, Sep. 19, 2022. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.mendix.com/press/low-code-software-development-from-crisis-technology-to-core-technology-says-new-mendix-sponsored-survey/>
- [4] Y. Luo, P. Liang, C. Wang, M. Shahin, and J. Zhan, “Characteristics and challenges of low-code development: The practitioners’ perspective,” in *Proc. ACM/IEEE Int. Symp. Empir. Softw. Eng. and Meas. (ESEM)*, Bari, Italy, Oct. 2021, doi: 10.1145/3475716.3475782.

- [5] Gartner, Inc., "Risk and opportunity index: low-code application platforms," Gartner, Accessed: Oct. 3, 2025. [Online]. Available: <https://www.gartner.com/en/documents/5459763>
- [6] Readiness IT, "Accelerating telco industry transformation with low-code solutions," Readiness IT. Accessed: Sep. 5, 2025. [Online]. Available: <https://readinessit.com/insights/low-code-solutions-telco-industry/>
- [7] H. White, "A low-code future for telecom," ISE Mag, Accessed: Sep. 5, 2025. [Online]. Available: <https://www.isemag.com/network-reliability-testing-and-assurance-cybersecurity-safety/cybersecurity/article/14305002/a-low-code-future-for-telecom>
- [8] K. K. Pasunuri, "Collaboration between Humans and AI in Telecom Network Automation," *Int. J. Sci. Res. Appl.*, vol. 14, no. 1, pp. 766-772, Feb. 2025, doi: 10.30574/ijrsra.2025.14.1.0121.
- [9] Cisco, "Intent-based networking," Cisco, Accessed: Oct. 3, 2025. [Online]. Available: <https://www.cisco.com/site/us/en/solutions/intent-based-networking/index.html>
- [10] S. Sharma, "How can AI UI capture intent?," UX Collective, Accessed: Sep.30, 2025. [Online]. Available: <https://uxdesign.cc/how-can-ai-ui-capture-intent-d7a6d0393ded>
- [11] Guac Design, "AI + UX design: Complete guide to building trust in AI products," Guac Design, Accessed: Sep. 5, 2025. [Online]. Available: <https://www.guac.design/blog/ai-ux-design-complete-guide-to-building-trust-in-ai-products>
- [12] D. Walsh, "Why people favor AI in certain domains but not others," MIT Sloan School of Management, Accessed: Sep. 12, 2025. [Online]. Available: <https://mitsloan.mit.edu/ideas-made-to-matter/why-people-favor-ai-certain-domains-not-others>
- [13] S. Sousa, J. Cravino, P. Martins, and D. Lamas, "Human-centered trust framework: An HCI perspective," 2023, *arXiv:2305.03306*.
- [14] X. Qin, et al., "AI aversion or appreciation? A capability-personalization Framework and a meta-analytic review," *Psychological Bulletin*, vol. 151, no. 5, pp. 580-599, 2025. doi: 10.1037/bul0000477.
- [15] T. A. Bach, A. Khan, H. Hallock, G. Beltrão, and S. Sousa, "A systematic literature review of user trust in AI-enabled systems; A HCI perspective," *arXiv:2304.08795*.
- [16] Edelman, "2024 Edelman Trust Barometer: Key insights around AI," Edelman Research Report, Mar. 2024, Accessed: Sep. 12, 2025. [Online]. Available: <https://www.edelman.com/sites/g/files/aatuss191/files/2024-03/2024%20Edelman%20Trust%20Barometer%20Key%20Insights%20Around%20AI.pdf>
- [17] Microsoft, "HAX Design Library," Microsoft, Accessed: Sep. 30, 2025. [Online]. Available: https://www.microsoft.com/en-us/haxtoolkit/library/?content_type%5B0%5D=guideline

- [18] Microsoft, "Responsible AI standard: Marketing assets and guidelines," Microsoft Global Design Center, 2024, Accessed: Sep. 12, 2025. [Online]. Available: <https://marketingassets.microsoft.com/gdc/gdct!Njpb/original>
- [19] J. Nielsen, "10 usability heuristics for user interface design," Nielsen Norman Group, Accessed: Sep. 29, 2025. [Online]. Available: <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [20] B. Shih, "I tried 50 AI tools—here's how AI is changing the way we design," UX Collective, Accessed: Oct. 3, 2025. [Online]. Available: <https://uxdesign.cc/i-tryed-50-ai-tools-heres-how-ai-is-changing-the-way-we-design-65f82aa87164>

08

A SHAP-Driven Framework for Misclassification Analysis and Feature Optimization in ML Models

Abstract

The increasing complexity of high-performing machine learning (ML) models often results in “black-box” systems, whose decision-making processes are opaque, particularly when they misclassify instances. This article introduces a unified, SHAP-driven approach to enhance ML model explainability by first providing deep insights into misclassifications and then leveraging these insights for feature set optimization, including pre-optimized industrial models. The approach begins by examining misclassifications using SHAP values, which help identify patterns and group errors into clusters.

This analysis reveals how specific features contribute to these mistakes and highlights data conditions that contribute to errors. Based on these insights, the method introduces a feature selection strategy that uses SHAP-based scores to refine the model’s feature set. By focusing on the most impactful features, it aims to improve the model’s interpretability while preserving or enhancing its accuracy. A special technique for handling LightGBM models further optimizes this process, ensuring faster results in exchange for performance.

Authors

Henrique Oliveira Silva, Altice Labs, henri1silvafcc@gmail.com

Maria Manuel Castro, Altice Labs, maria-m-castro@alticelabs.com

Luís Cortesão, Altice Labs, luis-m-cortesao@alticelabs.com

Keywords

Explainable AI (XAI), SHAP, Machine Learning, Misclassification Analysis, Feature Selection, Model Interpretability

DOI

10.57650/INNOVATION/2025.08



Introduction : the challenge of opacity in high-stakes AI

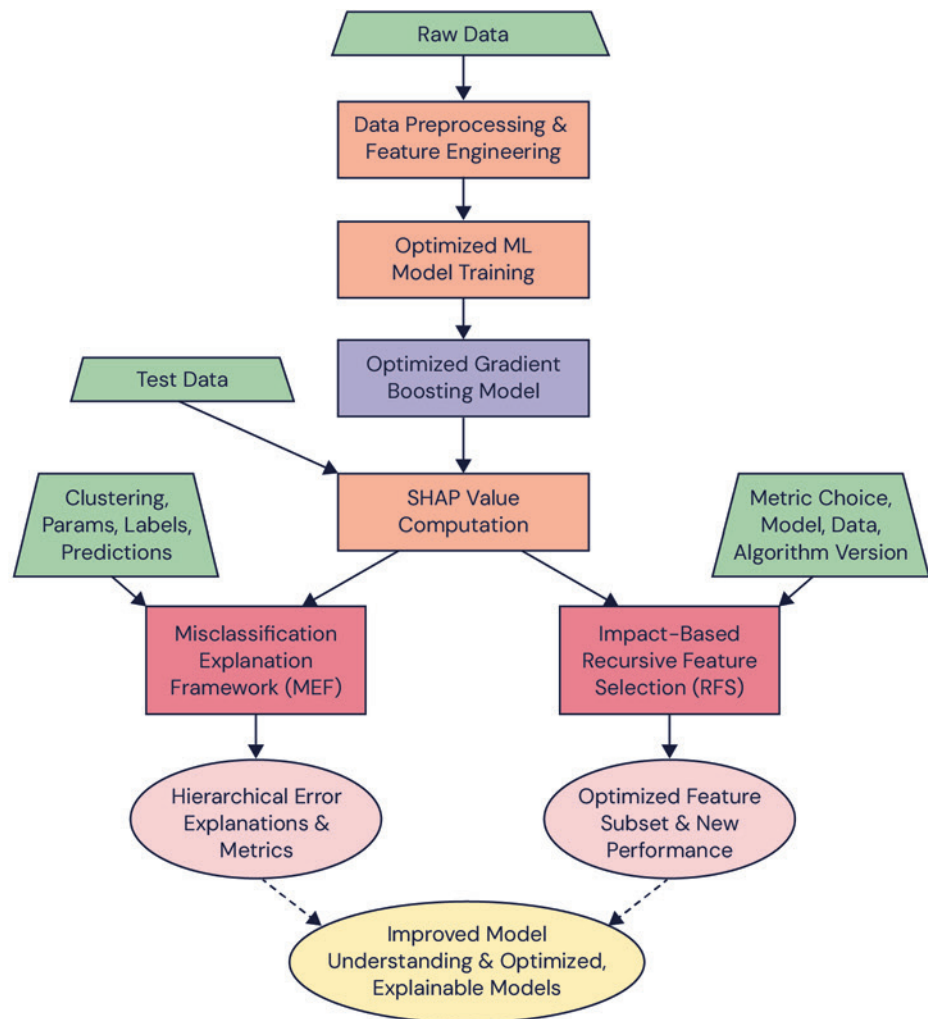
The proliferation of high-performing machine learning (ML) models, such as gradient boosting machines and deep neural networks, has enabled remarkable advances across numerous industries. However, as models become deeper and more complex, their internal decision logic often turns opaque, creating so-called “black-boxes” systems. This opacity is not merely an academic concern; it poses a critical barrier to the responsible deployment of AI in high-stakes domains like healthcare and telecommunications.

The business and engineering implications of this challenge are profound. A lack of transparency erodes user trust, complicates regulatory compliance with mandates like the EU’s General Data Protection Regulation (GDPR) [1] and its “right to explanation”, and hinders effective model lifecycle management. When a model fails — for instance, by misclassifying a critical diagnosis — it is exceptionally difficult to debug the root cause or implement targeted improvements without understanding why the error occurred.

Although Explainable AI (XAI) offers numerous techniques to address this, a persistent gap remains. Many existing methods provide either highly localized, instance-specific explanations that fail to reveal systemic issues or overly broad global summaries of feature importance, which miss the nuanced conditions that lead to specific error types. There remains a need for a systematic methodology that diagnoses coherent patterns of model failure and transforms those diagnostics into actionable improvements.

To bridge this gap, this article presents a unified, SHAP-driven methodology composed of two complementary components: the Misclassification Explanation Framework (MEF) which analyzes model errors through hierarchical clustering and differential SHAP analysis, and the Impact-Based Recursive Feature Selection (RFS) strategy which refines the model's feature set based on a new SHAP-derived "Net Impact" score (Figure 1). Together, these components create a structured pathway from error diagnosis to feature optimization, improving both interpretability and predictive robustness in industrial-scale ML systems.

Figure 1: Overall Conceptual Diagram of the Developed Methodologies



Foundational methodology: quantifying feature contributions with SHAP

SHAP brings transparency to machine learning by turning feature contributions into clear, numeric explanations

Our framework is grounded in a robust and theoretically consistent concept from the field of cooperative game theory: SHapley Additive exPlanations (SHAP) [2]. SHAP provides a unified way to interpret model predictions by assigning a precise attribution value, or "SHAP value", to each input feature for a given decision. This value quantifies the marginal contribution of that feature, indicating how much it pushed the model's output away from a baseline expectation, thereby providing a clear explanation of the prediction's composition.

A major strength of SHAP lies in its ability to generate explanations at multiple granularities. At the local level, it provides a detailed breakdown for a single prediction, revealing the specific drivers of that outcome. By aggregating these local SHAP values across many instances, one can derive global insights, such as a more robust measure of overall feature importance that accounts for feature interactions and non-linear effects. This duality is critical for our two-pronged approach of specific error diagnosis and overall model optimization.

We selected SHAP as the foundational engine for our methodologies due to several key properties that make it particularly suitable for complex industrial applications. First, its strong theoretical guarantees, namely local accuracy (the sum of feature attributions equals the model's output for that prediction), and consistency (a feature's attribution value does not decrease if its contribution to the model's performance increases), ensure that the explanations are both faithful and reliable. Furthermore, the existence of highly efficient implementations, such as "TreeSHAP" for tree-based ensemble models like LightGBM [3] and XGBoost [4], makes its application computationally feasible for the large-scale, industrial models evaluated in this work.

Framework 1: the Misclassification Explanation Framework (MEF)

The first component of the developed methodology is the MEF, a systematic process designed to move beyond aggregate error metrics and provide deep, interpretable insights into the underlying causes of model failures. The primary objective of this framework is to identify and characterize coherent patterns of misclassification, allowing developers to understand not just that errors are occurring, but precisely why and under what data conditions.

Instance clustering for error pattern identification

The MEF starts by isolating a specific confusion pair from the model's performance, such as all instances where the true class was "A" but the model incorrectly predicted "B". To identify meaningful sub-patterns within this error group, the framework employs agglomerative hierarchical clustering. A critical aspect of this step is the use of Gower's distance as the similarity metric. This choice is deliberate, as Gower's distance is specifically designed to handle mixed data types (both numerical and categorical features), which are common in real-world datasets. By clustering misclassified instances based on their original feature values, the MEF effectively groups together errors that occur under similar data conditions, forming distinct error clusters that represent specific failure modes of the model.

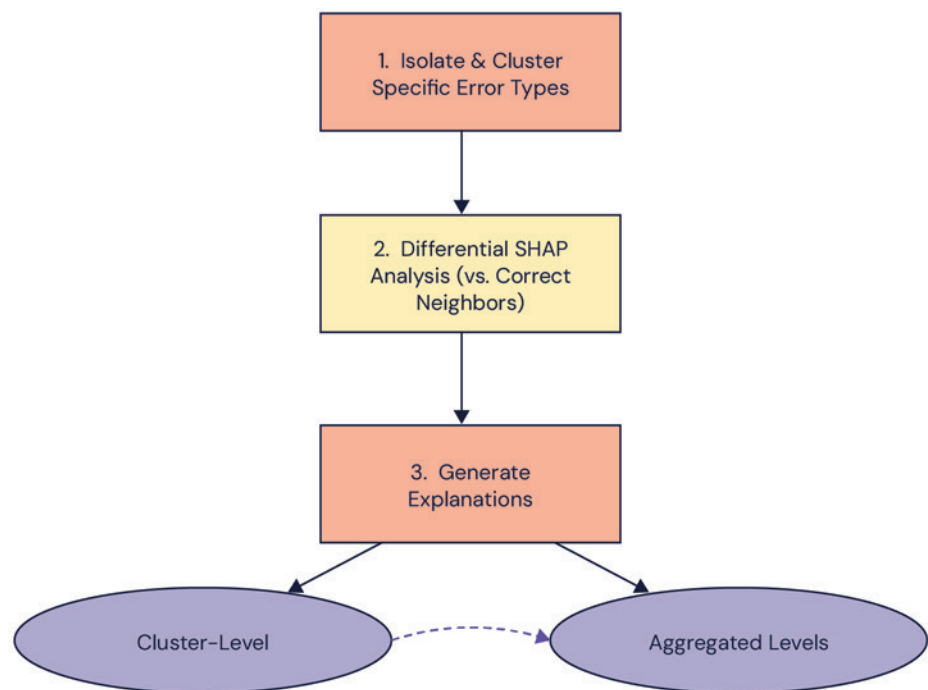
Differential SHAP analysis

Once a cohesive error cluster is identified, the MEF seeks to understand what makes it different from instances the model handled correctly. To achieve this, it

samples a comparative group of “neighbor” instances that share the same true class (“A”) but were correctly classified by the model. These neighbors are selected based on their feature-space proximity to the error cluster, ensuring a relevant and fair comparison.

The core of the diagnostic process is a differential analysis of SHAP values between the error cluster and its correctly classified counterpart. By comparing the mean SHAP values for each feature across these two groups, the MEF pinpoints the features whose influence diverges significantly when the model errs. This analysis reveals the features that are most responsible for driving the misclassification, either by pushing the prediction towards the wrong class or by failing to provide a strong enough push towards the correct one (Figure 2).

Figure 2: Conceptual Diagram of the MEF



Hierarchical explanation generation

The final output of the MEF is a set of hierarchical, human-readable explanations. At the most granular level, for each identified error cluster, the framework generates a report containing two key outputs:

1. **Ranked Feature Impacts:** a list of the features whose differential SHAP analysis revealed them to be the most influential drivers of that specific error.
2. **Localized Rule Conditions:** a set of IF-THEN-like statements derived from the common feature value ranges within the cluster, that describe the data sub-region prone to this specific failure.

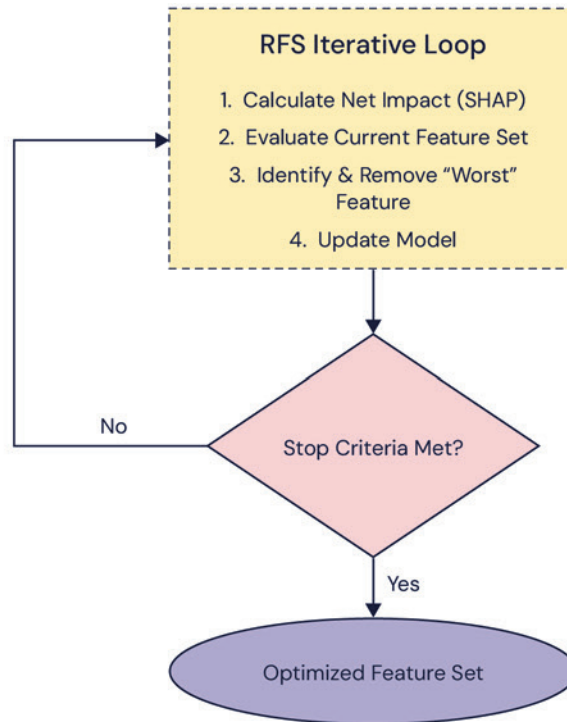
These granular insights are then aggregated to provide higher-level summaries for entire misclassification pairs (e.g., all A → B errors), for all errors originating from a single true class, or globally across all model misclassifications. This hierarchical structure allows stakeholders to perform both deep-dive diagnostics on specific failure modes and high-level analysis of systemic error patterns.

MEF outcomes empower stakeholders to investigate individual failure modes and understand systemic error patterns

Framework 2: the Recursive Feature Selection (RFS) strategy

Building upon the diagnostic capabilities of the MEF, the second component of our methodology is the RFS strategy (Figure 3). The RFS is an automated framework designed for feature set optimization, with the dual objectives of enhancing model parsimony and improving (or maintaining) predictive performance. It moves beyond traditional feature selection methods by leveraging a nuanced understanding of each feature’s contribution to both correct and incorrect predictions.

Figure 3: Conceptual Diagram of the RFS



The “Net Impact” heuristic

The cornerstone of the RFS strategy is a new heuristic called the “Net Impact” (NI) score, which quantifies the overall contribution of a feature to model performance according to a user-defined metric (e.g., accuracy, precision, recall, or F1-score). To calculate the NI score, we first disaggregate a feature’s SHAP values across the entire dataset based on the correctness of the predictions. This creates several SHAP-based impact components, such as “Correct Positive Push” (the sum of positive SHAP values for true positives) and “Misclassified Positive Push” (the sum of positive SHAP values contributing to an incorrect class).

These components are then aggregated in a metric-specific formula to compute the final NI score. For instance, the accuracy-focused NI score for a feature positively weights its contributions to correct predictions while penalizing its contributions to incorrect ones. By tailoring the calculation to a specific performance metric, the RFS allows for targeted optimization aligned with the detailed goals of the modeling task.

RFS iteratively refines machine learning model’s feature set by using a SHAP-derived “Net Impact” score to identify and remove features that least contribute to predictive performance

The recursive elimination algorithm

The RFS strategy employs a recursive elimination algorithm that iteratively refines the feature set. The process begins with a baseline model trained on the full set of features. It then enters an iterative loop where each cycle consists of the following steps:

- 1. Performance Evaluation:** At the start of each iteration, the algorithm computes the NI score for every feature in the current active set. The sum of these scores defines the total NI for the current feature subset, which is logged to a history tracker along with the feature set itself.
- 2. Candidate Selection:** To identify a candidate for removal, the algorithm performs a heuristic evaluation. For each feature, it projects the resulting total NI score if that feature were to be removed. This projection is calculated by summing the NI scores (computed in the previous step) of all other features in the current set. The feature whose removal yields the highest projected total NI is selected as the elimination candidate for the current iteration.
- 3. Elimination and Model Update:** The identified candidate feature is then removed from the active set. The model is subsequently updated to prepare for the next iteration, either through full retraining on the reduced feature set (Standard RFS) or via approximation.

This iterative process continues until a stopping criterion is met, such as reaching a minimum number of features or when no hypothetical removal improves the projected NI score. Crucially, because the elimination at each step is based on a heuristic projection that cannot fully account for complex feature interactions, the algorithm does not assume the final feature set is optimal. Upon termination, it reviews the entire logged history and returns the feature subset that achieved the highest actual total NI score during any iteration of the process.

To address the computational cost of full model retraining, we also developed an efficient “Tree-Dropping” variant for LightGBM models. This variant performs the model update step by leveraging the initial full model and dynamically selecting only the decision trees that remain valid after a feature’s removal, significantly reducing runtime at the cost of an approximated evaluation.



Empirical validation on real-world industrial models

Table 1: Summary of Real-World Industrial Datasets used for Empirical Evaluation

To evaluate the practical utility and generalizability of our frameworks, we conducted an empirical evaluation of three distinct, real-world industrial datasets from Altice Labs. The datasets, drawn from healthcare and telecommunications domains, differed in scale, feature composition, and problem complexity. Importantly, our methodologies were applied to pre-existing, highly optimized LightGBM and XGBoost models, providing a stringent test of their ability to add value beyond standard optimization practices. The key findings are summarized in **Table 1**.

#	Domain	Classification Task	Size	# Features	Features Types	# Classes
1	Healthcare	Diagnosis of "Disease X"	974	50	48 numerical, 2 categorical	3
2	Telecommunication	Need for technical intervention (binary)	98999	19	18 categorical, 1 numerical	2
3	Telecommunications	Failure Cause (multiclass)	193739	17	All categorical	305

Finding 1: quantitative performance of the MEF

The MEF demonstrated a consistent ability to generate explanations for a significant proportion of model errors across all datasets. To move beyond simple coverage counts and formally evaluate the quality of these explanations, we introduced two new metrics: Adjusted Explanation Coverage (AEC), which penalizes the looseness of the underlying clustering; and the AEC Efficiency Ratio, which measures the retained explanatory power after penalization.

A sensitivity analysis on the large telecommunications dataset (Dataset C), which contained nearly 20,000 misclassified instances, revealed a clear and important trade-off, as summarized in **Table 2**. When using relaxed clustering parameters, the framework achieved a very high raw coverage of 86.4%. However, the penalty for this looseness reduced the AEC score to 0.6174, resulting in a corresponding AEC Efficiency Ratio of 0.7143. This indicated that while many errors were covered, their explanations were derived from less cohesive, and therefore less semantically precise, error clusters.

In contrast, by selecting tighter clustering parameters, the framework achieved a slightly lower raw coverage of 78.5%. Despite explaining fewer instances, this configuration was penalized less severely due to the high cohesion of its clusters, yielding a superior AEC score of 0.6545. This translated to a significantly higher AEC Efficiency Ratio of 0.8333. This quantitative comparison validates the MEF's ability to not only characterize large volumes of model failures but also to provide a tunable mechanism for balancing the breadth of coverage with the diagnostic precision of the generated insights. For detailed error analysis, prioritizing the configuration with the higher AEC and Efficiency Ratio provides more reliable and actionable results.

Table 2: MEF Sensitivity Analysis on Dataset C. Trade-off Between Coverage and Precision/Reliable explanations

Configuration	Explained Instances	Std. Coverage	AEC	AEC Efficiency Ratio
Relaxed Clustering	17.203	0.8643	0.6174	0.7143
Tighter Clustering	15.631	0.7854	0.6545	0.8333

Finding 2: actionable diagnostic value of the MEF

The ultimate test of an explanation framework is its ability to generate actionable insights for domain experts. To this end, we conducted a structured qualitative review of the MEF’s outputs with a telecommunications specialist, focusing on the highly complex, 305-class model from Dataset C. The session was designed to simulate a real-world model auditing scenario.

First, the specialist was provided with an introductory guide explaining how to interpret the framework’s output. Subsequently, they were presented sequentially with three distinct, high-efficiency cluster explanations. The diagnostic power of the MEF was most strikingly demonstrated with one specific explanation. It detailed a common error where tickets with a true cause of “Unidentified Root Cause” were being misclassified as “Physical Infrastructure Issue”. The MEF provided the following (masked) details for a large cluster of 781 such errors:

Top Impactful Features for this Error:

- SERVICE CATEGORY: +237% vs. average
- INFRASTRUCTURE FLAG: +140% vs. average
- ASSIGNED OFFICE: +110% vs. average

Example Rule Generating this Confusion:

IF

- SERVICE CATEGORY is “GPON”
- ORGANIZATIONAL GROUP is in [“OLT Team”, “PON”]
- SPECIFIC PROBLEM is in [“Out of Service DISS”, “Out of Service DST DISS”]
- CUSTOMER IMPACT is “Confirmed”
- INFRASTRUCTURE FLAG is “Missing”
- EQUIPMENT TYPE is “MSAN”
- VENDOR is in [“ALCATEL”, “PTIN”]
- CARD TYPE is in [“AC16SXG”, “GLT4-A”, “TG16GF”]
- RESPONSIBLE TEAM is “Missing” or “NOC CorpNet”

THEN → Model predicts “Physical Infrastructure Issue”

The specialist’s diagnostic process, triggered by this detailed rule, was immediate and insightful. They recognized that the combination of “SERVICE

CATEGORY”, “EQUIPMENT TYPE”, and “VENDOR” described a very specific hardware context. However, these were only secondary indicators. The crucial insight came from what was not present. The specialist explained that for this context, there should have been an automated system flag, a feature that was completely missing from the dataset, which would have correctly categorized the issue as “Unidentified” root cause.

In the absence of this critical flag, the model was forced to make a “best guess” based on the available, correlated hardware features, leading it to consistently default to the incorrect “Physical Infrastructure Issue” prediction. The MEF did not just explain the model’s behavior; its detailed rule provided the exact context needed for the expert to identify a fundamental data deficiency. This finding is particularly significant because the model and dataset had already undergone extensive prior revision, yet this critical data gap remained undiscovered. The session underscored the MEF’s capability to function as a powerful diagnostic instrument, enabling experts to translate a model’s statistical patterns into actionable insights about data quality and system design.

Finding 3: RFS for performance enhancement and validation

The RFS strategy demonstrated its dual utility as both an optimization and a verification tool across the different datasets, with its effectiveness depending on the maturity of the initial feature set.

For the healthcare classification model (Dataset A), which started with a relatively large set of 50 features, the standard RFS framework consistently identified opportunities for improvement. The results, summarized in **Table 3**, show that when guided by the accuracy metric, the RFS selected a more parsimonious 45-feature subset. This led to a statistically significant increase in both average accuracy (from 0.809 to 0.832) and weighted F1-score (from 0.797 to 0.817). This result highlights the RFS’s ability to refine even previously optimized models by identifying and removing features with a subtle negative impact on generalization.

Table 3: RFS Results on Dataset A (Healthcare). Both RFS variants outperform baseline model

Model Version	# Features	Accuracy	Weighted F1-Score
Baseline (Original)	50	0.809	0.797
RFS Standard (Accuracy-Guided)	45	0.832	0.817
RFS Tree-Dropping (Accuracy-Guided)	40	0.821	0.813

Furthermore, the computationally efficient “Tree-Dropping” variant for this dataset identified a 40-feature subset that still outperformed the original 50-feature baseline, achieving an average accuracy of 0.821 and a weighted F1-score of 0.813. While marginally less performant than the standard RFS result, it delivered

this improvement while completing its run substantially faster (22.28s vs. 28.28s). This validates its utility for rapid exploration and for scenarios where computational resources are constrained, offering a clear trade-off between peak performance and efficiency.

Conversely, when applied to the two highly tuned telecommunications models (Datasets B and C), which had already been reduced to lean sets of 19 and 17 features, respectively, the RFS framework consistently concluded that the original feature sets were optimal. In all configurations, the algorithm determined that no single feature removal would be beneficial according to its Net Impact heuristic. This “no-change” outcome is a significant finding, serving as a robust, data-driven validation of the existing models’ parsimony and effectiveness. It showcases the framework’s value not only as an optimization tool but also as a powerful verification instrument for providing confidence in mature, production-level ML systems.

Discussion: implications and future work

The empirical findings from this research carry significant implications for both the practical application of machine learning and the advancement of the XAI field. The developed frameworks offer a structured approach to move beyond surface-level performance metrics towards a deeper, more actionable understanding of model behavior.

Practical implications for ML operations

For data scientists and ML engineers, the two frameworks provide complementary tools to enhance the model lifecycle. The MEF serves as a powerful diagnostic instrument, enabling targeted debugging by pinpointing the specific data conditions and feature interactions that lead to errors. As demonstrated by our qualitative evaluation, this can directly inform data curation and feature engineering efforts. The RFS strategy offers a data-driven method for model simplification and optimization. Its ability to produce more parsimonious models without sacrificing – and, in some cases, improving – performance leads to systems that are not only more efficient but also inherently easier to audit and maintain.

Methodological contributions to XAI

From a research perspective, this work contributes three novel elements to the Explainable AI field. These contributions were originally developed as part of a Master’s thesis [5]. First, the MEF introduces a structured, hierarchical methodology for misclassification analysis that integrates instance clustering with differential feature attribution. Second, we propose the AEC and AEC Efficiency Ratio as new metrics for evaluating the quality of clustered explanations, explicitly accounting for the trade-off between explanation coverage and the precision of the underlying instance groups. Finally, the “Net Impact” score,

developed for the RFS strategy, represents a new metric-specific heuristic for SHAP-informed feature selection, offering a targeted alternative to methods based solely on global feature importance.

Future directions

While this research establishes a strong foundation, several promising avenues for future work remain. A primary direction is the extension and validation of these frameworks across a broader range of model architectures, particularly deep neural networks, where feature attribution and interaction are equally critical but more complex to dissect. Furthermore, conducting more extensive and formalized user studies with diverse panels of domain experts would be invaluable for rigorously assessing the clarity, actionability, and ultimate impact of the generated explanations across different contexts. Finally, continued optimization of the computational performance of both frameworks, perhaps by exploring more efficient clustering algorithms or SHAP approximation techniques, is an important direction to enhance their scalability for extremely large datasets or near real-time analysis requirements.



Conclusion

The increasing prevalence of complex machine learning models in critical sectors has instigated an urgent need for methods to render their decision-making processes more transparent. This work confronted the “black-box” challenge by developing and validating a unified methodology to enhance the explainability of classification models, focusing specifically on the analysis of misclassifications and the optimization of feature sets. Our research was grounded in the principle of leveraging a deep, nuanced understanding of feature contributions, derived from SHAP values, to both correct and incorrect predictions.

We introduced two primary contributions: the MEF and the RFS strategy. The MEF provides a multi-faceted diagnostic tool that combines instance clustering and differential SHAP analysis to generate hierarchical insights into why models make specific errors. The RFS strategy, in turn, operationalizes these insights, using a novel “Net Impact” heuristic to iteratively identify and prune detrimental features, thereby improving model performance and parsimony.

The empirical evaluations across diverse, real-world industrial datasets confirmed the practical utility of these frameworks. The MEF demonstrated its value

as a powerful diagnostic instrument, allowing specialists to rapidly identify actionable data deficiencies. Concurrently, the RFS strategy successfully refined an already-optimized model by improving its predictive performance with a reduced feature set and validated the robustness of other highly tuned models. By offering structured approaches to dissect model failures and optimize model inputs, this research provides a tangible step towards making complex machine learning systems more transparent, accountable, and reliable. The continued development and adoption of such explainability techniques is paramount as AI systems become increasingly integral to decision-making in society.

References

- [1] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," *Harv. J. Law & Technol.*, vol. 31, no. 2, pp. 841–887, 2017.
- [2] S. M. Lundberg and S.-In Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA, Dec. 2017, pp. 4768–4777.
- [3] Y. Shi et al., *lightGBM*: A fast, distributed, high-performance gradient boosting framework, GitHub repository, 2025. [Online]. Available: <https://github.com/Microsoft/LightGBM>
- [4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining (KDD'16)*, New York, NY, USA, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785
- [5] H. O. Silva, "Explainability in machine learning models: A shap-driven framework for misclassification analysis and feature selection," M.S. thesis, Fac. Eng., Univ. Porto, Porto, Portugal, 2025. [Online]. Available: <https://hdl.handle.net/10216/168350>



09

Network Configuration and Operation via Intent-Based Networking

Abstract

Intent-Based Networking (IBN) introduces a higher level of abstraction to network control by focusing on high-level objectives (the what) while delegating low-level configuration (the how) to automated processes. Evolving from policy-based network management, IBN adds a high-level layer that allows business requisites to be expressed in a human-understandable format. These intents are automatically interpreted, eventually decomposed, and implemented by network management entities, enabling automated configuration and operations. By abstracting network behaviour into intent-based policies, IBN simplifies management, improves efficiency, accelerates service deployment, and reduces human error. Network operators can specify high-level expectations for quality of service, security, or performance without dealing with configuration complexity. Monitoring control loops check fulfillment, according to the expected outcomes expressed in the Intents.

IBN is expected to rely heavily on Artificial Intelligence (AI) and Machine Learning (ML) to validate Intents, translate them into actions, and verify fulfillment. Recent research has been particularly focused on using Large Language Models (LLM) for this translation.

IBN concepts began emerging around 2014 with early Software Defined Networking (SDN) platforms. Today, 3GPP has a mature IBN strategy [1], [2] for both RAN and core domains, and the Broadband Forum¹ is developing related work through a dedicated Tiger Team, which Altice Labs is following. Within the P2030 ALBATROZ project, IBN will be implemented in the AGORA SDN Management and Control (M&C) framework and assessed in selected use cases.

Authors

Francisco Fontes, Altice Labs, fontes@alticelabs.com

Rui Calé, Altice Labs, cale@alticelabs.com

Keywords

Intent, AI, LLM, TM Forum, 3GPP

DOI

10.57650/INNOVATION/2025.09

¹ <https://www.broadband-forum.org/>

IBN relies on AI and ML to turn high-level intents into precise, dependable network actions

Introduction

In the early Internet times (pre-2010s), network management was largely manual, relying on Command Line Interface (CLI) configurations. In the Telco domain, Simple Network Management Protocol (SNMP) became the reference protocol and architecture as a first step into distributed and model-based network management, relying in agents operating at managed devices and enabling managers to build increasingly complete and operable models of the network. Concepts like policy-based management and network automation began to emerge, laying the groundwork for future abstraction. With the rise of SDN (2010–2015), centralized control and programmability were introduced, separating the control and data planes. SDN inspired the idea of expressing what the network should do, rather than how to do it, an early form of “intent”. IBN emerges around 2015 to 2017, when vendors like Cisco began promoting IBN as a next-gen evolution of SDN. Cisco’s Digital Network Architecture (DNA) was one of the first commercial implementations of IBN. IBN was defined as a system that:

- Accepts high-level business goals (intents);
- Translates those business goals into policies;
- Automates deployment and continuously verifies its compliance.

As a result of the growing interest in IBN and consolidating existing work, the Internet Engineering Task Force/Internet Research Task Force (IETF/IRTF) published Request for Comments (RFC) 9315 [3] in 2022, formally defining “intent” in networking. At the same time, academic and industry research expanded and enriched the concept, exploring AI and ML integration, Natural Language Processors (NLP), and closed-loop automation.

More recently, IBN is being increasingly integrated with AI, using ML and natural language interfaces, resulting in IBN making the first steps into Telecoms and cloud providers’ operations to manage complex, dynamic environments, especially in 5G, Internet of Things (IoT), and Edge Computing (EC). Current Autonomous Networks (AN), a long-term pursued target, build on IBN principles to self-configure, self-optimize, and self-heal. In that scope, TM Forum defined Intent Driven Autonomous Networks (IDAN)².

AN emerges in a context where operators pursue scale, reliability, and efficiency. Achieving a high rate in the three areas has been proved hard. While scale and reliability are common today, this is normally achieved at the cost of efficiency and waste of resources. As referred, IBN associated with emerging AI technologies may be the missing element to achieve highly efficient autonomous networks. Only by moving beyond policies and task-driven automation, toward intent-based autonomy, where business intents guide every phase of the network lifecycle, can network operation complexity evolve beyond human capacity to manage it using common tools efficiently. This aligns with the growing relevance of communication networks, particularly those based on wireless connectivity, in a very competitive scenario. In the face of an observable

² <https://www.tmforum.org/catalysts/projects/C23.0.563/intentdriven-autonomous-networks-phase-iii>

decrease in Average Revenue Per User (ARPU), deployed resources must be used with maximum effectiveness.

Building on 5G, 6G is undergoing initial studies under 3GPP auspices and is planned to emerge around 2030 for commercial usage. It is expected to incorporate automation in most of its operations to achieve total autonomy. 5G already includes components and features aimed at AI exploitation, like the Network Data Analytics Function (NWDAF), in Core scope, and Service Management and Orchestration (SMO), in the RAN scope. 6G will go one step further, as it will possess:

- **AI-Native Architecture:** AI will be embedded at every layer, from radio access to service orchestration, for predictive and adaptive networking.
- **Self-Learning Network:** Radios and nodes will autonomously learn from their environment how to optimize communication parameters.

This aligns with IBN. Current trends show a strong dependency of IBN on ML and AI to translate Intents into lower and technical instructions, and, thereafter, to guarantee that committed expectations are met.

TM Forum and 3GPP, as the entities with the most advanced and relevant work in the area, see their work being used as main references, with the former focusing on closed-loop control and management, and the latter on the network mechanisms and architecture.





IBN shift networks from manual commands to goal-driven orchestration, defining outcomes without prescribing steps

Intents

Definition

Intents, as the name implies, are declarative instructions expressing what is intended to be performed, without stating how it shall be done, in opposition to imperative ones. It is up to the Intent interpreter, or engine, to translate Intents into instructions and actions, suitable to accomplish the expressed intents' expectations.

RFC 9315 helps distinguishing Intent from traditional Policy, aiming to standardize terminology and guide future research and engineering efforts. In this RFC, an Intent is defined as a set of operational goals (that should be met) and outcomes (that are supposed to be delivered), defined in a declarative manner, without specifying how they should be achieved or implemented. Thus, Intents are relevant and exist at the network and services layer, not individual devices.

Operational goals are defined at a high level, without specifying how those outcomes should be achieved or how goals should specifically be satisfied, and without the need to enumerate specific events, conditions, and actions. The following are two simple examples of Intents definition in natural language and their respective translation into network configurations.

Example 1: Natural-Language to Enterprise Network Policy

- **Intent (User Expression):**
 - "Block social media access for all employees from 9am to 5pm, but allow YouTube for the marketing team."
- **Translation (Policy Abstraction):**
 - Define source = Employee VLANs

- Define group exception = Marketing VLAN
- Policy: Deny traffic to social media domains from source VLANs; permit access to YouTube for the exception group
- **Resulting Network Configurations:**
 - Cisco Access List (ACL):


```
ip access-list extended BLOCK_SOCIAL_MEDIA
deny tcp any social_media_IPs any time-range WORK_HOURS
permit tcp any youtube_IPs marketing_subnet
!
time-range WORK_HOURS
periodic weekdays 9:00 to 17:00
```

Example 2: 5G Network Slicing Intent

- **Intent (Network Operator):**
 - “Launch a low-latency, high-throughput network slice for autonomous vehicles in zone A.”
- **Translation (Slice Descriptor):**
 - Slice type: enhanced Mobile Broadband/Ultra Reliable Low-Latency Communications (eMBB/URLLC)
 - Location: Zone A cell sites
 - Latency: <10ms, Throughput: >1Gbps
- **Automated Network Configurations:**
 - Core: Establish 5G slice in Session Management Function/Access and Mobility Management Function (SMF/AMF) with relevant slice identifiers and QoS templates
 - RAN: Adjust scheduling, tune resource blocks
 - Edge: Route slice traffic to edge compute nodes

In this second example it is possible to identify the Intent decomposition into different networking domains (Core, RAN, and Edge), where specific configurations will be done by the respective managers, or new domain-specific Intents are defined and sent.

Besides vendor CLI commands, Intents can also be mapped onto models, e.g., using Yet Another Next Generation (YANG), a standardized data modelling language developed by the IETF for defining, manipulating, and validating configuration and state data in network devices and services. For example, the IETF RFC 8299 [4] defines a YANG Data Model for Layer 3 Virtual Private Networks (L3VPN) service delivery, which allows IBN controllers to provision L3VPNs based purely on high-level requirements, automating the creation of complex, multi-vendor configurations.

Lifecycle

To unlock the full potential of IBN, starting from the Intent (i) expression, the network system must support four additional critical functions: (ii) translation, (iii) resolution, including decomposition, (iv) activation and (v) assurance, to ensure seamless validation and automation of user-defined intents. These functions, when described sequentially, are known as the Intent lifecycle:

1. Intent Expression

Who: Intent Owner (e.g. user or business system)

What: High-level goals stated in natural language or structured formats (YAML, JSON)

Example: "Ensure latency < 10ms for video traffic"

2. Intent Translation

Who: Intent Handler (e.g., orchestrator or AI agent)

What: Converts human-readable Intent into machine-executable policies or service orders

Tools: NLP engines and LLM (see section "Usage of LLM in IBN"), TMF 921A Application Programming Interface (API) [5] (see section "TM Forum")

3. Intent Resolution

What: Maps translated Intents to available resources and services

Includes: Conflict detection, feasibility checks, and domain decomposition

4. Intent Activation

What: Executes the resolved Intent across network domains

API Used: TMF641 (Service Order) [6], TMF702 (Resource Activation) [7], 3GPP slice orchestration

5. Intent Assurance

What: Monitors fulfilment and compliance with the original Intent

Includes: Telemetry, SLA tracking, closed-loop feedback

Feedback Loop: Reports success/failure to the Intent owner and refines configuration if needed

Standardization

For the telecom world, two organizations, TM Forum³ and 3GPP⁴, are of especial relevance, guiding networks and systems architecture definition, protocols and

3 TM Forum Home | TM Forum; <https://www.tmforum.com>

4 3GPP – The Mobile Broadband Standard; <https://www.3gpp.org>

Table 1: Summary of TM Forum and 3GPP positioning on IBN

information exchange, both having IBN related complementary work done. While TM Forum emphasizes business-level Intents and service orchestration across domains, often using AI-native toolkits and open APIs, 3GPP focuses on network-level Intents for mobile services, with 5G as their major focus. **Table 1** summarizes TM Forum and 3GPP IBN positioning.

Organization	Focus Area	Role in IBN	Key Contributions	Strengths
TM Forum	Telecom operations, service management	Business and operational layer of IBN	Open Digital Architecture (ODA) with Intent interfaces Intent APIs and lifecycle models Catalyst projects demonstrating real-world IBN use cases	Strong in-service orchestration Business-driven intent modelling Vendor-neutral frameworks
3GPP	Cellular networks (5G/6G)	Network and service layer abstraction	Intent-driven management for 5G Advanced and 6G Intent-CSC and CSP models for service slicing	Deep integration with mobile network functions Standardized interfaces for RAN/core automation

TM Forum

TM Forum is a global industry association focused on digital transformation for telecoms and service providers. It develops open standards, frameworks, and APIs to help operators manage complex networks and services. Its flagship model is the ODA⁵, which promotes modular, interoperable systems.

TM Forum defines an AN as “a network that self-configures, self-heals, self-optimizes and can sense, think, and act”, associating five network autonomy levels, from 0 (fully manual operations) with no automation, to 5 (fully self-running networks) with no human intervention. To help operators consistently measure their progress against these levels, TM Forum has developed evaluation tools and methodologies, such as the Autonomous Network Level Evaluation Tool (ANLET)⁶. Intents appear in TM Forum in the context of the TMF IDAN framework, being defined as a formal specification of expectations, including requirements, goals, and constraints, communicated to a technical system. This enables autonomous systems to adapt their behavior and generate solutions beyond static policies.

⁵ <https://www.tmforum.org/open-digital-architecture/>

⁶ <https://www.tmforum.org/resources/guidebook/gb1059-autonomous-network-level-evaluation-tool-anlet-v1-0-0/>

TM Forum’s Intent-based architecture introduces a structured and systematic approach to autonomous network and service management through the concepts of “Intent Owners” and “Intent Handlers”, carefully organized across a hierarchical model of three distinct layers (Business, Services, and Resources), which are detailed in **Table 2**.

Table 2: TM Forum layers and roles for IBN

Layer	Description	Example Intent Owner	Example Intent Handler
Business	High-level goals tied to customer experience, revenue, or strategic outcomes	CxO, Product Manager	Service Orchestrator
Services	Intent related to service performance, availability, or configuration	Service Manager	Resource Orchestrator
Resources	Low-level infrastructure and network elements that support services	Resource Manager	Network Controller

This layered approach enables closed-loop automation and Intents negotiation between owners and handlers at each level.

Under the TM Forum Intent Ontology (TIO) [8], which introduces the formal models that precisely determine how an Intent and its associated reports are expressed, TM Forum TMF921 defines the standardized Intent Management API, which enables intent-driven automation and efficient management in telecom and digital services environments, with the following key capabilities:

- **Setting Intent:** An “intent owner” (like a business customer) can use the API to specify desired outcomes for a service, such as service duration, quality, or location.
- **Reporting on Intent:** An “intent handler” (typically a network or service management system) uses the API to report on the status and progress of fulfilling that intent.
- **Negotiating Intent:** The API facilitates a dialogue between the owner and handler to negotiate and refine the intent requirements.

TMF 921 is part of the TM Forum Open API⁷ initiative, which creates standardized APIs to improve interoperability and operational agility in the telecommunications and digital services industry. It specifies RESTful API operations for clearly expressing, reporting and negotiating operational intents between an intent owner (e.g., a business stakeholder) and an intent handler (e.g., a network management system).

- **Intent Management:** Its specific purpose is to manage, via Create/Read/Update/Delete (CRUD) operations, the Intent behind a service or network configuration, focusing on what the user or system desires to achieve rather than the low-level technical steps required.

⁷ <https://www.tmforum.org/oda/open-apis/>

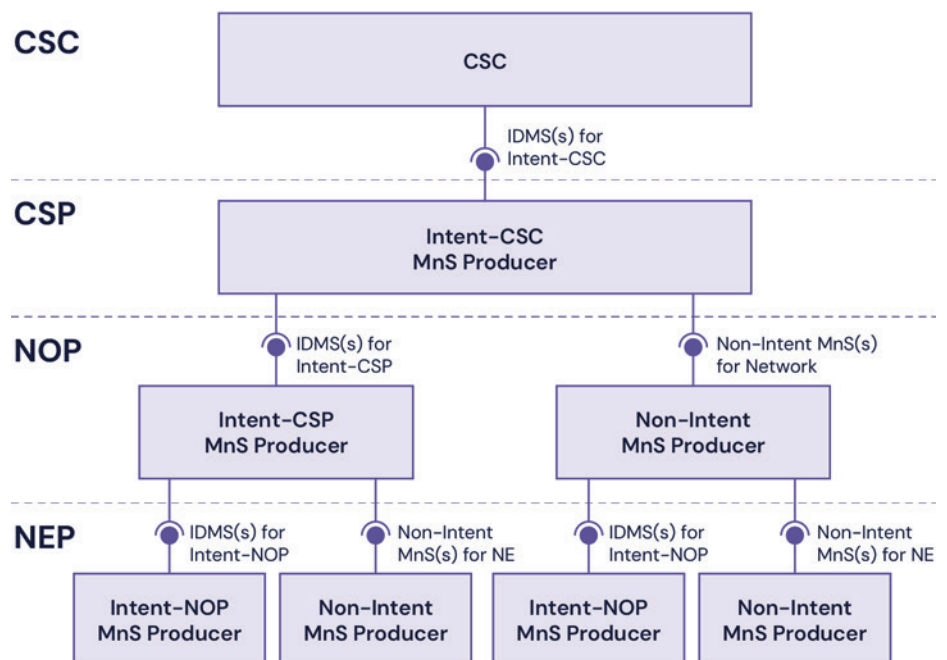
- **RESTful Interface:** It uses the Representational State Transfer (REST) architectural style, a common standard for network application development that utilizes simple Hypertext Transfer Protocol (HTTP) methods like POST, GET, PUT, and DELETE.

3GPP

3GPP presents a consolidated work in the IBN area, reflected in TS 28.312, introducing the Intent Driven Management Service (IDMS), part of the Service Based Management Architecture (SBMA). In 3GPP's SBMA, Management Service (MnS) Producers and Consumers are key concepts used to define MnS interactions, within telecom networks, e.g. in 5G. In particular, TS 28.312 defines a standardized framework for intent-driven management and orchestration in LTE and 5G networks, presenting 3GPP Stage 2 and Stage 3 "definitions for Intent-Based Management". It enables network operators and service consumers to express high-level operational intents, which the management system translates into concrete actions and policies using automation and intelligent orchestration.

Similarly to TM Forum, an Intent in the 3GPP context is a set of expectations (requirements, goals, and constraints) expressed to the network, focusing on "what" must be achieved instead of "how" to achieve it. Taking the roles related to 5G networks and network slicing management⁸, Intents will go through a translation process, from the higher Communications Service Consumer (CSC) layer to execution in the Network Equipment Provider (NEP) lower layer, where the upper layer acts as MnS Consumer and the lower layer as MnS Producer. Such a possible translation process is presented in **Figure 1**.

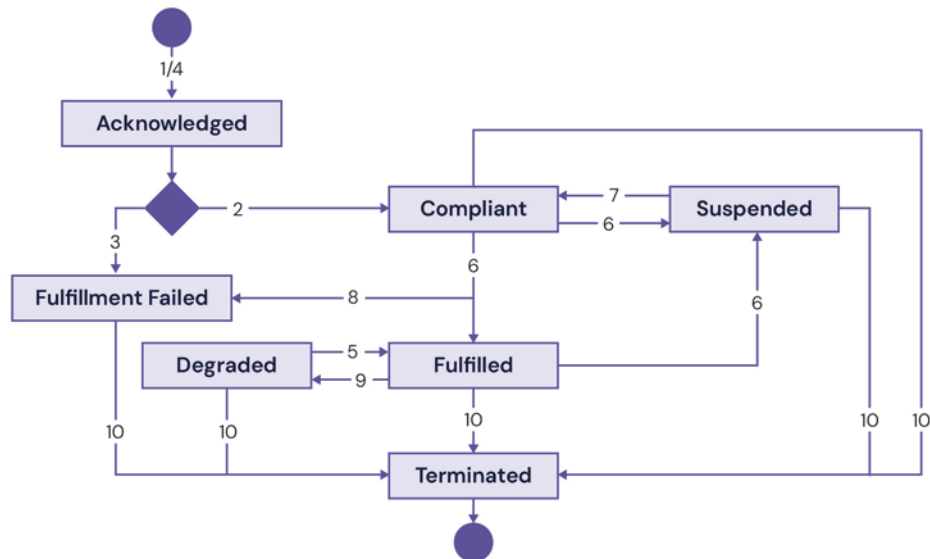
Figure 1: 3GPP hierarchy of MnS Producers and Consumers



⁸ <https://www.tmforum.org/opendigitalframework/intent-based-automation/>

3GPP defines Intent operations (Create, Delete, Modify, Query, Activate, and Deactivate) and identifies Intent states (Acknowledged, Compliant, Fulfilment Failed, Fulfilled, Suspended, Compliant, Degraded, and Terminated), as shown in **Figure 2**. Transition details between states can be found in [1]. From “Acknowledged” (intent instance is created), to “Terminated” (the MnS consumer requested the Intent to be deleted), “Compliant” (feasibility check for the intent is successful) and “Fulfilled” (the MnS producer considers that the intent, expectation or target has been fulfilled as stated by the MnS consumer), are successful Intent intermediate states.

Figure 2: 3GPP defined Intent states



In TS 28.312, 3GPP identifies and describes the following 5G management and orchestration use cases, covering intent-driven management across multiple layers and components of the 5G ecosystem, from radio and edge to core, supporting automation and closed-loop orchestration:

1. **Radio Network Delivery:** Management intents for providing radio network resources in specific geographical areas;
2. **Radio Service Provisioning:** Expectations for delivery and quality of radio services (e.g. voice, data and messaging);
3. **Edge Service Support:** Management of edge services, including deployment and resource allocation close to end users;
4. **Coverage Performance:** Ensuring and optimizing 5G coverage requirements and seamless connectivity;
5. **Radio Network Performance:** Guaranteeing network Key Performance Indicators (KPI) such as throughput, latency, reliability, and capacity fulfilment;
6. **End-to-End Network Optimization:** Management of resources across the entire network to optimize application and service delivery;
7. **RAN Energy Saving:** Expectations for applying energy-efficient techniques in Radio Access Network operations;
8. **Core Network (5GC) Management:** Handling intents related to the configuration and operation of the 5G Core Network.

Table 3: Comparison of TM Forum and 3GPP positioning in IBN based on TMF921 and TS 28.312.

Comparison

As IBN reference documents from TM Forum and 3GPP, organizations' positioning can be made based on the analysis of TMF921 and TS 28.312. As shown in **Table 3**, their differences stem from their different scopes, focus areas, and abstraction levels, despite their complementary roles in intent-driven network management:

	TM Forum TM921	3GPP TS 28.312
Scope and Domain Focus	Targets broad intent management API applicable across multiple network domains and critical business processes in telecom OSS/BSS environments	Focuses specifically on mobile network orchestration and management , including radio access, core network, and edge, with details aligned to 3GPP mobile system architectures
Abstraction Levels and Models	Defines API operations and behavior for intent expression, negotiation, and lifecycle management but has less emphasis on detailed network function models	Provides comprehensive information models and data structures describing mobile network components and capabilities, supporting closed-loop automation with finer granularity
Standardization Maturity and Adoption	APIs have broader adoption in OSS/BSS integration and diverse commercial API ecosystems	Is newer and more focused on mobile network vendors and 3GPP-centric orchestration solutions
Intent Handling Features	Emphasizes dynamic interaction and negotiation flows between intent users and handlers but ultimately relies on underlying implementations for those specifics	Incorporates more detailed mechanisms for conflict resolution, intent feasibility checking, and network-specific intent fulfilment aligned with 3GPP orchestration standards
Cross-domain Integration	Is designed for cross-domain, multi-vendor orchestration , comprehensively including BSS, OSS, and network layers	Covers mainly mobile network domains , with ongoing work needed to harmonize with broader OSS/BSS frameworks

In essence, while TMF921 provides the standardized intent API interface for broad telecom scenarios, 3GPP TS 28.312 delivers detailed mobile network intent orchestration models. Work remains to fully harmonize their data models, expand cross-domain use cases, and unify intent handling approaches for comprehensive autonomous network management.

AI role in IBN

AI plays a foundational role in IBN by enabling automation, intelligent translation of network intents, continual adaptation and self-optimization across complex

networking environments. IBN requires continuous decision-making, adaptation, and optimization, natural domains for AI, bringing the following benefits:

- **Simplifies Service Provisioning:** Customers can express their needs in a high-level, business-oriented way, rather than with complex technical parameters. Intent inference using NLP is a good example.
- **Drives Automation:** Enables automated “zero-touch” operations by allowing systems to understand and fulfil business intents directly. In fact, Intents require automation, as their fulfilment occurs over a closed loop, and automation is often obtained from analyzing results and making decisions, two areas strongly impacted by AI.
- **Increases Agility and Interoperability:** Promotes a standardized way for systems to communicate, reducing the cost, complexity, and time-to-market for new services.

These are some AI functions in IBN:

1. **Intent Translation and Automation:** AI uses NLP and ML to understand high-level business or operational goals (Intents), automatically translating these into precise, actionable network policies and configurations.
2. **Policy Enforcement and Adaptation:** AI engines can automatically deploy network settings (such as Quality of Service, routing, or security rules) and adjust them in response to real-time context and demands.
3. **Continuous Monitoring and Optimization:** AI-driven analytics constantly monitor network performance, proactively detect anomalies, predict potential issues, and suggest or implement optimizations to keep networks aligned with user-defined intents and changing requirements.
4. **Self-Healing and Predictive Maintenance:** Machine learning helps networks predict failures, initiate automated self-healing, and maintain high availability, especially in distributed, high-performance computing environments.
5. **Context Awareness and Resource Allocation:** AI enables sophisticated resource management, using algorithms like reinforcement learning and predictive models to allocate bandwidth and prioritize traffic for critical workflows such as AI model training or real-time inference.

Among these, perhaps the most challenging step is the Intents translation, expressed in natural language, into network policies and configurations, as this is the focus of much research and scientific work (e.g., [9]).

Translating business intents into network policies

Hereafter, a more detailed walkthrough of a real-world scenario for an IBN system is shown, explaining how a high-level business intent becomes an actionable network configuration with the help of AI:

Step 1: State the Business Intent

The business declares: “We need videoconferencing prioritized during company-wide meetings across all sites.”

Step 2: Intent Capture and NLP Interpretation

The IBN platform receives the intent from a dashboard, chatbot, or API. AI, using NLP, extracts key requirements:

- Identify “videoconferencing” traffic/application
- Time window for prioritization (meeting hours)
- Scope: “all sites” refers to all network segments

Step 3: Semantic Mapping and Contextualization

AI parses the business language (“prioritize”, “videoconferencing”, “company-wide meetings”) and:

- Maps “videoconferencing” to traffic types/protocols (e.g., Zoom, Teams, and SIP/RTP)
- Associates “prioritize” with Quality of Service (QoS) enhancements (e.g., higher bandwidth allocation, lower latency)
- Sets time-based policies for “company-wide meetings”

Step 4: Policy Generation

The system automatically generates policies to:

- Tag videoconferencing packets for priority routing;
- Allocate higher minimum/maximum bandwidth for videoconferencing apps;
- Apply traffic shaping rules during defined meeting windows;
- Push configurations to all switches/routers/firewalls spanning “all sites”.

Example Policy (abstract):

```
{
  "application": "videoconferencing",
  "priority": "high",
  "bandwidth": {
    "min": "10Mbps",
    "max": "100Mbps"
  },
  "time": "Monday 9-11am",
  "targets": "site_A, site_B, site_C"
}
```

Actual network configuration is then derived (the system translates this into Command Line, API calls, or config templates for each device).

Step 5: Validation and Simulation

- IBN system checks policies against current resources and network state to ensure feasibility
- Simulates policy results: flagging capacity, identifying potential conflicts, or highlighting risks

LLM are emerging as the intelligence layer of intent-based networking, translating human language into automated, context-aware network orchestration

Step 6: Automated Implementation

- Upon successful validation, the platform deploys the policies network-wide using APIs, orchestration, or programmable controllers
- Devices update traffic flows and priorities automatically

Step 7: Monitoring & Assurance

- Real-time analytics verify that videoconferencing is receiving priority exactly as intended
- If network state drifts or meetings expand, the system can auto-adjust or alert IT staff

Usage of LLM in IBN

An LLM is a type of AI system designed to understand (interpret complex queries, summarize documents, translate languages, and answer questions) and generate human-like language. This is the focus of relevant research work and may rapidly transform IBN by serving as the core intelligence for translating, decomposing, and operationalizing natural language intents into concrete network configurations and policies [9], e.g., TM Forum-compliant formats. This simplifies service orchestration and enhances accessibility.

With this in mind, LLM are likely to play a key role in IBN, as the technology shows remarkable performance in interpreting human language as well as doing code generation, enabling the translation of high-level intents expressed by humans into low-level network configurations. Besides translation, LLM may well contribute to other complementary areas:

1. **Natural Language Interface:** LLM allow users to specify intents using plain language (not technical policy syntax), reducing the barrier for non-experts to manage complex networks [10].
2. **Intent Decomposition:** LLMs can break down high-level, potentially ambiguous requests (“prioritize video conferencing during meetings”) into structured steps or sub-intents, which match the required configuration actions at every hierarchy in the management system [11].
3. **Policy Synthesis:** After decomposing the intent, LLMs generate the appropriate policy abstractions and translate them into device-level or controller-level instructions (JSON, YAML, NSD, CLI, etc.), making the deployment of network changes seamless and scalable [12].
4. **Contextual Reasoning:** Compared to classic, template-based approaches, LLM reason based on context (considering current network state, historical outcomes, and organizational business rules) to produce policies that are accurate and relevant [12].
5. **Conflict Detection and Validation:** LLM are used to detect policy conflicts (e.g. contradictory flows), and validate intent feasibility before changes are pushed network-wide [12].

There are already some LLM proposed, like NetIntent [13], IntentLLM [14], and

Table 4: Comparison of LLM for IBN

Emergence [15], which interpret plain-language Intents, progressively decompose them, and auto-generate correct configurations, which are summarized in **Table 4**.

Feature	NetIntent	IntentLLM	Emergence
Focus Area	SDN-based IBN automation	5G/6G core network intent extraction	6G-native intent orchestration
LLM Role	Translates and benchmarks intents using IBNBench	Extracts intents from natural language and integrates with NWDAF	Uses few-shot learning and human-in-the-loop refinement
Architecture	Modular with SDN controller integration (ODL, ONOS)	Custom LLM tuned for telecom domain	End-to-end LLM pipeline aligned with TMF921A
Intent Feedback	Dynamic re-prompting and conflict detection	Real-time feedback via analytics functions	Closed-loop refinement with human feedback
Standards Alignment	TM Forum APIs, SDN/NFV	3GPP NWDAF, 5G slicing	TM Forum Autonomous Networks (IG1251), TMF921A
Use Cases	Automated service provisioning, benchmarking LLMs	Autonomous 5G slice management, SLA enforcement	6G service orchestration, smart city automation
Strengths	Strong benchmarking and SDN integration	Telecom-specific LLM with real-time analytics	Full-stack autonomy and TM Forum compliance

In summary, LLM are central to the new generation of IBN systems, making intent lifecycle management (translation, decomposition, validation, deployment, and assurance) more user-friendly, powerful, and adaptive than ever before:

- **Dynamic Adaptation:** LLM can learn from past interactions and network outcomes (human feedback loop), thus improving translation accuracy and intent fulfilment over time.
- **Plug-and-Play for New Scenarios:** LLM excel at “few-shot” or “zero-shot” learning, enabling IBN systems to handle new business intents without explicit reprogramming.
- **Broader Applicability:** LLM are bridging the gap between user intent and advanced SDN/Network Functions Virtualization (NFV) programmability, unlocking automation for operations, security, assurance and design [16].

Conclusions and future work

IBN, bolstered by recent advancements in AI and LLM for intent translation, is rapidly maturing. These technologies now offer a level of reliability and interpretability

The convergence of AI, LLM, and 5G is making intent-driven automation the new standard in networking

that instills confidence in their practical adoption. Simultaneously, the widespread deployment of 5G across diverse economic sectors and the arrival of 6G within the next few years makes the integration of IBN not just advantageous, but essential.

Altice Labs is actively exploring and evaluating the most effective pathways to embed IBN capabilities across its product portfolio. As a comprehensive solution provider, Altice Labs delivers technologies spanning the entire communications stack, from advanced wired access (e.g., multi-PON architectures) and wireless networks (e.g., 5G small cells), to service platforms and Business & Operations Support Systems (BSS/OSS). Products designed for the areas of network connectivity, network orchestration and operations management stand to benefit most directly from the adoption of intent-driven automation and closed-loop control.

In parallel, Altice Labs' leadership in API-driven innovation has already been proven by implementing GSMA CAMARA APIs to expose network capabilities, deploying AI-powered bots in live operational environments, and integrating intelligent automation into several BSS solutions. These internal capabilities, combined with external market dynamics, position Altice Labs in the front line to evolve toward IBN-enabled architectures.

And this evolution is already underway. Altice Labs is actively participating in national and international research initiatives focused on IBN, contributing to the development and prototyping of intent-aware systems. Through these efforts, Altice Labs continues to reinforce its role as a catalyst for innovation in the telecom sector, bridging cutting-edge research with real-world deployment.

References

- [1] 3GPP, "Management and orchestration; Intent driven management services for mobile networks," Rep. 28.312, 2018. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3554>
- [2] 3GPP, "Study on enhanced intent driven management services for mobile networks," Rep. 28.912, 2021. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3969>
- [3] A. Clemm, L. Ciavaglia, L. Z. Granville, and J. Tantsura. "Intent-Based Networking – Concepts and Definitions," RFC 9315, Internet Research Task Force (IRTF) Datatracker. [Online]. Available: <https://datatracker.ietf.org/doc/rfc9315/>
- [4] Q. Wu, S. Litkowski, L. Tomotaki, and K. Ogaki. "YANG Data Model for L3VPN Service Delivery," RFC 8299, Internet Engineering Task Force (IETF) Datatracker. [Online]. Available: <https://datatracker.ietf.org/doc/rfc8299/>
- [5] TM Forum, "Intent Management API Profile," London, U.K., Rep. TMF921A v1.1.0, 2022. [Online]. Available: <https://www.tmforum.org/resources/how-to-guide/tmf921a-intent-management-api-profile-v1-1-0/>

- [6] TM Forum, "Service Ordering Management API User Guide," London, U.K., Rep. TMF641 v4.1.1, 2021. [Online]. Available: <https://www.tmforum.org/resources/standard/tmf641-service-ordering-management-api-user-guide-v4-1-1/>
- [7] TM Forum, "Resource Activation API User Guide," London, U.K., Rep. TMF702 v4.0.1, 2020. [Online]. Available: <https://www.tmforum.org/resources/standard/tmf702-resource-activation-api-user-guide-v4-0-0/>
- [8] TM Forum, "TM Forum Intent Ontology (TIO) – References," London, U.K., Rep. TMF292R v3.6.0, 2024. [Online]. Available: <https://www.tmforum.org/resources/standard/tr292r-tm-forum-intent-ontology-tio-references-v3-6-0/>
- [9] K. Dzeperoska, J. Lin, A. Tizghadam, and A. Leon-Garcia, "LLM-based policy generation for intent-based management of applications," 2024, arXiv:2402.10067.
- [10] A. Mkrache and A. Ksentini, "LLM-enabled Intent-driven Service Configuration for Next Generation Networks," in *Proc. 2024 IEEE 10th Int. Conf. Netw. Softw. (NetSoft)*, Saint Louis, MO, USA, 2024, pp. 253–257, doi: 10.1109/NetSoft60951.2024.10588881.
- [11] A. Bekri, A. Abane, A. Battou, and S. Bensalem, "Bridging Language Models and Formal Methods for Intent-Driven Optical Network Design," 2025, arXiv:2509.22834.
- [12] M. A. Habib et al., "LLM-Based Intent Processing and Network Optimization Using Attention-Based Hierarchical Reinforcement Learning," in *Proc. 2025 IEEE Wireless Commun. Netw. Conf. (WCNC)*, Milan, Italy, 2025, pp. 1–6, doi: 10.1109/WCNC61545.2025.10978505.
- [13] M. K. Hossain and W. Aljoby, "NetIntent: Leveraging Large Language Models for End-to-End Intent-Based SDN Automation," 2025, arXiv:2507.14398.
- [14] L. Dinh, S. Cherrared, X. Huang, and F. Guillemin, "Towards End-to-End Network Intent Management with Large Language Models," 2025, arXiv:2504.13589.
- [15] K. Dzeperoska, A. Tizghadam, and A. Leon-Garcia, "Emergence: An Intent Fulfillment System," in *IEEE Commun. Mag.*, vol. 62, no. 6, pp. 36–41, Jun. 2024, doi: 10.1109/MCOM.001.2300270.
- [16] F. Liu, B. Farkiani, and P. Crowley, "Large Language Models for computer networking operations and management: A survey on applications, key techniques, and opportunities," in *Comput. Netw.*, vol. 271, Art. no. 111614, Oct. 2025, doi: 10.1016/j.comnet.2025.111614.



10

Powering the Future of Connectivity with Integrated Edge and AI in 5G Small Cells

Abstract

Altice Labs is developing a new generation of advanced small cells that marks a strategic evolution in the 5G network infrastructure. Designed with an all-in-one but highly modular architecture, this solution enables the deployment of 5G networks, private or public, tailored to the specific requirements of each application scenario. Its flexible software framework allows customized network functions and services to be integrated according to the needs of different verticals, easing rapid adaptation to diverse environments.

This solution can also incorporate on-cell local edge computing capabilities, contributing to ultra-low latency interactions for mission-critical applications, including real-time execution of Machine Learning (ML) and Artificial Intelligence (AI) algorithms. Additionally, direct integration with Passive Optical Networks (PON) reduces network complexity and lowers deployment costs. These features position the solution as a key enabler for verticals requiring high-performance, customizable, secure, and localized connectivity.

A standout use case will be demonstrated in the healthcare sector, in partnership with the Champalimaud Foundation, under the ALBATROZ project. This deployment showcases the small cell's potential in supporting demanding medical applications through distributed intelligence and edge-enabled 5G communications.

Authors

Samuel Madail, Altice Labs, samuel-r-madail@alticelabs.com

Diogo Marques, Altice Labs, diogo-c-marques@alticelabs.com

Pedro Gouveia, Champalimaud Foundation, pedro.gouveia@fundacaochampalimaud.pt

Keywords

5G, Small Cells, Edge Computing, Healthcare

DOI

10.57650/INNOVATION/2025.10



Introduction

The deployment of 5G networks is accelerating worldwide, driven by the need for enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC). While traditional macrocell-based architecture remains the backbone of public mobile networks, small cells have emerged as a critical enabler for densification and targeted coverage, especially in localized or high-demanding environments.

Private and public 5G networks demand flexible, cost-effective, and easily deployable solutions capable of offering high performance, strong security, and local control of data. To meet these requirements, Altice Labs is developing a new generation of advanced small cells based on an all-in-one but modular architecture. This approach emphasizes software flexibility and customization, allowing network functions and services to be adapted to the specific needs of each deployment scenario and application domain. Additionally, the solution supports integration with optical accesses and edge computing infrastructures, enabling low-latency processing and intelligent services orchestration at the network edge.

The healthcare sector provides a compelling context for demonstrating the benefits of this technology. In hospital and clinical environments, reliable and secure 5G connectivity can support a wide range of innovative applications, from real-time monitoring and remote diagnostics to AI-assisted image analysis and autonomous medical devices, all requiring deterministic latency and local data handling. Under the ALBATROZ project [9], in collaboration with the Champalimaud Foundation (CF), several use cases will illustrate how such an adaptable small cell platform can address the stringent communication demands of modern healthcare, contributing to improved efficiency, safety, and patient outcomes.

Current challenges

Building on their widespread adoption in 4G, small cells have become a central element of 5G deployment strategies, as operators and enterprises seek to meet the customers' growing demand for coverage, capacity, and performance. Beyond their traditional role in radio densification, small cells are now key enablers for localized and private 5G networks, offering dedicated connectivity tailored to specific environments and verticals.

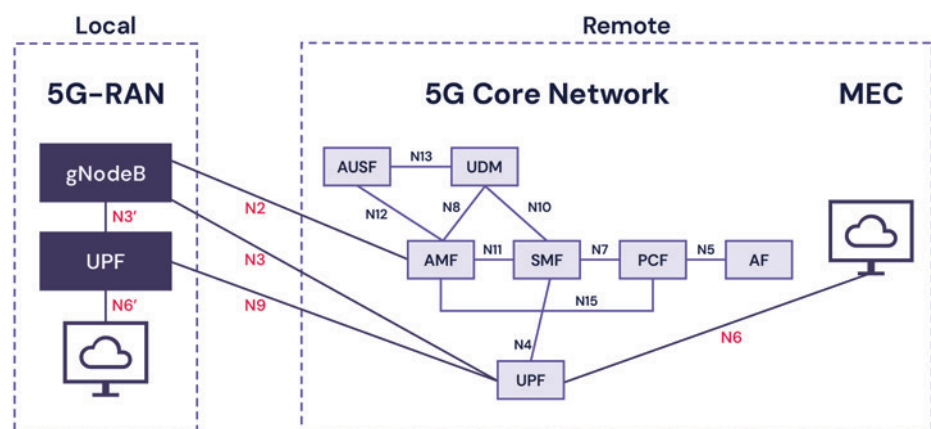
Conventional small cell designs typically maintain a clear separation between the Radio Access Network (RAN) and all the core network elements, requiring connection to centralized infrastructures. While this approach is efficient in public mobile networks, it introduces several challenges for private or enterprise scenarios:

- 1. Integration complexity:** Connecting to centralized core and service platforms requires specialized infrastructure and expertise;
- 2. Latency limitations:** Dependence on remote processing, constrains real-time responsiveness, limiting support for low-latency use cases;
- 3. Cost and scalability:** Centralized core architectures can entail significant investment and operational costs to address smaller or distributed RAN deployments;
- 4. Vertical-specific demands:** Sectors such as healthcare, manufacturing, and logistics require on-premises data processing, enhanced security, and predictable quality of service.

Recent research and industrial efforts have therefore focused on distributing network intelligence and services toward the edge [1], supported by frameworks such as 3GPP's Non-Public Networks (NPN) and ETSI's Multi-access Edge Computing (MEC) [2]. These developments enable architectures in which applications and User-Plane Functions (UPF) can be deployed close to end devices, allowing local traffic breakout and low-latency services execution.

It is, however, essential to distinguish between service centralization and core centralization. Full deployment of a standalone 5G core within each small cell is technically feasible, but only relevant in specific, isolated scenarios. In most practical cases, the focus lies on hosting edge services or localized network functions, for instance, a UPF co-located with the small cell (**Figure 1**), to achieve ultra-low latency and data sovereignty without replicating the entire core infrastructure.

Figure 1: Network topology to run workloads locally in the picocell



Despite continuous progress, commercially available solutions that seamlessly combine RAN, edge computing, and flexible local service hosting remain limited. This gap motivates the development of next-generation, software-customizable small cells, such as those pursued by Altice Labs, capable of adapting to the diverse requirements of vertical industries. In particular, the healthcare sector presents a strong use case [3], where reliable, low-latency, and secure communication infrastructures are essential to support advanced medical applications and improve patient outcomes.

Solution overview

Building on the expertise and insights gained from the first generation of 5G small cell Radio Units (RU) for private networks, Altice Labs has designed and developed a new all-in-one solution that represents a significant step forward in both functionality and efficiency. Unlike the 5G RU, part of the O-RAN Alliance architecture, which primarily targeted private network deployments, this new generation has been conceived to serve both private and public environments through an all-in-one design. The integration of a full stack into a single and compact platform simplifies deployment and management, reduces hardware and operational costs, and enables rapid setup.

This evolution goes beyond traditional connectivity by integrating advanced features that transform how small cells are deployed and managed, including:

- **AI/ML-driven optimization**, executed locally on the unit, enabling use cases like load balancing, interference mitigation, and energy-saving strategies for more efficient operation;
- **On-cell Local edge computing**, unlocking ultra-low-latency applications and localized data processing;
- **Zero-Touch Provisioning (ZTP)**, automating installation and configuration while minimizing operational effort;
- **Software-Upgradable & Future-Proof**, ensuring seamless 4G/5G evolution and keeping the system always up to date;
- **Native PON connectivity**, allowing direct integration with PON without requiring dedicated mobile backhaul;
- **Embedded 5G core**, enabling a “private network in a cell” that delivers end-to-end functionality without external dependencies. Other breakout options, like UPF (data path), are also possible.

These innovations turn all-in-one small cells into a disruptive option for flexible, cost-efficient, and intelligent 5G deployments.

How small cells fit into modern mobile networks

The Altice Labs small cell solution is designed to complement the existing mobile network, providing enhanced capacity and coverage where it is most needed. While

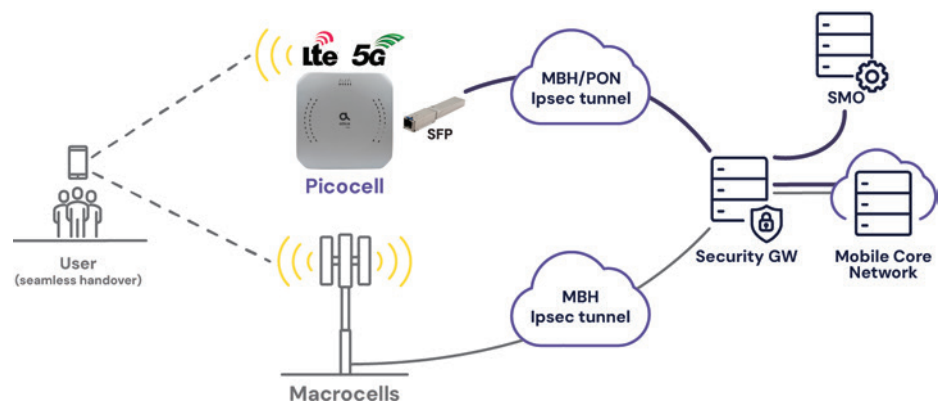
With Altice Labs' AGORA SMO, small cells and diverse technologies converge under unified, secure management

macrocells continue to deliver wide-area coverage, small cells are strategically deployed to boost capacity and improve service quality in high-traffic hotspots.

The solution ensures seamless coexistence and coordination with the current network infrastructure, as shown in **Figure 2**. From an architectural standpoint, the small cell connects to the operator's transport network through standardized interfaces, enabling flexible backhaul options. To ensure data integrity and protection across these links, secure backhaul mechanisms such as IPsec tunnels can be used, protecting both user and control plane traffic. Through these connections, the small cell establishes communication with an existing mobile core network, becoming an integral part of the corresponding RAN domain. The small cell is capable of interacting with the macro base stations and other neighboring cells to ensure mobility continuity, handover management, and interference mitigation.

Regarding management, Altice Labs' Service Management and Orchestration (SMO) solution (i.e., AGORA [12]) can manage small cells as well as other technologies, enabling unified management (wireless and wired backhaul), bringing benefits in problem resolution and end-to-end provisioning. As a result, the solution supports simplified deployment (via ZTP), remote monitoring, and software updates, which are essential for scalable and cost-effective network expansion.

Figure 2: Altice Labs' Small Cell integration in an existing mobile network



Additionally, the solution leverages shared PON infrastructure wherever possible, optimizing resource utilization and reducing deployment costs. This combination of wide-area coverage, targeted capacity enhancement, and cost efficiency enables operators to meet growing demand while maintaining a high-quality user experience.

Enabling 5G with FTTx

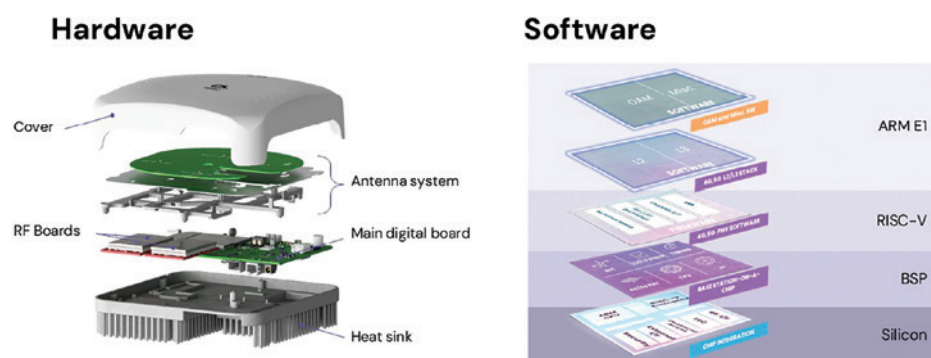
When combined with all-in-one small cells, Fiber to the x (FTTx) and PON infrastructures become a game-changer. Unlike dedicated mobile transport backhaul typically used for macrocells, PON offers a significantly lower connection cost (up to 70% reduction compared to traditional solutions) while being widely available across urban and rural areas. The PON ubiquity dramatically expands the economic viability of 5G, enabling operators and enterprises to deploy localized 5G coverage wherever fiber access is present. By leveraging existing PON infrastructure, all-in-one small cells can deliver high-performance 5G at a fraction of the cost, making new deployment scenarios both technically and financially attractive.

Modular and reconfigurable design

The solution's modular hardware and software architecture enables a high degree of flexibility and scalability. Each component can be independently configured, upgraded, or replaced, allowing the system to adapt to different scenarios, changing network demands, and evolving technology standards. This reconfigurable design not only simplifies maintenance and future expansion but also ensures that operators can deploy capacity exactly where and when it is needed. By combining modularity with a software-driven approach, the solution supports an evolutionary path, allowing for continuous performance improvements and feature enhancements over time.

As shown in **Figure 3**, the small cell is composed of distinct layers. Each layer can be individually optimized or upgraded, ensuring adaptability to various deployment scenarios. For instance, different RF boards and antenna systems can be integrated to support different frequency bands, allowing the deployment of small cells tailored to specific spectrum requirements.

Figure 3: Exploded view of hardware and software modules



The software architecture is modular, taking full advantage of the hybrid processor architecture of the System on Chip (SoC), integrating both Reduced Instruction Set Computing – V (RISC-V) and Advanced RISC Machines (ARM) cores. The RISC-V cores are specifically optimized for 4G and 5G physical-layer workloads, enabling the simultaneous, concurrent execution of both. Complementing them, the ARM cores open the door to virtualization, allowing execution of virtualized L2 and L3 functions for 4G and 5G, along with OAM software. This virtualization allows software modules to be independently updated or optimized for different network scenarios. In addition, the SoC includes hardware that can be used for AI acceleration, preparing the RAN for AI scenarios.

Edge intelligence capabilities

One of the defining features of the proposed solution is its ability to execute applications within the small cell (on-cell edge computing). This is particularly relevant for AI and ML workloads that require real-time processing.

Examples of potential applications include:

- **Healthcare:** Real-time image analysis for diagnostic support.

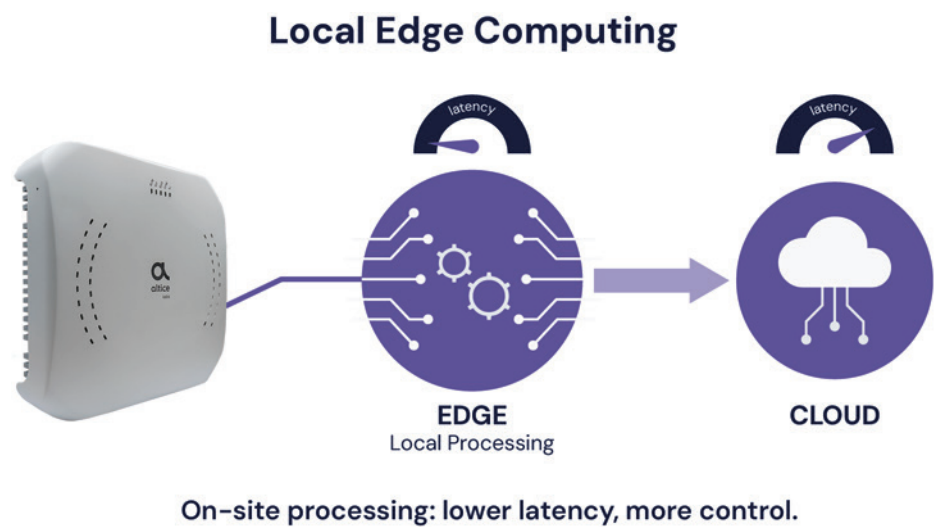
On-cell edge computing cuts latency while safeguarding local data, representing a dual gain in speed and trust

- **Industry 4.0:** Predictive maintenance and anomaly detection in manufacturing.
- **Transportation:** Safety-critical decision-making in connected vehicles.

As illustrated in **Figure 4**, by optionally embedding computational resources directly within the small cell, latency is minimized and data remains within the local environment, enhancing both performance and privacy. This makes the solution especially attractive for verticals with stringent requirements on responsiveness and security.

Comparing a solution with common Edge Computing, an on-cell edge processing can reduce the End-to-End (E2E) latency to less than half. For example, in AGV control within a factory, the positional error for 20ms of latency is between 10cm and 0.5m (depending on speed). However, with a local UPF, the positional error decreases to less than 2cm, which is an enormous operational gain, even if the control processing is trivial.

Figure 4: Reducing latency with local processing within the small cell



In Altice Labs' solution, resources can be provided for on-cell edge computing (with a local UPF), namely 5 cores (10threads) @ARM Neoverse E1 CPU. These third-party applications can also use resources from RISC-V and HW blocks available in the SoC.

5G small cells in diverse deployment contexts

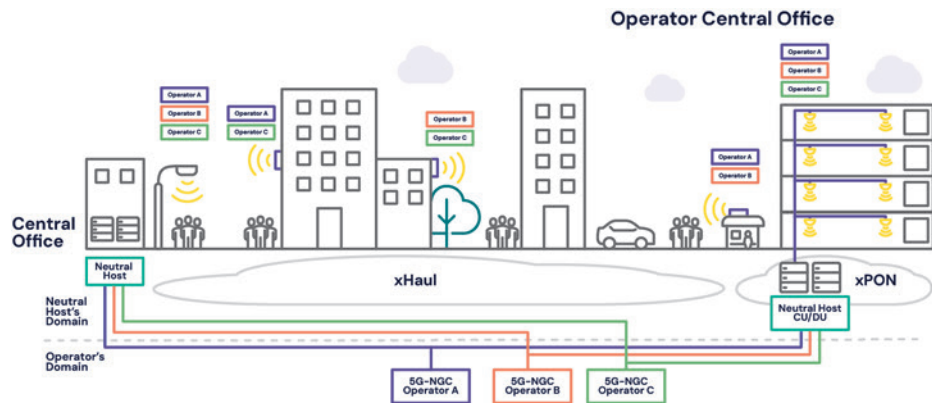
The 5G small cells deployment approach is relevant in multiple practical use cases, such as Mobile Network Operators (MNO), Neutral Host Providers (NHP), and private network owners. Small cells aim to promote 5G cell densification and extend the coverage of their current mobile networks. MNO rely on small cell deployment to capitalize on the following:

- **Cost-effective densification,** for a more manageable and less expensive 5G network densification through the deployment of outdoor and indoor 5G small cells.

- **Enhanced flexibility**, since small cells offer the flexibility to quickly increase network capacity and coverage, while seamlessly integrating with existing technologies, reducing the complexity of network expansions.

NHP, a third-party non-operator entity, is expected to play a key role in deploying 5G small cells in urban centers, historic downtowns, or public buildings (Figure 5). Considering the lack of business cases for large MNO to invest in their own network densification or local regulation constraints, an opportunity is presented for NHP to deploy a network to be rented to the different MNO and potentially reduce operators' Operating Expenditure (OPEX) and Capital Expenditure (CAPEX).

Figure 5: Neutral Host architecture enabling multi-operator 5G connectivity



The rise of private 5G marks a new era of secure, end-to-end connectivity for enterprises and public entities

Large enterprises, businesses, and public entities are increasingly turning to private 5G networks to ensure robust security and guaranteed Quality of Service (QoS). These independent, end-to-end small to medium-sized networks utilize 5G small cells to provide reliable connectivity in key contexts, such as industrial centers, large corporations, and facility owners or municipalities.

Cell offload and coverage increase: local hotspot

Small cells serve as an effective tool for solving modern network challenges, especially in high-demand areas, such as stadiums, airports, and dense city centers. By intelligently offloading traffic from overloaded macro cells, they ensure consistent, high-quality service for every user, even during peak events. This not only significantly improves user experience but also optimizes overall network performance and efficiency.

Beyond permanent installations, small cell solutions also offer remarkable flexibility. Their plug-and-play nature and minimal footprint make them the ideal choice for rapid, temporary deployments. Whether it is for supporting seasonal demand or one-off events, they can quickly provide reliable, high-performance coverage, exactly where and when needed.

Additionally, small cells can be strategically deployed to extend coverage into areas where macro signals struggle to penetrate, such as underground stations, shopping malls, or office complexes. In these scenarios, small cells act as localized hotspots, ensuring that users remain connected in otherwise challenging environments while maximizing spectrum utilization and network performance (Figure 6).

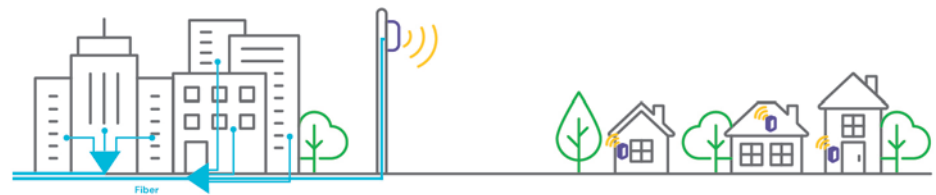
Figure 6: Cell offload and coverage increase scenario



Fixed Wireless Access (FWA)

FWA uses wireless technology to deliver broadband services to homes and businesses without the need for new cables. It is often more affordable and quicker to deploy than wired alternatives, since it leverages existing mobile infrastructure and requires minimal installation. With plug-and-play equipment, households can get connected rapidly without waiting for fiber rollout, while getting competitive speeds. Most importantly, FWA supported by small cells can extend reliable connectivity to suburban, rural, or underserved areas where last-mile fiber is either cost-prohibitive or time-consuming (**Figure 7**).

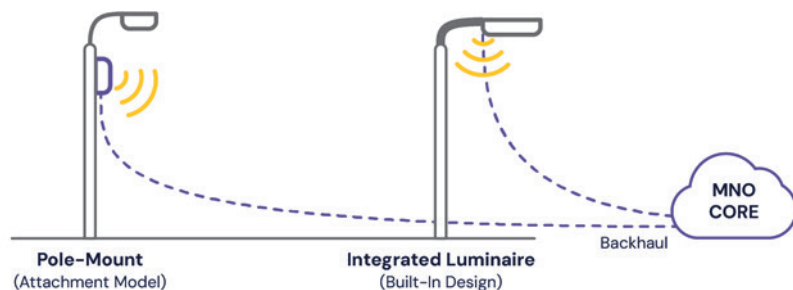
Figure 7: FWA deployment scenario



Lampsite small cells

Lampsite small cells are compact 5G access points designed to be seamlessly deployed within existing street lighting infrastructure (**Figure 8**). By embedding these units directly into lampposts, operators can extend network coverage and capacity in dense urban areas without adding visible clutter to the streetscape. This discreet integration not only preserves city aesthetics but also accelerates deployment, since power and mounting structures are already in place.

Figure 8: Lampsite small cells deployment options



Such solutions are especially valuable in high-demand zones like city centers, commercial districts, and transport hubs, where traditional network infrastructure

can be difficult to install and often faces public resistance. Leveraging lampposts as hosting sites reduces both deployment time and cost, while ensuring reliable connectivity to support growing data traffic and emerging smart city services.

Altice Labs offers solutions for these different scenarios, providing an efficient, scalable, and integrated xHaul architecture. Agnostic to the transport technology adopted, it offers a complete end-to-end 5G small cell solution, using 5G small cell solutions developed to handle fully equipped 5G and Beyond 5G (B5G) mobile networks.

As part of the Invisible 5G project [13], Altice Labs had already demonstrated the advantages of discreetly and harmoniously integrating small cells into urban furniture, such as utility poles, lighting poles, or even technical cabinets.

Case study: the ALBATROZ project

The ALBATROZ project, “Altice Labs Advanced Technology on Radio, Optics and Virtualization”, aims to advance Altice Labs’ network solutions in line with emerging telecommunications technologies, including 5G small cells, to enable innovative solutions for private and operator networks. The CF, a leading private biomedical research and clinical care institution, joined as an external partner to test and demonstrate the application of private 5G networks in healthcare, focusing on patient treatment workflows, particularly in breast cancer care.

The 5G/Edge architecture driving dynamic Digital Twins in breast cancer surgery

Real-time digital twins transform surgical planning and tumor localization, enabled by edge-driven private 5G networks

In this context, medical imaging analysis plays a critical role in diagnosis and treatment decision-making. Traditional workflows require local GPU-enabled workstations for AI-assisted image analysis, which is resource-intensive. Furthermore, surgical planning demands collaborative visualization of patients’ digital twins [14], and the operating room requires precise, real-time image guidance for tumor localization. The creation of these digital twins relies on real-time data capture and integration, giving them a dynamic and predictive nature, in contrast to static 3D anatomical models. The shift from static 3D anatomical models to dynamic digital twins in breast cancer surgery is fundamentally an architectural necessity enabled by 5G and Edge Computing [15]. The key difference lies in the network’s ability to handle the volume, velocity, and latency requirements of real-time patient data. These scenarios require ultra-low latency, reliable connectivity, and secure local processing, making them ideal use cases for edge-enabled private 5G networks. Within the ALBATROZ project, three different use cases are planned to be demonstrated using Altice Labs’ small cell solution. A brief description of each one follows:

1. **AI-assisted Medical Imaging:** AI algorithms for Magnetic Resonance Imaging (MRI), mammography, and Computed Tomography (CT) scans can be centralized using 5G connectivity. Instead of requiring every radiologist to maintain a local GPU workstation, hospitals can deploy shared

GPU resources accessed seamlessly and securely over the 5G network for real-time analysis;

- 2. Collaborative Immersive Visualization:** Virtual Reality (VR) and Multiple Mixed Reality (MR) wireless devices connect to centralized imaging resources to collaboratively plan surgeries using digital twins [16];
- 3. Mixed Reality Surgical Guidance:** In the operating room, MR devices provide real-time tumor localization over the patient's anatomy [17]. On-cell edge computing within small cells for specific time-critical tasks reduces latency, enabling precise visualization and improving surgical accuracy.

The advantage of using 5G, even when wired connections could perform the same tasks, is the added mobility, scalability, flexibility, and redundancy it provides, which ensures real-time, secure, and resilient connectivity without the constraints of physical infrastructure.

These deployment scenarios highlight the transformative potential of private 5G and edge intelligence in healthcare, enabling new clinical workflows and operational efficiency.



Conclusion

The new small cell solution from Altice Labs represents a significant step towards a converged, high-performance platform for both private and public 5G networks. This solution simplifies the deployment of private 5G systems, enables mission-critical low-latency applications, and creates powerful new opportunities for public operators. By reducing complexity, lowering costs, and enabling local intelligence, it can accelerate adoption in sectors such as:

- **Healthcare** (telemedicine, surgery assistance);
- **Industrial IoT** (automation, robotics);
- **Smart cities** (public safety, energy management);
- **Defense and security** (secure tactical networks).

The small cell features a modular hardware design that supports multiple deployment options and enables on-cell edge processing, allowing workloads to run directly on the cell for low-latency, high-performance operation. In addition, it integrates an AI-capable chip, preparing the RAN for advanced AI-driven applications and intelligent network optimization.

The versatility of this solution accelerates 5G roll-out and allows operators to respond more flexibly

to evolving customer demands, positioning these intelligent small cells as a game-changer for the future of mobile connectivity.

The healthcare use case under the ALBATROZ project will demonstrate the practical benefits and transformative potential of this approach. As industries increasingly adopt private 5G, compact and intelligent network elements like the one described here will play a pivotal role in shaping the digital infrastructure of the future.

References

- [1] S. Eswaran and P. Honnavalli, "Private 5G networks: a survey on enabling technologies, deployment models, use cases and research directions," in *Telecommun. Syst.*, vol. 82, pp. 3–26, 2023, doi: 10.1007/s11235-022-00978-z.
- [2] A. S. Sidibé, S. Katta, and J. C. Requena, "Performance Study of 5G Indoor Small Cells for Industrial MEC," in *Proc. 2024 Joint Eur. Conf. Netw. Commun. & 6G Summit (EuCNC/6G Summit)*, Antwerp, Belgium, 2024, pp. 1–6, doi: 10.1109/EuCNC/6GSummit60053.2024.10597043.
- [3] B. Chen et al., "Construction and Application of a Private 5G Standalone Medical Network in a Smart Health Environment: Exploratory Practice From China," in *J. Med. Internet Res.*, vol. 26, Art. no. e52404, 2024, doi: 10.2196/52404.
- [4] European Telecommunications Standards Institute (ETSI), "5G; Telecommunication management; Study on system and functional aspects of energy efficiency in 5G networks," ETSI, Rep. TR 132 972 V18.0.0. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/QKD/001_099/015/02.01.01_60/gs_QKD015v020101p.pdf
- [5] O-RAN Alliance, *Selected O-RAN Specifications Relevant to AI/ML and RIC Architecture*, Alfter, Germany, 2021–2025.
- [6] AI-RAN Alliance, "AI-RAN Alliance: Vision and Mission White Paper," 2024. [Online]. Available: <https://ai-ran.org/publications/>
- [7] G. Nencioni, R. G. Garroppo, and A. F. Olímí, "5G Multi-Access Edge Computing: A Survey on Security, Dependability, and Performance," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3288334.
- [8] 5GAmericas, "Energy Efficiency and Sustainability in Mobile Communications Networks," 2023. [Online]. Available: <https://www.5gamericas.org/energy-efficiency-and-sustainability-in-mobile-communications-networks/>
- [9] Altice Labs (2024) ALBATROZ Project. Technical Annex. P2030
- [10] M. Sauter, *From GSM to LTE-Advanced Pro and 5G*, 4th ed. Hoboken, NJ, USA: Wiley, 2021. [Online]. Available: <https://www.oreilly.com/library/view/from-gsm-to/9781119714675>

- [11] P. Jangale and A. Sayed, "Analysis of Hybrid DAS–Small Cell Architectures for Dense 5G Deployments," *Int. J. Lead. Res. Publ. (JLRP)*, vol. 5, no. 9, pp. 1–9, 2024.
- [12] Altice Labs, "AGORA Network Management Solution for Altice Labs products" Altice Labs, Aveiro, Portugal. [Online]. Available: https://www.alticelabs.com/wp-content/uploads/2025/03/BR_AGORA_ALB_EN.pdf
- [13] R. Caldeirinha, J. Reis, T. Oliveira, R. Bispo, S. Madail, J. Salgado, "Evolving invisible 5G small-cells to 6G smart radio environments," *InnovAction*, no. 8, pp. 22–35, 2023.
- [14] R. Cremese et al., "Virtual reality–augmented differentiable simulations for digital twin applications in surgical planning," *Sci. Rep.*, vol. 15, Art. no. 24377, 2025, doi: 10.1038/s41598-025-99608-2.
- [15] P. F. Gouveia et al., "Augmented Reality in Breast Surgery Education," *Breast Care*, vol. 18, no. 3, pp. 182–186, Jun. 2023, doi: 10.1159/000529587.
- [16] R. Cremese et al., "Virtual reality–augmented differentiable simulations for digital twin applications in surgical planning," *Sci. Rep.*, vol. 15, Art. no. 24377, 2025, doi: 10.1038/s41598-025-99608-2.
- [17] P. F. Gouveia et al., "Augmented Reality in Breast Surgery Education," *Breast Care*, vol. 18, no. 3, pp. 182–186, Jun. 2023, doi: 10.1159/000529587.



11

Driving the Expansion of the 5G Ecosystem through Impactful Success Stories

Abstract

Altice Labs operates a 5G laboratory at its facility in Aveiro, Portugal, dedicated to testing advanced network and service solutions before market release. The setup demonstrates Enterprise 5G in real-world private deployments, featuring seamless mobility across multiple cells and integrating innovative capabilities such as 5G-based positioning.

Altice Labs plays a key role in European and national research initiatives, including the European Commission (EC) project IMAGINE-B5G [1] and the National Recovery and Resilience Plan (RRP) NexGen Mobility [2]. In those, Altice Labs manages and supports projects and pilots from external companies using its laboratory facilities, helping them elevate their products to higher Technology Readiness Levels (TRL) and fostering valuable synergies. In doing so, Altice Labs contributes significantly to the national industry while building relationships with potential future partners.

This publication outlines the local facilities and showcases key pilots and projects conducted on the testbed in recent months. It highlights features such as 5G-based positioning in application domains like automotive, public safety, smart cities, transport, logistics, and security, emphasizing results that benefit both external companies and Altice Labs.

Authors

Carlos Marques, Altice Labs, carlos-martins-marques@alticelabs.com

Álvaro Gomes, Altice Labs, AGomes@alticelabs.com

Miguel Freitas, Altice Labs, miguel-r-freitas@alticelabs.com

Keywords

5G-NR, 3GPP, O-RAN, 5G Positioning, V2x, PPDR, Smart Cities, Transport, Logistics, Energy

DOI

10.57650/INNOVATION/2025.11

Introduction

The evolution of mobile telecom networks, driven by the disruptive capabilities of 5G, presents a unique opportunity to lead innovations across various verticals, including public safety, smart cities, transport, logistics, energy, V2X, and Industry 4.0.

To support this transformation, Altice Labs established a 5G laboratory as part of European and national research initiatives. A major goal is to help industry and academic partners increase the TRL of their solutions. Two notable examples of this effort are the IMAGINE-B5G [1] and NextGen Mobility [2] projects, funded by the European Commission and Portugal's National RRP, respectively. These initiatives have enabled several companies to enhance their products, solutions, and use cases.

The following sections provide an overview of the 5G laboratory and highlight the successful outcomes of selected pilots conducted within the scope of these two projects.

5G Laboratory

Altice Labs' 5G lab in Aveiro tests advanced network and service solutions before market rollout

Altice Labs has established a 5G laboratory at its facility in Aveiro, Portugal, designed to evaluate advanced network and service solutions prior to market deployment. The company develops broadband access technologies, with Passive Optical Network (PON) as a key strategic product line, alongside advanced 5G solutions.

The 5G laboratory is built upon the Open Radio Access Network (O-RAN) architecture, featuring distributed Radio Units (RU) across the facility, covering the complete Aveiro Campus, indoors and outdoors. The Next Generation Node B (gNB) is disaggregated into three components: the RU, the Distributed Unit (DU), and the Centralized Unit (CU), which can be deployed in multiple combinations. The available RAN solutions include ASOCS Cyrus 2.0 [3], srsRAN Enterprise 5G [4], both commercial solutions, and OAIBOX Max [5], an open-source solution, also following the O-RAN split 7.2. Multiple RU units are operating on the campus: (i) Altice Labs indoor and outdoor [6], (ii) ASOCS indoor [7], (iii) Benetel outdoor [8], and (iv) Ettus USRP [9].

Figure 1 illustrates the architecture of the 5G laboratory:

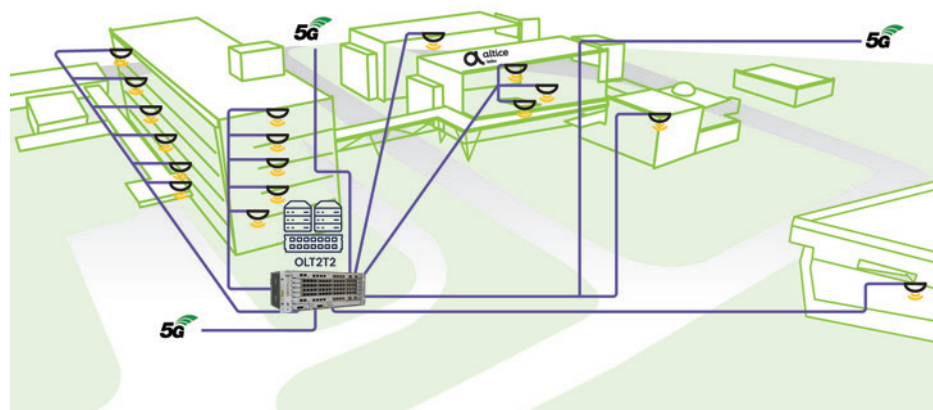


Figure 1: 5G Laboratory architecture



For experimental activities, 20 MHz of 5G spectrum in the higher part of the N78 band is available, provided by MEO (Altice Portugal's mobile operator).

The 5G laboratory supports several 5G standalone (SA) core solutions, including Raemis Druid [10], a commercial 5G SA, as well as Open5GS [11] and OAIBOX Max [5], both open-source platforms that support all major core components. Raemis and Open5GS follow a VM-based containerized deployment via Proxmox [12].

Additionally, the 5G laboratory supports Mobile Edge Computing (MEC) and 5G positioning capabilities, exploiting the 5G Core Location Management Function (LMF) and the New Radio Positioning Protocol A (NRPPa).

5G Pilots

The IMAGINE-B5G project [1], part of the Smart Network and Services (SNS) initiative, aims to implement an advanced, accessible, secure, and programmable end-to-end (E2E) 5G platform for large-scale trials and pilots across Europe. By leveraging the best breed of B5G features, the project fostered collaborations across key 5G verticals, with the following enterprises:

1. Public Protection and Disaster Recovery (PPDR): OneSource [13] and RedZinc [14];
2. Automotive: AtoBe [15];
3. Transport and logistics: YouShip [16];
4. Open RAN and 5G-based location: srsRAN [4].

The NexGen Mobility project [2] focuses on supporting the development, testing, and validation of mobility and smart infrastructure products and services. By providing human resources and 5G infrastructure, Altice Labs is targeting the evolution of the technological readiness of participating small and medium-sized enterprises (SMEs) and fostering collaborations across key 5G verticals, with the following enterprises:

1. Smart Cities: Omniflow [17];
2. Transport and logistics: Youship [16] and Cibersur [18];
3. Automotive: OneSource [13];
4. IoT: JSIO [19];
5. Forestry: Bold Robotics [20], [21], [22];
6. Security: Quantumnova [23] and SafetyScope [24];
7. 5G: Allbesmart [25].

These collaborations span a wide spectrum of current and emerging applications, demonstrating the versatility and impact of 5G innovation across diverse industry sectors.

IMAGINE-B5G

Pilot Name: **Software Radio Systems Platform Extension (SRS-5G)** [26], [27]

Company Name: **srsRAN**

Vertical Sector: **RAN**

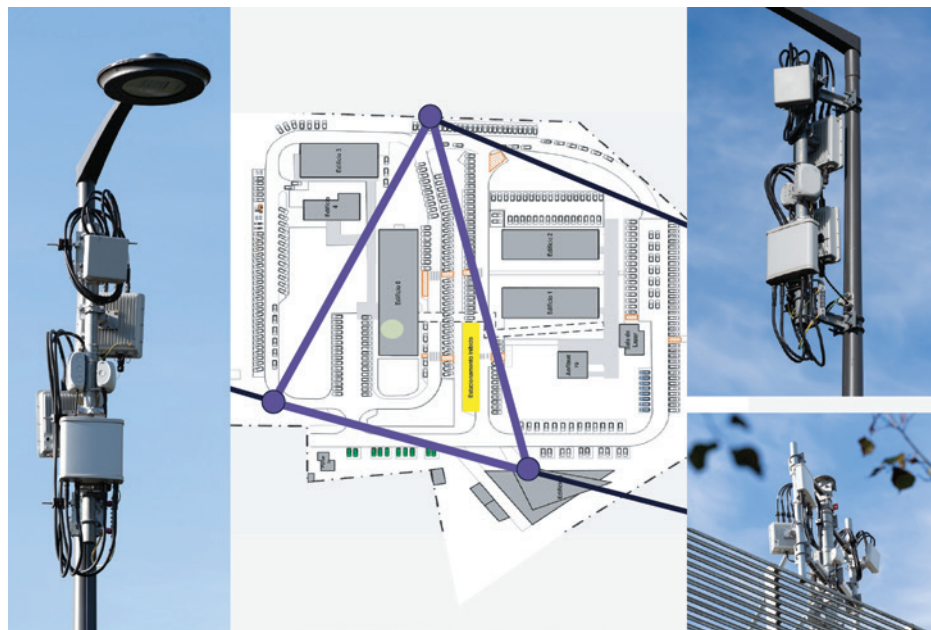
Use Case Description:

The SRS-B5G project, with the collaboration of Capgemini, aimed to extend the IMAGINE-B5G facility at Altice Labs with a complete multicell Open RAN deployment based on the srsRAN solution. This deployment incorporated advanced 5G RAN positioning features developed, including NRPPa protocol, and the methods Enhanced Cell-ID (eCID) and UpLink Time Difference of Arrival (UL-TDoA). These positioning methods outperform standard Global Navigation Satellite System (GNSS) in Time To First Fix (TTFF) and offer a compelling alternative, especially in environments where satellite signals are limited or unavailable, making them highly relevant across several verticals.

This extension also provides a standard-compliant E2 interface and supports the Key Performance Measurement (KPM) and RAN Controller (RC) Service Models (SM), including RAN Intelligent Controller (RIC) functionality.

Notably, the first implementation and testing of these capabilities by srsRAN was carried out at Altice Labs, marking a significant milestone in the development of 5G-based positioning, as shown in **Figure 2**. This work was highlighted in a press release [27], showcasing the pioneering nature of this Open RAN deployment and its advanced positioning capabilities in multi-cell scenarios.

Figure 2: srsRAN deployment at Altice Labs campus (cell sites for location calculation)



Pilot Name: **5G Edge V2X Enhanced Revenue (5G-EVER)** [28]

Company Name: **A-to-Be**

Vertical Sector: **Transport and Logistics**

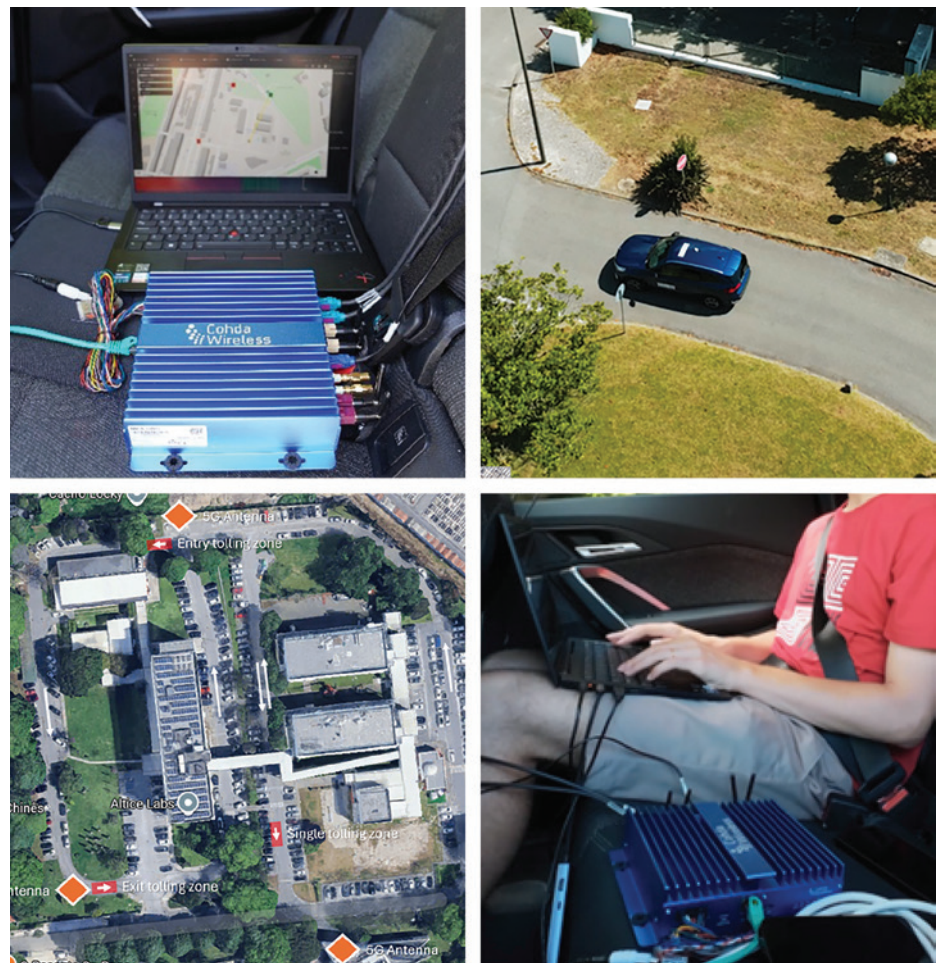
Use Case Description:

The 5G-EVER project introduces a novel approach to smart mobility by enabling vehicles to autonomously activate and pay for transport services through Vehicle-to-Network (V2N) communications. This technology replaces traditional roadside infrastructure with a digital alternative that controls 5G connectivity, edge computing, and advanced positioning.

Led by A-to-Be, the project builds upon existing V2X payment solutions and adapts them to operate over 5G networks. By integrating with Multi-access Edge Computing (MEC), transactions can be processed locally and in real-time, ensuring reliability and responsiveness.

As shown in **Figure 3**, the test facility at Altice Labs offered an ideal environment, combining 5G connectivity with edge infrastructure and enhanced positioning capabilities, which were enabled by SRS-5G. This 5G feature allowed the team to test realistic tolling scenarios, from open-road single-point tolls to more complex entry-exit systems, while addressing one of the key challenges in V2X payments: precise and trustworthy vehicle localization.

Figure 3: 5G-EVER trial (a: OBU; b: vehicle in the campus; c: cells' location in the campus and the tooling zones; d: testing environment in the vehicle)



Pilot Name: **Autonomous delivery service using 5G and GNSS localization fusion (ADS-5G)** [29]

Company Name: **YouShip**

Vertical Sector: **Transport and Logistics**

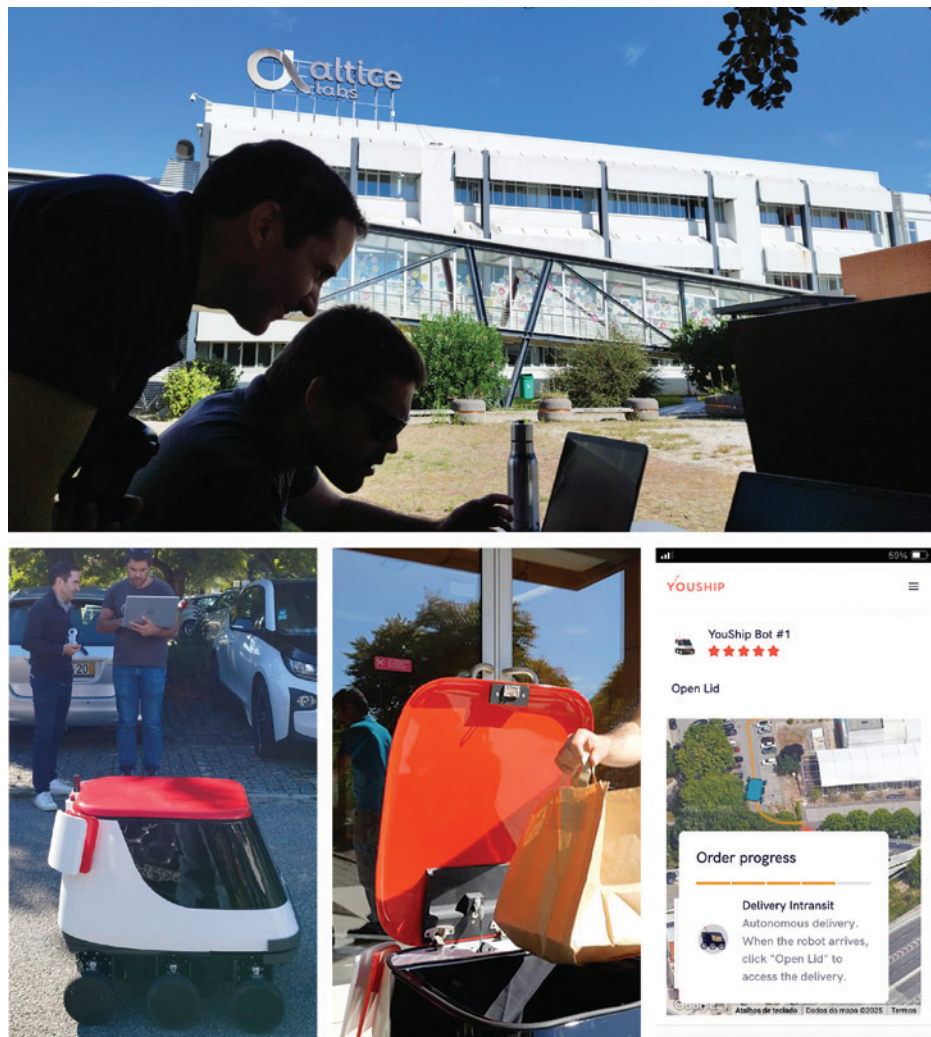
Use Case Description:

The ADS-5G project aims to enhance localization accuracy and ensure continuous positioning by combining 5G and GNSS positioning. **Figure 4** shows the ADS-5G trial at the Altice Labs campus, where an Autonomous Delivery Robot (ADR) developed by YouShip delivers small packages across indoor and outdoor routes, addressing key challenges in the transportation and logistics sector.

Equipped with GNSS technology, Artificial Intelligence (AI), and machine vision algorithms, the robot safely and swiftly navigates autonomously, in short routes or following a designated person, avoiding obstacles and offering a smooth delivery experience.

The project takes advantage of 5G technology to significantly enhance localization accuracy, enable low-latency communication, and ensure reliable communication across both indoor and outdoor environments.

Figure 4: ADS-5G trial (a: Final trial at the Altice Labs campus; b: ADR showcase; c: ADR; d: Delivery service live tracking)



Pilot Name: **Situational Awareness Framework Enabling Robust Emergency Response for Urban Flood Warnings** (SAFER-FLOW) [30], [31], [33], [34], [35], [36]

Company Name: **OneSource**

Vertical Sector: **PPDR**

Use Case Description:

The SAFER-FLOW Vertical Experiment aims to enhance PPDR by integrating advanced Beyond-5G technologies to address complex and challenging flood scenarios. This project tackles key emergency response challenges through real-time data collection and smarter decision-making, aiming to enable fast, informed, coordinated disaster response.

Ultra-low latency and high reliability make 5G a catalyst for real-time, data-driven decisions

SAFER-FLOW uses Internet of Things (IoT) sensors to enhance PPDR. It monitors flood levels, issues early warnings, and aids evacuation. Drones provide aerial views, while AI and Machine Learning (ML) map water courses and assess severity. With ultra-low latency and high reliability, 5G enables real-time data transmission from IoT sensors, drones, and first responders, supporting faster, more informed decision-making. Advanced network slicing and edge computing optimize data processing and dynamic resource allocation, ensuring more efficient and adaptive responses to evolving disaster conditions.

The trial, illustrated in **Figure 5**, took place in Aveiro, Portugal, utilizing Imagine-B5G's local facility, comprising IT-Aveiro, Altice Labs, Capgemini, and Ubiwhere, for integration and demonstration to end-users and stakeholders such as the Ílhavo Municipality, Civil Protection, and the local fire brigade. The city's vulnerability to urban flooding made it an ideal facility to validate the platform's disaster response capabilities.

Figure 5: SAFER-FLOW Trial (a: Dashboard of command Control Centre and 5G wearable sensors; b: 5G Water Level Sensor; c: AR point of view with drone video; d: AI/ML Watercourse Detection e: 5G Water Level Sensor showcase)



Pilot Name: **6G Enabled Integrated Care (CRITICAL)** [30], [37]

Company Name: **RedZinc**

Vertical Sector: **PPDR**

Use Case Description:

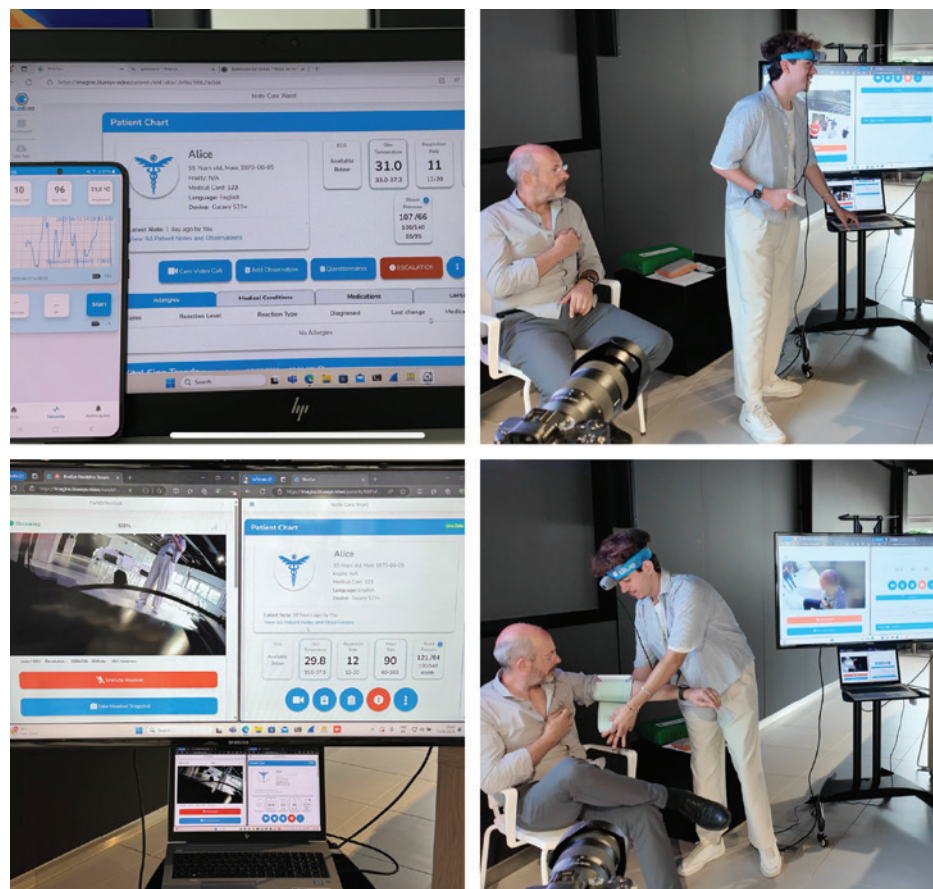
By combining 5G and MEC, Altice Labs helps transform emergency healthcare with secure, efficient, and dedicated connectivity

The CRITICAL project aims to enhance emergency medical care by enabling paramedics to communicate with medical professionals in real time. This is achieved using several wearable sensors alongside live video streaming, allowing emergency responders to transmit vital signs and patient data to specialists before reaching the hospital. This solution speeds up diagnosis and treatment for critical patients, while enabling medical professionals to assess the severity of the situation in advance.

In addition, the system supports treating non-critical patients at the scene, reducing hospital strain. With the use of Altice Labs 5G technology and Mobile Edge Computing (MEC), patient privacy is ensured, and enables emergency healthcare in an efficient, prioritized way through a dedicated slice.

The use case was validated through three trials, one of which is represented in **Figure 6**. The most recent involved PPDR users from Aveiro's firefighters, during which a paramedic simulated an emergency response, transmitting real-time video and sensor data to a remote medical team. The feedback from participating firefighters was especially valuable, offering insights and suggestions to further improve the solution.

Figure 6: CRITICAL Trial (a: Hospital BlueEye Dashboard and Paramedic BlueEye Care app; b: Biometric Sensor Showcase by 5G; c: Hospital BlueEye Dashboard with Live POV video and Patient vital signs; d: Biometric device applied to patient)



NextGen Mobility

Pilot Name: **V2X/D2D Communications**

Company Name: **Allbesmart** [38]

Vertical Sector: **Automotive**

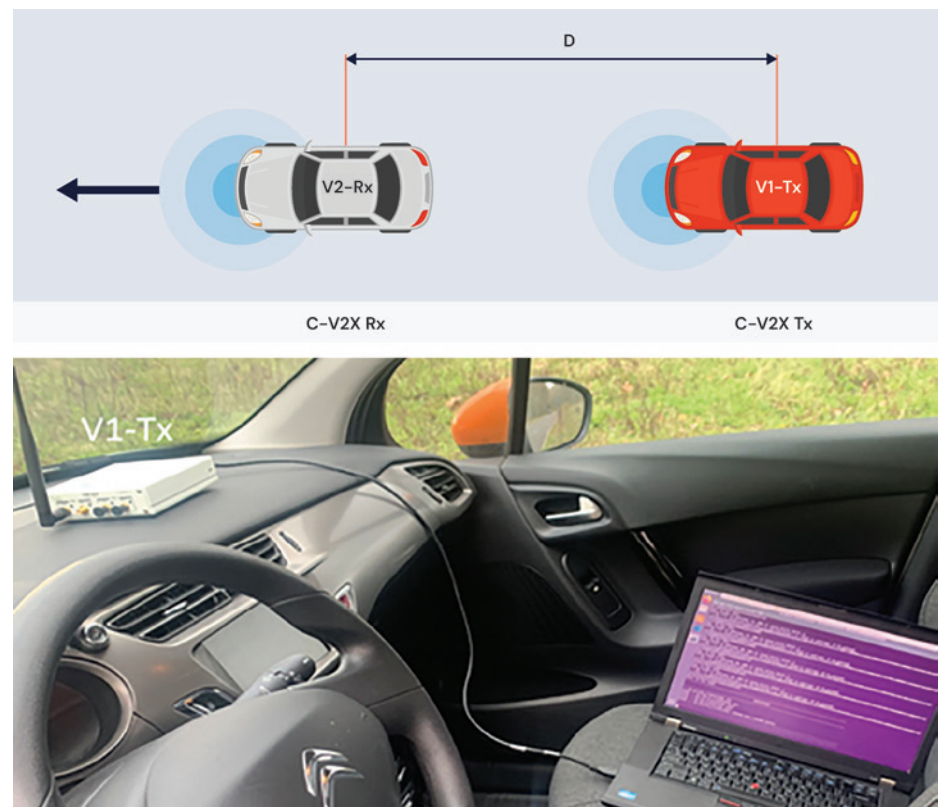
Use Case Description:

Cooperative Intelligent Transport Systems (C-ITS) enable direct communication between vehicles, roadside infrastructure, and other road users to share real-time data about the road environment. Therefore, this system overcomes some limitations of public mobile networks, as it requires no additional infrastructure besides the On-Board Units (OBU) for Vehicle-to-Vehicle (V2V) communication, and significantly reduces latency. This last characteristic is mandatory for such demanding scenarios.

To support this, the pilot implemented the PC5/Sidelink radio interface for Mode 4 [39], using Software Defined Radio (SDR), enabling direct V2V communication between two OBU.

The developed prototype successfully performed discovery, synchronization, and data transfer processes between C-V2X devices. C-ITS services were tested with a 95% interoperability success rate, demonstrating a high degree of compatibility with C-V2X implementations from different manufacturers. Additional tests were performed to evaluate radio coverage and immunity to interference from ITS-G5 systems in adjacent frequency bands. **Figure 7** represents one of the V2X/D2D Communications trials.

Figure 7: V2X/D2D Communications trial



Pilot Name: **Augmented Reality in Automotive Driving**

Company Name: **OneSource** [40]

Vertical Sector: **Automotive**

Use Case Description:

Most modern vehicles are equipped with video cameras and local processing capabilities. Through video analytics and/or high-definition road maps, many can identify and alert drivers to road signs, thus increasing driving safety. However, most vehicles currently on the road lack this capability, creating a business opportunity to retrofit them using existing technologies.

In this pilot, Microsoft HoloLens, an Augmented Reality/Mixed Reality (AR/MR) headset developed and manufactured by Microsoft, was used, with connectivity between the handset and the Mobile Edge Computing (MEC) provided by a Mobile Computing Unit equipped with WLAN and mobile connectivity (2G/4G/5G). The AR component can be observed in **Figure 8**.

Specific models were trained to identify traffic signs, with processing outsourced from the vehicle to the MEC at the Altice Labs Data Center. This approach offers several benefits, such as greater processing capacity, shared resources across multiple users, and the elimination of the need for frequent hardware and algorithms updates, allowing the service to evolve and improve over time.

Figure 8: Augmented Reality in Automotive Driving Trial



Pilot Name: **Smart Lamp Pole**

Company Name: **Omniflow** [41], [42]

Vertical Sector: **Smart Cities**

Use Case Description:

The Omniflow Smart Lamp Pole is powered by wind and solar energy, featuring integrated energy storage and grid connectivity. Its versatile design allows for the simultaneous integration of multiple applications, ranging from public safety systems to environmental condition monitoring. Installing smart luminaires on existing poles offers a sustainable solution that benefits urban spaces without requiring full infrastructure replacement.

In this pilot, Omniflow evolved its solution to support 5G connectivity. Remote wireless access for monitoring, configuration, and data collection is essential for integrating the smart lamp with broader smart city systems. The implemented 5G modem supports multi-SIM functionality, enabling failover between non-public 5G networks (Altice Labs 5G laboratory) and public ones (MEO). The 5G technology met the system's requirements, particularly in accessing air quality monitoring data collected by the smart lamp, as demonstrated in **Figure 9**.

Figure 9: Smart Lamp Pole Trial



Pilot Name: **QS-P Client command line**

Company Name: **QuantumNova**

Vertical Sector: **Security**

Use Case Description:

A robust cryptographic system is essential for ensuring the confidentiality, integrity, and authenticity of communications. With the emergence of quantum computing, many traditional cryptographic algorithms are becoming vulnerable, underscoring the urgent need for Post-Quantum Cryptography (PQC), algorithms designed to withstand attacks from both classical and quantum computers. Despite its promise, PQC introduces new challenges: it often requires significantly greater computational power, memory, and energy. These demands make its adoption particularly difficult in IoT environments, where devices are inherently resource-constrained.

To address this, the pilot project focused on developing and testing a PQC solution specifically optimized for such limited-resource IoT systems.

A secure tunnel with bilateral authentication and confidentiality guarantees was implemented over a point-to-point (server-client) connection between a streaming platform and a video camera. This tunnel leverages PQC [43] methods, specifically Module Lattice Key Encapsulation Mechanism (ML-KEM) and Module Lattice Digital Signature Algorithm (ML-DSA), integrated into a VPN system based on OpenVPN. Additionally, client infrastructure was developed to ensure post-quantum security for the tunnel, along with a CLI client for the QS-P Network (whose trial is illustrated in **Figure 10**), enabling seamless integration with IoT devices.

To evaluate the solution in a wireless scenario, a connection between an IoT device and a server was established using the IoT's device 5G modem and the Altice Labs 5G laboratory, successfully streaming high-quality video from the camera.

Figure 10: QS-P Client command line Trial

The screenshot displays the Altice Labs QS-P Client interface. At the top, there are four summary cards: 'CONNECTED SINCE' (2025-02-18 18:28:00), 'LAST TIME ONLINE' (0h 49m 24s), 'TOTAL BYTES IN' (965.45 GB), and 'TOTAL BYTES OUT' (121 GB). Below these is a search bar for 'Public IP' and a table of connections. The table has columns for Public IP, Local IP, Location, Connected Since, Bytes Received, Bytes Sent, Last Ref., and Time Online.

Public IP	Local IP	Location	Connected Since	Bytes Received	Bytes Sent	Last Ref.	Time Online
103.140.07.03:55067	13.88.0.13	Porto, PT	2025-02-18 18:28:00	965.45 GB	121 GB	Tue, 18 Feb 2025 18:28:04 GMT	0h 49m 24s
95.69.3.220:4102	13.88.0.13	Porto, PT	2025-02-17 18:49:34	174 TB	6.69 GB	Fri, 18 Feb 2025 18:58:04 GMT	23h 34m 30s
95.69.3.220:55789	13.88.0.13	Porto, PT	2025-02-17 12:52:42	6.17 TB	12.38 GB	Mon, 17 Feb 2025 18:58:24 GMT	5h 43m 42s

Conclusions

The 5G pilot initiatives carried out at Altice Labs have clearly demonstrated the significant potential of advanced connectivity to drive innovation across multiple sectors. By enabling startups and SMEs to not only validate but also enhance and optimize their products and solutions, these pilots have played a crucial role in increasing their TRL.

The continuous evolution of Altice Lab's infrastructure has proven essential in supporting increasingly complex and demanding scenarios, ensuring that the environment remains at the forefront of technological experimentation. At the same time, potential strategic partnerships were identified, both commercial and research-oriented, to generate value and long-term collaboration.

Importantly, these efforts are not limited to immediate outcomes. By contributing to the definition of Key Performance Indicators (KPI) and relevant target values, Altice Labs is actively shaping the roadmap toward 6G, ensuring that future networks are built on solid, real-world insights [35]. Furthermore, the identification of emerging technologies and new players within the national and European innovation landscape reflects a broader commitment to expand and strengthen the 5G ecosystem.

Altice Labs continues to affirm its role as a trusted enabler of innovation, fostering meaningful collaboration between academia, industry, and end-users. These 5G pilots are a testament to the organization's strategic vision and its ongoing contribution to the digital transformation of society.

Altice Labs' 5G pilots drive innovation, elevate technology readiness, and shape the roadmap toward 6G

References

- [1] Imagine Beyond 5G. "Home Page." Imagine-B5G. Accessed: Oct. 17, 2025. [Online]. Available: <https://imagineb5g.eu/>
- [2] Next Gen Mobility. "Home Page." Next Gen Mobility. Accessed: Oct. 17, 2025. [Online]. Available: <https://nextgenmobility.pt/>
- [3] ASOCS Ltd., "CYRUS Industrial 5G Network." ASOCS. Accessed: Oct. 17, 2025. [Online]. Available: <https://asocscscloud.com/cyrus/>
- [4] Software Radio Systems. "srsRAN Enterprise 5G." SRS. Accessed: Oct. 17, 2025. [Online]. Available: <https://srs.io/srsran-enterprise-5g/>
- [5] Allbesmart, "Choose the right OIABOX for your needs." OIABOX. Accessed: Oct. 17, 2025. [Online]. Available: <https://oaibox.com/#products>
- [6] Altice Labs. "5G Small Cells." Altice Labs. Accessed: Oct. 17, 2025. [Online]. Available: <https://www.alticelabs.com/products/5g-small-cells/>
- [7] ASOCS Ltd. "Home Page." ASOCS. Accessed: Oct. 17, 2025. [Online]. Available: <https://asocscscloud.com/>
- [8] Benetel. "5G Split 7.2x Outdoor O-RU." Benetel. Accessed: Oct. 22, 2025. [Online]. Available: <https://benetel.com/ran650/>
- [9] Ettus Research. "Products." Ettus Research. Accessed: Oct. 22, 2025. [Online]. Available: <https://www.ettus.com/products/>
- [10] Druid Software. "Raemis – Cellular Network Technology." Druid Software. Accessed: Oct. 22, 2025. [Online.] Available: <https://druidsoftware.com/raemis-cellular-network-technology/>
- [11] Open5GS. "Home Page." Open5GS. Accessed: Oct. 22, 2025. [Online]. Available: <https://open5gs.org/>
- [12] Proxmox Server Solutions. "Proxmox Virtual Environment." Proxmox. Accessed: Nov. 3, 2025. [Online]. Available: <https://www.proxmox.com/en/products/proxmox-virtual-environment/overview>
- [13] One Source. "Home Page." OneSource. Accessed: Oct. 23, 2025. [Online]. Available: <https://onesource.pt/>
- [14] RedZinc Services. "Home Page." RedZinc. Accessed: Oct. 23, 2025. [Online]. Available: <https://redzinc.net/>
- [15] A to Be. "Home Page." A to Be. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.a-to-be.com/>
- [16] YouShip. "A following and carrying delivery robot." YouShip. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.you-ship.com/>
- [17] Omniflow. "Home Page." Omniflow. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.omniflow.io/>

- [18] Cibersur. "Home Page." Cibersur. Accessed: Oct. 23, 2025. [Online]. Available: <https://cibersur.eu/>
- [19] JSIO. "Home Page." JSIO. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.jsio.pt/>
- [20] Bold Robotics. "Home Page." Bold Robotics. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.boldrobotics.pt/>
- [21] TestBed | NextGen Mobility, "Já conhece a BOLD ROBOTICS LLC?" LinkedIn. Accessed: Nov. 5, 2025. [Online]. Available: https://www.linkedin.com/posts/nextgen-mobility_nextgenmobilitychallenge-inovaaexaeto-startups-activity-7272558319855153152-_MRp
- [22] Altice Labs. "Altice Labs drives innovation in mobility with 5G network solution." AlticeLabs. Accessed: Nov. 5, 2025. [Online]. Available: <https://www.alticelabs.com/blog/altice-labs-drives-innovation-in-mobility-with-5g-network-solution/>
- [23] QuantumNova. "Home Page." QuantumNova. Accessed: , 2025. [Online]. Available: <https://quantumnova.xyz/>
- [24] SafetyScope. "Home Page." SafetyScope. Accessed: Oct. 23, 2025. [Online]. Available: <https://www.safetyscope.eu/>
- [25] Allbesmart, Lda. "Home Page." Allbesmart. Accessed: Oct. 23, 2025. [Online]. Available: <https://allbesmart.pt/>
- [26] Imagine Beyond 5G. "SRS-B5G at Imagine-B5G Open Call #1." Imagine-B5G. Accessed: Oct. 23, 2025. [Online]. Available: <https://imagineb5g.eu/srs-b5g-at-imagine-b5g-open-call-1/>
- [27] SoftwareRadio Systems. "SoftwareRadio Systems Deploys srsRAN Enterprise 5G at Altice Labs' Aveiro Innovation Hub." SRS. Accessed: Oct. 26, 2025. [Online]. Available: <https://srs.io/srs-enterprise-5g-altice-labs-deployment/>
- [28] Imagine Beyond 5G. "5G-EVER in Imagine-B5G Open Call #3." Imagine-B5G. Accessed: Oct. 26, 2025. [Online]. Available: <https://imagineb5g.eu/5g-ever-in-imagine-b5g-open-call-3/>
- [29] Imagine Beyond 5G. "ADS-5G in Imagine-B5G Open Call #3." Imagine-B5G. Accessed: Oct. 26, 2025. [Online]. Available: <https://imagineb5g.eu/ads-5g-in-imagine-b5g-open-call-3/>
- [30] Imagine Beyond 5G, "Use cases, Ecosystem Analysis and KVs/KPIs for OC1 and OC2." Rep. no. D1.1, 2024. Accessed: Oct. 26, 2025. [Online]. Available: <https://drive.google.com/file/d/1J-jbfr7-SJgbcYkQKlu2ZgqwaEuyWZx/view>
- [31] Imagine Beyond 5G, "Vertical Trials and Pilots Demonstrations over IMAGINE-B5G Platform." Rep. no. D3.3, 2024. Accessed: Oct. 26, 2025. [Online]. Available: https://drive.google.com/file/d/1QTbPdIJe4oF_zMCtH2DwNYuSr2aK-JkJ/view
- [32] Imagine Beyond 5G, "Intermediate Validation of B5G Core Technologies." Rep. no. D4.5, 2024. Accessed: Oct. 26, 2025. [Online]. Available: <https://drive.google.com/file/d/13AJzIwx2J9GEKIGpsMiayhuncCeokpmf/view>

- [33] Imagine Beyond 5G. "SAFER-FLOW at Imagine-B5G Open Call #1." Imagine-B5G. Accessed: Nov. 3, 2025. [Online]. Available: <https://imagineb5g.eu/safer-flow-at-imagine-b5g-open-call-1/>
- [34] 6G SNS, "SNS Trials & Pilots." Brochure No.1, 2025. [Online]. Accessed: Nov. 3, 2025. Available: https://smart-networks.europa.eu/wp-content/uploads/2025/06/sns_tps_brochure_may25_final.pdf
- [35] I. Mesogiti et al., "6G KPIs – Definitions and Target Values," *Zenodo*, 2025, doi: 10.5281/zenodo.15011631.
- [36] Altice Labs. "5G Innovation: Spotlight on SAFER-FLOW and DCA Pilot Projects." Altice Labs. Accessed: Nov. 5, 2025. [Online.] Available: <https://www.alticelabs.com/blog/5g-innovation-spotlight-on-safer-flow-and-dca-pilot-projects/>
- [37] Imagine Beyond 5G. "Project CRITICAL in Imagine-B5G Open Call #2." Imagine-B5G. Accessed: Nov. 5, 2025. [Online]. Available: <https://imagineb5g.eu/project-critical-in-imagine-b5g-open-call-2/>
- [38] TestBed | NextGen Mobility, "A inovação continua a ganhar força com mais um piloto apoiado pela nossa Test Bed!" LinkedIn. Accessed: Nov. 5, 2025. [Online]. Available: https://www.linkedin.com/posts/nextgen-mobility_nextgenmobilitychallenge-inovaaexaeto-startups-activity-7354475786805927936-93xH
- [39] L. Miao, J. J. Virtusio, and K.-L. Hua, "PC5-Based Cellular-V2X Evolution and Deployment," *Sensors*, vol. 21, no. 3, p. 843, 2021, doi: 10.3390/s21030843.
- [40] TestBed | NextGen Mobility, "Mais uma empresa apoiada no âmbito da nossa Test Bed!" LinkedIn. Accessed: Nov. 5, 2025. [Online]. Available: https://www.linkedin.com/posts/nextgen-mobility_nextgenmobilitychallenge-inovaaexaeto-startups-activity-7351547823366012928-jbNO
- [41] TestBed | NextGen Mobility, "Mais uma empresa apoiada no âmbito da nossa Test Bed!" LinkedIn. Accessed: Nov. 5, 2025. [Online]. Available: https://www.linkedin.com/posts/nextgen-mobility_nextgenmobilitychallenge-inovaaexaeto-startups-activity-7253034471539523584-dWuO
- [42] Altice Labs. "Installation of a smart pole on the Altice Labs campus." Altice Labs. Accessed: Nov. 5, 2025. [Online.] Available: <https://www.alticelabs.com/blog/installation-of-a-smart-pole-on-the-altice-labs-campus/>
- [43] Cloudflare. "What Is Post-Quantum Cryptography?" Cloudflare. Accessed: Nov. 5, 2025. [Online.] Available: <https://www.cloudflare.com/learning/ssl/quantum/what-is-post-quantum-cryptography/>



12

The Next Step for NG-PON2: 50G TWDM-PON

Abstract As broadband demand accelerates and digital services become increasingly bandwidth-intensive, the evolution of Next-Generation Passive Optical Networks (NG-PON2) is entering a pivotal phase. This article explores the transition toward 50G Time and Wavelength Division Multiplexed PON (TWDM-PON) as the natural progression of NG-PON2. We examine the technological advancements enabling 50G TWDM-PON, including enhanced optical components, advanced modulation formats, and improved digital signal processing. The paper also compares the performance, scalability, and deployment considerations of 50G TWDM-PON. This assessment provides a forward-looking perspective on how 50G TWDM-PON can serve as a cornerstone for future-proof fiber access networks.

Authors **Cláudio Rodrigues**, Altice Labs, claudio-e-rodrigues@alticelabs.com
Fernando Guiomar, Instituto de Telecomunicações (IT), guiomar@av.it.pt
Francisco Rodrigues, PICadvanced, francisco@picadvanced.com
Jacklyn Reis, PICadvanced, jacklyn.reis@picadvanced.com
João Santos, PICadvanced, joao.santos@picadvanced.com
Ana Tavares, PICadvanced, ana@picadvanced.com
António Teixeira, Instituto de Telecomunicações and PICadvanced, teixeira@av.it.pt

Keywords PON, NG-PON2, 50G TWDM-PON

DOI 10.57650/INNOVATION/2025.12



Introduction

Broadband networks are undergoing a paradigm shift as digital ecosystems demand unprecedented capacity and reliability. Applications such as 8K video streaming, virtual reality (VR), industrial Internet of Things (IoT), and 5G/6G mobile backhaul require symmetric bandwidth and deterministic latency. While Gigabit Passive Optical Network (GPON) and 10-Gigabit-capable symmetric PON (XGS-PON) have served well for the past decade, their aggregate capacity is insufficient for emerging services. NG-PON2 introduced TWDM-PON with up to 40 Gb/s aggregate capacity, but this is no longer adequate for the projected traffic growth beyond 2030.

The ITU-T G.9804.x series addresses this challenge by defining higher-speed PON systems, including 50G Time Division Multiplexing PON (TDM-PON) and 50G TWDM-PON [1]. These systems promise scalable architectures, coexistence with legacy technologies, and support for diverse applications. However, achieving 50 Gb/s per channel over cost-sensitive optical distribution networks (ODN) introduces significant technical challenges, particularly in modulation, dispersion management, and coexistence.

Standards framework: ITU-T G.9804.1

ITU-T G.9804.1 provides the baseline requirements for higher-speed PON systems:

- **Nominal Line Rate:** 50 Gb/s per wavelength channel, symmetric or asymmetric configurations.

- **Reach and Split Ratio:** minimum 20 km reach with 1:64 split; extendable to 60 km and 1:256 split using reach extenders.
- **Coexistence:** mandatory support for GPON, XGS-PON, and NG-PON2 via coexistence elements (CEX).
- **Service Requirements:** support for residential broadband, enterprise connectivity, and mobile fronthaul/backhaul with latency <1.5 ms and synchronization accuracy for 5G.
- **Spectral Flexibility:** operation in O-band for low dispersion, with provisions for C/L-band extensions.

These requirements ensure interoperability, scalability, and smooth migration from legacy systems [1].

System requirements for 5G TWDM-PON

According to ITU-T G.9804.1 [1], the 5G TWDM-PON system must support:

- Multiple wavelength channel pairs multiplexed on the same fiber, each operating in TDM/Time Division Multiple Access (TDMA) mode;
- Tunable Optical Network Units (ONU) capable of operating on any wavelength channel under Optical Line Terminal (OLT) control;
- Service differentiation by assigning distinct wavelength channels to services with different characteristics and Quality of Service (QoS) requirements;
- Pay-as-you-grow scalability by adding wavelength channels as demand increases;
- Non-disruptive ONU activation on service-critical channels;
- Wavelength service protection with restoration times ≤ 50 ms;
- Channel bonding for ONU service rates beyond a single channel;
- Load balancing across available wavelength channels;
- Rogue ONU Mitigation: Multi-wavelength techniques to isolate misbehaving ONU.

Architecture and migration

Logical architecture

The 5G TWDM-PON architecture builds upon NG-PON2 but introduces higher-speed channels and advanced flexibility. At the OLT, multiple channel terminations (CT) are connected via a wavelength multiplexer (WM), enabling simultaneous operation of multiple wavelength pairs over a single fiber [2]. Each wavelength channel operates in TDM/TDMA mode, supporting dynamic bandwidth allocation and coexistence with legacy systems.

A single fiber becomes a shared medium for multiple wavelength channels, coordinated through the multiplexer

Key architectural features [1], [2]:

- **Tunable ONU:** ONU can dynamically tune to any available wavelength channel under OLT control, reducing inventory complexity and enabling colorless ONU deployment.
- **Channel Bonding:** Advanced ONU can aggregate multiple wavelengths to exceed the maximum service rate of a single channel, supporting enterprise and mobile backhaul scenarios.
- **Spectral Flexibility:** The system can operate in O-band for low dispersion or extend to C/L bands for long-reach applications using Erbium Fiber Doped Amplifiers (EDFA).
- **Protection and Resilience:** ITU-T G.9804.1 introduces Type B, C, and wavelength protection schemes, leveraging wavelength tunability for rapid failover (<50 ms) in case of OLT card or fiber failure.

The architecture supports pay-as-you-grow scalability, allowing operators to start with a minimal wavelength set and expand capacity by adding channels as demand increases. This modular approach reduces initial capital expenditure (CAPEX) while ensuring future-proof growth.

Migration scenarios

Migration from legacy PON systems to 50G TWDM-PON is critical for operators with extensive installed ODN. ITU-T G.9804.1 outlines several strategies [1]:

- **Brownfield Deployment:** Coexistence with GPON, XGS-PON, NG-PON2, and 50G TWDM-PON on the same ODN using CEx, as depicted in **Figure 1**. This approach minimizes service disruption and leverages existing splitter-based infrastructure.
- **Hybrid Coexistence:** For triple coexistence (GPON, XGS-PON, and 50G TWDM-PON), hybrid WDM/TDM techniques are employed. Legacy systems occupy their wavelength bands, while TWDM-PON uses separate bands or shared spectrum with guard bands.
- **Full Migration (Greenfield):** In new deployments or complete overhauls, operators can implement 50G TWDM-PON exclusively, enabling optimized wavelength plans, higher split ratios (up to 1:256), and advanced features like channel bonding and slicing.
- **Stepwise Upgrade:** Operators may adopt a phased approach:
 1. Upgrade from GPON to XGS-PON.
 2. Introduce 50G TWDM-PON alongside XGS-PON.
 3. Gradually retire legacy systems as customer migration completes.

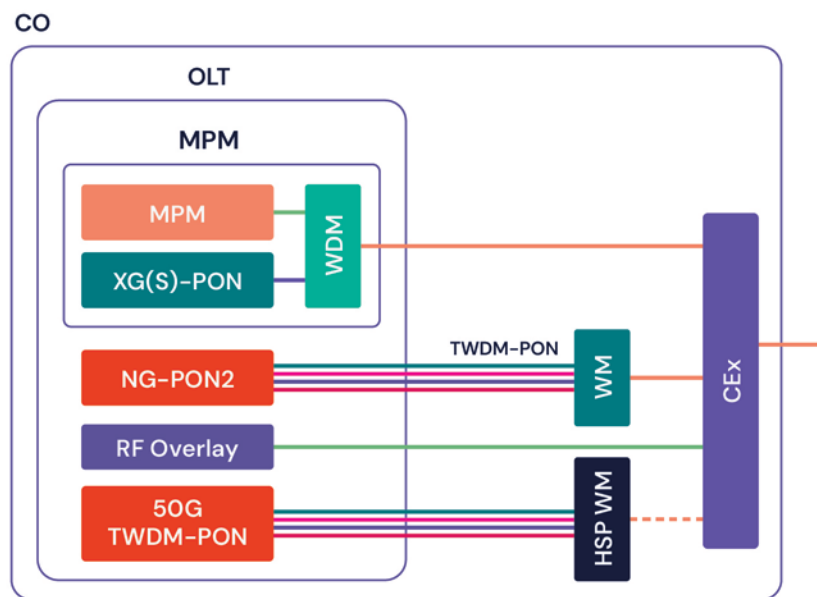
The deployment of 50G TWDM-PON introduces technical and operational challenges that must be addressed to ensure reliable performance and seamless migration:

- **Spectrum Management:** The O-band spectrum allocated for PON systems is inherently limited, and the coexistence of multiple technologies such as

GPON, XGS-PON, and TWDM-PON further constrains available wavelengths. Careful wavelength planning and allocation strategies are essential to prevent interference and maintain guard bands for legacy systems. Advanced wavelength routing and dynamic tuning mechanisms at the OLT and ONU level can mitigate these constraints.

- **Guard Band Requirements:** Coexistence with RF-video overlays and legacy optical filters necessitates the introduction of guard bands between adjacent wavelength channels. These guard bands reduce spectral efficiency and require precise optical filtering to avoid crosstalk. The challenge is compounded in scenarios involving triple coexistence, where multiple generations of PON technologies share the same optical distribution network.
- **Operational Complexity:** Managing a heterogeneous environment that includes GPON, XGS-PON, and 5G TWDM-PON demands sophisticated OLT scheduling algorithms and inventory control systems. Operators must implement dynamic bandwidth assignment (DBA) across multiple wavelength channels while ensuring service-level agreements for diverse applications such as residential broadband, enterprise connectivity, and mobile fronthaul. Additionally, provisioning colorless ONU and supporting wavelength agility introduce further complexity in network planning and maintenance.

Figure 1: Coexistence scenario



Physical layer and performance

The physical layer of 5G TWDM-PON is designed to meet stringent requirements for capacity, reach, and coexistence while maintaining compatibility with legacy ODN. ITU-T G.9804.1 specifies that each wavelength channel shall support a nominal line rate of approximately 50 Gb/s in both downstream and upstream directions, with symmetric or asymmetric configurations [1]. This baseline ensures that even a single channel can deliver service rates up to 40 Gb/s per ONU, accommodating high-bandwidth applications such as enterprise connectivity and mobile fronthaul.

5G TWDM-PON meets stringent demands for reach and coexistence, ensuring legacy ODN remain future-proof

The aggregate capacity of a TWDM-PON system scales with the number of wavelength channels deployed. For example, a configuration with four wavelengths operating at 50 Gb/s each achieves an aggregate capacity of ~200 Gb/s over a single fiber. Research demonstrations have shown that coherent modulation formats, such as single-polarization quadrature phase-shift keying (SP-QPSK), can double the per-channel rate to ~100 Gb/s, enabling aggregate capacities exceeding 400 Gb/s with four wavelengths [3], [4]. It should be noted that this extrapolation is based on experimental literature and is not part of ITU-T G.9804.1 requirements.

The standard mandates a minimum reach of 20 km for general applications and 10 km for latency-sensitive scenarios such as 5G fronthaul [1]. An extended reach of up to 60 km is achievable through mid-span reach extenders, provided that the optical power budget is carefully managed. The optical loss budget must account for fiber attenuation, splitter losses, connector insertion losses, and penalties due to chromatic dispersion and nonlinear effects. ITU-T G.9804.1 recommends compatibility with legacy fiber types (ITU-T G.652 and G.657) to ensure smooth migration [1].

Higher split ratios improve infrastructure sharing but impose stricter optical budget constraints. The standard supports a minimum split ratio of 1:64, with provisions for 1:256 in TWDM-PON deployments [1]. Achieving these ratios requires advanced optical amplification or reach extension techniques, particularly in scenarios that combine high split ratios with long reach.

The operating wavelength band for 50G TWDM-PON is primarily the O-band (1260–1360 nm) due to its low chromatic dispersion, which simplifies Intensity Modulation / Direct Detection (IM/DD) transmission. However, the C-band and L-band remain attractive for extended reach applications because of their compatibility with EDFAs and the possibility of repurposing the NG-PON2 ecosystem and therefore enabling a “true” evolved TWDM-PON. Migration to these bands introduces challenges such as higher dispersion and filter coexistence with legacy systems, necessitating advanced Digital Signal Processing (DSP) for the IM/DD implementations or even full coherent technology [5].

Key performance indicators include [1]:

- **Bit Error Ratio (BER):** Compliance with BER thresholds at the Forward Error Correction (FEC) decoder input (typically 10^{-3} pre-FEC) ensures post-FEC BER of 10^{-12} or better.
- **Dynamic Range:** Defined by the ratio of receiver overload to sensitivity, ensuring robust operation under varying optical power conditions.
- **Latency:** Sub-millisecond latency for broadband services and $<200 \mu\text{s}$ for fronthaul applications.

While IM/DD employing Pulse Amplitude Modulation with 4 levels (PAM4) remains the most cost-effective solution for near-term deployments, its scalability is limited by dispersion and bandwidth constraints. Coherent detection offers superior dispersion tolerance and reach, enabling higher per-channel rates and advanced features such as polarization multiplexing and probabilistic constellation shaping [6], [4].

Technology options

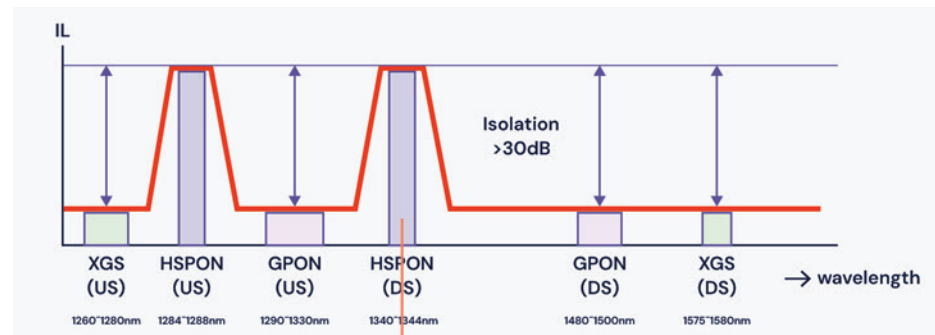
IM/DD transmission

IM/DD remains the most practical solution for near-term 50G TWDM-PON deployments due to their simplicity, cost-effectiveness, and compatibility with existing ODN. Unlike coherent systems, IM/DD does not require local oscillators or complex DSP for phase recovery, making it attractive for mass-market access networks. However, scaling IM/DD to 50 Gb/s per channel introduces significant challenges related to chromatic dispersion, bandwidth limitations, and inter-channel interference.

Experimental validation

As shown in **Figure 2**, recent experimental work [7] demonstrated a multi-wavelength IM/DD system operating at 50 Gb/s per channel in the O-band with 100 GHz channel spacing, achieving transmission over 20 km of standard single-mode fiber (SSMF) without any optical or electrical dispersion compensation.

Figure 2: 50G TWDM O band vs C/L band –50G-PON O band for 50G TWDM



Frequency (THz)	Wavelength (nm)
223	1344,36
223,1	1343,75
223,2	1343,15
223,3	1342,55
223,4	1341,95
223,5	1341,35
223,6	1340,75
223,7	1340,15

The setup employed external modulated lasers (EML) with integrated semiconductor optical amplifiers (SOA) to boost launch power. Key findings include:

- **Single-Channel Performance:** Back-to-back eye diagrams showed an extinction ratio (ER) of 5.9 dB without a wavelength multiplexer (WM) and 5.5 dB

with WM insertion. Detuning tests revealed that ER degradation remained within 0.5 dB for spectral excursions up to ± 34 GHz, indicating robust tolerance to wavelength drift and compatibility with NG-PON2 control systems.

- **Multi-Channel Operation:** Tests with two and three channels confirmed negligible penalties from four-wave mixing (FWM) and linear crosstalk. ER variation on the probing channel was less than 0.3 dB across back-to-back, 10 km, and 20 km fiber scenarios. This validates the feasibility of dense wavelength packing (100 GHz spacing) for TWDM-PON without resorting to complex nonlinear mitigation techniques.
- **Coexistence with Legacy Technologies:** Integration with GPON and XGS-PON via a CEx resulted in penalties below 0.5 dB on OLT sensitivity and overload, confirming that 50G IM/DD channels can coexist with legacy systems without service disruption.

Advantages of IM/DD transmission

IM/DD transmission offers several practical benefits that strongly support its widespread adoption, including:

- **Cost Efficiency:** IM/DD leverages mature laser technology, which is widely available and optimized for O-band operation. By avoiding the need for expensive coherent optics, local oscillators, and high-complexity DSP, IM/DD significantly reduces both CAPEX and operational expenditure (OPEX), making it ideal for large-scale deployments in access networks.
- **Low System Complexity:** Unlike coherent systems, IM/DD does not require phase recovery, polarization tracking, or advanced digital signal processing for carrier synchronization. This simplicity translates into lower power consumption, smaller form factors, and reduced hardware requirements at the ONU, which is critical for cost-sensitive residential and business applications.
- **Compatibility with Existing Infrastructure:** IM/DD systems operate in the O-band (1260–1360 nm), where chromatic dispersion is minimal, enabling high-speed transmission without complex dispersion compensation. This characteristic allows seamless reuse and coexistence of ODNs designed for GPON and XGS-PON, facilitating smooth migration and coexistence with minimal changes to the outside plant.

Limitations of IM/DD transmission

Despite its practical benefits, IM/DD transmission also presents technical limitations that must be considered:

- **Dispersion and Bandwidth Constraints:** At line rates of 50 Gb/s, even the relatively low chromatic dispersion in the O-band (approximately 0 ps/nm·km) combined with finite component bandwidths (typically around 40 GHz) introduces signal distortion and eye closure. These impairments necessitate advanced equalization techniques and optimized filtering to maintain acceptable BER over standard single-mode fiber spans.

- **Scalability Beyond 50 Gb/s:** Extending IM/DD systems beyond 50 Gb/s per channel is inherently challenging due to the limitations of intensity modulation formats and the absence of phase information. While higher-order modulation schemes such as PAM4 can increase spectral efficiency, they require sophisticated DSP for dispersion compensation and error correction, which adds cost and complexity to ONU designs.
- **Power Budget Constraints:** Achieving high split ratios (e.g., 1:128 or 1:256) and extended reach scenarios (up to 60 km) imposes stringent optical power requirements. These conditions often demand the use of optical amplification or mid-span reach extenders, which can increase deployment costs and operational complexity.

Future enhancements

To overcome these limitations and extend the viability of IM/DD transmission for next-generation PON systems, ongoing research focuses on several key areas:

- **Advanced Modulation Formats:** Techniques such as PAM4 and duobinary signaling, combined with FEC and DSP-based equalization, are being explored to mitigate dispersion and improve tolerance to bandwidth constraints [5].
- **Enhanced Optical Components:** Development of high-bandwidth EML, low-loss wavelength multiplexers, and improved sensitivity and high bandwidth photodetectors will enable better signal integrity and higher aggregate capacity.
- **Machine-Learning-Based Equalization:** Adaptive algorithms leveraging machine learning are emerging as promising solutions for real-time impairment compensation, offering improved performance under dynamic network conditions without excessive computational overhead.



Coherent detection, once reserved for long-haul transport, now challenges IM/DD as PON races beyond 50 Gb/s

Coherent transmission and detection

Originally reserved for high-capacity transport links in long-haul and metro networks, coherent detection is now emerging as a practical candidate for next-generation PON exceeding 50 Gbps. This shift is driven by its superior spectral efficiency, modulation flexibility, and robustness against fiber impairments. However, adopting coherent technology in access networks marks a major paradigm change, as the entire PON ecosystem has historically been designed around IM/DD. Consequently, sourcing, interoperability, and technological readiness must be carefully addressed. Moreover, the associated higher power consumption, hardware complexity, and cost remain key challenges compared to the simplicity, maturity, and efficiency of IM/DD systems.

Simplified coherent PON: Single-Polarization QPSK approach

Simplified coherent transceivers have been investigated as a viable path toward future PON implementations, with most of the research focusing on reducing receiver complexity [8]. While the advantages of coherent detection are well established, the technical challenge for its penetration in access networks lies in translating those capabilities into system architectures that meet the stringent cost and power constraints of ONU.

QPSK modulation enables reuse of proven 100G components, lowering entry costs for next-gen PON

A particularly effective direction toward simplification is the sacrifice of one of the available axes in the modulation vector space, the polarization, allowing the design of receivers that recover only a single polarization component [9]. This principle led to the development of SP-QPSK as a candidate technological basis for 50G TWDM-PON. By removing polarization diversity, the system eliminates what is essentially duplicated hardware and DSP, resulting in a simpler and more power-efficient coherent front-end. Additionally, the choice of QPSK as the modulation format (2 bits per symbol) allows the reuse of optoelectronics developed for the first generation of long-haul and metro coherent deployments, such as 100 G ZR and related standards. This reuse paves the way for lower entry cost and easier technological availability, as those components are now mature and being replaced by higher-order coherent solutions.

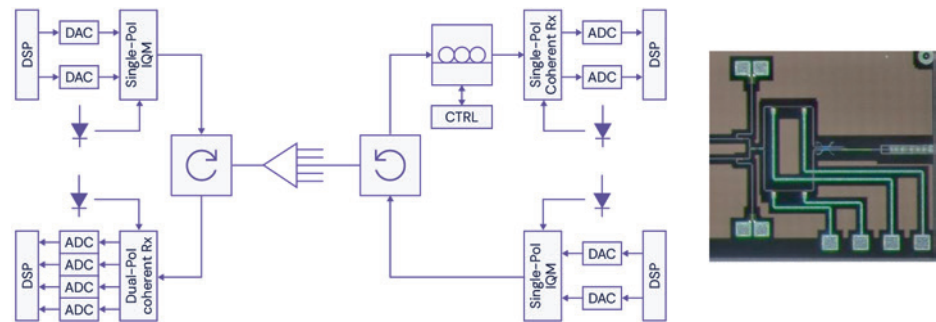
In the SP-QPSK scheme, information is encoded onto a single optical polarization using a conventional IQ modulator operating at 25 GBaud. Operating in the C-band further enhances deployment feasibility, as it enables the reuse of mature optical components and wavelength plans from NG-PON2, particularly wavelength mux/demux filters.

Two main hardware implementations have been proposed to implement SP-QPSK in practice. The first, represented in Figure 3, employs analog polarization control (APC) at the receiver, where the transmitter emits a single fixed polarization, while the receiver uses a slow feedback loop to track and align to the incoming polarization state [10]. This approach avoids the need for dual-polarization transmission, minimizing both optical and DSP complexity. It also eliminates the 3 dB power penalty associated with polarization duplication, resulting in a more power-efficient design. Furthermore, this configuration offers flexibility for network evolution: a fully coherent

dual-polarization transmitter could still be used at the OLT, with the second polarization left unmodulated or reserved for future “pay-as-you-grow” service upgrades.

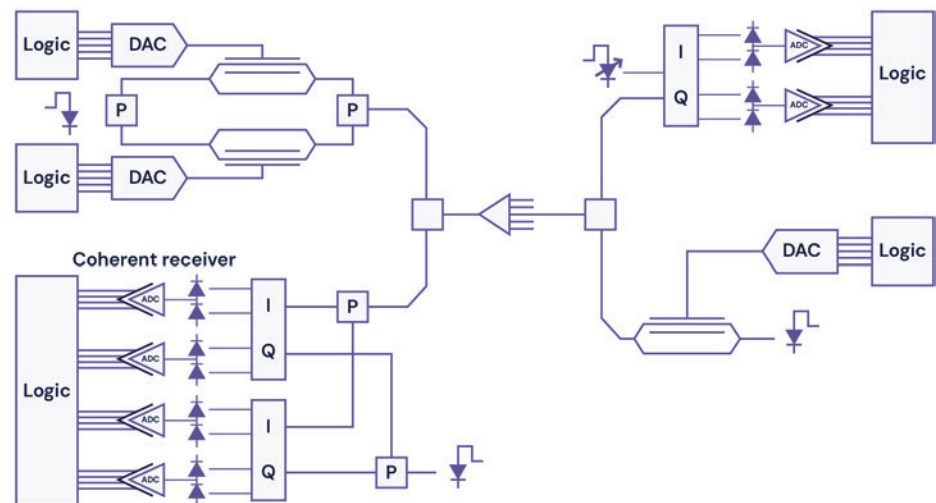
Although already demonstrated with discrete optics and benchtop equipment, the advent of Integrated Photonics and the promise of optics with increased functionality density now enable the APC scheme, which is being actively explored with a prototype developed on an open-access Silicon photonics platform, as depicted in **Figure 3**.

Figure 3: Left: Analog Polarization Control reference implementation for SP-QPSK; Right: Picture of fabricated APC structure



The second approach relies on polarization-time block coding, most notably the Alamouti scheme, where the same information is transmitted across both polarizations in consecutive symbol periods [11]. This configuration ensures polarization-independent reception, removing the need for active polarization tracking in the receiver. However, it requires a fully coherent dual-polarization transmitter, since both polarization branches must be modulated with the same data. Consequently, the transmitted optical power is effectively divided between the two polarizations, introducing an inherent 3 dB penalty that can be critical in power-budget-limited PONs. **Figure 4** shows a reference implementation of this approach.

Figure 4: Alamouti reference implementation of SP-QPSK



Both architectures pursue the same goal of minimizing ONU complexity while maintaining the advantages of coherent detection. The APC-based implementation prioritizes hardware simplicity and optical-power efficiency through photonic integration, whereas the Alamouti approach favors polarization independence and operational robustness through dedicated DSP.

Positioning between IM/DD and full coherent solutions

At 50 Gbps per wavelength, the options for IM/DD transmission in the C-band are extremely limited as the impact of Chromatic Dispersion cannot be tolerated without the use of dispersion compensation modules and/or pre-/post-equalization algorithm [5]. While advanced IM/DD formats such as PAM4 and vestigial-sideband single-carrier (vSSB) are under study, their reach and spectral-efficiency trade-offs remain incompatible with the cost and operational targets of PON systems. A coherent transition phase based on SP-QPSK sits between the IMDD limitations and the full Coherent systems with the balancing of the coherent advantages, particularly the extended power budget, with the power/cost reference imposed by IMDD.

Between the coherent implementations, and since all the flavors are similar in terms of performance and both require developments of new DSP chips, particularly the burst mode DSP algorithms, one of the decision vectors is the power consumption of each approach. Although exact numbers are not available due to IP rights and the always-evolving microelectronics ecosystem, a study was conducted to derive the relative costs of each implementation, with the results being depicted in **Table 1**. The projections are normalized to the Alamouti-based OLT.

Table 1: Power Breakdown of APC, Alamouti and Full Coherent flavours of SP-QPSK

Power Breakdown (W)	Alamouti	APC	Full-Coh
OLT	A (DP)	0.85 x A (SP)	0.95 x A (DP)
ONU	0.63 x A (SP)	0.69 x A (SP)	0.95 x A (DP)
New Development	ASIC	Pol. Ctrl. Optics	Tailor for 20km
Burst Mode			

Both simplified variants are expected to consume less power than full coherent receivers. The APC version achieves the lowest total power, as it relies on a single-polarization transmitter and minimal control electronics. The Alamouti-coded version offers polarization independence but requires a dual-polarization transmitter and hence higher driving power, as well as the implementation of a dedicated Alamouti DSP chip.

Conclusions

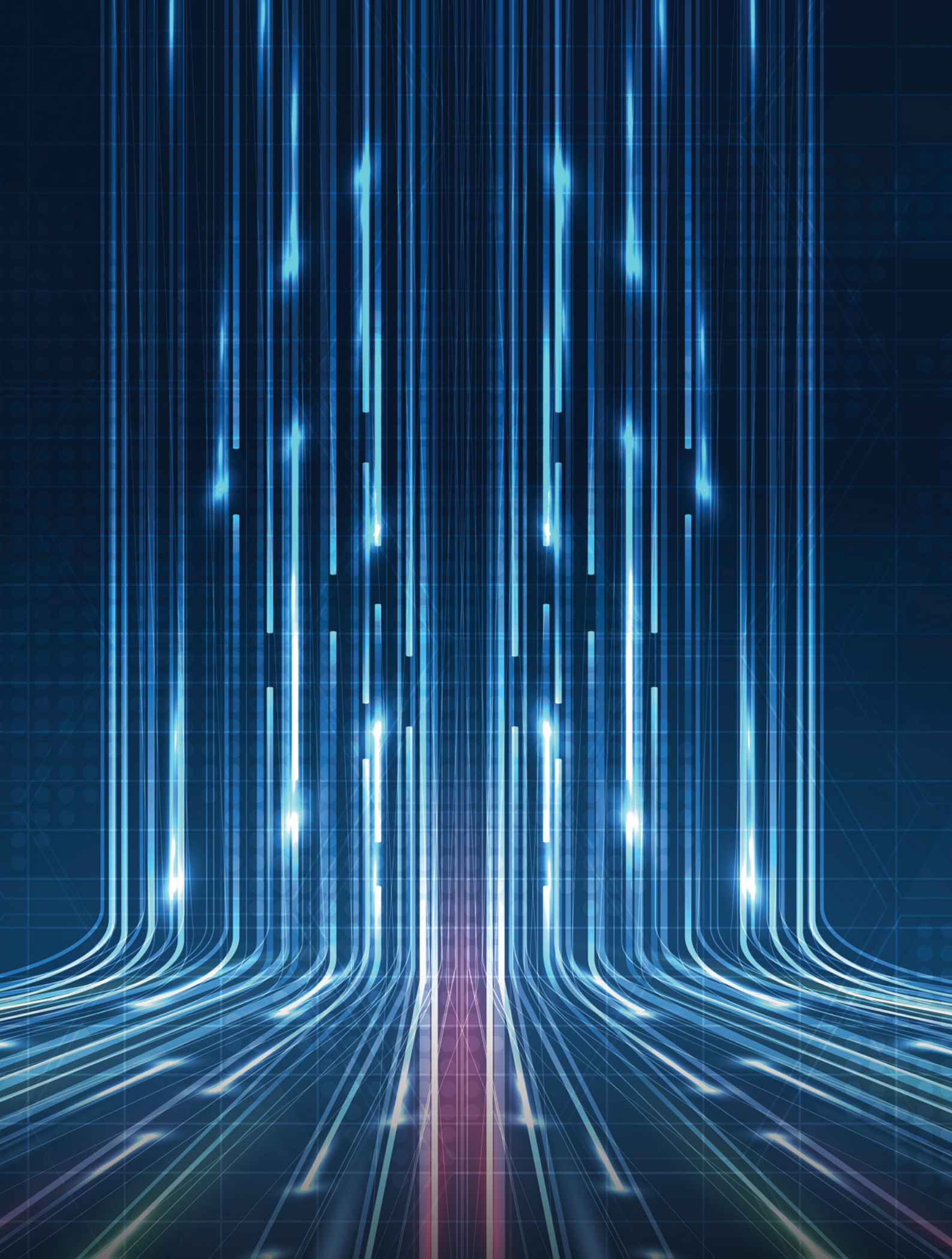
Overall, SP-QPSK provides a scalable and robust foundation for the transition toward coherent access networks, bridging today's IM/DD technology with the fully coherent architectures envisioned for future high-capacity PON. It overcomes the dispersion and sensitivity limitations of IM/DD while maintaining viability within the stringent power and cost constraints of access systems, leveraging the mature 25-GBaud coherent ecosystem and existing C-band infrastructure.

Meanwhile, IM/DD remains a practical and cost-efficient solution for short-reach O-band deployments where dispersion is negligible, and where the single wavelength 50G standard is implemented.

This work was funded by Projeto Albatroz (Projeto: 01775, P2023, Aviso MPr-2023-11 – I&D Empresarial – RCI). The authors acknowledge the support of all institutions and individuals who contributed to this article.

References

- [1] ITU-T, "ITU-T Recommendation G.9804.1 – Higher speed passive optical networks – Requirements," 2019.
- [2] ITU-T, "ITU-T Recommendation G.9804.3 – 50-Gigabit-capable passive optical networks (50G-PON): Physical media dependent (PMD) layer specification," 2021.
- [3] M. Li et al., "400 Gb/s real-time coherent PON based on a silicon photonic integrated transceiver," *Opt. Express*, vol. 30, pp. 47847–47855, 2022, doi: 10.1364/OE.476600.
- [4] S. Luo et al., "112-Gb/s/λ Downstream Transmission for TDM-PON with 31-dB Power Budget using 25-Gb/s Optics and Simple DSP in ONU," in *Proc. Opt. Fiber Commun. Conf. Exhib. (OFC)*, San Diego, CA, USA, 2020, pp. 1–3.
- [5] J. Chen et al., "50-km C-Band Transmission of 50-Gb/s PAM4 Using 10-G EML and Complexity-Reduced Adaptive Equalization," *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–10, Feb. 2019, Art no. 7200410, doi: 10.1109/JPHOT.2018.2886296.
- [6] J. Li et al., "Demonstration of 200 Gb/s/λ PON in O-band with high-bandwidth TFLN modulator," *J. Lightw. Technol.*, vol. 41, pp. 3594–3602, 2023.
- [7] A. Tavares et. al, "Multichannel 50 Gb/s per channel IM/DD O-Band 100GHz spacing high speed PON coexisting with legacy technologies," *InnovAction*, no. 9, pp. 10–17, 2024.
- [8] A. Teixeira, A. Shahpari, R. Ferreira, F. P. Guiomar and J. D. Reis, "Coherent Access," in *Proc. Opt. Fiber Commun. Conf. Exhib. (OFC)*, Anaheim, CA, USA, 2016, pp. 1–3.
- [9] M. S. Erkiling et al., "Comparison of low-complexity coherent receivers for UDWDM-PONs (λ-to-the-user)," 2017, arXiv:1711.02043.
- [10] J. K. J. K. Perin, A. Shastri, and J. M. Kahn, "Data Center Links Beyond 100 Gbit/s per Wavelength," *Opt. Fiber Technol.*, vol. 44, pp. 69–85, 2018.
- [11] M. S. Faruk, X. Li, and S. J. Savory, "Experimental demonstration of 100/200-Gb/s/λ PON downstream transmission using simplified coherent receivers," in *Optical Fiber Communication Conference*, (Optica Publishing Group, 2022), pp. Th3E–5.
- [12] IEEE 802.3 Ethernet Working Group. "IEEE P802.3ca 50G-EPON Task Force Physical – Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks." IEEE802.org. Accessed: Sep, 2025. [Online.] Available: <https://www.ieee802.org/3/ca/>



13

VHSP Technology Assessment: IMDD vs. Coherent PON

Abstract

As the demand for ultra-high-speed broadband continues to grow, the evolution of Passive Optical Network (PON) technologies is critical to meeting future capacity and performance requirements. This article presents a comprehensive assessment of Very High-Speed PON (VHSP) technologies, focusing on a comparative analysis between Intensity Modulation Direct Detection (IMDD) and Coherent PON approaches. We evaluate both technologies across key dimensions, including spectral efficiency, reach, cost, complexity, and power consumption. IMDD, known for its simplicity and cost-effectiveness, is contrasted with Coherent PON, which offers superior performance in terms of capacity and reach but at higher implementation complexity. This study provides insights into the trade-offs and deployment scenarios best suited for each technology.

Authors

Cláudio Rodrigues, Altice Labs, claudio-e-rodrigues@alticelabs.com

Fernando Guiomar, Instituto de Telecomunicações (IT), guiomar@av.it.pt

Francisco Rodrigues, PICadvanced, francisco@picadvanced.com

Jacklyn Reis, PICadvanced, jacklyn.reis@picadvanced.com

João Santos, PICadvanced, joao.santos@picadvanced.com

Ana Tavares, PICadvanced, ana@picadvanced.com

Bruno T. Brandão, Instituto de Telecomunicações and Universidade de Aveiro, bruno.brandao@ua.pt

Paulo P. Monteiro, Instituto de Telecomunicações, paulom@av.it.pt

António Teixeira, Instituto de Telecomunicações and PICadvanced, teixeira@av.it.pt

Keywords

PON, VHS-PON, IMDD, Coherent

DOI

10.57650/INNOVATION/2025.13

Introduction

The exponential growth in broadband demand, driven by bandwidth-intensive applications such as 8K video streaming, cloud gaming, and immersive Augmented Reality/Virtual Reality (AR/VR) experiences, is pushing access networks to their limits. Passive Optical Network (PON) technologies have evolved from Gigabit PON (GPON) to 10-Gigabit-capable symmetric PON (XGS-PON) and 50G-PON, now toward Very High-Speed PON (VHSP), targeting capacities beyond 50 Gbps per wavelength.

This paper presents a comparative assessment of two key approaches for VHSP: Intensity Modulation Direct Detection (IMDD) and Coherent PON.

VHSP overview

VHSP represents the next evolutionary step in optical access networks, targeting capacities beyond 50 Gb/s per wavelength over a point-to-multipoint fiber infrastructure [1]. Unlike previous generations, the defining metric for VHSP is its system service capacity, which refers to the aggregate ability to transport upper-layer service data units across multiple Optical Network Units (ONU), rather than simply the nominal line rate [2].

Current standardization efforts by ITU-T and IEEE envision overall service capacities of at least 100 Gb/s or 200 Gb/s, depending on cost-performance trade-offs [3], [4]. While 100 Gb/s may serve as a transitional baseline, it is widely acknowledged that this capacity will likely be insufficient for future endpoint service tiers, especially considering emerging applications such as immersive media and cloud services [5]. Consequently, 200 Gb/s is considered a more sustainable target for next-generation deployments [4].

The VHSP roadmap builds upon the ITU-T G.sup.VHSP Supplement, which outlines potential system requirements and candidate technologies for PON systems exceeding 50 Gbit/s per wavelength between the optical line termination (OLT) and the ONU.

To achieve these goals, VHSP leverages advanced techniques such as digital signal processing (DSP) and coherent optics to overcome the inherent limitations of traditional IMDD systems [6], [7]. These innovations aim to balance scalability, cost, and performance, while ensuring compatibility with existing fiber infrastructure.

Service use case requirements for VHSP systems

The design of VHSP systems must accommodate a diverse set of service use cases, each characterized by distinct bandwidth, latency, and operational requirements. These use cases span residential broadband, enterprise connectivity, and mobile network transport functions such as backhaul, midhaul, and fronthaul.

Residential services typically demand downstream rate tiers ranging from 5 Gb/s to 50 Gb/s, with an optional target of 100 Gb/s for future scalability. Symmetry coefficients vary between 1.0 (fully symmetric) and 0.25 (highly asymmetric),

For fronthaul networks, every microsecond counts: delays must stay below 100 μs

reflecting the predominance of downstream traffic. Enterprise services exhibit higher capacity requirements, with rate tiers extending up to 200 Gb/s, and generally favor symmetric or moderately asymmetric configurations to support bi-directional traffic flows [8].

Mobile and fixed wireless access (FWA) backhaul and midhaul scenarios introduce stringent latency and bandwidth guarantees. Backhaul links require up to 50 Gb/s per endpoint, while midhaul may scale to 100 Gb/s, driven by centralized radio access network (C-RAN) architectures. Fronthaul applications (both macrocell and small cell) are the most demanding in terms of latency, with allowable one-way delays approaching 100 μs, necessitating near-zero startup and burst tolerance. These scenarios also exhibit strict requirements for bandwidth assurance and minimal service interruption during wavelength channel changes [8].

Dynamic bandwidth allocation (DBA) is supported across most use cases, except macrocell fronthaul, where deterministic performance is paramount. Oversubscription is permissible in residential and midhaul contexts but generally prohibited in fronthaul deployments to maintain service integrity. Startup and burst tolerances decrease progressively from residential (4 s and 2 s, respectively) to fronthaul (near-zero), reflecting the need for rapid bandwidth provisioning in mobile transport [8].

VHSP systems must deliver a flexible architecture capable of supporting heterogeneous service profiles. This includes accommodating high-capacity tiers for enterprise and midhaul, ultra-low latency for fronthaul, and dynamic resource allocation for residential and backhaul services. These requirements underscore the importance of advanced scheduling, wavelength management, and physical layer innovations in next-generation PON systems.

Coexistence scenario requirements for VHSP systems

The deployment of VHSP systems must ensure compatibility with existing and emerging optical access technologies to facilitate smooth migration and protect operator investments. **Table 1** consolidates the coexistence scenarios that VHSP systems are expected to support, either through the implementation of a multi-port module (MPM) or via a dedicated coexistence element [8].

Table 1: VHSP coexistence scenario requirements

Coexistence scenario reference	GPON	XGS-PON /10G-EPON	NG-PON2	50G-PON	VHSP
CES 1			Yes		Yes
CES 2		Yes		Yes	Yes
CES 3	Yes		Yes		Yes
CES 4	Yes	Yes		Yes	Yes
CES 5	Yes			Yes	Yes

Coexistence scenarios (CES) span interactions with legacy systems such as GPON, XGS-PON/10G-EPON, and Next-Generation Passive Optical Networks (NG-PON2), as well as next-generation solutions including 50G-PON and VHSP itself. The table indicates that VHSP must coexist with multiple technologies across five defined scenarios (CES1 – CES5). For example, CES 1 and CES 3 involve coexistence with NG-PON2 and VHSP, while CES 2, CES 4, and CES 5 are considered spectrally compatible and can be supported by a single coexistence element. This compatibility reduces complexity and cost for operators deploying multi-service networks [8].

From a spectral perspective, coexistence requires careful wavelength allocation and management to mitigate interference, including second-order effects such as out-of-band crosstalk. Wavelength allocation bands for each technology highlight the importance of precise planning in shared fiber environments. Additionally, initial design choices, such as the selection of narrow or reduced upstream options for GPON, can significantly influence subsequent coexistence capabilities.

VHSP systems must incorporate flexible coexistence strategies to enable simultaneous operation with legacy and next-generation PON technologies. This approach ensures service continuity, minimizes operational disruption, and supports incremental upgrades aligned with market and technology evolution.

Candidate technologies for VHSP

Multi-access PON with direct detection technologies

Direct Detection (DD) has historically been the predominant method for optical signal detection in fiber communication systems, particularly before 2010. DD operates by converting optical intensity variations directly into electrical signals at the receiver, without requiring phase or frequency recovery. This simplicity has made DD-based systems highly cost-effective, balancing performance and implementation ease. Consequently, DD remains widely deployed for intra-data center interconnects and continues to underpin all generations of commercial PON, including 50G-PON.

Within the DD paradigm, IMDD is the most prevalent modulation approach. IMDD encodes information by varying the optical intensity at the transmitter and employs direct detection at the receiver. Its advantages, such as straightforward architecture, low cost, and minimal processing complexity, have driven its adoption across multiple PON generations.

Overview of NRZ-OOK and PAM4

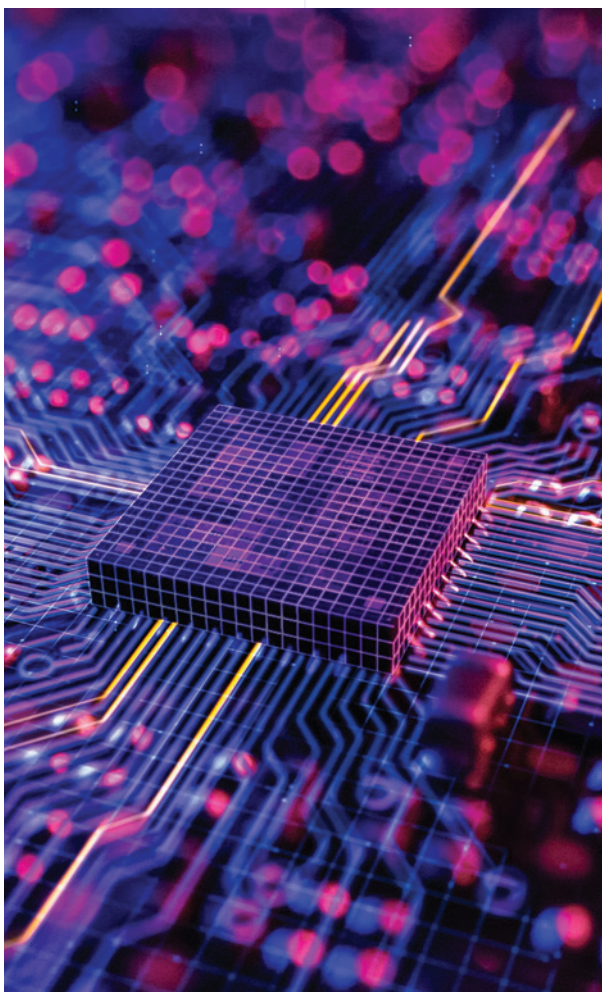
Historically, Non-Return-to-Zero On-Off Keying (NRZ-OOK) has been the primary modulation format in IMDD-based systems. NRZ-OOK offers simplicity and robust receiver sensitivity, making it well-suited for PON deployments. However, as capacity demands have grown, the data center market has increasingly adopted

NRZ-OOK continues to lead as the modulation choice for VHSP systems above 50 Gb/s per wavelength

Pulse Amplitude Modulation with four levels (PAM4) to achieve higher bit rates without requiring proportionally higher component bandwidth. PAM4 reduces signal bandwidth requirements and improves dispersion tolerance compared to NRZ-OOK at equivalent bit rates, enabling its widespread use in Ethernet standards up to 400G and beyond [8].

Despite these advantages, PAM4 faces significant challenges in PON environments. Its lower signal-to-noise ratio (SNR) tolerance, sensitivity to multipath interference (MPI), and stringent linearity requirements result in substantial receiver sensitivity penalties. Theoretical modulation penalties for PAM4 relative to NRZ-OOK are approximately 4.8 dB, with practical implementation penalties adding several additional decibels. For example, a 240 Gb/s (120 GBd) PAM4 IMDD Time Division Multiplexing (TDM) PON receiver may exhibit a sensitivity deficit of ~7 dB compared to an NRZ-OOK receiver at the same symbol rate. Even at 60 GBd, PAM4 incurs a penalty of ~4 dB, comprising modulation complexity and non-DSP compensated impairments. These penalties pose significant challenges for achieving the high-power budgets required in PON systems, often exceeding 32 dB [8].

Given these constraints, NRZ-OOK remains the preferred modulation format for VHSP systems targeting capacities beyond 50 Gb/s per wavelength. While PAM4 and other higher-order schemes may offer future opportunities, their adoption will depend on advances in DSP, component linearity, and MPI mitigation [9].



Overview of Optical Duobinary (ODB)

ODB represents another IMDD-compatible modulation format designed to enhance dispersion tolerance. ODB introduces a controlled phase shift of 180° between consecutive “one” bits surrounding a “zero” bit, mitigating inter-symbol interference caused by chromatic dispersion [9]. This approach reduces signal bandwidth and can be implemented using a low-pass or delay-and-add filter at the transmitter, followed by a Mach-Zehnder Modulator (MZM). ODB can be received using direct detection, offering a potential pathway for extending IMDD performance in VHSP systems.

Overview of Digital Chromatic Dispersion Pre-Compensation (DCPC)

DCPC is a transmitter-side technique designed to mitigate chromatic dispersion (CD) in optical fiber links. DCPC employs DSP at the transmitter to generate a complex optical launch signal whose properties are the inverse of those induced by accumulated CD in the fiber [7]. After linear propagation, the signal arrives at the receiver largely undistorted, requiring only approximate knowledge of the fiber length.

However, in the nonlinear regime, self-phase modulation (SPM) introduces penalties that degrade DCPC performance, particularly at higher CD values. Consequently, operation near the fiber's zero-dispersion wavelength is preferred to minimize these effects [8].

The DCPC scheme uses DSP-based finite impulse response (FIR) filters to drive an IQ modulator, although other modulators, such as MZM, may be considered. At the receiver, detection occurs via a conventional photodetector, similar to IMDD PON systems. To achieve VHSP capacities, DCPC PON could employ two wavelength channels operating at approximately 120 GBd. A key characteristic of DCPC is its optimization for a specific accumulated CD value, determined by FIR filter coefficients. Deviations from this optimized value introduce path penalties, necessitating multiple coefficient sets to cover the full CD range.

Overview of Single Sideband (SSB)

SSB modulation boosts VHSP performance by halving optical spectral width compared to conventional formats

SSB modulation is an attractive candidate for VHSP systems due to its ability to improve transmission performance by reducing the optical spectral width by approximately 50% compared to conventional intensity modulation formats [9]. This spectral efficiency enhancement mitigates dispersion-induced fading, which typically constrains IMDD chromatic dispersion tolerance at the high baud rates required for VHSP.

A practical method for generating SSB signals involves the use of an IQ modulator driven by an appropriately conditioned signal. In a typical implementation, the in-phase (I) input receives an NRZ drive signal, while the quadrature (Q) input is driven by the same NRZ signal after undergoing a Hilbert transform. Alternative modulators, such as MZMs, may also be considered. At the receiver, detection is performed using a standard photodetector, similar to IMDD PON systems [8].

Multi-access PON with coherent technologies

Coherent detection represents a transformative approach for achieving the key objectives of VHSP systems, including ultra-high capacity, extended link budgets, and enhanced operational flexibility. Techniques such as time-division multiplexing (TDM), time-division multiple access (TDMA), frequency division multiplexing (FDM), frequency division multiple access (FDMA), and digital sub-carrier multiplexing (DSCM) can be combined with coherent detection to enable scalable multi-access architectures. Both single-carrier and subcarrier-based coherent PON implementations are under consideration for next-generation deployments [10], [11], [12].

Overview of Coherent PON

Unlike direct detection, coherent detection reconstructs the full optical field of the transmitted signal, including amplitude, phase, and polarization. This capability allows the use of advanced modulation formats, where information is encoded across the I and Q components of the carrier and both polarization states.

Figure 1: General schematic of the coherent transmission system architecture

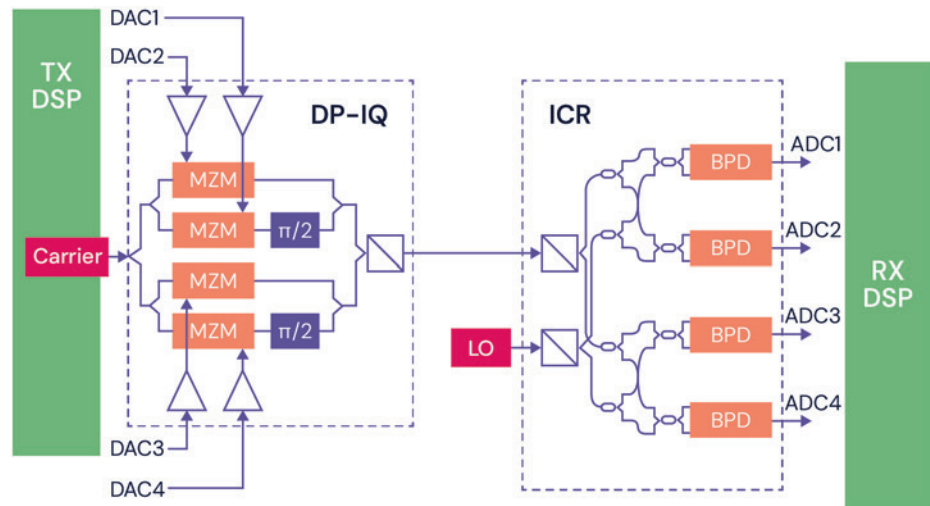


Figure 1 illustrates the architecture of a typical coherent transmission system [8], [13]. Key transceiver elements include:

- **Dual-polarization IQ modulator (DP-IQM):** modulates an optical carrier across four dimensions of light (I/Q for X and Y polarizations).
- **Local Oscillator (LO):** provides the carrier wave for coherent mixing; wavelength alignment with the received signal is critical.
- **Integrated Coherent Receiver (ICR):** combines the received signal and LO in a 90° optical hybrid, converting outputs into electrical signals via balanced photodetectors (BPD).
- **DSP blocks with Analog-to-Digital Converter/Digital-to-Analog Converter (ADC/DAC):** generate advanced modulation formats at the transmitter and perform dispersion compensation, timing recovery, and other functions at the receiver.
- **Amplifiers and TIAs:** provide initial amplification of electrical signals.

Coherent detection offers significant advantages over IMDD in terms of spectral efficiency, dispersion tolerance, and receiver sensitivity. Experimental results suggest sensitivity gains of up to 10 dB compared to IMDD under full-coherent reception assumptions [14], [15], [16], [17]. Furthermore, coherent systems can compensate for impairments such as chromatic dispersion and polarization-mode dispersion (PoMD), enabling operation in the C-band for extended reach and higher link budgets compared to IMDD systems typically confined to the O-band.

Market trends and standardization

Coherent technology, traditionally deployed in core and metro networks since 2010 for 100 Gb/s and beyond, is now penetrating short-reach applications, such as Data Center Interconnect (DCI). Industry initiatives by the Optical Internetworking Forum (OIF) and the Institute of Electrical and Electronics Engineers (IEEE) have accelerated this trend, introducing interoperable pluggable modules for coherent interfaces: OIF 400ZR (2020), OIF 800ZR and 800LR (2024–2025), and IEEE 802.3dj B400G projects [8], [10], [11], [12].

Table 2: Status of coherent standards as of August 2025

Table 2 summarizes the status of coherent standards as of August 2025, highlighting baud rates, modulation formats, and Forward Error Correction (FEC) schemes. These developments indicate a clear trajectory toward cost reduction and broader adoption of coherent transceivers, with OIF already initiating work on 1.6T coherent optics for DCI and metro applications, targeting completion by 2027.

	IEEE	OIF	IEEE	OIF	OIF	IEEE
Item	800G LR1	800LR	800G ER1	800ZR	400ZR/ER	100GBASE-ZR
Maximum Reach (km)	10	10	40	80	80/40	80
Operating frequency	O band	O band	C band	C band	C band	C band
Baud rate (GBd)	123.6	123.6	118.2	118.2	~60	27.9525
Format	16QAM	16QAM	16QAM	16QAM	16QAM	DP-DQPSK
FEC	KR4 Physical-coding-sublayer FEC (KP4)	KP4	Open FEC (OFEC)	OFEC	CFEC	Staircase FEC (SC-FEC)

DSP-lite solutions for cost-effective and power-efficient Coherent PON

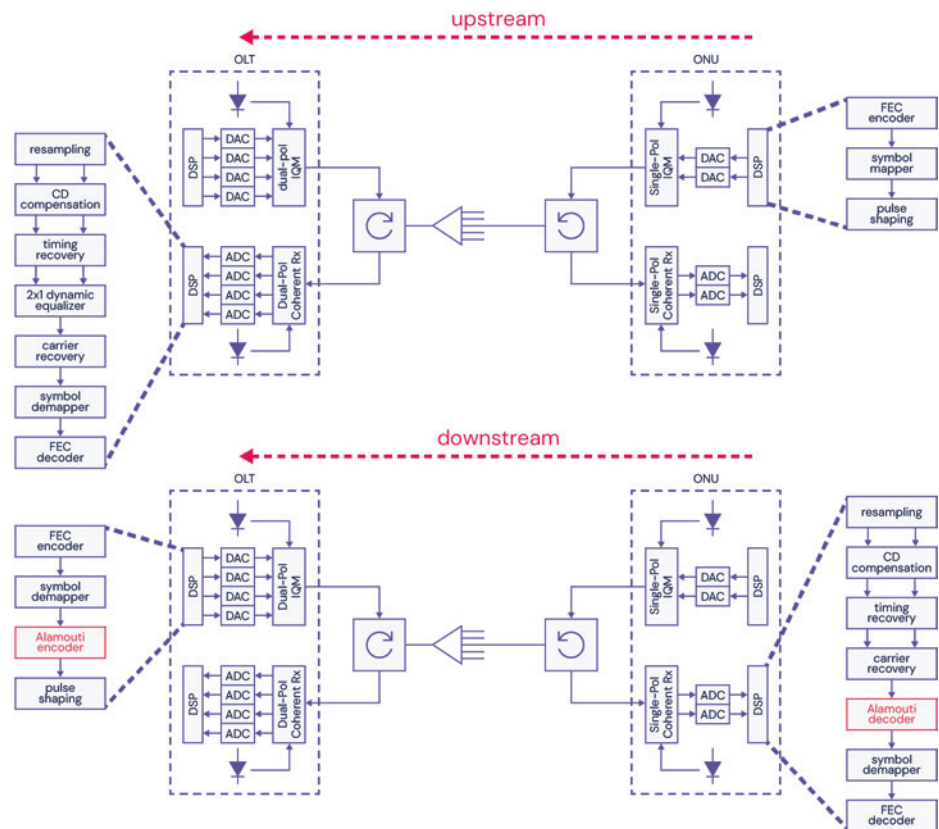
The main challenge associated with the adoption of coherent communications for next-generation PON lies on the high-complexity of traditional DSP algorithms/subsystems that are required to enable the successful modulation and decoding of upstream and downstream information [18].

Aiming to circumvent that challenge, several recent proposals have been put forward, targeting the development and validation of a DSP-lite solution that might enable a cost-effective and power-efficient coherent PON [19], [20]. One potential technical approach that has recently gained traction is the adoption of a single-polarization coherent transmission paradigm, which enables relaxation of the hardware requirements at the ONU, while also simplifying the DSP stack [20]. Such an approach is depicted in more detail in **Figure 2**, which highlights both the upstream and downstream use cases.

From the inspection of **Figure 2**, the main advantages of this single-polarization coherent PON architecture become apparent, including the following factors:

- At the ONU side, a simplified hardware architecture can be employed, using a single-polarization receiver and single-polarization IQ modulator, which roughly reduces the required number of hardware components by half. This is expected to also translate into similar advantages in terms of power consumption, footprint, and cost of the ONU terminal.

Figure 2: Architecture of a simplified coherent PON system supporting single-polarization transmission



- At the OLT, a traditional dual-polarization coherent transceiver architecture is still employed, leveraging the maturity of this technology and its wide availability in the market. Given that the cost and power consumption at the OLT is significantly less critical than that of the ONU, the use of full dual-polarization transmission/reception can be accommodated without major implications on the power/cost efficiency of the whole PON system.
- By employing single-polarization transmission/detection, the DSP complexity at the ONU side can also be significantly reduced.
 - To generate the transmitted signal, typical FEC encoding is first applied according to the selected coding scheme, followed by symbol mapping – using Quadrature Amplitude Modulation (QAM) formats, typically limited to Quadrature Phase Shift Keying (QPSK) signaling – and pulse shaping (optional or simplified pulse shaping to reduce complexity) is required;
 - To detect and decode the received signal, the DSP stack includes a digital resampling stage (optional, depending on the ADC configuration), chromatic dispersion compensation (also optional, depending on the link length and operating baud-rate), timing recovery (mandatory, but similar to what is typically needed for IMDD systems), carrier recovery (to compensate for the frequency and phase deviation between transmitter and local oscillator lasers), Alamouti decoding (required to retrieve the single-polarization information from the transmitted dual-polarization signal), symbol demapping (required to retrieve the transmitted symbols of the utilized QAM format) and FEC decoding (matched with the utilized encoding option).

- At the OLT side, standard dual-polarization coherent DSP can be employed, with two key additional subsystems:
 - An Alamouti encoder at the transmitter side to enable the single-polarization decoding at the ONU;
 - A multiple-input single-output (MISO) equalizer at the receiver, to retrieve the transmitted single-polarization signal generated by the ONU from the dual-polarization receiver at the OLT.

Although this single-polarization coherent PON architecture provides a promising starting point for the development of a commercially competitive solution, there are still important challenges that must be tackled within the next couple of years, including the following key factors:

- Devise novel architectural solutions that can reuse the transmitter laser at the ONU as a local oscillator, thereby halving the number of active laser sources, consequently reducing power consumption, footprint, and cost [21]. However, this requires an efficient management of upstream and downstream traffic with very tightly packed channels. It might also be a potential opportunity to explore digital multicarrier solutions [22].
- Introduce optical polarization control into the ONU terminal, thereby avoiding the need for Alamouti encoding/decoding, which can greatly simplify the DSP stack. On the other hand, this required the introduction of efficient PIC-level solutions to track and control the signal polarization, simultaneously representing an important opportunity for novel disruptive developments, as well as a significant technical challenge.



Multi-access PON based on IMDD-Coherent Hybrid PON

Overview of IMDD-Coherent Hybrid PON

IMDD-Coherent Hybrid PON combines IMDD and coherent technologies in different transmission directions to achieve VHSP capacity beyond 50 Gb/s. One configuration uses IMDD for upstream and coherent detection for downstream, enabling high downstream bandwidth and improved chromatic dispersion tolerance through DSP-based compensation. Conversely, another configuration applies IMDD for downstream and coherent detection for upstream, requiring careful design to maintain the link budget and transmission distance. These hybrid approaches leverage asymmetrical rates, zero-dispersion wavelengths, and advanced modulation schemes to balance cost, complexity, and performance [8].

Transmitter options include Directly Modulated Laser (DML), Electro Absorption Modulated Laser (EML) with Semiconductor Optical Amplifier (SOA), dual-drive EML, Single Polarization Mach-Zehnder (SP-MZ) modulator with SOA, SP-IQ MZ modulator with SOA, and DP-IQ MZ modulator with SOA. Receiver options range from direct detection with DSP to simplified coherent receivers (polarization-insensitive heterodyne) and full coherent receivers. Each configuration offers trade-offs in complexity, cost, and performance, with downstream coherent schemes providing superior chromatic dispersion tolerance and upstream IMDD requiring optimized design for sensitivity and reach.



Comparative Analysis

To evaluate the suitability of different VHSP architectures, a comparative analysis was conducted across five critical dimensions: spectral efficiency, reach, cost, complexity, and power consumption. These parameters directly influence the scalability, economic viability, and operational performance of next-generation optical access networks.

- **Spectral Efficiency** reflects the ability to transmit more data per unit of optical spectrum, which is essential for maximizing fiber capacity and supporting future service tiers (Bits per second per Hz).
- **Reach** determines the maximum transmission distance without regeneration, impacting deployment in rural and metropolitan environments.
- **Cost** encompasses both capital expenditure (CAPEX) and operational expenditure (OPEX), including transceiver complexity and ecosystem maturity.

- **Complexity** relates to hardware sophistication and DSP requirements, influencing integration and maintenance challenges.
- **Power Consumption** is increasingly critical for sustainable network design, as higher processing demands can significantly increase energy usage.

Table 3: Comparative analysis: IMDD vs. Coherent vs. Hybrid PON

Table 3 summarizes the relative performance of IMDD PON, Coherent PON, and IMDD-Coherent Hybrid PON across these dimensions, based on the technical characteristics and deployment considerations discussed in previous sections.

	IMDD PON	Coherent PON	IMDD-Coherent Hybrid PON
Spectral Efficiency	Low – Limited to OOK/PAM4 (1–2 bits/symbol). Requires multiple wavelengths for scaling.	High – Advanced formats (QPSK, 16QAM) enable 4–8 bits/symbol. Supports dense wavelength packing.	Moderate to High – Coherent in one direction boosts efficiency; IMDD limits overall gain.
Reach	Short to Medium – Typically 20–40 km; dispersion limits at high baud rates.	Long – >60 km feasible with DSP-based CD and PoIMD compensation.	Intermediate – Coherent downstream improves reach; IMDD upstream may require zero-dispersion wavelengths.
Cost	Low – Mature ecosystem, simple optics, minimal DSP. Ideal for cost-sensitive FTTH.	High – Requires LO lasers, high-speed ADC/DAC, DSP ASICs. Cost trending down with pluggables.	Medium – Adds complexity on one side; ONU cost can remain low if IMDD retained downstream.
Complexity	Low – Simple Tx/Rx, minimal DSP.	High – Full DSP stack, polarization diversity, phase recovery.	Medium – Complexity concentrated at OLT; ONUs can remain simple.
Power Consumption	Low – No heavy DSP; efficient for large-scale deployments.	High – DSP and coherent optics increase power draw significantly.	Medium – Coherent DSP only in one direction; better than full coherent but higher than IMDD

Hybrid PON delivers a practical balance: coherent detection for critical performance, IMDD for cost-sensitive paths

As key Insights, IMDD remains attractive for cost-sensitive, short-reach deployments but struggles with scalability and dispersion at >50 Gb/s. Coherent PON delivers unmatched performance in spectral efficiency and reach, enabling operation in C-band and advanced modulation formats, but at higher cost and power. Hybrid PON offers a pragmatic compromise: leveraging coherent detection where performance is critical (e.g., downstream) while retaining IMDD for cost-sensitive directions (e.g., upstream).

In terms of Spectral Efficiency, IMDD systems typically employ simple modulation formats such as OOK or PAM4, which offer limited spectral efficiency (up to 2 bits/symbol for PAM4). This constraint becomes critical as operators seek to maximize fiber capacity. Coherent PON, on the other hand, leverages advanced modulation schemes like QPSK and 16QAM, achieving significantly higher

spectral efficiency (up to 4–8 bits/symbol). This enables better utilization of available spectrum and supports higher aggregate throughput.

Regarding the maximum transmission distance, IMDD is inherently dispersion-limited. Without complex compensation techniques, its reach is typically constrained to 20–40 km for high-speed links. Coherent systems incorporate DSP-based impairment mitigation, allowing for extended reach beyond 60 km, even under challenging conditions. This makes coherent PON attractive for rural deployments and scenarios requiring long feeder segments.

IMDD benefits from mature, low-cost optics and minimal DSP requirements, making it ideal for cost-sensitive residential deployments.

Conversely, coherent PON introduces complexity through local oscillator lasers, high-speed ADC/DAC, and DSP blocks. These components increase the bill of materials and operational complexity, though integration trends (e.g., silicon photonics) are expected to reduce costs over time. IMDD systems consume less power due to their simplicity and absence of heavy DSP.

Coherent systems, while offering superior performance, demand significant processing power for DSP and cooling, impacting operational expenses and sustainability goals.

Discussion

The choice between IMDD and coherent PON is not binary, but rather depends on operator priorities. IMDD remains attractive for its low cost and simplicity, while coherent PON is the clear winner for performance-critical applications.

Looking ahead, several developments and trends are expected to influence this ongoing discussion:

- DSP integration and Application-Specific Integrated Circuit (ASIC) development will reduce coherent complexity and power consumption.
- Silicon photonics and pluggable optics will drive down cost for coherent network solutions.
- Standardization efforts will shape ecosystem readiness.

At present, a number of challenges remain to be addressed:

- Balancing CAPEX and OPEX in large-scale deployments.
- Ensuring interoperability and backward compatibility with existing PON worldwide infrastructure.

Conclusion

This paper presented a comprehensive assessment of VHSP technologies, comparing IMDD and coherent PON across key dimensions. IMDD offers simplicity and cost-effectiveness, making it suitable for mass-market FTTH deployments. Coherent PON, while more complex and costly, delivers unmatched

performance in terms of reach and spectral efficiency, positioning it as the technology of choice for premium and long-reach applications.

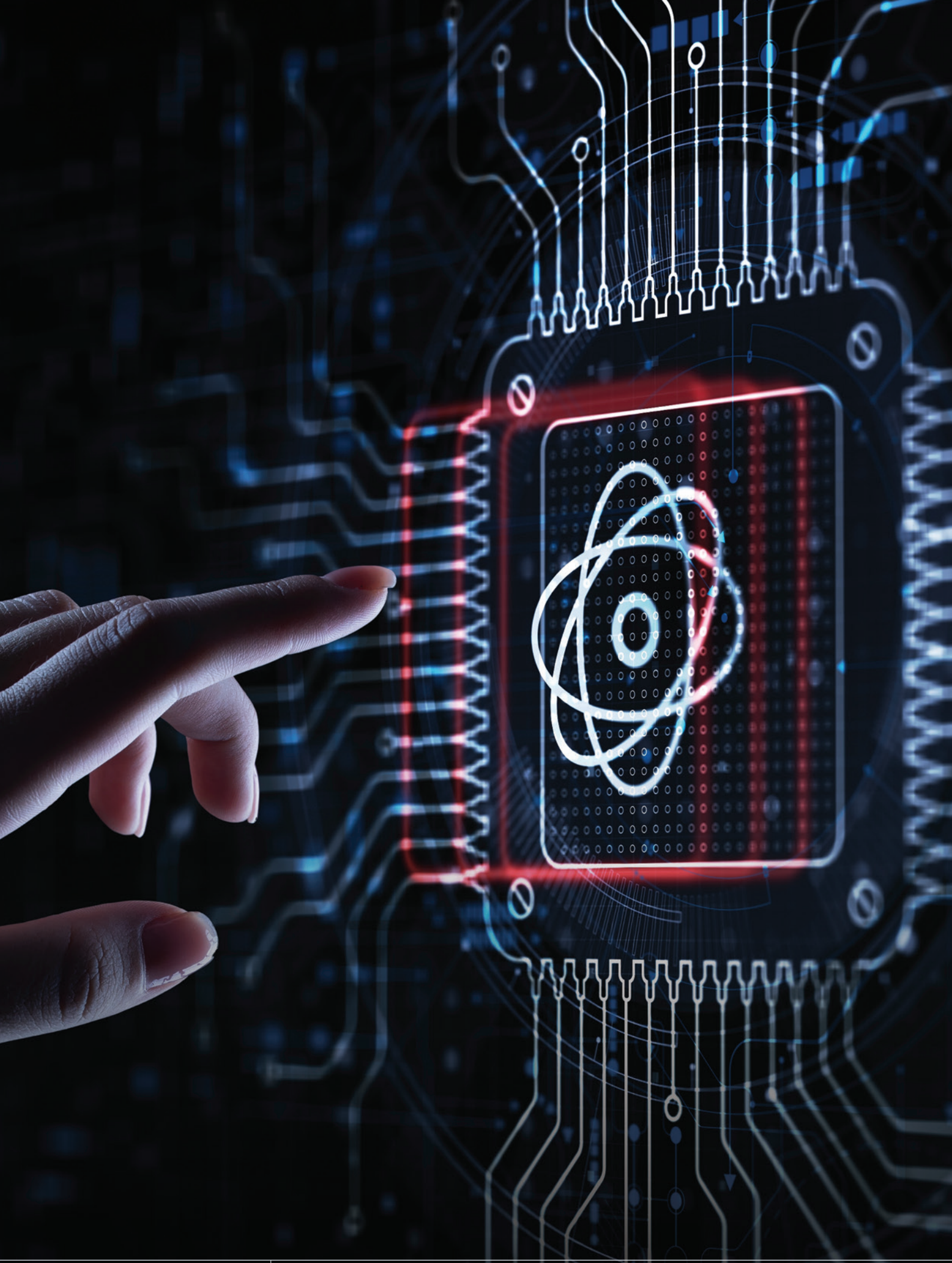
Operators should adopt a scenario-based approach, leveraging IMDD for cost-sensitive segments and coherent PON for high-capacity, performance-critical use cases. Future innovations in DSP and photonic integration will further influence this balance, paving the way for scalable, energy-efficient VHSP networks.

This work was funded by Projeto Albatroz (Projeto: 01775, P2023, Aviso MPr-2023-11 – I&D Empresarial – RCI). The authors acknowledge the support of all institutions and individuals who contributed to this article.

References

- [1] ITU-T, "Recommendation ITU-T G.9804.1- Higher speed passive optical networks – Requirement," 2024.
- [2] D. Nettet, "Progress on Very High Speed PON in ITU-T," in *Proc. Opt. Fiber Commun. Conf. (OFC) 2025*, Tech. Dig. Ser., 2025, Art. no. Th1J.5, doi: 10.1364/OFC.2025.Th1J.5.
- [3] IEEE, "IEEE 802.3ca 25G/50G EPON RX PMD Values," 2019.
- [4] Broadband Forum, "State of PON 2025: Where We Are and the Roadmap for Next-Generation Fiber Access," 2025.
- [5] Y. Xu, X. Guan, W. Jiang, X. Wang, W. Hu, and L. Yi, "Low-complexity end-to-end deep learning framework for 100G-PON," *J. Opt. Commun. Netw.*, vol. 16, no. 11, pp. 1093–1103, Nov. 2024, doi: 10.1364/JOCN.532742.
- [6] I. B. Kovacs, M. S. Faruk, P. Torres-Ferrera, and S. J. Savory, "Simplified coherent optical network units for very-high-speed passive optical networks," *J. Opt. Commun. Netw.*, vol. 16, no. 7, pp. C1–C10, Jul. 2024, doi: 10.1364/JOCN.514867.
- [7] P. Torres-Ferrera, G. Rizzelli, V. Ferrero, e R. Gaudino, "100+ Gbps/λ 50 km C-Band Downstream PON Using CD Digital Pre-Compensation and Direct-Detection ONU Receiver," *J. Lightw. Technol.*, vol. 38, no. 24, pp. 6807–6816, 2020.
- [8] ITU-T, "Work item G Suppl. VHSP: PON transmission technologies above 50 Gb/s per wavelength," ITU-T Study Group 15, Q2/15, Work Programme item ISN 21543, Oct. 2025. [Online]. Available: https://www.itu.int/ITU-T/workprog/wp_item.aspx?isn=21543
- [9] T. Bo, H. Kim, Z. Tan, and Y. Dong, "Optical Single-Sideband Transmitters," *J. Lightw. Technol.*, vol. 41, no. 4, pp. 1163–1174, Feb. 15, 2023, doi: 10.1109/JLT.2022.3212473.
- [10] OIF, "Implementation Agreement 400ZR," 2020.
- [11] OIF, "800ZR and 800LR Implementation Agreements," 2024–2025.

- [12] IEEE P802.3dj Task Force, "Standard for Ethernet Amendment: Media Access Control Parameters for 1.6 Tb/s and Physical Layers and Management Parameters for 200 Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s Operation," IEEE Std. 802.3 (Amendment P802.3dj), work in progress, Jul. 2025. [Online]. Available: <https://ieee802.org/3/dj/index.html>
- [13] E. Ip, A. P. T. Lau, D. J. F. Barros, e J. M. Kahn, "Coherent Detection in Optical Fiber Systems," *Opt. Express*, vol. 16, no. 2, pp. 753–791, 2008, doi: 10.1364/OE.16.000753.
- [14] J. Li, X. Zhang, M. Luo, C. Yang, Z. He, and X. Xiao, "SOA Pre-Amplified 200 Gb/s/ λ PON Using High-Bandwidth TFLN Modulator," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, J. Leuthold, C. Harder, B. Offrein, and H. Limberger, Eds., Technical Digest Series (Optica Publishing Group), 2022, paper Tu4C.5.
- [15] D. Che, P. Iannone, G. Raybon, and Y. Matsui, "200 Gb/s Bi-Directional TDM-PON With 29-dB Power Budget," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Bordeaux, France, 2021, pp. 1–3, doi: 10.1109/ECOC52684.2021.9606049.
- [16] G. Li, S. Xing, Z. Li, J. Zhang, e N. Chi, "200-Gb/s/ λ Coherent TDM-PON With Wide Dynamic Range of >30 -dB Based on Local Oscillator Power Adjustment," in *Proc. Opt. Fiber Commun. Exhib. (OFC)*, San Diego, CA, USA, 2022, pp. 1–3. G. Li et al., "Advances in Coherent Optics for Short-Reach Applications," Optica, 2025.
- [17] A. Hraghi, G. Rizzelli, A. Pagano, V. Ferrero, and R. Gaudino, "Analysis and Experiments on C Band 200G Coherent PON Based on Alamouti Polarization-Insensitive Receivers," *Opt. Express*, vol. 30, no. 26, pp. 46782–46797, 2022, doi: 10.1364/OE.472470.
- [18] A. Shahpari et al., "Coherent Access: A Review," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1050–1058, Feb. 2017, doi: 10.1109/JLT.2016.2623793.
- [19] I. B. Kovacs, M. S. Faruk, P. Torres-Ferrera, and S. J. Savory, "Simplified coherent optical network units for very-high-speed passive optical networks," *J. Opt. Commun. Netw.*, vol. 16, no. 7, pp. C1–C10, Jul. 2024, doi: 10.1364/JOCN.514867.
- [20] M. S. Erkilinc et al., "PON transceiver technologies for ≥ 50 Gbits/s per λ : Alamouti coding and heterodyne detection [Invited]," *J. Opt. Commun. Netw.*, vol. 12, no. 2, pp. A162–A170, Feb. 2020, doi: 10.1364/JOCN.12.00A162.
- [21] J. Segarra, V. Sales, V. Polo, J. Tabares, and J. Prat, "Flexible coherent UDWDM-PON with dynamic user allocation based on limited-tunability lasers," *J. Opt. Commun. Netw.*, vol. 12, no. 9, pp. D27–D35, Sep. 2020, doi: 10.1364/JOCN.391023.
- [22] [22] Z. Hu et al., "Experimental assessments of a coherent DSCM and SOA-based WDM metro-access network architecture with enhanced flexibility and capacity," *Opt. Lett.*, vol. 50, pp. 1325–1328, 2025, doi: 10.1364/OL.553535.



14

Securing Today Against Tomorrow's Quantum Threats

Abstract

As quantum computers evolve, they will be used to break today's public-key cryptography. It is a matter of when, not if, it will happen. There is also enough evidence of "Harvest Now, Decrypt Later" (HNDL) attacks, mostly by nation-states, to make this a problem for today, not just for an undetermined date in the future when quantum computers are powerful enough.

Quantum cryptography, such as Quantum Key Distribution (QKD), utilizes quantum principles to address this problem, but these solutions are still too immature and expensive for widespread deployment. The immediate response lies in adopting Post-Quantum Cryptography (PQC), new "classical" algorithms and protocols that can provably withstand the assault of quantum computers.

Given this sense of urgency, various regional (and not necessarily aligned) organizations are setting objective targets and defining plans for transformation. But what are the tools that a systems architect or a developer can pick today to be "Quantum-safe" or "Quantum-resistant"? And what is the actual value of these labels?

This article dives into the possibilities at hand and those being studied to keep communications secure in a post-quantum world and their application to different scenarios. It also discusses the role that telecommunications platforms can or soon will have to play in this context.

Authors

André Grilo, QuantumNova, andre.grilo@qnova.xyz

António Filipe, QuantumNova, antonio.filipe@qnova.xyz

Mafalda Nunes, Altice Labs, mafalda-g-nunes@alticelabs.com

Miguel Freitas, Altice Labs, miguel-r-freitas@alticelabs.com

Rui Calé, Altice Labs, Cale@alticelabs.com

Keywords

Post-Quantum Cryptography, Quantum Cryptography, Hybrid Cryptography, Quantum-safe, Quantum-resistant

DOI

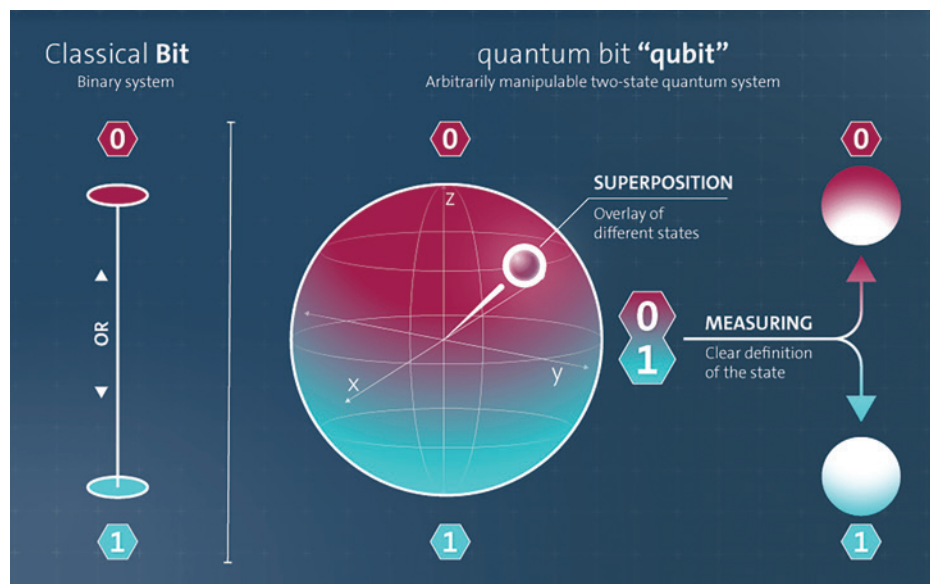
10.57650/INNOVATION/2025.14

Introduction

The digital infrastructure that supports modern society, from banking systems and healthcare records to national defense and global communications, relies heavily on cryptographic protocols that were designed before quantum computing was a practical concern. These protocols, particularly those based on asymmetric cryptography like Rivest–Shamir–Adleman (RSA) [1] and ECC (Elliptic Curve Cryptography) [2], [3], have long been considered secure due to the computational difficulty of problems like integer factorization and discrete logarithms. However, the emergence of quantum computing threatens to undermine this foundation.

Quantum computing represents a radical shift in computational capability. Unlike classical computers, which process information in binary bits (0s and 1s), quantum computers use quantum bits, or “qubits” [4], [5]. **Figure 1** illustrates this distinction, showing a representation of classical bits and quantum bits.

Figure 1: Classical Bit and Quantum Bit representation [6]



Thanks to the principle of superposition [7], a qubit can exist in a combination of the states $|0\rangle$ and $|1\rangle$ at the same time, allowing it to represent and process far more information than a classical bit. Qubits can also exhibit entanglement, a phenomenon in which the state of one qubit becomes intrinsically linked to the state of another, enabling powerful correlations that classical systems cannot replicate. Together, these properties allow quantum computers to perform certain calculations exponentially faster than their classical counterparts, particularly the computationally difficult problems that underpin modern cryptography. Shor’s algorithm [8], for example, demonstrates that a sufficiently powerful quantum computer could efficiently break RSA and ECC, compromising the confidentiality of today’s secure communications.

While such machines are not yet widely available, the trajectory of research and investment, driven by nation-states and technology leaders, suggests that their arrival is a matter of when, not if. This forthcoming milestone, in which a quantum computer can break modern cryptography, is often referred to as the “Q-Day” or

“Year to Quantum” (Y2Q). While the exact date remains uncertain, some sources estimate it could occur in as little as five years [9], [10]. Security agencies are establishing tight timelines for transitioning to quantum-resistant cryptography [11], [12], [13], with milestones between 2025 and 2035.

The urgency for this transition is not just theoretical: the race to quantum supremacy is now a geopolitical contest, with implications for national security, economic competitiveness, and technological sovereignty. The international response has been fragmented, with varying levels of investment, regulation, and urgency across regions, as evidenced by the Post-Quantum Cryptography Coalition’s International PQC Requirements analysis [14].

Intensifying the quantum threat is a practice known as “Harvest Now, Decrypt Later” (HNDL) or “Store Now, Decrypt Later” (SNDL), in which adversaries, often state-sponsored, intercept and store encrypted data today with the expectation of decrypting it once quantum capabilities mature [15]. Sensitive communications, intellectual property, and classified information are all potential targets, particularly those with long-term strategic or economic value. This means that current cryptographic standards can no longer guarantee long-term confidentiality, i.e., data considered secure today may be compromised in the future, exposing organizations and individuals to breaches long after the original interception.

Despite the growing awareness, the gap between recognition and action remains striking: a 2024 Entrust study found that while 61% of organizations plan to migrate to post-quantum cryptography within five years, only 41% are actively preparing today [16]. On the public web, F5 Labs reports that just 8.6% of the top one million most popular websites support hybrid PQC key exchange, with adoption in critical sectors like banking as low as 2.9% [17].

Seeking to help close the gap between awareness and action, this article explores the cryptographic responses to the quantum threat, the engineering challenges of building quantum-resilient systems, the global landscape of PQC initiatives, and the pivotal role of telecommunications platforms in securing the post-quantum future. The approaches of Quantum Nova as a Security Solutions Provider and of Altice Labs as a Communication Solutions Provider are also highlighted.

Cryptographic responses to the quantum threat

From theory to necessity: the quantum era is rewriting the rules of cybersecurity

With the quantum era rapidly approaching, the need for quantum-resilient cryptographic systems has shifted from a theoretical concern to an operational necessity. The cybersecurity community is actively pursuing two broad avenues: quantum cryptography, which uses quantum mechanics to perform various cryptographic tasks, and PQC, which involves designing classical algorithms resistant to quantum attacks.

Quantum cryptography offers elegant solutions like QKD, which leverages the fact that any attempt to eavesdrop on a quantum channel disturbs the transmitted qubits and can therefore be detected. However, its practical deployment remains limited due to high cost, complexity, and infrastructure requirements [18].



PQC, on the other hand, presents a more immediate and scalable path forward, with the American National Institute of Standards and Technology (NIST) Post-Quantum Cryptography standardization process already delivering candidate algorithms for global adoption [19]. Hence, PQC is generally accepted as the right approach to offer widely applicable and timely mechanisms for the immediate transition to a quantum-ready world.

Nevertheless, transitioning to quantum-resilient cryptographic systems demands careful and strategic planning. To prepare effectively for the post-quantum era, organizations must first identify which of their digital assets are most vulnerable to quantum attacks and prioritize remediation accordingly. This begins with a comprehensive cryptographic inventory, mapping where and how encryption is used across systems, networks, and data stores. Assets should then be classified by sensitivity, exposure, and required confidentiality lifespan, with particular attention to long-lived data that could be targeted under an HNDL strategy. By combining this classification with an assessment of cryptographic agility, organizations can develop a phased migration plan that addresses the highest-risk areas first [20], [21].

Engineering quantum-resilient systems

The emergence of quantum computing represents both a technological leap and a pressing security challenge. While the potential of quantum computing to solve complex problems is immense, its ability to break widely used cryptographic primitives poses an existential risk to digital security. Engineering Quantum-Resilient Systems requires more than just adopting new cryptographic algorithms. It demands a holistic approach that spans architecture, software engineering practices, system design, and cross-industry collaboration.

Post-Quantum Algorithms: PQC represents the first line of defense in the transition to quantum resilience. NIST has been leading the standardization process, with algorithms such as CRYSTALS-Kyber (for key encapsulation) and CRYSTALS-Dilithium (for digital signatures) emerging as frontrunners. Other approaches, like Falcon and SPHINCS+, address different trade-offs in efficiency and security. These algorithms rely on mathematical problems believed to be resistant to quantum attacks, such as lattice-based, hash-based, code-based, and multivariate polynomial problems. Their adoption will provide the cryptographic foundation for future systems, but widespread implementation requires careful consideration of performance and interoperability.

After many years in testing, on August 13, 2024, Crystals-Kyber, Crystals-Dilithium, and Sphincs+ were announced by NIST as Federal Information Processing Standards (FIPS) and became ML-KEM, ML-DSA, and SLH-DSA, respectively [22].

Hybrid Approaches: Given the uncertainty surrounding the maturity of PQC algorithms, many organizations are deploying hybrid cryptographic schemes. These combine traditional algorithms such as RSA or ECC with PQC counterparts, ensuring backward compatibility and defense-in-depth. For example, a hybrid TLS handshake may include both an elliptic-curve Diffie-Hellman exchange and a lattice-based key exchange, protecting against both classical and quantum adversaries.

This incremental approach allows systems to adopt PQC gradually, without forcing an immediate and risky “rip and replace” of existing cryptographic infrastructure. The hybrid approach is a great option for the transition. Conventional algorithms are “time-proven” and work well for current cyberthreats, while PQC algorithms are recent, meaning they could have unknown pitfalls yet to be discovered. The combination of both allows for PQC algorithms to mature in a safe environment.

Performance Trade-offs: Quantum-safe algorithms often introduce higher computational and memory costs compared to their classical predecessors. Lattice-based schemes, for instance, may require larger key sizes and longer handshake times, impacting real-time systems such as mobile applications or IoT devices. System architects must evaluate latency, throughput, and power consumption implications before deployment. For example, in high-frequency trading environments, even microsecond delays can be unacceptable. Engineering resilient systems, therefore, requires benchmarking across hardware, networks, and application layers to strike a balance between security strength and performance efficiency.

Cryptographic Agility and Modular Design: One of the most critical engineering practices is building cryptographic agility into systems. This means designing applications, libraries, and protocols such that cryptographic algorithms can be swapped out without major reengineering. Modular cryptographic frameworks, abstraction layers, and algorithm negotiation protocols are key in keeping an agile architecture. Standards such as TLS 1.3 already incorporate mechanisms for cipher suite negotiation, but many legacy protocols and custom implementations do not. Without agility, organizations risk being locked into obsolete algorithms or facing expensive and disruptive upgrades when new PQC standards evolve.

Integration with Legacy Systems: Enterprises today run on decades of accumulated infrastructure: legacy databases, mainframes, industrial control systems, and embedded devices. Engineering quantum-resilient solutions requires seamless integration with these environments. Not all legacy systems can be directly upgraded; some may require cryptographic gateways, wrappers, or middleware to enable secure communication with quantum-safe components. Migration strategies must prioritize critical systems (e.g., those handling sensitive financial transactions or classified government data) while ensuring business continuity. Backward compatibility and phased rollouts are essential to minimize operational risk. If a system is too dependent on years of patched infrastructure, or is looking to migrate to a quantum safe environment with ease, a good approach is to have a parallel solution to the system that may gap critical communications from quantum attacks, even if they are just HNDL. Solutions that aid in this transition can be found in companies such as QuantumNova [23]. The company’s main goal is to help companies

with whatever infrastructure, be it legacy or state of the art, prepare for this growing quantum world. Providing an adaptable software solution, QuantumNova can, with minimal setup, evolve a company's network to be quantum safe.

Hardware Acceleration and Optimization: The performance overhead of PQC can be mitigated by hardware acceleration. Chipmakers and vendors are beginning to explore quantum-safe cryptography at the hardware level. For example, Intel and ARM are experimenting with instruction set extensions that accelerate lattice-based operations. Similarly, Field-Programmable Gate Arrays (FPGA) and Graphics Processing Units (GPU) can be used to offload computationally expensive cryptographic primitives. Cloud providers may offer hardware-optimized PQC as a service, ensuring that developers can deploy secure systems without sacrificing performance. Hardware-software co-design will play a critical role in scaling PQC across consumer devices, data centers, and edge environments.

Developer Tooling – PQC Libraries and Software Development Kits (SDK): Developers today have access to an expanding ecosystem of open-source and commercial PQC libraries:

- **Open Quantum Safe (OQS)** and its C library `liboqs` provide reference implementations of candidate PQC algorithms, along with integrations for Transport Layer Security (TLS), Secure Shell (SSH), and Virtual Private Network (VPN) protocols [24].
- **Bouncy Castle PQC** extends the popular cryptographic library to include post-quantum primitives, making it accessible to Java developers [25].
- Other toolkits, such as Microsoft's **PQCrypto-VPN** or **Google's CECPQ2** experiments in Chrome, demonstrate how PQC can be tested in real-world applications.

These libraries accelerate experimentation and deployment, enabling developers to build quantum-resilient applications today, even before final standards are fully ratified.

Major Players: The push toward quantum resilience involves industry, academia, and governments. Internet giant **Cloudflare** supports and privileges PQC connections in (human) HTTPS traffic, with currently around 50% of the traffic to use an ECC+PQC connections [26]; **IBM** has a leading role in quantum computing hardware, but is also heavily engaged in developing PQC standards and integration within enterprise systems; **Google** has conducted early PQC experiments in Chrome (e.g., CECPQ1/CECPQ2) and continues to research hybrid and quantum-safe communication protocols; Microsoft offers PQC support in Azure and contributes to open-source PQC libraries; large Cloud providers like **AWS**, **Oracle** and **Alibaba Cloud** are exploring quantum-safe encryption in their platforms to protect customer workloads. These players are shaping both the algorithms and deployment models that enterprises will rely on.

Case Studies and Real-World Applications: Around the world, (i) **Government and Defense** agencies in the U.S., EU, and Asia are prioritizing quantum resilience for classified communications and critical infrastructure. The NSA's Commercial National Security Algorithm Suite 2.0 (CNSA 2.0) mandates a transition to PQC within the next decade; (ii) **Banks, payment providers, and other financial services** face risks from HNDL attacks. Institutions like Mastercard and JPMorgan

are piloting PQC in secure messaging and transaction channels; (iii) **Cloud and Enterprise Platforms** are integrating PQC into VPNs, identity management systems, and API gateways; (iv) **Consumer Applications** are also onboard. Messaging platforms such as **Signal** and **iMessage** are exploring PQC-enabled end-to-end encryption, ensuring long-term confidentiality of personal communications. These consumer-facing deployments will likely drive public awareness and adoption of quantum-safe solutions. More specific Telco business Use Cases can be found ahead, under “Telecommunications in the post-quantum era”.

Compliance, Testing & Validation efforts: National compliance frameworks (e.g., U.S. CNSA 2.0, EU guidelines) are starting to appear to push PQC adoption [27]. Interoperability testing (e.g., IETF hackathons, NIST SP 1800-38C draft) highlights practical issues like certificate size, hybrid compatibility, and software support [28]. Tools such as `pkilint` assist in validating PQC certificate implementations [29].

Interoperability & Fragmentation Risks: Divergent national preferences and asynchronous adoption schedules risk creating incompatibilities [30]. Also, legacy devices and bandwidth-constrained systems may struggle with larger PQC key sizes, despite the effort to keep PQC processes compatible with those kinds of devices [31].

Overall, cryptographic agility (the ability to update algorithms seamlessly) is becoming a key design principle [32].

Global landscape of PQC initiatives

As the quantum threat looms, governments, regional alliances, and international organizations are defining strategies to enable a smooth and secure transition to PQC. The following list provides an overview of current efforts and challenges to achieve this goal.

National & Regional Initiatives

In the USA, after a multi-year process, NIST released the first three PQC standards (FIPS 203, 204, 205) in August 2024 [33], and Congress issued the Quantum Computing Cybersecurity Preparedness Act, which mandates federal agencies to inventory quantum-vulnerable systems and prepare migration plans [34]. The Office of Management and Budget of the White House estimates ~\$7.1B will be required between 2025–2035 for PQC migration across U.S. federal systems [35].

Around the world, many countries’ governments have issued loosely aligned guidelines and roadmaps for this transition (Canada, UK, Japan, Australia, and others) [36], [37].

The European Union has agreed to a Coordinated Roadmap (2025), where the EU Member States commit to harmonized migration milestones [11]. This roadmap categorizes use cases by their level of quantum risk. In short, systems that require long-term confidentiality (10+ years) or where a breach would have a severe impact are considered high-risk. Use cases with significant confidentiality needs or

long transition timelines (over 8 years) fall into the medium- to high-risk category. By contrast, applications with limited confidentiality requirements and shorter migration paths are treated as low-risk. **Figure 2** illustrates how these categories guide prioritization in the EU's transition strategy.

Figure 2: EU coordinated roadmap for the transition to PQC



First Steps	Next Steps
<ul style="list-style-type: none"> • Identify and involve stakeholders. • Support mature cryptographic asset management. • Create dependency maps. • Perform quantum risk analysis. • Include the supply chain. • Create a national awareness and communication program. • Share knowledge and get involved with the NIS CG work stream on PQC. • Develop a timeline and an implementation plan. 	<ul style="list-style-type: none"> • Support cryptographic agility and a quantum-safe upgrade path. • Allocate resources for the transition. • Adapt certification schemes. • Evolve the rules. • Look for opportunities within the ecosystem. • Considering transversal activities throughout the creation and implementation of the roadmap. • Implement pilot use cases and contribute to testing centres.

International Organizations & Industry Coalitions

Besides national and regional authorities, various organizations lead the transition in different aspects: the Post-Quantum Cryptography Alliance (PQCA), Linux Foundation initiative that publishes migration roadmaps, develops open tools, and fosters collaboration among industry and government [38]; the World Economic Forum (WEF) engages governments, academia, and businesses in building governance frameworks for PQC adoption [39]; Standards Bodies like the Internet Engineering Task Force (IETF), European Telecommunications Standards Institute (ETSI), and others, are developing hybrid and composite schemes, migration guides, and interoperability testing for PQC in protocols like TLS and X.509 [40], [41], [42].

This landscape shows a clear global momentum: PQC standards are finalized, national deadlines are set, and interoperability challenges are being actively addressed. Yet the success of this transition will depend on coordinated action, cryptographic agility, and timely compliance testing.

Telecommunications in the post-quantum era

In the race toward a post-quantum era, telecom companies must lead the charge in securing communication

As central players in the information transfer ecosystem, telecom companies are under growing pressure to help shape a secure, post-quantum future. Besides having to meet the same challenges as other critical organizations, they manage a highly regulated, standards-based business. Regulators, standards bodies (NIST, 3GPP, ETSI, IETF), and industry organizations such as the GSM Association (GSMA) are cooperating to set up the basis for the transition.

Particularly relevant is the GSMA Taskforce on PQC [43], which started back in 2022 and now joins together more than 50 vendors and 20 operators. It involves a comprehensive approach by (1) identifying Use Cases, both for operators' internal resources and for customer-facing services; (2) providing common guidelines to the industry; and (3) collaborating with the Standards Definition Organizations, like NIST, ETSI, IETF, and the 3rd Generation Partnership Project (3GPP).

These Use Cases are important not only to operators, as they describe concrete situations to be addressed, but also to solution providers, as they define requirements their products must meet, to guarantee a timely and rich supply chain, providing the solutions that will feed the transition.

The next table shows an abbreviated list of these use cases [44]:

Table 1: List of PQC use cases affecting 5/6G and IoT on a Telco context

Network Operator Use-Case	Customer impacting Use-Case
Protection of the interface between base stations and security gateway	Virtual private network (VPN) services (VPN)
Virtualized network functions	Software-Defined Wide Area Network (SD-WAN) services
Cloud infrastructure	Internet of Things (IoT) smart meters
SIM (physical)	IoT / Automotive
eSIM (remote provisioning)	Lawful intercept
Devices and firmware update	Privacy of customer data
Concealment of the subscriber public identifier	
Authentication and transport security in cellular networks	

Besides these general Use Cases, the GSMA has, to date, already issued specific official documents on PQC in IoT [45] and for 5G Roaming Use Case [46].

The GSMA highlights that the migration to PQC should be carefully prioritized based on a combination of technical, operational, and strategic factors rather than approached as a uniform transition across all systems. The goal is to ensure

that the most critical assets and functions are protected first while maintaining system stability and operational continuity.

According to the GSMA [44], several key aspects guide this prioritization:

- **System criticality** – Systems whose compromise would lead to severe consequences should be updated first.
- **Security properties** – The sensitivity of data and the type of protection required matter. For example, confidentiality is crucial for data vulnerable to HNDL attacks, while authenticity becomes paramount for long-term digital signatures and legal documents.
- **Update complexity and timing** – Some updates, such as firmware or authentication mechanisms, may take longer and should be prioritized early to ensure continuity.
- **System exposure** – Public-facing or internet-connected systems are at higher risk and therefore should be prioritized ahead of internal or isolated systems.
- **Regulatory guidance** – Recommendations from cybersecurity agencies and regulators may also influence which systems transition first.

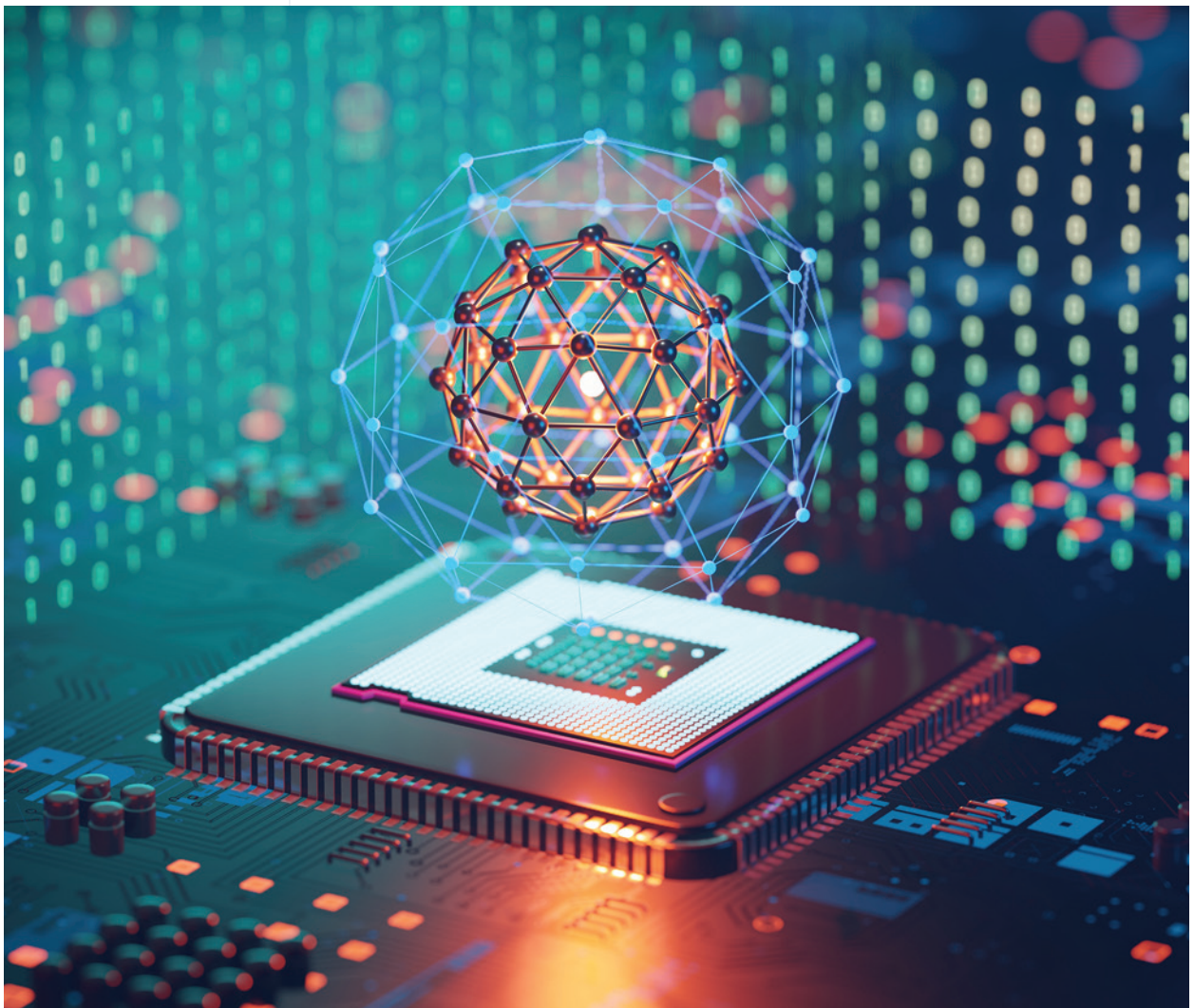
In summary, a risk-based, phased approach to PQC migration is advised, balancing urgency, feasibility, and regulatory compliance to achieve a secure and efficient transition to quantum-resistant systems. Transitioning to post-quantum cryptography must involve multiple stakeholders, from standards bodies to operators, each contributing to a coordinated migration effort. The established international roadmaps emphasize standardization, coordination, and phased implementation to achieve a secure and smooth global migration to post-quantum cryptography (albeit they are not absolutely aligned, which brings added complexity to international/global service providers). The roadmap proposed by GSMA identifies the following phases:

- **Standards Development:** The process should begin with the establishment of key PQC standards by organizations such as NIST and IETF, covering IPsec/TLS protocols, PKI systems, key exchange and hybrid mechanisms, and MACsec/MKA frameworks.
- **National Guidelines:** Next, national authorities issue guidelines, migration recommendations, and regulations to ensure consistent and secure adoption across sectors.
- **Vendor Implementation:** Vendors then update their products to support PQC, integrating new algorithms into VPNs, TLS stacks, cryptographic libraries, PKI systems, and network security protocols.
- **Operator Deployment:** Network operators proceed to deploy PQC-enabled solutions, ensuring updated IPsec/TLS, PKI, MACsec, and digital signature algorithms are integrated into operational environments.
- **Proof of Concept (PoC):** Integration and end-to-end testing phases validate interoperability and performance across the ecosystem before full rollout.
- **Migration:** Finally, systems are migrated to PQC, marking the transition to quantum-resistant infrastructure.

On the standards path for telecommunications, besides the already mentioned role of NIST in standardizing PQC algorithms, The ITU-T is providing a global, overarching view; The European Telecom Standards Institute (ETSI) Cyber-Security Committee hosts a Quantum Safe Cryptography workgroup [42], which has been publishing Technical Reports and Specifications since 2017; the 3GPP, in Release 20, is introducing studies on PQC in mobile network specifications for 5G-Advanced and 6G [47]; the IETF is standardizing the protocols that run across the Internet (TLS, IPsec, SSH, etc.).

Altice Labs towards quantum readiness

Altice Labs has long recognized the transformative impact that quantum computing will bring to the field of cybersecurity and has been actively preparing for this new era. Aligned with the roadmap presented before and recognizing the importance of trials and proof-of-concept implementations, it is currently embracing a hybrid strategy: engaging in PQC initiatives while also closely following developments in QKD.



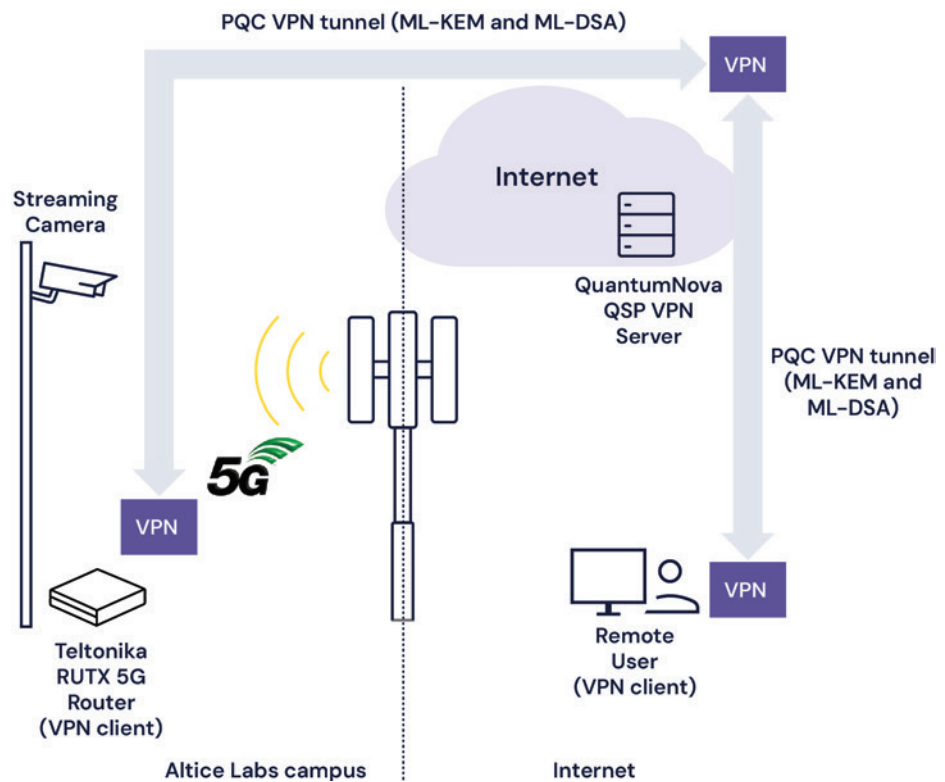
Altice Labs is pursuing a hybrid path: advancing post-quantum cryptography while pioneering quantum key distribution

In practice, Altice Labs’ work on PQC has focused on establishing partnerships and conducting pilot trials with companies that specialize in quantum-safe technologies, fostering hands-on experience and knowledge exchange. On the QKD side, the company actively participates in several European and nationally funded projects, both learning from and contributing to the advancement of the field. These collaborations are built on the company’s established expertise, most notably in Software Defined Networking (SDN), while also extending its contributions to new and complementary domains, helping to drive the integration and practical application of QKD technologies into real-world communication infrastructures and networks.

Altice Labs PQC pilot initiatives

A notable example of Altice Labs’ efforts in the post-quantum cryptography domain was a pilot project carried out in May 2025, in collaboration with QuantumNova, a company specializing in quantum-safe cybersecurity solutions. The pilot successfully established a post-quantum VPN tunnel (QuantumNova Q-SP network) between the private 5G network of the Altice Labs campus and the internet, demonstrating the feasibility of securing next-generation communications against quantum-enabled threats. In the pilot, a point-to-point client-server connection was established for a camera streaming within the private 5G network, via a commercial off-the-shelf (COTS) 5G router (Teltonika RUTX50), secured through a tunnel that provided mutual authentication and confidentiality using post-quantum cryptography methods, specifically Module-Lattice Key Encapsulation Mechanism (ML-KEM) and Module-Lattice Digital Signature Algorithm (ML-DSA), which were implemented within a VPN system based on OpenVPN (Figure 3).

Figure 3: Altice Labs and Quantum Nova PQC VPN tunnel setup

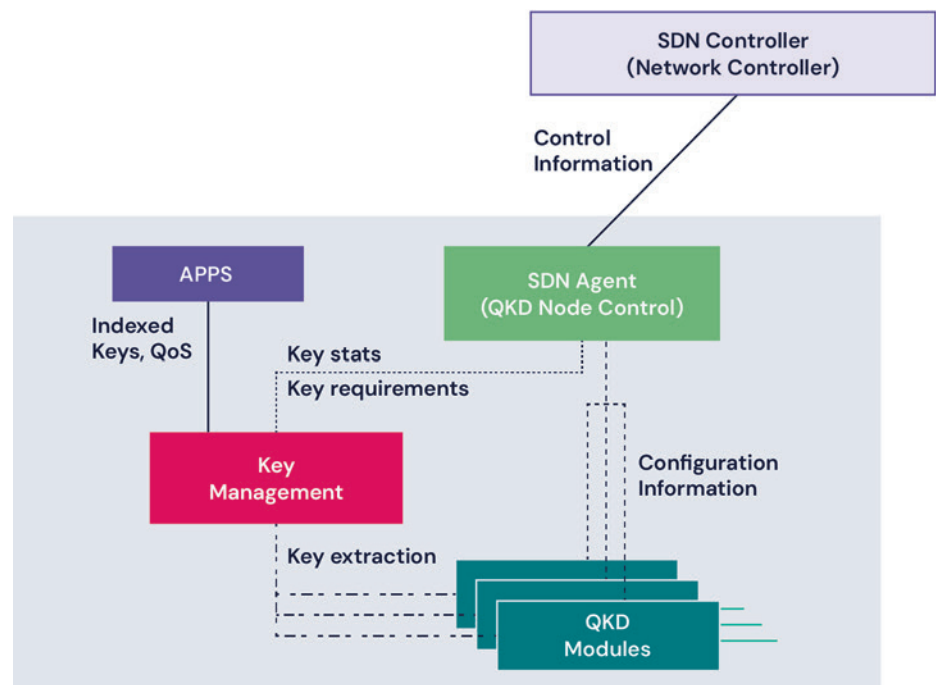


The pilot setup showed that using the QS-P networks (which went through a lot of iteration and initial research to build) to integrate PQC into an existing system requires minimal effort, proving that the transition to quantum-safe solutions can be smooth and practical, even taking advantage of already existing hardware. The successful demonstration helped move the QuantumNova QS-P network client product from TRL3 to TRL5.

Altice Labs QKD pilot initiatives

Besides PQC, Altice Labs is involved in Quantum Cryptography efforts, namely on QKD. One notable initiative in the QKD domain where Altice Labs is actively involved is the Portuguese Quantum Communications Infrastructure (PTQCI) project [48], which builds on prior expertise from the DISCRETION project and forms part of the broader EuroQCI initiative [49]. The project is establishing the first segment of the Quantum Communication Infrastructure (QCI) in Portugal, connecting sovereign points in the Lisbon metropolitan area. PTQCI brings together national stakeholders, including the National Security Authority (GNS – Gabinete Nacional de Segurança), the Portuguese government, telecom operators, industry, and academia, to deploy an advanced quantum network that integrates with existing communication infrastructures and prepares for future expansion across mainland Portugal and the islands, as well as to Spain and the rest of Europe. In this project, Altice Labs is designing, developing, and implementing the entire SDN architecture and its components to support the QKD network, compliant with the ETSI QKD 015 standard [50], as illustrated in **Figure 4**.

Figure 4: QKD enabled SDN node architecture [50]



Another initiative under way in Altice Labs is the exploration of its Passive Optical Network (PON) product as a feasible quantum communications infrastructure, sharing media with classical communications [51].

Altiice Labs' work in quantum communications emphasizes that PQC and QKD are in fact complementary approaches, and that the urgency in implementing the former will probably be followed by the added security of combining it with the latter. By exploring both approaches in parallel, the research bridges cutting-edge academic developments with real-world deployments, ensuring that future networks are secure, resilient, and ready for the post-quantum era. These efforts also have a direct impact on the company's product portfolio, helping to shape solutions that are future-proof and aligned with emerging security requirements.

Final thoughts and future directions

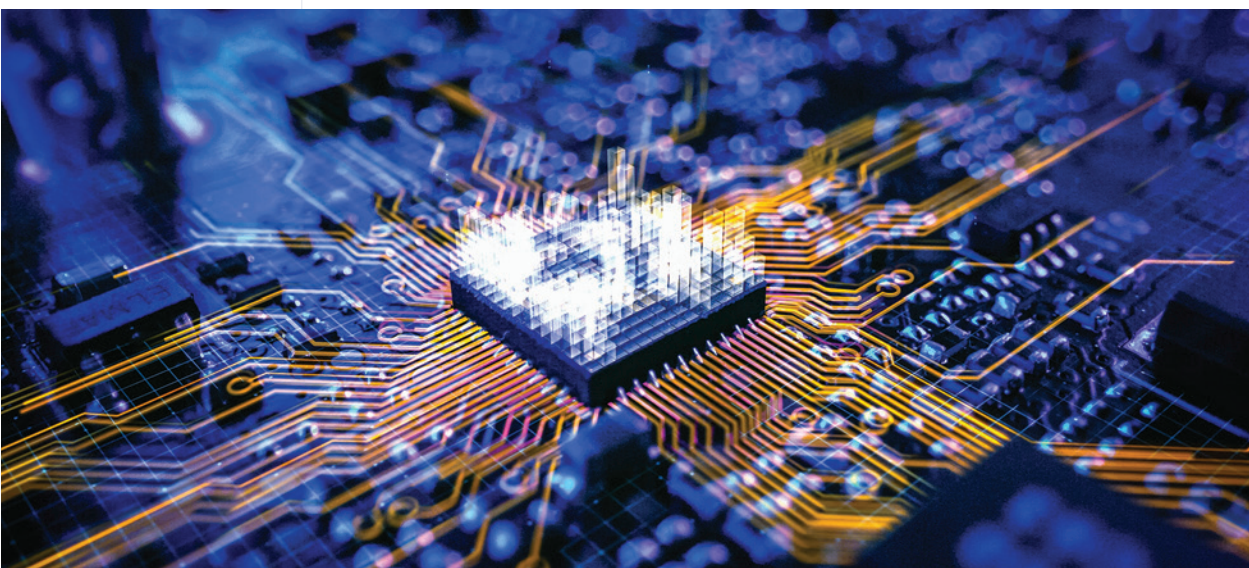
Quantum computing challenges long-standing cryptographic foundations and demands a coordinated response

The arrival of quantum computing marks a turning point for cybersecurity, challenging long-standing cryptographic foundations and demanding a proactive and coordinated response. As the industry transitions toward quantum-resistant systems, collaboration across standardization bodies, vendors, operators, and research centers becomes essential to ensure a secure and orderly migration.

Importantly, this is not just a migration to new algorithms. The dynamics will be different, and changes to cryptographic systems are likely to be frequent. Hence, cryptoagility needs to be unlocked with the quantum migration in these entities, paving the way to an ever-evolving world in the cybersecurity scope.

Telecommunications operators have a pivotal role in this process, as they provide the backbone of global connectivity and must guarantee that future networks remain both interoperable and resilient in the face of emerging quantum threats.

As the quantum era approaches, preparing today is the key to securing tomorrow. The shift to quantum-resistant cryptography is not merely a technological upgrade but a strategic necessity, one that will define the trust and resilience of digital communications for decades to come. By anticipating change, fostering collaboration, and embracing innovation, the industry can ensure that the foundations of cybersecurity remain strong in a world transformed by quantum technology.



References

- [1] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-key Cryptosystems," MIT Laboratory for Computer Science, Cambridge, MA, USA, Tech. Rep. MIT-LCS-TM-082, Apr. 1977. Accessed: Sep. 12, 2025. [Online]. Available: <https://hdl.handle.net/1721.1/148910>
- [2] V. Miller, "Use of Elliptic Curves in Cryptography," in *Proc. Advances in Cryptology – CRYPTO '85*, Santa Barbara, CA, USA, Aug. 18–22 1985, pp. 417–426. [Online]. Available: https://www.researchgate.net/publication/221355411_Use_of_Elliptic_Curves_in_Cryptography
- [3] N. Koblitz, "Elliptic Curve Cryptosystems," *Math. Comput.*, vol. 48, no. 177, pp. 203–209, Jan. 1987, doi: 10.1090/S0025-5718-1987-0866109-5.
- [4] Quantum Inspire. "Superposition and entanglement." Quantum Inspire. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.quantum-inspire.com/kbase/superposition-and-entanglement/>
- [5] Domain of Science, "The Map of Quantum Computing – Quantum Computing Explained." (Dec. 3, 2021). Accessed: Sep. 12, 2025. [Online Video]. Available: <https://www.youtube.com/watch?v=-UlxHPiEVqA>
- [6] Devopedia, "Qubit," Devopedia. Accessed: Sep. 12, 2025. [Online]. Available: <https://devopedia.org/qubit>
- [7] D. J. Griffiths, *Introduction to Quantum Mechanics*, 3rd Edition. Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [8] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Santa Fe, NM, USA, 1994, pp. 124–134, doi: 10.1109/SFCS.1994.365700.
- [9] Cloud Security Alliance. "Quantum-safe Security Working Group." Cloud Security Alliance. Accessed: Sep. 12, 2025. [Online.] Available: <https://cloud-securityalliance.org/research/working-groups/quantum-safe-security>
- [10] Indigo Vault. "Understanding Y2Q: The Quantum Threat to Encryption." Indigo Vault. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.indigovault.com/insights/understanding-y2q-the-quantum-threat-to-encryption/>
- [11] European Commission. "A Coordinated Implementation Roadmap for the Transition to Post-Quantum Cryptography." European Commission. Accessed: Sep. 12, 2025. [Online.] Available: <https://digital-strategy.ec.europa.eu/en/library/coordinated-implementation-roadmap-transition-post-quantum-cryptography>
- [12] National Security Agency, "Announcing the Commercial National Security 2.0," National Security Agency, Fort Meade, MD, USA, Sep. 2022. Accessed: Sep. 12, 2025. [Online]. Available: https://media.defense.gov/2025/May/30/2003728741/-1/-1/0/CSA_CNSA_2.0_ALGORITHMS.PDF

- [13] Canadian Centre for Cyber Security, "Roadmap for the migration to post-quantum cryptography for the Government of Canada (ITSM.40001)," Government of Canada, Ottawa, ON, Canada, Jun. 2025. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.cyber.gc.ca/en/guidance/roadmap-migration-post-quantum-cryptography-government-canada-itsm40001>
- [14] P. Kampanakis, M. Manzano, D. Connolly, S. Jamshidi, and A. Bosarge. "International PQC Requirements." Post-Quantum Cryptography Coalition. Accessed: Sept. 12, 2025. [Online.] Available: <https://pqcc.org/international-pqc-requirements>
- [15] R. DuBose e M. M. Rao. "Harvest now, decrypt later: Why today's encrypted data isn't safe forever." HashiCorp. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.hashicorp.com/pt/blog/harvest-now-decrypt-later-why-today-s-encrypted-data-isn-t-safe-forever>
- [16] K. Kadet. "Post Quantum Cryptography Awareness is High, But Widespread Preparation Lags, Finds 2024 Global Entrust Report." Entrust. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.entrust.com/company/news-room/post-quantum-cryptography-awareness-is-high-but-widespread-action-lags-finds-2024-global-entrust-report>
- [17] D. Warburton and M. Heath. "The State of Post-Quantum Cryptography (PQC) on the Web." F5 Labs. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.f5.com/labs/articles/the-state-of-pqc-on-the-web>
- [18] N. Aquina et al., "A critical analysis of deployed use cases for quantum key distribution and comparison with post-quantum cryptography," *EPJ Quantum Technol.*, vol. 12, Art. no. 51, 2025, doi: 10.1140/epjqt/s40507-025-00350-5
- [19] National Institute of Standards and Technology. "Post-Quantum Cryptography." NIST. Accessed: Sep. 12, 2025. [Online]. Available: <https://csrc.nist.gov/projects/post-quantum-cryptography>
- [20] National Institute of Standards and Technology. "Migration to Post-Quantum Cryptography." NIST. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.nccoe.nist.gov/crypto-agility-considerations-migrating-post-quantum-cryptographic-algorithms>
- [21] National Cyber Security Center, "Timelines for migration to post-quantum cryptography." UK National Cyber Security Center. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.ncsc.gov.uk/guidance/pqc-migration-timelines>
- [22] National Institute of Standards and Technology. "Announcing Approval of Three Federal Information Processing Standards (FIPS) for Post-Quantum Cryptography." NIST. Accessed: Sep. 12, 2025. [Online]. Available: <https://csrc.nist.gov/news/2024/postquantum-cryptography-fips-approved>
- [23] QuantumNova. "QuantumNova – Post-Quantum Security." QuantumNova. Accessed: Sep. 12, 2025. [Online]. Available: <https://quantumnova.xyz/>

- [24] Open Quantum Safe. "Open Quantum Safe – software for the transition to quantum-resistant cryptography." Open Quantum Safe. Accessed: Sep. 12, 2025. [Online]. Available: <https://openquantumsafe.org/>
- [25] Bouncy Castle. "Latest NIST PQC Standards and more – Bouncy Castle Java 1.79." Bouncy Castle. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.bouncycastle.org/resources/latest-nist-pqc-standards-and-more-bouncy-castle-java-1-79/>
- [26] Cloudflare Radar. "Adoption & Usage Worldwide." Cloudflare Radar. Accessed: Sep. 12, 2025. [Online]. Available: <https://radar.cloudflare.com/adoption-and-usage#post-quantum-encryption-adoption>
- [27] National Security Agency, "Announcing Commercial National Security Algorithm Suite 2.0 (CNSA 2.0) Cybersecurity Advisory," National Security Agency, Fort Meade, MD, USA, Sep. 2022. Accessed: Sep. 12, 2025. [Online]. Available: <https://media.defense.gov/2022/Sep/07/2003076874/-1/-1/0/CNSA-2.0-FACT-SHEET.PDF>
- [28] National Institute of Standards and Technology, "PQC migration preliminary draft: NIST SP 1800-38C." NIST. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.nccoe.nist.gov/sites/default/files/2023-12/pqc-migration-nist-sp-1800-38c-preliminary-draft.pdf>
- [29] C. Bonnell. "PKIint and the Path to Interoperable, Quantum-Safe PKI." DigiCert Blog. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.digicert.com/blog/pkilint-and-the-path-to-interoperable-quantum-safe-pki>
- [30] Keyfactor. "Post-quantum interoperability in action: How collaboration is shaping the PQC future of cybersecurity." Keyfactor. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.keyfactor.com/blog/post-quantum-interoperability-in-action-how-collaboration-is-shaping-the-pqc-future-of-cyber-security/>
- [31] SafeLogic. "PQC standards & global roadmaps." SafeLogic. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.safelogic.com/compliance/pqc-standards>
- [32] E. Barker et al., "Considerations for Achieving Crypto Agility: Strategies and Practices," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST CSWP 39 2pd, 2025, doi: 10.6028/NIST.CSWP.39.2pd.
- [33] National Institute of Standards and Technology. "NIST releases first 3 finalized post-quantum encryption standards." NIST. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards>
- [34] U.S. Congress, "Quantum Computing Cybersecurity Preparedness Act (Public Law 117-260)," 117th Congress, 2022. [Online]. Available: <https://www.congress.gov/bill/117th-congress/house-bill/7535/text>
- [35] C. Burgan. "White House: Agencies Need \$7.1 B to Transition to PQC." MeriTalk. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.meritalk.com/articles/white-house-agencies-need-7-1b-to-transition-to-pqc/>

- [36] SafeLogic. "Post-Quantum Cryptography (PQC) Compliance Standards: Key Compliance Mandates Shaping the Future of Cryptography." SafeLogic. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.safelogic.com/compliance/pqc-standards>
- [37] GSMA. "Post-Quantum Government Initiatives by Country and Region." GSMA. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.gsma.com/newsroom/post-quantum-government-initiatives-by-country-and-region>
- [38] Post-Quantum Cryptography Coalition. "Post-Quantum Cryptography (PQC) Migration Roadmap." Accessed: Sep. 12, 2025. [Online]. Available: <https://pqcc.org/wp-content/uploads/2025/05/PQC-Migration-Roadmap-PQCC-2.pdf>
- [39] World Economic Forum. "Why the New NIST Standards Mean Quantum Cryptography May Just Have Come of Age." World Economic Forum. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.weforum.org/stories/2024/10/quantum-cryptography-nist-standards>
- [40] K. Kwiatkowski. "Guidance for Migration to Post-Quantum Cryptography." Internet Engineering Task Force (IETF). Accessed: Sep. 12, 2025. [Online]. Available: <https://www.ietf.org/archive/id/draft-kwiatkowski-pquip-pqc-migration-00.html>
- [41] T. K. Reddy and H. Tschofenig. "Post-Quantum Cryptography Recommendations for TLS-Based Applications," Internet Engineering Task Force (IETF) Datatracker. Accessed: Sep. 12, 2025. [Online]. Available: <https://datatracker.ietf.org/doc/draft-ietf-uta-pqc-app/>
- [42] European Telecommunications Standards Institute (ETSI). "Quantum-Safe Cryptography (QSC)." ETSI. Accessed: Oct. 3, 2025. [Online]. Available: <https://www.etsi.org/technologies/quantum-safe-cryptography>
- [43] GSMA. "Post-Quantum Telco Network Task Force." GSMA. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.gsma.com/solutions-and-impact/technologies/security/post-quantum-telco-network-task-force/>
- [44] GSMA. "Post-Quantum Cryptography Guidelines for Telecom Use Cases." Rep. PQ.03, v1.0, Feb. 2024. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.gsma.com/solutions-and-impact/technologies/security/wp-content/uploads/2024/08/PQ.03-Post-Quantum-Cryptography-Guidelines-for-Telecom-Use-v1.0.pdf>
- [45] GSMA, "Post-Quantum Cryptography in IoT Ecosystem," GSMA, White Paper, Nov. 2025. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.gsma.com/solutions-and-impact/technologies/internet-of-things/wp-content/uploads/2025/11/Post-Quantum-Cryptography-in-IoT-Ecosystem.pdf>
- [46] GSMA, "Post-Quantum Cryptography for 5G Roaming Use Case," GSMA, Rep. PQ.05, Mar. 2019. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.gsma.com/solutions-and-impact/technologies/security/wp-content/uploads/2019/03/PQ.05-Post-Quantum-Cryptography-for-5G-Roaming-use-case-1.pdf>

- [47] M. C. Soveri, "Study on Transitioning to Post Quantum Cryptography (PQC) in 3GPP," 3GPP, Rep. 33.703 (TR), Draft. Accessed: Sep. 12, 2025. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4422>
- [48] Portuguese Quantum Communication Infrastructure. "Home Page". PTQCI. Accessed: Sep. 12, 2025. [Online]. Available: <https://www.ptqci.pt/>
- [49] A. N. Pinto, C. Bastos, G. C. Ângelo, L. Maia, R. Chaves, and R. Calé, "Quantum Key Distribution for secure communications", *Innovaction*, no. 8, pp. 58–73, 2023.
- [50] European Telecommunications Standards Institute (ETSI), "Quantum Key Distribution (QKD); Control Interface for Software Defined Networks," ETSI, Rep. GS QKD 015 V2.1.1. Accessed: Nov. 13, 2025. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/QKD/001_099/015/02.01.01_60/gs_QKD015v020101p.pdf
- [51] D. Cruz, N. J. Muga, S. Mantey, A. Tavares, and C. Rodrigues, "Impact of Passive Optical Network technology on polarization encoded Discrete Variable Quantum Key Distribution", *Innovaction*, no. 9, pp. 94–105, 2024.



15

Shaping the Future of Data: Why Gaia-X and IDSA are Crucial for a Stable and Innovative Digital Europe

Abstract

In an increasingly digitalized world, data drives innovation and economic growth, but its concentration in tech giants like Amazon, Google, and Meta raises concerns about sovereignty, security, dependence, and ethics. Centralized data models often compromise privacy and organizational control, highlighting Europe's urgent need for secure and sovereign data infrastructures to maintain digital autonomy and competitiveness.

This article introduces Gaia-X and the International Data Spaces Association (IDSA) as cornerstone initiatives designed to transform how data is shared and governed. Gaia-X defines common standards for transparent, federated, and user-controlled data ecosystems, while IDSA provides the technical framework for secure and trusted data exchange. Together, they promote interoperability, ethical governance, and a fairer digital market that supports innovation across industries, from healthcare to manufacturing.

Understanding these initiatives is crucial for anyone interested in the future of digital society, as they lay the foundation for a sovereign and resilient digital landscape where data is harnessed responsibly for collective benefit.

Authors

Bruno Santos, Altice Labs, bruno-g-santos@alticelabs.com

Tiago Lameirão, University of Trás-os-Montes and Alto Douro, al70636@alunos.utad.pt

Miguel Mesquita, Altice Labs, mesquita@alticelabs.com

Filipe Cabral Pinto, Altice Labs, filipe-c-pinto@alticelabs.com

Keywords

Gaia-X, International Data Spaces Association, Data Sovereignty, Interoperability, Digital Economy, Data Space

DOI

10.57650/INNOVATION/2025.15



Introduction

The digitalization of society has turned data into a pillar of modern economies, fostering innovation in artificial intelligence, machine learning, and personalized services. However, this rapid evolution has also exposed weaknesses in data management across the globe. Predominant platforms owned by non-European actors are non-interoperable, based on black box algorithms, which foster data monopolies and loss of sovereignty [1]. Traditional data ecosystems have centralization aspects in storage, processing, and monetization of large datasets. This implementation has driven targeted advertising and predictive analytics, which, in some cases, have led to privacy breaches, creating public distrust [2]. Moreover, the concentration of power raises antitrust concerns, reducing competition and innovation from smaller companies.

In Europe, the pursuit of digital sovereignty, where a user can control data flows while preserving core values like privacy and transparency, has become a strategic imperative. Since non-European cloud providers handle a significant portion of European data [3], this dependency undermines Europe's competitiveness and exposes critical sectors, such as manufacturing and energy, to risks. Therefore, barriers that prevent seamless data sharing can cause ethical dilemmas surrounding the use of data without consent. This was exacerbated during the COVID-19 pandemic, when rapid data exchanges caused vulnerabilities for contact tracing and cross-border collaboration [4].

Tackling these issues requires a paradigm shift toward federated systems that prioritize sovereignty, security, and fairness. That is where both projects Gaia-X and IDSA come in, representing this shift by offering frameworks that decentralize control while enabling value creation from data. Gaia-X and IDSA emerge as crucial efforts to address these challenges by creating a decentralized, trustworthy data ecosystem that empowers users and businesses alike [5]. By fostering interoperability and security, they not only mitigate risks associated with centralized data control but also pave the way for innovation that aligns with European regulations, such as the General Data Protection Regulation (GDPR). This article explores the foundations, mechanisms, and implications of Gaia-X and IDSA, underscoring their role in securing a stable and innovative digital future for Europe.

Gaia-X: a federated data infrastructure

Overview

Gaia-X, whose framework is represented in **Figure 1**, is a European initiative aiming to establish a federated and secure data infrastructure as the foundation for a

By fostering secure cross-border data sharing, Gaia-X reduces reliance on American and Chinese cloud platforms

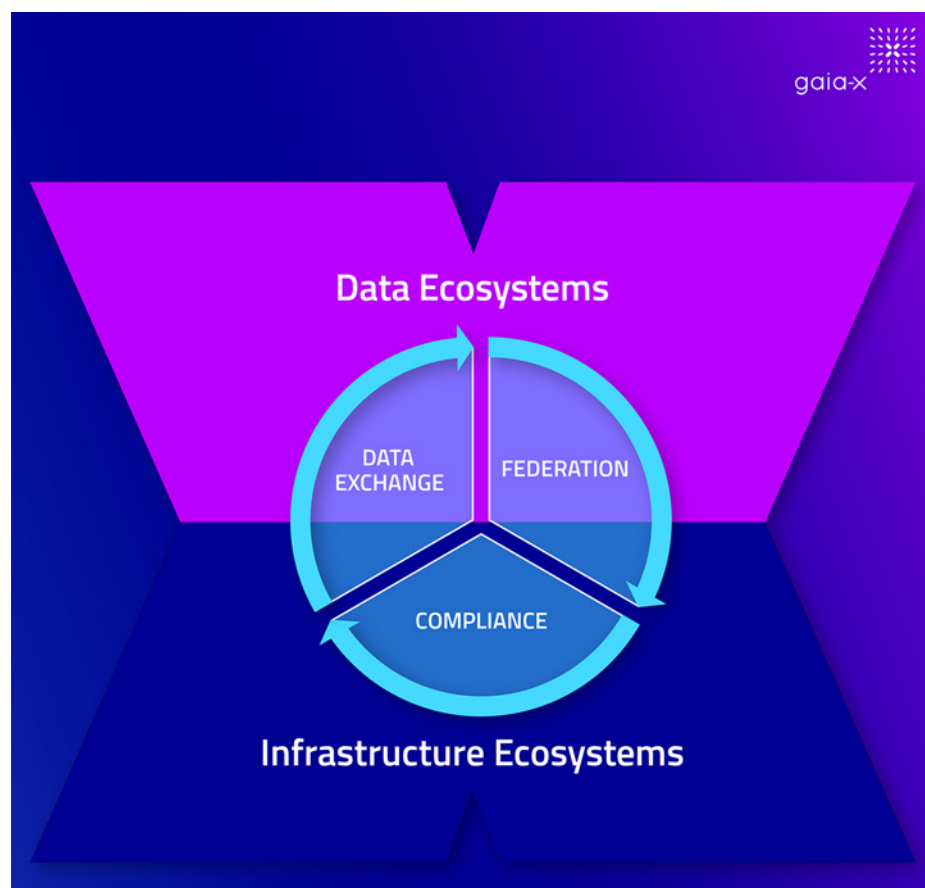
trusted digital economy. Launched in 2019 and rooted in the German and French governmental seal, its mission is to establish a trusted, decentralized network where data and services can be shared securely. Gaia-X sets common rules, standards, and governance so that existing cloud and edge services can interoperate while ensuring data sovereignty, transparency, and openness, reflecting the European aims [6].

Gaia-X aims to make data sharing across organizations, sectors, and borders secure, trustworthy, and sovereign, enabling Europe to build strong digital ecosystems that compete globally, reducing dependence on global hyperscalers. Innovation and the resulting value creation will be cornerstones of fostering European leadership in the digital economy, making it autonomous from American and Chinese cloud platforms [7].

Gaia-X focuses on three major concepts:

- **Federation** – Enabling a network of independent actors, such as edge and cloud providers, data holders, or data consumers, under common governance and rules, without being centralized under a single and closed platform.
- **Compliance** – Ensuring that participants and services meet the Gaia-X rules, standards, and trust framework, in alignment with European regulations, guaranteeing the entire ecosystem remains securely transparent and interoperable.
- **Data Exchange** – Enabling trusted, secure, and sovereign sharing of data between participants within a federated ecosystem, ensuring that the terms, rights, and technical conditions of data sharing are transparent, verifiable, and compliant with Gaia-X principles.

Figure 1: Gaia-X Framework [8]



Gaia-X deployment

Digital Clearing Houses ensure transparent, standards-compliant data exchange while preventing vendor lock-in

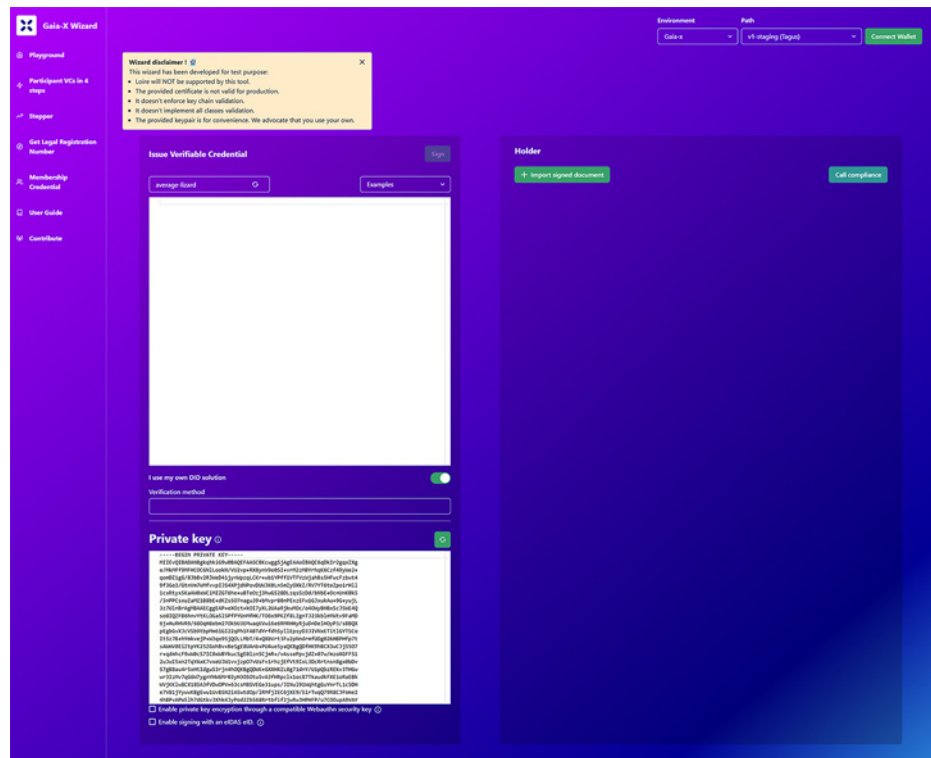
The recent Gaia-X architecture revolves around Digital Clearing Houses, which are decentralized nodes that verify compliance with standards and facilitate interoperability. These nodes ensure that data exchanges are transparent and auditable, using open-source tools to prevent vendor lock-in. Participants, including businesses and governments, retain sovereignty over their data, deciding access rights and usage policies [9].

The current version of Gaia-X, Loire, seeks to lower the ambitions of the previous Gaia-X release, Tagus, whose broader objectives ultimately proved to be unrealistic. The initial Gaia-X requirements were too wide in scope, leading to implementations that were incomplete or overly complex [10], [11]. To address this, the Digital Clearing House was devised to achieve a more compact set of specifications, based on the registry, compliance, and notary implementations. The Loire version aims to achieve a visible improvement in performance while reducing resources compared with the previous version [12].

Gaia-X will thus be responsible for managing all issues related to security in the transactions going through the platform. It will handle the issuance of credentials for all actors involved, whether they are application providers or clients. There is also a strong component to ensure compliance between the content within the platform and the possible external clients. These compliance assurances will provide seamless connections, minimizing errors during the interactions. The Digital Clearing House will not cover data exchange issues among actors within the Gaia-X ecosystem, as this topic, originally included in the early specifications, was dropped in later versions and is now expected to be handled by the IDSA [12].

Figure 2 presents the wizard running in the Tagus version environment.

Figure 2: Gaia-X Wizard [13]



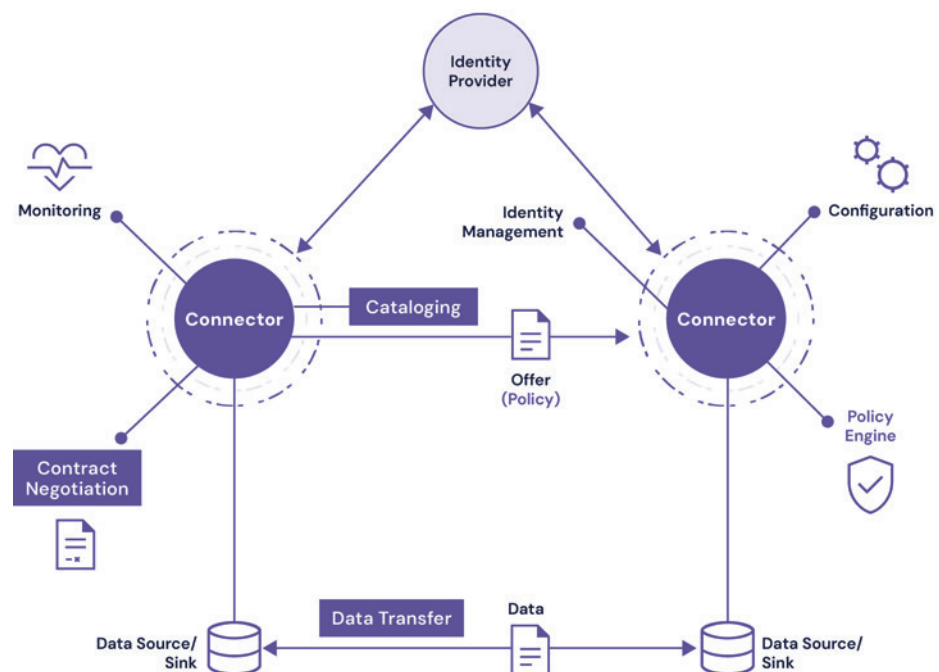
IDSA: enabling secure data spaces

Overview

A data space (DS) is a trusted environment where different entities can share data while maintaining full control of its usage [14]. IDSA aims to drive innovation in the future of data exchange in Europe and beyond. It is a non-profit alliance of more than 190 entities (including companies, research organizations, and institutions) dedicated to developing standards and architectures for data spaces, ensuring data sovereignty and promoting a fair and secure digital economy [15]. Since 2017, the business and research communities have been jointly determining a trusted, open architecture for the data economy, enabling peer-to-peer data exchange across all sectors of activity. This approach ensures true data sovereignty through usage control and enables data providers to unlock the full value of their data within trusted partnerships [16].

IDSA focuses on developing the technical foundations for sovereign data sharing. The International Data Spaces Reference Architecture Model (IDS-RAM) provides a blueprint for creating data spaces where organizations control their data's lifecycle. The architecture includes connectors, like the ones represented in **Figure 3**, that enforce usage policies, ensuring that data is shared only under predefined conditions [17]. The Data Space Protocol (DSP) defines standardized interactions that enable seamless integration across platforms and support cross-industry applications, from smart cities to agriculture, while preserving data sovereignty and building trust through certification schemes that reduce risks in data exchange [17]. With members across various sectors, IDSA's standardization efforts can collaborate with bodies such as ISO to extend their global impact [15], fostering the vision of a fair data economy that aims at solving interoperability and security challenges.

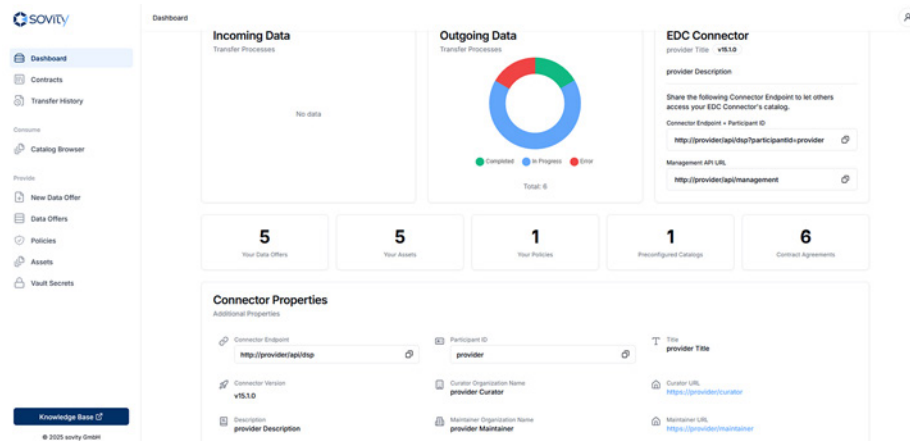
Figure 3: Data Connectors [18]



IDSA deployment

One of the tools that enables a federated ecosystem aligned with IDSA-based specifications is Sovity, whose Community Edition is represented in **Figure 4**. Altice Labs has been keen on exploring this technology's core innovation, which is an open-source extension of the EDC that natively integrates IDSA standards. This connector acts as a sovereign gateway, allowing users and organizations to deploy customizable data offers with granular policies for data access, transfer, and usage (without compromising control), supported by catalogs for service exposure [19]. A key point of the Sovity platform is the enablement of robust mechanisms for managing and revoking data-sharing agreements in sovereign ecosystems, allowing users to seamlessly manage contracts by defining their specific parameters.

Figure 4: Sovity Community Edition



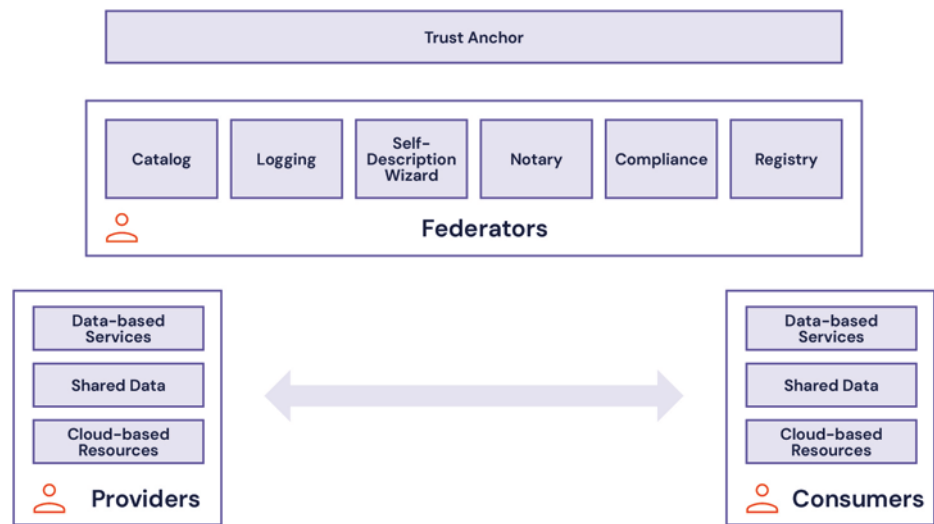
By democratizing these data-based controls, Sovity not only streamlines administrative overhead but also strengthens trust in federated data economies, empowering entities to participate securely.

Synergies between Gaia-X and IDSA

By aligning Gaia-X's framework with IDSA's standards, Europe builds a cloud model rooted in sovereignty and trust

Gaia-X and IDSA are not advancing in isolation; their integration amplifies their strengths and may achieve a real framework for data sovereignty. IDSA's technical architecture aligns with Gaia-X's federated model, creating a smooth fit between principles of self-sovereign storage and interoperable data exchange [3]. This synergy establishes a comprehensive ecosystem: Gaia-X provides the overarching framework and verification, while IDSA supplies the tools for secure data flows (see **Figure 5** for a high level approach to their integration) [20]. Thus, Gaia-X is a framework that enables cloud and data service providers to interoperate under shared technical and ethical standards, promoting transparency and facilitating data sovereignty. Unlike other cloud services, Gaia-X does not provide cloud applications, infrastructure, or storage. Instead, it establishes a model for trust and interoperability, ensuring that users can securely share and manage data across multiple providers without vendor lock-in [21]. IDSA, in turn, manages data by providing standards for creating secure and sovereign "data spaces" where organizations can share information while retaining full control over their data [20].

Figure 5: Gaia-X and IDSA: a high level approach



Future architectures that integrate both frameworks will consist of a set of participants having several roles. On one hand, data and data services providers will share their specific offers through catalogs, providing services according to detailed service-level agreements and following the framework’s rules for data security and sovereignty. On the other hand, service consumers will claim these offers, getting the benefits of data services for their businesses. Finally, federators will participate in the building of operational support services to ensure a smooth link between interested peers, while maintaining secure data exchanges [22].

Challenges and future outcomes

Building a European-based data sovereignty infrastructure and mindset encompasses a set of complex questions with answers that are not always straightforward. The first major challenge is breaking Europe’s dependency on external hyperscalers. Currently, Europe lacks the infrastructure and service maturity to compete with platform-based offerings from the United States and China [23]. Existing services fall short of market needs, and the transition from closed platforms to open systems is hindered by the absence of interoperability standards, making agile migration nearly impossible. To achieve long-term data sovereignty, Europe must invest in a secure, interoperable, and federated infrastructure [24].

Operational and governance challenges are equally critical. In a pan-European ecosystem of federated platforms, robust support systems will be essential to ensure the reliability and efficiency of distributed infrastructure. Governance will be complex, shaped by the diverse agendas and priorities of individual member states. To foster innovation and prevent the rise of new monopolies, it is vital to build inclusive data ecosystems that enable seamless collaboration across sectors, sizes, and national borders [25].

Another key concern in building the ecosystem is ensuring active engagement from all communities. On one side are the data and service providers, and on the other, the consumers of these offerings. A third group, the federators, delivers support services within this environment [22]. All participants aim to benefit from

a collaborative ecosystem that offers a seamless experience from service discovery to contract completion. Delivering a frictionless journey will be essential to attract and retain engagement while driving business growth across Europe. This vision relies on a federated, secure, interoperable data infrastructure aligned with European regulations, moving away from closed platforms that monopolize control and limit market openness [6].

Gaia-X and IDSA are steadily advancing as strategic responses to current data sovereignty challenges. Gaia-X has already demonstrated the viability of its framework through real-world use cases in healthcare, manufacturing, and public administration [26]. A standout example is Catena-X, an automotive pilot project under Gaia-X that enables secure data sharing across supply chains [27]. Meanwhile, IDSA's updated reference architecture (DS-RAM 5.0) is gaining traction among telecom operators, proving its capability to support cross-border data exchange with strong privacy safeguards [17].

However, the long-term success of both initiatives hinges on their ability to foster cross-sector collaboration and widespread adoption. The EU's commitment, evidenced by the €15.3 billion allocated through Horizon Europe for digital transformation between 2021 and 2027 [28], underscores the region's dedication to this vision. If Gaia-X and IDSA can establish clear compliance standards, robust governance, and inclusive participation models, they have the potential to reshape Europe's digital landscape, enhancing sovereignty and setting a global benchmark for ethical, secure, and innovative data ecosystems.

With €15.3 billion fueling Horizon Europe, Gaia-X and IDSA stand poised to reshape sovereignty into a global benchmark



Conclusion

Europe is positioning itself at the forefront of data space innovation. Achieving digital sovereignty depends on building a federated, interoperable, and trustworthy data infrastructure. By moving away from proprietary, centralized platforms, Europe can transform data from a vulnerability into a strategic asset, empowering businesses, governments, and citizens alike.

More than European initiatives, Gaia-X and IDSA are global benchmarks. These frameworks balance innovation with accountability in cloud-based services, ensuring that data owners retain full control. This paves the way for fair, dynamic digital ecosystems that fuel entrepreneurship and growth. Global adoption is essential to standardize interoperability, drive responsible progress, and deliver economic and societal value.

Altice Labs is committed to shaping a secure digital future. In a world defined by geopolitical tensions and commercial rivalries, the company champions data sovereignty as the foundation of a safe and trustworthy digital society.

Acknowledgments

This research has been partially funded by the 6G-PATH project through the Smart Networks and Services Joint Undertaking (SNS JU) and co-funded by the European Union under the EU Horizon Europe research and innovation program (GA no. 101139172). The authors of this work would like to acknowledge the contributions of 6G-PATH project partners. This research was also partially funded by the ED-X Test Bed (11078), supported under the Portuguese Recovery and Resilience Plan (RRP). The authors gratefully acknowledge the contributions and support of the ED-X Test Bed Consortium Partners.

References

- [1] A. Braud, G. Fromentoux, B. Radier, and O. Le Grand, "The Road to European Digital Sovereignty with Gaia-X and IDSA," *IEEE Netw.*, vol. 35, no. 2, pp. 4–5, Mar./Apr. 2021, doi: 10.1109/MNET.2021.9387709.
- [2] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach," *The Guardian*, Mar. 17, 2018. [Online]. Available: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- [3] AIT Austrian Institute of Technology, "Digital Sovereignty Review," AIT, Vienna, Austria, Rep. EFA23-AIT-LAB, Jan. 2024.
- [4] European Commission, "European Health Data Space: A step towards a more digital and sustainable Union," *EC Press Release*, May 2022. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2711

- [5] B. Otto, "GAIA-X and IDS," Zenodo, Jan. 01, 2021, doi:10.5281/zenodo.5675897.
- [6] B. Otto, "A federated infrastructure for European data spaces," *Commun. ACM*, vol. 65, pp. 44–45, 2022, doi: 10.1145/3512341.
- [7] J. Jussen, F. Möller, J. Schweihoff, A. Gieß, G. Giussani, and B. Otto, "Issues in inter-organizational data sharing: Findings from practice and research challenges," *Data Knowl. Eng.*, vol. 150, p. 102280, 2024, doi: 10.1016/j.datak.2024.102280.
- [8] Gaia-X. "Gaia-X Framework." Gaia-X. Accessed: Oct. 28, 2025. [Online]. Available: <https://gaia-x.eu/gaia-x-framework/>
- [9] Federal Ministry for Economic Affairs and Energy, "GAIA-X: The European project kicks off the next phase," BMWi, Berlin, Germany, 2020.
- [10] A. Bott et al., "Challenges in implementing Gaia-X for industrial applications," *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 120, pp. 163–168, 2025, doi: 10.1515/zwf-2025-1023.
- [11] B. Gerlach, H. Hesse, S. Fischer, and M. Leucker, "Gaia-X-Med: Experiences with building dataspaces for medical applications," *Future Int.*, vol. 16, p. 463, 2024, doi: 10.3390/fi16120463.
- [12] R. Adler-Nissen and K. Eggeling, "The discursive struggle for digital sovereignty: Security, economy, rights and the cloud project Gaia-X," *JCMS: J. Common Market Stud.*, 2024, doi: 10.1111/jcms.13594.
- [13] Gaia-X Wizard. "Home Page." Gaia-X Wizard. Accessed: Oct. 28, 2025. [Online]. Available: <https://wizard.lab.gaia-x.eu/>
- [14] J. Marino, L. Camiciotti, F. Cheinasso, A. Olivero, and F. Risso, "Enabling compute and data sovereignty with infrastructure-level data spaces," in *Proc. 3rd Eclipse Secur., AI, Archit. Model. Conf. Cloud to Edge Continuum*, 2023, doi: 10.1145/3624486.3624509.
- [15] H. Pettenpohl, M. Spiekermann, and J. R. Both, "International Data Spaces in a nutshell," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, B. Otto, M. ten Hompel, and S. Wrobel, Eds., Cham, Switzerland, Springer, 2022, pp. 29–40, doi: 10.1007/978-3-030-93975-5_3.
- [16] B. Arnold, C. Lange, C. Gillmann, and S. Decker, "A service architecture for dataspaces," 2025, arXiv:2507.07979, doi: 10.48550/arxiv.2507.07979.
- [17] International Data Spaces Association, "Reference Architecture Model," IDSA, 2025. [Online]. Available: <https://internationaldataspaces.org/offers/reference-architecture/>
- [18] International Data Spaces Association (IDSA). "Advancing interoperability: the Dataspace Protocol." IDSA. Accessed: Oct. 28, 2025. [Online]. Available: <https://internationaldataspaces.org/offers/dataspace-protocol/>
- [19] P. Wittenburg, G. Strawn, C. Bianchi, D. Broeder, and U. Schwardmann, "Layered dataspaces in GIDS," *Open Conf. Proc.*, 2025, doi: 10.52825/ocp.v5i.1194.

- [20] J. Pampus, B. Jahnke, and R. Quensel, "Evolving data space technologies: Lessons learned from an IDS connector reference implementation," in *Int. Conf. Adv. Inf. Syst. Eng.*, Cham, Switzerland: Springer, 2022, pp. 366–381, doi: 10.1007/978-3-031-19762-8_27.
- [21] H. Tardieu, "Role of Gaia-X in the European Data Space Ecosystem," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, B. Otto, M. ten Hompel, and S. Wrobel, Eds., Cham, Switzerland, Springer, 2022, pp. 41–59, doi: 10.1007/978-3-030-93975-5_4.
- [22] S. Meneguzzo, A. Favenza, V. Gatteschi, and C. Schifanella, "Integrating a DLT-based data marketplace with IDSA for a unified energy dataspace: Towards silo-free energy data exchange within GAIA-X," in *Proc. 5th Conf. Blockchain Res. Appl. Innov. Netw. Serv. (BRAINS)*, 2023, pp. 1–2, doi: 10.1109/brains59668.2023.10316796.
- [23] E. Celeste, "Digital sovereignty in the EU: Challenges and future perspectives," in *Data Protection Beyond Borders*, 2020, doi: 10.5040/9781509940691.ch-013.
- [24] D. Firdausy, P. De Alencar Silva, M. Sinderen, and M. Iacob, "Towards a reference enterprise architecture to enforce digital sovereignty in International Data Spaces," in *2022 IEEE 24th Conf. Bus. Inform. (CBI)*, 2022, vol. 01, pp. 117–125, doi: 10.1109/cbi54897.2022.00020.
- [25] V. Damjanovic-Behrendt and W. Behrendt, "Governance mechanisms for federated digital platform ecosystems," 2020, doi: 10.5281/zenodo.3951097.
- [26] Gaia-X Hub Germany, "Funding Projects," 2025. [Online]. Available: <https://gaia-x-hub.de/en/funding-projects/>
- [27] Catena-X, "Business Areas in the Catena-X ecosystem," 2025. Accessed: Oct. 28, 2025. [Online]. Available: <https://catena-x.net/overview-use-case-cluster/>
- [28] Horizon Europe, "Horizon Europe - funding digital technology and research," 2025. Accessed: Oct. 28, 2025. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/activities/horizon-europe-funding-digital>

ACRONYMS AND TERMS

2	2G	Second Generation	B	BBF	Broadband Forum
3	3GPP	3 rd Generation Partnership Project		BER	Bit Error Ratio
4	4G	Fourth Generation		BPD	Balanced Photodetector
5	5G	Fifth Generation		BSS	Business Support System
6	6G	Sixth Generation	C	CAPEX	Capital Expenditure
A	A2A	Agent-to-Agent Protocol		CD	Chromatic Dispersion
	ACL	Access List		CDN	Content Delivery Network
	ACP	Agent Communication Protocol		CES	Coexistence Scenarios
	ADC	Analog-to-Digital Converter		CEx	Coexistence Elements
	ADK	Agent Development Kit		CF	Customer Feedback
	ADR	Autonomous Delivery Robot		C-ITS	Cooperative Intelligent Transport Systems
	ADS	Autonomous Delivery Service		CLI	Command Line Interface
	AEC	Adjusted Explanation Coverage		CNSA	Commercial National Security Algorithm
	AI	Artificial Intelligence		COTS	Commercial Off-the-Shelf
	AIIA	Altice International Innovation Award		CPE	Customer Premises Equipment
	AMF	Access and Mobility Management Function		C-RAN	Centralized Radio Access Network
	AN	Autonomous Networks		CRUD	Create, Read, Update, Delete
	ANLET	Autonomous Network Level Evaluation Tool		CSC	Communications Service Consumer
	APC	Analog Polarization Control		CSP	Communications Service Provider
	API	Application Programming Interface		CT	Channel Termination
	AR	Augmented Reality		CU	Centralized Unit
	ARM	Advanced RISC Machines		CxO	Cognitive eXperience Operation
	ARPU	Average Revenue Per User	D	DAC	Digital-to-Analog Converter
	ASA	Average Speed of Answer		DAG	Directed Acyclic Graph
	ASR	Automatic Speech Recognition		DBA	Dynamic Bandwidth Assignment
	AURA	AI for Understanding and Resolving Anomalies		DCI	Data Center Interconnect
				DCPC	Digital Chromatic Dispersion Pre-Compensation
				DD	Direct Detection
				DNA	Digital Network Architecture

	IPSec	Internet Protocol Security		MNO	Mobile Network Operator
	IPTV	Internet Protocol Television		MnS	Management Service
	IRTF	Internet Research Task Force		MPI	Multipath Interference
	ISDA	International Data Spaces Association		MR	Mixed Reality
	IT	Information Technology		MRI	Magnetic Resonance Imaging
	IVR	Interactive Voice Response		MZM	Mach-Zehnder Modulator
<hr/>					
K	KPI	Key Performance Indicator		N	NAT Network Address Translation
	KPM	Key Performance Measurement			NC No-Code
<hr/>					
L	L3VPN	Layer 3 Virtual Private Networks			NEP Network Equipment Provider
	LC	Low-Code			NFV Network Functions Virtualization
	LightGBM	Light Gradient Boosting Machine		NG-PON2	Next-Generation Passive Optical Networks
	LLM	Large Language Model			NHP Neutral Host Provider
	LMF	Location Management Function			NI Net Impact
	LO	Local Oscillator			NIST National Institute of Standards and Technology
	LOF	Loss of Frame			NLP Natural Language Processor
	LOS	Loss of Signal			NLU Natural Language Understanding
<hr/>					
M	M&C	Management and Control			NMS Network Management System
	MAS	Multi-Agent Societies			NOC Network Operations Center
	MCP	Model Context Protocol			NPN Non-Public Networks
	MDU	Multi-Dwelling Units			NRPPa New Radio Positioning Protocol A
	MEC	Mobile Edge Computing		NRZ-OOK	Non-Return-to-Zero On-Off Keying
	MEF	Misclassification Explanation Framework			NSD Network Service Descriptor
	MILP	Mixed Integer Linear Programming			NWDAF Network Data Analytics Function
	ML	Machine Learning		<hr/>	
	ML-DSA	Module-Lattice Digital Signature Algorithm		O	OBU On-Board Unit
	ML-KEM	Module-Lattice Key Encapsulation Mechanism			ODA Open Digital Architecture
	mMTC	massive Machine-Type Communications			ODB Optical Duobinary
					ODN Optical Distribution Network

	ODP	Optical Distribution Points		RAN	Radio Access Network
	OI	Open Innovation		RC	RAN Controller
	OIF	Optical Internetworking Forum		REST	Representational State Transfer
	OLT	Optical Line Terminal		RFS	Recursive Feature Selection
	ONOS	Open Network Operating System		RIC	RAN Intelligent Controller
	ONT	Optical Network Terminal		RISC-V	Reduced Instruction Set Computing – V
	ONU	Optical Network Unit		ROI	Return on Investment
	OPEX	Operating Expenditure		RRP	Recovery and Resilience Plan
	OQS	Open Quantum Safe		RSA	Rivest–Shamir–Adleman
	O-RAN	Open Radio Access Network		RU	Radio Unit
	OSS	Operations Support System			
	OTT	Over-the-Top			
<hr/>					
P	PAM4	Pulse Amplitude Modulation with Four Levels		SA	Standalone
	PoC	Proof of Concept		SBMA	Service Based Management Architecture
	PoIMD	Polarization-Mode Dispersion		SDK	Software Development Kit
	PON	Passive Optical Network		SDN	Software Defined Networking
	PPDR	Public Protection and Disaster Recovery		SDR	Software Defined Radio
	PQC	Post-Quantum Cryptography		SDU	Single-Dwelling Units
	PQCA	Post-Quantum Cryptography Alliance		SD-WAN	Software-Defined Wide Area Network
	PTQCI	Portuguese Quantum Communications Infrastructure		SHAP	SHapley Additive exPlanations
<hr/>					
Q	Q	Quadrature		SIM	Subscriber Identity Module
	QAM	Quadrature Amplitude Modulation		SLA	Service Level Agreement
	QCI	Quantum Communication Infrastructure		SLM	Small Language Model
	QKD	Quantum Key Distribution		SM	Service Model
	QoS	Quality of Service		SMEs	Small and Medium-Sized Enterprises
	QPSK	Quadrature Phase Shift Keying		SMF	Session Management Function
<hr/>					
R	R&D	Research & Development		SMO	Service Management and Orchestration
				SMS	Short Message Service
				SNDL	Store Now, Decrypt Later
				SNMP	Simple Network Management Protocol
				SNR	Signal-to-Noise Ratio
				SNS	Smart Network and Services

	SNS JU	Smart Networks and Services Joint Undertaking		USB-C	Universal Serial Bus Type-C
	SOA	Semiconductor Optical Amplifiers		UWB	Ultra-Wideband
	SoC	System on Chip		UX	User Experience
	SPM	Self-Phase Modulation			
	SP-QPSK	Single-Polarization Quadrature Phase-Shift Keying			
	SRS	Software Radio Systems			
	SSB	Single Sideband			
	SSH	Secure Shell			
	SSMF	Standard Single-Mode Fiber			
	STB	Set-Top Box			
T	TAI	Trustworthy AI		V	V2N Vehicle-to-Network
	TDM	Time Division Multiplexing			V2V Vehicle-to-Vehicle
	TDMA	Time Division Multiple Access			VHSP Very High-Speed PON
	TDM-PON	Time Division Multiplexing PON			VoIP Voice over Internet Protocol
	TINA	Telecommunications Information Networking Architecture			VPN Virtual Private Network
	TIO	TM Forum Intent Ontology			VR Virtual Reality
	TLS	Transport Layer Security			Vssb Vestigial-Sideband Single-Carrier
	TreeSHAP	Tree SHapley Additive exPlanations			
	TRL	Technology Readiness Levels		W	WEF World Economic Forum
	TTD	Time-to-Detect			WM Wavelength Multiplexer
	TTFF	Time To First Fix			WMS Windless Media Server
	TTK	Trouble Ticket			
	TTS	Text-to-Speech		X	XAI Explainable AI
	TWDM-PON	Time and Wavelength Division Multiplexed PON			XGBoost eXtreme Gradient Boosting
					XGS-PON 10-Gigabit-Capable Symmetric PON
				Y	Y2Q Year to Quantum
					YAML YAML Ain't Markup Language
					YANG Yet Another Next Generation
				Z	ZTP Zero-Touch Provisioning
U	UI	User Interface			
	UL-TDoA	UpLink Time Difference of Arrival			
	UPF	User-Plane Functions			
	URLLC	Ultra Reliable Low-Latency Communications			

Have you missed previous editions of InnovAction?

Every year, Altice Labs publishes *InnovAction*, a curated publication that provides our customers, partners, and innovation ecosystem with the latest insights into our knowledge, expertise, and technological developments, showcasing the latest cutting-edge technology in the market.

For a company that has a very clear investment strategy in telecommunications, advanced network and connectivity solutions, digital products and services, a publication like *InnovAction* is particularly relevant. More than a valuable marketing tool, it is a method to share the most relevant technological developments and knowledge across the tech community, reinforcing our unwavering commitment to technological innovation.

If you missed previous editions of this annual publication, scan the QR code below to access all the *InnovAction* magazines.



Acknowledgments

Year after year, *InnovAction* is the result of powerful teamwork. Several individuals from Altice Labs get involved and, along with external partners, contribute their knowledge and time to grant this magazine's quality and ensure each edition has a diversified collection of articles.

Thus, we would like to thank all authors for their valuable input and for sharing the knowledge gained from their research and work.

We also want to express our gratitude to all technical and editorial reviewers, who did a meticulous job and contributed to improving the quality and excellence of all articles and, therefore, of the publication as a whole.

Finally, a special word of acknowledgment to our designers for the excellent job in creating an appealing and coherent design across the entire magazine.

Edition

December 2025



altice
labs



www.alticelabs.com