



# InnovAction

#9 | December 2024

### Editors-in-Chief

João Paulo Firmeza  
Ana Margarida Almeida  
Daniela Rei  
Pedro Carvalho

### Editorial Board

Ana Margarida Almeida  
Daniela Rei  
Dulce Teles  
Fausto de Carvalho  
Filipe Cabral Pinto  
Francisco Fontes  
Luís Cortesão  
Nuno Silva  
Pedro Carvalho  
Pedro Neves  
Ricardo Jorge Ferreira  
Vitor Marques

### Art Coordinator and Graphic Edition

Cátia Pinto  
Carlos Figueiredo

### Manuscript editing and proofreading

Miguel Neto

### Reviewers

Bruno Cabral  
Filipe Cabral Pinto  
Francisco Fontes  
Gonçalo Machado  
Hélder Alves  
João Nifo  
Jorge Gonçalves  
José Machado  
José Salgado  
Luís Azevedo  
Luís Castro  
Luís Cortesão  
Luís Reis  
Miguel Biscaia  
Miguel Mesquita  
Nuno Silva  
Paulo Mão Cheia  
Pedro Carvalho  
Rui Calé  
Victor Marques  
Vitor Mirones

### Title

Innovation  
Altice Labs journal

### Edition

2023

### Design

Altice Labs

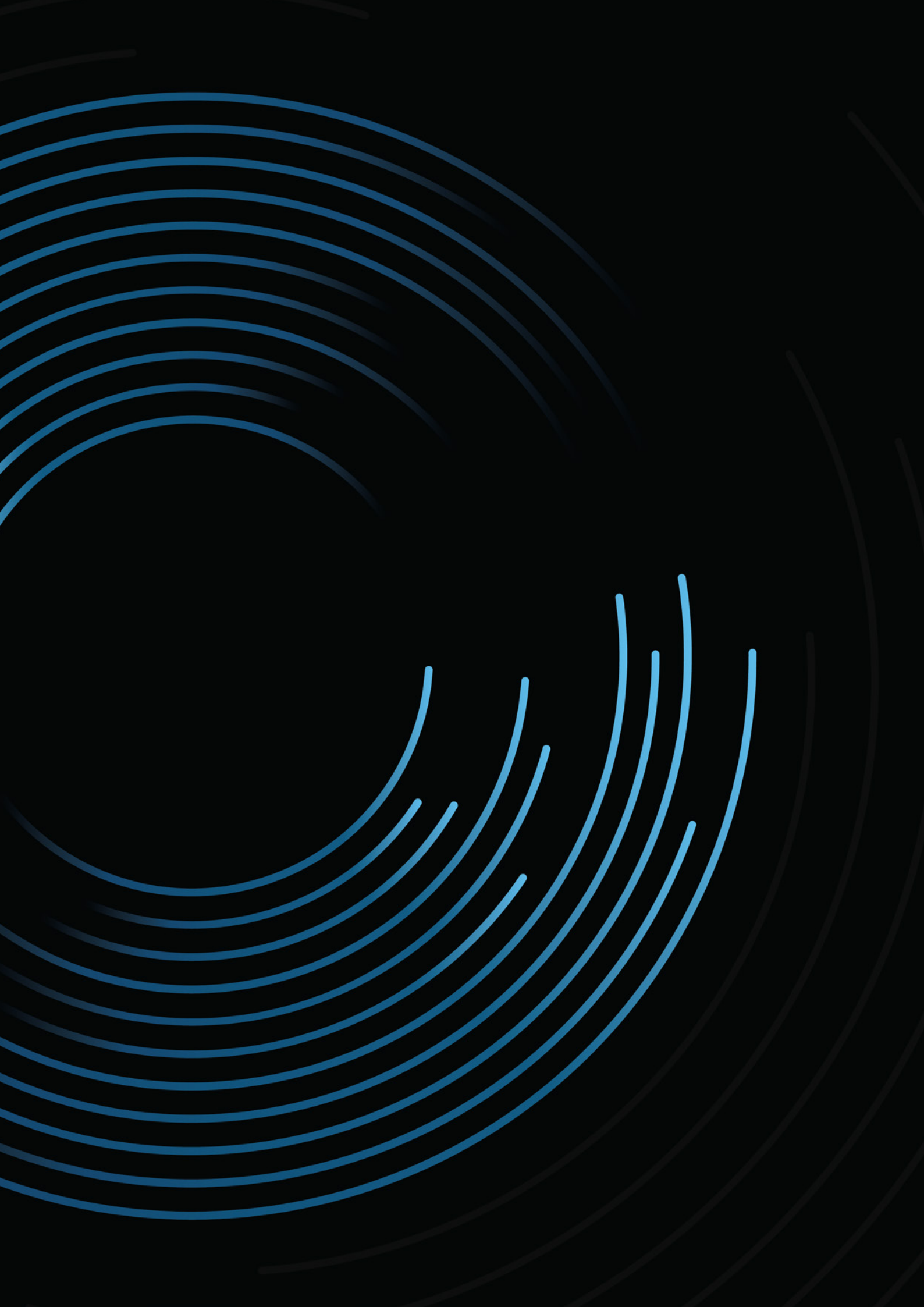
### Legal deposit

512167/23

### ISSN | DOI

2975-8467 | 10.57650/INNOVATION







# Opening note

As we turn the page to a new chapter, I am pleased to introduce the latest edition of InnovAction. This annual magazine not only highlights the remarkable achievements and advancements made within Altice Labs but also reflects the shared commitment of our community to shaping the future of technology while improving people's lives.

On a year defined by dynamic change, we find ourselves navigating the complexities of evolving global challenges and the rapid transformation of the technological landscape. Through our collective efforts, we've explored how innovation can connect people, transform industries, and create solutions that matter.

The stories within these pages celebrate the spirit of collaboration and the resilience of our teams. They capture the essence of what makes Altice Labs unique: our ability to merge creativity with technical excellence while converting knowledge into action.

This edition takes a deep dive into groundbreaking advancements, such as a multichannel high-speed PON solution coexisting with legacy technologies, the evolution of 5G Radio Unit architecture, and the impact of Passive Optical Network technology on quantum signals. We explore the future of disaggregated and virtualized access networks, the sharing of resources among virtual operators, and the potential of blockchain in federated network ecosystems.

Sustainability also takes center stage with smart energy tools and AI-driven automation redefining efficiency. Our focus on DevSecOps practices reflects our commitment to embedding security in every aspect of development, while new methods for automatic FTTH network design showcase the growing role of AI in network planning.

We trust this publication will provide valuable insights and inspiration, offering a glimpse into how we're preparing for the future—by fostering talent, engaging with the international innovation community, and staying true to our mission of being a global catalyst for change.

We invite you to explore this InnovAction 2024 as a window into the future we are building together.

Thank you for joining us on this voyage of innovation and discovery.

**João Paulo Firmeza**  
General Manager of Altice Labs  
[PFirmeza@alticelabs.com](mailto:PFirmeza@alticelabs.com)



# Editorial note

As 2024 comes to an end, we acknowledge the adverse global context while recognizing our responsibility to be agents of change and hope. At Altice Labs, we will continue to blaze a pioneering trail in the telco industry, striving to develop cutting-edge products that connect people and revolutionize society. Our technology empowers us to overcome obstacles, adapt seamlessly to new contexts, and respond effectively to emerging challenges.

In this journey, InnovAction stands not only as a means of sharing valuable information but also as a reflection of our efforts in tackling pressing topics ranging from connectivity and cybersecurity to artificial intelligence and energy sustainability.

In a world where connectivity demands increase exponentially, communication service providers must be able to adapt and grow in tandem to continue delivering the best possible service to their clients. As a core element for digital transformation, we have five engaging themes in this area:

- **Evolving Passive Optical Networks (PONs)** to accommodate multiple wavelength channels were object of a study that shows good viability perspectives for the next generation of multi-wavelength PONs;
- **The significant role of FPGA devices** in developing Altice Labs 5G Radio Unit (RU) products was highlighted as it enriches the mobile network portfolio;
- **Altice Labs' solution to leverage the potential of cloud-based environments** and the

micro-services principles in access networks was explored as it allows us to quickly adapt the devices to new service models and features, while also providing new interoperability points, adding the potential for decoupling the hardware vendor from the software's;

- **The Fixed Access Network Sharing (FANS) paradigm** emerged as the structure that permits usage optimization of the underlying network infrastructure, now shared by different Virtual Network Operators (VNO);
- **Improving the fiber network inventory** was one of the main concerns tackled in this issue, as well as the necessity of giving autonomy to field technicians.

Alongside connectivity, cybersecurity becomes of paramount importance. More than ever, it is essential to consider security in all its angles to create a safer digital society. The generalized concern with this field is reflected in three of our articles' topics:

- **The DevSecOps methodology** is presented as a crucial strategy to minimize potential security risks in the software development life cycle, not without acknowledging its requirement for a strong organizational commitment;
- **Blockchain technology** is highlighted as it uses its immutability, transparency, and decentralization to address the shortcomings of conventional security systems, reducing the risks of



data breaches, identity theft, and unauthorized access to a resource or service;

- **The impact of PON technology on Discrete Variable Quantum Key Distribution (DV-QKD)** is investigated, underscoring how Quantum Communication is able to mitigate today's challenges if capable to work with fiber-based network solutions.

In this edition, the subject of Artificial Intelligence (AI) applied to different domains is a key topic. We have prepared a total of five articles addressing AI themes:

- **Genetic algorithms** are proved to be an effective method for addressing the complexities involved in FTTH GPON network design, drastically reducing the time of the process;
- **The impact of AI in software engineering** is the subject of a reflection where AI tools are acknowledged for allowing users to focus on the more complex and creative aspects of the development process;
- **Neural networks applied to enhance predictive models**, focusing on anticipating top-up propensity and account balances before top-up in prepaid mobile services, reaching promising results;
- **GenAI makes hyper-personalized experiences** possible, reshaping customer experience and how businesses run, driving innovation across industries that unlocks new opportunities for human connection;

- **A machine learning-based algorithm that detects fires** in their early stages while being lightweight enough to be used in a low-power device, such as a single-board computer, is presented.

In recent years, energy sustainability has become a primary concern for Communications Service Providers. This focus has inspired us to prepare an article on **Dynamic Energy Management**, exploring the technologies and software tools that can optimize the use of electric vehicle batteries, energy communities, and peer-to-peer (P2P) applications for buying and selling energy.

All in all, we delved into complex and stimulating topics, sharing new information and knowledge in an engaging and captivating manner. Thanks to the meticulous efforts of our passionate collaborators and partners, we hope to have crafted a magazine that sparks your creativity and inspires you to take action and innovate.

May this magazine fuel your inspiration!

**Ana Margarida Almeida**

[ana-margarida-almeida@alticelabs.com](mailto:ana-margarida-almeida@alticelabs.com)

**Pedro Carvalho**

[PCarv@alticelabs.com](mailto:PCarv@alticelabs.com)

**Filipe Cabral Pinto**

[filipe-c-pinto@alticelabs.com](mailto:filipe-c-pinto@alticelabs.com)

# Contents

## 01

Multichannel 50 Gb/s per channel IM/DD O-Band 100GHz spacing high speed PON coexisting with legacy technologies

(p. 10)

## 02

5G Radio Unit based on an RFSoc FPGA

(p. 18)

## 03

The disaggregation and virtualization of the access network

(p. 34)

## 04

Sharing network resources between virtual operators in access networks

(p. 46)

## 05

Overbooking Maneuver solution for OTBs

(p. 60)

## 06

Introducing DevSecOps practices in Altice Labs

(p. 70)

## 07

Network Federated Ecosystem with Blockchain

(p. 84)

## 08

Impact of Passive Optical Network technology on polarization encoded Discrete Variable Quantum Key Distribution

(p. 94)

## 09

Automatic FTTH network design

(p. 106)



## 10

AI-driven development:  
what may the future  
bring?

*(p. 118)*

## 11

From development to  
deployment: neural  
network implementation  
in telco using PySpark

*(p. 134)*

## 12

Machine-learning aided  
fire detection systems –  
an opportunity for early  
response and improved  
safety

*(p. 150)*

## 13

How LLMs are reshaping  
customer experience:  
a vision of enhanced  
interactions and a  
brighter future

*(p. 164)*

## 14

Smart energy tools

*(p. 176)*

# 01

Multichannel 50 Gb/s per channel IM/DD O-Band 100GHz spacing high speed PON coexisting with legacy technologies



# Multichannel 50 Gb/s per channel IM/DD O-Band 100GHz spacing high speed PON coexisting with legacy technologies

We demonstrate a multiwavelength intensity modulation of 50 Gb/s per channel PON. It propagated through 20 km of fiber without any electrical or optical compensation, with negligible penalties. The impact on the coexisting technologies was <0.5 dB.

## Authors

**Ana Tavares, PICadvanced and Department of Electronics, Telecommunications and Informatics (DETI), University of Aveiro, [ana@picadvanced.com](mailto:ana@picadvanced.com)**

**Cláudio Rodrigues, Altice Labs, [claudio-e-rodrigues@alticelabs.com](mailto:claudio-e-rodrigues@alticelabs.com)**

**Miguel Masanas, Department of Signal Theory and Communications, Polytechnic University of Catalonia, [miquel.masanas@upc.edu](mailto:miquel.masanas@upc.edu)**

**João Santos, PICadvanced and Department of Electronics, Telecommunications and Informatics (DETI), University of Aveiro, [joao.santos@picadvanced.com](mailto:joao.santos@picadvanced.com)**

**Francisco Rodrigues, PICadvanced, [francisco@picadvanced.com](mailto:francisco@picadvanced.com)**

**Josep Prat, Department of Signal Theory and Communications, Polytechnic University of Catalonia, [josep.prat@upc.edu](mailto:josep.prat@upc.edu)**

**António Teixeira, PICadvanced and Department of Electronics, Telecommunications and Informatics (DETI), University of Aveiro, [teixeira@picadvanced.com](mailto:teixeira@picadvanced.com)**

## Keywords

High Speed PON; 50G-PON; 50G TWDM-PON



## Introduction

Passive Optical Networks (PONs) play a crucial role in modern telecommunications and networks. They are the vehicle for the broad offering of high-speed internet access, cost efficiency, and scalability. Standardization committees, together with operators, are aiming to achieve several hundred Gb/s per port, using the existing fiber network and available wavelength bands [1], [2], [3]. For this next stage, the trend is O-band due to its low attenuation and chromatic dispersion, which is critical for high data rates [4], [5].

Following the efforts in the specification of the NG-PON2, for PON to accommodate multiple wavelength channels, it is required to assess conventional wavelength division multiplexing (WDM) metrics. These metrics include the wavelength multiplexer (WM) effect on signal integrity, maximum spectral excursion (MSE), maximum tuning error (MTE), and the impact of multiplexing several channels [6], [7].

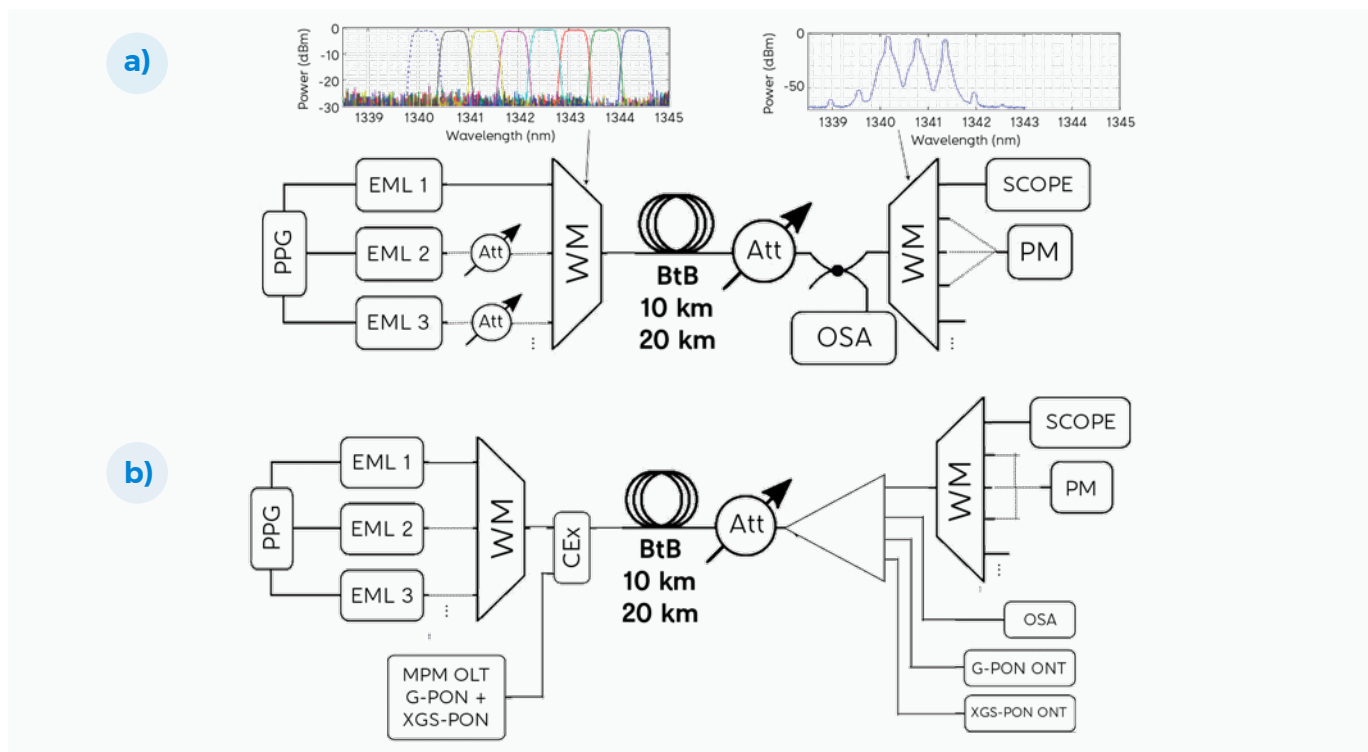
In this work, we demonstrate multiwavelength 50 G-PON with 100 GHz channel spacing, where we focus on evaluating and testing the performance of off-the-shelf external modulated lasers (EML) and TOCAN-based O-band 50 Gb/s transmitters within a Time and Wavelength Division Multiplexing (TWDM) scheme. First, we examine the modulation quality by testing the effect of (de) multiplexing operations through a WM, studying the eye extinction ratio (ER) against signal-multiplexer frequency detuning ( $\Delta f$ ). Secondly, we investigate the effects of four-wave mixing (FWM) and inter-channel interference (ISI) by introducing up to two additional channels and studying the ER penalty as a function of channel spacing. Finally, we also tested the impact of these channels on a network with current PON technologies.

## Experimental set-up

The EML lasers used in this work have an integrated Semiconductor Optical Amplifier (SOA) to boost the power and were modulated with  $2^{15}-1$  PRBS patterns at 49.7664 Gb/s through a 40 GHz bias tee. The characterization of the signal was made through the ER offline analysis of the waveshape. We used three of these lasers at 1340.15, 1340.75, and 1341.35 nm, modulated with double-sideband (DSB) (optical spectra as inset in **Figure 1a**), with 6.2 dBm of output modulated power. The channels are multiplexed by a WM (inset of the transfer function in **Figure 1a**), which consists of an 8-channel O-band 100 GHz ITU grid mux/demux with 60 GHz bandwidth (BW). The spectrum measurements were taken with an Optical Spectrum Analyzer (OSA) with 0.1 nm resolution. In all the experiments we have taken the back-to-back (BtB) measurements and the 10 and 20 km of standard single-mode fiber (SSMF). The received optical power (ROP) was adjusted using a variable optical attenuator (VOA).

After the WM demultiplexer, a 70 GHz real-time oscilloscope is used with a 59 GHz BW photodetector while a power meter (PM) monitors the cross talk power. In the scope, before eye plotting and analysis, a 0.75 symbol rate 4th order Bessel filter is applied as in ITU-T G.9804.3 standard [2]. EML1, at CH1 of the WM, is the reference probing channel in all setups.

We made the experiments in three phases: single channel characterization, multiple channels impact on the first channel, and, finally, the effect of this combination in other technologies. For the single channel characterization, only EML1 from **Figure 1a** was used (EML2 and 3 were off). In the multiple channel scenario, different configurations were used: EML1+EML2, EML1+EML3, and EML1+EML2+EML3. Finally, this last combination of EML1+EML2+EML3 was inserted into a PON scenario with current technologies (G-PON and XGS-PON) to evaluate its impact; as depicted in **Figure 1b**, the 50 Gb/s channels entered a Coexistence Element (CEX) together with the current PON technologies, sent to the network and then de-multiplexed.

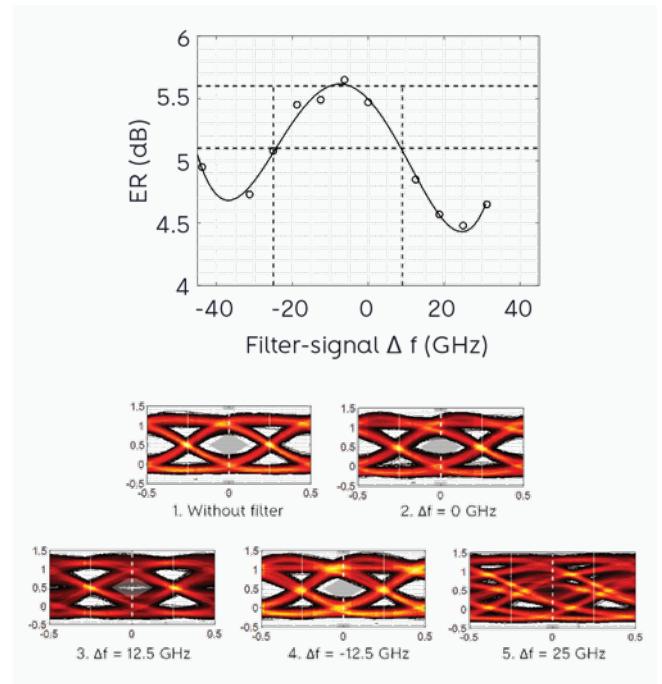


**FIGURE 1** – a) Schematic of the single and multichannel setup; b) multi-technology coexistence trial setup. WM - Wavelength mux with the profile in the inset on the left; OSA; PM; Pattern generator (PPG)

# Experimental results

## Single channel

The first test aimed at testing the robustness of the signal to the reduced bandwidth of each WM channel (60 GHz). In B2B we analyzed the signal ER, and the results are presented in **Figure 2**. The B2B eye diagram without WM, inset #1, displays an ER of 5.9 dB, while, in the presence of the WM, at nominal 0 GHz of detuning (inset #2), the eye shows an ER of 5.5 dB. The signal was shifted from -40 to 40 GHz, and it gets degraded as we detune due to signal distortion. However, it improves as we get to the point where it gets into optical single-sideband operation. The remaining insets are given as a visual reference of the signal performance. From the results, and considering a maximum tolerance of 0.5 dB in the signal ER, one can define a maximum spectral excursion for this signal of around 34 GHz, which is rather compatible with the existing control systems used in NGPON2.



**FIGURE 2** – Impact of the detuning of a single 50Gb/s signal with respect to the central wavelength of WM channel.  $\Delta f$  is the difference between the WM design frequency



## Multichannel

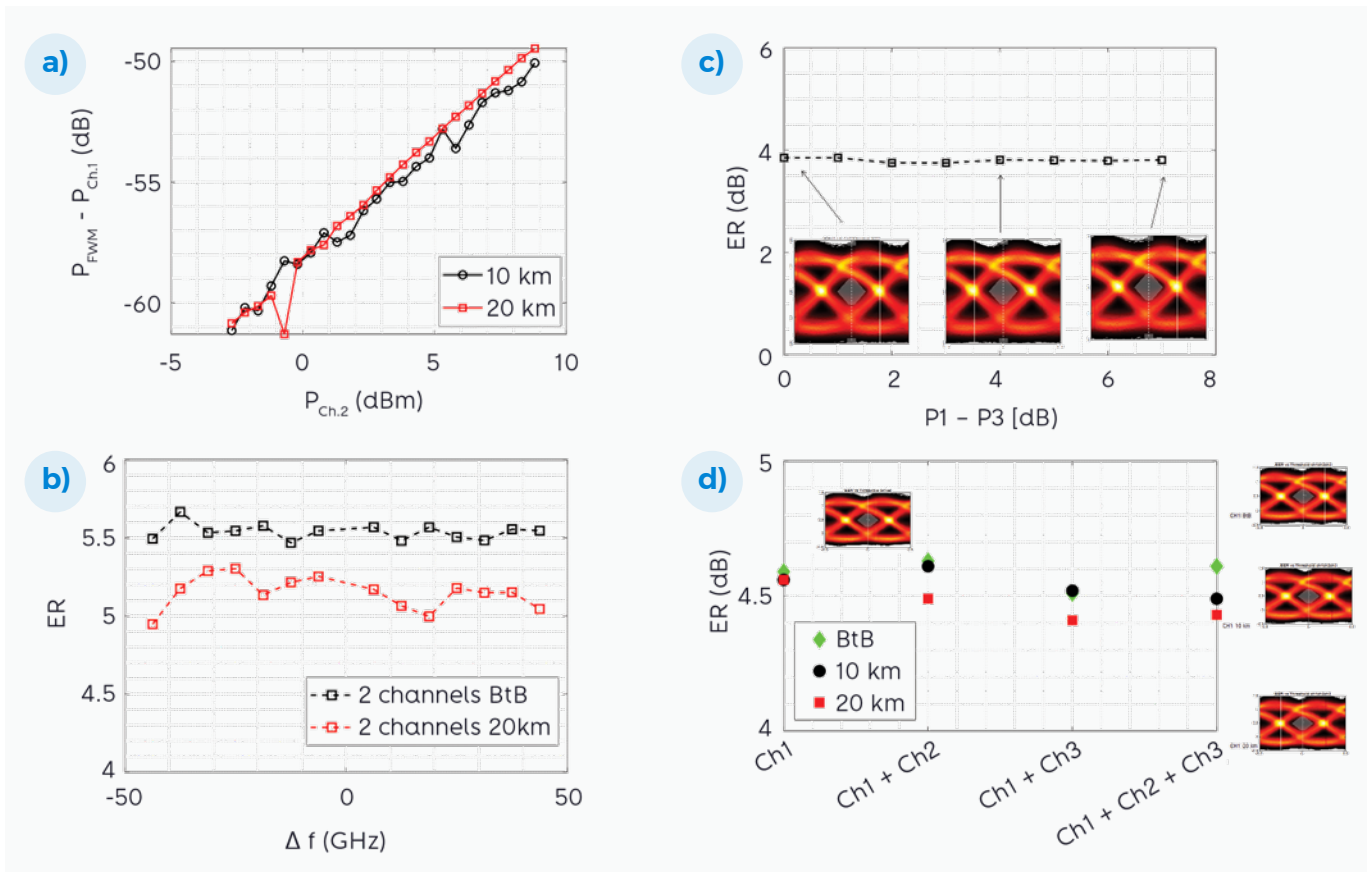
In this section, and for all the multichannel analysis, channel 1 was used as a probing channel. We used the combinations of the three available EMLs to i) observe the impact of cross talk and FWM with two or three channels; ii) test the impact of using the 100 GHz spacing when compared to the 200 GHz, using two channels CH1+CH2 and CH1+Ch3 (without CH2).

For the i) case, CH1+CH2 were used with fiber, and the FWM power was evaluated by monitoring the optical power on the third WM output while sweeping CH2 optical power. The results, displayed in **Figure 3a**), show that FWM power is very residual. This setup was also used to measure the impact of the linear cross talk between channels; to do this, the CH2 wavelength was varied inside the filter bandwidth, and the impact on CH1 was measured.

Results are shown in **Figure 3b**), where it can be seen that the impact was very small, with some of the minor variations also coming from the temperature tuning of the lasers.

Still maintaining the same setup but adding a third laser at CH3, resulting in CH1+CH2+CH3 with fiber, the penalty of the FWM on CH1 was also studied. To do so, CH3 optical power was varied, and the impact on CH1 was evaluated through ER measurements. A negligible impact was observed, as can be seen in **Figure 3c**).

Regarding case ii), CH1 ER was measured in the presence of CH2 only (CH1+CH2), CH3 only (CH1+CH3), and both (CH1+CH2+CH3), in BtB, 10 and 20 km of fiber. **Figure 3d**) shows the results, together with a single channel (CH1 only), where can be seen that CH1 ER is little affected by the presence of the other channels (CH1 ER variation is less than 0.3 dB).



**FIGURE 3** – Results obtained with the multichannel scenario. a) and b) Two channels; c) Three channels; d) Combination of one, two, and three channels

## Coexistence with legacy technologies

A CEx was added to the scenario in order to (de)multiplex the current PON technologies with the multiwavelength 50 Gb/s PON channels. The sensitivity and overload of an XGS-PON and G-PON OLT, equipped with an MPM C+ Class, were measured according to ITU-T G.984.5 and 2. These measurements were conducted in back-to-back configurations and 20 km of SSMF, as shown in the setup schematic in **Figure 1b**). The impact of the introduction of the new PON signals was analyzed, and there was a penalty of less than 0.5 dB in the two technologies, showing good compatibility and desired interoperability of PON technologies, including the novel proposed in this article 50G TWDM PON.

## Conclusions

We have demonstrated a multiwavelength IM/DD 50 Gb/s per channel system without any optical or electrical compensation at O-band. The impacts of the linear and nonlinear effects on the performance of such system were drafted, showing good viability perspectives to be presented as an evolutive path for the next generation of multiwavelength PONs. The study was made for the downstream path in order to observe the fundamental limits of the proposed technology. Burst mode will be studied in the upcoming work.

A coexistence scenario was also considered, and no impact was observed on the OLT sensitivity and overload, as well as on G-PON and XGS-PON terminals with the introduction of the proposed 50 Gb/s multiwavelength system. 🌐



## References

- [1] Ed Harstead, D. van Veen, V. Houtsma, and P. Dom, "Technology Roadmap for Time-Division Multiplexed Passive Optical Networks (TDM PONs)", *J. Lightw. Technol.*, vol. 37, no. 2, Jan., 2019, doi: 10.1109/JLT.2018.2881933.
- [2] 50-Gigabit-Capable Passive Optical Networks (50G-PON): Physical Media Dependent (PMD) Layer Specification, Rec. ITU-T G.9804.3, International Telecommunications Union, Geneva, Switzerland, 2021.
- [3] IEEE P802.3ca 50G-EPON Task Force: Physical Layer Specifications and Management Parameters for 25Gb/s and 50Gb/s Passive Optical Networks, IEEE Std 802.3ca-2020, Institute of Electrical and Electronics Engineers, 2020.
- [4] J. Li et al., "Demonstration of 200 Gb/s/ $\lambda$  PON in O-band with high-bandwidth TFLN modulator," *J. Lightw. Technol.*, 2023, early access, doi: 10.1109/JLT.2023.3256361.
- [5] S. Luo et al., "112-Gb/s/ $\lambda$  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.*, San Diego, CA, USA, 2020, pp. 1-3, doi: 10.1364/OFC.2020.Th3K.4.
- [6] C. Li et al., "Demonstration of symmetrical 50-Gb/s TDM-PON in O-band supporting over 33-dB link budget with OLT-side amplification," *Opt. Exp.*, vol. 27, no. 13, pp. 18343–18350, 2019, doi: 10.1364/oe.27.018343.
- [7] Song, Yingxiong et al., "Simulation and experimental investigation of nonlinear effects in 5G fronthaul transmission system based on WDM-PON architecture". *Optical Fiber Technology*, vol. 65, p. 102628, 2021, doi: 10.1016/j.yofte.2021.102628.

# 02

## 5G Radio Unit based on an RFSoc FPGA



The Field Programmable Gate Array (FPGA) team at Altice Labs delivers embedded digital designs for communication products. Adopting these devices in the product life cycle allows for adding new features or correcting issues without hardware redesigns, thereby eliminating additional costs.

This paper highlights the significant role of FPGA devices in developing Fifth Generation (5G) Radio Unit (RU) products, aligned with 3rd Generation Partnership Project (3GPP) specifications, which define waveforms, protocols, interfaces, and the overall architecture used in current mobile networks and have been instrumental since the Third Generation (3G).

An overview of the architecture is provided, describing the main functions of the AMD Ultrascale+ RFSoc Digital Front End (DFE) FPGA [1] device family 'xczu67dr' for Altice Labs' 5G RU development. The embedded digital design, developed with Instituto de Telecomunicações - Aveiro, and supported by AMD IP cores, is illustrated with our modules. Each module's behavior is written and simulated using Very High-Speed Integrated Circuit Hardware Description Language (VHDL) code and Matlab [2] before final implementation. The embedded digital design implements physical layer functionality aligned with standards, enabling our 5G RU to interface and communicate in 5G networks.

---

## Authors

**Nelson L. Arqueiro, Altice Labs, [nelson-l-arqueiro@alticelabs.com](mailto:nelson-l-arqueiro@alticelabs.com)**

**Pedro C. Marques, Altice Labs, [pedro-c-marques@alticelabs.com](mailto:pedro-c-marques@alticelabs.com)**

**Francisco F. Serôdio, Altice Labs, [francisco-f-serodio@alticelabs.com](mailto:francisco-f-serodio@alticelabs.com)**

**Manuel J. Almeida, Altice Labs, [manuel-j-almeida@alticelabs.com](mailto:manuel-j-almeida@alticelabs.com)**

**Joaquim F. Serra, Altice Labs, [jserra@alticelabs.com](mailto:jserra@alticelabs.com)**

**Luís F. Almeida, Instituto de Telecomunicações, [luisfilipealmeida@ua.pt](mailto:luisfilipealmeida@ua.pt)**

**Samuel S. Pereira, Instituto de Telecomunicações, [samuel.s.pereira@ua.pt](mailto:samuel.s.pereira@ua.pt)**

**Arnaldo S.R. Oliveira, Instituto de Telecomunicações, [arnaldo.oliveira@ua.pt](mailto:arnaldo.oliveira@ua.pt)**

---

## Keywords

5G-NR; O-RAN; O-RU; O-DU; FPGA; RFSoc; DFE; OFDM; gNodeB; eCPRI; RoE; PTP; 3GPP



## Introduction

FPGA devices became a relevant component of the main modules of diverse technologies, responding to an increasingly high demand for complex, performant technology product solutions. Interface diversity, reconfigurability, and capability to process high data rates make such devices the perfect solution to be aligned with 5G applicable standards.

The 5G New Radio (NR) may operate in different frequency ranges to support distinct scenarios, providing a higher capacity and a broader range. To contribute to that flexibility, the 5G-NR Radio Access Network (RAN) architecture is composed of various logical nodes called gNodeB, and these are divided into three distinct components: the Central Unit (CU), the Distributed Unit (DU), and the RU. This division is fundamental to the modular architecture of 5G, providing flexibility and scalability [3] in network deployment and management. This translates into different possible topologies, such as point-to-point, point-to-multipoint, and ring architecture.

The Open Radio Access Network (O-RAN) [3] aims to standardize interfaces across RAN units to allow for interoperability and flexibility among different vendor CUs, DUs, and RUs. The O-RAN fronthaul interface (RU to DU) is implemented by the

embedded AMD O-RAN FPGA IP core and supports eCPRI, IEEE 1914.3 (NGFI), IEEE 1588 Precision Time Protocol (PTP), Synchronous Ethernet and Network OAM. The ability of eCPRI to run on Ethernet is a significant advantage in urban areas and indoor environments where 5G coverage will be required. With such functionalities, the implemented RU is compliant with the O-RAN specification [3] and mandatory to operate on the 5G mobile network.

## 5G Concepts

### ORAN Overview

As mentioned above, each 5G gNodeB is composed of three different units: CU, DU, and RU. The functionalities each of these components is responsible for depend on deployment requirements, corresponding to different functional splits. The 3GPP organization defines eight different options (**Figure 1**) for these splits [4], each tailored to a different deployment scenario, balancing considerations of cost and complexity. These options allow operators to optimize their network architecture based on specific needs, such as latency requirements, coverage areas, and capacity.

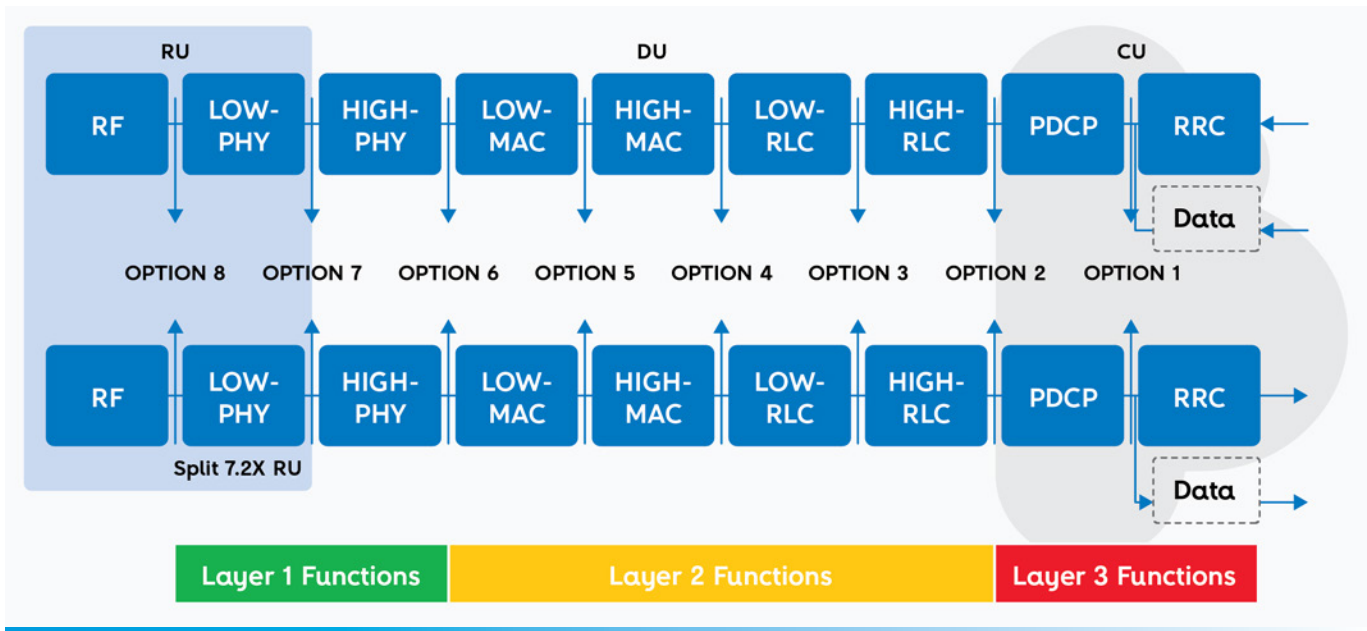


FIGURE 1 – 5G-NR Functional Splits [4]

The presented RU is based on split 7.2x defined by the O-RAN specifications [3], splitting the physical (PHY) layer functionality into the High-PHY in the DU and the Low-PHY in the RU, making it an O-RU.

### Time and frequency structure

A crucial aspect to understand the following sections is the 5G Time/Frequency frame structure [5]. In the time domain, the 5G radio frame is composed of 10 subframes (Figure 2). The number of

slots per subframe varies with the adopted numerology, but the symbols per slot are constant, with a fixed value of 14 Orthogonal Frequency-Division Multiplexing (OFDM) symbols. Each subframe has a fixed length of 1 ms, meaning that as numerology increases, the symbol period decreases.

On the frequency domain, each symbol is divided into various frequency sections that are called Resource Blocks (RBs), and the number of RBs varies according to the used bandwidth. For example, for a signal with a bandwidth of 100 MHz, each symbol

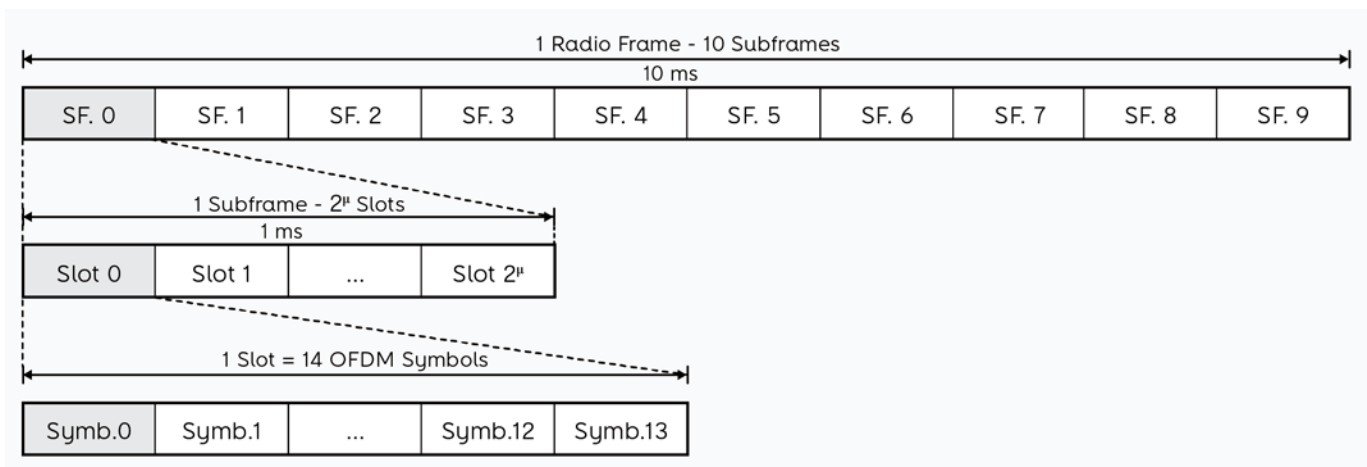


FIGURE 2 – 5G Radio Frame [5]

is composed of 275 RBs, with two being used for guard band purposes, leaving 273 RBs to be allocated for useful information. Each RB is composed of 12 Resource Elements (REs) or Subcarriers (SCs), each containing a pair of IQ samples.

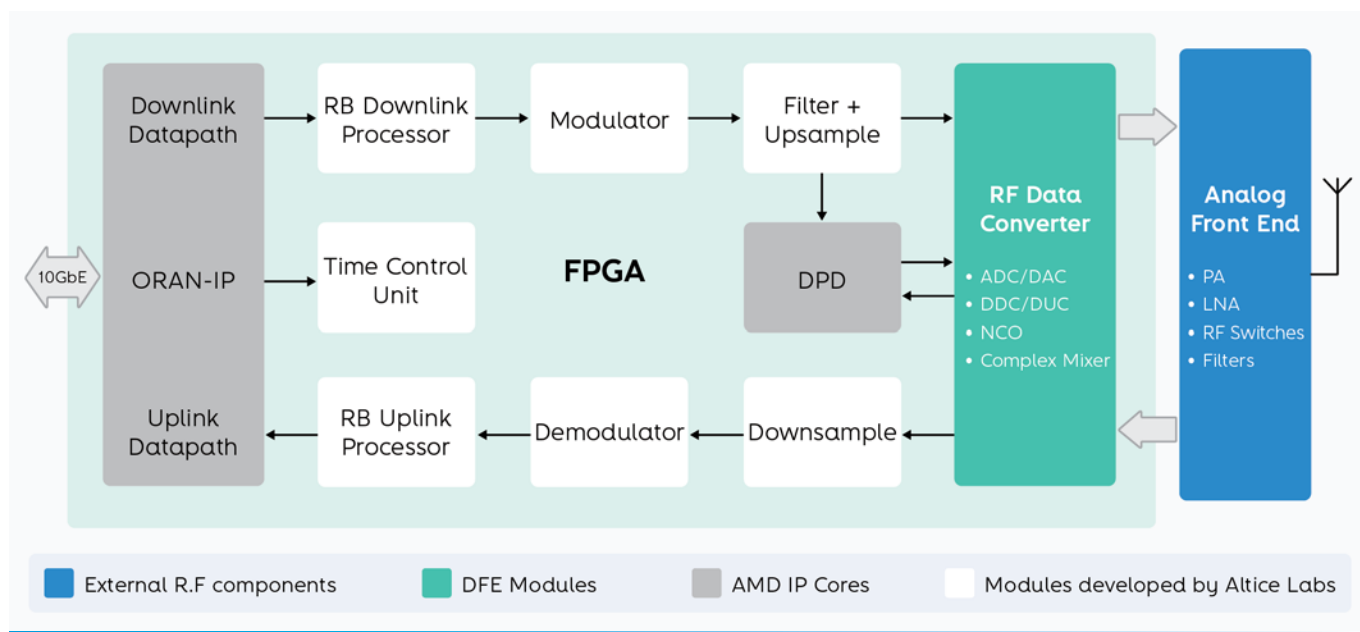
## O-RU FPGA implementation

**Figure 3** illustrates all the modules involved in the implemented O-RU solution. Each module guarantees the correct data rate processing, achieving the expected traffic performance for the 5G mobile network. The DU 10 GbE interface on the left (eCPRI interface) is directly connected to the O-RAN-IP core block, being in charge of the ingress and egress data traffic processing, with the frame format according to the IEEE standards. It also supports the PTP, which guarantees network synchronization between the RU and the DU, with a high degree of accuracy, based on periodic Master/Slave timing messages exchanged over the network to the left. These frames carry timing adjustment information that is able to compensate network latency and

jitter for all the network nodes. These messages are detected by the O-RAN interface that then applies all the required timing adjustments extracted from those messages to achieve, in some scenarios, nanosecond-level accuracy.

The selected 'xczu67dr' Zynq Ultrascale+ RFSoc DFE [1] device has embedded Analog to Digital Converters (ADCs) and Digital to Analog Converters (DACs), working together with a hardened radio subsystem that facilitates an integrated and complete RF solution. The RF-DACs can achieve a maximum rate of 10.0 GSPS and 5.9 GSPS from the RF-ADCs [6], depending on the selected target device. The RF data converters also incorporate Digital Down Converters (DDCs) and Digital Up Converters (DUCs) that include programmable interpolation and decimation rates, a Numerically Controlled Oscillator (NCO), and a complex mixer. The RF Data Converter IP core is fully configurable to support a different mix of signals and required bandwidths of the 5G spectrum.

The analog front end includes the Power Amplifier (PA), Low Noise Amplifier (LNA), Radio Frequency (RF) switches, and analog filters, making the solution compliant with the strict requirements of the 3GPP specification for 5G-NR.



**FIGURE 3** - O-RU FPGA Block Diagram

## Fronthaul Interface

As mentioned before, the implemented fronthaul interface is O-RAN compliant, and it houses various functionalities, which include the 10G Ethernet Medium Access Control/Physical Layer (MAC/PHY) and the O-RAN packet processing interface, amongst many others (see **Figure 4**). The O-RAN interface module [3] must handle the traffic received from the Ethernet MAC in the downlink (DL) direction and transmit the packets carrying frequency domain samples in the uplink (UL) direction, according to the O-RAN specifications.

For the UL direction, the interface must be able to process O-RAN Distributed Unit (O-DU) requests, forward the UL samples after being compressed, framed with eCPRI, and then sent to the Ethernet MAC.

The fronthaul interface is a component of major importance to the RU, since it is responsible for receiving and transmitting the frequency domain samples in a packet-based fashion to and from the DU and to ensure that these packets are O-RAN compliant, which allows the RU to interoperate with diverse O-DUs. This compliance involves a plethora of specifications, the most crucial being the ones related to protocols applicable to the packets carrying samples, which must use an eCPRI or IEEE1914.3 Radio over Ethernet (RoE) layer, along with an O-RAN specific one. A brief overview of the roles of these components follows.

## 10G Ethernet MAC/PHY

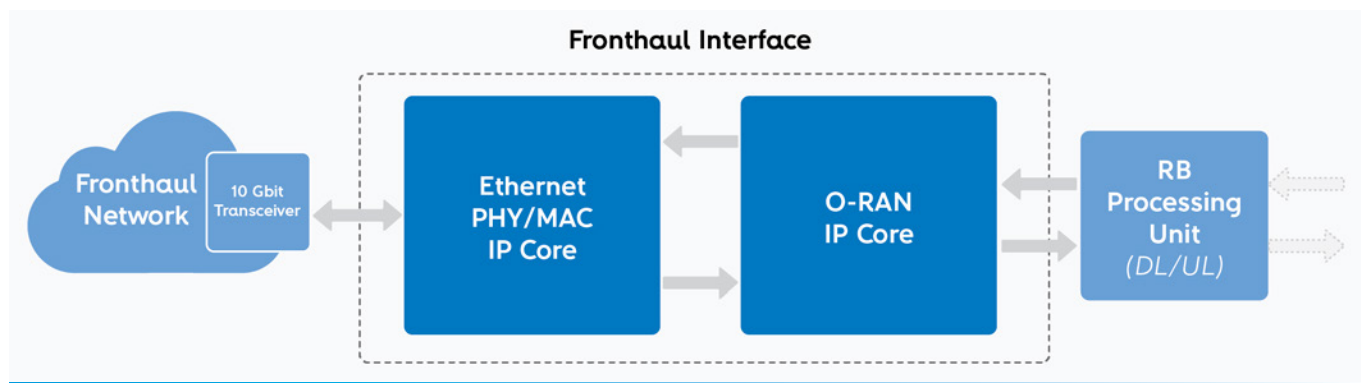
The 10G Ethernet MAC/PHY is responsible for handling the transmission and reception of data at 10Gbit/s. The MAC organizes data into packets, handles addressing, ensures error-free delivery, and manages the flow of data to avoid congestion. The PHY contains various complex functionalities, such as the physical transmission of data over the physical medium, converting data into signals for transmission, and the inverse operation for reception. These functionalities are implemented by the embedded AMD's Ethernet MAC IP Core, together with our developed logic.

The O-RAN interface module [7] must not only handle the traffic received from the Ethernet MAC in the DL direction but also transmit the packets carrying frequency domain samples in the UL direction, according to the O-RAN specifications.

In the DL direction, the interface conducts the filtering process of the O-RAN packets, their dissection, and the subsequent decompression, before forwarding them to the modulator. As for the UL direction, the interface must be able to process O-DU requests, forward the UL samples after being compressed, framed with eCPRI, and then send them to the Ethernet MAC.

## RB Processing Units

In the DL direction, the RB Downlink Processor samples received from the O-RAN interface are processed and sent to the modulator, which expects symbols that are 273 RBs wide. If a symbol does

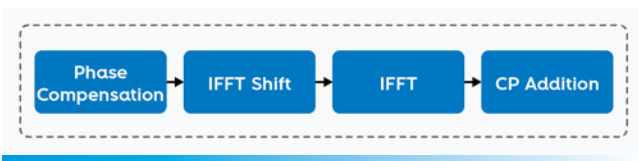


**FIGURE 4** – Fronthaul interface diagram

not occupy all its RBs, padding is added to fill the gaps, ensuring full symbols are delivered. In the reverse direction, the RB Uplink Processor outputs full symbols, but the O-DU may only require specific portions. To accommodate this, the system uses circular memory to store the symbols, allowing only the requested sections to be accessed while new symbols overwrite the previously read ones. If partial symbols are requested, padding is added to complete them before sending them to the O-RAN interface.

## Modulator

In 5G OFDM systems, the data stream is divided into multiple parallel substreams, each modulated into a different subcarrier, each orthogonal to others, closely spaced in frequency. In the DL, the O-RAN-IP core delivers the baseband signal to the modulator, where distinct Layer 1 functionalities are implemented. **Figure 5** illustrates the distinct functionality stages provided by the modulator, where a brief description of each one is presented in the following subsections.



**FIGURE 5** – Modulator - Functional Diagram

### Phase Compensation

This block was developed using the Matlab tool and then exported to VHDL for implementation and validation purposes. The phase compensation procedure is the first processing stage of the modulator block. In 5G, phase compensation is a technique used to correct phase or offset errors introduced during transmission and reception. Ignoring channel distortions and noise for simplicity, the signal at the receiver side can be described as presented in [8] Annex E:

$$s_L(p, \mu)(t) \cdot e^{j2\pi(f_0 - f_1)t} = s_L(p, \mu)(t) \cdot e^{j\theta l} \cdot e^{j2\pi(f_0 - f_1)(t - t_{CP,L,\mu})}$$

$$= s_L(p, \mu)(t) \cdot e^{j\theta l} \cdot e^{j2\pi(f_0 - f_1)t_{CP,L,\mu}} \cdot e^{j2\pi(f_0 - f_1)t}$$

$s_L(p, \mu)(t)$ , signal on antenna port  $p$  and subcarrier spacing  $\mu$

$\theta L = 2\pi(f_0 - f_1)t_{CP,L,\mu}$ , symbol subjected to a phase shift  $\theta L$

$t_{CP,L,\mu} = t_{start,L,\mu} + N_{CP,L,\mu}T_c$  is the time at the end of the OFDM symbol  $l$

$f_0$ , represents the central frequency of the RF signal at the transmitter

$f_1$ , represents the central frequency of the RF signal at the receiver

Each received OFDM symbol is subjected to a phase shift, resulting in phase changes between successive OFDM symbols, leading to possible frequency error tracking and a weak channel estimation process. To avoid such problems, the solution proposed by 3GPP in TS 38.211 clause 5.4 [8] is to apply a phase pre-compensation term  $e^{-j2\pi f_0 t}$  at the transmitter and the term  $e^{j2\pi f_0 t}$  at the receiver. These two terms together provide the required correction  $e^{-j\theta l}$ . This procedure helps to maintain signal integrity, enhances spectral efficiency, and improves the reliability of communication systems.

### The Inverse Fast Fourier Transform (IFFT)

The Fast Fourier Transform (FFT) algorithm is too computationally demanding to be efficiently executed solely on software; an FPGA implementation is more power efficient.

At this stage, the input data symbols modulated onto the subcarriers are in the frequency domain and must be converted to the time domain, being this operation the main IFFT role. The Inverse zero-frequency Shift (IFFTSHIFT) function is a mathematical operation that shifts the zero-frequency component of a signal to the center of the spectrum, avoiding phase discontinuities when

performing Inverse Fourier Transforms (IFT). This shift algorithm, here developed using VHDL, basically divides the signal into two halves and swaps their positions. The applied Inverse Discrete Fourier Transform (IDFT) and Discrete Fourier Transform (DFT) algorithms can be described by the following Equation 1 and Equation 2:

$$1) \quad x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{j2\pi \frac{kn}{N}}$$

$$2) \quad \text{where, } X(K) = \sum_{k=0}^{N-1} x(n)e^{j2\pi \frac{kn}{N}}$$

$X(k)$  represents the  $k$ -th frequency component of the DFT

$x(n)$  represents the time-domain signal

$N$  represents the number of samples in the sequence

$e$  is the base of the natural logarithm

$j$  represents the imaginary unit

$k$  represents the frequency index in the frequency domain

In our case, the IFFT is configured for 4,096 samples, according to the selected numerology ( $\mu = 1$ ). The output samples of the IFFT core correspond to a time-domain waveform that represents a part of the signal.

### Cyclic Prefix (CP) Addition

One of the causes of Intersymbol Interference (ISI) is the multipath propagation effect, which occurs when the RF signal from a transmitter reaches the receiver via multiple paths, resulting in the overlapping of adjacent symbols in a communication channel, making it difficult for the receiver to accurately distinguish between them. To mitigate the ISI impact, an interval guard between symbols is created. The objective is to add CP, which is a copy of the end portion of the OFDM symbol, appended to the beginning. The CP helps

maintain orthogonality by providing additional "space" between OFDM symbols, reducing the interference between them [9].

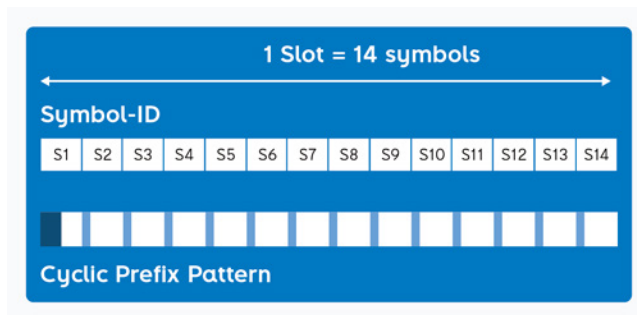


FIGURE 6 – Cyclic Prefix Symbol Pattern for  $\mu = 1$

The 3GPP specifies two types of CPs, Normal Cyclic Prefix (NCP) and Extended Cyclic Prefix (ECP), the latter being only used for the 60kHz subcarrier spacing. If normal CP (NCP) is used, the CP of the first symbol is present every 0.5 ms and is longer than the other symbols, as illustrated in Figure 6. The CP length for different subcarriers can be calculated using the following formula.

$$\text{CP length} = \begin{cases} 512k \times 2^{-\mu} & \text{if extended cyclic prefix} \\ 144k \times 2^{-\mu} + 16k & \text{if } l = 0 \text{ or } l = 7 \text{ for normal cyclic prefix} \\ 144k \times 2^{-\mu} & \text{if } l = 0 \text{ and } l = 7 \text{ for normal cyclic prefix} \end{cases}$$

$$\text{where, } k = \frac{T_s}{T_c} = \frac{\Delta f \times N_f}{\Delta f_{ref} \times N_{ref}} = \frac{\Delta f_{ref} \cdot n_{FFT}}{15 \times 10^3 \times 2048}$$

$l$  is the respective symbol

$\mu$  is the numerology

$k$  is a constant used to relate the 5G-NR basic time unit

$T_s$  is the slot duration

$T_c$  is the subframe duration

$f$  is the subcarrier spacing

$N_{ref}$  is the number of subcarriers

$f_{ref}$  is the reference frequency

Numerology	Subcarrier spacing	CP Long ( $\mu\text{s}$ )	CP Normal ( $\mu\text{s}$ )
$\mu = 0$	15 kHz	l=0 or 7 NCP=5.2	NCP=4.69
$\mu = 1$	30 kHz	l=0 or 14 NCP=2.86	NCP=2.34
$\mu = 2$	60 kHz	l=0 or 28 NCP=1.69 ECP=4.17	NCP=1.17 ECP=4.17
$\mu = 3$	120 kHz	l=0 or 56 NCP=1.11	NCP=0.59
$\mu = 4$	240 kHz	l=0 or 112 NCP=0.81	NCP=0.29

TABLE 1 – CP Length versus Numerology in 5G-NR

Based on the presented CP length formula, **Table 1** summarizes the duration of the normal and the long cyclic prefix per numerology. Each numerology has two long symbols per 1 ms, and these longer symbols are generated by increasing the duration of the normal cyclic prefix.

## Filtering

Due to the 5G modulator’s finite quantization, the generated signal presents a substantially high out-of-band leakage. This is due to the required trade-off between resource usage by the modulator and waveform integrity and is a problem because the out-of-band leakage would be amplified and radiated, violating the masks of 5G. A combination of two Finite Impulse Response (FIR) filters and a digital pre-distortion is employed to minimize this impact. The FIR filters used can be generally described by Equation 3.

$$3) \quad y[n] = b_0x[n] + b_1x[n - 1] + \dots + b_Nx[n - N] = \sum_{i=0}^N b_i x[n - i]$$

The filtered result,  $y[n]$ , can be obtained by multiplying the input signal,  $x[n]$ , by the filter coefficients, with an impulse response shown in **Figure 7**. Depending on the number of coefficients of the filter and the desired cut-off frequency, a sharper or wider filter can be obtained. Adjusting these parameters allows an out-of-band leakage minimization without compromising the total resources of the filter implementation. A generic implementation of these filters can be seen in **Figure 8**.

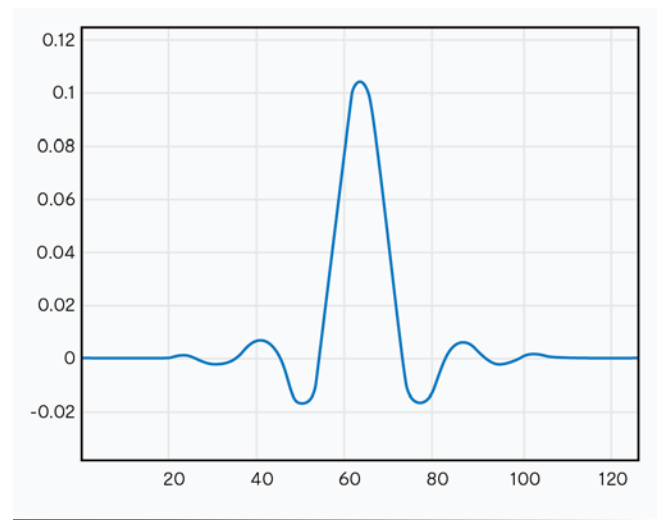


FIGURE 7 – Impulse response

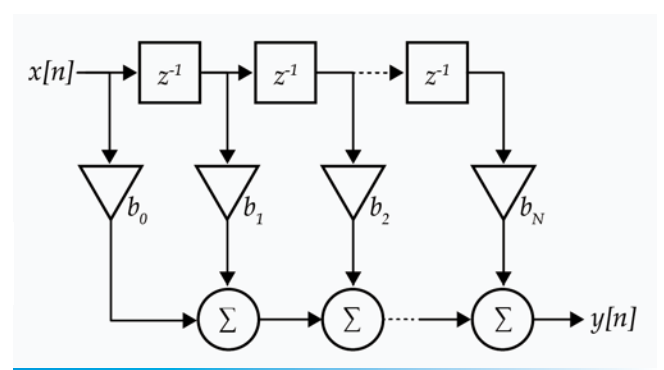


FIGURE 8 – Generic filter implementation

As mentioned, a combination of two filters is employed in series. The first one minimizes out-of-band leakage, and the second adjusts the sampling rate of the modulated waveform to the input of the digital pre-distortion block. The latter also acts as an interpolator.

## Demodulator

This section describes the main functionalities of the 5G Demodulator. The platform used to develop and simulate the behavior of this module was based on the Matlab/Simulink tool [10]. For hardware validation, the design was implemented in the “xczu67dr” AMD UltraScale+ FPGA with the support of the Vivado [11] environment design tool.

### Functionality

The main function of the Demodulator is to process the UL raw data stream received from the ADC/ antenna and extract the two main data and synchronization channels:

- Physical Random Access Channel (PRACH): Dedicated channel for establishing UL synchronization and the first step for the User Equipment (UE) attachment procedure with the network;
- Physical UL Shared Channel (PUSCH): Main UL channel that carries the user data and signaling.

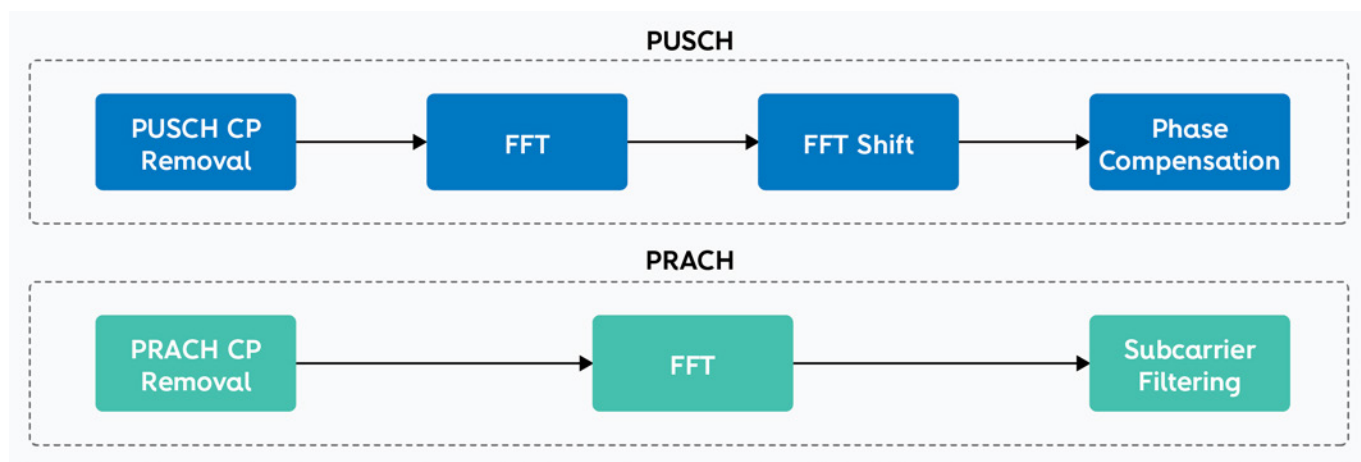
The current implementation of the Demodulator is parameterizable and can be used to process different bandwidth and Physical Resource Blocks (PRB) configurations relative to distinct Numerologies, as presented in **Table 1**.

### Operation

The Demodulator takes the I/Q data samples directly from the ADC integrated into the AMD RF Data Converter, being then split between the PUSCH and PRACH datapaths. Below are described the internal blocks that constitute the PRACH and PUSCH processing paths, also pictured in **Figure 9**:

- CP Removal: The first step for both UL channels is to remove the respective (CP) from the OFDM symbol;
- FFT: Computes the DFT using a FFT algorithm. I/Q samples are transformed from the time domain into the frequency domain. FFT size is dependent on bandwidth and PRB configuration;
- FFT Shift: Required only in the PUSCH datapath, reorders the FFT output by shifting the zero-frequency component to the center of the data vector;
- Phase compensation: Required only in the PUSCH datapath, analogous to the DL module of the same name previously explained in section 5.1;
- Subcarrier filtering: Required only in the PRACH datapath, filters the 139 valid subcarriers relative to the B4-PRACH [8] short preamble.

Each UL channel output is then fed to the O-RAN interface to be sent to the fronthaul.



**FIGURE 9** – Simplified view of the Demodulator internal functions

## Digital Pre-Distorsion

The Digital Pre-Distortion (DPD) and Crest Factor Reduction (CFR) modules are part of the DFE processing that occurs before the analog converters in our RU. **Figure 10** illustrates the transmit (DL) data path. The main purpose of the DPD is to adjust variations in the Power Amplifiers (PAs). This method involves inserting a nonlinear function between the input signal and the amplifier to produce a linear output. For that, an observation ADC is used to capture the PA's output signal, which is returned to the DPD module. This information helps the DPD system to adjust the input signal to counteract the amplifier's nonlinearities and distortions. The DPD continuously adjusts variations in the PAs' nonlinearity, caused by factors such as time, temperature, and different operating channels.

The CFR module minimizes the risk of a transmission-side PA getting saturated when transmitting an OFDM signal, a common occurrence in 5G systems. This saturation risk is due to the high Peak to Average Power Ratio (PAPR) of the OFDM waveform. This module identifies the highest peak in a symbol and reduces it without changing the phase shift. It transmits the same Quadrature Amplitude Modulation (QAM) symbols with modified amplitude, if the amplitude exceeds a specified threshold.

The functionalities described above enable inexpensive PAs to function in their nonlinear regions, minimizing distortions and thus resulting in higher output power and power efficiency improvements. Given the vast number of wireless base stations worldwide, improving PA efficiency may significantly lower the energy and cooling costs for cellular operators. This solution is designed for base stations in mobile technologies and similar applications, using a combination of hardware and embedded software processes to implement DPD correction.

## Test setup and results

To verify that the O-RU, based on the 'xczu67dr' FPGA device, operates as intended, it is essential to conduct comprehensive end-to-end testing and confirm O-RAN interoperability, RF signal transmission and reception, bit error rate, and measure and check key performance indicators. To achieve this, the test procedures were divided into two setups: (i) one that uses an O-DU emulator together with a Vector Signal Generator (VSG) [12] and a Vector Signal Analyzer (VSA) [13] to validate the O-RU's O-RAN interoperability, modulation, demodulation, and RF performance in a controlled and isolated manner; (ii) a second one validating the O-RU's 5G connection to a real O-DU, verifying if a UE is able to obtain a 5G connection.

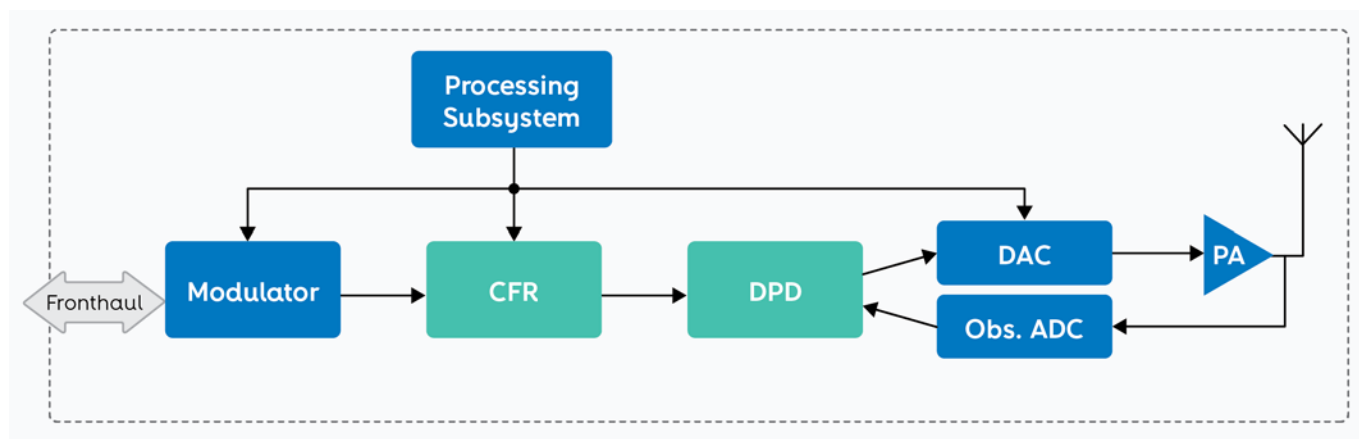


FIGURE 10 – DFE Transmit Path

## Functionality validation and integration

As mentioned above, this setup aims to validate the O-RU's functionalities by using an O-RAN compliant O-DU emulator [14] to generate and receive traffic. The VSG equipment injects an RF signal into the O-RU while the VSA is used to check the transmitted 5G signal. RF transmission in both directions is carried out using RF cables, eliminating any potential wireless-related issues and focusing on validating the implemented functionalities. The diagram below (**Figure 11**) depicts the adopted setup.

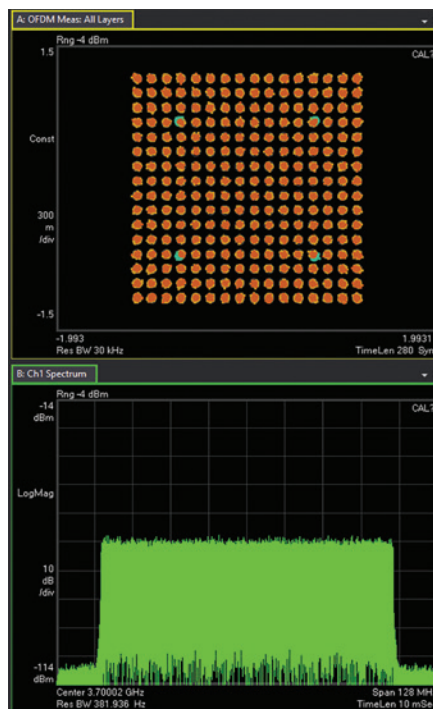


**FIGURE 11** – Test setup diagram

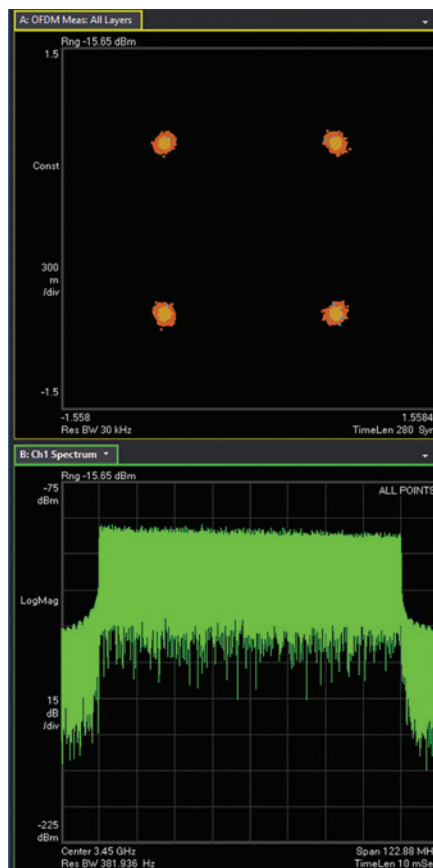
For UL validation, the O-DU emulator must generate UL request packets, while the VSG injects a 5G signal into the RF chain. The O-RU should handle the request traffic, perform data processing, and transmit the data back to the O-DU emulator. The latter then processes the received O-RAN packets, and the signal can be analyzed on the VSA, where the VSG-injected signal's constellation should be visible; for DL validation, the same process was used.

## Validation results

For the DL test, a 256 QAM 5G signal occupying 273 RBs was generated, loaded onto the O-DU emulator, and then transmitted. The signal was successfully demodulated on the VSA connected to the RF port. A constellation with a low bit error rate was observed, validating the DL/UL capabilities as depicted in **Figure 12**. For the reverse direction, the same procedure was used, now with a 4 QAM 5G signal. This process was also successfully completed with a constellation with a low bit error rate, illustrated in **Figure 13**.



**FIGURE 12** – Setup test results, DL



**FIGURE 13** – Setup test results, UL



## Interoperability and UE connection

The second setup was used to test an end-to-end 5G connection interoperability with an O-DU and a UE (see **Figure 12** and **Figure 13**). Unlike the wired tests in the Keysight setup, the wireless transmission was here adopted.

While the first setup was focused on validating the individual components of the O-RU performing end-to-end testing, here the objective was to ensure that the O-RU interoperates with a commercial O-DU (ASOCS's O-DU [15]) and verifying if the 5G signal is detected by the UE, allowing it to attach and use services. If the signal is detected and a connection is established with success, the UE should have Internet access. This approach validated the O-RU in a complete, more realistic environment.

## Results

Once the setup was operational and both the O-DU and O-RU were properly configured, the UE was powered on. A Customer Premise Equipment (CPE) was used, instead of a cellphone, as shown in **Figure 14**. It successfully detected the 5G signal and connected to the network, establishing a 5G PDU session required for the exchange of data related to devices connected to that CPE via Wireless Local-Area Network (WLAN) (**Figure 16**).

Power consumption measurements have also been conducted with two 5G-RU implementations but with different chipsets (the xczu67DR and its previous generation, the xczu47DR) in the same scenario. Power was measured between the AC/DC converter and the input of 5G-RU equipment on 48VDC and its Root Mean Square (RMS) value. The selected 'xczu67dr' FPGA proved to be 10% more power efficient compared to the previous chip generation.

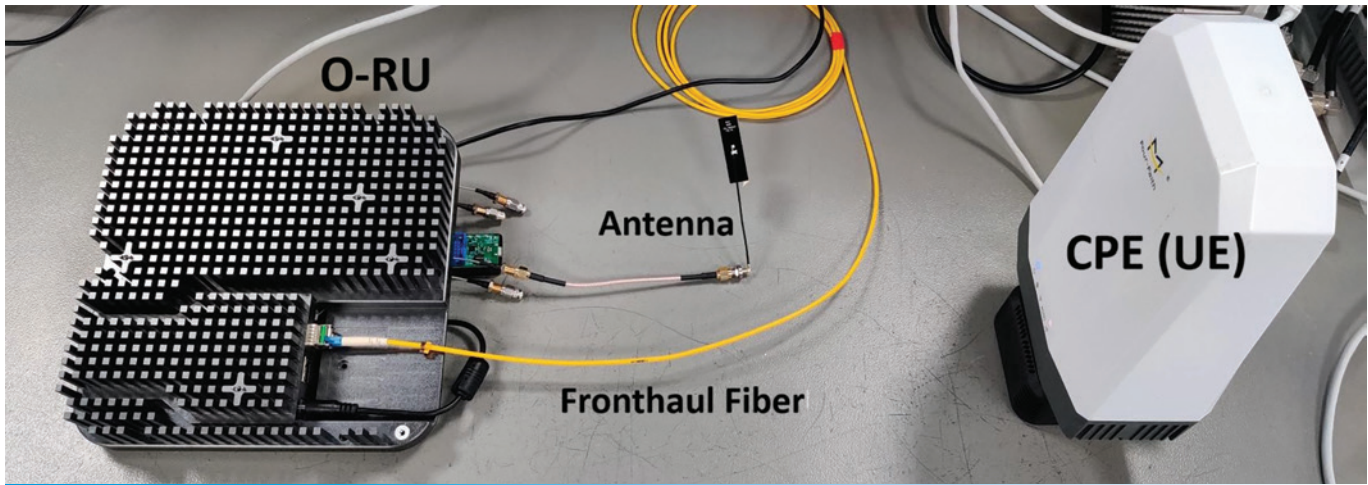


FIGURE 14 – O-RU connected to ASOCS O-DU with a CPE being used as UE (wireless setup)

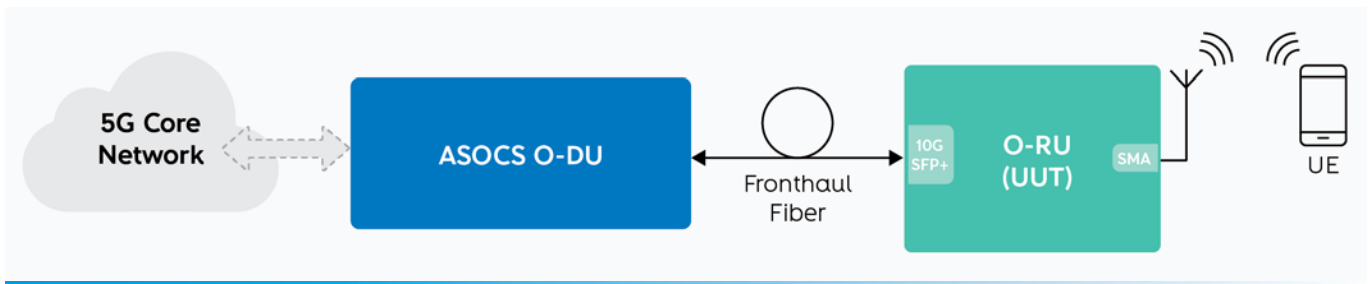


FIGURE 15 – Controlled Test Setup Diagram (ASOCS’s Setup)

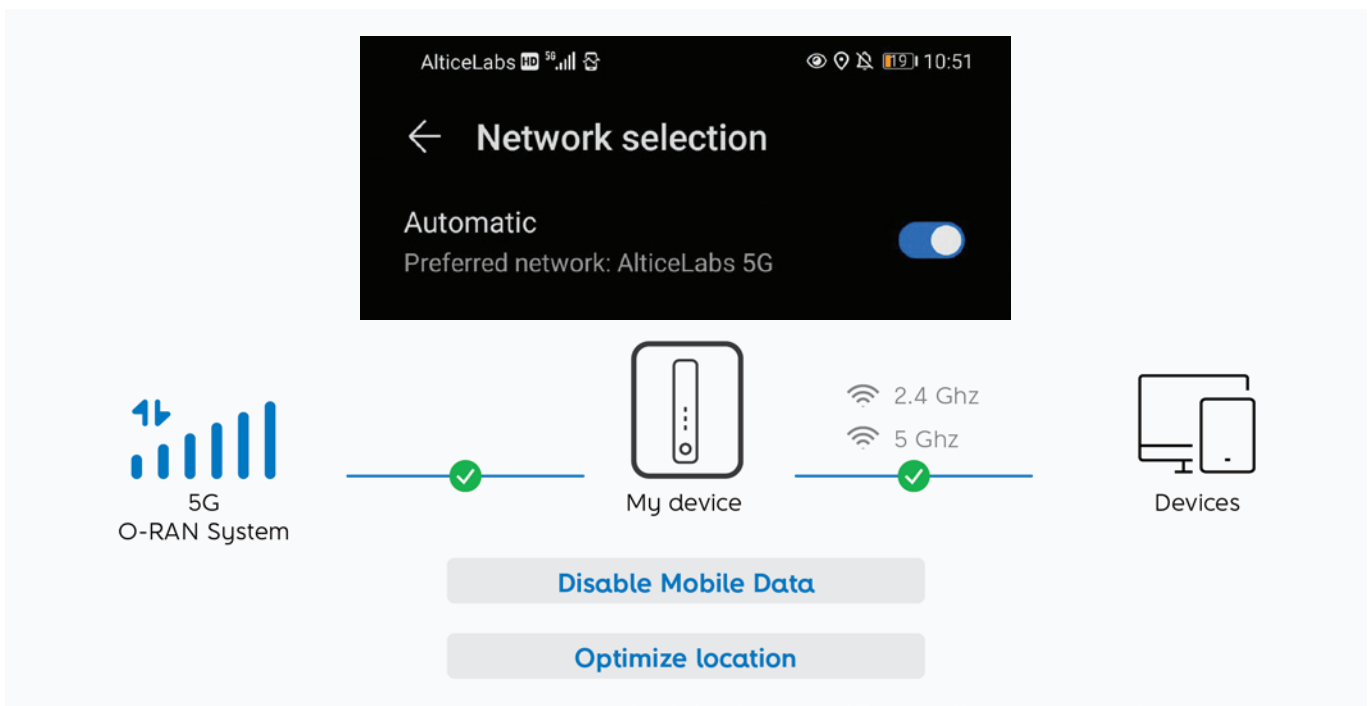


FIGURE 16 – CPE connection to Altice Labs 5G private testbed network, with one device connected to the Internet

## Conclusions

The Zynq Ultrascale+ RFSoc DFE platform enables the development of flexible, high-performance, and power-efficient radio solutions, through the integration of a fully hardened digital front-end subsystem, in addition to the direct RF. The available cores are a plus to decrease time to market and guarantee the final performance of the present solution. The selected 'xczu67dr' FPGA is more power efficient than the previous generation making this device an important player to support our solution and enable end-to-end communication infrastructure upgrades.

The presented work demonstrates that the selected Zynq Ultrascale+ RFSoc DFE 'xczu67dr' device, with the embedded IPs and our developed modules, enabled our 5G RU solution to be part of a real 5G mobile network. All these features

provide high flexibility to accommodate a broad range of use cases, supporting numerous carrier and band configurations. Compliance with the 3GPP specification and with the O-RAN is important and mandatory for our RU since it guarantees the 5G communication infrastructure conformity.

Besides the presented tests, a wide range of distinct signals were used to test and validate our RU performance, configured with a wide range of data patterns and various types of modulations. The wide range of conformity tests demonstrated the robustness of our RU, as well as its ability to be in line with the specifications required in 5G networks. The proof of concept (PoC) was successfully achieved with the connection between our RU and commercial CPEs, with bandwidth signals of 100MHz and 20MHz. All in all, we demonstrated that the developed RU was able to establish communication with commercial CPEs in both directions, DL and UL. [🔗](#)



## References

- [1] "Zynq UltraScale+ RFSoc DFE Data Sheet: Overview," XILINX, San Jose, California, U.S., Rep. no. DS883, March 30, 2022. Accessed: Oct. 14, 2024. [Online]. Available: <https://docs.amd.com/v/u/en-US/ds883-zynq-rfsoc-dfe-overview>.
- [2] "MATLAB." mathworks.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://mathworks.com/products/matlab.html>.
- [3] O-RAN Alliance, "Control, User and Synchronization Plane Specification (O-RAN.WG4.CUS.0-v10.00)," Technical Report, 2022.
- [4] O. Andersson. "Functional Splits: the foundation of an Open 5G RAN." 5gtechnologyworld.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.5gtechnologyworld.com/functional-splits-the-foundation-of-an-open-5g-ran>.
- [5] "5G/NR – Frame Structure." sharetechnote.com. Accessed: Oct. 14, 2024. [Online] Available: [https://www.sharetechnote.com/html/5G/5G\\_FrameStructure.html](https://www.sharetechnote.com/html/5G/5G_FrameStructure.html).
- [6] "Zynq UltraScale+ RFSoc RF Data Converter v2.4 Gen 1/2/3," XILINX, San Jose, California, U.S., Rep. no. PG269, Nov. 30, 2020. Accessed: Oct. 14, 2024. [Online]. Available: <https://docs.amd.com/v/u/2.4-English/pg269-rf-data-converter>.
- [7] AMD, "O-RAN Radio Interface," PG370. v3.1., Oct. 2023.
- [8] 3GPP, "NR; Physical channels and modulation," TS 38.211, July 2020.
- [9] "5G Cyclic Prefix (CP) Design." techplayon.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.techplayon.com/5g-nr-cyclic-prefix-cp-design/>.
- [10] "Simulink." mathworks.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.mathworks.com/products/simulink.html>.
- [11] "AMD Vivado™ Design Suite." amd.com. Accessed: Sept. 18, 2024. [Online]. Available: [www.amd.com/en/products/software/adaptive-socs-and-fpgas/vivado.html](http://www.amd.com/en/products/software/adaptive-socs-and-fpgas/vivado.html).
- [12] "N5182B MXG X-Series RF Vector Signal Generator, 9 kHz to 6 GHz." keysight.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.keysight.com/us/en/options/N5182B/mxg-x-series-rf-vector-signal-generator-9-khz-6-ghz.html>.
- [13] "N9020B MXA Signal Analyzer, 10 Hz to 50 GHz." keysight.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.keysight.com/us/en/product/N9020B/n9020b-mxa-signal-analyzer-multi-touch-10-hz-50-ghz.html>.
- [14] "S5040A Open RAN Studio Player and Capture Appliance." keysight.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://www.keysight.com/us/en/product/S5040A/open-ran-studio-player-and-capture-appliance.html>.
- [15] "Industrial 5G Network - Run your high-performance private network on your own terms." asocscloud.com. Accessed: Oct. 14, 2024. [Online]. Available: <https://asocscloud.com/cyrus/>.

# 03

## The disaggregation and virtualization of the access network



In a world where connectivity demand increases exponentially, fixed network operators must be able to adapt and grow at the same pace to continue delivering the best possible service to their clients. The use of Software Defined Networks (SDN) principles and the virtualization of the physical network are seen as a way to increase the flexibility and agility of the network. The disaggregation of the network devices' functionalities into software components that can run in a cloud environment aims to shorten the technology and service updates lifecycle, to quickly adapt the devices to new service models and features, and to provide new points of interoperability, adding the potential for decoupling the hardware vendor from the software vendor. However, this new paradigm raises new challenges, such as how to deal with legacy devices while migrating to the next-generation networks without increasing operational complexity.

This article describes how the AGORA SDN Manager and Controller (SDN M&C) solution leverages the potential of cloud-based environments and the micro-services principles in access networks to tackle these challenges. The evolution of Altice Labs access nodes will also be described, contributing to a smooth migration path and enabling multiple deployment options, where the operator can choose between running network functions embedded in the devices or disaggregated in a cloud environment. This article also provides an overview of the most relevant standards, fundamental for interoperability.

---

## Authors

**André Domingos Brízido, Altice Labs, [andre-d-brizado@alticelabs.com](mailto:andre-d-brizado@alticelabs.com)**

**Nuno Ricardo Farinha, Altice Labs, [nuno-r-farinha@alticelabs.com](mailto:nuno-r-farinha@alticelabs.com)**

**Miguel Ângelo Miranda, Altice Labs, [miguel-a-miranda@alticelabs.com](mailto:miguel-a-miranda@alticelabs.com)**

**Licínio Miguel Pereira, Altice Labs, [licinio-miguel-pereira@alticelabs.com](mailto:licinio-miguel-pereira@alticelabs.com)**

---

## Keywords

SDN; NFV; Access Networks; PON; CloudCO

## Introduction

The adoption of SDN principles and Network Functions Virtualization (NFV), complemented by well-defined standards, is seen by network operators as a means to address several challenges:

- Increases the scalability and flexibility of the network: virtualizing network functions and decoupling them from dedicated specialized hardware enables the possibility of running some of these functions in Commercial Off The Shelf (COTS) servers installed in distributed data centers, depending on several requirements, such as latency. This virtualization enables the possibility to dynamically instantiate and scale up and down resources according to the demand. Another benefit of running network functions in an Information Technology (IT) infrastructure arises from the possibility to reduce/eliminate software upgrade downtime. Each network function can be upgraded independently, and the traffic can be seamlessly rerouted to the newly upgraded function only when it is ready. Furthermore, this shortens the cycles for technology and service updates;
- Avoids vendor lock-in by having a broad supplier ecosystem with highly interoperable components: the decoupling of network functions from dedicated hardware brings a very high potential for interoperability, since what was typically a monolithic software package running in a specialized network equipment can now be broken down into modular interoperable components, each with a separated lifecycle. This interoperability will only be possible with a well-defined set of standard interfaces. In the past recent years, the Broadband Forum (BBF) [1] has been working on the definition of these interfaces;
- Quickly adapts the network to new service models and features: one of the key components of an SDN network is the SDN controller. This controller should provide an open and

extensible Northbound Interface (NBI) that enables automation and easy integration with northbound systems, abstracting the details of the underlying network. SDN networks also typically implement concepts of Zero-Touch Provisioning (ZTP), automating many tasks that were typically done via manual configuration of the network devices;

- Provides an evolutive and non-disruptive solution: the SDN controller plays a key role in the coexistence between legacy and new nodes, by providing unified management chains and workflows. Furthermore, the evolution of the network elements from monolithic software into modular architectures eases the introduction of new features.

Despite the advantages of the SDN/NFV approach, operators will have to deal with new challenges, such as:

- The increased initial deployment cost, due to the requirements of additional computer infrastructure that enable the running of virtualized network functions;
- The increased complexity. The virtualized network functions that were previously running inside a single device, now need to be orchestrated and their lifecycle handled separately.

The focus of this article is on PON-based access networks. In traditional PON access networks, an operator usually has Optical Line Terminations (OLTs) from multiple vendors, each with proprietary management protocols and/or data models under the control of multiple vendor-specific Network Management System/Element Management System (NMS/EMS). In the SDN/NFV paradigm, the role of the multiple NMS/EMS is performed by an SDN controller, the Access SDN M&C, which can control the entire multi-vendor network. Some of the OLT specific network functions can be disaggregated from the physical OLT (pOLT) and run as virtualized functions. This set of disaggregated functions is called D-OLT (Disaggregated OLT).

## Architecture and standards

In 2015, the Open Networking Foundation (ONF) publicly demonstrated the Central Office Re-Architected as a Datacenter (CORD) project [2]. The CORD platform is built with commodity servers, white-box switches, and specialized hardware, such as OLTs that provide the optical connections to the end subscribers. In terms of software, CORD integrates multiple open-source projects to deliver “a cloud-native, open, programmable, agile platform for network operators to create innovative services” [3]. A specialization of the CORD platform gave origin to the Software Enabled Broadband Access (SEBA) project [4], a specialization for residential access and wireless backhubs.

One of the key aspects of the SEBA project is the use of white-box OLTs and switches controlled by the OpenFlow protocol [5] via the Open Network Operating System (ONOS) SDN Controller. Since the OpenFlow protocol was designed essentially to control ethernet switches, a software component called Virtual OLT Hardware Abstraction (VOLTHA) was introduced to provide a “set of abstract APIs that enable the PON network to appear as a programmable Ethernet switch to an SDN controller” [4].

Despite the advantages of SEBA, such as the use of white-box hardware and being a highly programmable cloud-native open-source platform, and the fact that it has been deployed in some network operators [6], it has some limitations that, in our opinion, have prevented more widespread adoption, such as:

- The fact of being fully based on open source requires network operators to have knowledge of how to deploy, troubleshoot, and maintain each component or, otherwise, rely on a system integrator. Furthermore, the software may require modifications to adapt to each operator's workflow [6];
- It requires additional IT servers and Ethernet switches besides the OLT hardware, which

increases the deployment cost in specific simple scenarios that can be handled by a single vendor-branded chassis-based OLT;

- Limited scalability when compared to some chassis-based OLTs: 1,024 ONUs per VOLTHA stack, 10,240 on 10 stacks across the same infrastructure [7]. This limit was later increased to 4,096 ONUs per VOLTHA stack [8]. However, a single large OLT chassis can support over 24,000 ONTs without any additional IT infrastructure (12-line cards, each with 16 PON ports and using a 1:128 split ratio);
- It does not provide an out-of-the-box integration with existing non-white-box devices already deployed in the field;
- The Openflow protocol has some limitations in the implementation of advanced packet processing functionalities such as Access Control List (ACL), Quality of Service (QoS), and firewalls [9];
- The NBI of the SEBA stack and the South-bound Interface (SBI) between VOLTHA and the OLT hardware are not yet defined in an official standard.

In December 2023, the ONF announced its dissolution, transferring its open-source projects to the Linux Foundation (LF), which created project-directed funds to support the work of the technical projects [10]. The LF Broadband Directed fund is now responsible for continuing the development of SEBA and VOLTHA.

As mentioned before, a well-defined set of standard interfaces is very important to guarantee interoperability and fully leverage the potential of SDN/NFV networks. The BBF is the reference Standards Development Organization (SDO) in access networks, and, in 2018, it published the TR-384 *Cloud Central Office Reference Architectural Framework* [11], which lays the foundation for the Cloud Central Office (CloudCO) architecture – designed for next-generation central offices built on SDN/NFV principles. Besides other benefits, the CloudCO architecture, which has already been

addressed in a previous article [12], provides a migration path and coexistence of devices already deployed in the network [13], widening the scope of the CO architecture to range from legacy equipment to new disaggregated/virtualized and white-box approaches.

The CloudCO architecture is depicted in **Figure 1**, where a clear separation between the access segment, the edge segment, and the NFV infrastructure (NFVI) can be seen. Each segment has an SDN M&C responsible for delivering traditional Fault, Configuration, Accounting, Performance, and Security (FCAPS) functionality but also for providing disaggregated control plane functionalities.

Some major components include the CCO Domain Orchestrator (CCODO), whose role is to orchestrate all the network segments and provide a high-level view of the CCO network. The End-to-End Service Orchestrator (E2ESO) provides the E2E view of the operator network and coordinates multiple CCODOs.

Focusing on the access segment, the BBF published TR-477 [14], introducing the concept of Disaggregated OLT (D-OLT) and specifying interfaces, functional blocks, and requirements for moving functionalities from the devices into the cloud. The purpose of the Broadband Access Abstraction (BAA) component, described in TR-484 [15], is to define an abstraction layer that provides a unified standard set of interfaces for managing multi-vendor and multi-technology access nodes. This BAA component can support traditional chassis-based Access Nodes, as well as white-boxes or disaggregated Access Nodes compliant with the TR-477 specification.

Another specification with major relevance for PON deployments is the TR-451 Virtual ONU Management Control Interface (vOMCI) [16], which defines the interfaces and components to virtualize the OMCI functionality, usually embedded within the OLTs, and one of the OLT functions whose disaggregation was first addressed by the BBF. The OMCI protocol [17] is used by OLTs to configure Optical Network Units (ONUs). Although it is a standard protocol, the implementation may vary

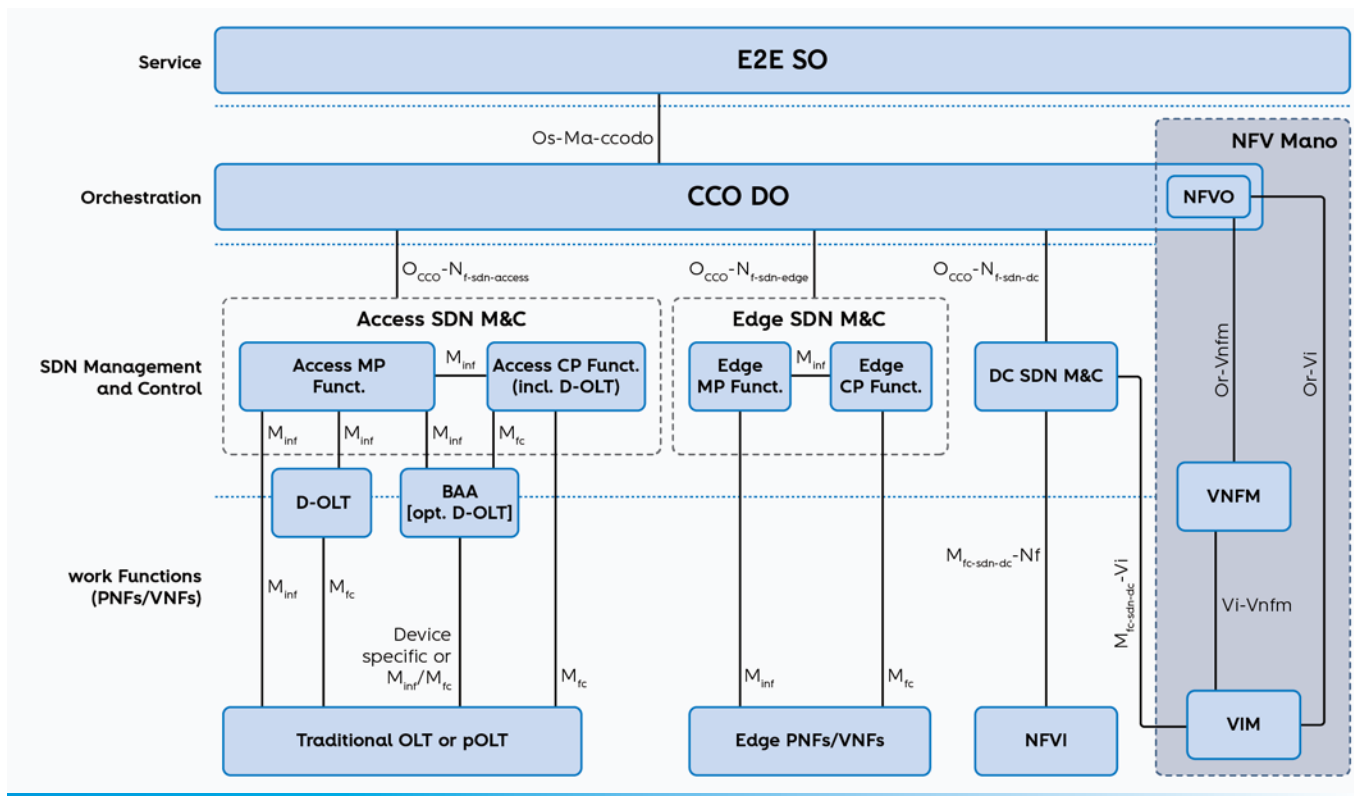


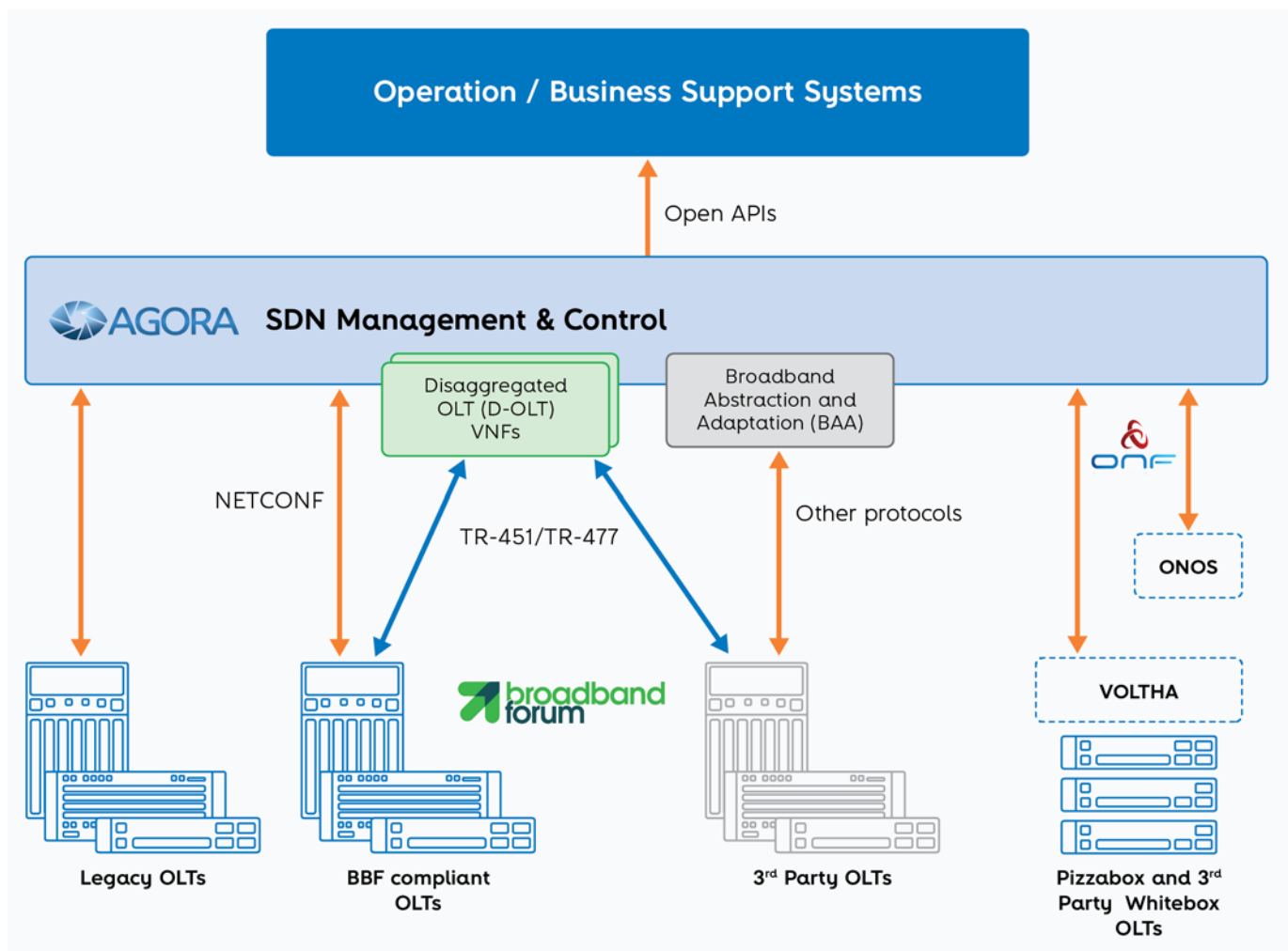
FIGURE 1 – BBF’s Cloud CO architecture [14]

from vendor to vendor or even between devices of the same vendor. As a result, software upgrades on the OLTs are often required when new ONU types are introduced. With TR-451, the OMCI stack is removed from the OLT. The support of new ONU devices is thus possible by simply instructing the OLT to use a new remote/virtualized OMCI stack (the mentioned vOMCI function) to configure them.

Since the TR-384 specification publication, the CloudCO architecture has suffered some adaptations, and several standards have been published to fully define the components and interfaces of the architecture. Recent work includes the evolution to autonomous networks such as the Automated Intelligent Management (AIM) related specifications [18], [19] and Intent-Based Networking (IBN), currently under development.

## Altice Labs' Solution for the Access Network

Altice Labs' strategy is aligned with the BBF's CloudCO Reference Architecture [12] and aims to answer network operators' challenges by providing an evolutionary approach with the support of multi-vendor, multi-access technologies and guaranteeing the coexistence between legacy and new-generation devices. This architecture is supported by the AGORA access controller and Altice Lab Access Nodes, and it is depicted in **Figure 2** for PON-based access network scenarios.



**FIGURE 2** – Altice Labs architecture for PON based access networks

AGORA provides a complete all-in-one management solution. It is compatible with BBF-compliant OLTs and with third-party non-BBF-compliant OLTs through a BAA layer. Altice Labs' solution is also compatible with ONF's VOLTHA NBI, allowing the integration of any VOLTHA-compatible white-box OLT.

The AGORA controller exposes an open northbound API that simplifies network operation, leveraging an intent-based approach, and provides easy integration with northbound systems. This API also enables network programmability that speeds up the delivery of new services and allows custom automation.

Altice Labs Access Nodes are designed to be future-proof, either in hardware by providing high

bandwidth capability and support for upcoming new PON technologies or in software by providing a modular architecture that allows offloading some functionalities to run in a cloud environment as defined in TR-477 (Disaggregated OLT) and TR-451 (vOMCI). This distributed architecture is supported by standard interfaces and models.

## AGORA Access Controller

AGORA is Altice Labs' Software Defined Networking Management and Control suite (SDN M&C), composed of software services that provide an all-in-one management solution, fulfilling the network operator's needs for access networks. The main components and features are depicted in **Figure 3**.

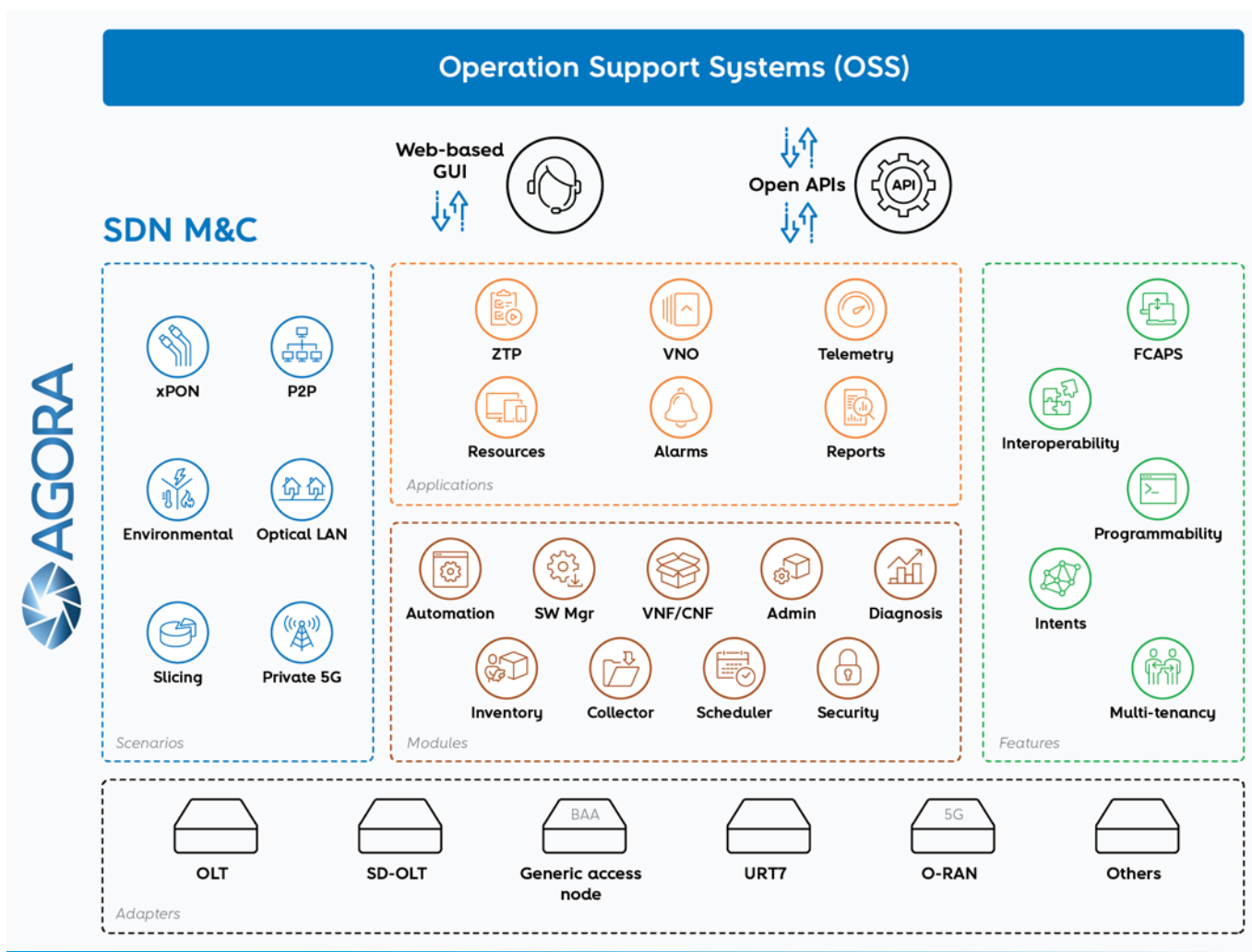


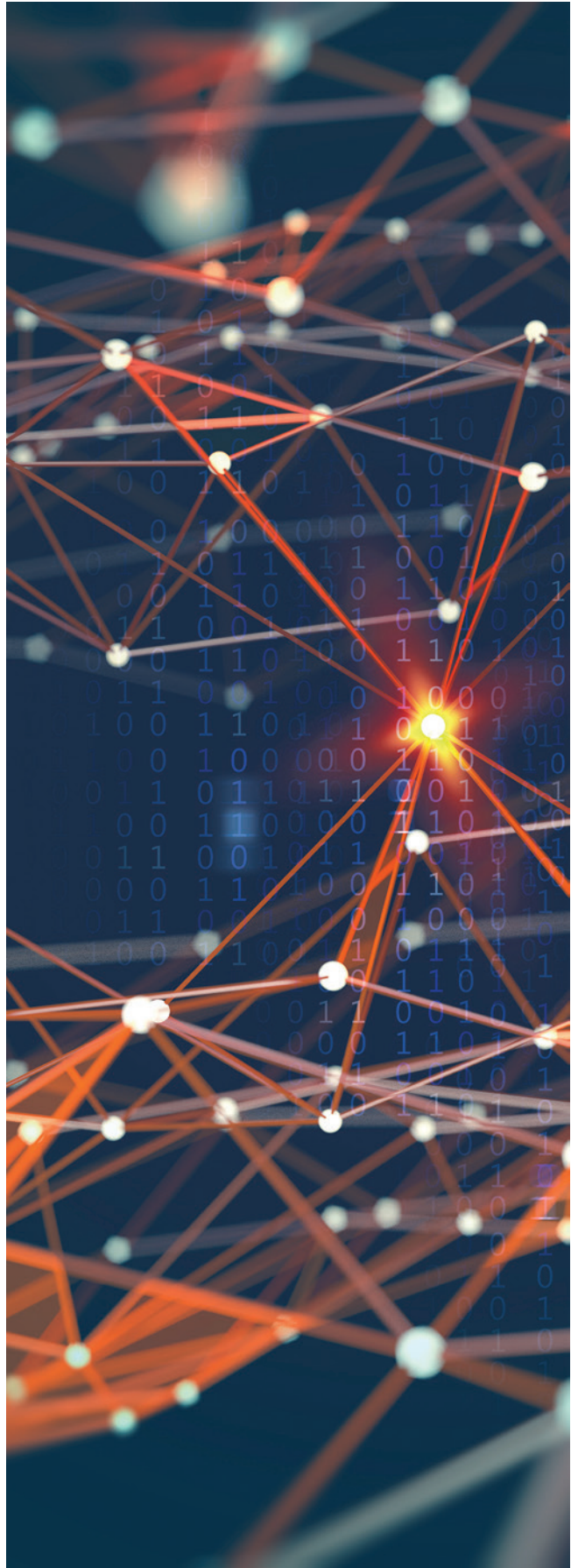
FIGURE 3 – AGORA Access Controller

AGORA's features include network element software (SW) management, service lifecycle management, and monitoring with health check and self-healing mechanisms, as well as high-availability deployments with geo-redundancy. At the same time, AGORA provides a complete FCAPS experience to the network owner:

- **Fault Management** for Status, Events, and Alarms handling and reporting, including alarm filtering, correlation and forwarding;
- **Configuration Management** for Network Discovery, Inventory Management, Resource Configuration, and Service Provisioning, including zero-touch commissioning automation tools;
- **Accounting Management** to track the usage of network resources and services for auditing and usage monitoring purposes;
- **Performance Management** for bulk telemetry data handling and reporting;
- **Security Management** to ensure user management with single sign-on Authentication, granular Authorization control, and proper Auditing reports.

As it was previously mentioned, AGORA provides a full-featured open REST API that abstracts the underlying network, ensures high network programmability, and provides automation and easy integration with northbound systems. This API allows seamless management of all the network elements deployment options, regardless of the OLT protocol/information models, virtualization level, and PON technologies.

For wholesale deployments or any other type of network sharing scenarios, AGORA also provides dedicated services that allow the Infrastructure Provider (InP) to manage and control the slicing of its physical network, and each Virtual Network Operator (VNO) to have their own FCAPS experience through a set of management interfaces. This feature is explained in detail in the article "Sharing network resources between virtual operators in access networks" of this issue of *InnovAction*.



AGORA is a cloud-ready modular solution with many components developed as cloud-native services. As the product evolves, new services are added, and existing services are migrated to a cloud-native approach, increasing the overall flexibility and reliability of the solution. The AGORA solution supports deployment in all types of scenarios, such as bare-metal dedicated servers, Virtual Machines, or private/public clouds, such as Amazon AWS.

## Modular OLT software architecture

The Altice Labs OLTs' software follows a modular approach that enables a smooth migration towards the adoption of Disaggregated OLTs (D-OLT),

compatible with TR-477. The operator can choose between having all the functions embedded (not requiring an additional infrastructure), virtualizing all the functions (benefiting from additional flexibility), or virtualizing only some of the functions that are considered to bring added value at a particular moment (e.g., support new ONU types through an external vOMCI function). Each individual function can be managed and upgraded separately, either as an embedded or virtualized function. Furthermore, an isolated area is available where additional/third-party applications can be deployed as containers, extending the OLT capabilities. The interaction between these applications and the core space is made through dedicated APIs. **Figure 4** depicts this flexibility with the major OLT functional blocks.

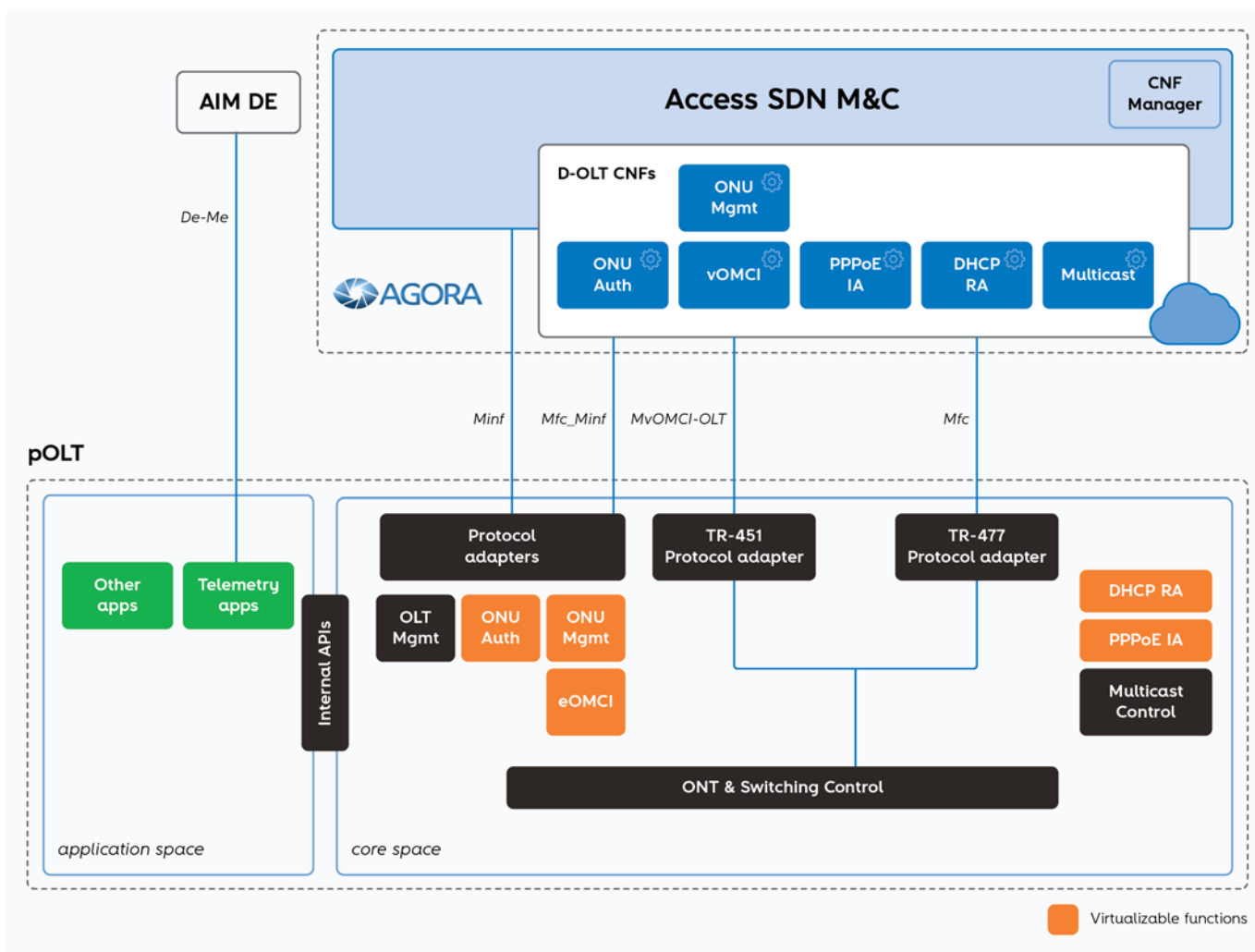


FIGURE 4 – OLT modular software architecture

Most of the network functions defined in TR-477 are supported in both virtualized or embedded mode, such as the DHCP Relay Agent (DHCP RA), the PPPoE Intermediate Agent (PPPoE IA), and the ONU Authentication (ONU Auth). The ONU Management (ONU Mgmt) can also be done using an embedded OMCI (eOMCI) stack or with a TR-451 compliant vOMCI function.

Altice Labs provides implementations for each of the D-OLT virtualized functions, which are designed to be standard-compliant and cloud-native, taking advantage of the scale-in/scale-out mechanism inherent to cloud-based environments.

The pOLT interacts with the virtualized functions through a set of independent protocol adapters that can be upgraded separately in case of standards evolution. The support of the D-OLT functionalities is enabled by an alternate/optional software package supported by Altice Labs' OLTs.

Regardless of the virtualization level that the operator chooses, the AGORA access controller provides a unified management API with consistent workflows. The Cloud-Native Network Function (CNF) Manager allows the operator to easily manage the virtualized functions lifecycle.

An example of third-party applications that can be deployed by the operator in the application space of the OLT are applications that gather telemetry information to feed Automated Intelligent Management Decision Elements (AIM-DE) as defined in TR-486 [18].


## Conclusion

In this article, some of the challenges faced by network operators and how they can be solved

through SDN/NFV technologies were presented. The status of the main reference open-source projects and standards was also summarized, with a focus on the central office. The BBF's CloudCO architecture is being adopted by many operators and vendors due to a well-defined set of interfaces that does not mandate a single implementation and provides a migration path with coexistence between legacy and new generation nodes.

Altice Labs' solution for the access network was designed to leverage the benefits of SDN/NFV technologies with the BBF standards in mind. As previously mentioned, the AGORA SDN M&C provides a unified open API that streamlines the management of all types of access nodes while providing features such as automation, zero-touch operations, and intent-based management.

At last, Altice Labs' access nodes are also designed to provide a modular software architecture that makes functional disaggregation and compliance to BBF standards possible, also on chassis-based OLTs. This gives the operator the liberty to choose the best option to suit their needs and deployment scenario: a fully embedded approach to minimize initial investments or a more flexible approach to run virtual functions in the cloud. Either option is supported by the same hardware and can be changed later by a software update and/or reconfiguration.

Altice Labs is actively contributing to the industry standards and best practices, while continuously evolving its line of products for access networks, making it the right partner for operators looking for a state of the art approach for implementing their network, while safeguarding investments already made. 

## References

- [1] "CloudCO Technical Reports," broadband-forum.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://www.broadband-forum.org/projects/cloud/technical-reports>.
- [2] Saurav Das, "CORD: Leaf-Spine Fabric with Segment Routing," wiki.onosproject.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://wiki.onosproject.org/display/ONOS15/CORD%3A+Leaf-Spine+Fabric+with+Segment+Routing>.
- [3] Open Networking Foundation, "CORD," opennetworking.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://opennetworking.org/cord/>.
- [4] Open Networking Foundation, "SEBA/VOLTHA," opennetworking.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://opennetworking.org/voltha/>.
- [5] Open Networking Foundation, "OpenFlow Switch Specification Version 1.5.1," TS-025, 2015.
- [6] A. Campanella, "SEBA/VOLTHA: SDN-Enabled Broadband Access Projects – A Year in Review (Part 1)," opennetworking.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://opennetworking.org/news-and-events/blog/seba-voltha-sdn-enabled-broadband-access-projects-a-year-in-review-part-1/>.
- [7] Open Networking Foundation, "VOLTHA 2.8 Release Notes," docs.voltha.org. Accessed: Sept. 20, 2024. [Online]. Available: [https://docs.voltha.org/master/release\\_notes/voltha\\_2.8.html](https://docs.voltha.org/master/release_notes/voltha_2.8.html).
- [8] Open Networking Foundation, "VOLTHA 2.10 Release Notes," docs.voltha.org. Accessed : Sept. 20, 2024. [Online]. Available: [https://docs.voltha.org/master/release\\_notes/voltha\\_2.10.html](https://docs.voltha.org/master/release_notes/voltha_2.10.html).
- [9] D.-F. Huang, C. Chen, and M. Thanavel, "Fast packet classification on OpenFlow switches using multiple R\*-tree based bitmap intersection," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, Taipei, Taiwan, 2018, pp. 1–9. doi: 10.1109/NOMS.2018.8406227.
- [10] The Linux Foundation, "ONF Merges Market Leading Portfolio of Open Source Networking Projects into the Linux Foundation," linuxfoundation.org. Accessed: Sept. 20, 2024. [Online]. Available: <https://www.linuxfoundation.org/press/onf-merges-market-leading-portfolio-of-networking-projects-into-linux-foundation>.
- [11] "TR-384: Cloud Central Office (CloudCO) Reference Architectural Framework," BBF, Fremont, California, USA, Rep. TR-380, 2018.
- [12] A. Brízido, G. Gaspar, M. Santos, R. Calé, and V. Mirones, "Redesigning the network edge for a new era," *InnovAction*, pp. 102-115, 2020.
- [13] "TR-408 Cloud CO Migration and Coexistence," BBF, Fremont, California, USA, Rep. TR-408, 2020.
- [14] "TR-477 Cloud CO Enhancement - Access Node Functional Disaggregation," BBF, Fremont, California, USA, Rep. TR-477, 2024.
- [15] "TR-484 Access Network Abstraction," BBF, Fremont, California, USA, Rep. TR-484, 2022.
- [16] "TR-451 vOMCI Specification," BBF, Fremont, California, USA, Rep. TR-451, 2022.

- [17] International Telecommunication Union Standardization Sector, "ITU-T G.988 ONU management and control interface (OMCI) specification," 2022.
- [18] "TR-486 Interfaces for AIM," BBF, Fremont, California, USA, Rep. TR-486, 2023.
- [19] "TR-436 Access & Home Network O&M Automation/Intelligence," BBF, Fremont, California, USA, Rep. TR-436, 2021.

# 04

## Sharing network resources between virtual operators in access networks



In today's world, traffic demand continues confronting network operators' ability to scale and grow physical networks at a rate befitting that demand, without incurring large capital expenditure (CapEx) investments and greater operational complexity. This is challenging because operators cannot scale in an efficient way and achieve their full revenue potential. Another problem is the necessity of compliance with an environment of greater regulatory burden imposed by regulators that require increasing competitiveness between operators.

The Fixed Access Network Sharing (FANS) paradigm aims to resolve those issues by providing a framework that allows usage optimization of the underlying network infrastructure, now shared by different Virtual Network Operators (VNO). This paradigm also raises some interesting questions, such as how the Infrastructure Provider (InP), owning the physical shared resources, can properly manage those in a way that they are abstracted into a virtual network layer, allowing each VNO to operate on it, without affecting others VNOs and allowing physical network growth more efficiently.

Altice Labs is tackling these issues with its own FANS solution, which is part of the AGORA suite of products. This solution also has the added value of incorporating network provision automation and providing a simplified API for both InP and VNO to manage the supporting physical and virtual network resources, respectively. In this article, the overall architecture of the solution is presented, some technical challenges are identified, and a demonstration of both InP and VNO service managers is provided.

---

## Authors

**André Rodrigues Almeida, Altice Labs, [andre-r-almeida@alticelabs.com](mailto:andre-r-almeida@alticelabs.com)**

**André Domingos Brízido, Altice Labs, [andre-d-brizado@alticelabs.com](mailto:andre-d-brizado@alticelabs.com)**

**Nuno Ricardo Farinha, Altice Labs, [nuno-r-farinha@alticelabs.com](mailto:nuno-r-farinha@alticelabs.com)**

---

## Keywords

InP; VNO; Network Sharing; AGORA



## Introduction

Traditional fixed broadband access networks deployment models often lack the flexibility to adapt to changing needs and traffic patterns, making it difficult to scale incrementally as demand grows. With such growth and demand, telecommunications companies (TELCO) need to adapt and evolve their network infrastructure for a rapidly changing and evolving digital landscape. These changes entail the need for increased traffic rates and strict availability requirements that only next-generation Passive Optical Network (PON) technology can provide. Consequently, to expand and maintain this network infrastructure and to support the increasing number of subscribers, greater investments are needed. With the rise in costs, also arises the need to increase the efficiency of utilization of access network resources, namely Optical Line Terminators (OLT) and fiber connections. The solution comes in the form of FANS, which allows for the emergence of VNO taking advantage of network virtualization techniques.

The AGORA Access Controller provides a complete network management solution for a simpler centralized access network operation of devices developed by Altice Labs. It provides different modules, each with its own purpose and functionalities, for a more scalable and efficient way of solving a particular problem that may arise from network operators. Some of the most recent AGORA's modules support network virtualization, providing the ability of an InP to manage multiple VNOs, each with its own virtual network on top of the same shared physical network infrastructure.

In this article, the motivation and challenges of InP operators to support network virtualization and network sharing models are explored. Altice Labs' solution for the abovementioned challenges is also described, along with the advantages of its use over traditional broadband network deployments.

## The road to network sharing through virtualization

### Business perspective

The number of Internet accesses is constantly growing and, as such, more and more fixed broadband access connections are needed to satisfy that demand. It is predicted that PON infrastructure for Fiber To The Home (FTTH) deployments, cable broadband access equipment, and fixed wireless Customer Premises Equipment (CPE) will all increase until 2026 [1]. All these predictions reveal a scenario where traditional TELCO companies need to invest large amounts of money to grow their networks. On the other hand, more and more strict regulations are being enforced by developed countries to increase competition between currently established players, and incentives are provided to facilitate the entry of new players into the market. In fact, according to the Organization for Economic Co-operation and Development (OECD), a scrutiny and regulatory environment has been developing in these countries over the last few years. This is due to the wave of Mergers and Acquisitions (M&A) that occurred in these countries in the last decade between fixed and mobile network operators, resulting in bigger companies in the market, each possessing a more concentrated network infrastructure. Still, this process was done with some safeguards imposed by regulators, as a requirement for those mergers to be successful. To increase competition in the services provided in this new telecommunication landscape, most regulators demanded commitments from these operators to create conditions for the entrance of VNOs as new players [2]. With all these hurdles, a new business model emerged, in which these incoming players enter into an agreement with network providers to access network resources at wholesale rate, without the need to own the underlying network infrastructure. This can be

achieved through a virtual network abstraction of the underlying physical resources (hence the designation of “Virtual Networks Operators”) and the leverage of FANS model. The FANS model comes with a myriad of benefits that extend beyond the conventional telecommunications approach, as shown in **Figure 1**.



**FIGURE 1** – FANS model benefits

Telco businesses can achieve significant cost efficiency by avoiding the substantial capital investment (CapEx) required for building and maintaining network infrastructures. This allows a more efficient resource allocation, allowing time and resources to be more focused on innovation for the development of more customer-centric solutions. It allows for easier market accessibility that fosters a more diverse and dynamic market, which in turn leads to healthy competition by providing consumers with a range of choices and tailored services that lead to more competitive pricing and innovation in service offerings and customer experiences. It offers much more agility and flexibility, enabling businesses to adapt swiftly to market demands, whether it be scaling operations more quickly and with less effort or pivoting service offerings to be more in line with the market reality of the moment.

VNOs also have the option of specializing in specific market segments, offering more personalized and niche services that cater to unique consumer needs. Finally, this model also facilitates risk mitigation, allowing companies to enter the market without the high risk imposed by the heavy infrastructure investments and to experiment with more refined strategies based on a loopback feedback effect of market response, which allows for more informed decisions being made before committing to substantial capital investments.

## Technical perspective

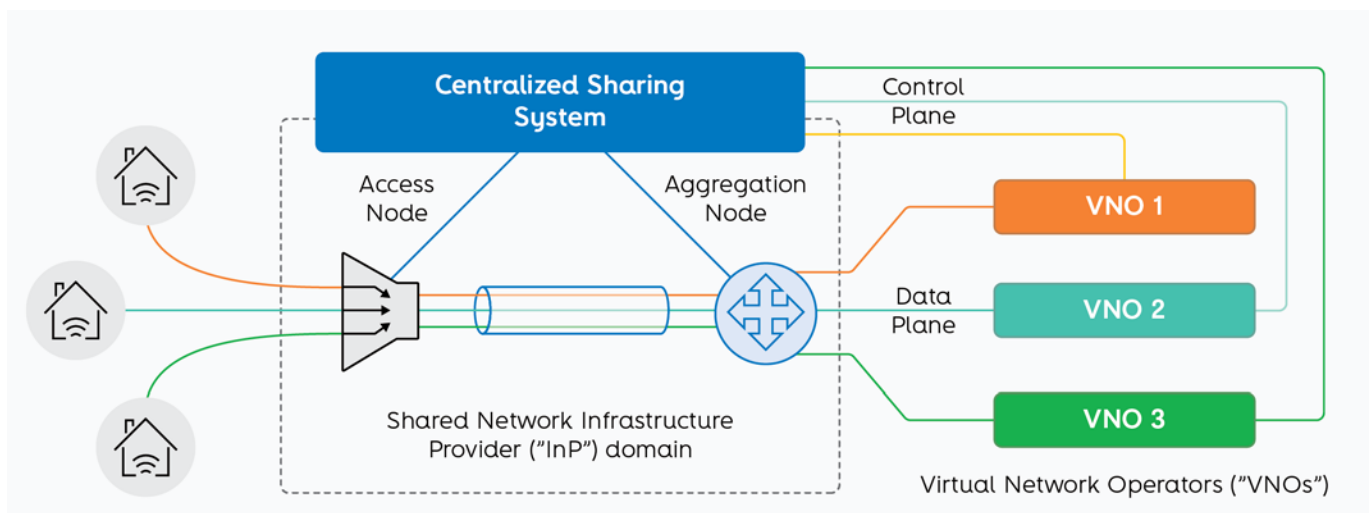
### FANS paradigm

From a technical point of view, the transition from a traditional fixed network architecture to a virtual architecture, which empowers VNOs with enough features and capabilities to manage their own virtual network, is a very difficult endeavor. One entity that dedicates itself to addressing these challenges is the Broadband Forum (BBF), which aims to accelerate the growth and shape of broadband technology by providing technical specifications and standards for the industry [3]. Some of the most important contributions on the topic of fixed broadband access networks sharing, defined by BBF in the TR-370 [4] and TR-386 standards [5], are as follows:

- **Centralized System model:** where a centralized system is responsible for enabling physical access resource sharing amongst several VNOs and mediating communication performed by it to the physical access node. This model does not require any changes to the existing network deployed by the InP, and it works well with both legacy and new types of network devices;
- **Virtual Node model:** designed to enable the partitioning of the InP's physical access node into multiple virtual access nodes, each associated to one VNO. In contrast with the previous model, each virtual access node is responsible for resource management control for each VNO in a decentralized way. However, this mode requires Multiprotocol Label Switch (MPLS) tunneling, VxLAN tunneling, or VLAN tunneling with Q-in-Q up to a stack of three VLANs [4], features that not all types of access nodes may support.

For the purpose of this article, the FANS Centralized System model is the one that interests us the most, and **Figure 2** illustrates its main components.

This figure illustrates the physical domain belonging to the InP, hosting the physical network to be shared by "slices" and the virtual domain, with the virtual network of each VNO on top of the physical network. The represented Centralized Sharing System (CSS) is the core of the FANS sharing



**FIGURE 2** – Network sharing slices [6]

model, enabling the InP to execute network partitioning and logically divide the resources between the VNOs. This management system supports multi-tenancy, where each VNO is a separate tenant. Using this sharing technique, it is possible to separate the functions and responsibilities of the control and management plane from the data plane. Some functions reserved to the control plane consist in resource allocation management of virtual network elements, assigning and mapping them to physical resources, mediating backhaul bandwidth allocation through quota allocation, and diagnostic signaling. On the other hand, the data plane at the physical network level remains unchanged and is responsible for ensuring packet forwarding across the network and traffic segregation between different operators through service VLANs, and quality of service (QoS) assurance.

The CSS glues the previously discussed FANS use cases and actors together by providing a

management interface to the InP, a southbound interface to control network devices, and a northbound management interface to each VNO. Required configuration and operation actions can be executed by the InP on behalf of the VNO, or the VNOs may perform those operations by themselves, depending on their capacity to do so. It can also support the automation of certain operations between the InP and VNOs, such as the configuration of customers' Optical Network unit (ONU) and service provision. Finally, the CSS also needs to guarantee that any VNO user or customer cannot access any data related to a different VNO and that all resources will be shared in accordance with regulatory constraints and contractual agreements [6].

Figure 3 illustrates the physical domain network provided by InP, sliced by two distinct VNOs in their virtual domain.

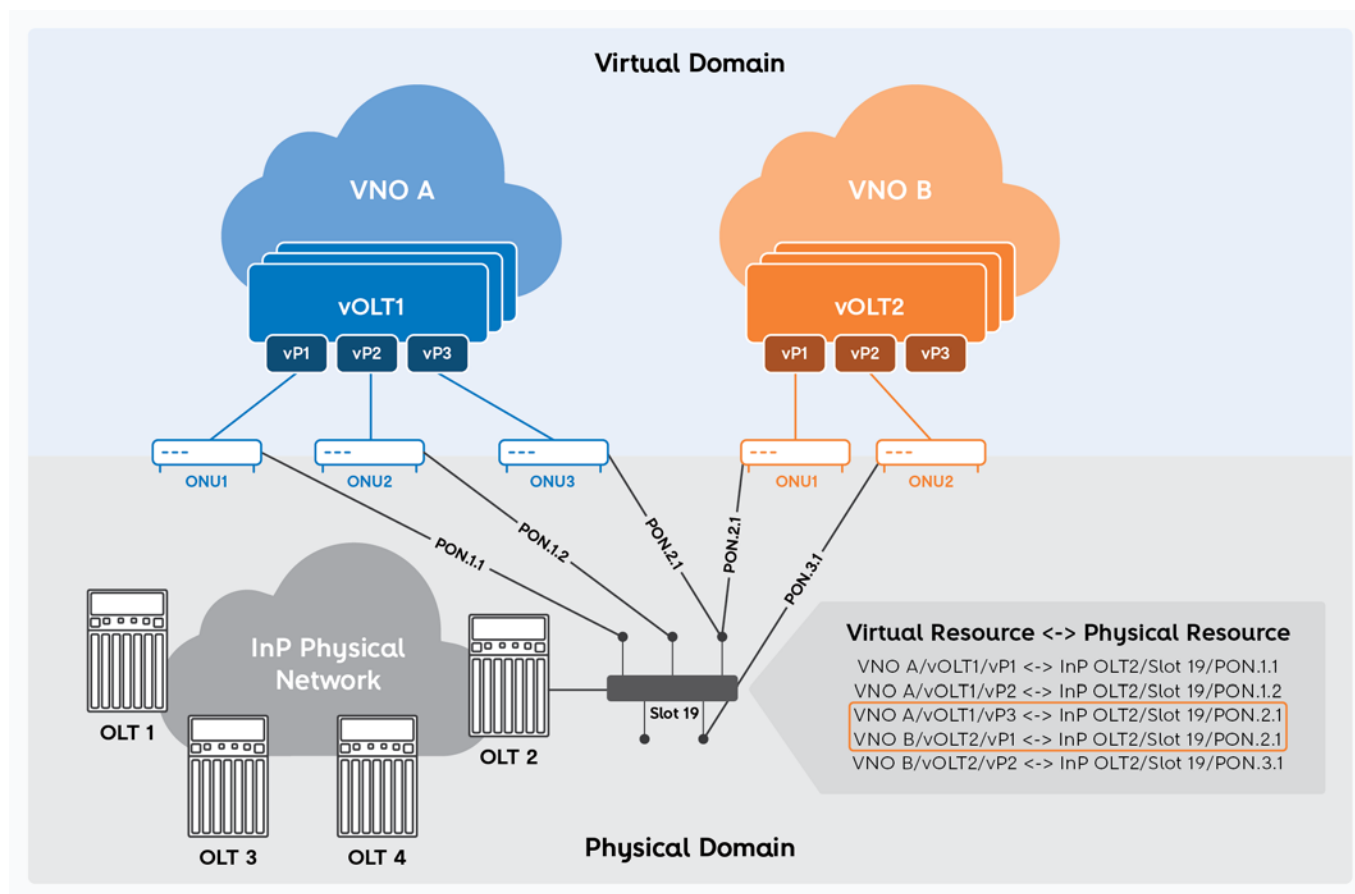


FIGURE 3 – Physical network slice

In this example, the same physical port, in this case PON.2.1, is shared between two distinct VNOs (A and B). Both VNO A and VNO B are using a slice of the resources of the physical OLT2, seeing them as virtual vOLT1 and vOLT2, respectively. Similarly, the physical PON port slices are mapped as virtual vPort3 in vOLT1 (VNO A) and vPort1 in vOLT2 (VNO B).

In summary, these are the main responsibilities of each stakeholder in a network system that follows the FANS centralized model:



- Leases its physical infrastructure to VNOs;
- Enables physical resource sharing and carries out the required partitioning, mapping it to virtual resources;
- Ensures backhaul bandwidth mediation through quota allocation to optimize resources;
- Guarantees security and privacy between different VNO users;
- Provides a northbound interface to the VNO for data and control.



- Uses the network resources provided by InP;
- Operates and manages its own virtual network;
- Can provide specific tailored services through its own network to the end-user;
- Virtual network monitoring through alarmistic and reports systems.

## QoS assurance

A relevant aspect to take into consideration is the presence of mechanisms that allow each VNO to meet certain levels of QoS assurance to its end users. For a VNO to be able to provide customized services, multiple differentiated service level agreements (SLA) must be in place, depending on the scenario it is addressing. In one scenario, the VNO may want to cover real-time applications, such as VoIP and video conferencing. For that, low packet loss, low latency, and low jitter are important QoS metrics. In other scenarios, the VNO may want to prioritize uptime and availability for critical applications, while others, like non-critical services such as residential broadband, may prioritize capacity utilization for resource optimization. There are also those that may prioritize network scalability to handle network growth quickly and effortlessly when there is a traffic demand increase.

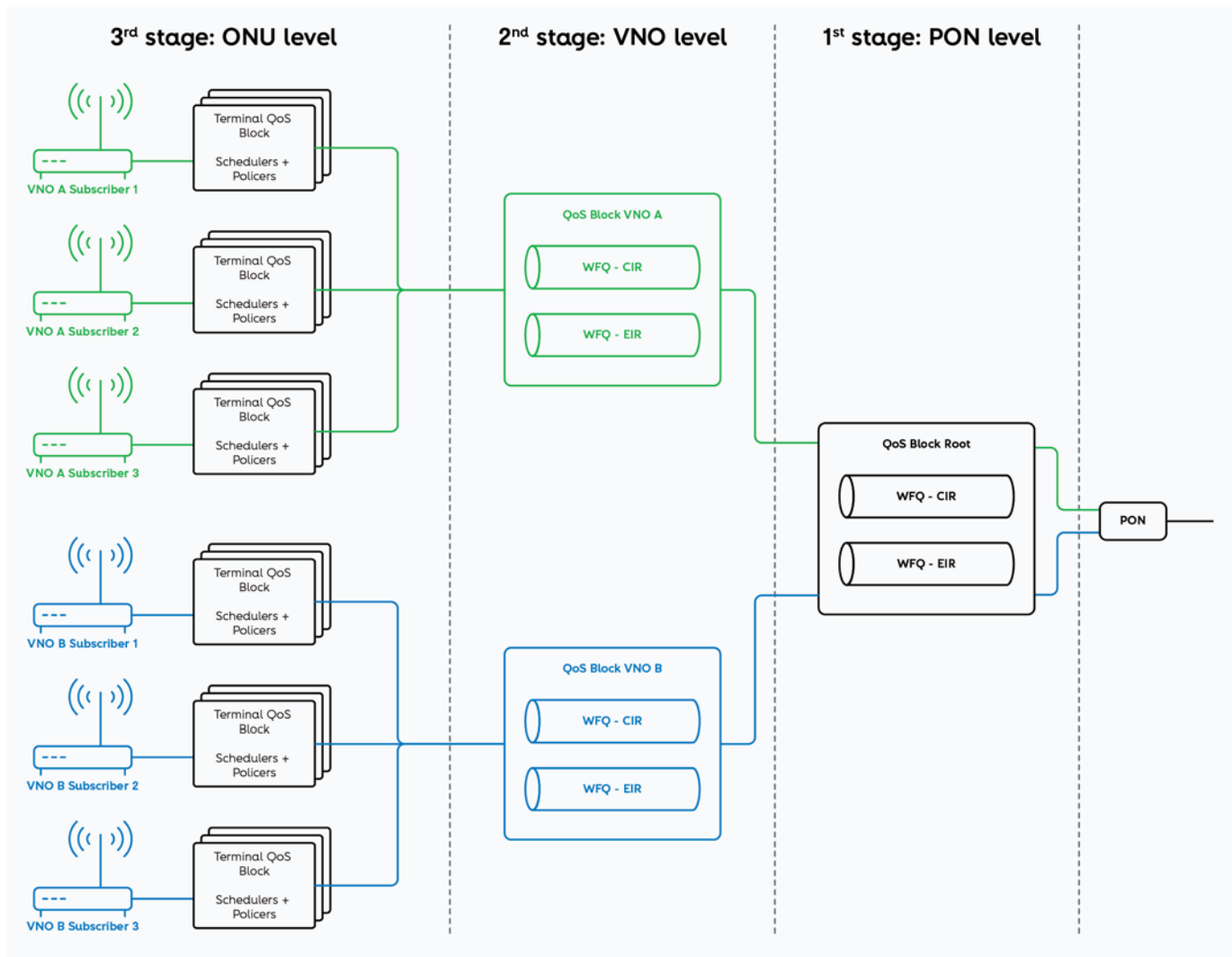
The CSS must be able to handle these vast myriads of scenarios by enabling different QoS parametrizations during service provision so that they can be enforced in an isolated way for each VNO, according to its unique use case. For PON, enforcing QoS compliance levels in the upstream transmission is easier, since the mechanism of upstream dynamic bandwidth allocation (DBA) uses time division multiplexing (TDM) as a fundamental way of transmitting upstream data. This allows for efficient sharing of the upstream channel with predictive and fair opportunities for the different ONUs based on their traffic demand. Guaranteeing in a predictive way, QoS levels for downstream transmission are more complicated, however. Because of the nature of broadband networks, particularly in Digital Subscriber Line (DSL) and fiber optic networks, downstream transmission often utilizes statistical multiplexing to handle varying data rates, handle traffic patterns efficiently, minimize wasted capacity, and maximize the use of available bandwidth [7]. The drawback of this method is that we cannot easily guarantee consistent data information rate, packet loss, and latency.

In traditional fixed access networks, QoS is typically achieved through appropriate queue management

that separates traffic and schedulers that police incoming traffic rates. However, sharing the available InP resources amongst the several VNOs adds another layer of complexity, and the mentioned QoS management methods are no longer sufficient, because it is also required to separate and isolate incoming traffic from different VNOs. To solve this problem, new methodologies emerged, such as the three-stage Hierarchical Quality of Service (H-QoS) [8].

AltiCe Labs OLT devices have their own H-QoS implementation, which our CSS leverages to guarantee minimum QoS levels per VNO in the data plane level. **Figure 4** illustrates a generic three-stage hierarchical QoS solution.

This example shows downstream transmission traffic segregated by VNA A in green color and VNO B in blue color. This architecture allows for rate limits to be established in three different stages, such as Committed Information Rate (CIR), as minimum bandwidth that is assured, and Excess Information Rate (EIR), as remaining bandwidth that can be used in competition as a first come first served basis. The first stage is done at the PON level, where aggregated rate limits can be established based on the maximum capacity of the PON port. The second stage is done at the VNO level, and it is here that we can guarantee minimum bandwidth capacity, which allows QoS assurance per VNO. Both levels are controlled and managed by the InP, where rate limits are configured, considering



**FIGURE 4** - Three-stage H-QoS architecture

the contract established with each VNO. Finally, the last stage is configured at the ONU level, and it is controlled by the VNO, where it has the freedom to use appropriate scheduler and policer function combinations, such as increasing and decreasing priorities, shaping the traffic delivered to its final subscribers, this way covering a multitude of traffic use cases, including data, control, signaling, voice, video, and others.

## AGORA FANS virtualization solution

As stated before, there are several challenges that need to be surpassed when offering access network sharing management solutions to VNOs, particularly when the InP already has legacy devices deployed in the physical network, and it is essential to reconcile the management of such devices with new ones, supporting new PON technologies. This chapter describes Altice Labs' own implementation of the FANS model, including the overall high-level architecture of the system and the actions to be executed by the InP and VNO users during their operations.

### InP and VNO systems

Altice Labs developed a new centralized service manager, named VNO Manager, that runs on top of its Access Controller solution called "Resource Manager". It has also developed a complete VNO system with its own services that allow the VNO to manage and monitor its virtual environment. **Figure 5** describes, with some detail, the InP and VNO architecture and some of the main functions and responsibilities of these two different system stakeholders.

The InP system includes the following service managers that assure the operationalization of the FANS model for the VNO:

- **VNO Manager:** this is a central service, essential for this model to work. It acts as an

orchestrator, where the InP is able, through a graphical or REST interface, to leverage the following functions:

- Manage VNO entities: create new ones, configure, and monitor them during their lifecycle. It also allows the InP to establish, for each VNO, different levels of restriction policies for its users;
- Partition of the physical network resources, such as OLT equipment and PON ports, in different logical slices to be used by each VNO in their respective virtual networks as vOLT and vPorts;
- Manage VNO's ONU devices, planning them before field installation. When those ONUs are installed, notifications generated and received by the Alarm Monitor are used to automatically detect and register that ONU in the system. There are automation parametrizations at the VNO level that automate the ONU assignment to the respective network slice;
- Enable the creation of service intent rules, which can be used by intent workers to automate the translation of high-level business policies into low-level network configurations;
- Support of optional ONU quota management, allowing the InP to impose limits on the VNO slice size, i.e., on the number of ONUs each VNO can activate in the same PON port:
  - A minimum committed value for each VNO can be defined, depending on the maximum split ratio capacity of the physical port, ensuring a minimum number of ONUs for a specific VNO on that shared PON;
  - An excess number can also be defined, allowing each VNO to gain new ONU slots in the PON, in competition with other VNOs. These values must be configured in compliance with pre-agreed negotiations between InP and VNO;

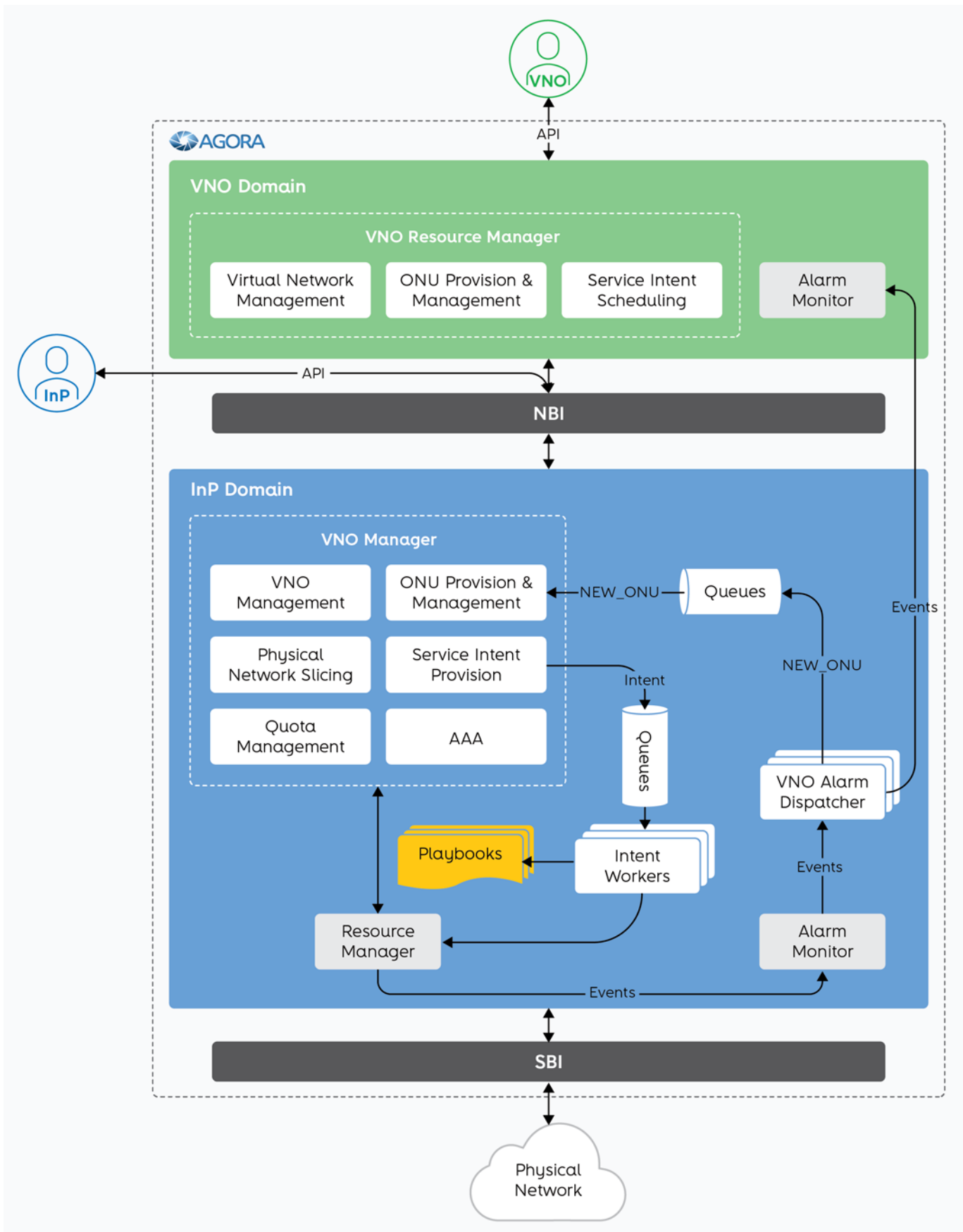


FIGURE 5 – High-level architecture of the InP and VNO systems

- These configurations can be done at three different levels. The first two are the VNO and OLT levels, where these configurations act as default values to be applied to each new vPort created in the network, without affecting the current managed ONU capacity already in use by the VNO. The last level allows for a finer control by overriding previous configurations at vPort level. This enables the resource optimization of the physical network, since the PON can be used to its full capacity.
- Finally, InP is also responsible for enforcing authentication, authorization, and accountability (AAA) functions, to ensure the privacy of each VNO user, in its own domain, from other VNO users. In VNO Manager, the InP can create new VNO users, managing their respective accounts by blocking them and monitoring them.
- **Intent Workers:** modules that can be horizontally scaled, used to process service intents scheduled by the VNO user and serviced by a queue. These service intent events contain information that matches a personalized playbook, tailored for the specific VNO that originated those intents. These playbooks contain all the detailed network configurations that meet the VNO-specific requirements to provide connectivity services to their end users. The scope of this service is not limited to service intents processing, with the possibility to be expanded and include other type of intents;
- **VNO Alarm Dispatcher:** is an isolated module that, like the previous one, can be horizontally scaled and is responsible for these functions:
  - Receive alarm events from the physical network;
  - Filter alarms by specific VNO virtual network;
  - Translate the alarms to a simplified virtual alarm model that can be easily understood by each VNO;
  - Finally, it is responsible for forwarding those virtual alarms to the alarm monitor of the VNO domain.

The VNO domain is much simpler and nimbler. Each VNO user can operate two services, through their respective graphical and REST interfaces:

- **A VNO Resource Manager:** which is the virtual counterpart of the management system of the InP, in which the VNO user has the capacity to operate a subset of the functionality the InP has. These functions are as follows:
  - Monitoring and administratively managing all virtual resources, such as vOLTs and vPorts;
  - Provide the same functions as the InP regarding ONU provision and ONU registration;
  - Finally, it can schedule the provision of its own service intents, previously negotiated and delivered by the InP, and monitor the progress of that provision and the result of that schedule.
- **An Alarm monitor:** which allows the VNO user to monitor opened alarms that affect its virtual network, in a simplified manner. Those alarms correspond to the translated alarms forwarded by InP VNO Alarm Dispatcher.

**Figure 6** illustrates the graphical interface of the InP VNO Manager service:

Name	Administrative state	Additional information	Creation time	Update time	Auto-provision	vOLTs	vPorts	vOLTs	Actions
Adriano	In service	---	19/09/2022 12:10:50 PM	19/09/2022 12:10:50 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adriano	In service	---	19/09/2022 12:21:40 PM	19/09/2022 12:21:40 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Argon	In service	Empty Scenario	19/09/2022 2:49:28 PM	19/09/2022 2:49:28 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bardhan	In service	---	19/09/2022 12:22:52 PM	19/09/2022 12:22:52 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bardhan	In service	LDN Scenario	19/09/2022 4:17:45 PM	19/09/2022 10:52:30 AM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Buenos	In service	---	19/09/2022 11:52:23 AM	19/09/2022 11:52:23 AM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Carlson	In service	---	19/09/2022 12:12:26 PM	19/09/2022 12:12:26 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Carlsson	In service	---	19/09/2022 12:22:16 PM	19/09/2022 12:22:16 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Carlson	In service	Operator contact: 023865789	19/09/2022 2:17:14 PM	19/09/2022 2:17:14 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Carlson	In service	---	19/09/2022 12:17:50 PM	19/09/2022 12:17:50 PM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**FIGURE 6** - InP graphical interface

The first menu allows the InP an overall view of all the VNOs it can manage. For each VNO, there is a dedicated, detailed page where the InP can

manage the corresponding network slice from a top-down view, i.e., creating the virtual network elements first and then mapping them to the physical resources it wants to slice. The option of slicing the network from a bottom-up perspective is also possible, making it easier for the InP to perform the network slicing in a massive way for several VNOs without the need to do it one at a time.

Figure 7 illustrates the graphical interface of the VNO Resource Manager service:

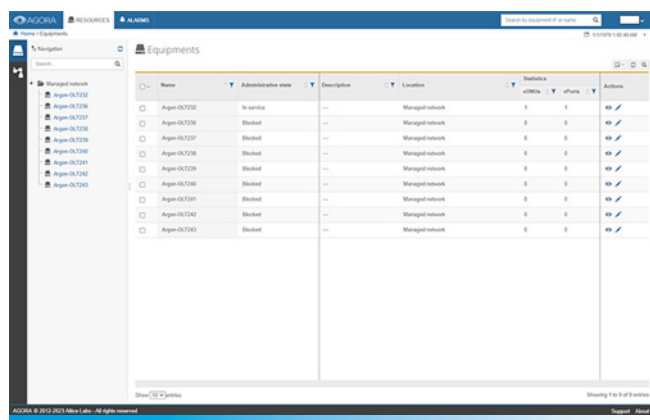


FIGURE 7 – VNO graphical interface

Here, every VNO can operate and monitor its virtual network and perform all the functions already described.

## InP and VNO operational overview

After detailing the implemented architecture of the solution and its main functions, here it is described the required operational procedures realized by both InP and VNO. The respective workflow is illustrated in Figure 8, being visible this can be executed in different temporal phases.

The pre-provision phase occurs when the InP does much of the preliminary work, such as VNO creation and configuration, quota parametrization, provisioning planned ONUs by their respective serial number, creation of service intents and their respective playbooks, and configuration of VNO

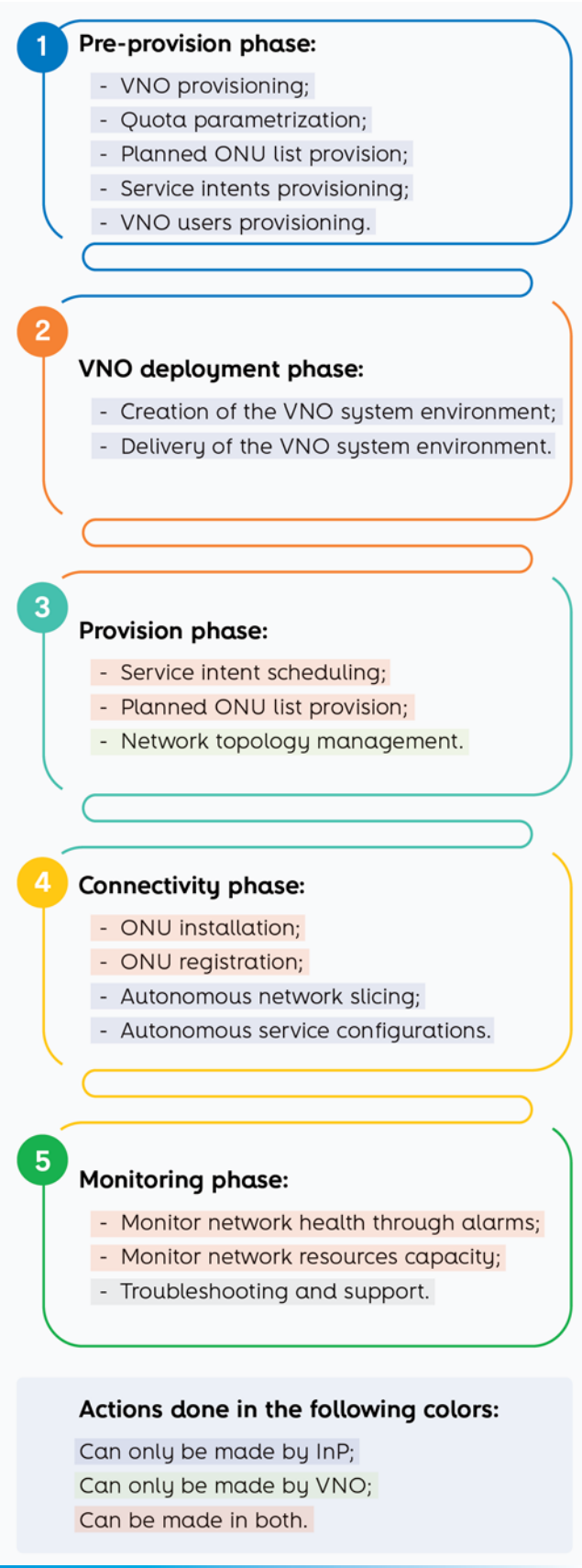



FIGURE 8 – Actions performed during the various operational phases

users. In the next phase, InP creates and delivers the system environment to the VNO, so that it can start its operations. Because most of the things the VNO can do, the InP can do too, the provision phase entails a mix of actions done by the two different stakeholders of the system. As such, both InP and VNO can trigger service intent scheduling in this phase.

The connectivity phase entails the installation of the ONU at the end user premises, in which the serial number is then matched with one of the planned ONU lists previously provisioned. In this stage, the necessary autonomous processes of the VNO Manager are responsible for performing the network slicing, where virtual resources are created and mapped on top of physical resources, and triggering the event for service intent workers to do the necessary network level configurations so that the connectivity service can be established to the end user. The final phase of these operational procedures is a more passive one, where both stakeholders monitor the health and capacity of the system and act upon it, if necessary.

## Wrapping-up

As a final remark, we are confident that the proposed CSS solution, supported by Altice Labs' AGORA, provides the means to answer the most important issues raised by the FANS model specification defined by BBF. The possibility of providing the VNO the ability to enter the market without the need for chassis-based solutions to connect network subscribers is satisfied by providing vOLTs and vPorts mapping on a physical slice, removing the constraint the traditional operators have to incrementally scale their operations because of the fixed size port count of traditional FTTH networks. This, in turn, allows VNOs to be more cost-efficient and to scale their virtual networks as demand grows. Furthermore, the solution is compatible with existing FTTH deployments since it does not require special device capabilities to perform network sharing, but it can leverage advanced features such as H-QoS to provide better control over the network resources.

Finally, this solution also provides an interface that allows each VNO the autonomy to manage its slice of the physical network in a centralized way. 



## References

- [1] "2022.02.04 - Open Broadband News." broadband-forum.org. Accessed: Aug. 02, 2024. [Online]. Available: <https://www.broadband-forum.org/2022-02-04-open-broadband-news?cn-reloaded=1>.
- [2] OECD, "Emerging trends in communication market competition," *OECD Digital Economy Papers*, no. 316, Sep. 2021, doi: 10.1787/4ad9d924-en.
- [3] "About BBF." broadband-forum.org. Accessed: Aug. 05, 2024. [Online]. Available: <https://www.broadband-forum.org/about-bbf>.
- [4] "TR-370 Issue 2: Fixed Access Network Sharing - Architecture and Nodal Requirements," *BBF*, Fremont, California, USA, Rep. TR-301i2a3, 2024.
- [5] "TR-386: Fixed Access Network Sharing - Access Network Sharing Interfaces," *BBF*, Fremont, California, USA, Rep. TR-386i2, 2024.
- [6] "MR-432: Fixed Access Network Sharing (FANS)," *BBF*, Fremont, California, USA, Rep. MR-432,2018.
- [7] J. Huang, C. W. Tan, M. Chiang, and R. Cendrillon, "Statistical Multiplexing Over DSL Networks," in *IEEE INFOCOM 2007 - 26<sup>th</sup> IEEE International Conference on Computer Communications*, Anchorage, AK, USA, 2007, pp. 571-579, doi: 10.1109/INFOCOM.2007.73.
- [8] F. Slyne, B. Cornaglia, M. Boselli, and M. Ruffini, "3-Stage Hierarchical Quality of Service for Multi-tenant Passive Optical Networks," in *Optical Network Design and Modeling: 23<sup>rd</sup> IFIP WG 6.10 International Conference*, Athens, Greece, 2019, pp. 193-203, doi: 10.1007/978-3-030-38085-4\_17.

# 05

## Overbooking Maneuver solution for OTBs



Achieving a uniform and controlled occupation of the network through constant updating of the inventory of clients and Optical Terminal Box (OTB) ports is a challenge, especially considering the high adoption of Fiber-to-the-Home (FTTH) networks, which demands control of port occupancy, their availability for new activations, and more precise control of port inventory. Given this fact, Fibrasil's Optical Terminal Box (OTB) Overbooking Maneuver solution, developed by Fibrasil and powered by Open Labs' NOSSIS suite, provides more autonomy to field technicians, allowing them, in situations of high occupancy, to seek installation alternatives in a controlled manner and keeping the port occupancy in the inventory up to date.

---

## Authors

**Alessandro Gentil, FiBrasil, [alessandro.gentil@fibrasil.com.br](mailto:alessandro.gentil@fibrasil.com.br)**

**Alexandre Oberg Pereira da Cruz, FiBrasil, [alexandre.oberg@fibrasil.com.br](mailto:alexandre.oberg@fibrasil.com.br)**

**Atila Araujo Branco, FiBrasil, [atilaab@fibrasil.com.br](mailto:atilaab@fibrasil.com.br)**

**Daniel Soeiro dos Santos, FiBrasil, [daniel.soeiro@fibrasil.com.br](mailto:daniel.soeiro@fibrasil.com.br)**

**Thalles Santana da Silva, FiBrasil, [thalles.silva@fibrasil.com.br](mailto:thalles.silva@fibrasil.com.br)**

---

## Keywords

OTB; OLT; Neutral Networks; GPON; API



## Introduction

The adoption of neutral FTTH [1] networks is advancing rapidly worldwide, reflecting a growing appreciation for high-speed connectivity and telecommunications infrastructure. In Brazil, for example, there has been a notable reduction in network construction volumes and an increase in the adoption of neutral FTTH networks by fiber broadband service providers.

The Neutral Network business depends on the following key success factors:

- **High-quality infrastructure:** a highly available, secure, controlled, and managed network;
- **Accessibility and neutrality:** ensuring that the network is accessible to different service providers and end users, regardless of their size or location, promotes a competitive and inclusive environment;
- **Sustainable business model:** having a financial model that allows the operation and maintenance of the network in a sustainable manner is crucial. This may include fair access tariffs, strategic partnerships, and efficient resource management;
- **Customer service and support:** providing effective technical support and good customer service to service providers and end users is important for resolving problems quickly and maintaining satisfaction;
- **Operational efficiency:** operating efficiently, with well-defined processes and effective resource management, contributes to reducing costs and improving service quality.

Ensuring maximum efficiency in the use of network resources, allowing consistent resource utilization, can help to create a successful neutral network that benefits all market participants. In this article, we propose a set of functionalities to control port occupancy, their availability for new activations, and more precise control of port inventory.

We will call this set of functionalities “Overbooking Maneuvers.” The objective is to maintain the network with the best ratio of maximum occupancy to maximum port availability, with the decision-making power over port use being shared by the inventory system and field technicians.

FiBrasil's OTB Overbooking Maneuver solution, powered by Open Labs' NOSSIS suite, provides more autonomy to field technicians, allowing them, in high occupancy situations, to seek installation alternatives in a controlled manner and keeping the port occupancy in the inventory up to date.

## Macro description of the solution

A major challenge for broadband service providers (ISPs) and neutral Gigabit Passive Optical Network (GPON) fiber networks is to achieve uniform and controlled network occupancy through constant updating of the inventory of customers and ports. This challenge is even greater in standard Passive Optical Network (PON) networks, since such networks do not offer native port occupancy management solutions.

The Overbooking Maneuver solution (**Figure 1**) allows field technicians to choose the best OTB

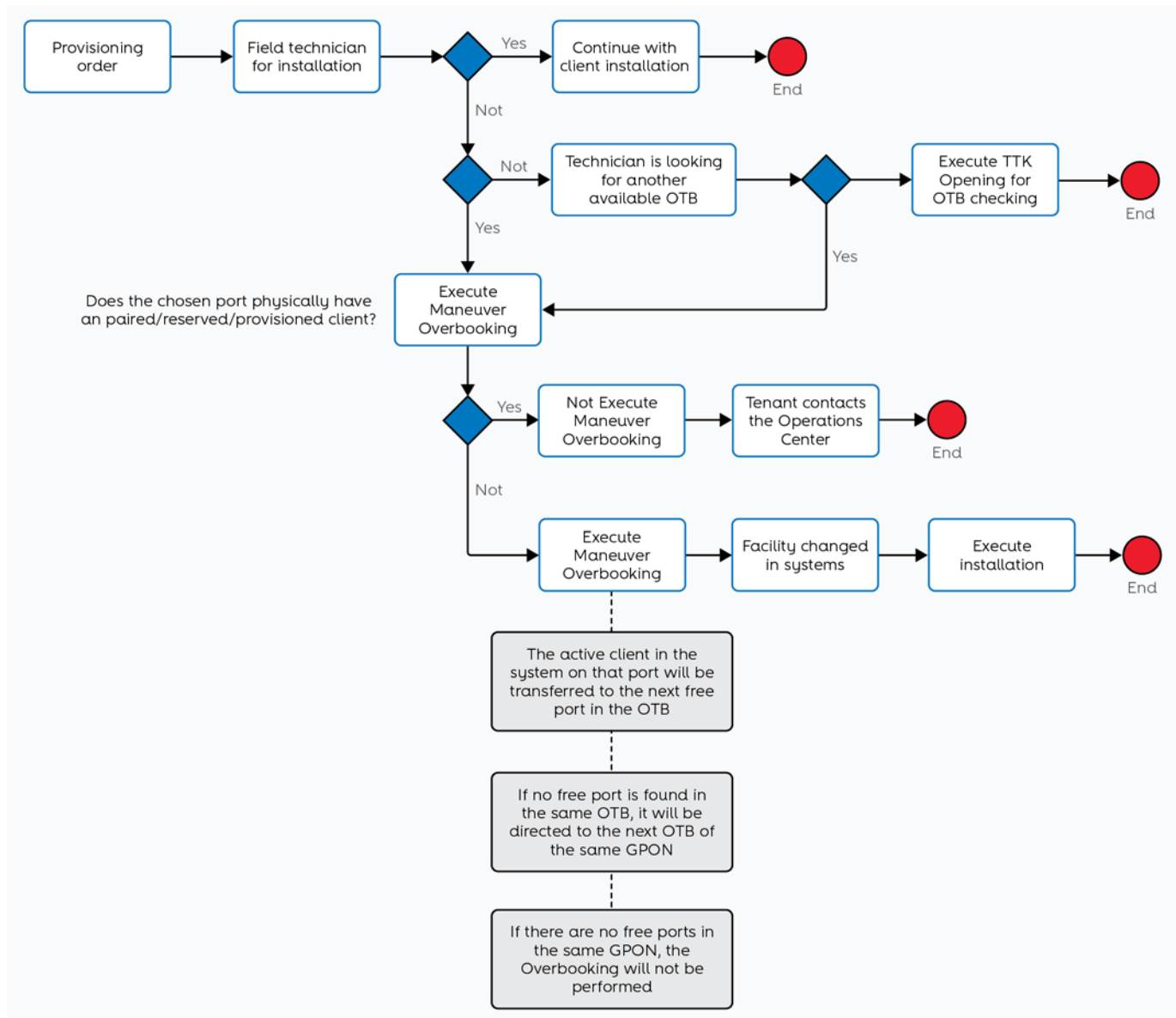


FIGURE 1 – Overbooking Maneuver flowchart

port for executing a customer installation and to identify free ports on site. This last process can occur when, upon arriving at the installation site, the technician finds the originally assigned port is occupied or the OTB is not the most suitable or closest to the requesting customer. After the field technician identifies the appropriate port and OTB, the system applies dynamic rules in two steps:

**1<sup>st</sup> stage of automatic maneuver:** Automatic transfer of the busy service to a free port on the same PON port of a given Optical Line Terminal (OLT), freeing up the desired port, relying on technical inspection.

**2<sup>nd</sup> stage of automatic maneuver:** After completing the first stage, the system (FiBrasil Integration x FulfillmentOne x Netwin) performs the desired service maneuver for the port chosen by the technician.

Continuous tuning and integration of the NOSSIS suite includes:

- Optional addition of the "OTB Port" field in the Maneuver action in the integration flows, FulfillmentOne, and Netwin;
- Marking services as "overbooked" within the previous flow;
- New operation in TMF638 Service Inventory Application Programming Interface (API) developed by Open Labs to list OTBs and their free ports powered by the same PON – used by FiBrasil integration orchestration;
- Enhancements to the "Ports Provisioned" report to reflect information on services driven by overbooking.

The Overbooking Maneuver solution involves the Netwin Inventory and FulfillmentOne modules of the NOSSIS suite, as can be seen in **Figure 2**. Requests for these maneuvers are entered through the FiBrasil Portal.

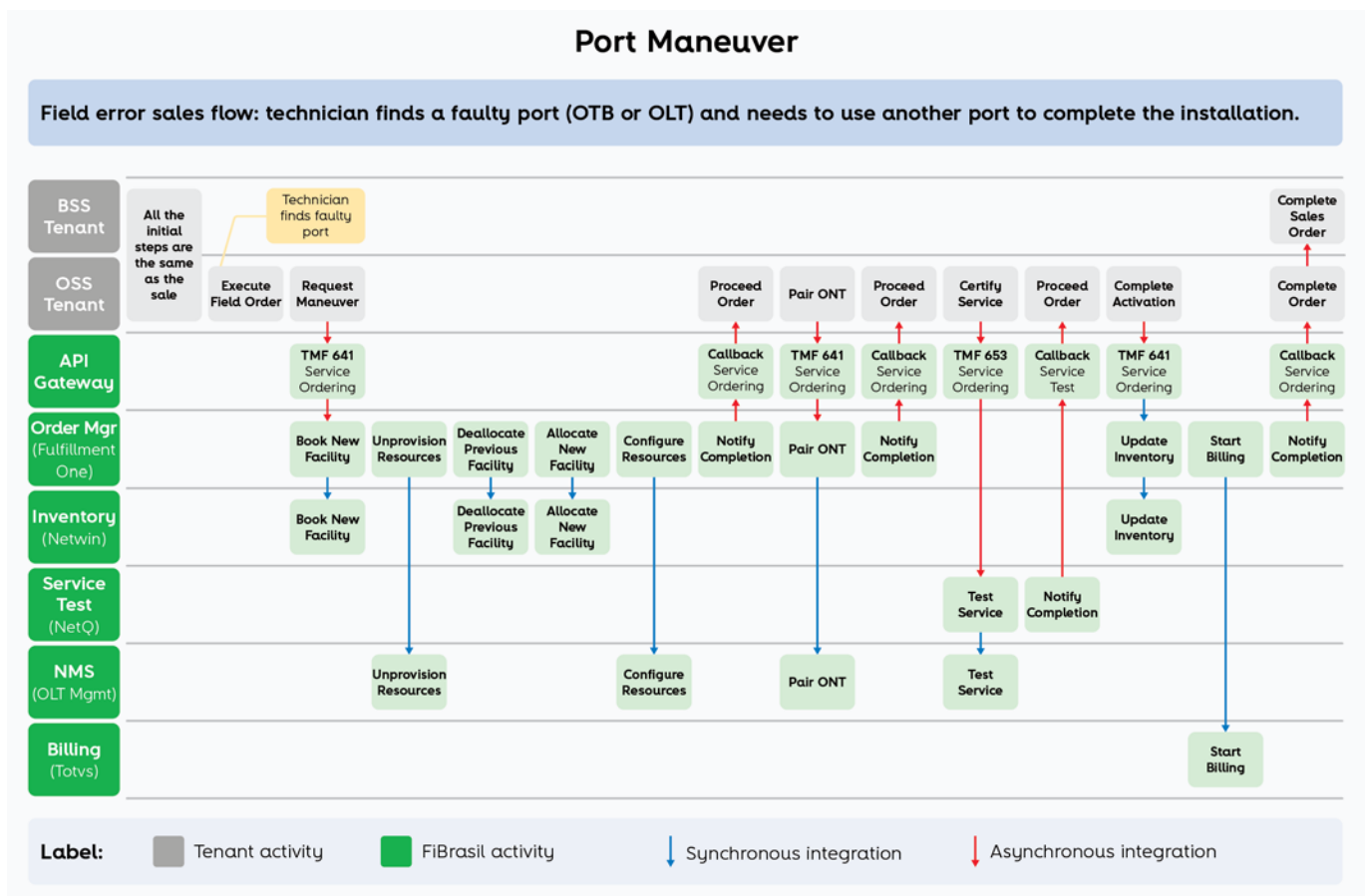


FIGURE 2 – Maneuver process flow

## Inventory (Netwin)

Netwin is the module of the NOSSIS suite responsible for all levels of inventory information of physical resources associated with infrastructures (Outside Plant and Inside Plant), covering the various physical and logical levels of the network for service management, thus providing general end-to-end coverage.

In the Outside Plant section, inventory information is based on georeferenced information and is used to manage the inventory of the physical network and its infrastructures.

In the Inside Plant section, we have functionalities for obtaining, editing, and manipulating information according to the inventory of equipment, the technical room plan, and physical connectivity, in addition to other information that can also be manipulated in the fulfillment module, such as address management, address query, service qualification, network planning, GPON network design, etc.

Additionally, Netwin provides GPON resource allocation capabilities, allocating resources to services under FulfillmentOne requests.

## Fulfillment (FulfillmentOne)

The FulfillmentOne module includes the resource provisioning part. Through a service order, it is possible to allocate resources in the inventory and to provision them in the network, thus configuring the resources according to the expected configurations for each access and provisioning the Customer Premises Equipment (CPEs) in the OLTs. The provisioning is based on Open APIs.

Our SOM (Service Order Management) system coordinates the workflow of activities distributed across various systems such as inventory system, activation system, and workforce management system, among others, until the customer's request is fully met, also allowing a status check of each activation order.





## Integration (APIs)

All systemic integration, whether between internal modules or in communication with tenant systems, is based on the TM Forum's APIs standard, which allows the integration of various systems and platforms, facilitating the automation of operations and the implementation of new services. With the adoption of TM Forum APIs, developers are able to create more robust and scalable solutions, ensuring interoperability between different components of the IT ecosystem. Because they are open and widely known, they facilitate adoption by potential tenants.

Below, we present a detailed list of the APIs used, highlighting their main functions and how they contribute to the optimization of maneuvers.

### Service Ordering API TMF641 [2]

This API is intended for the control of the orders created, making it possible to create configuration orders for the FiBrasil network, the most common orders being provisioning, pairing, unpairing, activation, modification, disconnection, change of address, and port maneuver. This API also provides query operations for the generated orders, making it possible to identify their status.

### Service Inventory API TMF638 [3]

This API corresponds to the query of facilities of an already provisioned service, finding and listing the service elements in the network inventory. For the Overbooking Maneuver solution, this API received a new operation, the OTB port query, which lists the available OTBs according to the location parameters provided, as well as the free ports of these OTBs.

### Service Test Management API TMF653 [4]

This API is intended to perform Service Certification to ensure that what is provisioned in the inventory is correctly configured on the network. After the execution of service orders, for example, provisioning, pairing, or port maneuvering, it is used to

certify that all parameters are consistent with the specification of the service to be provided.

On FiBrasil's website [5], it is possible to find the complete list of available APIs, with examples of requests and responses from its operations, at the following link: <https://fibrasil.com.br/catalogo-de-apis/>.

## Overbooking Maneuver solution: step-by-step

**Case:** at the time of the installation of a service, if the technician identifies that the port assigned by the system to perform the installation is already physically occupied in the OTB, the overbooking maneuver allows the technician autonomy to choose the OTB and the port to be used.

### Overbooking Maneuver solution: step-by-step

**1. Installation start (Figure 3):** the technician tries to install a service and finds Port 1, designated by the system, is already occupied in the OTB.

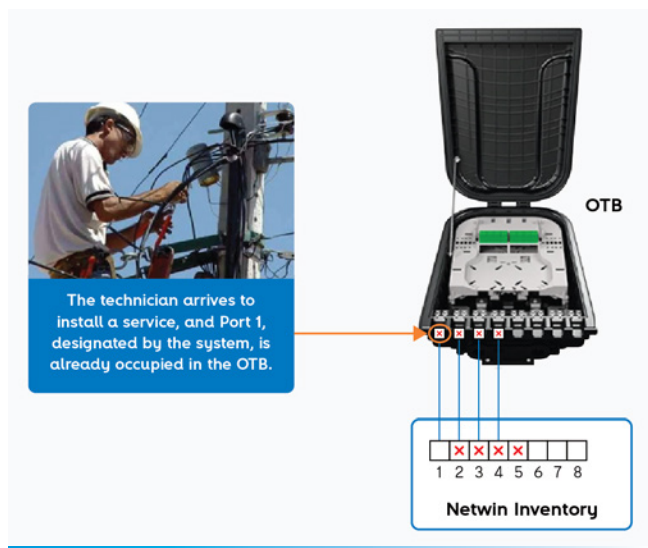


FIGURE 3 – Installation start

**2. Technician identifies free port (Figure 4):** the technician verifies that Port 5 is free at the OTB. Requests the Maneuver from the service office, which performs it via the Portal, informing the OTB ID and Port ID.

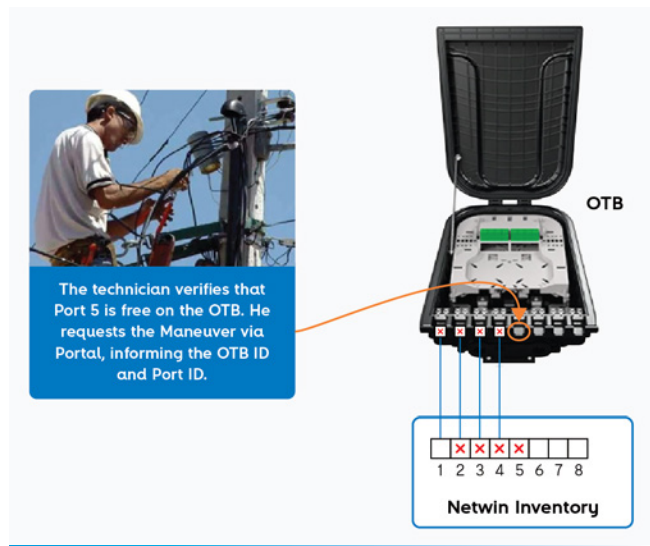


FIGURE 4 – Technician identifies free port

**3. Execution of the Overbooking Maneuver (Figure 5):** during the execution of the Maneuver, if the system identifies that the port to be used, in this example Port 5, is incorrectly occupied in the inventory (Netwin), it executes a new Maneuver, releasing Port 5

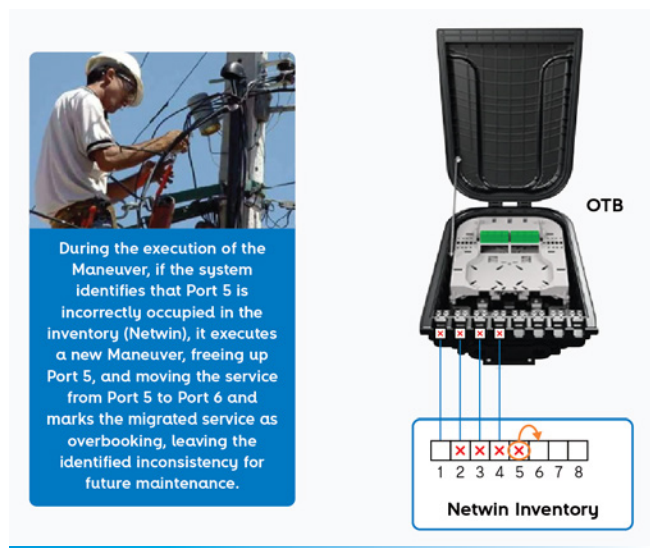
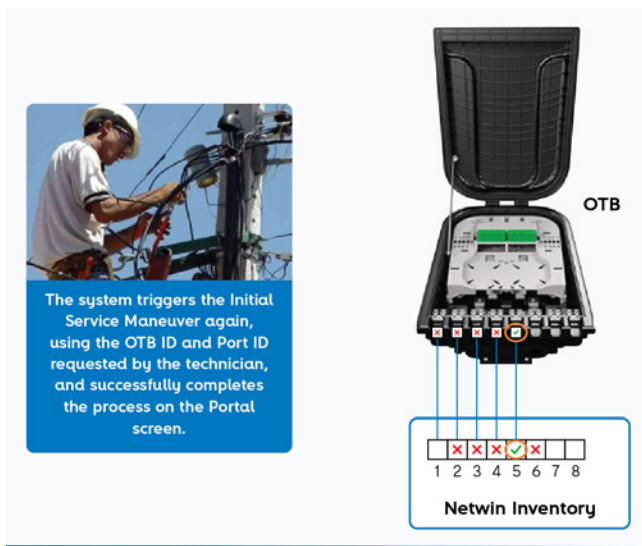


FIGURE 5 – Execution of the Overbooking Maneuver

and moving the service from Port 5 to a free port as long as it is on the same PON port, here exemplified by Port 6, and marks the migrated service as overbooking, leaving the inconsistency identified for future maintenance. This marking helps in the process of improving the network's occupancy inventory.

**4. Service installation completer (Figure 6):** the system triggers the initial Service Maneuver again, using the OTB ID and Port ID requested by the technician, and successfully terminates the process on the Portal screen.



**FIGURE 6** – Service installation completer

The port information and maneuvered ports are updated in the Netwin inventory. It guarantees that the inventory is updated to the next installation or repair in this OTB.

**Exception cases:** this solution also manages detailed exception flows, such as the lack of free ports for the overbooking maneuver, ensuring transparency and effectiveness in operations.

## Conclusions

With the integration between FulfillmentOne, Netwin, and Fibrasil's integration layer, we increase

responsiveness and operational efficiency, ensuring agile and accurate management of service maneuvers in the GPON network infrastructure, reducing the problems of undue disconnections and lack of precision of the inventory. The autonomy offered in this delivery provides agility and a significant reduction in the time required to carry out field activities.

This solution not only optimizes operational processes, but also strengthens Fibrasil and Open Labs' position as a leader in advanced telecommunications solutions, empowering neutral fiber providers like FiBrasil to reach new heights of performance and customer satisfaction. 🌐

## Acknowledgements

Thanks to the team without whom this project would not have been successful.

- Andreia Marques Lobato Mateus
- Antonio Wagner Mota Cavalcanti
- Briane Cristina Pereira
- Dayane Cavalcanti de Oliveira
- Gisele Cristina Serafim de Oliveira Guimarães
- Isabella Marques de Barros Torres
- Laura Silva Soares Gomes
- Renata Luana Silva de Andrade
- Silvani Mendes de Souza
- Solomar Izaias
- Sylvio Augusto Alcantara Nogueira

---

## References

- [1] ITU-T Recommendation G.984.1, "General Characteristics of Gigabit-capable Passive Optical Networks (GPON)," 2003.
- [2] "TMF641 Service Ordering Management API User Guide v4.1.1." tmforum.org. Accessed: Sep. 13, 2024. [Online]. Available: <https://www.tmforum.org/resources/specifications/tmf641-service-ordering-management-api-user-guide-v4-1-1/>.
- [3] "TMF638 Service Inventory API User Guide v5.0.0." tmforum.org. Accessed: Sep. 13, 2024. [Online]. Available: <https://www.tmforum.org/resources/specifications/tmf638-service-inventory-api-user-guide-v5-0-0/>.
- [4] "TMF653 Service Test Management API User Guide v4.1.0." tmforum.org. Accessed: Sep. 13, 2024. [Online]. Available: <https://www.tmforum.org/resources/specification/tmf653-service-test-management-api-user-guide-v4-1-0/>.
- [5] "Catálogo de APIS." fibrasil.com.br. Accessed: Sep. 13, 2024. [Online]. Available: <https://fibrasil.com.br/catalogo-de-apis/>.

# 06

## Introducing DevSecOps practices in Altice Labs



The adoption of the DevSecOps methodology addresses the urgent need to incorporate security into the software development life cycle, which results in a more efficient process, greater robustness against threats and vulnerabilities, and a product of higher quality. Ignoring security during the development process can result in significant risks, such as data breaches and substantial financial losses. However, it is also necessary to acknowledge the lack of clear guidance on how to implement DevSecOps and select the appropriate security tools.

The present article aims to offer guidance on the current security-related problems and solutions at each stage of the software development life cycle, as well as to serve as a guide for DevSecOps implementation and security tools selection. Furthermore, some of these concepts and third-party tools were applied to create an internal solution that integrates into Altice Lab's Continuous Integration/Continuous Delivery (CI/CD) pipelines to perform security analysis. As part of a pilot, this solution is being validated with selected projects. This article highlights the usefulness of the ongoing pilot, including the effectiveness of our approach and the improvements implemented based on real-world feedback.

---

## Authors

**Pedro Souto, Altice Labs,** [pedromgsouto@gmail.com](mailto:pedromgsouto@gmail.com)

**Mafalda Nunes, Altice Labs,** [mafalda-g-nunes@alticelabs.com](mailto:mafalda-g-nunes@alticelabs.com)

**Ricardo Cadime, Altice Labs,** [ricardo-v-cadime@alticelabs.com](mailto:ricardo-v-cadime@alticelabs.com)

**Pedro Neves, Altice Labs,** [pedro-m-neves@alticelabs.com](mailto:pedro-m-neves@alticelabs.com)

---

## Keywords

Security; DevSecOps; SecDevOps; DevOps; CI/CD; Software Development; Pipelines; Organizations; Secure SDLC



## Introduction

For the past years, many organizations across the globe have chosen to deviate from the classic software development methodology for the sake of adopting a newer and more advantageous one [1]. The development process for software used to be linear and sequential, using the classic waterfall approach. This method was rigid and slow, which made it challenging for businesses to adapt to the shifting consumer needs and market demands. As a result, businesses started implementing more agile methods of software development, such as Scrum and Lean, which emphasized the incremental and iterative delivery of software.

Nonetheless, despite this adoption of agile approaches, there was still a barrier between the development and operations teams. While operations oversaw ensuring the stability and dependability of the current systems, development focused on delivering new features and capabilities. The distribution of software was delayed by this division, which frequently resulted in disputes, as well as security vulnerabilities and breaches [2].

The DevOps approach was introduced to address these issues, combining both the software development process and the IT operations silos that were formerly separated [3]. The adoption of these new principles and methodologies allowed companies to reach new velocity and agility levels previously unattainable [1]. From a DevOps point of view, automation is key. In this approach, the automation of the software development process is intrinsically connected to the concept of CI/CD, which refers to the pipeline that stands between the development team and the client, ensuring an automated process of building, testing, and deploying the application. This automation process has accelerated software development and release, making it faster and more reliable than ever before.

However, despite the numerous advantages of the DevOps approach, the systems development life cycle (SDLC) still lacked the integration of security in every phase. On one hand, manual security

processes can be a bottleneck for the fast release of software, since they are complex and take time. On the other hand, security is indispensable and, when ignored, can not only cost companies huge sums of money, but also make both the stakeholders and clients lose confidence in the brand.

In addition to their own SDLC, modern organizations also depend on their software supply chain, which encompasses the development, distribution, and maintenance of the software they use, supporting a variety of activities and processes. The security of the supply chain must be ensured to safeguard against cyber threats and data breaches, which can have serious financial and reputational repercussions for organizations.

In 2023, Sonatype reported the discovery of 245,000 malicious packages, which is twice the total number of malicious packages identified in the four previous years combined [4]. Furthermore, IBM's 2023 study revealed several crucial statistics: organizations that adopted DevSecOps experienced a 38.4% cost savings per data breach, translating to approximately US\$ 1.68 million – making it the top cost mitigator [5]. Data breaches related to compromised supply chains accounted for 27% of all breaches, with the average cost reaching US\$ 4.63 million [5]. Additionally, 51% of organizations planned to increase security investments after experiencing a data breach [5].

Recent high-profile incidents, such as the Solarwinds attack [6] and the CrowdStrike incident [7], as well as the approval of stricter regulations (e.g., the European GDPR and NIS2 Directive, the US FISMA, and CCPA), put tremendous pressure on the need for various industries to ascertain the security of their products and services. These industries responded by trying to integrate automated security processes into the DevOps approach, complementing the CI/CD pipelines and allowing the shift to a DevSecOps paradigm, giving meaning to the term “Continuous Security.”

The shift to DevSecOps presents some interesting challenges, such as how to secure both the applications being developed and the CI/CD pipelines themselves. Another important and relevant

concern is how to flawlessly integrate these security methods into existing DevOps processes without disturbing their fast pace and responsiveness.

In a 2022 study that explored the challenges faced in adopting DevSecOps [8], researchers found that industries considered the most challenging factors to be the lack of automated testing tools for security in DevOps and the process of identifying software defects found in operations monitoring and feeding them back to development, both top-rated by 85% of the industries surveyed.

Industries are increasingly acknowledging the necessity of integrating robust security measures throughout the various phases of the SDLC. This recognition is highlighted by the alarming statistics, the increasing volume of attacks, and the emerging strict legislative frameworks discussed earlier. To this end, there has been a progressive adoption of automated security measures within existing DevOps practices, aiming to significantly enhance the security posture of software development processes. However, industries face significant challenges in this transition to DevSecOps, primarily due to the lack of standard tools designed to facilitate this change. The objective of the work presented in this article is to tackle this exact issue, enabling the transition to DevSecOps with the ultimate aim of bolstering the security of the entire software development process and software supply chain.

## DevSecOps-enabled SDLC

The growing significance of software security in the development process has drawn a lot of attention in recent years to the adoption of DevSecOps principles and practices. However, incorporating security measures into the quick and agile development processes poses multiple difficulties. A security-focused extensive analysis of current best practices and industry standards for each phase of the DevSecOps life cycle was performed and consolidated [9]. In this section, some of the

key security concerns and best practices that must be addressed in each phase are identified. **Figure 1** illustrates these phases, integrating the Development (Dev), Security (Sec), and Operations (Ops) aspects into a unified model.

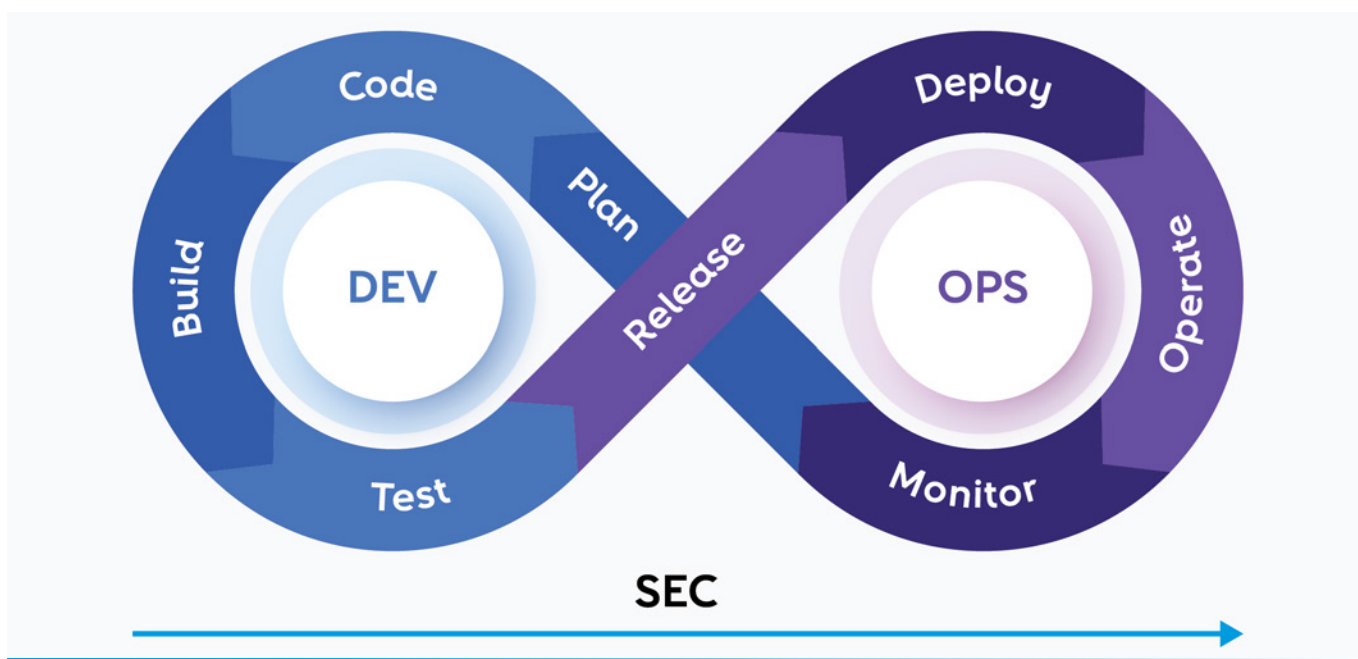
The Development phases reflect the initial stages, such as planning, coding, building, and testing. The Operations phases represent the latter stages in which the application is released, deployed, operated, and monitored. Security traverses and intertwines with all phases, ensuring it is not an afterthought, but rather a constant focus pervading every aspect of the development and operations processes. This also emphasizes the “shift left” security approach in DevSecOps, where security is considered earlier in the life cycle to identify and address issues as soon as possible.

The feedback loop format of the figure indicates that the DevSecOps process is iterative and cyclical. After each cycle, feedback is taken from monitoring and applied to improve the next iteration, promoting continuous improvement and learning. This ensures that the development, operation, and security application aspects are constantly updated and improved, leading to higher quality and more secure software.

Transversal security concerns in DevSecOps include selecting the proper automation tools, access control, knowledge sharing, pipeline security, secrets management, and fostering a security-aware culture. Key practices involve Infrastructure as Code (IaC), which automates the setup of computing environments and resources, and Configuration as Code (CaC), which ensures consistent configurations across environments. Managing various environments, whether on-premises, cloud, or hybrid, whether single-tenant or multi-tenant, also presents unique security challenges.

In the planning, analysis, and design phase, threat modeling is used to identify potential security threats. Security and privacy risk assessments focus on protecting data and ensuring data privacy. Change impact analysis predicts the potential effects of system modifications, while abuse case analysis anticipates system misuse. These steps inform the definition of security requirements for the team to follow. Additionally, secure code training sharpens developers’ security skills.

In the coding phase, practices like source code versioning and signed commits ensure the traceability and authenticity of code changes. IDE plugins for Static Application Security Testing



**FIGURE 1** – DevSecOps-enabled SDLC

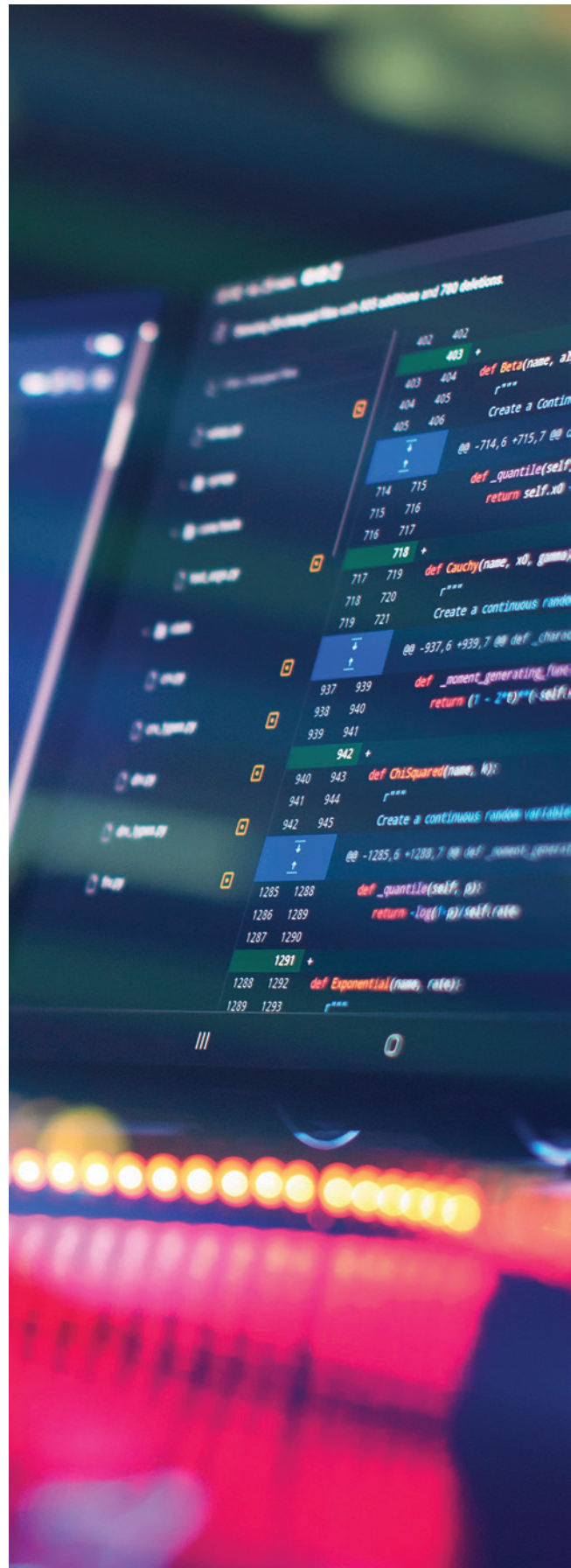
(SAST) and Software Composition Analysis (SCA) improve code security. Automated tools aid code reviews for quality assurance, while commit hooks enforce coding standards. Secret scans detect sensitive data, preventing leaks. Workstation security policies protect development environments.

The build phase involves employing secure build techniques to create reliable, secure software. It also includes the use of signed dependencies to verify the integrity and authenticity of third-party software components, ensuring a trusted software supply chain.

The testing phase of DevSecOps integrates several strategies for robust security. Secrets injection handles sensitive data during testing. SAST, SCA, and secret scans detect vulnerabilities in code and potential data leaks. Dynamic Application Security Testing (DAST) and Interactive Application Security Testing (IAST) identify security threats in running applications.

The release phase is the final checkpoint in DevSecOps before deployment, making rigorous security testing crucial. Techniques like SAST, SCA, and secret scans guarantee the security of the code. The package's digital signature/encryption, coupled with secure transfer, ensures its safety throughout the process. A Software Bill of Materials (SBOM) provides full transparency of the package's contents. Extensive DAST and IAST performed in Quality Assurance (QA) or staging environments that closely mimic the production environment ensure the utmost security of the application while avoiding disrupting operations. Penetration testing (pen testing) in these environments allows to identify potential vulnerabilities that could be exploited in a real-world scenario.

In the deployment phase, runtime protection implementations intend to safeguard applications while they are running. Moreover, host security measures and vulnerability scans (possibly aided by DAST and IAST tools) ensure the deployment environment's security. Additional targeted penetration testing helps to further identify potential vulnerabilities.



The operation phase involves strategies for continued security and maintenance. Log collection provides valuable insight into application behavior and potential security incidents. Runtime Application Self-Protection (RASP) and Web Application Firewalls (WAF) actively safeguard the application. Cloud Native Application Protection Platforms (CNAPP) offer holistic protection in cloud environments. Bug bounty programs incentivize the identification of vulnerabilities, while regular patching ensures the mitigation of identified security risks.

In the monitoring phase, CNAPPs provide continuous oversight of applications in cloud environments. Security Information and Event Management (SIEM) systems collect and analyze security-related data for threat detection. Extended Detection and Response (XDR) integrates various security tools for a comprehensive defense strategy. Continuous vulnerability monitoring identifies and addresses emerging security threats, ensuring the ongoing security of the application.

## Altice Labs system development process

This section explores some key aspects of the SDLC as practiced by Altice Labs. Understanding these practices provides a foundation for suggesting improvements based on the DevSecOps methodology. This approach enhances not just the security embedded within the SDLC but also within the software supply chain. The integration of DevSecOps can strengthen Altice Labs' ability to combat security risks and ensure the development of secure, efficient, and robust software solutions.

**Agile Practices** - Altice Labs originally operated under the Waterfall model of software development. This model, while well-structured, lacks the flexibility and iterative nature found in Agile methodologies. Recognizing these limitations, Altice Labs has integrated Agile practices, specifically Scrum, into their development process. Looking towards the future, Altice Labs anticipates further optimizing its software

development and delivery processes through the adoption of the DevSecOps methodology.

**Custom Components** - Altice Labs has been transitioning from proprietary solutions and custom on-premises components, such as registries, to more widely accepted and standardized solutions like GitHub Registries and Google Cloud Artifact Registries. However, the transition is challenging due to the tight coupling and integration between custom components. Moving towards these standardized solutions is a strategic move that not only aligns with industry best practices but also addresses the challenges of maintaining and enhancing custom solutions, particularly in terms of security.

**Knowledge Centralization** - currently, Altice Labs maintains both internal and external Server instances of Jira and Confluence, leading to increased costs and potential inefficiencies. A migration process is ongoing to consolidate them into single instances of Cloud Jira and Confluence. This centralization aims to reduce costs, enhance collaboration, ensure data consistency, and improve overall operational efficiency.

**Workstation Security** - current security practices at Altice Labs involve access control measures and two-factor authentication (2FA), with varying privileges for users based on job roles. Antivirus software, firewall policies, and full disk encryption are also used. While SAST plugins for secure code development and SCA plugins for dependency security scanning are already in use, they have not been standardized or uniformly adopted across the entire organization. The goal moving forward is to establish a more consistent and widespread application of these tools.

**Source Code Management** - Altice Labs has been transitioning from Apache Subversion (SVN) to Git, specifically GitHub, for source code management. Git provides multiple appealing features, such as distributed version control, a robust security model, and commit signing to ensure code authenticity. GitHub provides some additional security features that will be used, such as protected branches and pull request reviews, making sure the requirements for code review, approval, and status checks are met.

**Security Tests** - Altice Labs plans to standardize the use of security testing tools across projects, integrating system tests after each build, and utilizing SAST, SCA, DAST, and IAST. There is also a special focus on QA-intensive security testing, as this environment closely mimics the production environment and does not cause service disruption.

**Continuous Integration** - currently, CI is used for code integration, automated building, and testing. In the past, continuous integration was predominantly managed by Jenkins, but a transition to GitHub Actions has been taking place.

## Tools analysis

Several security tools on the market can help organizations and individuals safeguard their systems and data [10]. Security analysis tools identify possible vulnerabilities in systems and applications, thus helping to identify and stop security attacks and breaches. There are different kinds of security analysis tools, including SAST [11], SCA [12], and DAST [13], among others. Each of these types of tools has a distinct purpose, and they can be used in conjunction to build a robust CI/CD pipeline [10].

Free security analysis tools are very helpful for people and companies who might not have the dedicated funds to invest in pricey security solutions.

It is also vital to keep in mind that while these tools can be helpful, they might not be as feature-rich as their paid counterparts, even though they offer numerous capabilities.

Various free security analysis tools were compared in this section to build a secure CI/CD pipeline that is both efficient and economical. Furthermore, these tools were chosen with deep consideration for Altice Labs' current tech stack to address any existing security gaps within its systems. This way, security measures can be implemented without disrupting the established workflows, aligning with the principles of DevSecOps, which advocates for the seamless integration of security into the development and operations processes.

## Types of security analysis

As previously mentioned, there are various types of security analysis, three of which are the focus of this work: SAST, SCA, and DAST. **Table 1**, presented below, highlights and compares these analyses' key characteristics.

In this table, the type of analysis denotes the level of system knowledge during testing (white, black, or gray box). The requirements field lists what is analyzed by the tool. The SDLC phases pinpoint when the tool can be implemented in the development life cycle. The targets indicate the scopes to which the analysis is applied. Lastly, the findings refer to the type of security issues the analysis can detect.

	SAST	SCA	DAST
Type	White-Box	Gray-Box	Black-Box
Requirements	Source Code	Metadata / Source Code	Running Application
SDLC phases	Code; Test; Release	Code; Test; Release	Test; Release; Deploy
Targets	Source Code, IaC, and CaC	Open-source SW, 3 <sup>rd</sup> party libraries, and dependencies	Running applications and Infrastructure
Findings	Bugs, vulnerabilities, secrets, and code smells	Known vulnerabilities, licensing issues, and outdated libraries	Runtime vulnerabilities and security weaknesses

**TABLE 1** – Types of security analysis

This comparative understanding aids in choosing the right tools for integration into the CI/CD pipeline, ensuring robust and cost-effective security measures. These tools can be used in combination for a comprehensive security analysis.

## Tool comparison criteria

When researching free security analysis tools, a combination of academic literature and web-based resources were consulted to compile a list of potential candidates for each type of security analysis (SAST, SCA, and DAST). The next step was to filter the options based on the criteria specified in **Table 2**.

## Tool comparison results

Filling the security gaps in any organization requires careful selection and testing of tools that are compatible with the existing tech stack. In

this context, multiple free tools were meticulously examined to address the specific security needs of Altice Labs.

This involved a comprehensive evaluation of tool features, compatibility, ease of integration, and their effectiveness in identifying and mitigating vulnerabilities. The goal was to assemble a toolset that could not only enhance Altice Labs' security posture but also integrate seamlessly with their existing development processes.

The outcome of this rigorous comparison and testing process is illustrated in **Figure 2**. This figure presents the security tools that were ultimately selected for inclusion in Altice Lab's security toolbox. The tools are organized according to the type of security analysis they perform, providing a clear representation of how they collectively contribute to a more secure and robust SDLC at Altice Labs.

Criterion	Description
<b>Last version's release date</b>	The tool's latest version release date should be less than six months old, ensuring active maintenance and up-to-date features and capabilities.
<b>Last commit date</b>	The tool's latest commit should be within the last 3 months, ensuring active development and timely addressing of bugs and vulnerabilities.
<b>GitHub stars</b>	The tool should have more than 1,000 GitHub stars, indicative of its popularity, large user base, reliability, and effectiveness.
<b>Free access for private repositories</b>	The tool should offer free access to its features for private repositories, thereby supporting secure development without incurring additional costs.
<b>Comprehensive free features</b>	The tool should incorporate as many free functionalities as possible, allowing the concentration of the needed capabilities within a minimalist toolset, thereby promoting efficiency and cost-effectiveness.
<b>Code privacy</b>	The tool should not upload or share the source code externally, ensuring the privacy and confidentiality of the proprietary code.
<b>CI/CD integration capability</b>	The tool should be designed to seamlessly integrate into CI/CD pipelines, thereby promoting continuous and automated security checks during the software development life cycle.

**TABLE 2** – Tool comparison criteria

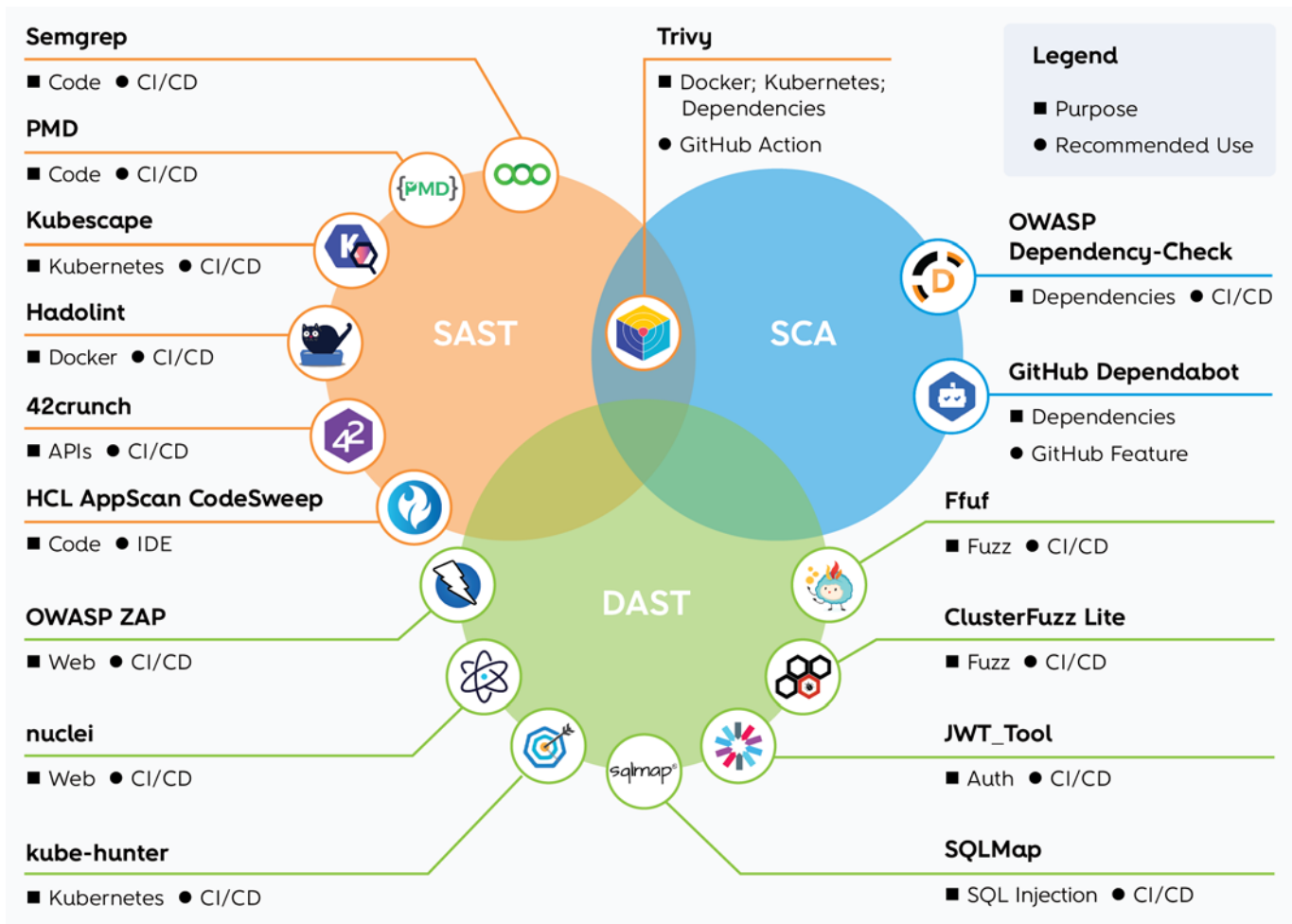


FIGURE 2 – Tool comparison results

## Security in CI/CD at Altice Labs

In an enterprise setting, maintaining robust, reliable, and secure development processes is critical, especially as modern software development relies heavily on CI/CD pipelines. These pipelines streamline the development process, automating workflows and improving efficiency. GitHub Actions [14] have proven to be an excellent resource for building these pipelines, allowing for seamless automation directly with GitHub repositories. Given that Altice Labs already utilizes GitHub as both a code repository and Docker image registry, integrating GitHub Actions into the CI/CD pipeline became an obvious choice to ensure smooth and automated integration and deployment processes.

However, while CI/CD pipelines improve development speed and reliability, they must also account for security. By embedding security checks directly into the pipeline, enterprises can integrate continuous security practices that mirror their continuous development and deployment efforts. This approach allows for the early detection and remediation of security vulnerabilities, thereby minimizing both the cost and the disruption of addressing issues late in the development cycle. As such, Altice Labs recognized the need to incorporate security automation to enhance the security posture of its software development processes.

To address the growing need for integrated security within Altice Labs' CI/CD pipelines, the Systems Vulnerability Detection Tool (SVDT) was introduced. The SVDT brings together the selected open-source

SAST and SCA tools, offering a comprehensive security solution embedded directly into the development workflow. Designed for flexibility, the SVDT can be used in multiple ways:

- Standalone Docker image – this is the base of the SVDT, intended for it to be used anywhere;
- Pre-commit hook – runs the previously mentioned image to catch vulnerabilities before the code is even committed;
- GitHub Action – facilitates the usage of the previously mentioned image in CI/CD pipelines based on GitHub workflows.

By embedding security checks at various stages, the SVDT helps automate the detection of vulnerabilities throughout the system's development lifecycle.

The SVDT performs SAST to identify vulnerabilities in the source code, such as exposed secrets, insecure configurations, and potential SQL injections. This ensures that security issues can be detected and addressed early in the development process, minimizing the risk of introducing vulnerabilities as the codebase evolves. In addition to analyzing source code, the SVDT also incorporates SCA, which scans project dependencies for known vulnerabilities. This component helps ensure that third-party libraries and open-source components used in the software are secure and compliant with licensing requirements, reducing both security risks and legal exposure.

By integrating both SAST and SCA within the CI/CD pipeline via GitHub Actions, the SVDT delivers continuous, automated security checks that align with the agile development practices at Altice Labs, ensuring that security is an ongoing concern throughout the entire development lifecycle.

## SVDT Pilot

Given the diverse range of projects and development environments at Altice Labs, it is essential to evaluate the effectiveness and adaptability of the SVDT across different departments and teams. The goal is to validate the tool's suitability before

expanding its use throughout the organization. This involves ensuring that the SVDT can seamlessly integrate into existing workflows, tackle unique security challenges, and deliver tangible improvements in vulnerability detection.

To achieve this, Altice Labs initiated a pilot program to test the SVDT in a variety of existing projects. The pilot aims to assess the tool's performance across different environments, measure its impact on security practices, and gather feedback for further refinement. Through this structured approach, Altice Labs will be able to determine whether the SVDT meets its goals of enhancing security while supporting the speed and flexibility of modern development practices.

The pilot is organized in four phases to evaluate the tool's effectiveness, adaptability, and security enhancements for diverse projects. Each phase was designed to progressively introduce, deploy, and assess the SVDT.

**Phase 1 - Project Selection:** in July 2024, various departments across the organization nominated projects to participate in the pilot. Among these projects, some were selected in a way that ensured company representativity, as well as technological and requirements diversity, aiming to guarantee that the SVDT would be broadly applicable across the organization.

**Phase 2 - Pilot Kick-off and Goals Setting:** the pilot began at the end of July 2024 with a kick-off meeting to introduce the pilot team, the SVDT, and the overall plan. This was followed by individual meetings with each project team to define specific goals, tailoring the tool's usage to the unique needs and expectations of each project.

**Phase 3 - Execution:** upon the pilot goals setting, at the beginning of August, teams were encouraged to autonomously start integrating the SVDT into their projects using the available documentation and provide feedback. This feedback was instrumental in shaping the subsequent development of the tool, which directly influenced a new version of the SVDT, addressing bugs and adding requested functionalities. In mid-September, a

workshop was conducted to ensure teams were aligned on how to best use the SVDT and to gather more comprehensive feedback. During the workshop, teams shared further insights, which reinforced the feedback previously shared and provided additional perspectives on the tool's strengths and areas for improvement. The SVDT is now being actively integrated into the CI/CD pipelines of the selected projects, with the teams gradually using the tool at various stages of their development lifecycle, based on their specific project needs. The implementation varies across projects, including security analysis using pre-commit hooks and GitHub Actions (e.g., full periodic scans, differential analysis during testing, full scans during release). Until November, regular check-in meetings will be conducted with all participating teams to track the progress of the pilot and collect feedback. These meetings, aided by surveys, will help us gather insights on the tool's usability, its impact on existing workflows, and suggestions for further improvements.

**Phase 4 - Evaluation:** in December, evaluation meetings and a final survey will be conducted to assess the overall success of the SVDT pilot. This evaluation will be based on predefined qualitative and quantitative metrics:

- **Qualitative:** The evaluation will focus on the tool's usage across different stages (pre-commit, CI/CD pipeline, and periodic analysis), as well as its ease of use. Feedback will be gathered on the clarity of the documentation, the ease of interpreting results, and any potential areas for improvement;
- **Quantitative:** The evaluation will consider the number of critical vulnerabilities detected throughout the pilot, the scan duration and efficiency, and the SVDT's performance when compared with security tools already used by each project. These metrics will provide a clear indication of the tool's impact on both security and development speed.

If pilot results confirm SVDT readiness, the tool will be made available, and its usage spread across the entire organization in the beginning of 2025.

## Conclusions

In the intricate landscape of software development and supply chain security, the adoption of a comprehensive DevSecOps approach emerges as a crucial strategy. This study aims to provide a thorough guide for securing the SDLC and software supply chain, as well as to foster a holistic understanding of DevSecOps, equipping organizations with the required knowledge to strengthen their defenses against potential security risks.

The selection of suitable tools is a cardinal part of this process, requiring rigorous standards and exhaustive analyses. The empirical validation provided by the Altice Labs case study further underlines the importance of tool selection and highlights the efficacy of a well-executed DevSecOps approach. By integrating strategic tools into the CI/CD pipeline, Altice Labs is able to identify and mitigate a substantial number of potential security issues, thereby attesting to the robustness and effectiveness of DevSecOps.

However, the path to a successful DevSecOps implementation is not solely a matter of effective tool selection; it requires a strong organizational commitment. To fully realize the benefits of DevSecOps, organizations must be prepared to invest time, resources, and energy into refining their development processes, fostering a security-centric culture, and ensuring continuous learning and improvement.

In conclusion, this research elucidates the multifaceted benefits of integrating DevSecOps into software development and supply chain processes. By highlighting the practical measures, potential challenges, and a real case study, it serves as a beacon for organizations embarking on their DevSecOps journey. The future of secure software development hinges on the wider adoption of such robust practices, and we hope this work will contribute to accelerating that shift. 🌐

## References

- [1] R. W. Macarthy and J. M. Bass, "An Empirical Taxonomy of DevOps in Practice," in *2020 46<sup>th</sup> Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Portoroz, Slovenia, 2020, pp. 221-228, doi: 10.1109/SEAA51224.2020.00046.
- [2] M. Jaatun, D. Cruzes, and J. Luna, "DevOps for Better Software Security in the Cloud Invited Paper," in *Proceedings of the 2017 Conference*, 2017, pp. 1-6, doi: 10.1145/3098954.3103172.
- [3] M. S. Khan, A. W. Khan, F. Khan, M. A. Khan, and T. K. Whangbo, "Critical Challenges to Adopt DevOps Culture in Software Organizations: A Systematic Review," *IEEE Access*, vol. 10, pp. 14339-14349, 2022, doi: 10.1109/ACCESS.2022.3145970.
- [4] Sonatype, "9<sup>th</sup> Annual State of the Software Supply Chain Report | Sonatype." sonatype.com. Accessed: Sep. 12, 2024. [Online]. Available: <https://www.sonatype.com/state-of-the-software-supply-chain/introduction>.
- [5] IBM, "Cost of a Data Breach Report 2023." ibm.com. Accessed: July 15, 2024. [Online]. Available: <https://www.ibm.com/reports/data-breach>.
- [6] CIS, "The SolarWinds Cyber-Attack: What You Need to Know." cisecurity.org. Accessed: Sep. 16, 2024. [Online]. Available: <https://www.cisecurity.org/solarwinds>.
- [7] Forbes, "CrowdStrike Reveals What Happened, Why—And What's Changed." forbes.com. Accessed: Sep. 16, 2024. [Online]. Available: <https://www.forbes.com/sites/kateoflahertyuk/2024/08/07/crowdstrike-reveals-what-happened-why-and-whats-changed/>.
- [8] M. A. Akbar, K. Smolander, S. Mahmood, and A. Alsanad, "Toward Successful DevSecOps in Software Development Organizations: A Decision-Making Framework," *Information and Software Technology*, vol. 147, p. 106894, Feb. 2022, doi: 10.1016/j.infsof.2022.106894.
- [9] "Secure software development life cycle." hdl.handle.net. Accessed: Sep. 16, 2024. [Online]. Available: <https://hdl.handle.net/10773/41105>.
- [10] F. Mateo Tudela, J.-R. Bermejo Higuera, J. Bermejo Higuera, J.-A. Sicilia Montalvo, and M. I. Argyros, "On Combining Static, Dynamic and Interactive Analysis Security Testing Tools to Improve OWASP Top Ten Security Vulnerability Detection in Web Applications," *Applied Sciences*, vol. 10, no. 24, p. 9119, Dec. 2020, doi: 10.3390/app10249119.
- [11] A. Fatima, S. Bibi, and R. Hanif, "Comparative study on static code analysis tools for C/C++," in *2018 15<sup>th</sup> International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, 2018, pp. 465-469.
- [12] N. Imtiaz, S. Thorn, and L. A. Williams, "A comparative study of vulnerability reporting by software composition analysis tools," in *Proceedings of the 15<sup>th</sup> ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, Bari, Italy, 2021, pp. 1-11.

- [13] T. Rangnau, R. v. Buijtenen, F. Fransen, and F. Turkmen, "Continuous Security Testing: A Case Study on Integrating Dynamic Security Testing Tools in CI/CD Pipelines," in *2020 IEEE 24<sup>th</sup> International Enterprise Distributed Object Computing Conference (EDOC)*, Eindhoven, Netherlands, 2020, pp. 145-154.
- [14] "GitHub Actions Feature." github.com. [Online]. Accessed: Sep. 16, 2024. [Online]. Available: <https://github.com/features/actions>.

# 07

## Network Federated Ecosystem with Blockchain



Traditional centralized identity management systems may have security vulnerabilities that can compromise privacy when two or more network entities attempt to do business. To tackle this matter, Blockchain technology presents a transformative digital solution to these issues by enabling decentralized creation, verification, and control of digital identities and ownership in a Network Federated Ecosystem (NFE). With this technology, regulatory entities can issue immutable, transparent, and decentralized credentials that are registered in a federated facility through Distributed Ledger Technology (DLT), thus enhancing privacy and security. As such, Blockchain uses its immutability, transparency, and decentralization to address the shortcomings of conventional systems, reducing the risks of data breaches, identity theft, and unauthorized access to a resource and/or service. The consensus mechanisms enabled by Blockchain facilitate the establishment of secure testbeds, promoting the enablement of Sixth Generation (6G) Radio Access Network (RAN) devices.

---

## Authors

**Bruno Santos, Altice Labs, [bruno-g-santos@alticelabs.com](mailto:bruno-g-santos@alticelabs.com)**

**André Gomes, Altice Labs, [andre-d-gomes@alticelabs.com](mailto:andre-d-gomes@alticelabs.com)**

---

## Keywords

NFE; Blockchain; DLT; 6G, SUNRISE-6G; Telecommunications Networks

## Introduction

Advancements in telecommunications networks have been driven by data generated from technologies such as Radio Intelligent Surfaces, Edge Computing, 5G networks, Internet of Things (IoT) devices, and Machine Learning algorithms. However, challenges persist in managing, sharing, and securing this data, as it must be transparently and trustfully accessible to all actors within the telecommunications ecosystem [1]. To address these challenges, the concept of NFE with Blockchain has emerged as a transformative solution, promising to revolutionize data governance, sharing, and security across ecosystems composed of diverse telecommunications network infrastructures [2].

An NFE is an established network of several conglomerates that participate in a global ecosystem to perform business with other verified partners. Each participant in this ecosystem can be an individual, organization, or even a device, contributing to and benefiting from the shared pool of data and services [1], [3]. They can share these data, resources, and other services, while the resulting transactions are recorded on the blockchain, ensuring transparency and trust across all partners. Since this ecosystem is also built on the principle of data privacy, sensitive information can remain confidential [2].

In this way, 6G is expected to emerge as a key enabler for the intelligent digital society, as it will enable an ever-increasing number of devices, supporting the expansive growth of IoT. In a federated ecosystem, this means that more Communication Service Providers (CSP) can participate, contributing to and accessing shared resources and data [3]. With expected higher reliability and stability in communications, this will ensure a higher level of consistency and dependable interactions between entities belonging to a federated ecosystem, bolstering trust and operational efficiency [4].

Former European projects have proposed novel federation approaches to address the need for

ubiquitous connectivity in 6G networks and to develop frameworks for service deployment and experimentation. However, these previous initiatives faced several limitations, such as low adoption due to the non-standardized nature of their solutions, restricted scope, and insufficient scalability to meet the 6G requirements for an orchestration continuum throughout the end-to-end network and across all planes [5].

A new European project, named SUNRISE-6G [6], is currently trying to tackle this matter by proposing a scalable, open, and federated approach to experimentation and vertical application deployment in a European conglomerate of 6G infrastructures that provides access to a comprehensive library of 6G enablers.

## Evolution of the federated ecosystem through SUNRISE-6G

The objective of SUNRISE-6G is to create a new experimentation facility for 6G networks by federating platforms and enablers across Europe into a unified infrastructure for testing, validation, and vertical application deployment. Inspired by the "network of networks" concept of 6G, SUNRISE-6G seeks to integrate several partner infrastructures into a massively scalable, internet-like architecture.

The project aims to establish a pan-European federation of 6G test infrastructures that will support Testing-as-a-Service (TaaS) workflows and tools. Within these infrastructures runs a diverse catalog of 6G resource enablers, cross-domain device, and vertical application onboarding enablers. These services will be served to a marketplace portal, serving as an entry point for service/resource consumers and infrastructure manufacturers.

Altice Labs' role in SUNRISE-6G is, alongside other partners, to design, develop, and integrate DLT with east/westbound Application Programming Interfaces (API) into the Federation Platform of its

architecture. This layer relies on a DLT software solution based on blockchain networks and open-source technology. The blockchain network ensures trust for network resources and devices in a decentralized environment, establishing immutable, non-repudiable transactions, using DLT for multi-stakeholder and multi-domain 6G infrastructures.

The communications between the existing domains that involve Blockchain operations, such as the request and issuance of Verifiable Credentials (VCs), will be supported by the innovative Verifiable Organizations Network (VON) [7], which will implement load-balancing mechanisms and operator wallet registration. Thus, members of the consortium can manage and regulate access granting of other newly onboarded members into the marketplace. These can then register their marketplace network resources, such

as Spectrum, Cloud Capacity, Edge Computing, Network, and Storage Capacity. **Figure 1** showcases the interactions between the onboarded resource operators (also known as stakeholders or service/resource providers), DLT infrastructure, and network resources in the marketplace of the federation platform.

The operators and network resources are identified by their unique VCs, whereas these VCs can provide control access mechanisms to, i.e., a marketplace portal and a type of resource. These are rooted on a Blockchain through a DLT so that entities in all domains (Cloud, Edge, Network, etc.) are interoperable across several testbeds. Besides managing service providers and their network resources, this technology also makes it possible to issue VCs to an operator’s RAN device.

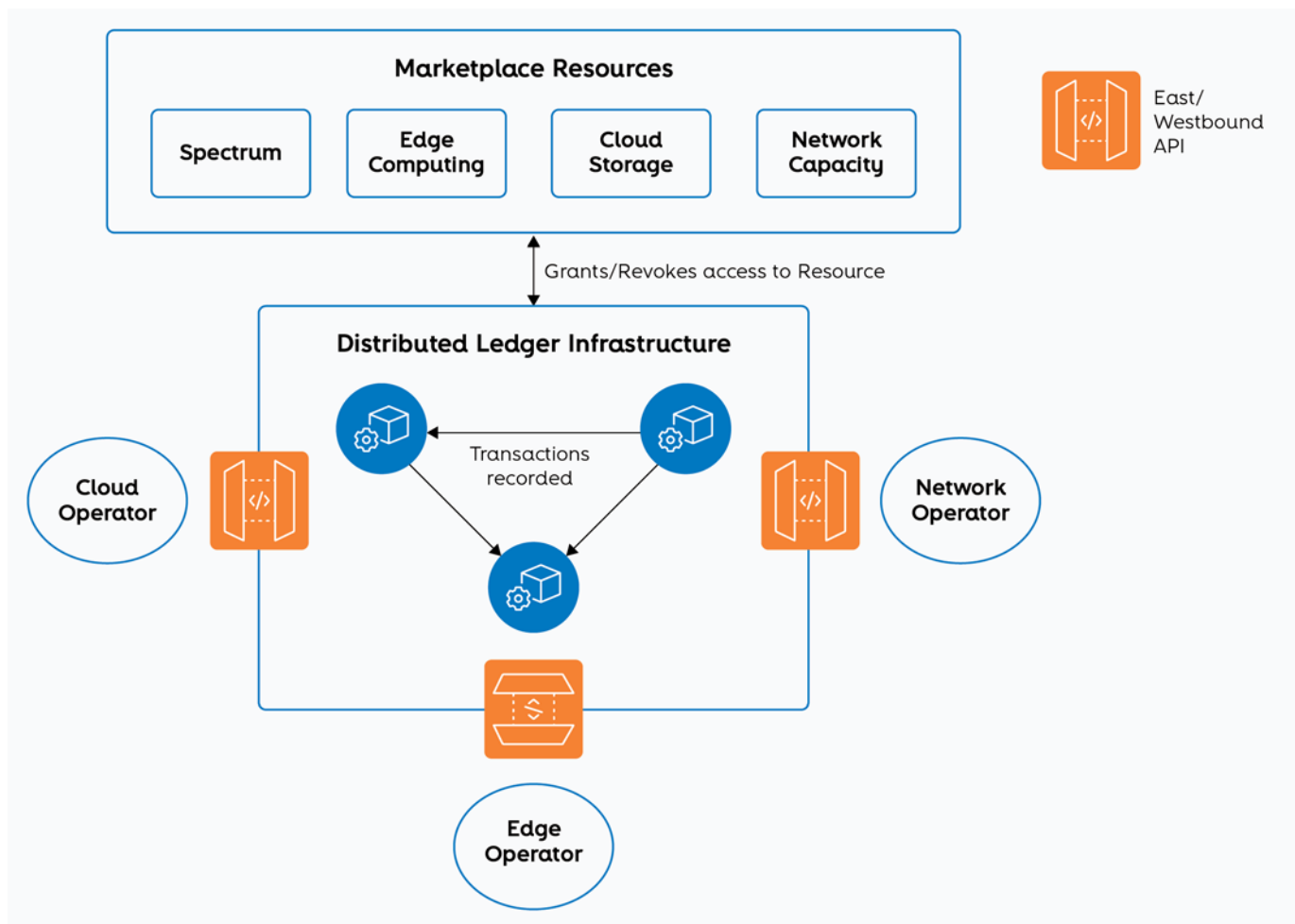


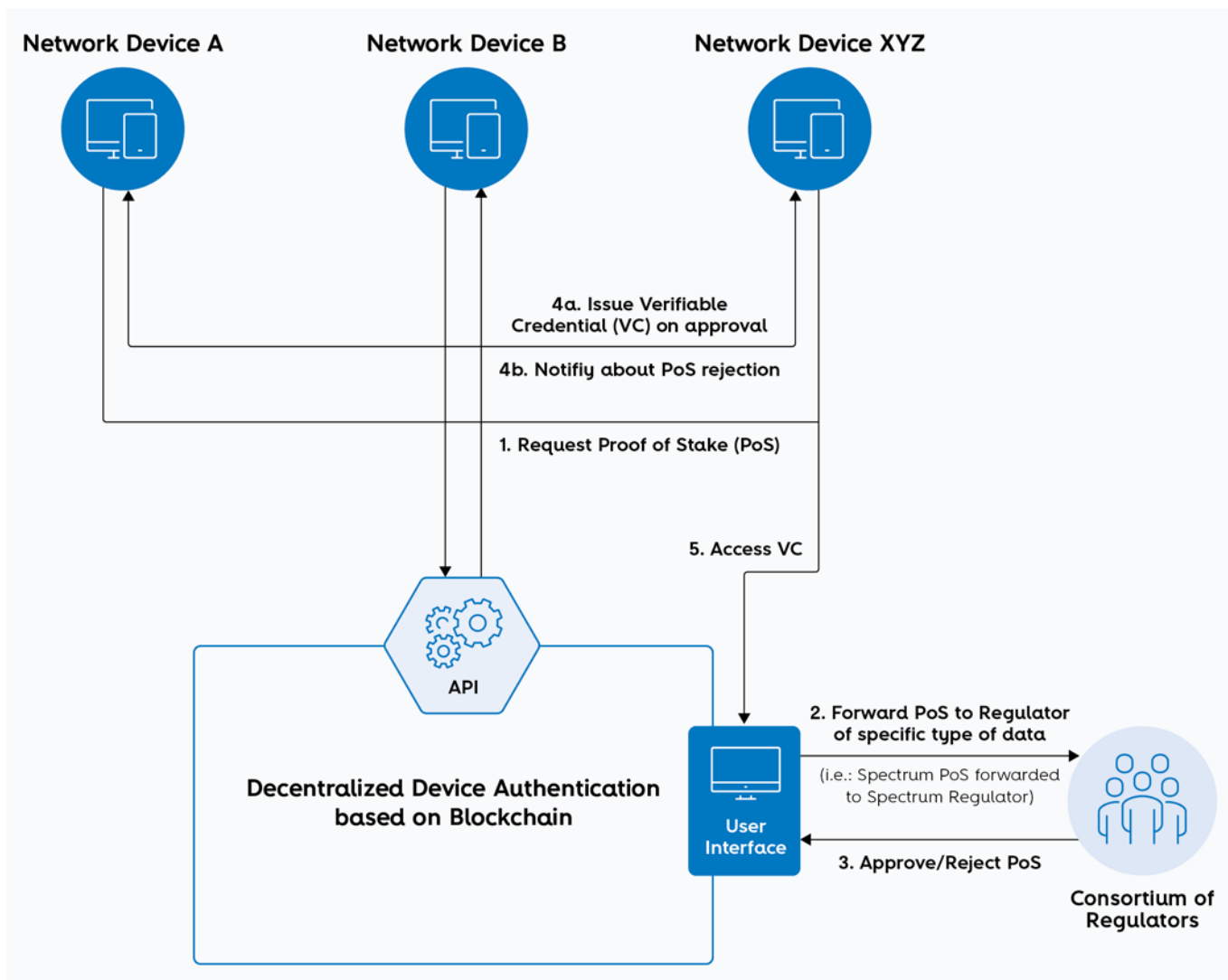
FIGURE 1 - Multi-stakeholder federation high-level architecture

**Figure 2** displays how Altice Labs’ component enables the process of VCs issuance to an operator’s network device that wants access to a marketplace resource.

The simplified steps for issuing and obtaining a VC are as follows:

1. A Proof-of-Stake (PoS) is requested by a representative of the service provider. The user can follow the status of the request on a marketplace portal;
2. The request is sent to the Regulator of the type of resource to be marketed, where a unique identifier is created to globally identify the proposal;

3. After receiving notification, the Regulator analyzes the information associated with the proposal through the portal;
  - a) The Regulator either approves or rejects the PoS claim;
  - b) In the event of approval, a VC will be generated and recorded as a transaction on the Blockchain;
4. The Regulator decides on the PoS:
  - a) The Regulator either approves or rejects the PoS claim;
  - b) In the event of approval, a VC will be generated and recorded as a transaction on the Blockchain;
5. The Credential is then transmitted to the service provider via the blockchain agents of the two entities, where it can consult the VC via the marketplace portal.



**FIGURE 2** – Operator Verifiable Credential issuance steps



## Network federation challenges: state of the art

6G is envisioned to foster a revolution in digital transformation and will accelerate the building of smart societies, leading to quality-of-life improvements. However, it also comes with a unique set of challenges due to the unprecedented capacity, efficiency, and infrastructure flexibility requirements [8].

As interest grows, so does the number of network devices, each of which with its specific security and authorization policies. This leads to an increase in the complexity of networks, as well as in the number of connected devices and users [9]. To track these devices, the concept of a Blockchain-based network device management framework has emerged [10].

Blockchain is a DLT that allows data records to be stored and shared in a decentralized and secure network. This technology is essentially a digital ledger, which monitors transactions and information in a permanent, transparent, and tamper-proof manner. In this way, it offers transparency, security, and reliability, which makes it promising for solving trust and security problems in various industries and sectors [11], [12], [13].

The use of permissioned DLT infrastructures facilitates the tokenization of telecommunication resources [14] (i.e., spectrum, virtualized radio access, virtualized edge/core, etc.). Direct Smart Contracts [15] execution, which are connected to these resource tokens, responds to the requirements of vertical industries, whilst unblocking the collaboration between multiple parties not relying on pre-established trust [16].

The evolution towards 6G requires the telco infrastructures to adopt multi-tenant architectures and

to provide Edge integration in a global marketplace [17]. This multi-party collaboration in a dynamic environment where operators and service providers often need to employ third-party service chains to satisfy a contract is a necessity of next-generation networks [18]. The resource consumers interact with a marketplace portal to obtain the set of product offers available and, upon acquiring a resource or service that satisfies their need, a Smart Contract is established between the two parties. All Smart Contracts are recorded on the Marketplace’s DLT. **Figure 3** showcases how the third-party services interact with the marketplace platform.

While this approach may be sufficient to run a federated marketplace, it leaves the customer to search over provided offers to find the ones that best match some scenario or optimization criteria, which quite often may imply more complex trade-offs. In addition, the customer also has the challenge of filtering configurations of several selected resources spread across multiple domains. For this reason, the marketplace should employ zero-touch

capabilities to enable acquiring and configuring a customized subset of resources that best satisfy consumer expectations [19].

Another crucial requirement for the marketplace is its trustworthiness. Without the trusted environment to trade resources among multiple stakeholders, the whole aim of the marketplace could be jeopardized. These requirements can be fulfilled by anchoring the marketplace resource in Distributed Ledger Technologies, mainly because security, trustworthiness, and distributed structure are intrinsic to a DLT, as previously mentioned [20].

Furthermore, to digitally manifest the resource trade between multiple stakeholders, the marketplace should make use of VCs. These enable the marketplace to automatically execute, control, or document legally relevant events and actions according to the terms of a contract or an agreement. The solution needs to allow for a granular approach to heterogeneous resources’ fulfillment terms across multi-stakeholders to formulate customized product offers [21], [22], [23].

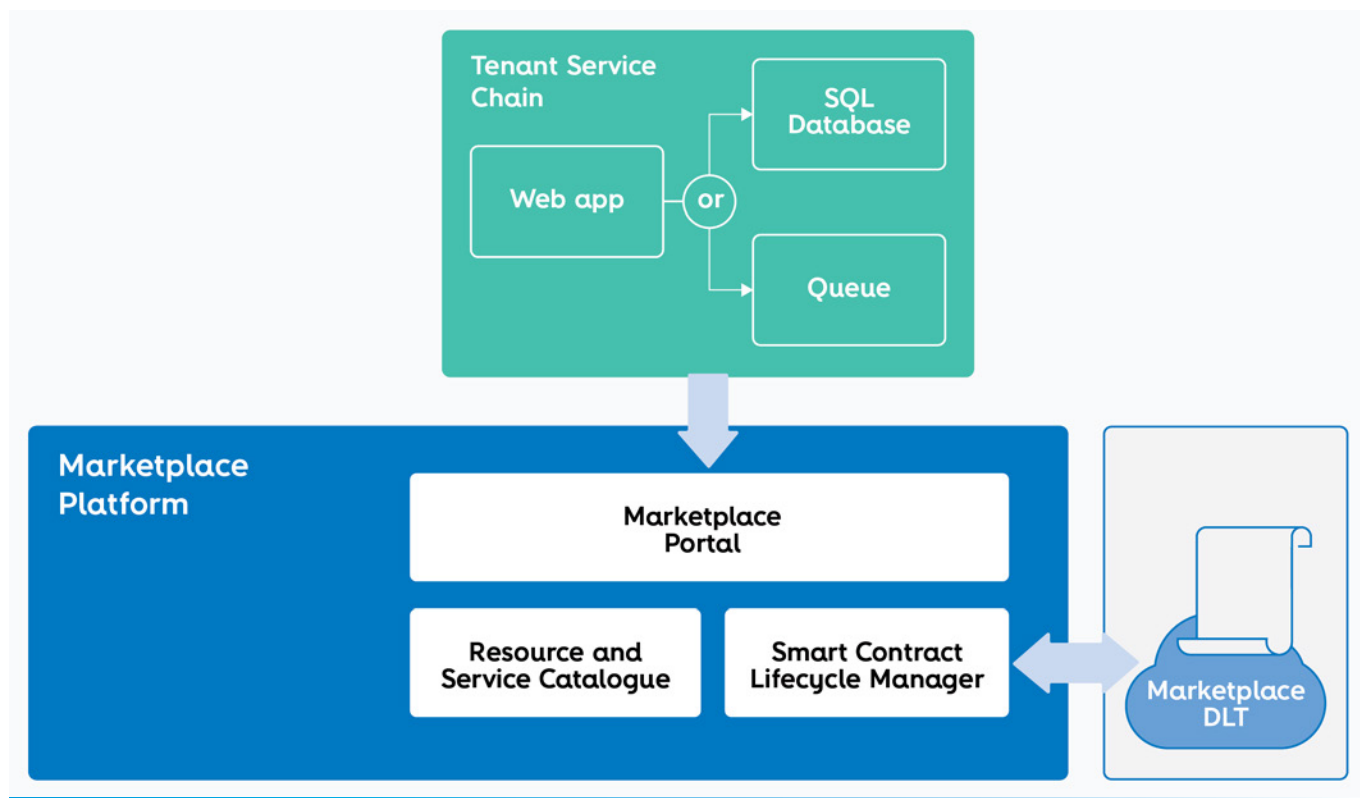


FIGURE 3 – Marketplace platform with DLT

## Conclusion

New European projects are looking at innovative ways to implement architectures with a single, unifying, and openly controllable framework that manages a federated pool of networking and computational resources. By offering an architected Business Layer based on Distributed Ledger Technologies, the SUNRISE-6G project offers a new method for NFE orchestration and control challenges. This approach is supposed to contribute to the following competitive advantages:

- **Immutability:** blockchain is known for its immutability [12], [13], [15]. This means that once registered on the Blockchain, information about partners cannot be easily tampered with or falsified. This contributes to the security of authentication;
- **Secure transactions with VCs:** VCs can be used to automate actions in device interactions [21], [22], [23]. For example, a network device can be configured to perform certain tasks only if its digital signature is legitimate;
- **Privacy:** the proper implementation of Blockchain can allow the authentication of third parties without exposing sensitive information about them, protecting their privacy and security [11], [14], [15], [16];
- **Decentralization:** one of the advantages of Blockchain is its decentralized nature [12], [13], [16]. In practice, this means that for each type of information/technology, there is a specific regulator.

With Altice Labs' contribution to device management in an NFE, the project can ensure that businesses network-wise can operate efficiently, securely, and reliably. A correct configuration of network devices and their continuous management, whilst defining access permissions for the devices' users, will help build a dynamic and decentralized marketplace. This, in return, supports dynamically adding and removing infrastructure resources as they are needed, regardless of their type or location, to provide a sustainable and multi-party system able to cope with heterogeneous service tenants' requirements. 🌐

## Acknowledgement

The work presented in this paper is supported by the SUNRISE-6G project that has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101139257.

## References

- [1] D. Saraswat et al., "Blockchain-Based Federated Learning in UAVs Beyond 5G Networks: A Solution Taxonomy and Future Directions," *IEEE Access*, vol. 10, pp. 33154–33182, 2022, doi: 10.1109/ACCESS.2022.3161132.
- [2] A. Alghamdi et al., "Blockchain Empowered Federated Learning Ecosystem for Securing Consumer IoT Features Analysis," *Sensors*, vol. 22, no. 18, 2022, doi: 10.3390/s22186786.
- [3] Z. Li, V. Sharma, and S. P. Mohanty, "Preserving Data Privacy via Federated Learning: Challenges and Solutions," *IEEE Consumer Electronics Magazine*, vol. 9, no. 3, pp. 8–16, 2020, doi: 10.1109/MCE.2019.2959108.
- [4] G. Chandrasekaran, A. Kumar, K. Subramaniam, N. Sunil, and R. Challa, "EDgE: Enhancing Predictability in Decentralized Network Slices Through Federated Learning—A Multi Layered Ecosystem," *IEEE Access*, vol. 12, pp. 100957–100971, 2024, doi: 10.1109/ACCESS.2024.3429631.
- [5] B.M. Khorsandi et al., "Deliverable D1.3: Targets and requirements for 6G - initial E2E architecture," 2022.
- [6] "SUNRISE-6G," sunrise6g.eu. Accessed: Aug. 6, 2024. [Online.] Available: <https://sunrise6g.eu/>.
- [7] Government of British Columbia, "Digital trust," digital.gov.bc.ca. Accessed: Aug. 6, 2024. [Online.] Available: <https://digital.gov.bc.ca/digital-trust/home/>.
- [8] M. W. Akhtar, S. A. Hassan, R. Ghaffar, H. Jung, S. Garg, and M. S. Hossain, "The shift to 6G communications: vision and requirements," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, p. 53, 2020, doi: 10.1186/s13673-020-00258-2.
- [9] S. Nižetić, P. Šolić, D. López-de-Ipiña González-de-Artaza, and L. Patrono, "Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future," *J Clean Prod*, vol. 274, p. 122877, 2020, doi: 10.1016/j.jclepro.2020.122877.
- [10] S. Gong, E. Tcydenova, J. Jo, Y. Lee, and J. H. Park, "Blockchain-Based Secure Device Management Framework for an Internet of Things Network in a Smart City," *Sustainability*, vol. 11, no. 14, 2019, doi: 10.3390/su11143889.
- [11] K. Kiania, S. M. Jameii, and A. M. Rahmani, "Blockchain-based privacy and security preserving in electronic health: a systematic review," *Multimed Tools Appl*, vol. 82, no. 18, pp. 28493–28519, 2023, doi: 10.1007/s11042-023-14488-w.
- [12] Y. Chen, Y. Lu, L. Bulysheva, and M. Yu. Kataev, "Applications of Blockchain in Industry 4.0: a Review," *Information Systems Frontiers*, 2022, doi: 10.1007/s10796-022-10248-7.
- [13] U. Bodkhe et al., "Blockchain for Industry 4.0: A Comprehensive Review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020, doi: 10.1109/ACCESS.2020.2988579.
- [14] B. Hanneke, O. Hinz, J. Pfeiffer, and W. M. P. van der Aalst, "The Internet of Value: Unleashing the Blockchain's Potential with Tokenization," *Business & Information Systems Engineering*, vol. 66, pp. 411–419, 2024, doi: 10.1007/s12599-024-00883-6.

- [15] S. N. Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa, and A. Bani-Hani, "Blockchain smart contracts: Applications, challenges, and future trends," *Peer Peer Netw Appl*, vol. 14, no. 5, pp. 2901–2925, 2021, doi: 10.1007/s12083-021-01127-0.
- [16] W. Serrano, "Verification and Validation for data marketplaces via a blockchain and smart contracts," *Blockchain: Research and Applications*, vol. 3, no. 4, p. 100100, 2022, doi: 10.1016/j.bcra.2022.100100.
- [17] N. Afraz, M. Ruffini, and H. Ahmadi, "Hyperledger Blockchain-Based Distributed Marketplaces for 5G Networks," *Wireless Blockchain: Principles, Technologies and Applications*, pp. 117–136, 2022. doi: 10.1002/9781119790839.ch5.
- [18] A. Fernández-Fernández et al., "Multi-Party Collaboration in 5G Networks via DLT-Enabled Marketplaces: A Pragmatic Approach," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Porto, Portugal, 2021, pp. 550–555. doi: 10.1109/EuCNC/6GSummit51104.2021.9482487.
- [19] V. Theodorou et al., "Blockchain-based Zero Touch Service Assurance in Cross-domain Network Slicing," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Porto, Portugal, 2021, pp. 395–400. doi: 10.1109/EuCNC/6GSummit51104.2021.9482602.
- [20] B. Nour, A. Ksentini, N. Herbaut, P. A. Frangoudis, and H. Moun gla, "A Blockchain-Based Network Slice Broker for 5G Services," *IEEE Networking Letters*, vol. 1, no. 3, pp. 99–102, 2019, doi: 10.1109/LNET.2019.2915117.
- [21] P. N. Mahalle, G. Shinde, and P. M. Shafi, "Rethinking Decentralised Identifiers and Verifiable Credentials for the Internet of Things," in *Internet of Things, Smart Computing and Technology: A Roadmap Ahead*, Cham: Springer International Publishing, 2020, pp. 361–374, doi: 10.1007/978-3-030-39047-1\_16.
- [22] C. Brunner, U. Gellersdörfer, F. Knirsch, D. Engel, and F. Matthes, "DID and VC: Untangling Decentralized Identifiers and Verifiable Credentials for the Web of Trust," in *Proceedings of the 2020 3<sup>rd</sup> International Conference on Blockchain Technology and Applications*, in ICBTA '20, New York, NY, USA: Association for Computing Machinery, 2021, pp. 61–66, doi: 10.1145/3446983.3446992.
- [23] Z. A. Lux, D. Thatmann, S. Zickau, and F. Beierle, "Distributed-Ledger-based Authentication with Decentralized Identifiers and Verifiable Credentials," in *2020 2<sup>nd</sup> Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, Paris, France, 2020, pp. 71–78. doi: 10.1109/BRAINS49436.2020.9223292.

# 08

## Impact of Passive Optical Network technology on polarization encoded Discrete Variable Quantum Key Distribution



With the advent of Quantum Computing threatening public key cryptography, Quantum Key Distribution (QKD) is bringing quantum communications from somewhere in the distant future to our business reality. And for a Telco, that reality is also one of enormous investment in fiber infrastructure. New services will be highly welcome, but the capacity to support them must be assessed. So, before we even consider the introduction of quantum-based security in our fiber-based network solutions, a question needs to be answered: can a Passive Optical Network (PON) share the fiber with a very low power, highly sensitive technology where single photons are the raw material for communications? This article investigates the impact of PON technology on Discrete Variable Quantum Key Distribution (DV-QKD), using Gigabit Passive Optical Network (GPON) and 10 Gigabit Symmetric Passive Optical Network (XGS-PON) technologies over a fiber loop between Altice Labs and the Instituto de Telecomunicações (IT).

---

## Authors

**Diogo Cruz, Altice Labs, Instituto de Telecomunicações and University of Aveiro, [diogovcruz@ua.pt](mailto:diogovcruz@ua.pt)**

**Nelson J. Muga, Instituto de Telecomunicações and University of Aveiro, [muga@ua.pt](mailto:muga@ua.pt)**

**Sara Mantey, Instituto de Telecomunicações and University of Aveiro, [smantey@ua.pt](mailto:smantey@ua.pt)**

**Ana Tavares, PicAdvanced, [ana@picadvanced.com](mailto:ana@picadvanced.com)**

**Cláudio Rodrigues, Altice Labs, [claudio-e-rodrigues@alticelabs.com](mailto:claudio-e-rodrigues@alticelabs.com)**

---

## Keywords

BB84; CV-QKD; DV-QKD; GPON; ODN; PON; QKD; SpRS; XGS-PON



## Introduction

Advancements in the telecommunications sector, such as the expansion of PONs as a medium for transmitting information and the increasing volume of information being transmitted, have made it essential that the technology that secures the confidentiality of said information keeps up with the threats to that same confidentiality.

Currently, when encrypting information, two types of cryptography are used: symmetric cryptography, mostly used to encode the message being sent, using an encryption key that is used to encrypt and decrypt the message (hence, symmetric), and asymmetric cryptography, which generates two encryption keys, one public and one private, that are used to share a symmetric key used to decode the message [1], [2], [3].

The challenge modern cryptography faces is the advent of a quantum computer. Quantum computers leverage quantum mechanics properties such as superposition to exponentially increase processing power with each qubit (a qubit would be equivalent to a bit in classical computing) that they have available [3], [4]. Symmetric cryptography should be able to weather the threat of a quantum computer capable of running Groover's algorithm. However, asymmetric cryptography will most likely collapse against a quantum computer capable of running Shor's algorithm. It is expected that the 1024-bit version of the Rivest-Shamir-Adleman (RSA) algorithm, a widespread asymmetric algorithm, could be broken by a 2000 qubit quantum computer [3].

To address this challenge, new solutions are necessary, and Post-Quantum Cryptography presents two paths: to use more complex problems that do not present the same vulnerabilities of current asymmetric cryptography or to exploit the properties of quantum mechanics to ensure that it is not possible to eavesdrop on the encryption keys used. This last approach is referred to as QKD.

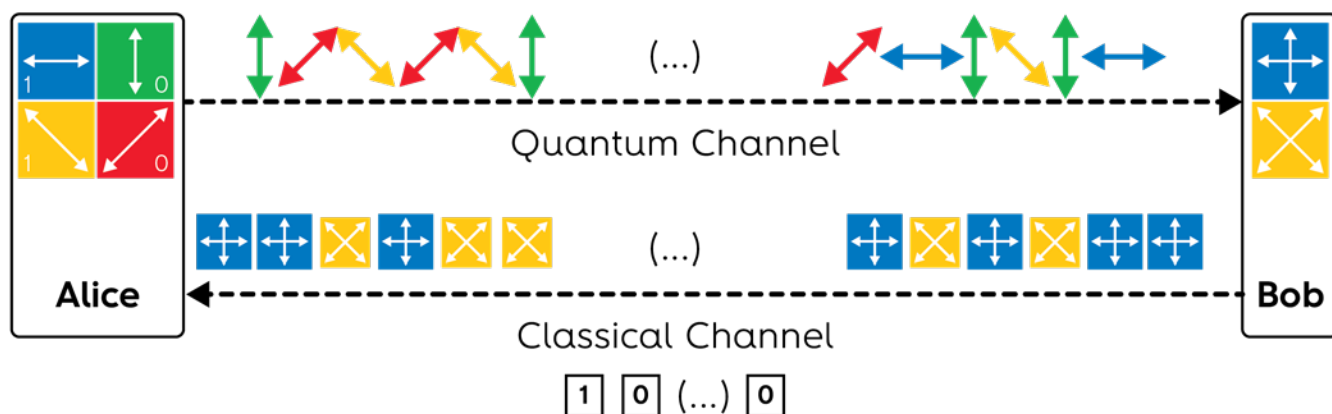
QKD protocols are categorized depending on how the information is encoded and processed to result in

a secure key exchange. If the protocol uses continuous properties of light, such as its amplitude or phase, it is categorized as Continuous Variable Quantum Key Distribution (CV-QKD). CV-QKD is faster, and the technology is potentially ready to be integrated into photonic integrated circuits. However, due to its sensitivity to noise, it has only been implemented for medium transmission distances [5], [6]. On the other hand, when the QKD protocol encodes the information on the properties of individual photons, such as their polarization or their phase state, is categorized as DV-QKD. DV-QKD protocols generally have a lower Secret Key Rate (SKR) than CV-QKD, but they are capable of being implemented over longer transmission distances, since they are more resistant to the noise issues that affect CV-QKD [5], [6].

## The Bennet-Brassard 84 protocol

The Bennet-Brassard 1984 (BB84) protocol (schematically represented in **Figure 1**) is the first QKD protocol proposed and is proven to be theoretically secure [7]. It works by using the polarization (or angle) of single photons to encode the information needed to generate a secure cryptographic key between two parties [8].

The security of the BB84 protocol (a DV-QKD protocol) relies on the uncertainty principle and on the quantum no-cloning theorem, which says that it is not possible to make an exact copy of an unknown



Alice's data	1	1	0	1	0	1	...	1	0	1	0	1	0	
Alice's basis	$\oplus$	$\otimes$	$\otimes$	$\otimes$	$\otimes$	$\oplus$	...	$\otimes$	$\oplus$	$\oplus$	$\otimes$	$\oplus$	$\oplus$	
$\theta$	$0^\circ$	$45^\circ$	$135^\circ$	$45^\circ$	$135^\circ$	$0^\circ$	...	$45^\circ$	$90^\circ$	$0^\circ$	$135^\circ$	$0^\circ$	$90^\circ$	
	$\oplus$	$\oplus$	$\otimes$	$\oplus$	$\otimes$	$\otimes$	...	$\oplus$	$\otimes$	$\oplus$	$\otimes$	$\oplus$	$\oplus$	Bob's basis
	1	0	0	1	0	0	...	0	0	1	0	1	0	Bob's result
Same basis?	✓	✗	✓	✗	✓	✗	...	✗	✗	✓	✓	✓	✓	
Sifted key	1	-	0	-	0	-	...	-	-	1	0	1	0	
Error correction	✓		✗		✓		...			✗	✓	✗	✗	Was the qubit sent?
Private key			0				...			1		1	0	

**FIGURE 1** – Illustration of the different steps that comprise the BB84 protocol

quantum state, such as a qubit carried by a photon. Consequently, when someone tries to intercept the information being sent, they will inevitably leave a trace [2].

In the BB84 protocol, two parties, traditionally named Alice and Bob, use two sets of linearly polarized basis to prepare and measure qubits that are employed to generate an encryption key:

- $\oplus$ , in which bits “1” and “0” correspond to qubits codified with polarization angles ( $\theta$ ) of  $0^\circ$  and  $90^\circ$ , respectively, and are represented by  $|\leftrightarrow\rangle$  and  $|\updownarrow\rangle$ .
- $\otimes$ , in which bits “1” and “0” correspond to qubits codified with polarization angles ( $\theta$ ) of  $45^\circ$  and  $135^\circ$ , respectively, and are represented by  $|\nearrow\rangle$  and  $|\nwarrow\rangle$ .

For example, to encode the bit “1”, Alice would choose either  $|\updownarrow\rangle$  or  $|\nwarrow\rangle$ , while Bob would randomly choose one of the bases with which he would measure the qubits. If Bob picks the same basis as Alice, he will see the exact bit Alice sent. If not, there will be a 50% chance to measure the right bit.

The implementation of the QKD protocol between Alice and Bob uses two channels: a quantum channel, used by Alice to send her qubits to Bob, and a classical channel (assumed to be insecure, but authenticated), used by Bob to communicate to Alice his choice of bases after he has received all of Alice’s qubits. Alice then compares her choice of bases with Bob’s choice. After discarding the instances in which they chose different bases, Alice can use the remaining set of bases to generate the sifted key. Subsequently, Alice communicates to Bob the intervals of time in which they chose the same basis so that Bob has access to the sifted key [2].

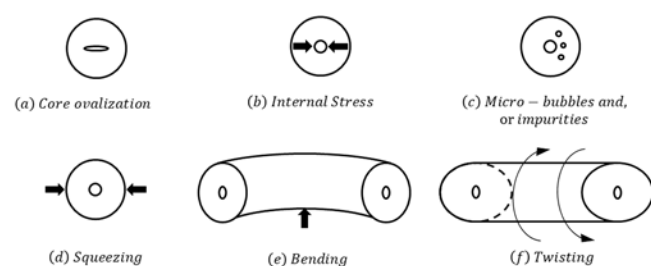
The QKD sifted keys undergo two classical operations. The first is error correction, comprising a stage where Bob transmits a subset of his bits from the sifted key to Alice. Alice then compares Bob’s subset with her own, and if the error rate is below 11%, the protocol can proceed to the next stage [7]. Here, classical error correction algorithms are used

to correct measurement errors and establish the same key on both sides. The second classical operation is privacy amplification. Here, from a partially secure key, a smaller key that meets the error threshold can be distilled, reducing the amount of information an eavesdropper could have obtained [7], [9], [10].

## Quantum Signals and coexistence with PON technology

To maximize the capacity of optical networks, multiple signals are transmitted through the same fiber in a process called coexistence. Even though each of the optical signals has its own modulation format or wavelength, coexistence is possible thanks to the large bandwidth that optical fibers provide. Techniques such as Wavelength Division Multiplexing (WDM) make it feasible to transmit multiple data streams through the same fiber, where each signal is characterized by its respective wavelength, increasing the efficiency of the PON [11].

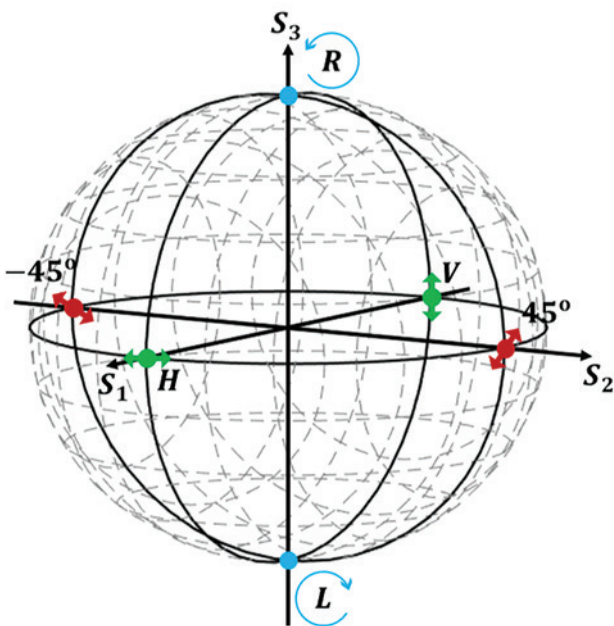
When a polarized optical signal is transmitted through an optical fiber, the polarization of the photons in the quantum channel can be affected by the characteristics of the fiber, such as birefringence (**Figure 2**). This includes both intrinsic factors, like core ovalization, internal stress, and impurities, as well as extrinsic factors, such as squeezing, bending, and twisting [12].



**FIGURE 2** – Birefringence mechanisms potentially affecting the state of polarization (adapted from [12])

When represented in the Stokes formalism, a photon's polarization can be described using four parameters,  $S_0$ , which represents the total intensity of the light,  $S_1$ , which quantifies the light with linear horizontal against linear vertical polarization,  $S_2$ , which measures the quantity of linear +45° or 135° polarized light, and  $S_3$ , which describes the quantity of left or right circularly polarized light [12], [13].

Parameters  $S_1$ ,  $S_2$ , and  $S_3$ , when normalized, can be represented using the Poincaré sphere (see **Figure 3**), which provides a very convenient representation of all possible States of Polarization (SOP) that a photon can have.



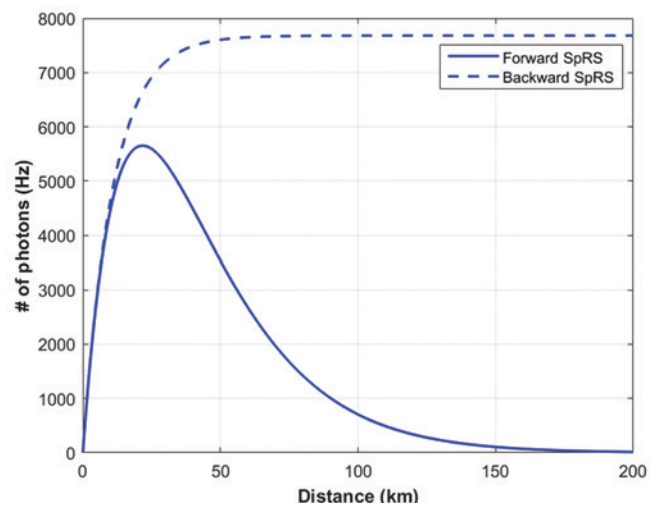
**FIGURE 3** – Poincaré sphere representation of light

The Poincaré sphere has a radius of 1. The points on the equator correspond to linearly polarized light, while the poles of the sphere represent right-hand and left-hand circularly polarized light. Points between the poles and the equator correspond to elliptically polarized light.

The quantum channel can also be affected by nonlinear phenomena, such as Four-Wave Mixing (FWM) and Spontaneous Raman Scattering (SpRS), caused by the higher power classical signals, introducing errors in the quantum system.

The focus of this study will be on SpRS, a phenomenon that results from the interaction of a classical signal with the molecules of the optical fiber core. This interaction produces photons with different wavelengths [14], [15], some of which may have wavelengths similar to those in the quantum signal, introducing errors in the protocol.

SpRS photons can scatter in all directions. However, because PONs use single-mode fibers, the only possible propagation directions are forward or backward. Using the equations provided in [16], the expected evolution of the photon count rate in the forward and backward directions should follow the trend represented in **Figure 4**.

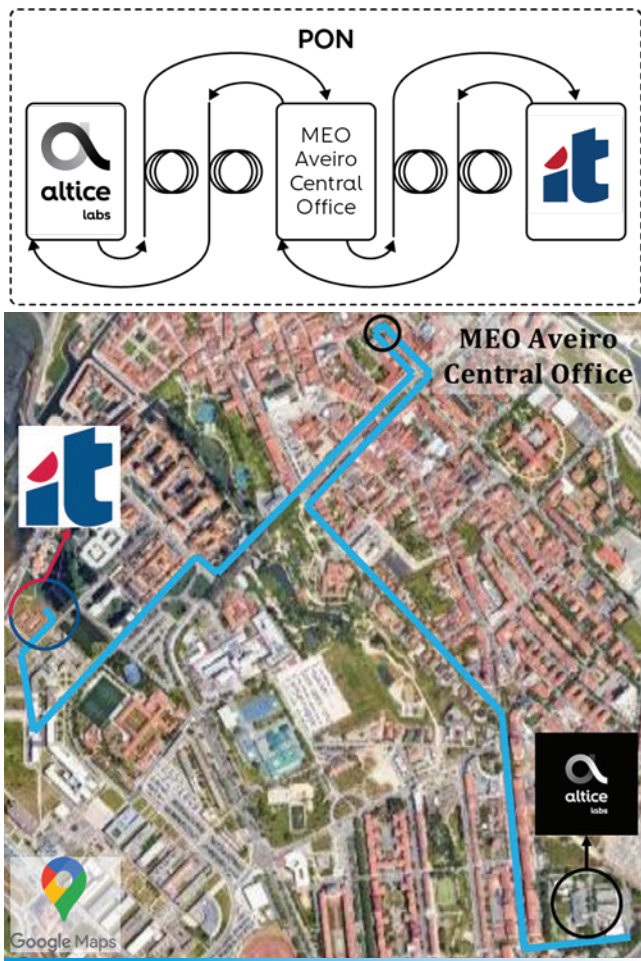


**FIGURE 4** – Representation of the photon count rate as a function of distance travelled, considering the cases of forward and backward propagation

## Characterization techniques

To test the coexistence capability of a legacy network to act as a quantum network, a small Optical Distribution Network (ODN) was established between Instituto de Telecomunicações (IT), located in the Campus of the University of Aveiro,

and Altice Labs, in Aveiro, via MEO Aveiro Central Office. The ODN was measured to have 7.39 km and is schematically represented in **Figure 5**.

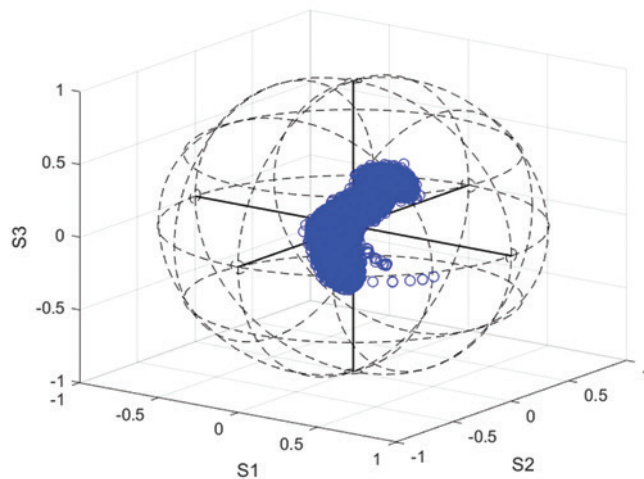


**FIGURE 5** – Schematic of the ODN established between Altice Labs and IT (on the top) and the satellite view of the optical fiber pathway, which span approximately 3.27 km (on the bottom)

## Assessment of Polarization Stability

To access the stability of the link, an optical signal emitted from an OSICS/TUNICS NetTest External Cavity Laser (ECL) located at IT’s Quantum Communications Lab was sent through the fiber. After completing a roundtrip on the ODN, the Stokes parameters  $S_1$ ,  $S_2$ , and  $S_3$  were measured using the ThorLabs PAX5710IR1-T, and their drift with time was measured for approximately 16 hours (the results

are displayed in **Figure 6**). A perfectly stable system would show a single point in the Poincaré sphere. The data points in **Figure 6** are in a relatively tight cluster with limited drift, suggesting that there is a reasonable state of polarization stability.

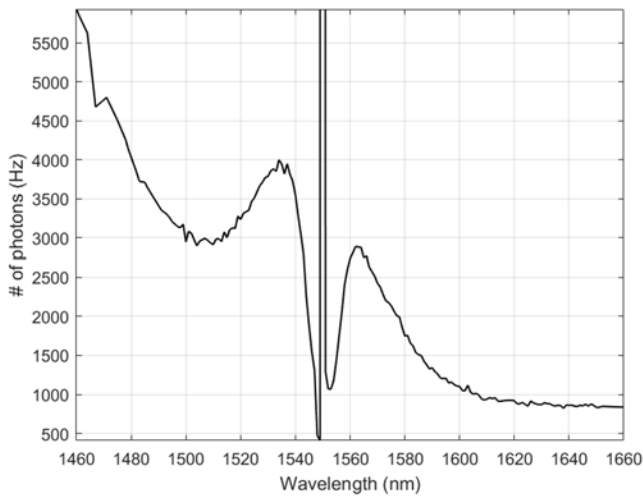


**FIGURE 6** – Poincaré sphere representation of the Stokes parameters measured in the Quantum Communications Lab, over a period of approximately 16 hours

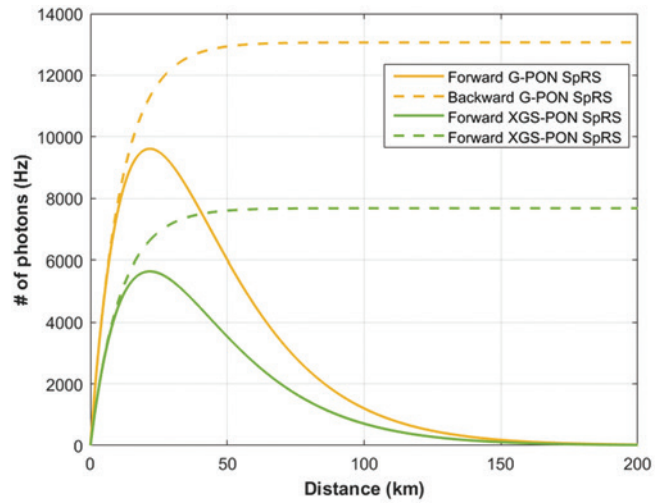
Due to how the PON was set up, it is complicated to compensate for the factors affecting the measurement. Within the Altice Labs and the IT, the fibers can still be exposed to variations in temperature and vibrations, even if the fiber patches are carefully secured to mitigate these factors.

## Spontaneous Raman Scattering assessment

To test the impact of SpRS of the quantum channel, the quantum signal was set to 1550.12 nm, and a tunable laser, Anritsu Osics Tunics ECL, model 1520 or Anritsu Osics Tunics ECL, model 1600, was used to predict how each wavelength could affect the quantum channel. The ECL was connected to the optical fiber via a set of three filters centered at 1550.12 nm. A Single Photon Detector (SPD), idQuantique id200, was used to measure the backward SpRS crosstalk to the quantum channel. The results obtained are displayed in **Figure 7**.



**FIGURE 7** – Experimental data of the evolution of the photon count rate as a function of wavelength, resulting from SpRS backscattering



**FIGURE 8** – Photon count rate to the quantum channel, centered around 1550.12 nm, resulting from forward and backward SpRS for G-PON (1490 nm) and XGS-PON (1577 nm)

Using these results and the equations in [16] that were used to predict the evolution of the photon count rate in **Figure 4**, the expected evolution of the photon count rate with distance traveled in the fiber for GPON and XGS-PON was obtained and is displayed in **Figure 8**.

Forward SpRS is expected to increase until approximately 22 km, at which point the SpRS starts to decrease. Backward SpRS increases until approximately 25 km, at which point SpRS saturates. These trends are true for both the GPON and XGS-PON signals.

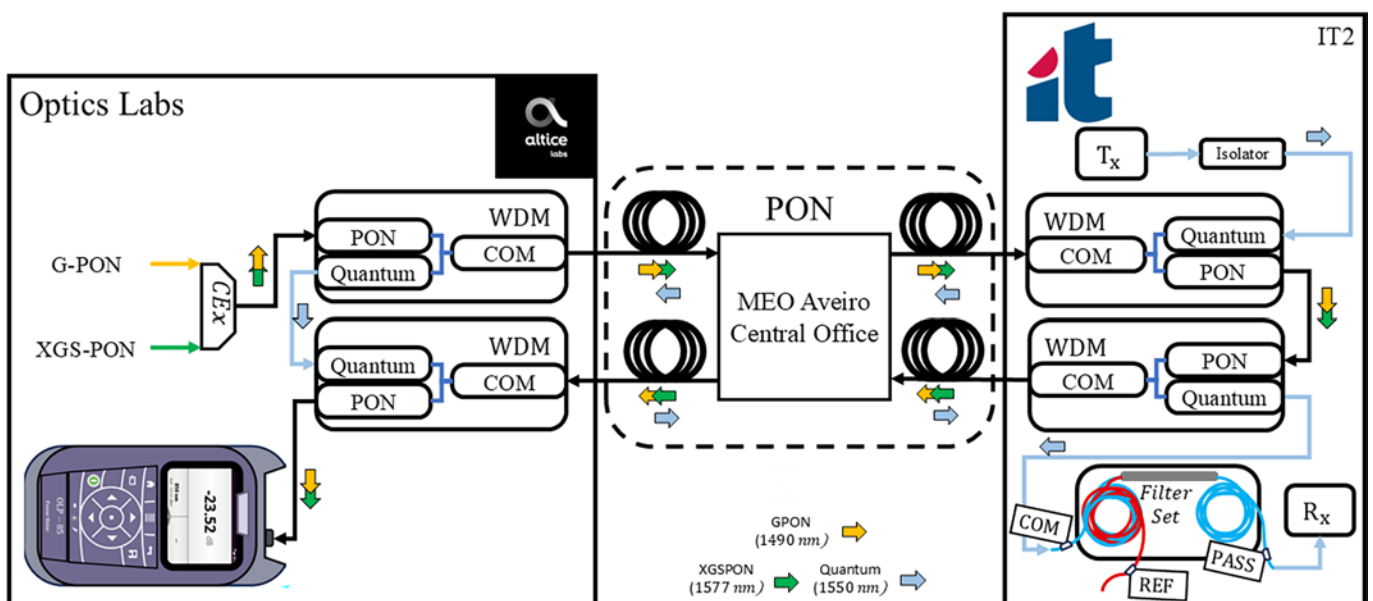


## Coexistence in Passive Optical Networks

To evaluate the possibility of coexistence between GPON and XGS-PON with a quantum signal, the Intrinsic Quantum Bit Error Rate (IQBER) was used as a figure of merit. To observe the impact a classical signal can have on the IQBER, a quantum signal with a constant polarization is sent through the fiber. The IQBER obtained from this measurement will serve as a reference for the cases of copropagation and will be referred to as Quantum IQBER. Then, the classical signals GPON (1490 nm) or XGSPON (1577 nm) are multiplexed into the PON, and their impact on the IQBER is evaluated. The experimental setup used is represented in **Figure 9**. The WDM equipment is centered at 1550 nm. On the IT's side, Tx and Rx comprise a source of polarized single photons and a pair of single photon detectors, respectively. The Filter Set is the same one used to measure Raman backscatter. At Altice Labs' side, the GPON and/or the XGS-PON signal is injected in the PON, and its power is measured after a round trip through the PON. It is also worth

noting that the architecture implemented for the PON only allows backward SpRS to reach the photon detectors.

To account for the random events that affect the quantum signal, each time the classical signal power was changed, the photon counts were measured over a period of two minutes, before the classical signal power was changed again. Over the two-minute period, the average, minimum, and maximum values of the IQBER were calculated. The results are displayed in **Figure 10**. The scenario with GPON alone (represented in **Figure 10.a**) shows moderate stability. The average (Avg) IQBER remains above the Quantum IQBER, and for power levels above -2 dBm, a trend towards an increase in IQBER with power sent to the fiber is noticeable. In the XGS-PON scenario (represented in **Figure 10.b**), the Avg IQBER is the most stable, remaining close to the Quantum IQBER, even for higher power values. The scenario with coexistence with both GPON and XGS-PON (represented in **Figure 10.c**) shows the lowest Quantum IQBER, but also the most noticeable trend for IQBER to grow with power sent to the fiber.



**FIGURE 9** – Schematic representation of the setup used to measure the photon count rate of the quantum signal and control the power of the classical signals sent to the ODN

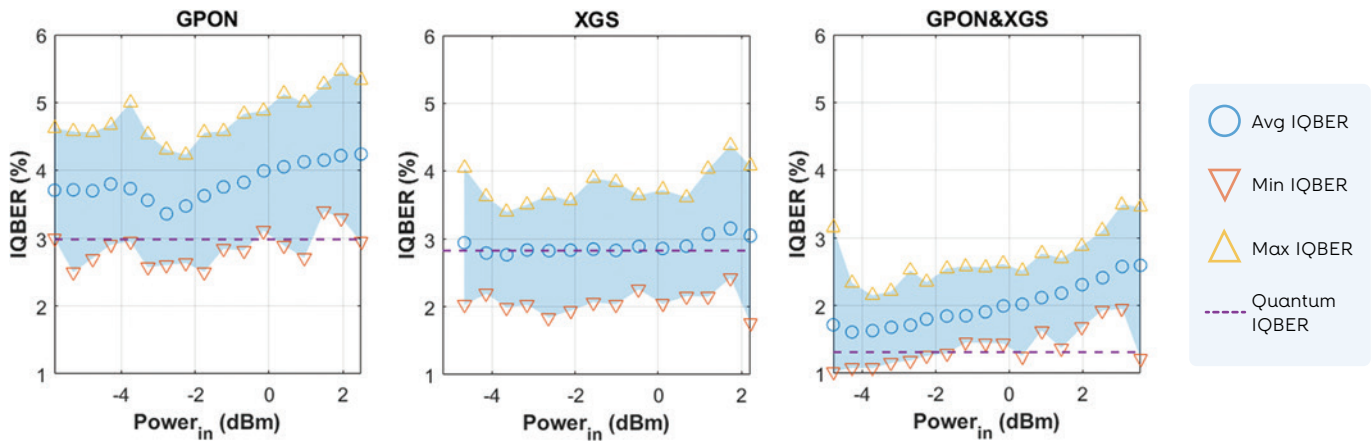


FIGURE 10 – IQBER of the quantum system as a function of power sent to the fiber

There are different values obtained for Quantum IQBER across the different scenarios. These values represent the IQBER achieved after optimizing the quantum channel’s optical visibility before classical signal propagation. Additionally, disturbances affecting the optical components of both the quantum receiver and transmitter can also introduce SOP instabilities. Together, these factors likely contribute to the observed variations in Quantum IQBER.

All the scenarios proved successful at working with IQBER below 11%. This confirms the effectiveness of the employed insulation of the quantum channel, even considering the relatively small size of the ODN. For 7.39 km, the SpRS should have a relatively small presence, based on Figure 8. Using a 25.3 km coil of standard single-mode fiber, the

setup was expanded to 32.7 km, and the scenarios were repeated to evaluate how SpRS would affect IQBER for a larger ODN. The results obtained for this scenario are displayed in Figure 11. The smaller number of experimental points is a consequence of the added attenuation to the ODN. It was not possible to recover the classical signals for as many values of power sent to the fiber.

Compared with the shorter ODN, the value of the Quantum IQBER has increased for all the scenarios, but the trend for IQBER to increase with power sent to the fiber is no longer evident. Instead, the IQBER appears to be largely independent of the power of the classical signals. This could be the result of the smaller range of experimental points in terms of optical power and the observed larger error margins.

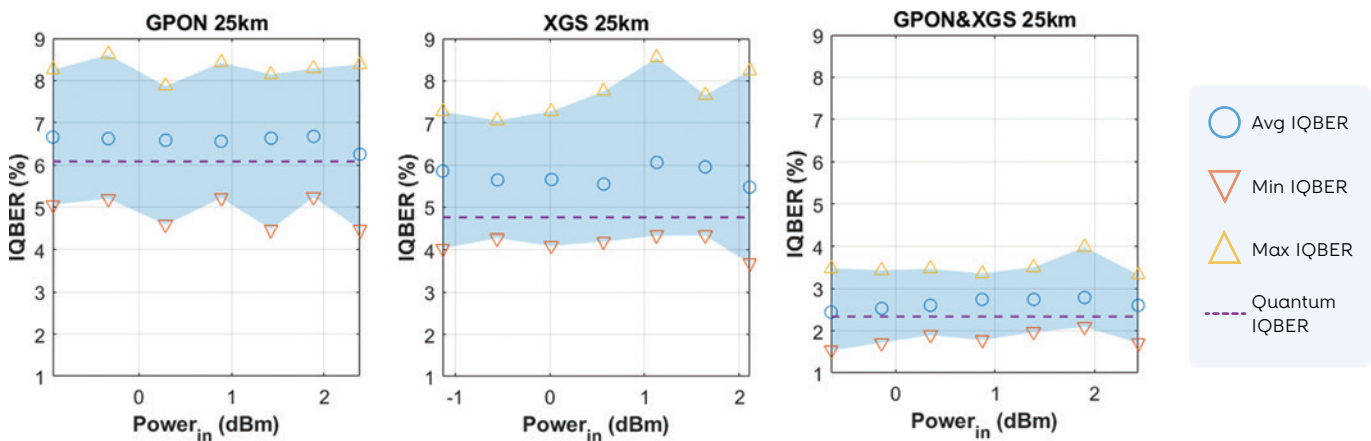


FIGURE 11 – IQBER of the quantum system as a function of power of classical signals sent to the ODN when a 25.3 km fiber is introduced in the upper branch of the ODN, as represented in Figure 10




## Conclusions

Regarding the characterization of the ODN, the polarization drift results suggest that the PON is stable and capable of providing a suitable medium to preserve the qubits. However, it is not completely isolated from factors that affect it. As for the Spontaneous Raman Scattering assessment, the evaluation suggested that GPON would generate more noise to the quantum channel than XGS-PON.

The evaluation of the possibility of coexistence shows that both the original and length extended configurations succeed at keeping the IQBER below 11% for all the scenarios tested, indicating the possibility for coexistence between classical and quantum signals in this kind of network.

The results discussed here suggest that PONs can adequately support quantum communications. Nevertheless, some constraints must be acknowledged, as the point-to-point nature of the ODN established does not reflect the point-to-multipoint nature of the currently deployed PONs. Therefore, future work should prioritize field trials in operational PON environments to gather more definitive data on the integration of DV-QKD protocols in real-world PONs.

Further research will also require a full implementation of a DV-QKD protocol, such as BB84. Achieving a successful DV-QKD protocol implementation involves more than maintaining an IQBER below 11%, since the secure key rate (SKR) is the ultimate factor determining the protocol's viability. While an IQBER of 11% might technically allow key exchange, it could result in a very low SKR, making the protocol impractical. Therefore, IQBER should serve as an indicator of protocol security, but it must be minimized to ensure the SKR is high enough for feasible implementation.

Future works should study the possibility of coexistence not only with established technologies, but also with emerging ones, such as 50G-PON. 

## References

- [1] W. E. Burr, "Selecting the Advanced Encryption Standard," *IEEE Security & Privacy*, vol. 1, no. 2, pp. 43-52, March-April 2003, doi: 10.1109/MSECP.2003.1193210.
- [2] A. M. Fox, *Quantum optics: an introduction*, vol. 15, Oxford, England, Oxford University Press, 2006.
- [3] V. Mavroeidis, K. Vishi, M. D. Zych, and A. Jøsang, "The Impact of Quantum Computing on Present Cryptography," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, 2018, doi: 10.14569/ijacsa.2018.090354.
- [4] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum Internet: Networking Challenges in Distributed Quantum Computing," *IEEE Network*, vol. 34, no. 1, pp. 137-143, Jan./Feb. 2020, doi: 10.1109/MNET.001.1900092.
- [5] I. W. Primaatmaja, C. C. Liang, G. Zhang, J. Y. Haw, C. Wang, and C. C.-W. Lim, "Discrete-variable quantum key distribution with homodyne detection," *Quantum*, vol. 6, p. 613, Jan. 2022, doi: 10.22331/q-2022-01-03-613.
- [6] M. K. O. Lasota and V. C. Usenko, "Robustness of entanglement-based discrete- and continuous-variable quantum key distribution against channel noise," *New Journal of Physics*, vol. 25, no. 12, p. 123003, Dec. 2023, doi: 10.1088/1367-2630/ad0e8c.
- [7] P. W. Shor and J. Preskill, "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol," *Physical Review Letters*, vol. 85, no. 2, pp. 441-444, July 2000, doi: 10.1103/physrevlett.85.441.
- [8] S. T. Mantey, N. A. Silva, A. N. Pinto, N. J. Muga, Design and implementation of a polarization-encoding system for quantum key distribution, *IOP Journal of Optics*, Vol. 26, No. 6, pp. 075704 - 075704, June, 2024, doi: 10.1088/2040-8986/ad535a.
- [9] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10, Ed., Cambridge, England, Cambridge University Press, 2010.
- [10] B. Yan, Q. Li, H. Mao, and X. Xue, "High-Speed Privacy Amplification Scheme Using GMP in Quantum Key Distribution," *IEEE Photonics Journal*, vol. 12, no. 3, pp. 1-13, June 2020, doi: 10.1109/JPHOT.2020.2987611.
- [11] X. Xu, et al., "Neuromorphic Computing Based on Wavelength-Division Multiplexing," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 29, no. 2: Optical Computing, pp. 1-12, March-April 2023, doi: 10.1109/JSTQE.2022.3203159.
- [12] S. T. Mantey, "Electro-Optic Polarization Control Techniques for Quantum Technologies," M. S. thesis, University of Aveiro, Aveiro, Portugal, 2021.
- [13] D. Goldstein, *Polarized Light*, 3<sup>rd</sup> ed., Boca Raton, Florida, U.S., CRC Press, 2017.
- [14] G. P. Agrawal, *Fiber Optics Communication*, 3<sup>rd</sup> ed., Hoboken, New Jersey, U.S. John Wiley & Sons, 2002.
- [15] G. P. Agrawal, *Nonlinear Fiber Optics*, 3<sup>rd</sup> ed., Cambridge, Massachusetts, U.S., Academic Press, Inc., 2001.
- [16] P. Eraerds, N. Walenta, M. Legre, N. Gisin, and H. Zbinden, "Quantum key distribution and 1 Gbps data encryption over a single fibre," *New Journal of Physics*, vol. 12, June 2010, doi: 10.1088/1367-2630/12/6/063027.

# 09

## Automatic FTTH network design



With the increasing demand for high-quality internet services, deploying GPON/FTTH (Gigabit Passive Optical Network/Fiber to the Home) networks is a significant challenge for internet providers due to the significant investments involved. Automated network design usage becomes more critical to aid with planning the network by minimizing the cost of planning and deployment.

In today's inventory environment, a designer typically takes around 1.5 days to create a network design for a map with approximately 128 homes (a distribution cell). The challenges include positioning the installation equipment, such as cabinets, poles, or ducts; determining the number of apartments and houses; planning cable pathways; defining the feeder and distribution networks; establishing the influence area (which houses are served by each Optical Distribution Point (ODP)); routing the cables; and setting up the connections.

This article presents an AI-based solution that relies on a Genetic Algorithm (GA) using a two-level representation to design the networks automatically. We will detail the business rationale used to implement the supported features (focused on the distribution cell) and the associated flexibility, allowing its easy adaptation and configuration to the specific rules and OpCo-related variations. We will also present the approach's validation by comparing the quality of the AI-generated solutions with the handmade designs deployed in the real world. Our results demonstrate that this method can significantly reduce both the time and cost associated with network design, providing better solutions than those currently available and positioning it as a valuable tool for GPON/FTTH network design.

---

## Authors

**Xavier Silva, Altice Labs**, [xavier-s-silva@alticelabs.com](mailto:xavier-s-silva@alticelabs.com)

**Alexandre Gazur, Altice Labs**, [alexandre-l-gazur@alticelabs.com](mailto:alexandre-l-gazur@alticelabs.com)

**João Guerrinha, Altice Labs**, [joao-t-guerrinha@alticelabs.com](mailto:joao-t-guerrinha@alticelabs.com)

**Luís Cortesão, Altice Labs**, [luis-m-cortesao@alticelabs.com](mailto:luis-m-cortesao@alticelabs.com)

**Pedro Antero Carvalhido, Altice Labs**, [pedro-a-carvalhido@alticelabs.com](mailto:pedro-a-carvalhido@alticelabs.com)

**Tiago Vieira, Altice Labs**, [tiago-s-vieira@alticelabs.com](mailto:tiago-s-vieira@alticelabs.com)

---

## Keywords

Automatic Network Design; AI; GPON

## Introduction

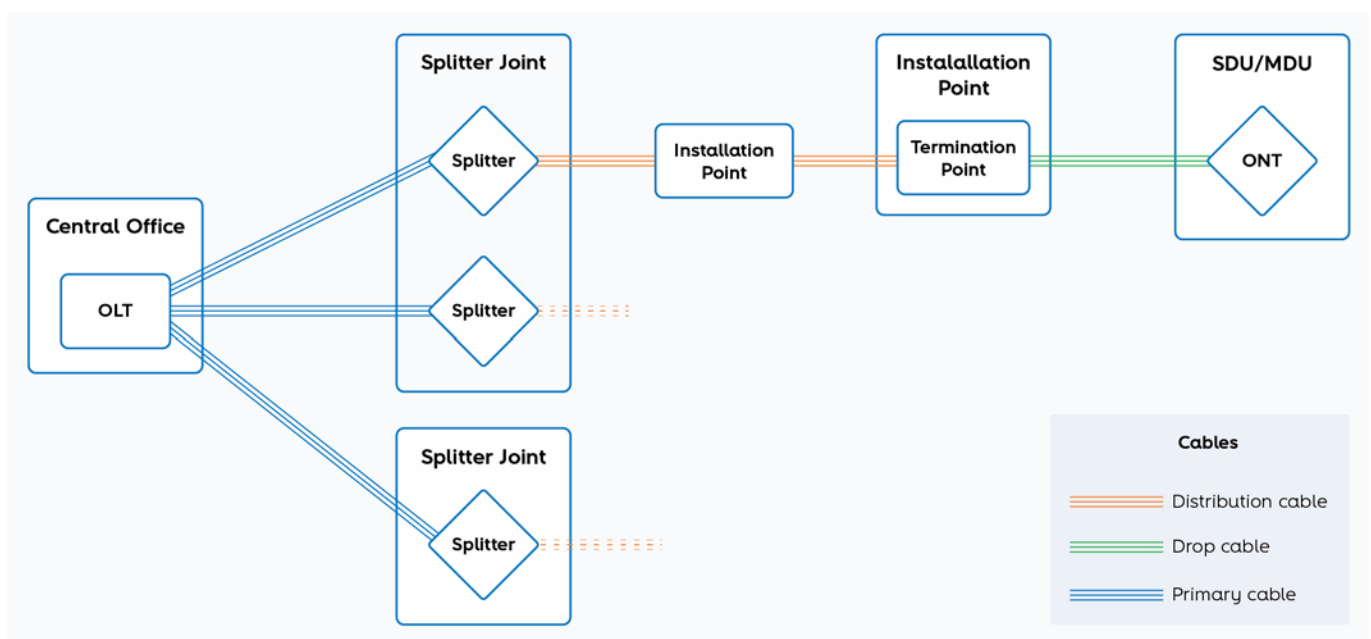
FTTH technology represents the cutting edge in broadband communication, offering unparalleled data transfer speeds and reliability by delivering optical fiber directly to the user's residence. Among the various FTTH architectures, the GPON is the most widely adopted due to its efficiency in delivering high-speed internet, voice, and video services over a single optical fiber. GPON networks utilize a point-to-multipoint topology, where a single optical fiber from the service provider's central office is split to serve multiple end-users, making it a cost-effective solution for large-scale deployments.

In a GPON network, the central office acts as the network's central distribution hub. It transmits data over optical fibers to multiple ONUs or Optical Network Terminals (ONTs) at the customers' premises. The latter are typically Multi-Dwelling Units (MDUs) in apartment buildings or Single-Dwelling Units (SDUs) in individual homes.

Between the starting and endpoints, the cables can be split (one fiber to many fibers) or fused (one fiber to one fiber). **Figure 1** illustrates a simplified representation of these types of networks.

The network's passive nature, which relies on optical splitters rather than powered devices, reduces operational costs and increases network reliability. Additionally, GPON infrastructure is highly adaptable; as technology advances, network operators can replace or upgrade equipment such as OLT and ONU without overhauling the entire cabling infrastructure. This flexibility ensures that the network can evolve and improve in quality without significant disruption or cost, making GPON a future-proof investment compared to older networking approaches.

Designing FTTH GPON networks is an intricate and time-consuming task that involves numerous parameters, such as the placement of splitters, routing of cables, termination points, cost, and respecting the optical-related criteria. Indicators such as bandwidth allocation, latency, and optical/power budget are critical in determining the network's effectiveness and require careful planning during the design phase. When performed by hand, it requires specialized knowledge in telecommunications and network engineering, and it is not uncommon for a person to spend over one business day manually designing a single network [1]. Given the ample solution space and the multiple conflicting objectives, traditional optimization



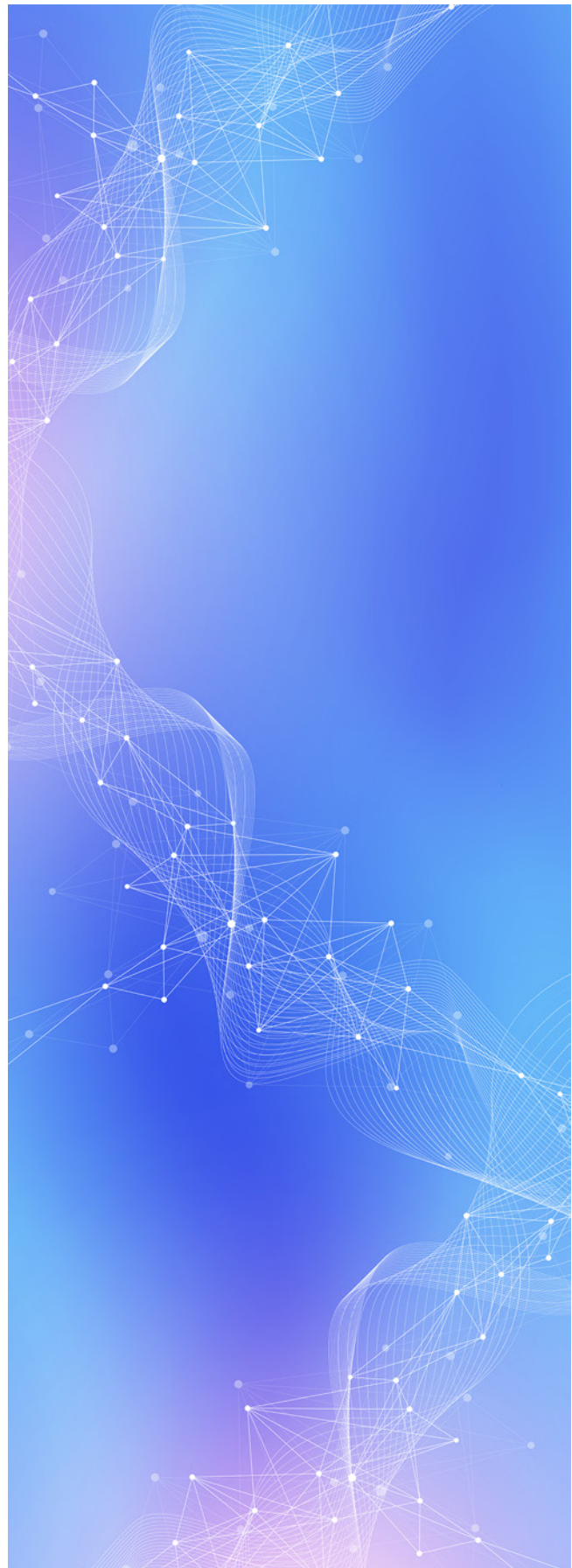
**FIGURE 1** – Simplified representation of a GPON FTTH network

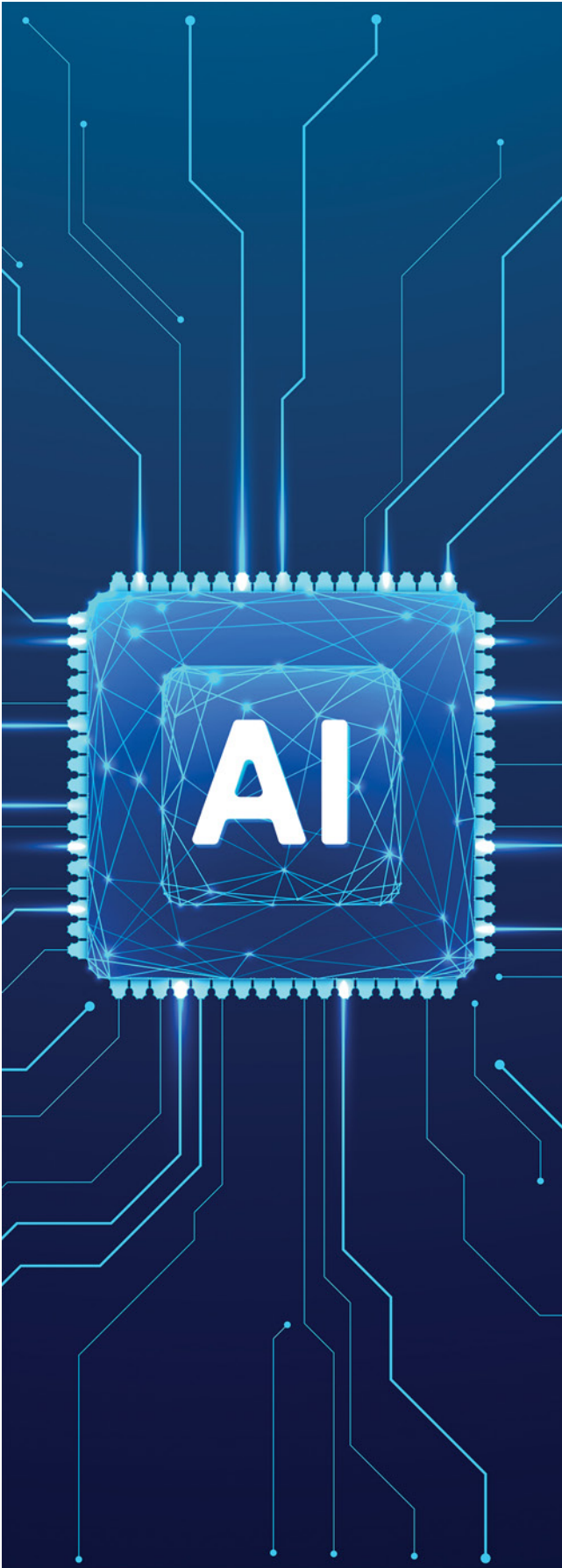
methods often fall short in addressing the complexity of GPON network design. AI-based approaches are a promising way to tackle this problem.

This article describes optimizing GPON network design, including crucial decisions regarding selecting and placing cables, equipment, and other network components. The resulting design is then handed over to teams responsible for physically constructing the network. While the ultimate vision is to extend this optimization process to cover the entire network from the central office to the end-user premises, this article concentrates on the distribution network as a continuation of a proof of concept. In GPON infrastructure, the feeder network refers to the segment from the central office to the first splitter, usually responsible for long-haul transmission. In contrast, the distribution network extends from the splitter to the end-user. By focusing on the distribution network, this research aims to refine the design process in a critical segment of the GPON infrastructure, laying the groundwork for future expansion into feeder network optimization.

A previous collaboration provided the first exploration and an initial proof of concept in this topic [2]. This early work provided valuable insights and laid the groundwork for the current study, both as a source of inspiration and a starting point. The lessons learned from this collaboration have contributed to advancing and refining the methodologies employed in this research, enabling a more sophisticated approach to optimizing GPON network design using GA.

This work thoroughly analyzes the existing literature on AI-based approaches for network design. It delves into the selected approach, exploring its potential to automate and optimize the complex task of GPON network design. The outcome is developing a comprehensive architectural software solution that integrates seamlessly with the existing Netwin architecture. This solution is a case study demonstrating AI's practical application and effectiveness in network design. By bridging the gap between theoretical research and real-world implementation, this article provides valuable insights and shows the potential for AI to revolutionize telecommunications infrastructure planning and design.





## Background

During the early development of this project, a review of existing commercial tools provided valuable insights and inspiration. Tools like Fast@Home [3] and X-brain [4] use AI to automate and optimize the planning of fiber networks. The first focuses on cost-efficient FTTB network design, while the latter supports high-level and low-level FTTH designs across large-scale deployments. These tools streamline decision-making by considering multiple scenarios and ensuring network designs meet specific project requirements while minimizing costs.

Other solutions, such as SpatialSUITE [5] and Comsof Fiber [6], combine heuristic approaches and AI to enhance fiber network design and planning. The first focuses on reducing costs and increasing productivity using GIS mapping and network inventory, while the latter provides precise control over network design, complete with detailed cost predictions. KSAVI GNI [7] complements these by automating the creation of network inventories and construction documents, further accelerating the network planning process.

Together, these tools highlight the industry's move toward sophisticated, automated solutions that significantly enhance the efficiency and accuracy of fiber network design.

## AI-based approaches to network design

In the context of modern network planning, the basic objective is to identify the most efficient and cost-effective network layout, which involves optimizing the placement of network components, such as cables, splitters, and termination points, all while considering various constraints, like the length of the cables or the maximum number of termination points. Traditional optimization techniques like Integer Linear Programming (ILP) and Mixed Integer

Linear Programming (MILP) have been widely used to address such problems. However, these methods can become impractical when dealing with the vast solution spaces and numerous variables inherent in complex network designs.

## Limitation of ILP and MILP

ILP and MILP are powerful mathematical optimization tools, particularly effective for problems with well-defined constraints and objective functions. They work by systematically exploring feasible solutions and identifying the optimal one.

However, as network design problems grow in complexity — with large numbers of potential equipment placements, routes, and connections — the computational burden of ILP and MILP increases exponentially. The vastness of the solution space and the sheer number of variables often result in excessive computation times and memory usage, rendering these methods less effective for large-scale network design [8].

## Alternative approaches: Simulated Annealing and Tabu Search

Given the limitations of ILP and MILP, alternative heuristic and metaheuristic approaches were studied to tackle the network design problem more efficiently.

Simulated Annealing (SA) is a probabilistic technique that starts with an initial solution and explores neighboring solutions by making random changes. Over time, the algorithm gradually reduces the probability of accepting worse solutions, focusing on finding a near-optimal solution. Despite its flexibility and ability to escape local minima, SA tends to be computationally intensive and slow, particularly for large-scale problems. Additionally, its tendency to discard reasonable solutions in pursuit of the global optimum can lead to suboptimal performance in scenarios where multiple reasonable solutions exist.

Tabu Search (TS) is another heuristic approach that systematically explores the solution space

by moving from one potential solution to another while avoiding previously visited solutions using a "tabu list." This algorithm is known for its efficiency and ease of implementation. However, like SA, TS focuses on finding the best solution, which may not be ideal for problems with multiple viable solutions. TS can also become time-consuming in large, complex problem spaces.

## Population-based approaches

Given the limitations of SA and TS, our focus shifted towards population-based approaches, which offer several distinct advantages for network design problems. Unlike single-solution methods, population-based algorithms explore multiple solutions simultaneously, making them well-suited for problems with large and complex solution spaces. These approaches maintain a diverse set of potential solutions, allowing for a more thorough exploration of the solution space and reducing the likelihood of getting stuck in local optima.

## Genetic Algorithm: a preferred solution

The GA, inspired by natural selection and evolution principles, emerged as an auspicious approach among population-based methods, in which a population of potential solutions evolves.

The key advantages of GA in network design are its flexibility and robustness. GAs are ideal for complex, multi-dimensional network planning problems with numerous constraints. By working with a population of solutions, GA allows for parallel exploration of the solution space, which speeds up the optimization process and increases the likelihood of finding high-quality solutions that single-solution methods like SA or TS may miss.

Additionally, GA's evolutionary nature means that the algorithm can continuously improve the solutions over successive generations. This iterative process is particularly useful in network design, where minor improvements in one part of the network can lead

to significant overall cost reductions. GA's diversity allows planners to explore various trade-offs between cost, performance, and other design criteria, selecting the most suitable solution.

## Using Genetic Algorithms in Network Design

After selecting the algorithmic approach, a software architecture was designed to enable the creation of an optimized network from a set of input data. The process begins by loading essential base data, including routes, homes, and installation points, as a graph. This graph forms the basis of the entire design process. The original graph undergoes a pre-optimization phase using the Steiner tree algorithm to enhance efficiency. This step reduces the graph to its most critical connections, minimizing unnecessary paths and simplifying the subsequent design stages.

The refined graph employs a genetic algorithm to generate the most suitable network configuration based on predefined cost, efficiency, and coverage objectives. The algorithm iteratively evolves potential designs, selecting and refining solutions that best meet the criteria. Once the optimal design is identified, the resulting graph is post-processed, during which additional information — such as details regarding the optical budget and network component quantities — is incorporated into the nodes.

### Network design as a Genetic Algorithmic problem

The solution incorporates key GPON network concepts crucial for creating a realistic and practical network design. These concepts include distribution and drop cables; different types of homes (namely SDU and MDU); termination points, which are critical components that connect the drop cable infrastructure at the client end to the main distribution network; installation points, such as

poles, where termination points can be installed; and splitter joints, which represents the starting node of the distribution network.

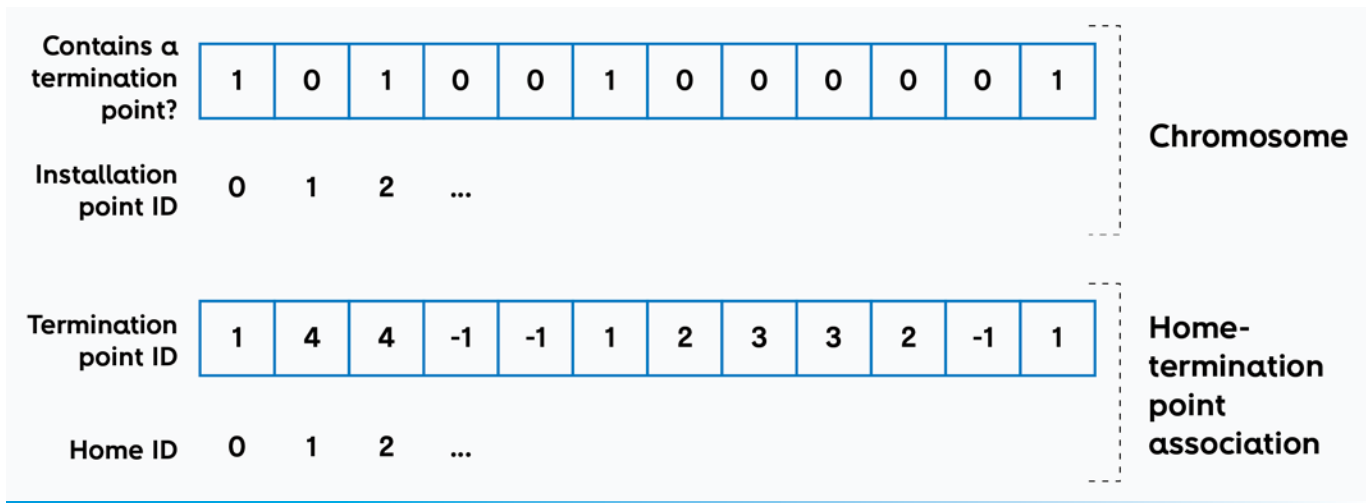
These elements are meticulously modeled within the GA framework to ensure that the resulting network design is both feasible and aligned with GPON standards.

To effectively represent the network design as a problem solvable by a GA, it is essential to conceptualize the problem regarding chromosomes and genetic operations. Each potential network configuration must be encoded as a chromosome, where individual genes represent specific design choices. This encoding enables the application of genetic operations, including mutation, selection, and crossover, which improves the solutions iteratively. By structuring the problem in this way, the GA can explore a vast space of potential network designs, iteratively refining them to meet the defined objectives better.

The network base, derived from raw data, represents all potential distribution cable routes within the service area. The algorithm identifies the optimal connections from homes to the main network, achieved through a chromosome that encodes the placement of termination points at installation points, as represented in **Figure 2**. Alongside this chromosome, an auxiliary map represents each home's allocation to its respective termination point. The chromosome undergoes standard genetic operations: mutation, selection, and crossover.

Once the chromosome representing termination point placements is generated and refined through genetic/variation operations, the auxiliary map is constructed to allocate homes to these termination points. This process involves several criteria to ensure the design's practicality and efficiency. Termination points must be appropriately sized for the connected homes, ensuring they do not exceed capacity, and drop cables must remain within the predefined length limits.

The variation operators (e.g., crossover and mutation) ensure that diversity is introduced into the populations by altering and combining genetic material. This ensures that a broader solution



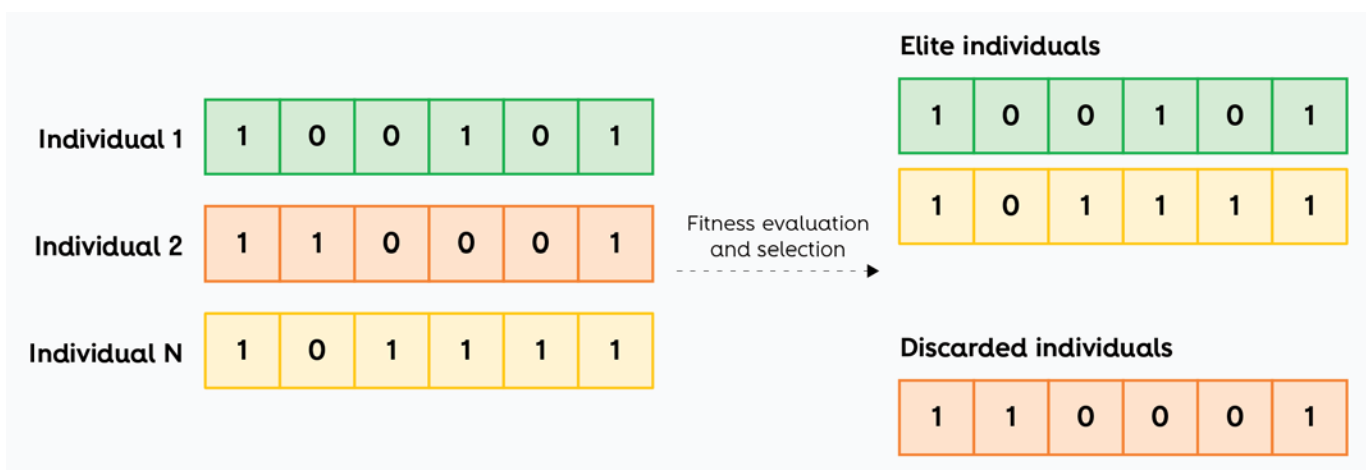
**FIGURE 2** – Representation of an individual in the implemented solution

space is explored and prevents premature convergence to suboptimal solutions. This integration of genetic and heuristic approaches results in a more robust and optimized network design.

With the termination points established, a search algorithm determines the appropriate types and quantities of cables required to fulfill the designed network. This includes calculating the required fiber count for each segment to ensure all home connections are accommodated. The algorithm measures the amount of drop and distribution cables required, considering the network's specific needs and the configurations provided. This step is essential to ensure the network's feasibility, cost-effectiveness, and efficient resource use.

The fitness function evaluates each solution based on the total cable length, client allocation to termination points, and the types and quantities of deployed termination points and splitters, assuring the evolution of the fittest individuals. The fitness value represents a measure of the network's efficiency and effectiveness, with the GA aiming to minimize this value over the generations and runs, as illustrated in **Figure 3**.

The solution with the lowest fitness value is considered the optimal design and is returned as the algorithm's final output. This approach guarantees that the network design meets all technical requirements while minimizing costs.



**FIGURE 3** – Fitness evaluation and selection over a set of individuals

## Netwin Automatic Network Design use case

In network design, particularly when applying a genetic algorithm, the fitness function is the cornerstone for evaluating and guiding the evolution of potential solutions. The fitness function assigns a numerical value to assess how suitable a network layout is based on predefined criteria. This value reflects the efficiency and cost-effectiveness of the network design, with lower fitness values indicating superior solutions. The goal is to minimize the fitness value, which implies a more cost-effective network design that still satisfies all requirements.

The fitness function in the Netwin Automatic Network Design system considers five critical factors, each contributing to the overall cost and effectiveness of the network:

- 1. Drop Cables:** these cables connect individual homes to the network. The fitness function considers both the length and cost per meter of these cables. The longer and more expensive the drop cables are, the higher the fitness value, resulting in a less optimal solution. Therefore, minimizing the length and cost of drop cables is crucial for reducing the overall network cost;
- 2. Distribution Cables:** these cables distribute the signal from the central junction, often called the Optical Splitter Joint (OSJ), to various points in the network. Similarly, the fitness function assesses the length and cost of distribution cables. Efficient use of distribution cables can significantly lower the fitness function by optimizing their routing and minimizing their length;
- 3. Termination Points:** termination points are critical components that distribute fibers from the network to individual homes. The fitness function considers both the quantity and cost of these termination points. Efficient placement and minimal usage of these costly components are key to lowering the fitness value;

**4. Splitters:** splitters divide the optical signal into multiple branches, allowing it to reach more homes. The fitness function assesses the number and cost of splitters. An optimal design strategically places splitters to minimize their number and cost while ensuring adequate signal distribution, thereby reducing the fitness value;

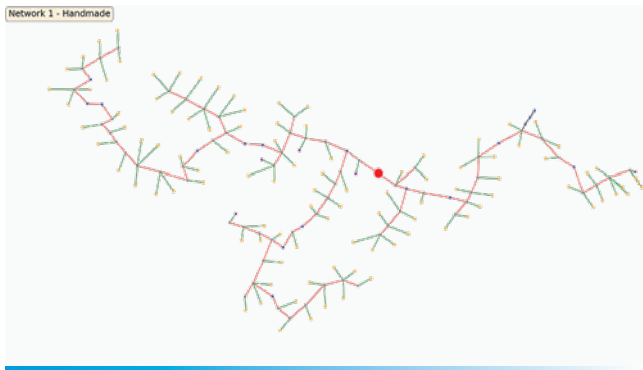
**5. Unallocated Homes Penalty:** the fitness function imposes a significant penalty for each home the algorithm fails to allocate to the network. This penalty ensures that the algorithm prioritizes connecting as many homes as possible, as failing to do so dramatically increases the fitness value. This ensures that the network design is both cost-effective and meets coverage requirements.

### Before and after: a visual comparison

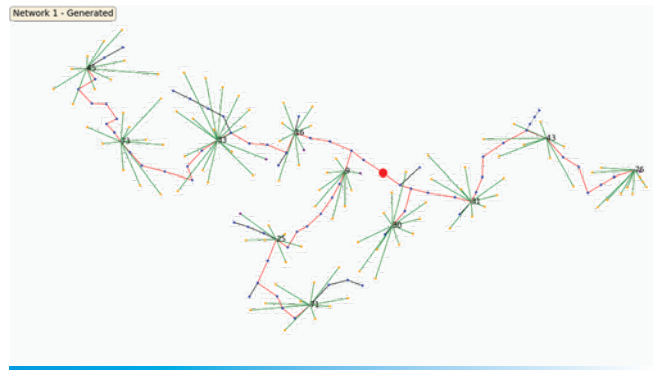
To illustrate the effectiveness of the Netwin Automatic Network Design system, we compare two network layouts: one created manually by a network designer and another generated automatically by the implemented solution.

As shown in **Figure 4**, Network A represents a manually designed network that, while functional, required many hours of labor to create and still resulted in a high fitness value of 60,091. This high cost is due to the suboptimal placement of cables, especially termination points and splitters, which significantly increased the fitness value.

In contrast, **Figure 5** showcases the optimized network layout generated by the genetic algorithm in less than one minute. The automated design process reduces the fitness value to an impressive 13,351, a 78% decrease in cost. This drastic reduction is achieved through the algorithm's ability to explore a vast solution space and iteratively improve the design by minimizing the length and cost of cables, optimizing the placement of termination points and splitters, and ensuring that all homes are connected.



**FIGURE 4** – Handmade design of Network A



**FIGURE 5** – Generated design of Network A

Below (**Table 1**) is the comparison between the two. The comparison between these two figures and their table values vividly demonstrates the power of genetic algorithms in network design. By automating the design process, the algorithm reduces costs and quickly produces a network layout that

meets all requirements far more efficiently than a manually created solution.

Below (**Table 2**) are more comparisons between the fitness values of real-life networks designed by engineers and those generated by the algorithm.

Parameter	Handmade	Generated
<b>Fitness value</b>	60091,86	13351,66
<b>Penetration rate</b>	1,00	1,00
<b>Distribution cable</b>	19956,00	8567,00
<b>Drop cable</b>	2748,93	5446,89
<b>Termination Points</b>	62	11
<b>SDU</b>	103	103
<b>MDU</b>	5	5
<b>Free Installation Points</b>	17	68
<b>Clients</b>	108	108

**TABLE 1** – Network A results comparison between handmade and generated designs

Network Name	Handmade Fitness Value	Generated Fitness Value	Cost Decrease
<b>Network B</b>	46819,54	11370,60	-76%
<b>Network C</b>	6485,65	3711,84	-43%
<b>Network D</b>	79019,13	76468,30	-3%
<b>Network E</b>	79369,46	34915,51	-56%

**TABLE 2** – Fitness value comparisons of various network designs

After stabilizing the solution, one could evolve it by fine-tuning the GA hyperparameters to achieve a feasible balance between runtime and fitness value. Hyperparameter tuning is crucial in optimizing the performance of AI-based solutions, as it directly impacts the algorithm's ability to explore the solution space effectively while maintaining computational efficiency [9]. The goal was to ensure that the GA could minimize the fitness value without excessively increasing the computational time required for each run.

## Conclusions and future work

GA has proven to be an effective method for addressing the complexities involved in FTTH GPON network design. As the results above clearly illustrate, GA is capable of efficiently managing the multidimensional optimization challenges inherent in such tasks.

The implemented tool significantly aids network engineers by drastically reducing the time required for planning and designing. Manual tasks that traditionally take hours or days can now be

completed in minutes or hours, depending on network complexity.

This approach is highly adaptable and can potentially be extended to other segments of the FTTH network. It can also incorporate additional network requirements, constraints, and objectives, offering flexibility to accommodate various deployment scenarios, regulatory constraints, and OpCo-specific norms.

This work establishes a solid foundation for ongoing research and development within the company, particularly in optimization. It also opens avenues for integrating other AI-based techniques and extending the solution to encompass the entire network, from the central office to the customer premises. Leveraging the scalability of the chosen algorithm, the proposed solution is ideal for large-scale network designs.

The main features of the existing stable version are focused on the distribution network design. However, it will swiftly evolve to explore the dynamic placement of Splitter Joints and central offices, extending the design to the feeder network. For that, it will rely on the analysis of the field topology to determine optimal routes from the main graph and identify the routes from a location instead of relying on a predefined scenario. 🔗



## References

- [1] C. G. Araújo, G. N. Gaspar, and J. T. Guerrinha, "Cognitive ops services," Altice Labs, Aveiro, Portugal, Jun. 2020. [Online]. Available: <https://www.alticelabs.com/shared-knowledge/whitepapers/whitepaper-cognitive-ops-services/>.
- [2] J. Correia et al., "Automatic Design of Telecom Networks with Genetic Algorithms," in *Applications of Evolutionary Computation: 26<sup>th</sup> European Conference, EvoApplications 2023*, Brno, Czech Republic, 2023, pp. 269-284, doi: 10.1007/978-3-031-30229-9\_18.
- [3] Esri, "Automated and Optimized Fiber Network Planning with GIS," esri.de, Accessed: Aug. 16, 2024. [Online]. Available: <https://www.esri.de/content/dam/distributor-restricted/esri-de/files/arcgis/produkte/fast-at-home/fast-at-home-en.pdf>.
- [4] X.-B. Fibre, "Optimize Your FTTH Planning With AI," x-brainfibre.io, Accessed: Aug. 16, 2024. [Online]. Available: <https://x-brainfibre.io/solution/>.
- [5] J. Arnold, "spatialSUITE: automated fiber network design & planning," synchronoss.com, Accessed: Aug. 16, 2024. [Online]. Available: <https://synchronoss.com/spatialsuite-unlocking-value-with-automated-fiber-network-design-and-planning/>.
- [6] IQGeo, "Automated fiber network planning and design software," iqgeo.com, Accessed: Aug. 16, 2024. [Online]. Available: <https://www.iqgeo.com/products/comsof-fiber>.
- [7] K. GNI, "Fiber optic network planning and design," ksavinetworkinventory.com, Accessed: Aug. 16, 2024. [Online]. Available: [https://ksavinetworkinventory.com/fiber-optic-network-planning-and-design/#geo-spatial\\_network\\_inventory](https://ksavinetworkinventory.com/fiber-optic-network-planning-and-design/#geo-spatial_network_inventory).
- [8] K. Metwally, A. Jarray, and A. Karmouch, "A cost-efficient QoS-aware model for cloud IaaS resource allocation in large datacenters," in *2015 IEEE 4<sup>th</sup> International Conference on Cloud Networking, (CloudNet)*, Niagara Falls, ON, Canada, Nov. 2015, pp. 38-43, doi: 10.1109/CloudNet.2015.7335277.
- [9] Monica and P. Agrawal, "A Survey on Hyperparameter Optimization of Machine Learning Models," in *2024 2<sup>nd</sup> International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2024, pp. 11-15, doi: 10.1109/ICDT61202.2024.10489732.

# 10

## AI-driven development: what may the future bring?



AI-driven development represents a transformative shift in software engineering, where Artificial Intelligence (AI) aims to automate and enhance various aspects of the software creation process. This cutting-edge technique leverages AI to aid in coding and handling complex tasks such as debugging, testing, deployment, and even optimization, thus simplifying the development process. They analyze vast amounts of code to identify patterns and suggest optimizations, minimizing the likelihood of errors and improving code quality, reducing development time/investment, and translating into operational improvements, quality, or support for problem-solving. In this article, we examine the current state of the art, present some data from relevant studies, and explore the current use of AI-assisted development at Altice Labs.

---

## Authors

**Filipe Delgado, Altice Labs, [filipe-j-delgado@alticelabs.com](mailto:filipe-j-delgado@alticelabs.com)**

**Ricardo Cadime, Altice Labs, [ricardo-v-cadime@alticelabs.com](mailto:ricardo-v-cadime@alticelabs.com)**

**Jorge Sousa, Altice Labs, [jorge-m-sousa@alticelabs.com](mailto:jorge-m-sousa@alticelabs.com)**

**Pedro Neves, Altice Labs, [pedro-m-neves@alticelabs.com](mailto:pedro-m-neves@alticelabs.com)**

---

## Keywords

Software creation process; Development lifecycle; AI-driven tools; Efficient coding practices; Improving code quality; Collaborative environments; Accelerated innovation; Reshaping the industry



## Introduction

In the dynamic field of software engineering, innovation and the search for new methods and execution strategies are critical factors. Today, we are witnessing a transformative shift driven by AI-powered development. This new approach leverages the capabilities of Artificial Intelligence, particularly Generative AI (Gen AI) and Large Language Models (LLMs), to revolutionize how code is understood, written, and improved. AI integrations within the development workflows are streamlining tasks such as coding, debugging, testing, optimization, code documentation, and deployment – resulting in greater efficiency and precision. In response to these advancements, companies and organizations are increasingly focusing on adopting AI-driven technologies

to enhance multiple aspects of their processes, particularly software development. By incorporating AI into their development pipelines, they aim to improve and optimize numerous stages of the software development lifecycle, from initial design to final deployment.

For software engineers, this new wave of AI tools enhances process efficiency and expands their capabilities – whether through automated code completions, bug detection, or testing automation. These innovations offer novel approaches that benefit all stakeholders by speeding up workflows, reducing overall costs, and delivering high-quality products.

The advantages of AI innovations are truly transformative, but it is important to acknowledge some limitations that may arise from complex codebases and legacy systems. Older systems often contain

intricate dependencies, outdated coding styles, and patterns, as well as architectural decisions that span different technological eras. To overcome or minimize the challenges posed by outdated infrastructures, organizations must first modernize by investing in cutting-edge infrastructures and robust data platforms, implementing strong data pipelines, and ensuring scalable computing resources. Updating methodologies and making them more agile will enable a more interactive development environment, allowing for frequent feedback and adjustments that enhance the software development cycle.

There have also been concerns that AI-assisted tools can generate problematic code. These models learn from vast datasets of existing code, which may contain biases, errors, and security concerns, potentially perpetuating these issues. This highlights the continued need for human oversight, which remains crucial in correcting inaccurate AI suggestions. Additionally, the complexity of “black-box” AI models – where it is not always clear how code suggestions are generated – makes debugging and informed decision-making more difficult. Addressing biases and security risks inherent in AI-generated code becomes essential. Rigorous testing, thorough code reviews, and secure training pipelines are necessary to avoid introducing vulnerabilities.

At the legislative level, the European Union (EU) is already working on regulations governing the use of AI, making it imperative for organizations to adopt mitigation strategies to address these concerns. Establishing a robust and responsible AI framework is no longer optional – it is a critical step toward ensuring the ethical use of AI in the future.

Despite these limitations, AI-assisted software development holds significant potential for businesses. When used properly, these tools can streamline development processes, reduce time-to-market, and enhance the overall quality of code, effectively augmenting the capabilities of human programmers [1]. In summary, the article begins by outlining the present situation in the AI-driven development field. It then emphasizes the advantages of AI-driven development and its implementation at Altice Labs. The paper ends with concluding remarks.

## State of the art

AI-driven development is transforming the software engineering landscape by automating and enhancing processes throughout the entire development lifecycle. By leveraging Artificial Intelligence (AI) and Machine Learning (ML), along with the understanding and generation capabilities brought by LLM models, developers are now able to build, test, and maintain software more efficiently, reducing the time and effort needed while increasing reliability and quality.

One of the most visible impacts of AI in software development is in AI-augmented code generation. Advanced models, such as OpenAI's Codex and Google's AlphaCode, are now capable of understanding natural language descriptions and generating relevant code snippets. These models are trained on large datasets of open-source code, enabling them to generate boilerplate code, offer contextual code completion suggestions, and even refactor existing code for better performance or readability. AI is helping to reduce the manual effort required for routine coding tasks, allowing developers to focus on more complex and creative aspects of software development. Tools like GitHub Copilot and Tabnine are leading examples of how AI is being integrated into code editors, offering developers real-time assistance as they write.

In addition to code generation, AI is also making significant improvements in automated testing, debugging, and documentation. One of the key applications here is the automatic generation of test cases based on the code being developed. This eliminates the need for developers to manually create tests, ensuring that applications are thoroughly tested without the time-consuming effort traditionally required. ML models are now able to detect patterns in code that are likely to introduce bugs or security vulnerabilities, sometimes even suggesting fixes automatically. This not only reduces the occurrence of bugs but also improves code quality early in the development process. Tools such as Diffblue Cover and DeepCode are examples of AI-driven solutions

that automatically generate tests or detect bugs with high accuracy, streamlining both the testing and debugging phases.

AI-powered DevOps, often referred to as AIOps, is another area where AI is driving change. In the realm of DevOps, AI is enhancing the efficiency of software delivery and operations management. For instance, continuous integration/continuous delivery (CI/CD) pipelines can now be optimized by AI, which predicts build failures, automates code merges, and manages dependencies. This minimizes human intervention and allows for faster software deployment. AI is also being used to monitor application performance and infrastructure health. By analyzing logs, performance metrics, and user behavior, AI can detect anomalies, predict outages, and even suggest corrective actions or auto-scale resources to ensure optimal application performance. Tools like Dynatrace and Splunk are pioneering the use of AI in this space, helping to create more reliable and responsive systems.

AI is also being applied at the earlier stages of the software development process, particularly in requirements engineering and software design. One major advancement is the usage of Natural Language Processing (NLP) techniques to analyze unstructured requirements, which are often written in natural language and extract key functional and non-functional requirements. This helps reduce ambiguity and ensures clarity in what the software should do. Additionally, AI is beginning to assist in software design by suggesting optimal architectures or design patterns based on a given set of requirements. This enables developers to make informed design choices with less effort and improves the overall structure of the system. Tools like ReqSuite RM are examples of how AI is being integrated into requirement management, helping teams define and analyze requirements more efficiently.

In terms of code optimization and refactoring, AI has proven to be a powerful tool for improving code performance and maintainability. AI-driven systems can analyze large codebases and suggest or automatically apply changes that enhance

performance, such as identifying and eliminating bottlenecks, optimizing algorithms, or improving memory management. These systems are also being used to refactor code, making it cleaner and easier to maintain without changing its core functionality. Tools like JetBrains IntelliJCode and SourceAI are in charge of helping developers optimize and refactor code efficiently.

AI is also transforming how teams collaborate and manage their projects. In AI-enhanced collaboration and project management, AI tools are used to forecast project timelines, predict potential bottlenecks, and automate task assignments. This helps teams stay on track and make more informed decisions about resource allocation and workflow. Furthermore, AI-powered tools are improving team communication by automating repetitive tasks, such as updating project statuses or sending reminders. Tools like Jira and Monday.com are leveraging AI to provide insights into project performance, making it easier for teams to manage large and complex projects.

Lastly, AI is playing a crucial role in software security through the integration of AI into DevSecOps practices. AI tools can analyze code to automatically detect security vulnerabilities and suggest fixes, helping organizations address potential threats before they become serious issues. In addition to vulnerability detection, AI is being used for threat hunting by analyzing logs, user behavior, and network traffic to identify unusual activity that may indicate a security breach. Tools such as Snyk and Fortify on Demand are applying AI to enhance software security, making it easier for developers to ensure their applications meet the highest security standards.

In short, AI can support various software engineering tasks:

- **Code Generation:** AI models like Codex and AlphaCode assist in generating boilerplate code, contextual code completions, and refactoring, reducing manual effort;
- **Automated Testing and Debugging:** AI-driven tools automatically create test cases and detect

security vulnerabilities, enhancing code quality early in development;

- **AI-Powered DevOps (AIOps):** AI optimizes CI/CD pipelines, predicts build failures, and monitors application performance, ensuring efficient software delivery and operations management;
- **Requirements Engineering and Software Design:** NLP techniques analyze unstructured requirements, while AI suggests optimal architectures and design patterns;
- **Code Optimization and Refactoring:** AI systems improve code performance and maintainability by identifying bottlenecks and refactoring code;
- **Collaboration and Project Management:** AI forecasts project timelines, predicts bottlenecks, and automates task assignments, enhancing team communication and workflow;
- **Software Security (DevSecOps):** AI detects security vulnerabilities, suggests fixes, and performs threat hunting to ensure high-security standards.

## Benefits of AI-driven development

Recent studies provide quantitative insights into the usability of generative AI systems for real business tasks. Three distinct studies [2], [3], [4] evaluated a variety of users across different domains – customer service agents in an enterprise software company [2], professionals writing business documents [3], and programmers coding a small project [4]. All the studies reached similar conclusions: productivity saw a significant boost (Figure 1), especially among less-skilled users. Additionally, some studies reported enhancements in the quality of the outputs.

According to [4] and [5], less experienced programmers gained slightly more from the AI tool.

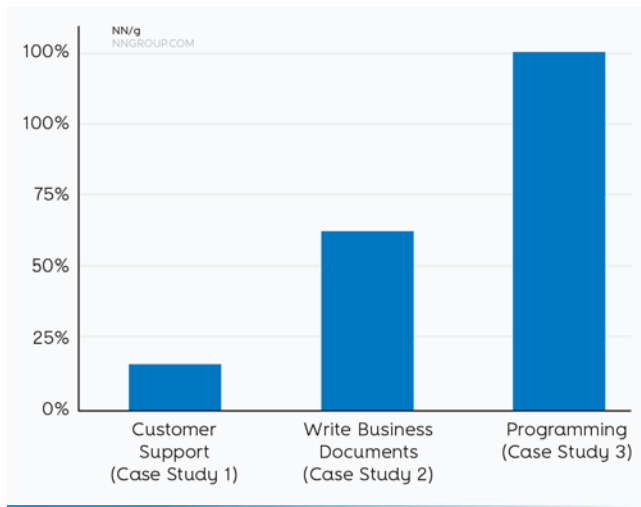


FIGURE 1 – Productivity Increases with AI (according to [2],[3],[4])

Additionally, those who spent fewer hours coding each day experienced greater benefits than those who coded for longer periods, with this second effect being significant. Together, these findings indicate that less-skilled programmers derive more advantages from AI compared to their more-skilled counterparts. This means that AI contributes to narrowing the gap between the most talented and the least talented employees.

An additional study from McKinsey [6] concluded that productivity improvement is much higher for less complex tasks than for high-complexity ones, as stated in Figure 2.

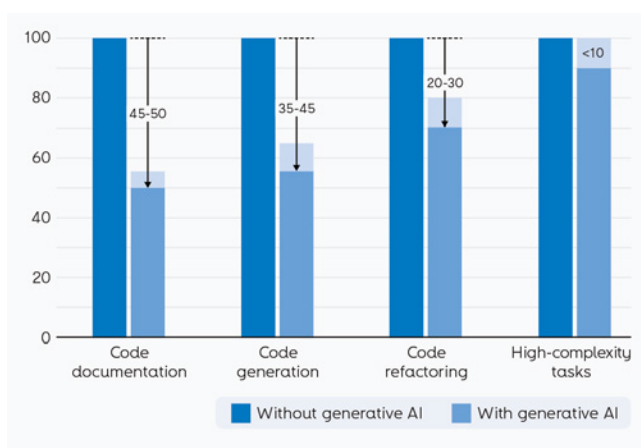


FIGURE 2 – Task completion time using generative AI (Reference: McKinsey study [5])

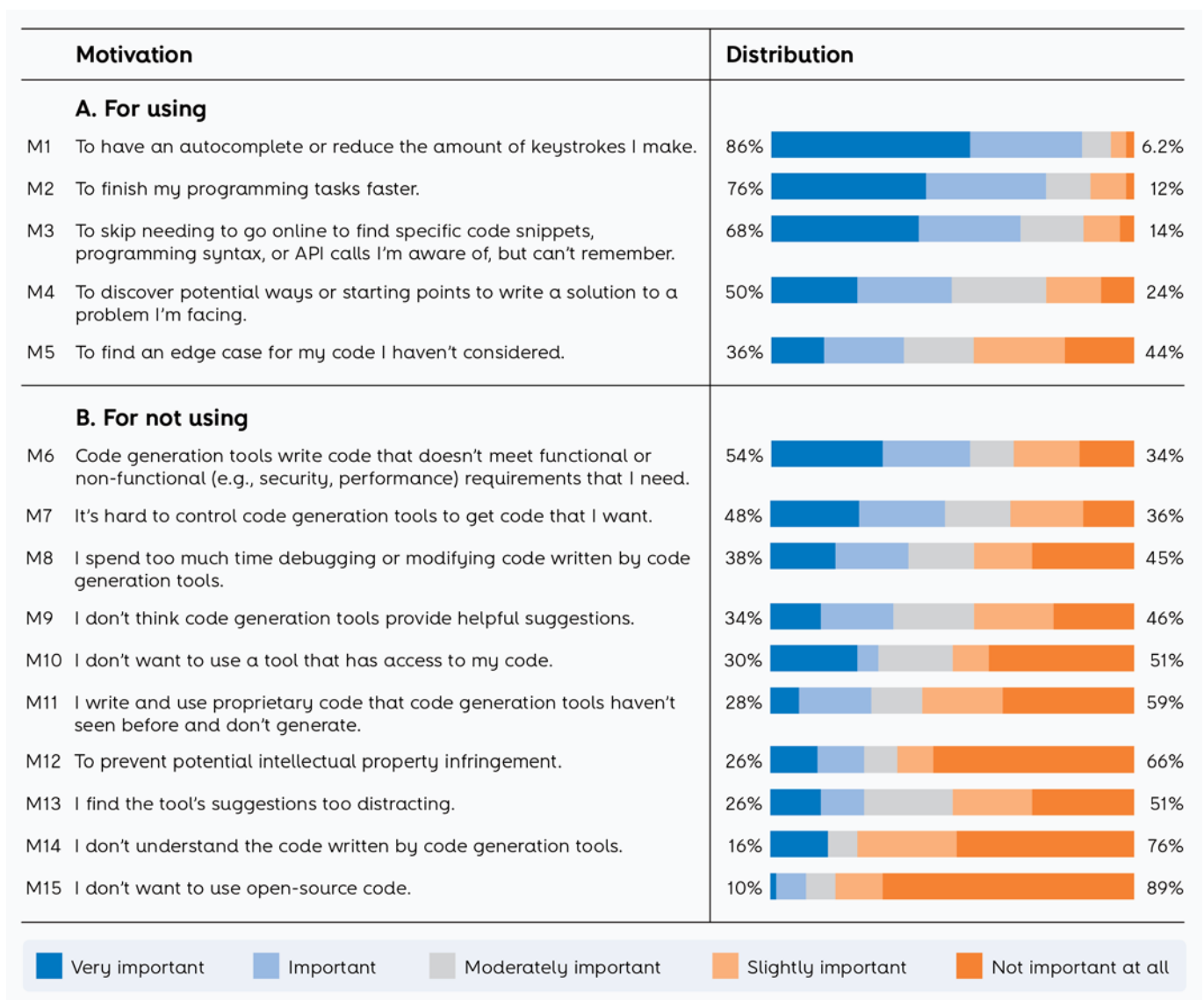
The same study shows that the developers' work satisfaction and motivation increased with AI tool usage, as they are now able to focus on meaningful and challenging work.

Carnegie Mellon also did an experiment with about 750 developers [7] that confirmed the above conclusions and detailed the relative importance of each aspect for the participants (**Figure 3**).

In summary, each study highlights a notable increase in productivity – ranging from 17% to 126% – along with improvements in quality, enhanced job satisfaction, and a more balanced distribution of skills across development teams.

## AI-driven development adoption at Altice Labs

GitHub Copilot is an AI-powered code completion tool developed by GitHub in collaboration with OpenAI. It aids developers by suggesting code snippets and entire functions based on the context of the code being written.



**FIGURE 3** – Motivation aspects, according to Carnegie Mellon study [7]

Its major features are:

- **Code Generation:** Copilot can generate boilerplate code, suggest improvements, fix errors, generate unit tests, and even help with documentation;
- **Language Support:** it supports multiple programming languages and integrates with popular code editors like Visual Studio Code;
- **Contextual Understanding:** Copilot's ability to understand the context of the code being written helps in providing relevant suggestions.

Here are some practical examples of using Copilot as a developer assistant.

## Code generation

GitHub Copilot offers a range of powerful code generation features that significantly enhance developer productivity. It provides real-time inline code suggestions, helping developers write code faster and with fewer errors. Copilot can generate entire functions or code blocks based on natural language prompts, making it easy to implement standard algorithms or boilerplate code. It also suggests refactoring and improvements to existing code, ensuring it adheres to best practices.

When errors occur, Copilot analyses the error messages and surrounding code to propose fixes, saving time and reducing frustration for developers.

For new projects, Copilot can scaffold a complete workspace, setting up the necessary files and configurations. It also helps generate database queries, simplifying interactions with databases.

Using a recent project as an example, at least a third of the code was generated by Copilot, including boilerplate code, HTTP client requests to integrate with other systems, location functions conversions from Python to Java, complex reactive invocations and composition, SQL query generation, and even implementation of APIs using an OpenAPI spec as input, in which the Copilot plugin automatically generates all the server endpoints and the required POJOs.



## Unit tests

GitHub Copilot is a valuable tool for generating unit tests. It automatically creates test code snippets, saving developers time and effort. By understanding the context of the code, it suggests relevant test cases that ensure comprehensive coverage. This helps developers to focus on refining and customizing tests rather than writing them from scratch. Additionally, it serves as an educational resource by providing examples of well-structured unit tests, which can help less experienced developers learn best practices and improve code quality.

In a recent project, about 30% of the unit tests were designed by Copilot. Although the test cases are very well defined, the developers found a large number of errors in their implementation; this seems to be related to the complexity of the toolkit in use

(Quarkus.io) and the complexity of the mocks that are required to cover a large code basis. See the example shown in **Figure 4**.

## Code improvement

As previously mentioned in the earlier chapters, Copilot can suggest or be asked to improve code. These improvements can be applied not only to the code style and structure but also to the performance of the written code, including SQL query performance.

In a recent project, this feature was useful for improving the readability of the code, its organization, and its performance since Copilot identifies smaller mistakes that can lead to huge performance issues, such as possible memory leaks and inefficient Java garbage collection.

The screenshot shows an IDE with a Java file named `TransformationModuleTest.java` and a `GitHub Copilot Chat` window. The chat window shows a user prompt: "Adding a test for maxAge=0 in TransformationModuleTest" and a response from GitHub Copilot. The response includes instructions on how to add a test method and provides the code for the new test method.

```

class TransformationModuleTest {
    // ...

    @Test
    void testGetLocationRetrieveResponseAnyAge() {
        //input data
        String lat = "40.1";
        String lon = "-7.1";
        int cellInRadius = 1800;
        int cellOutRadius = 1800;
        int cellStartAngle = 270;
        int cellStopAngle = 30;
        Instant time = Instant.now();

        //expected output data
        double expectedLat = 40.10555652210616;
        double expectedLon = -7.10417789811098;
        double expectedRadius = 966.4349781959106;

        SweLocationProperties locationProperties = new SweLocationProperties();
        locationProperties.setLat(lat);
        locationProperties.setLon(lon);
        locationProperties.setCellInRadius(cellInRadius);
        locationProperties.setCellOutRadius(cellOutRadius);
        locationProperties.setCellStartAngle(cellStartAngle);
        locationProperties.setCellStopAngle(cellStopAngle);
        locationProperties.setTime(time);

        // Call the method
        LocationRetrieveResponseDTO response =
            TransformationModule.buildLocationRetrieveResponse(locationProperties);

        // Verify the response
        assertEquals(OffsetDateTime.ofInstant(locationProperties.getTime(), ZoneId.systemDefault()), response.getTime());
        Area area = response.getArea();
        assertEquals("expected: 'CIRCLE'", area.getAreaType());
        Center center = area.getCenter();
        assertEquals(expectedLat, center.getLatitude());
    }
}

```

The chat window shows the following response from GitHub Copilot:

Adding a test for maxAge=0 in TransformationModuleTest

add a test for maxAge=0

3 steps completed successfully

To add a test for maxAge=0 in the TransformationModuleTest class, you can create a new test method that verifies the behavior of the retrieveLocation method when maxAge is set to 0. Here's how you can do it:

1. Set up the necessary mock behavior for the NOSSISLocationService.
2. Call the retrieveLocation method with maxAge=0.
3. Verify the expected behavior.

Here's the code for the new test method:

```

@Test
void testRetrieveLocationMaxAgeZero() {
    // Input data
    String msisdn = "+351999999999";
    int maxAge = 0;

    // Mock location properties
    SweLocationProperties locationProperties = new SweLocationProperties();
    locationProperties.setLat("40.630982");
    locationProperties.setLon("-8.648900");
    locationProperties.setCellInRadius(20000);
    locationProperties.setCellOutRadius(20000);
    locationProperties.setCellStartAngle(0);
    locationProperties.setCellStopAngle(70);
    locationProperties.setTime(Instant.now());
}

```

What other test cases can I add to the TransformationModuleTest class?

Ask Copilot a question or type '/' for commands

FIGURE 4 – Unit test generation with GitHub Copilot

## Error fixing

Currently, all IDEs provide syntax highlighting and error detection, but Copilot can assist developers when they want to validate if the code implements the desired algorithm. This was extremely helpful in a recent project when validating if the code was according to the specifications and if location algorithms were correctly implemented.

## Documentation generation

GitHub Copilot can generate explanations for code blocks, entire files, and even whole code workspaces, which is particularly useful for creating comprehensive project documentation. Developers can easily generate inline code comments that explain the purpose and functionality of specific code segments by using the GitHub extensions/plugins available for several IDEs (Visual Studio Code, IntelliJ Idea, etc.). In a recent project, for example, 95% of the Javadoc was written by the Copilot plugin, with the developers only having to perform some minor corrections to the generated documentation.

## Code review

Code review is a critical stage in the software development process, serving as a quality control measure to identify bugs, ensure coding standards, and improve the overall health of the codebase. It is a complex task that requires a reviewer to scrutinize code changes meticulously, often without the complete context of the entire development process. This lack of context can make it challenging to spot subtle issues or understand the implications of certain code changes. Moreover, developers, regardless of their experience, may not always have the comprehensive knowledge needed to identify all potential problems or areas for improvement, such as performance optimization, security vulnerabilities, or code maintainability. Effective code review thus demands a combination of broad technical knowledge, attention to detail, and a deep understanding of the project's architecture and objectives.



Regarding the available tools, we can use CodeRabbit as an example. CodeRabbit is an AI-driven tool designed to enhance the software development process, and it focuses on automating testing, review, and debugging, ensuring that code is robust and error-free.

Its major features are:

- **Automated Testing:** CodeRabbit can automatically generate test cases, identify potential bugs, and suggest fixes;
- **Performance Insights:** it provides insights into code performance and helps optimize resource usage;
- **Debugging Assistance:** CodeRabbit assists in debugging by pinpointing issues and suggesting solutions.
- Although we did not use this tool on a specific project, we performed some experiments to evaluate its capacities. As a result of the experiments, we can conclude that the tool has a great potential to detect issues, such as:
  - Critical issues (like possible null pointer exceptions, infinite loops, etc.);
  - Incorrect implementation of an algorithm, where the expected result is different from the one that was implemented (example: we want a method that calculates a sum, but it returns the subtraction);
  - Not using the best code practices (example: unorganized imports, unused variables, etc.).

- It suggests fixes and/or changes with accuracy, speeding up the application of the required changes;
- It is simple to configure, requiring only the registration of a GitHub repository for it to begin performing code reviews (no code is needed), and simple to use, providing intuitive commands to interact with it.

## Documentation generation and issue management

### JIRA

JIRA is the project management and issue-tracking tool adopted by our organization and has been in use for several years. In the most recent versions, it packages AI features that can boost a team’s productivity by reducing the number of repetitive tasks and generating content based on inputs made in natural language. Among the available features, we can highlight the following:

#### Search for issues using everyday language

This feature allows users to find issues using natural language instead of writing complex JIRA Query Language (JQL) queries (**Table 1**).

Input	Result
What are the most recent Bugs in project PRJ?	Jira will transform the input into a JQL query that will return the issues with type Bug belonging to project PRJ, sorting by the most recent.
What are the Internal Requirements for the API component in project PRJ?	Jira will transform the input into a JQL query that will return the issues with type = Internal Requirement, component = API, belonging to project PRJ.

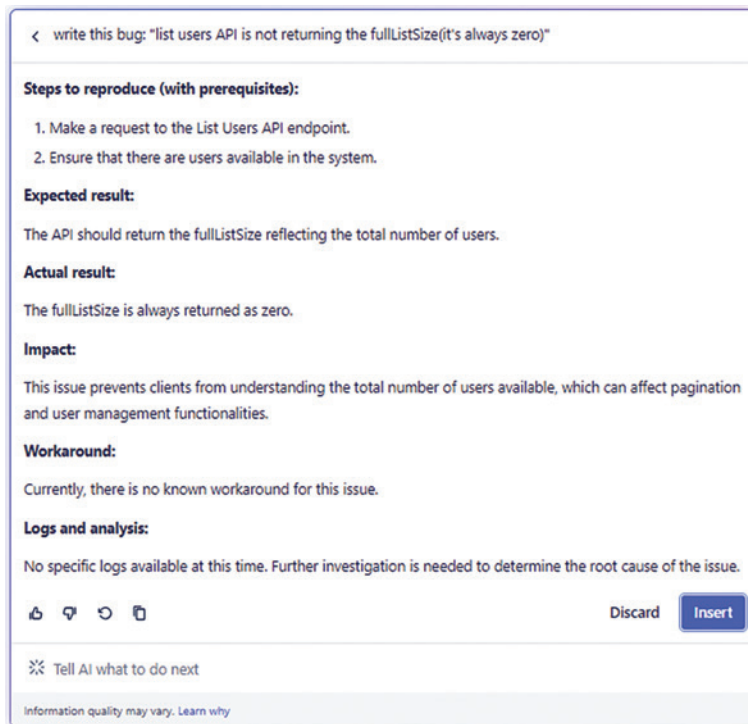
TABLE 1 – Examples of Atlassian Intelligence-generated JQL queries

## Generate and transform content

Writing an issue can be repetitive, especially when we need to follow a template and fill in all the required info repeatedly. Using this AI capability, we can ask JIRA, for example, to write a bug following the defined template, as illustrated in **Figure 5**.

## AI work breakdown

When designing a new feature, the teams must break it into smaller pieces, creating subtasks. We can use AI to do this, and JIRA will read the details of the issue and will create subtasks. **Figure 6** illustrates an example of subtasks for a FAQs page requirement.



**FIGURE 5** – Bug writing using Atlassian Intelligence



**FIGURE 6** – Atlassian Intelligence suggested subtasks



## Summarize issue comments

Another task that developers often do is reading comments on an issue to try to build a summary of the interactions it had during time. This can be achieved using AI: it will read the documents and summarize them (**Figure 7**).

## AI automation

The more recent versions of JIRA allow us to automate some tasks like performing automatic issue transitions based on rules, creating pages associated with issues, cloning issues, etc. Instead of manually building these automation flows, we can ask AI to do it. **Figure 8** shows an example of creating an automation that marks an issue as Resolved after a pull request is merged.

## CONFLUENCE

Confluence is a collaboration and documentation tool adopted by our organization and, like JIRA, has been in use for several years. Since it is developed by the same company (Atlassian), it also packages AI features, which we can highlight:

### Summarize pages and blogs

With this feature, we can ask to summarize content, so we do not need to read every information that is related to a page and/or comments.

### Q&A Search

Currently, when we want to obtain information about a specific topic, we need to search Confluence using keywords (ex.: we need information about a specific requirement, only knowing its name); then, we need to read all the pages that have content related to the keyword. Using AI, we can simply ask questions, and it collects info about the topic, summarizing it for us. See the example shown in **Figure 9**.

### AI automation

Like in JIRA, we can use AI to create automation in Confluence, as the example illustrated on **Figure 10**.

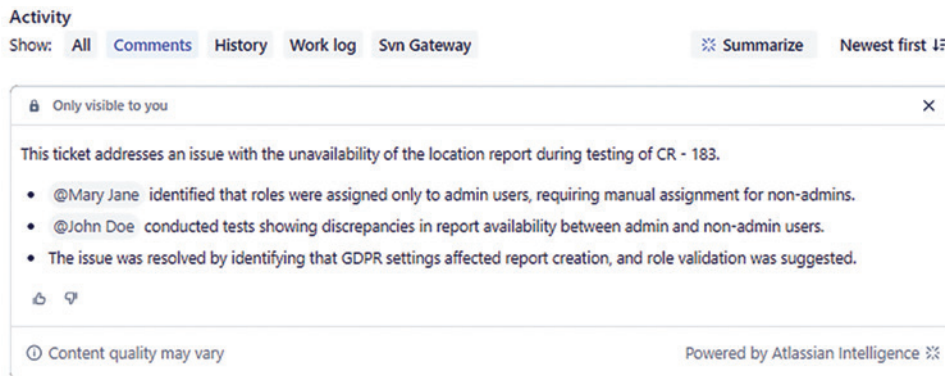


FIGURE 7 – Atlassian Intelligence issue summarization

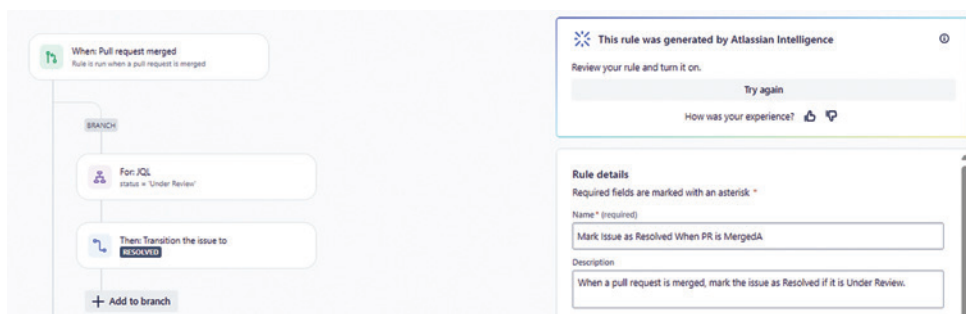


FIGURE 8 – Atlassian Intelligence generated automation

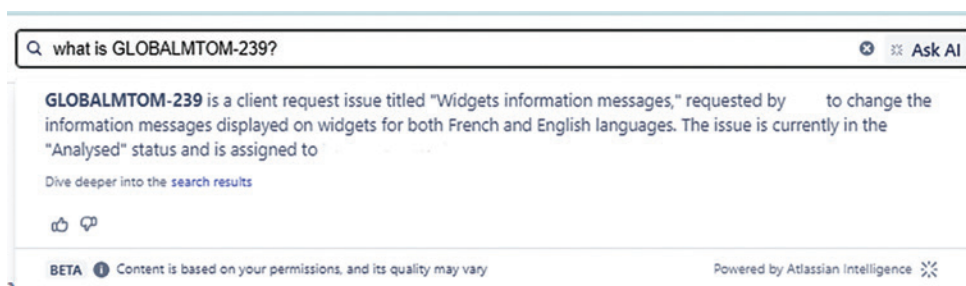


FIGURE 9 – Atlassian Intelligence “Ask AI” search

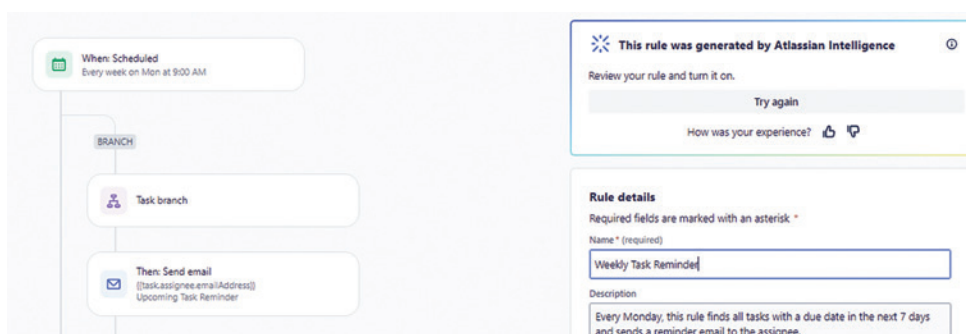


FIGURE 10 – Atlassian Intelligence automation creation in Confluence

## Conclusions

AI-driven development is revolutionizing the software engineering landscape. By automating routine tasks and providing intelligent insights, AI tools are enabling developers to focus on more complex and creative aspects of the development process. As these tools continue to evolve, they will undoubtedly play an increasingly vital role in shaping the future of software development.

Despite these advancements, there are challenges and limitations to AI-driven development. One major concern is data privacy and bias. Since AI models rely heavily on data for training, ensuring that the data used is unbiased and does not expose sensitive information is critical. Furthermore, integrating AI into the software development process requires specialized expertise, as developers need a solid understanding of both AI/ML technologies

and software engineering principles. This can make it difficult for organizations to fully adopt AI-driven tools. Additionally, AI systems need to be regularly updated to keep pace with new development practices and languages, requiring ongoing maintenance.

Looking ahead, the future of AI-driven development holds immense potential. Autonomous development pipelines that handle everything from coding to deployment with minimal human intervention are on the horizon. AI-driven collaborative agents could work alongside developers in real time, assisting with design, coding, and debugging tasks. Moreover, adaptive software that continuously improves itself post-deployment based on user behavior and performance metrics could soon become a reality. While challenges remain, the benefits of AI in software development are already reshaping how we build and maintain software, setting the stage for even more advanced and autonomous systems in the near future. [🔗](#)



## References

- [1] I. Ozkaya, "The Next Frontier in Software Development: AI-Augmented Software Development Processes," *IEEE Software*, vol. 40, no. 4, pp. 4-9, July-Aug. 2023, doi: 10.1109/MS.2023.3278056.
- [2] E. Brynjolfsson, D. Li, and L. Raymond, "Generative AI at work," *Working paper 31161*, Apr. 2023. doi: 10.3386/w31161.
- [3] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *MIT Economics Department Working Paper*, Mar. 13, 2023. [Online]. Available: [https://economics.mit.edu/sites/default/files/inline-files/Noy\\_Zhang\\_1\\_0.pdf](https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1_0.pdf).
- [4] S. Peng, E. Kalliamvakou, P. Cihon, and M. Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," *arXiv (Cornell University)*, Feb. 2023, doi: 10.48550/arxiv.2302.06590.
- [5] F. Dell'Acqua et al., "Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality," *Working Paper 24-013*, Sept. 2023, [Online]. Available: [https://www.hbs.edu/ris/Publication%20Files/24-013\\_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf](https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf).
- [6] B. K. Deniz, C. Gnanasambandam, M. Harrysson, A. Hussin, and S. Srivastava, "Unleashing developer productivity with generative AI," McKinsey Digital, [Online]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/unleashing-developer-productivity-with-generative-ai#/>.
- [7] J. T. Liang, C. Yang, and B. A. Myers, "A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and challenges," *arXiv (Cornell University)*, Sep. 2023, doi: 10.48550/arxiv.2303.17125.

# 11

## From development to deployment: neural network implementation in telco using PySpark



This study focuses on the application of predictive analytics within the telco sector, specifically on deploying Neural Network (NN) models for prepaid mobile services. Building on the laboratory phase, which highlights the superior performance of NN models over traditional Machine Learning (ML) models, such as Random Forest (RF) and Gradient Boosting Trees (GBT), in predicting customer behaviors like top-up propensity and balance before the next top-up, this study investigates their viability in a production environment. By leveraging PySpark, we test the scalability and efficiency of NN models for daily predictions.

While initial results show that, overall, NN models outperform traditional ML models, to fully assess the viability of these NN models in a production setting, we extend our testing period over a month to better evaluate their effectiveness compared to existing ML models. This research provides valuable insights into the challenges and potential of deploying advanced predictive models in the telco sector, offering a pathway toward more precise and data-driven marketing strategies.

---

## Authors

**Beatriz Mesquita, Altice Labs, [beatriz-s-goncalves@alticelabs.com](mailto:beatriz-s-goncalves@alticelabs.com)**

**Bernardo Duarte, Altice Labs, [bernardo-x-duarte@alticelabs.com](mailto:bernardo-x-duarte@alticelabs.com)**

**Francisco Silva, Altice Labs, [francisco-c-silva@alticelabs.com](mailto:francisco-c-silva@alticelabs.com)**

**Petia Georgieva, DETI/IEETA, University of Aveiro/Instituto de Telecomunicações, [petia@ua.pt](mailto:petia@ua.pt)**

---

## Keywords

Predictive analytics; Telco sector; NN models; Prepaid mobile services; ML models; RF; GBT; Customer behaviors; PySpark; Scalability; Efficiency; Production; Data-driven marketing strategies



## Introduction

Understanding and predicting customer behavior, particularly within prepaid mobile services, remains crucial for tailoring effective marketing strategies in the ever-evolving telco sector. Prepaid services, which allow customers to pay in advance for telco services like calls, SMS, MMS, and internet access, offer users control over their spending and free them from the contractual obligations associated with postpaid plans. However, the unpredictability of top-up events poses significant challenges for telco operators aiming to optimize marketing campaigns to meet individual customer needs [1].

This article builds on previous research [2] by extending the application of NNs to enhance predictive models in the prepaid mobile services domain. Specifically, we investigate the propensity of customers to top-up within specific time frames (two, three, or four days) and estimate account balances before top-up events. Accurate predictions in these areas enable telco operators to time their marketing communications more effectively, increasing the relevance and efficiency of promotional efforts [3].

Our primary objective is to assess the viability of deploying NNs in a production environment for these specific use cases. To achieve this, we leverage Elephas, a library that brings deep learning with Keras (a high-level neural networks API written in Python [4]), to PySpark, enabling the distributed training of NNs on large datasets. Elephas maintains Keras's simplicity and high usability, facilitating the fast prototyping of distributed models that can run across Spark's Resilient Distributed Datasets (RDD) and DataFrames. This study evaluates the performance of NN models trained with Elephas against established ML models, such as RF and GBT, which Altice Labs' Data Science team has successfully deployed. These traditional models are fully integrated and operational, proving their effectiveness in real-world applications.

Motivated by recent shifts in the usage of prepaid mobile services in Portugal [5], [6], this study addresses the need for more effective marketing strategies to

promote the adoption of these services. By deploying NN models using Elephas in a production setting, we aim to determine whether these models can sustain their superior performance over traditional ML models (as observed in laboratory conditions) and ultimately support telco operators in delivering more personalized and timely marketing campaigns.

This article outlines the deployment of these advanced predictive models, covering the design of data pipelines for data intake and transformation, including customer top-up and balance history, hyperparameter tuning to identify the optimal architecture (using Python), distributed training and prediction (with PySpark and Elephas), and the integration of performance indicators for ongoing model monitoring and oversight.

## State of the art

The problem of accurately predicting customer top-up behavior in the telco sector, especially for prepaid mobile services, is characterized by the limited number of publicly available studies. Companies closely guard industry research findings to maintain a competitive advantage, creating this gap. Nonetheless, we can gather valuable insights for our investigation by examining the results of available predictive model research.

P. M. Alves et al. [7] highlight the importance of detailed customer profiling in enhancing service quality and retention within the telco sector. Their study emphasizes the role of predictive analytics in understanding and anticipating customer behaviors, aligning with our research objectives of predicting top-up propensity and balance before the next top-up. While Alves et al. employed sliding window Multiple Linear Regression for monthly predictions, our research explores NNs to study if they can provide more accurate predictions over the specified timeframes.

The usage of Recency, Frequency, and Monetary Value (RFM) analysis in prior studies [8] to identify top-up patterns reflects our approach to feature

engineering, which also considers seasonal trends to refine predictive modeling.

Studies on churn prediction, though focusing on a different outcome, offer methodologies adaptable for predicting top-up behavior. For instance, Bharadwaj et al. utilized Logistic Regression and Multilayer Perceptron (MLP) models [9], and the success of Deep NNs in outperforming traditional models in churn prediction [10] suggests their applicability in our domain.

Techniques such as Mutual Information (MI) [11], SHapley Additive exPlanations (SHAP) [12], and resampling methods like Synthetic Minority Oversampling Technique (SMOTE) [13] play a crucial role in refining predictive models by tackling class imbalance and optimizing model performance.

## Distributed deep learning in predictive analytics

Advances in distributed deep learning have significantly improved the handling of large-scale datasets and enhanced training efficiency, which is critical for predictive analytics in the telco sector. Ravikumar and Sriraman [14] showcased the effectiveness of distributed deep learning using Elephas and PySpark for real-time pneumonia prediction, achieving faster training times and improved accuracy. Similarly, Al Jallad et al. [15] applied Long Short-Term Memory (LSTM) networks on Apache Spark for optimizing Intrusion Detection Systems (IDS), highlighting the ability to efficiently process and analyze massive datasets, which is essential for our predictive models.

The parallels between these studies and our work are evident, particularly in applying distributed deep learning to improve model scalability and performance. While the domains vary (healthcare, cybersecurity, and telco), the methodologies offer valuable insights into deploying advanced predictive models in production environments. The speed, efficacy, and scalability improvements demonstrated that predicting customer behaviors in the telco sector could benefit from similar advantages. Our study aims to extend these methodologies to

address the unique challenges of predicting top-up behavior and account balances, ultimately providing telco companies with actionable insights to tailor their marketing strategies effectively.

## Methodology

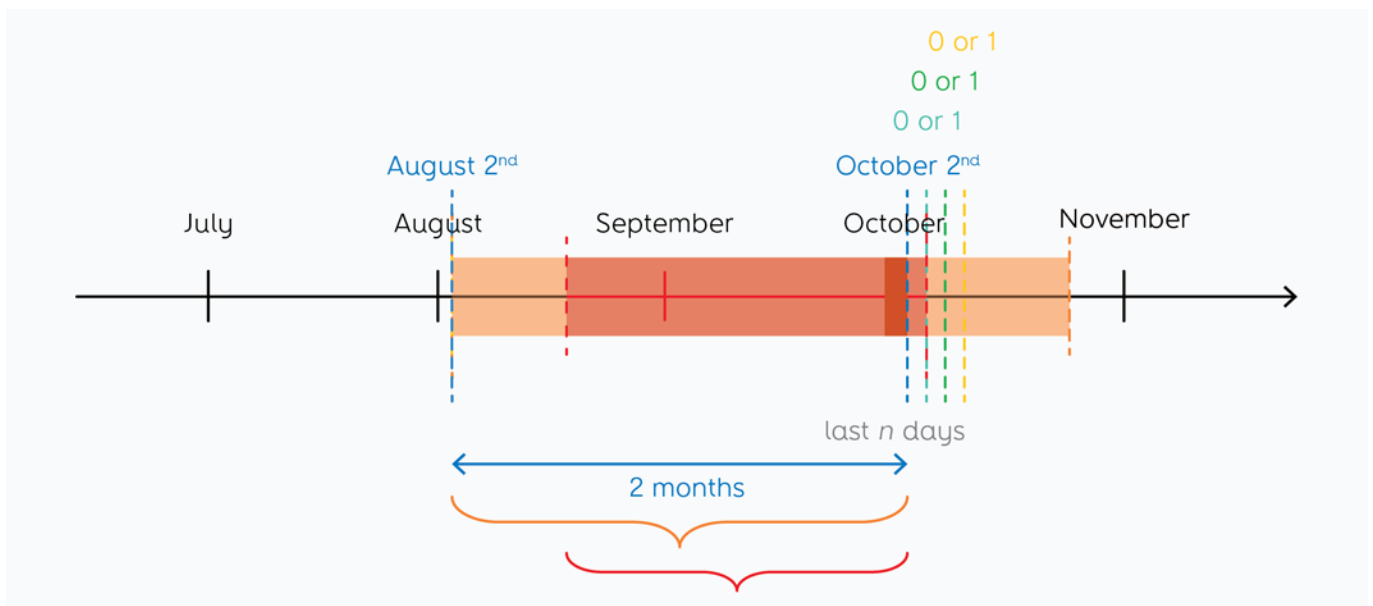
### Top-up propensity prediction in two, three, and four days (binary classification)

Our analysis utilizes 60 days of historical data, focusing on top-up and balance history to generate aggregated features for our model. We tailored these features to reflect customer engagement within this timeframe. This 60-day period balances inclusivity and relevance: more extended periods might capture less engaged customers, diluting predictive model efficacy, while shorter spans might exclude too many customers due to a smaller dataset. Therefore, we exclude customers who did not top-up in the 60 days before our reference day. We set a specific reference day, typically at the start of a month, to anchor our analysis consistently. The concept of a “top-up cycle”, represented by the

orange and red periods in **Figure 1**, is central to our feature construction. These periods correspond to two distinct clients, each with its own top-up cycle. It is defined as the time from a customer’s last top-up before the reference day to their first top-up after it. For feature construction, we consider the period up to the day before the reference day, as shown by the brackets in **Figure 1**, resulting in the concept of the “last top-up cycle”. The period following the reference day, including the day itself, is used for label construction.

Labels are assigned based on whether a customer tops up within a set interval (two, three, or four days) from the reference day, creating binary classification models. Customers topping up within this period are labeled “1”, and those who do not are labeled “0”. This process establishes a binary classification problem aimed at predicting the propensity of a customer topping up within the specified timeframes. In **Figure 1**, dotted lines mark the propensity periods: purple represents two days, green signifies three, and yellow indicates four, starting from the reference day.

Given the highly unbalanced nature of our binary classification datasets, we must carefully select our evaluation metrics. Accuracy, for instance, is unsuitable because it can misleadingly inflate



**FIGURE 1** – Methodology example for top-up propensity within the next two, three, or four days

performance by favoring the majority class, such as predicting all customers as “non-top-up”, yielding high accuracy despite failing to identify any “top-up” cases. Therefore, we use the F1-Score as our metric. It balances precision and recall through their harmonic mean, providing a more accurate assessment for the minority class in our classification problem.

## Balance before the next top-up prediction (regression problem)

This approach for feature temporal constraints is similar to the method used for top-up propensity prediction, where we focus on the last two months leading up to a chosen reference day. The features considered here are identical to those in the top-up propensity prediction, emphasizing the relevance of the “last top-up cycle”.

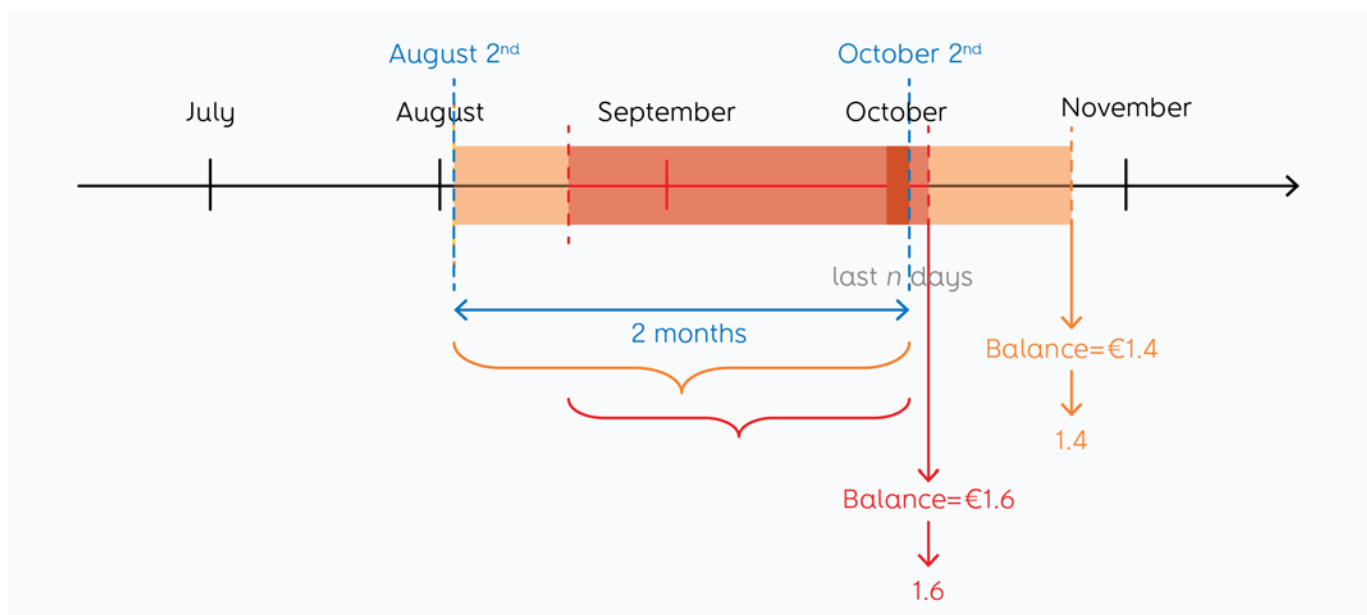
For balance prediction (**Figure 2**), we label data based on the customer’s account balance right before their first top-up following the reference day, which includes the day itself. Our model training involves analyzing the month after the reference day to identify each customer’s first top-up and predict the balance. This approach only considers customers who topped up in the two months before and the month following the reference day.

In this regression problem, we employ two distinct evaluation metrics: the Mean Absolute Error (MAE) [16] and the Hit Ratio. MAE calculates the absolute difference between actual and predicted balances before top-up, directly measuring prediction accuracy. Hit Ratio assesses the proportion of instances where the predicted balance exceeds the actual balance before the next top-up, indicating our model’s ability to anticipate top-ups based on balance evolution. However, relying solely on the Hit Ratio could be misleading. For instance, a model predicting a balance of €20 when the actual balance is €2 would still improve the Hit Ratio despite the prediction being significantly off. Therefore, we aim to balance these metrics, ensuring our predictions anticipate actual top-ups and accurately reflect the customers’ accurate balances.

## Daily performance testing

### Top-up propensity use case

We assess the model’s performance daily throughout the test month using a sliding window approach. Each day is a new reference point for the model to predict customer top-up propensity within the next two, three, or four days. This method offers a detailed evaluation of the model’s consistency and



**FIGURE 2** – Methodology for account balance before the next top-up

accuracy over a short period of one month. We monitor the model's F1-Score, precision, and recall daily to identify performance trends or patterns requiring attention.

## Balance before the next top-up use case

Again, we utilize a sliding window approach. We extend our analysis beyond the training set's constraints for balance prediction. For training, we considered only customers who topped up in the 60-day window before the reference day and the month after the reference day (including the reference day itself); for daily testing, we evaluated all customers who topped up at least once in the 60 days before the reference day. This adjustment allows us to observe the evolution of the predicted balance for each customer daily, offering a dynamic view of anticipated balances. Each day, we compare the predicted balance directly with the actual balance for customers who topped up on the reference day, aiming to measure our model's ability to anticipate imminent top-ups. We employ the Hit Ratio to assess our success in anticipating top-ups and the MAE to ensure the accuracy of our predicted balances.

## Data preprocessing

For our analysis, we strategically selected July 3<sup>rd</sup>, 2023, as the training reference day to capture a high activity phase, using data from May 4<sup>th</sup> to July 2<sup>nd</sup> (60 days), ensuring diverse customer interactions. The testing period spans from August 3<sup>rd</sup> to October 1<sup>st</sup>, with October 2<sup>nd</sup>, 2023, as the test reference date, examining performance across a different activity range.

Our datasets include 58 aggregated features related to customer top-up and balance behavior. For top-up propensity, we created datasets for two, three, and four-day predictions, with 58,472 entries for training and 60,428 for testing. The balance prediction dataset is smaller, with 28,559 entries

for training and 60,428 for testing. However, we only evaluated 25,242 customers for performance purposes due to the exclusion criteria focusing only on active customers in the subsequent month post-reference day.

We engaged in extensive feature engineering and hyperparameter tuning to optimize our NN architectures, including cross-validation. We tested various feature sets, including all 58 features, subsets selected via MI scores, and SHAP analysis for feature importance. We also explored normalization techniques to enhance data handling, such as min-max normalization and standardization.

For the top-up propensity case, we addressed class imbalance (approximately 5% top-up customers) by experimenting with SMOTE and undersampling. We also tested standard and weighted binary cross-entropy loss functions. We designed this comprehensive strategy to identify each use case's most effective NN architecture and feature set.

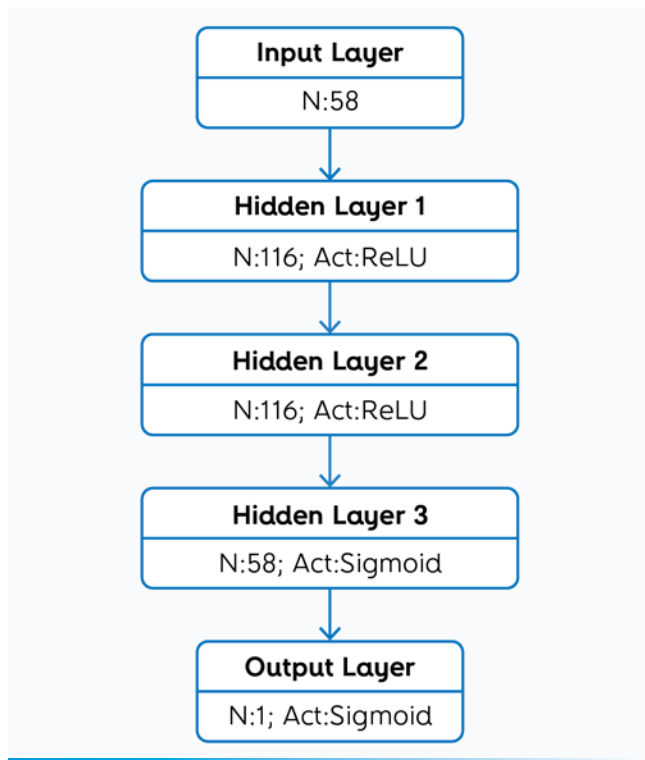
## Best-performing neural networks

### Top-up propensity use case

After extensive experimentation, we identified the most effective model architecture for predicting top-up propensity within two days. Applying this architecture to the three-day and four-day predictions revealed that different input features yielded optimal results for each timeframe. Min-max scaling and using the original class distribution without balancing techniques provided the best performance.

Using all 58 features was most effective for the two-day prediction. However, for the three-day and four-day predictions, selecting a subset of features through SHAP analysis improved performance, with 25 features for three-day and 29 for four-day predictions.

The optimal architecture, consistent across all intervals, is detailed in **Figure 3**. This model was trained for 100 epochs at a batch size of 64, using the Adam optimizer with a learning rate of 0.001 and weighted binary cross-entropy loss. L1 regularization with a strength of 0.001 was applied to the second hidden layer to prevent overfitting.

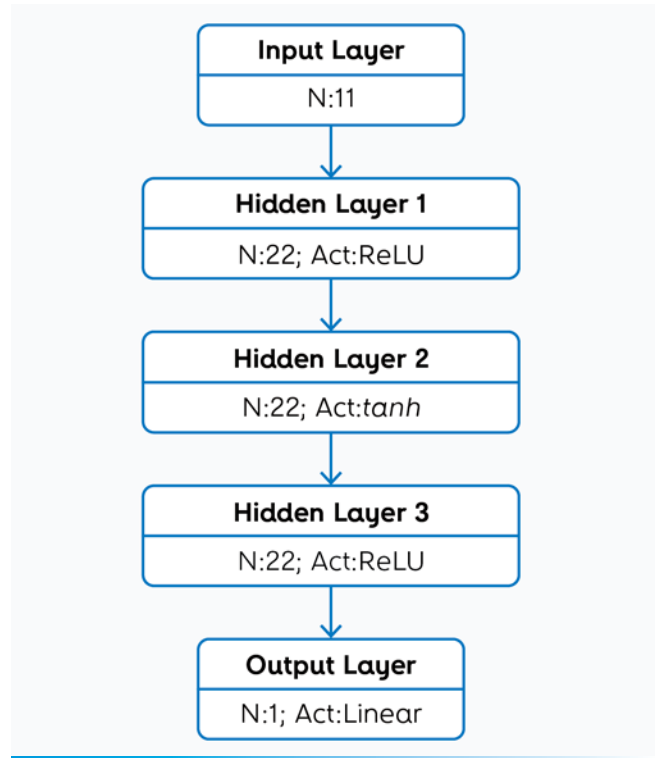


**FIGURE 3** – The optimal architecture of the NN for the top-up propensity use case showcasing layers, their number of neurons (N), and activation functions (Act)

## Balance before the next top-up use case

The model that performed best in the account balance prediction scenario uses a subset of 11 features selected based on MI scores, with min-max scaling for normalization. **Figure 4** shows the most effective architecture, trained over 150 epochs with a batch size of 32, using the Adam optimizer at a 0.001 learning rate. To combat overfitting, we applied L2 regularization with a strength of 0.01 in both hidden layers and implemented early

stopping with a patience of 30 epochs, reducing the total training epochs required.



**FIGURE 4** – The optimal architecture of the NN for the balance use case showcasing layers, their number of neurons (N), and activation functions (Act)

## Distributed modelling

After identifying the best architecture and hyper-parameters for each use case, the next step is to train the models and make predictions in a distributed manner using PySpark, preparing them for deployment. When utilizing Elephas, Keras models are initialized on the driver node, serialized, and then distributed to worker nodes along with the data and broadcasted model parameters. The Spark workers deserialize the model, train their allocated data chunks, and send the resulting gradients back to the driver. An optimizer updates the "master" model on the driver, which can operate synchronously or asynchronously. We use the

synchronous mode as the parallelization mode for both use cases, where the driver waits for all worker nodes to finish processing their data and sending gradients before updating the model. This approach ensures consistent updates but can be slower if some nodes take longer to complete their tasks [17], [14].

The basic model in Elephas is the SparkModel. To initialize a SparkModel, we pass a compiled Keras model along with parameters like update frequency and parallelization mode. The training process updates the model parameters at a frequency determined by the update frequency. In our implementation, we use batch update frequency for both use cases. This approach balances consistency in model updates with the need for efficiency in a distributed setting [17].

Elephas also supports distributed prediction and evaluation, which are crucial for deploying models at scale. Like the fit method, both the predict and

evaluate methods are designed to be compatible with the Keras Model API. Specifically, this compatibility allows the model to be serialized and sent out to worker nodes. At these nodes, the model simultaneously processes separate pieces of the testing data, enabling parallel execution of tasks [17]. Our deployment uses the predict method with an RDD, enabling distributed predictions across our data.

Currently, Elephas does not support the callback functionality [18] that is typically available in Keras. Callbacks are a powerful feature in Keras, allowing for real-time monitoring and control of the training process. They help track metrics like training and validation loss, implementing early stopping, saving model checkpoints, and dynamically adjusting learning rates. Due to this limitation, we cannot apply early stopping to our balance regression model during deployment as we did in the laboratory phase, as seen in **Section 5.2**.



# Daily performance evaluation of distributed neural networks and traditional machine learning models using PySpark

This section evaluates the daily performance of our optimized NN models trained using Keras and Elephas with PySpark. **Figures 3 and 4** depict the architecture of these models. We evaluated our models throughout October 2023 and compared these results with the ones obtained using traditional ML models, which are also implemented in PySpark and currently in production. This daily evaluation provides insights into the consistency and adaptability of the NNs under various conditions, offering a practical perspective on their real-world applicability.

Altice Labs' Data Science team previously employed an RF classifier for top-up propensity prediction and a GBT regressor for balance prediction, with their configurations detailed in **Table 1**. The NN and the traditional ML models utilized July 3<sup>rd</sup>, 2023, as the training reference day; however,

our methods diverged in feature selection and preprocessing strategies. While the traditional ML models used Information Value (IV) for feature selection and did not normalize features, we opted for MI for the balance prediction use case and SHAP analysis for the top-up propensity use case, alongside feature normalization.

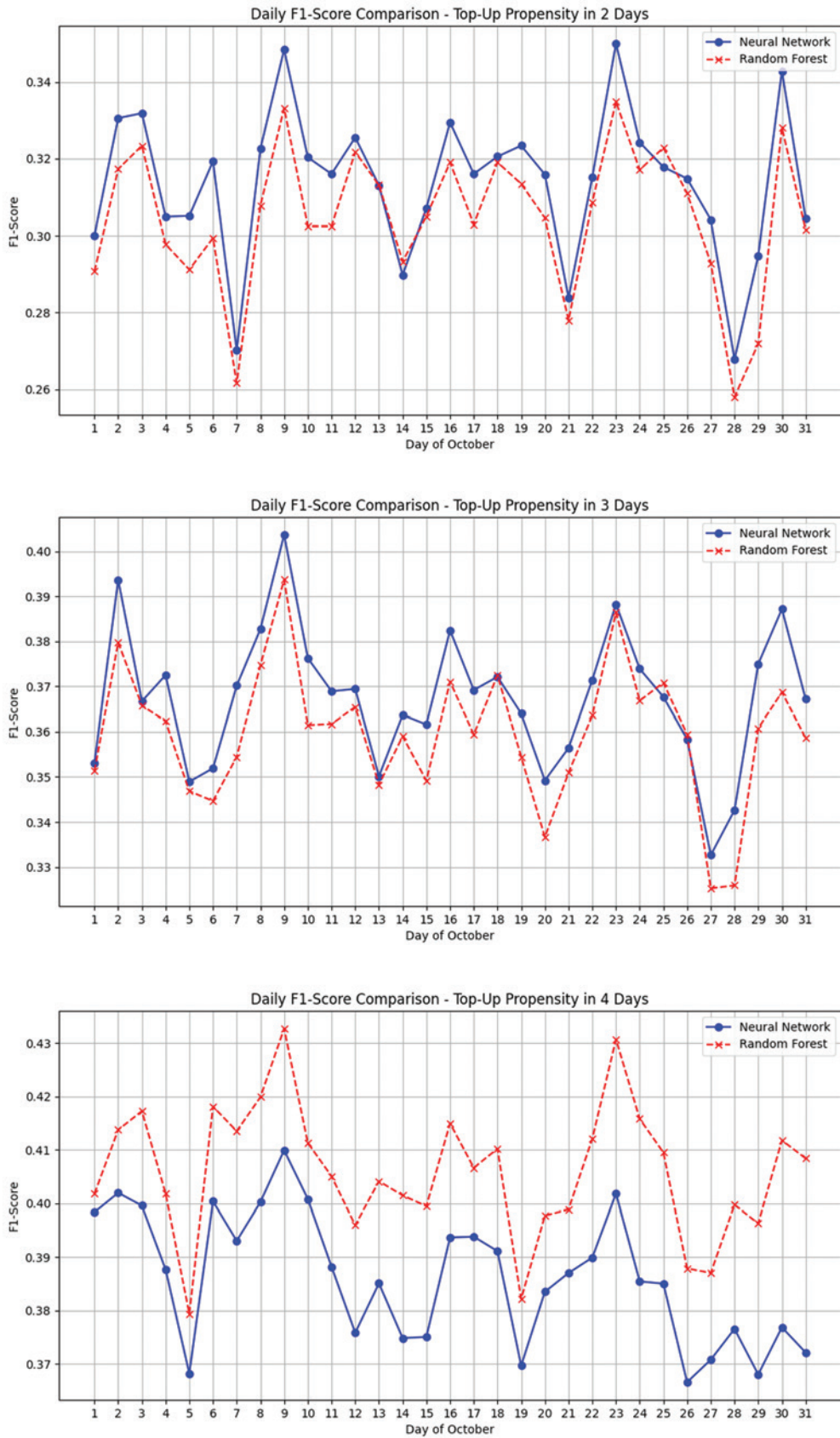
## Top-up propensity use case

**Figure 5** (on the next page) presents the daily performance comparison, in terms of F1-Score, between the NN model (blue) from **Figure 3**, trained using Elephas and Keras with PySpark, and the previously developed RF classifier (red) throughout October. Both models show a similar performance pattern, with higher F1-Scores at the start of the week and declines over weekends and on October 5<sup>th</sup>, a national holiday, suggesting a shared sensitivity to time-related behavioral shifts. This pattern, observed consistently across the two, three, and four-day propensity predictions, may be influenced by the selection of July 3<sup>rd</sup>, a Monday, as the training reference day, potentially biasing the models towards early-week behavior. Alternatively, it could reflect the naturally erratic customer behavior observed during weekends.

Our analysis reveals that for the two-day and three-day propensity predictions, the NN generally outperformed the RF classifier throughout October 2023,

Top-up Propensity Use Case		Balance Before The Next Top-up Use Case	
RF Classifier		GBT Regressor	
max_depth	11	learning_rate	0.1
criterion	entropy	criterion	friedman_mse
min_samples_split	2	min_samples_split	0.1
n_estimators	500	estimators	100
		loss	absolute
		max_depth	7
		min_samples_leaf	0.05

**TABLE 1** – Configuration details of traditional ML models



**FIGURE 5** – F1-Score obtained with the NN classifier (blue line) and with the RF classifier (red line) for the propensity use case over the testing month of October

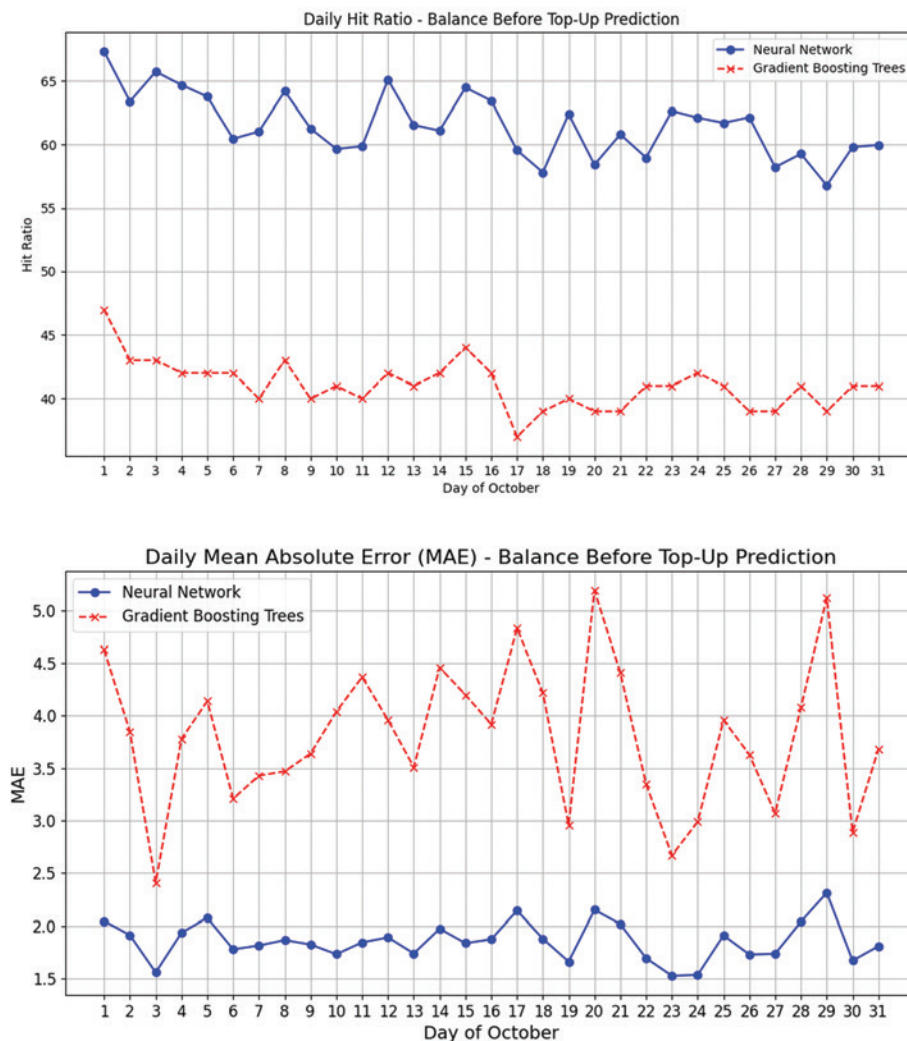
indicating an overall performance improvement with the NN. During the laboratory phase, when developing the NNs using Keras in Python, we observed a more significant performance gap between the NN and RF. However, when implementing the models with PySpark, this difference diminished slightly, though the NN still demonstrated improved performance.

In contrast, for the four-day propensity prediction, the NN did not surpass the RF classifier, consistently showing lower F1-Scores throughout the testing month. This outcome contradicts our findings from the laboratory phase, where the NN outperformed the RF across all propensity use cases. This discrepancy suggests that the NN developed with PySpark may not generalize as effectively

for the four-day propensity prediction. We must recall that we explicitly optimized the architecture and hyperparameters shown in **Figure 3** for the two-day propensity prediction. Then, we applied the same configuration to the three-day and four-day predictions without additional tuning due to computational constraints in the laboratory phase.

### Balance before the next top-up use case

In **Figure 6**, we compare the Hit Ratio (left) and MAE (right) between the GBT regressor (red) and our NN regressor (blue), aiming to balance these metrics optimally.



**FIGURE 6** – Hit Ratio (top plot) and MAE (bottom plot) obtained with the NN regressor (blue line) and the GBT regressor (red line) for the prediction of balance before the next top-up over the testing month of October

Throughout October, our NN regressor consistently achieved a higher Hit Ratio and significantly lower and more stable MAE than the GBT regressor, indicating more accurate and reliable predictions. Notably, both metrics were better at the start of the month but tended to worsen as the month progressed. This pattern may be due to both models being trained on data from July 3<sup>rd</sup>, suggesting they better capture customer behavior typical of the month's beginning and less adeptly generalize across the entire month. Alternatively, this trend could reflect the naturally erratic customer behavior often observed towards the end of the month.

Comparing these results with those obtained during the laboratory phase, we observed a notable improvement in the Hit Ratio: previously, it fluctuated between 42-52% during the test month, but with the NN trained with PySpark, it now varies approximately between 56-67%. For MAE, we observed slightly higher values. In our laboratory phase, MAE remained below €2 euros; in this case, the highest MAE reached approximately €2.3. Despite this slight discrepancy, we achieved better overall performance when training the NN with Keras and Elephas for the regression problem.

## Conclusions

This study explores predictive analytics within the telco sector, focusing on predicting top-up propensity and account balances before top-up in prepaid mobile services. We compared NNs trained using Keras and Elephas with PySpark to traditional ML models previously employed by the team, which are already in production.

Our research highlights the effectiveness of advanced feature selection methods, such as MI and SHAP values, in improving NN model performance during the laboratory phase while developing the NNs in Python. However, when transitioning to a distributed environment using PySpark, we observed variations in model performance that are important to note.

For the top-up propensity prediction, the distributed training environment reduced the performance gap for the two-day and three-day propensity predictions. Although NNs still outperformed traditional models, the improvement was slightly less pronounced. For the four-day propensity prediction, NNs could not surpass the RF model's performance. This could be attributed to the overhead and complexities of managing data across multiple nodes and the sensitivity of classification tasks, particularly with imbalanced datasets, to such variations.

In contrast, the balance prediction task improved performance with the NN model in the distributed setup. Regression models, which optimize for continuous output, proved more resilient to the subtle variations introduced by distributed training.

In terms of execution time, the distributed setting demonstrated significant advantages. For the top-up propensity prediction use case, training took approximately 4 minutes and 40 seconds in the local environment and around 4.45 seconds to make predictions on the test dataset. In the distributed environment, we reduced the training time to approximately 1 minute and 7 seconds, with predictions taking around 30 seconds. For the balance prediction use case, training took approximately 5 minutes, and the prediction was 1.50 seconds in Python. The training and prediction times were approximately 56 seconds and 26 seconds in the distributed setting, respectively.

Overall, our findings illustrate the potential of distributed deep learning to enhance predictive analytics within the telco sector. This study only evaluated performance over one test month. However, to fully assess the reliability and effectiveness of NNs in a distributed environment compared to traditional ML models, a more extended evaluation period, such as one year, would be necessary to observe performance consistency and adaptability over time.

## Future work

There are numerous avenues for future enhancements to our models. We applied the same NN architecture across the three prediction intervals (two, three, and four days) for the top-up propensity use case. A potential improvement could involve fine-tuning the NN architecture separately for each prediction period, enhancing performance by tailoring each model to its specific timeframe.

Additionally, adjusting the timing of training could address observed performance patterns. For instance, in the top-up propensity case, we noted superior performance at the beginning of the week compared to the end. Our choice of reference day for training might influence this variation. To explore this further, we could experiment with training models using reference days from the start and end of the week to capture a broader spectrum of

customer behaviors throughout the week. Similarly, for the balance prediction use case, which showed improved performance at the month's start and declined towards the end, setting training reference days at the beginning, middle, and end of the month might enable our models to capture the full range of monthly customer behaviors.

Integrating time series analysis into these use cases presents another promising direction. Transitioning from aggregated features to those that reflect the temporal evolution of customer behavior could allow using models like LSTM networks, potentially offering more profound insights into behavioral patterns.

These suggested research areas built upon our current findings are crucial for adapting to the dynamic nature of customer behavior in the telco sector. They have the potential to enhance the relevance and effectiveness of predictive models in this rapidly evolving industry. [🔗](#)



## References

- [1] Y. Zheng et al., "TelcoFlow: Visual exploration of collective behaviors based on telco data," in *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 843–852, doi: 10.1109/BigData.2016.7840677.
- [2] B. Mesquita, B. Duarte, F. Silva, and P. Georgieva, "Predictive Analytics for Customer Behavior - Top Up Propensity and Account Balance Predictions in Prepaid Mobile Services," in *2024 IEEE 22<sup>nd</sup> Mediterranean Electrotechnical Conference (MELECON)*, Porto, Portugal, 2024, pp. 1315–1320, doi: 10.1109/MELECON56669.2024.10608654.
- [3] S. Qiu, L. Wu, Y. Yang, and G. Zeng, "Offering the right incentive at the right time: Leveraging customer mental accounting to promote prepaid service," *Ann Tour Res*, vol. 93, p. 103367, March 2022, doi: 10.1016/j.annals.2022.103367.
- [4] F. Chollet et al., "Keras." github.com. Accessed: Aug. 10, 2024. [Online]. Available: <https://github.com/fchollet/keras>.
- [5] ANACOM, "Em 2022 o número de assinantes aumentou 3,4%," *anacom-consumidor.pt*. Accessed: Aug. 10, 2024. [Online]. Available: <https://www.anacom-consumidor.pt/-/em-2022-o-numero-de-assinantes-aumentou-3-4->.
- [6] ANACOM, "Serviços móveis - 2.o trimestre de 2024," *anacom.pt*. Accessed: Aug. 10, 2024. [Online]. Available: <https://anacom.pt/render.jsp?contentId=1792744>.
- [7] P. M. Alves, R. Â. Filipe, and B. Malheiro, "Telco customer top-ups: Stream-based multi-target regression," *Expert Syst*, vol. 40, no. 5, p. e13111, 2023, doi: 10.1111/exsy.13111.
- [8] P. M. Alves, R. Â. Filipe, and B. Malheiro, "Towards Top-Up Prediction on Telco Operators," in *Progress in Artificial Intelligence*, F. S. and L. N. and L. C. H. and R. L. P. Marreiros Goreti and Melo, Ed., Cham: Springer International Publishing, 2021, pp. 573–583.
- [9] S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra, and S. Kumar, "Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron (MLP)," in *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*, Bangalore, India, 2018, pp. 436–438, doi: 10.1109/ICGCIoT.2018.8752982.
- [10] N. Putu Oka H and A. Setyo Arifin, "Telecommunication Service Subscriber Churn Likelihood Prediction Analysis Using Diverse Machine Learning Model," in *2020 3<sup>rd</sup> International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)*, Medan, Indonesia, 2020, pp. 24–29, doi: 10.1109/MECnIT48290.2020.9166584.
- [11] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in *2012 15<sup>th</sup> International Multitopic Conference (INMIC)*, Islamabad, Pakistan, 2012, pp. 23–27, doi: 10.1109/INMIC.2012.6511498.
- [12] A. Dumitrache, A. Nastu, and S. Stancu, "Churn Prediction in Telecommunication Industry: Model Interpretability," *Journal of Eastern Europe Research in Business and Economics*, vol. 2020, Feb. 2020, Art. no. 241442.

- [13] A. Raj and D. Vetrithangam, "Prediction of customer churn using resampling and ensemble classification algorithms on telecom dataset," in *2023 11<sup>th</sup> International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP)*, Nagpur, India, 2023, pp. 1–7, doi: 10.1109/ICETET-SIP58143.2023.10151659.
- [14] A. Ravikumar and H. Sriraman, "Real-time pneumonia prediction using pipelined spark and high-performance computing," *PeerJ Comput Sci*, vol. 9, p. e1258, Mar. 2023, doi: 10.7717/peerj-cs.1258.
- [15] K. Al Jallad, M. Aljnidi, and M. S. Desouki, "Big data analysis and distributed deep learning for next-generation intrusion detection system optimization," *J Big Data*, vol. 6, no. 1, p. 88, 2019, doi: 10.1186/s40537-019-0248-6.
- [16] V. Plevris, G. Solorzano, N. Bakas, and M. Ben Seghier, "Investigation of performance metrics in regression analysis and machine learning-based prediction models," in *8<sup>th</sup> European Congress on Computational Methods in Applied Sciences and Engineering*, CIMNE, Oslo, Norway, 2022, doi: 10.23967/eccomas.2022.155.
- [17] M. Pumperla and D. Cahall, "Elephas: Distributed Deep Learning with Keras & Spark," *J Open Source Softw*, vol. 7, no. 80, p. 4073, Dec. 2022, doi: 10.21105/joss.04073.
- [18] "tf.keras.callbacks.Callback," tensorflow.org. Accessed: July 8, 2024. [Online]. Available: [https://www.tensorflow.org/api\\_docs/python/tf/keras/callbacks/Callback](https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/Callback).

# 12

Machine-learning aided fire detection systems – an opportunity for early response and improved safety



The focus of this work, in the context of the AlticeLabs@UA<sup>1</sup> project - SAFER-Fire, was to develop a machine learning model capable of detecting a fire in its early stages while being lightweight enough to be used in a low-power device, such as a single-board computer. Different algorithms were analyzed, and YOLO (You Only Look Once) was elected the best algorithm for accuracy and performance. The need for a consistent dataset to train and develop our models is highlighted, which required extending a public dataset and enriching it with carefully labeled samples for fires in early stages to improve the model's global accuracy. We then describe different iterations in terms of the choice of model parameters, as well as the challenge of deploying that model on a single-board computer.

This work aims to improve the safety of buildings and communities. It is directly linked to the sustainable development approach in the context of sustainable cities, and communities and innovation and infrastructure.

---

## Authors

**Carlos Couto, RISCO & DECivil, University of Aveiro, [ccouto@ua.pt](mailto:ccouto@ua.pt)**

**Rúben Santo, DFis, University of Aveiro, [rubensanto@ua.pt](mailto:rubensanto@ua.pt)**

**Hélder Caixinha, DigiMedia, DeCA, University of Aveiro, [caixinha@ua.pt](mailto:caixinha@ua.pt)**

**Jorge Ferraz de Abreu, DigiMedia, DeCA, University of Aveiro, [jfa@ua.pt](mailto:jfa@ua.pt)**

**Fausto de Carvalho, Altice Labs, [cfausto@alticelabs.com](mailto:cfausto@alticelabs.com)**

---

## Keywords

Fire and Smoke detection system; FireView; YOLO; Computer vision; Machine learning; Smart home applications

## Introduction

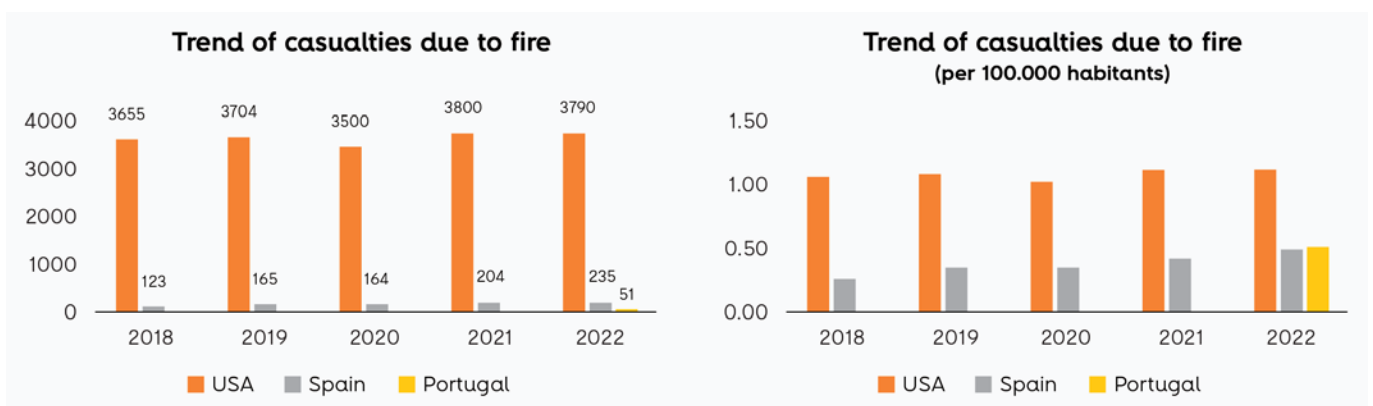
Fire is one of the most significant accidental hazards, responsible for a considerable loss of life, service disruptions, and loss of use of buildings, with profound economic and social impacts. It stands out among all extreme events that may affect a building during its lifespan due to its high frequency and severe consequences. It has one of the highest occurrence rates per 1,000 inhabitants and is associated with a notably high fatality rate [1]. **Figure 1** shows the trend of casualties due to fire in the USA, Spain, and Portugal between 2018 and 2022, according to the data gathered in the World Fire Statistics [1] report.

Different passive and active fire protection systems can be deployed in a building to reduce these impacts. Passive fire protection systems are those related to fire resistance and include, for example, improved building designs against fire, while active fire protection systems are those triggered by fire. The latter may include early fire detection systems (e.g., alarms), sprinklers, automatic fire doors, etc. In the field of early fire detection in residential and industrial buildings, existing systems are known to originate a large number of false alarms, which is particularly detrimental when a fire brigade is mobilized, so a trade-off between automation and efficiency must be sought to improve the false alarm/true event binomial [2]. Recent statistics in England show that the ratio of false alarms is, on average, above 40%.

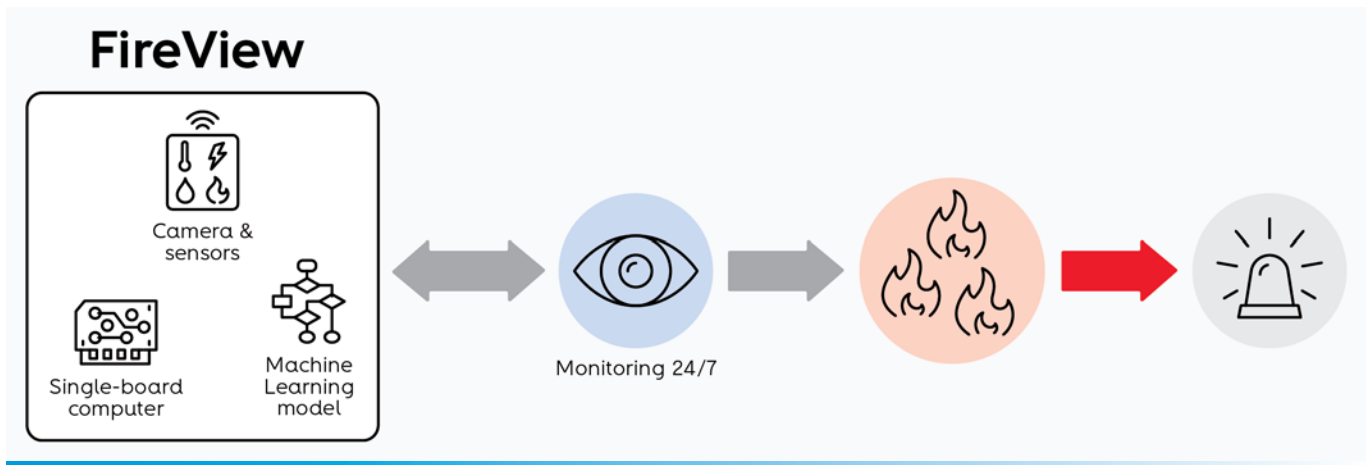
In this context, several opportunities exist to develop and incorporate state-of-the-art machine learning models to improve fire detection algorithms. In the ongoing SAFER-Fire project, we are developing machine-learning models capable of detecting fires at an early stage based on innovative algorithms that recognize complex fire-related patterns. Our approach merges computer vision techniques for pattern detection with state-of-the-art object detection machine learning models such as Ultralytics YOLOv8, aiming to deliver lightweight, fast, and accurate models that can run on a single-board computer, providing continuous 24/7 monitoring. This system was named “FireView”.

The core of FireView lies in its use of the YOLO object detection model, known for its balance of speed and accuracy, specifically, YOLOv8. In the development of our model, firstly, we analyzed a public dataset and its limitations and found the need to enlarge it with more samples. Indeed, to improve the model’s performance in early-stage fire detection, we added video frames of small-scale ignition events to ensure that our system was better suited to provide the desired early warnings. We then explored various configurations of YOLOv8 to identify the best trade-off between processing speed and accuracy. We included different model sizes and evaluated how they performed on both standard and expanded datasets.

The scheme of our FireView system is depicted in **Figure 2**, and this article focuses explicitly on details regarding the development of the machine learning model.



**FIGURE 1** – Trend of casualties due to fire (note: data for Portugal only for 2022) (source: [1])



**FIGURE 2** – FireView approach

The intended fire detection system will then be capable of issuing different triggers to enable immediate action, such as direct authorities’ notification, suppression systems’ activation (such as sprinklers), or it can continually evolve by integrating other Artificial Intelligence (AI) enhanced systems to, for example, determine the best fire escape routes in large buildings.

Therefore, our approach has the potential to deliver a cost-effective solution that can be readily available in high-importance buildings, linked to other active fire protection systems, and contribute to increased safety against the extreme event of a fire.

## Related work

The primary goal of a fire detection system is to detect early fire with fewer false positives [3]. With the recent advances in AI and computer vision algorithms, many researchers are using these methodologies to develop innovative fire detection systems. Among these, the convolutional neural networks (CNN) present a very appealing solution due to their high efficiency, high performance, and good recognition rates [4]. Since fire and smoke vary in texture, and as a result, the variance of the captured images is relevant in terms of color, angle, and lighting conditions, researchers have been using architectures that take advantage

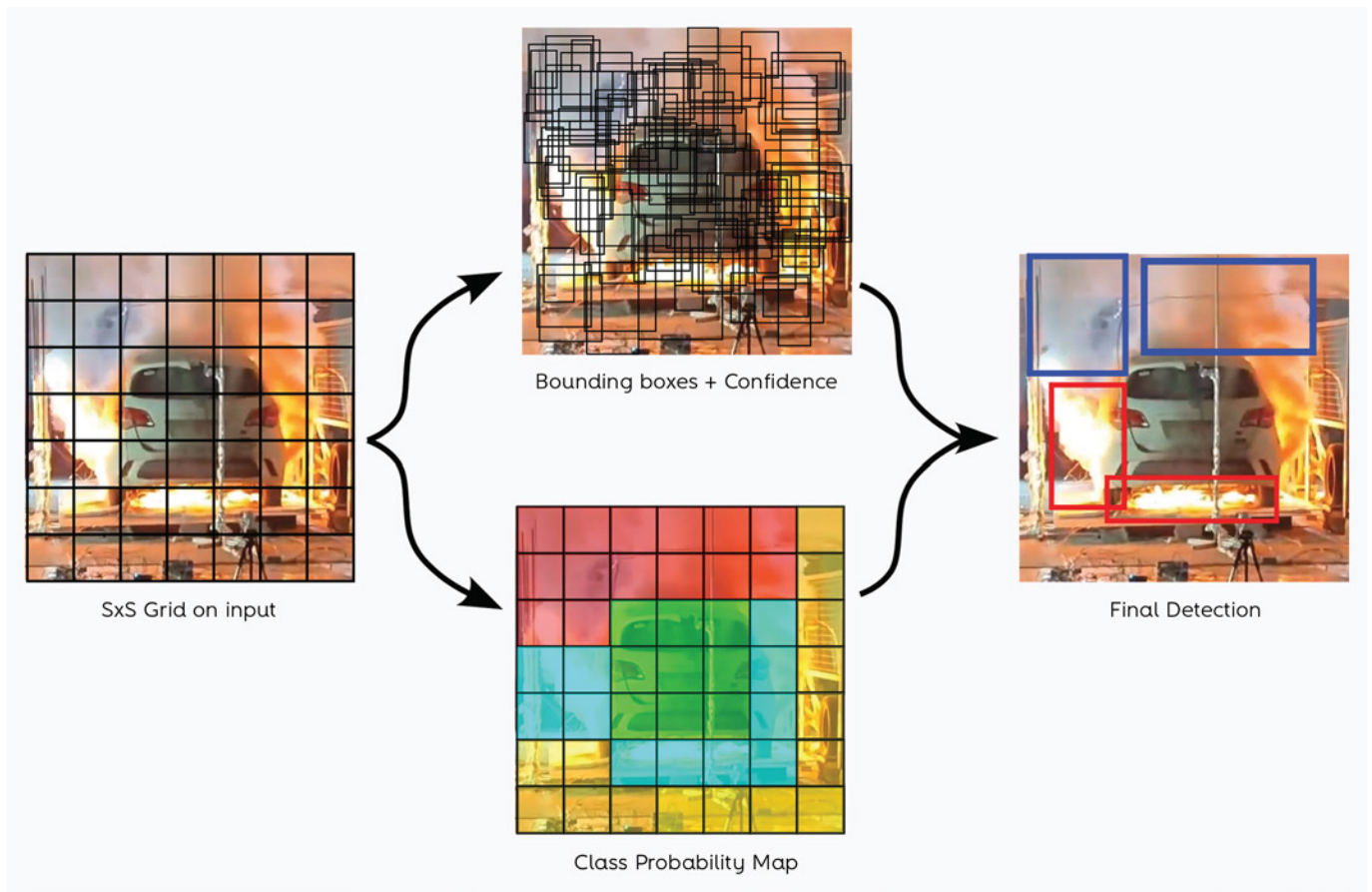
of such observation, for example, a dual channel network CNN [5] or a dark channel to detect smoke combined with a CNN [6]. However, these approaches increase the model size and complexity and are more computationally intensive to train and deploy. Another example tries to detect flames and smoke by movement, using spatiotemporal CNN (e.g., [7]) to simultaneously recognize smoke and estimate optical flow. Nevertheless, processing in real time is intensive and unsuitable for low-cost systems. Kim and Lee [8] and Zeng et al. [9] have implemented fire and smoke detection systems based on Faster Region-based Convolutional Neural Networks (R-CNN) also using motion (i.e., video sequences) to mimic human fire detection by identifying suspected fire regions. However, these networks struggle with changes in imaging angles, lighting, and scale, leading to higher false detection rates and significant computational costs, limiting real-time effectiveness. There are examples of more complex architectures that tackle the variance related to image and color transformations, such as Deep Multi-Scale CNNs (DMCNN) [10], achieving high accuracy with low false alarm rates. However, the additional model's complexity makes it unsuitable for deployment on low-end machines. The same issue is also observed in Recurrent Neural Networks (RNNs) [11], [12], where higher accuracy with low false detection rates is achieved at the cost of higher network complexity and slower detection capacity.

Reduced complexity CNNs can be applied to fire detection, are lightweight, offer fast results, and can be deployed on low-end machines. However, their shallow feature learning results in lower accuracy and higher false alarm rates in real-world scenarios. An example of the application of basic CNNs to flame and/or smoke is found in [13]. On the other hand, Deep CNNs are more complex architectures that can be used to identify suspected flame regions [14]. These models employ multiple layers to capture complex features, improving accuracy and minimizing false detection rates. As disadvantages, we can mention that the increased depth and complexity result in heavier networks, longer processing times, and limited deployment on low-end devices.

Finally, we observe a needed balance between the cost and efficiency of the machine learning algorithm. If the primary goal is to deploy a low-cost model to detect fire, it should be simultaneously accurate and fast.

## Methodology

The main drawback of the earlier discussed machine learning models is that the processing power required makes them too computationally intensive to be implemented in low-cost systems, as discussed in the previous section. Answering this issue, the YOLO algorithm proposed by Redmon et al. [15] presented a breakthrough in object detection, accomplishing the detection task with a single pass of the network (see **Figure 3**), as opposed to previous approaches that either used sliding windows followed by a classifier that needed to run hundreds or thousands of times per image or the more advanced methods that divided the task into two-steps, where the first step detects possible regions with objects or “regions proposals” and the second step run a classifier on the proposals. YOLO also uses a more straightforward output



**FIGURE 3** – Object detection based on the Yolo architecture

based only on regression to predict the detection outputs as opposed to other algorithms that use two separate outputs: a classification for the probabilities and a regression for the box coordinates.

In our approach, we made use of Ultralytics YOLOv8 to develop and train a machine-learning model capable of detecting a fire in its early stages.

## YOLOv8 model development

In developing our early fire detection system, we started by testing different versions of the YOLOv8 model—(n)ano, (s)mall, (m)edium, (l)arge, and e(x)tra-large—to find the right balance between accuracy and real-time performance. The YOLOv8 models scale in both parameters and, consequently, performance, with YOLOv8-nano (3.2M parameters) offering the fastest speed but lowest accuracy, YOLOv8-small (11.2M) providing a balance between speed and accuracy, and larger models like YOLOv8-medium (25.9M), YOLOv8-large (43.7M), and YOLOv8-extra-large (68.2M) delivering higher accuracy at the cost of slower processing.

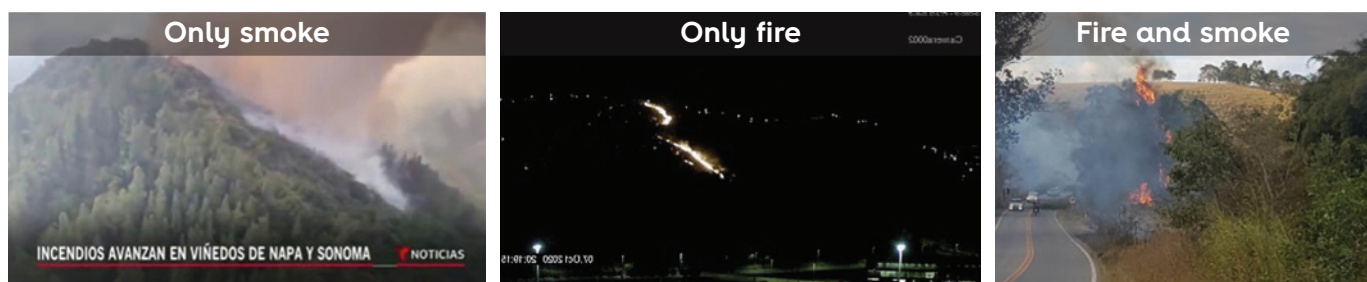
In addition to model size, different image resolutions were tested (by altering the `imgsz` parameter) to assess whether higher-resolution images lead to improved detection performance. Besides that, training was conducted using both a publicly available dataset and our expanded dataset, which included the additional frames of early-stage fires described in the previous section. The combination of these enriched datasets aimed to enhance the model's generalizability and robustness in detecting early fire stages across diverse scenarios.

## Existing fire dataset

To establish a baseline for our fire detection system machine learning model, we initially used the D-Fire dataset [16], a publicly available resource<sup>2</sup> with a diverse set of images categorized into four classes: "Only Fire," "Only Smoke," "Fire and Smoke," and "None". As highlighted in **Table 1**, the dataset consists of 1,164 images containing only fire, 5,867 images with only smoke, 4,658 images featuring both fire and smoke, and 9,838 images depicting neither fire nor smoke. Some example samples are plotted in **Figure 4**.

Class	Train	Test	# of samples
Only fire	944	220	1164
Only smoke	4681	1186	5867
Fire and smoke	3763	895	4658
No fire nor smoke	7833	2005	9838
<b>Total</b>	<b>17221</b>	<b>4306</b>	<b>21527</b>

**TABLE 1** – Classes of the DFire dataset [16] used for training and testing



**FIGURE 4** – Examples of images and their classes contained in the original dataset [16]

The dataset includes a total of 14,692 bounding boxes for fire instances and 11,865 bounding boxes for smoke instances. These bounding boxes were essential for training our YOLOv8 model, enabling it to learn to detect fire and smoke across various scenarios and conditions.

## Expanding the dataset (D-Fire-plus)

We have identified a significant limitation of the D-Fire dataset: it lacks sufficient samples of early-stage fires, where flames are small, and smoke may be minimal or dispersed. This shortcoming is critical because early detection accuracy is crucial for preventing small fires from escalating into major incidents. Therefore, to enhance the model's ability to provide early warnings, augmenting the dataset with additional samples specifically capturing these early-stage fires is necessary.

In the context of developing an early fire detection system using the YOLO object detection model, creating a highly specific and meticulously curated dataset is crucial. The key objective of such a system is to ideally detect fires before conventional sensors, such as smoke and temperature detectors, can identify the threat. This requires the YOLO model to be trained on images that accurately represent the initial stages of a fire, focusing on small flames that might otherwise be overlooked.

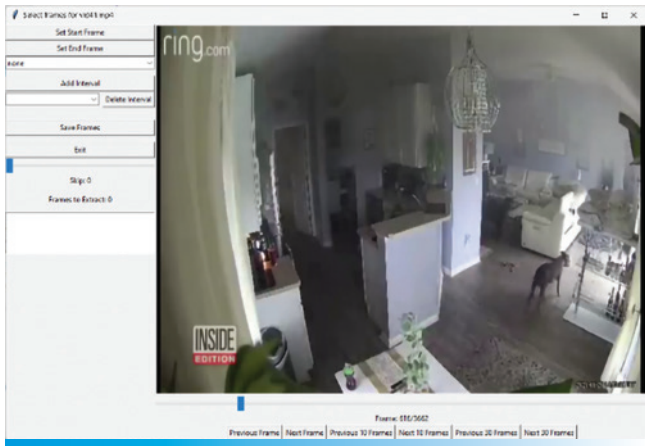
We built a dataset by extracting frames from a series of carefully selected YouTube videos to achieve this. These videos predominantly feature small-scale fires in their ignition stages. Our current

dataset comprises 5,441 frames sourced from 39 videos, with 754 frames meticulously labeled. This labeling process is crucial as it directly impacts the model's ability to learn and generalize from the data. **Table 2** shows the samples contained in the D-Fire-plus dataset, where the increase in the number of samples between parenthesis is indicated. It should be noted that the test subset remains equal to enable comparison between the two datasets.

The video frames were extracted using a custom-developed tool designed to provide precise control over the extraction process (see **Figure 5**). This tool allows the extraction of frames from a video directly into designated folders, ensuring that each frame is categorized appropriately. Users can define specific intervals of the video to be converted into frames, ignoring irrelevant sections, and set a step to determine the frequency of frame extraction within each interval. This feature is handy for capturing higher frame density in critical segments, such as those depicting early-stage fires. Additionally, the tool allows users to specify the class of extracted frames to be automatically saved into the correct folder. This organization is invaluable for the subsequent labeling process, as it streamlines the task of drawing bounding boxes by ensuring that the data is already sorted by class, making clear what each frame represents. This meticulous approach ensures that the dataset is both comprehensive and focused, capturing essential visual cues without overwhelming the model with redundant data.

Class	Train	Test	# of samples
Only fire	1423 (+479)	220	1643
Only smoke	4876 (+195)	1186	6062
Fire and smoke	3843 (+80)	895	4738
No fire nor smoke	7833	2005	9838
<b>Total</b>	<b>17975 (+754)</b>	<b>4306</b>	<b>22281</b>

**TABLE 2** – Expanding the original dataset to construct the D-Fire-plus



**FIGURE 5** – Interface of the tool to extract the frames of videos

Once the frames of interest were extracted, a rigorous labeling process was undertaken. We used the auto-labeling tool provided by Roboflow<sup>3</sup> to initially

label the frames, followed by a manual verification and correction process using YoloLabel<sup>4</sup>. This ensures that each frame is accurately annotated, which is vital for the YOLO model to learn effectively from the data.

Lastly, we applied data augmentation techniques, manipulating the values following a heuristic approach to observe their impact on model performance. We slightly adjusted image scales and shear angles to help the model learn under different conditions and applied perspective shifts to simulate different viewpoints. Randomly flipping images ensured that the model could recognize fire and smoke regardless of orientation. We also used mosaic augmentation to create more varied and complex training scenarios and mix-up augmentation to challenge the model with overlapping or obscured objects.



## Training

The training was performed on one LCA Navigator Cluster GPU computing node<sup>5</sup>. The node has 2x NVIDIA Tesla V100 16 GB, 2x Intel Xeon Gold 6148 (20-core) @ 2.40 GHz, and 96 GB DDR4-2666 R ECC. For the model #9 with 100 epochs (see **Table 3**), the training took around 8 hours. A Linux bash script was prepared to launch the training, using a Conda environment with the versions of Python 3.10.14, Numpy 1.26.4, Pytorch 1.12.1, Torchvision 0.13.1, Ultralytics 8.2.18, and Cuda Toolkit 11.3.1.

It should be noted that training was run using the NVIDIA CUDA technology to accelerate the training of the models. The correct version of the

packages to be installed in the environment is necessary since the NVIDIA drivers were set to version 510.47.03.

To determine the optimal training duration for our dataset, we conducted experiments using both 10 and 100 epochs. These trials highlighted that the model converges well before reaching 100 epochs, suggesting that an early stopping strategy could be more efficient. By implementing early stopping, we could prevent unnecessary computational expense, particularly in cases where the model achieves its best performance early in the training process.

To evaluate the performance of the models, the Precision and Recall metrics are given (**Table 3**) by setting a confidence level of 40%. Precision

#	Model	Imgsz	epochs	Dataset	Data aug?	Precision	Recall
1	L	640	10	D-Fire	No	0.6289	0.5558
2	M	640	10	D-Fire	No	0.6596	0.5852
3	M	1280	10	D-Fire	No	0.5889	0.5597
4	M	640	10	D-Fire-plus	No	0.6633	0.6027
5	N	640	10	D-Fire	No	0.6696	0.5946
6	N	640	100	D-Fire	No	0.7746	0.7131
7	N	1280	10	D-Fire	No	0.5900	0.5834
8	N	640	10	D-Fire-plus	No	0.7499	0.7142
9	N	640	10	D-Fire-plus	Yes	0.6684	0.6044
10	N	640	100	D-Fire-plus	Yes	0.7761	0.7031
11	N	640	10	D-Fire-plus	No	0.6684	0.6044
12	N	640	100	D-Fire-plus	No	0.6871	0.6114
13	N	1280	10	D-Fire-plus	No	0.5944	0.5753
14	S	640	100	D-Fire	No	0.7931	0.7295
15	S	1280	100	D-Fire	No	0.7597	0.7338
16	S	640	10	D-Fire-plus	No	0.6871	0.6114
17	S	640	10	D-Fire-plus	No	0.6871	0.6114
18	X	640	10	D-Fire	No	0.6156	0.5732

**TABLE 3** – Results for the models that were considered

quantifies the proportion of true positives among all positive predictions, assessing the model's capability to avoid false positives. On the other hand, Recall calculates the proportion of true positives among all actual positives, measuring the model's ability to detect all instances of a class. These metrics were evaluated considering the samples in the testing set, which are the same for the DFire and DFirePlus datasets, and higher values are better (maximum is 1.0).

For all the models, the values of the Recall are inferior to those of Precision. However, values above 0.7 are obtained consistently. When considering model size, although models S and N are smaller in size, they generally perform better compared to the larger models L, M, and X. This may be related to the fact that models with more parameters need bigger datasets to obtain better results. When examining the impact of image size, it becomes evident that increasing the resolution from 640 to 1280 does not consistently result in improved performance. To make valid comparisons between the two resolutions, additional training epochs are required for the models to reach convergence when larger image sizes are used. The comparison of model performance based on the dataset used for training indicates a clear advantage when utilizing the expanded dataset. Models trained on the D-Fire-plus dataset consistently exhibit superior Precision and Recall, suggesting improved detection capabilities and reduced false positives. Notably, the models trained on the expanded dataset demonstrate better performance in all direct comparisons—where only the dataset is varied while other parameters such as image size, epochs, and model configuration remain unchanged. This highlights the effectiveness of the expanded dataset in enhancing model accuracy and generalization. In terms of data augmentation, it is observed that sometimes it led to slight performance gains, though the improvements were inconsistent. This suggests that while data augmentation can be beneficial, it should be applied with a clear strategy in mind to maximize its effectiveness and avoid unnecessary computational overhead.



## Deployment in a single-board computer

Finally, to enable widespread adoption of the proposed real-time fire detection system, the main idea is that it can run on a single-board computer. We chose to target the Raspberry Pi 5 (RPi5) and benefit from its cost-effective and low-power consumption. On the other hand, the RPi5 provides an affordable and scalable solution that can be exploited in a real scenario. In addition, our proposed approach aims to integrate with other sensors to measure temperature and/or smoke to improve the detection capacity and reduce the false positive numbers.

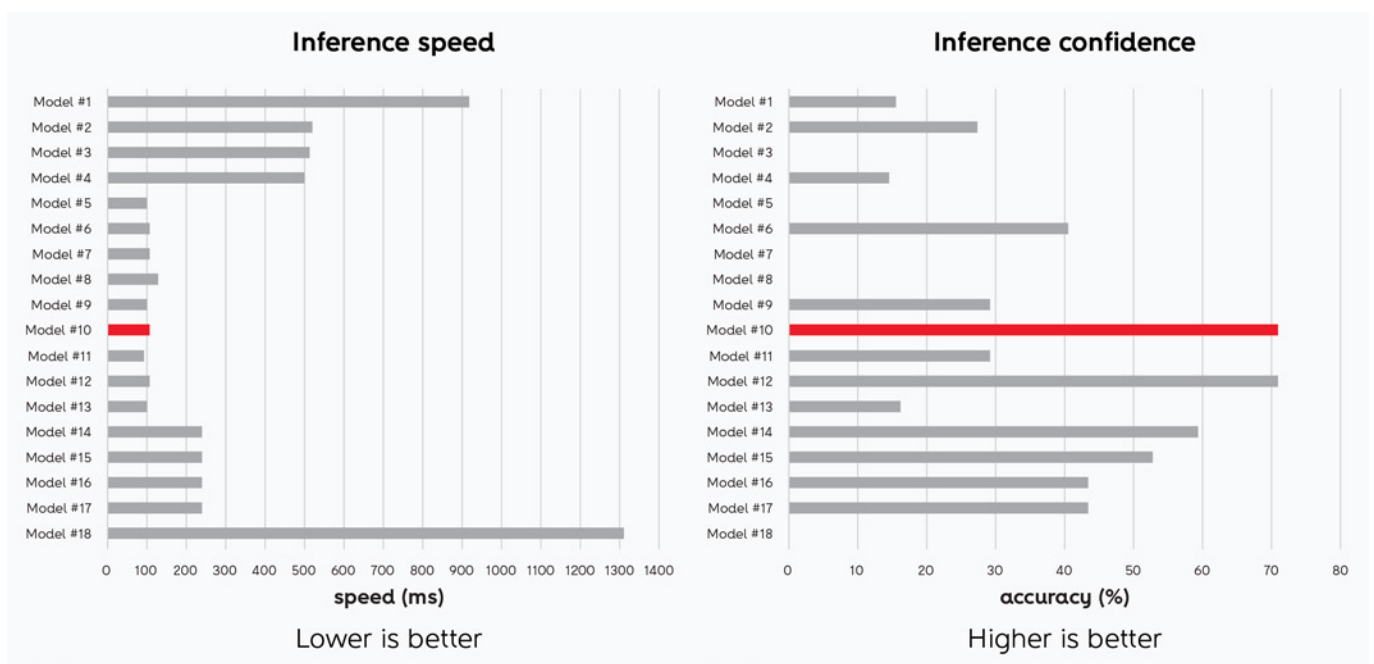
Upon measuring in the RPi5 the performance of the models developed in the previous section, it was clear that they underperformed. This is because these models are not well-suited and not fine-tuned to run on such low-power devices. It was thus necessary to convert our models so that they could be optimized and run on these platforms, and our choice was to use the NCNN framework developed by Tencent<sup>6</sup>. NCNN is known for its fast processing speed on mobile CPUs and was deemed feasible to use on the RPi5.

## Results

**Figure 6** shows the results obtained for different models in terms of inference speed and confidence (accuracy) using the NCNN format on a Raspberry Pi 5. These results were averaged for four tests that were initially selected to measure the performance of the models for different situations where a fire occurred (see **Figure 7**).

The findings show that the nano and small versions of the YOLOv8 model – the only two options capable of classifying an average of 10 and 15 frames per second, respectively – are the most feasible options for a real-time application. The fact that more complex models do not perform significantly better could be attributed to several factors, including the dataset size (larger models perform better on longer datasets) and the number of epochs used to train them, as more parameters require more epochs for convergence.

Regarding image resolution, our tests revealed that, while increased resolution can enhance the model's ability to detect fine details, the improvement in accuracy is often marginal compared to



**FIGURE 6** – Average inference speed and confidence for four tests



FIGURE 7 – Accuracy of the best model (model #10) for the four tests<sup>7</sup>

the substantial increase in computational demand. Therefore, to achieve faster processing without visibly compromising performance, we concluded that choosing a reduced resolution, namely images = 640, reveals the best option.

For the best-performing model (Model #10), **Figure 7** shows the bounding boxes and confidence level attained.

## Conclusions

This work successfully demonstrated the potential of using the YOLOv8 model to develop an effective early-response fire detection system. By expanding the dataset to include more early-stage fire scenarios, we addressed the limitations of existing datasets and enhanced the accuracy of the model in detecting small fires before they escalate. By carefully tuning model parameters, such

as selecting the optimal model size and applying strategic data augmentation, we balanced accuracy and real-time performance.

Moreover, the feasibility of deploying the model on a low-power device, such as the RPi5, was discussed and validated, highlighting its practical use in real-world applications. This combination of an enhanced dataset and a fine-tuned model offers a promising solution for rapid and reliable fire detection, paving the way for safer fire detection systems. 🔗

## Acknowledgements

This work was funded by AlticeLabs@UA, produced with the support of UC-LCA, and funded by FCT I.P. under the project Advanced Computing Project 2023.10774.CPCA.A1, platform Navigator.

## References

- [1] Center of Fire Statistics, "World fire statistics. Report No 29. Center of Fire Statistics of CTIF," International Association of Fire and Rescue Services, 2024.
- [2] J. Pincott, P. W. Tien, S. Wei, and J. K. Calautit, "Indoor fire detection utilizing computer vision-based strategies," *Journal of Building Engineering*, vol. 61, p. 105154, Dec. 2022, doi: 10.1016/j.jobbe.2022.105154.
- [3] A. Gaur et al., "Fire Sensing Technologies: A Review," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3191–3202, May 2019, doi: 10.1109/JSEN.2019.2894665.
- [4] S. Geetha, C. S. Abhishek, and C. S. Akshayanat, "Machine Vision Based Fire Detection Techniques: A Survey," *Fire Technol*, vol. 57, no. 2, pp. 591–623, Mar. 2021, doi: 10.1007/s10694-020-01064-z.
- [5] Z. Zhong, M. Wang, Y. Shi, and W. Gao, "A convolutional neural network-based flame detection method in video sequence," *SIViP*, vol. 12, no. 8, pp. 1619–1627, Nov. 2018, doi: 10.1007/s11760-018-1319-4.
- [6] X. Shi, N. Lu, and Z. Cui, "Smoke detection based on dark channel and convolutional neural networks," in *2019 5<sup>th</sup> international conference on big data and information analytics (BigDIA)*, Kunming, China, 2019, pp. 23–28, doi: 10.1109/BigDIA.2019.8802668.
- [7] Y. Hu and X. Lu, "Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features," *Multimed Tools Appl*, vol. 77, no. 22, pp. 29283–29301, Nov. 2018, doi: 10.1007/s11042-018-5978-5.
- [8] B. Kim and J. Lee, "A Video-Based Fire Detection Using Deep Learning Models," *Applied Sciences*, vol. 9, no. 14, p. 2862, Jul. 2019, doi: 10.3390/app9142862.
- [9] J. Zeng, Z. Lin, C. Qi, X. Zhao, and F. Wang, "An improved object detection method based on deep convolution neural network for smoke detection," in *2018 international conference on machine learning and cybernetics (ICMLC)*, Chengdu, China, 2018, pp. 184–189, doi: 10.1109/ICMLC.2018.8527037.
- [10] F. Yuan, L. Zhang, B. Wan, X. Xia, and J. Shi, "Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition," *Machine Vision and Applications*, vol. 30, no. 2, pp. 345–358, Mar. 2019, doi: 10.1007/s00138-018-0990-3.
- [11] M. Yin, C. Lang, Z. Li, S. Feng, and T. Wang, "Recurrent convolutional network for video-based smoke detection," *Multimed Tools Appl*, vol. 78, no. 1, pp. 237–256, Jan. 2019, doi: 10.1007/s11042-017-5561-5.
- [12] A. Filonenko, L. Kurnianggoro, and K.-H. Jo, "Smoke Detection on Video Sequences Using Convolutional and Recurrent Neural Networks," in *Computational Collective Intelligence*, vol. 10449, N. T. Nguyen, G. A. Papadopoulos, P. Jedrzejowicz, B. Trawinski, and G. Vossen, Eds., in *Lecture Notes in Computer Science*, vol. 10449, Cham: Springer International Publishing, 2017, pp. 558–566, doi: 10.1007/978-3-319-67077-5\_54.
- [13] Z. Xu, W. Wanguo, L. Xinrui, L. Bin, and T. Yuan, "Flame and Smoke Detection in Substation Based on Wavelet Analysis and Convolution Neural Network," in *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*, Suzhou, China: ACM, Mar. 2019, pp. 248–252, doi: 10.1145/3319921.3319962.

- [14] J. Gotthans, T. Gotthans, and R. Marsalek, "Deep Convolutional Neural Network for Fire Detection," in *2020 30<sup>th</sup> International Conference Radioelektronika (RADIOELEKTRONIKA)*, Bratislava, Slovakia, Apr. 2020, pp. 1–6. doi: 10.1109/RADIOELEKTRONIKA49387.2020.9092344.
- [15] J. Redmon, Divvala, Santosh, Girshick, Ross, and Farhadi, Ali, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
- [16] P. V. A. B. De Venâncio, A. C. Lisboa, and A. V. Barbosa, "An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices," *Neural Comput & Applic*, vol. 34, no. 18, pp. 15349–15368, Sep. 2022, doi: 10.1007/s00521-022-07467-z.

---

## Endnotes

- 
- 1 <https://www.alticelabs.com/altice-labs-ua/>
  - 2 <https://github.com/gaiasd/DFireDataset?tab=readme-ov-file>
  - 3 <https://roboflow.com/>
  - 4 [https://github.com/developer0hye/Yolo\\_Label](https://github.com/developer0hye/Yolo_Label)
  - 5 <https://www.uc.pt/lca/computing-resources/navigator-cluster/>
  - 6 <https://github.com/Tencent/ncnn>
  - 7 Top-left: <https://www.youtube.com/watch?v=JGIIciX2CNI>  
Top-right: <https://www.youtube.com/watch?v=JdEI7g2i1ZU>  
Bottom-left: <https://www.youtube.com/watch?v=puy6IUd4IQ0>  
Bottom-right: <https://www.youtube.com/watch?v=fhXyiFI4yvY>

# 13

How LLMs are reshaping customer experience: a vision of enhanced interactions and a brighter future



Generative AI (GenAI) is ushering in a transformative era of artificial intelligence, fundamentally redefining how machines engage with human language and adapt to complex tasks. While early systems relied on basic automation, today's AI models have transcended these limitations, utilizing data-driven insights to deliver more responsive, human-centered interactions. The breakthrough in Natural Language Processing (NLP) has enabled AI not only to understand but also to anticipate the needs and preferences of users, laying the groundwork for the advanced architectures that now drive the field.

GenAI models leverage powerful Retrieval-Augmented Generation (RAG) frameworks and agent architectures to deliver sophisticated, context-rich responses. RAG combines generative capabilities with real-time information retrieval, enabling systems to draw from vast knowledge bases to provide accurate, relevant answers. When integrated into agent architectures, these models can autonomously handle complex tasks, analyze information, and interact intuitively with users, facilitating more natural, meaningful engagement.

As these advancements become integral to daily life, GenAI's impact on society and business is profound. It promises new efficiencies, innovations, and ethical considerations, transforming industries while fostering deeper human-machine connections. Looking ahead, GenAI paves the way toward hyper-personalized experiences and seamless customer journeys, setting the stage for a future where intelligent, adaptive systems help us live, work, and connect in unprecedented ways. With its rapid evolution, GenAI stands as a catalyst for a brighter, more connected future, unlocking possibilities we are only beginning to imagine.

---

## Authors

**Jorge Miguel Sousa, Altice Labs, [jorge-m-sousa@alticelabs.com](mailto:jorge-m-sousa@alticelabs.com)**

**Pedro Nuno Rocha, Altice Labs, [pedro-n-rocha@alticelabs.com](mailto:pedro-n-rocha@alticelabs.com)**

---

## Keywords

GenAI; RAG; Agents; NLP; Personalization; Automation; Innovation; Ethics



## From rule-based to data-driven: the dawn of intelligent customer interactions

In the early days of customer interaction management, businesses relied heavily on rule-based systems [1] designed to automate simple tasks. These systems, such as Interactive Voice Response (IVR) [2] and basic scripted chatbots, stood for a significant technological advancement at the time. They allowed for automated responses to common customer inquiries, yet they were inherently limited by their reliance on rigid, pre-defined inputs. These early technologies struggled to accommodate the complexities of human language and the varied needs of customers, often leading to impersonal and frustratingly linear interactions.

Technically, these rule-based systems were built on a foundation of hard-coded decision trees and regular expressions (regex) [3] to process input. They relied on pattern-matching mechanisms to interpret user queries but were constrained by their inability to handle unstructured or ambiguous input. For example, a user interacting with a regex-based system needed to provide a query in an exact

format, as any deviation in wording would likely fail to match the pre-defined patterns. The decision-making logic behind these systems followed a strict if-then-else structure [4], where each possible input had a corresponding, fixed response or action.

This reliance on static rules meant these systems could only handle predictable scenarios. Complex queries or conversations, which required an understanding of context, were beyond the scope of what these systems could process. Even in cases where businesses tried to expand functionality, the complexity of managing and maintaining extensive branching logic in the scripts made the system prone to errors and inconsistencies. For instance, even adding a single new option required manually updating multiple parts of the system, increasing both maintenance costs and the likelihood of system bugs.

Furthermore, interactions with these systems were often highly structured, forcing users into a narrow set of choices. Customers often had to navigate multi-layered menus or input exact phrases to achieve their goals. This lack of flexibility in input interpretation and the rigid decision trees contributed to poor user experience, especially as customers began to expect more natural, conversational interfaces. The systems could not retain context across multiple turns in a conversation, resulting in disjointed and frustrating interactions, especially when compared to modern AI capabilities.

The dawn of Machine Learning (ML) and NLP [5] brought a revolutionary shift away from these rigid, rule-based systems. Unlike traditional scripted bots, ML-driven systems could learn from data, allowing them to interpret more diverse inputs, understand unstructured text, and make sense of queries that would have broken previous systems. For example, early Intent Classifiers [6] and Named Entity Recognition (NER) [7] models allowed systems to grasp the user's intent and extract key information, even when the input was not perfectly structured. This new capability laid the groundwork for the highly adaptive AI systems we see today, which are capable of real-time context retention and dynamic conversation management.

By integrating data-driven models instead of static rules, these systems were not only more robust but could also evolve over time. They learned from previous interactions and became more accurate and efficient, eliminating the need for manually updated scripts. This evolution also meant that AI-driven platforms could handle more complex customer journeys, offering flexibility, scalability, and the ability to personalize responses, which the rigid rule-based systems of the past could never achieve.

## The NLP Revolution: Understanding and Responding to the Voice of the Customer

In the preliminary stages of NLP, the predominant methods for customer interaction relied heavily on supervised learning [8] approaches. Models for intent classification were among the first to enable automation in customer service tasks. These models, which classified user utterances into pre-defined categories (intents), were typically based on Support Vector Machines (SVMs) [9] or logistic regression [10], leveraging carefully curated datasets of labeled customer inputs. These early classifiers used hand-crafted features such

as word frequencies and other linguistic properties to map customer input to specific intents like "billing inquiry" or "technical support." While functional, these models were constrained by their reliance on extensive feature engineering, making them labor-intensive and limited in scalability.

Alongside intent classification, NER systems emerged to extract key information from customer queries, identifying entities such as names, locations, dates, and product terms. Early NER models relied on algorithms like Conditional Random Fields (CRFs), which performed well for structured data extraction but required significant manual feature engineering and could struggle with complex, unstructured inputs. The advent of neural networks [11] marked a significant leap in this area. With their ability to capture sequential dependencies in text, Long Short-Term Memory (LSTM) [12] and Bi-LSTM [13] models improved the performance of both intent classification and NER tasks, allowing systems to understand context across a sentence or even multiple sentences. For example, Bi-LSTMs were particularly effective in NER as they could consider both the past and future words in a sequence to recognize entities more accurately.

However, early neural approaches, including those based on tokenization and lemmatization (reducing words to their base or root form), still faced challenges in understanding word relationships. These approaches, though improving model performance, did not fully address the need for richer contextual representations of language.

The next significant advancement came with the development of word embeddings [14] — a method for representing words in continuous vector space, which transformed NLP capabilities. Techniques like Word2Vec [15] and GloVe [16] captured the semantic relationships between words by mapping words with similar meanings to nearby points in vector space. This allowed systems to understand not only individual words but also their relationships in a much more meaningful way. For instance, Word2Vec's famous example of encoding relationships like "king - man + woman = queen" illustrated how these embeddings could capture subtle semantic and syntactic information that earlier bag-of-words models could not.

These embeddings significantly improved tasks like intent classification, where the AI could now recognize related words or phrases even if they were not explicitly included in the training data. Word embeddings also enhanced closed-dialogue systems commonly used in enterprise applications, where precise interpretation of technical language was crucial. For instance, an AI system equipped with word embeddings could more effectively understand technical jargon during a troubleshooting interaction, leading to faster and more accurate resolutions.

Despite these advancements, word embeddings like Word2Vec were still limited in their ability to fully understand context across longer texts. They represented each word independently, without accounting for the multiple meanings that a word could take depending on context. To address this, newer models like FastText [17] and StarSpace [18] expanded on word embeddings by incorporating sub-word information, helping the system deal with rare or out-of-vocabulary words. However, it was not until the development of transformer architectures that NLP truly revolutionized.

In spite of the improvements in word embeddings, models like Word2Vec and GloVe were still limited to word-level understanding, which could miss the broader context of entire sentences or phrases. This limitation led to the development of sentence-level embeddings, particularly with the introduction of the Universal Sentence Encoder (USE) [19].

USE was designed to generate embeddings for entire sentences, providing a richer, more nuanced understanding of language. Rather than relying on individual word relationships, USE creates a fixed-length vector representation for a sentence, capturing its overall meaning. This proved especially valuable for tasks like semantic similarity and paraphrase detection, where understanding the broader context of a message is more important than just the individual words. USE was built on either Deep Averaging Networks (DANs) [20] or transformer architectures [21] and trained on a wide variety of data, making it more flexible for different tasks than traditional word embeddings.

While USE marked a significant advancement by moving beyond word-level understanding, it was still limited in capturing contextual dependencies across long text sequences. For example, USE could understand a single sentence well but could not fully grasp how that sentence fitted into a larger conversation or multi-turn dialogue. This challenge paved the way for the development of even more sophisticated models, such as Bidirectional Encoder Representations from Transformers (BERT) [22], which would address the need for deeper contextual understanding — setting the stage for the next chapter in NLP's evolution.

## The rise of GenAI: unlocking the power of contextual understanding and generation

The introduction of BERT in 2018 transformed NLP using a bidirectional approach to context. This innovation allowed BERT to capture both preceding and following word sequences in a sentence, enabling more nuanced language understanding. While BERT excelled in tasks such as text classification, question answering, and named entity recognition, it was primarily focused on understanding and encoding text rather than generating it.

Building on the Transformer Architecture that underpinned BERT, OpenAI introduced the GPT (Generative Pre-trained Transformer) models. While BERT was a significant advancement in language understanding, GPT [23] represented the leap toward language generation, a crucial step in creating GenAI systems. Unlike BERT's bidirectional nature, GPT utilized a unidirectional architecture, which means it processed text sequentially from left to right. However, GPT's transformative capability to generate human-like text on demand was directly inspired by the core principles of the transformer architecture established by BERT.



One of the critical advancements in GPT technology was integrating Reinforcement Learning from Human Feedback (RLHF) [24]. This technique allows the model to be fine-tuned through human interaction, improving its ability to generate responses that align with human preferences. RLHF was crucial in making GPT's outputs more coherent, contextually appropriate, and reflective of real-world use cases, distinguishing it from traditional AI models that were purely rule-based or limited to supervised learning.

While BERT emerged in 2018, Large Language Models (LLMs) like GPT-3 only became widely available to the public in 2020, marking a significant milestone in the deployment of GenAI. It was GPT-3, released by OpenAI, that captured public attention with its remarkable ability to generate human-like text across a wide range of tasks, from content creation to code generation. The availability of such LLMs to developers and businesses for practical applications ushered in the broader adoption of GenAI technology. By 2022, LLMs [25] were further refined and integrated into mainstream products, allowing companies across various industries to

deploy these models at scale for customer interactions, content generation, and other applications.

Generative AI, powered by GPT-style models, unlocked new possibilities for real-time, dynamic customer experiences. Unlike earlier AI systems that relied on predefined scripts or rigid decision trees, GPT-3 and similar models can understand and generate content in response to complex customer queries, providing real-time solutions that feel more natural and personalized. This transformation was accelerated by the rise of RAG [26] architectures, which added a layer of fact-checking and domain-specific knowledge retrieval, ensuring that the generative responses remain accurate and contextually relevant. RAG architectures play a vital role as they combine generative capabilities with the precision of information retrieval, grounding the generative process in concrete, domain-specific knowledge. By retrieving relevant information from a knowledge base or a curated set of documents, RAG ensures that the generated content is not only coherent and engaging but also factually correct and aligned with the customer's specific context.



This is especially important in telecommunications, where customer interactions often involve complex technical details, service plans, and account-specific information. RAG empowers AI systems to access and use this information in real time, providing customers with accurate and personalized responses to their queries. The continuous evolution of its architectures is steadily improving the precision and reliability of GenAI-powered systems, making them increasingly invaluable for enhancing customer experience.

In this paradigm, RAG and LLMs coalesce to create AI systems that are not only more intelligent but also more adaptable to real-world, personalized customer interactions. For businesses like Altice Labs with platforms such as BOTSchool, this transition to GenAI has opened doors to sophisticated, context-aware customer engagement systems that respond to individual users' unique needs and preferences, ultimately driving higher satisfaction and operational efficiency of its clients.

## Agent architectures: automating complexity and empowering human connection

Autonomous Agents [27], fueled by the advancements in GenAI and RAG, are revolutionizing customer service. These intelligent Agents are not simply replacing Human Agents; they are augmenting their capabilities, automating complex workflows, handling massive datasets, and interacting seamlessly with humans. This allows Human Agents to focus on higher-value interactions that require empathy, creativity, and strategic thinking, ultimately fostering more profound customer connections. Ultimately, this shift is about enhancing human connection, not ending it.

One of the critical advantages of autonomous Agents is their ability to handle the scale and complexity

of modern customer service operations. They can process vast amounts of data, access multiple systems of record simultaneously, and personalize interactions based on individual customer profiles and history. In the telecommunications sector, for instance, these agents can diagnose technical issues, troubleshoot service disruptions, and even proactively identify and address potential problems before they affect the customer [28]. This proactive approach improves efficiency, reduces resolution times, and enhances the customer experience by providing personalized and timely support.

Furthermore, Agent architectures are continuously evolving to incorporate more sophisticated decision-making capabilities. By integrating with business logic and workflow management systems, these agents can automate end-to-end processes, such as onboarding new customers, processing service upgrades, and handling billing inquiries. This automation reduces operational costs and ensures consistency and compliance with internal policies and regulations. For example, an autonomous agent can automatically verify a customer's identity, check their eligibility for a service upgrade, and process the request without human intervention, ensuring a seamless and efficient experience. Altice Labs, with its BOTSchool [29] platform, enables businesses to build these sophisticated agent architectures. However, the development and deployment of autonomous agents require careful consideration of ethical implications and the importance of human oversight. Building customer trust is paramount, and businesses must ensure that these agents operate transparently and ethically, respecting customer privacy and data security.

## The implications on society and business

The rise of GenAI is not just transforming individual industries — it is reshaping the fabric of society and business globally. With the ability to process vast amounts of data, generate human-like responses, and automate decision-making, GenAI can

revolutionize how organizations work, dramatically improving efficiency and customer experiences while also challenging long-standing norms in the workforce and decision-making processes.

One of the most immediate societal impacts is its ability to augment human capabilities. In industries like healthcare, finance, telecommunications, and customer service, GenAI systems are increasingly being used to handle tasks that previously required human intervention. This shift is improving operational efficiency while enabling organizations to deliver more personalized, rapid, and accurate customer services.

Beyond operational improvements, GenAI is reshaping decision-making in ways that fundamentally change humans' role in business processes. In sectors such as finance and healthcare, AI systems can now analyze vast datasets to provide insights that would take humans days or weeks to uncover. This data-driven approach augments human decision-making, allowing faster, more informed choices leading to better outcomes. In business operations, AI can refine everything from supply chains to customer engagement strategies, identifying patterns and trends that would be invisible to human analysts.

However, the broader adoption of these technologies raises essential questions about the future of work and the role of humans in increasingly automated industries. While it can take over many routine, mundane tasks, this evolution demands that businesses rethink their workforce strategies. Employees will need to shift their focus toward roles that require more advanced skills, such as managing AI systems, making strategic decisions, and ensuring that AI runs within ethical and compliance boundaries. This shift presents challenges and opportunities, as businesses must invest in upskilling their workforce while leveraging its efficiencies.

Regarding societal impact, the integration of GenAI into business operations also influences how people interact with technology. The rise of AI-powered customer service agents, for example, is reshaping the nature of customer support, with intelligent systems, rather than humans, increasingly handling conversations. This shift is creating

a more immediate and responsive customer experience, as AI systems can deliver correct answers and perform actions in real time. However, it also raises concerns about the depersonalization of customer interactions and the potential reduction of human employment in specific sectors.

Despite these concerns, the adoption is accelerating across industries, driven by the need for increased efficiency, cost reduction, and enhanced customer experiences. Companies today rapidly integrate these technologies with customer information systems to streamline operations and improve service delivery.

In conclusion, the implications of GenAI on society and business are vast and far-reaching. By improving operational efficiency, enhancing decision-making, and reshaping customer interactions, GenAI is driving a new era of automation and innovation. As companies continue to integrate these technologies with their customer information systems, the future of business will likely involve a greater reliance on AI to augment human capabilities and transform how organizations deliver value to their customers. However, this also needs careful consideration of the broader societal impacts, ensuring that the evolution of AI aligns with ethical practices and human-centered values.

## The future: a vision of hyper-personalization and seamless customer journeys

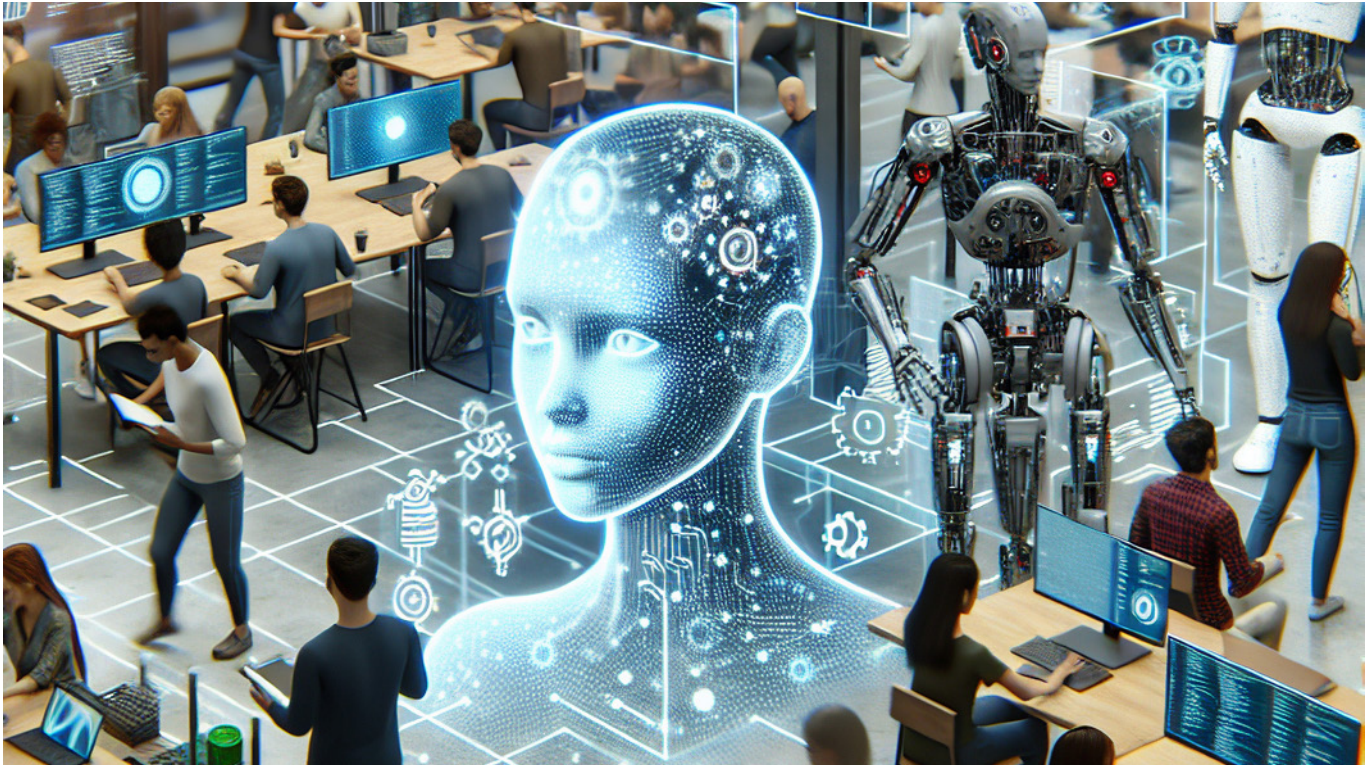
The future is promising, offering a world of hyper-personalized experiences and seamless customer journeys. As technologies like RAG architectures, in-context learning, and Agent-based systems continue to advance, AI will become even more deeply integrated into the fabric of our lives, predicting our needs and proactively providing solutions before problems arise. Imagine a future where your digital assistant understands your preferences and predicts your needs, seamlessly

managing your tasks and refining your experiences across various touchpoints.

One of the most exciting prospects is the development of fully autonomous AI systems that act as trusted advisors. These systems will go beyond simple question-and-answer interactions, offering proactive guidance, anticipating potential roadblocks, and delivering tailored solutions in real time. This level of personalized and proactive service will redefine customer expectations and create entirely new possibilities for customer engagement.

Another promising trend is the increasing personalization of customer interactions. GenAI will enable businesses to tailor every touchpoint to the individual customer, delivering customized content, offers, and support based on unique preferences, past behavior, and real-time context. Unlike traditional recommendation systems, which rely primarily on historical data and often lack real-time adaptability, GenAI brings a deeper contextual understanding, interpreting live behavioral cues, external factors, and personal preferences to deliver precisely timed, relevant recommendations. Imagine receiving not only product suggestions aligned with your interests but also dynamically generated content that speaks directly to your needs. Thanks to its advanced pattern recognition and anticipation capabilities, GenAI can proactively predict and address requirements before customers realize they have them. By continuously learning from interactions and adapting on the fly, GenAI does not merely recommend — it engages in a natural, conversational way, refining suggestions through real-time feedback to ensure they remain accurate and satisfying. This granular level of personalization will enhance customer satisfaction and foster more robust, meaningful relationships between businesses and their customers.

However, realizing this vision requires careful consideration of the ethical implications. As AI systems become more autonomous and integrated into our lives, ensuring they operate transparently and ethically, respecting customer privacy and data security, is paramount. Building and keeping customer trust will be crucial, and businesses must prioritize responsible AI practices to ensure these powerful technologies are used for society's benefit.



## Conclusion: GenAI – a catalyst for a brighter future

The integration of GenAI into customer information systems signifies more than just a technological advancement; it stands for a profound shift in how businesses work and societies interact. This is not merely an evolution but a revolution, reshaping customer experience, driving innovation across industries, and unlocking new opportunities for human connection. GenAI is not simply augmenting human capabilities; it is acting as a catalyst for a brighter future where technology empowers us to connect, communicate, and collaborate more meaningfully.

The journey from rigid, rule-based systems to the dynamic, context-sensitive AI we have today has been remarkable. We have witnessed how advancements in NLP, the rise of GenAI, and the development of autonomous Agents have transformed customer interactions, leading to more

personalized, efficient, and seamless experiences. The telecommunications sector, as explored throughout this article, serves as a prime example of how GenAI is revolutionizing customer service, automating complex processes, and delivering hyper-personalized support. Companies like Altice Labs, with its BOTSchool platform, and its clients like MEO are at the forefront of this transformation.

However, the true potential of GenAI extends beyond its technological capabilities. Its power lies in its ability to empower human connection. By automating routine tasks and freeing up human workers to focus on higher-value activities, GenAI allows us to dedicate our time and energy to what truly matters: building relationships, fostering creativity, and solving complex problems together. As we move forward, embracing responsible AI practices is crucial. We must ensure that these powerful technologies are used ethically and transparently, focusing on enhancing human potential and promoting a positive societal impact. The future is bright, and by working together, we can harness its transformative power to create a better future for all. 🌐

## References

- [1] C. Grosan and A. Abraham, "Rule-Based expert systems," in *Intelligent systems reference library*, vol.17, Springer, Berlin, Heidelberg, 2011, pp. 149–185, doi: 10.1007/978-3-642-21004-4\_7.
- [2] R. Corkrey and L. Parkinson, "Interactive voice response: review of studies 1989-2000," *Behav. Res. Methods Instrum. Comput.*, vol. 34, no. 3, pp. 342–353, Aug. 2002, doi: 10.3758/bf03195462.
- [3] A. V. Aho and J. D. Ullman, "Patterns, Automata, and Regular Expressions," in *Foundations of Computer Science*, W. H. Freeman, 1994, pp. 529–590.
- [4] P. Fodor, "Dialog Management for Decision Processes," in *Proc. 3<sup>rd</sup> Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, Poland, 2007, pp. 1-4. [Online]. Available: <https://arxiv.org/pdf/cs/0703072>.
- [5] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, Aug. 2011, doi: 10.1136/amiajnl-2011-000464.
- [6] K. Balodis and D. Deksnė, "FastText-Based intent detection for inflected languages," *Information*, vol. 10, no. 5, p. 161, May 2019, doi: 10.3390/info10050161.
- [7] N. Patil, A. Patil, and B. V. Pawar, "Named Entity Recognition using Conditional Random Fields," *Procedia Computer Science*, vol. 167, pp. 1181–1188, Jan. 2020, doi: 10.1016/j.procs.2020.03.431.
- [8] T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," in *Springer series in statistics*, New York, NY, U.S.: Springer, 2008, pp. 9–41. doi: 10.1007/978-0-387-84858-7\_2.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
- [10] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395–2399, May 2008, doi: 10.1161/circulationaha.106.682658.
- [11] S. Bozinovski, "Reminder of the first paper on transfer learning in Neural Networks, 1976," *Informatica*, vol. 44, no. 3, Sep. 2020, doi: 10.31449/inf.v44i3.2828.
- [12] J. Schmidhuber, "Long Short-Term Memory: 2003 Tutorial on LSTM Recurrent Nets," 2003, Available: <https://people.idsia.ch/~juergen/lstm2003tutorial.pdf>.
- [13] S. Ghosh, A. Ekbal, and P. Bhattacharyya, "Natural language processing and sentiment analysis: perspectives from computational intelligence," in *Computational Intelligence Applications for Text and Sentiment Data Analysis*, 2023, pp. 17–47. doi: 10.1016/b978-0-32-390535-0.00007-0.
- [14] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *Proc. 32<sup>nd</sup> Int. Conf. Mach. Learn.*, vol. 37, pp. 957–966, 2015.
- [15] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv (Cornell University)*, Sept. 2013, doi: 10.48550/arxiv.1301.3781.

- [16] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543.
- [17] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *arXiv (Cornell University)*, June 2016, doi: 10.48550/arxiv.1607.04606.
- [18] L. Wu, A. Fisch, S. Chopra, K. Adams, A. Bordes, and J. Weston, "StarSpace: Embed all the things!," *arXiv (Cornell University)*, Nov. 2017, doi: 10.48550/arxiv.1709.03856.
- [19] D. Cer et al., "Universal Sentence Encoder," *arXiv (Cornell University)*, April 2018, doi: 10.48550/arxiv.1803.11175.
- [20] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé III, "Deep unordered composition rivals syntactic methods for text classification," in *Proc. 53<sup>rd</sup> Annu. Meet. Assoc. Comput. Linguistics and 7<sup>th</sup> Int. Joint Conf. Nat. Lang. Process. (Volume 1: Long Papers)*, Beijing, China, 2015, pp. 1681–1691.
- [21] A. Vaswani et al., "Attention is all you need," *arXiv (Cornell University)*, Aug. 2023, doi: 10.48550/arxiv.1706.03762.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv (Cornell University)*, May 2019, doi: 10.48550/arxiv.1810.04805.
- [23] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, 2019.
- [24] L. Ouyang et al., "Training language models to follow instructions with human feedback," *arXiv (Cornell University)*, Mar. 2022, doi: 10.48550/arxiv.2203.02155.
- [25] OpenAI, "Introducing ChatGPT," Openai.com. Accessed: Nov. 12, 2024. [Online]. Available: <https://openai.com/index/chatgpt/>.
- [26] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks," *arXiv (Cornell University)*, Apr. 2021, doi: 10.48550/arxiv.2005.11401.
- [27] L. Wang et al., "A Survey on Large Language Model based Autonomous Agents," *arXiv (Cornell University)*, Apr. 2024, doi: 10.48550/arxiv.2308.11432.
- [28] "Improving IVR interactions through Conversational AI Platform," botschool.ai. Accessed: Nov. 12, 2024. [Online.] Available: <https://www.botschool.ai/contact-center-offload/>.
- [29] "BotSchool AI," botschool.ai. Accessed: Nov. 12, 2024. [Online]. Available: <https://www.botschool.ai/>.

# 14

## Smart energy tools



This paper explores the impact of dynamic electricity management on energy efficiency in Portugal, with a particular focus on the technologies and software tools that can be developed to optimize the use of electric vehicle batteries, energy communities, and peer-to-peer (P2P) applications for buying and selling energy. The paper also addresses the bidirectional charging functionality of electric vehicles, which allows for both charging and discharging energy back to the grid, contributing to the stability of the electrical system.

---

## Authors

**Higino Vieira, Altice Labs,** [higino.v.vieira@altice.pt](mailto:higino.v.vieira@altice.pt)

**Nuno Honório, Altice Labs,** [nuno-m-honorio@alticelabs.com](mailto:nuno-m-honorio@alticelabs.com)

---

## Keywords

Efficiency; Pricing; Optimization; Consumption; Electricity; Sustainability; Charging; Community; Renewable; Simulation



## Introduction

Throughout the centuries, energy has played a fundamental role in the development of societies, being essential for the advancement of various economic sectors. Electricity plays an important role in creating a more sustainable planet, facilitating the shift towards cleaner energy and improving energy efficiency. It serves as an effective energy source that can be produced from renewable sources, like solar and wind power, to reduce greenhouse gas emissions and decrease reliance on fossil fuels [1]. This transition is vital to address climate change and reduce our impact on the environment [2]. Moreover, electricity plays a role in driving progress in technology and infrastructure, enabling the emergence of smart grids, electric vehicles, and energy-efficient appliances [3]. These advancements not only help in lowering energy usage, but also contribute to economic prosperity and enhance overall well-being.

The electricity demand has been rising consistently over the years due to various interconnected reasons. Urbanization and industrial growth in developed nations have contributed to increased energy usage, as more individuals get electricity access and industries grow. Additionally, the widespread use of technologies and the human increasing dependence on electronic devices, like smartphones and data centers, are also driving up the demand for electricity. Furthermore, the move towards using electricity in transportation and heating systems opens new opportunities for decreasing carbon emissions, while also leading to higher electricity consumption [2].

Consumer behavior is a key factor in how electricity is used and in promoting the shift to more sustainable energy practices [4]. By embracing energy-efficient habits and technologies, consumers can reduce power usage and inherent costs, while also lessening the need for extensive energy production efforts in the long run. As a result, the electricity market is changing rapidly, reshaping how consumers make decisions by offering tools to promote energy-efficient usage, such as Dynamic pricing.

Dynamic energy pricing will help consumers make decisions regarding their energy consumption in accordance with hourly wholesale market prices. On the other hand, this feature helps to regulate and balance the electrical system, which is becoming increasingly difficult to manage. Also, the use of more and more renewable sources, which can be quite unpredictable as they directly depend on weather conditions, sets a challenge in the management of the electrical system. This inconsistency poses difficulties in sustaining a reliable energy provision and requires the use of advanced strategies for regulating the equilibrium between energy supply and demand. To manage these changes effectively and stabilize the situation, energy storage solutions like batteries also play an important role. A consumer or producer with batteries can store surplus energy created during peak renewable energy production and release it during times of supply or high demand. Although batteries offer benefits, their widespread adoption is stalled by challenges such as high costs, restricted storage capabilities, and a gradual decline in performance over time. The challenges involved in incorporating batteries into the power grid reinforce the importance of advancements in energy storage technology to maintain a dependable electricity supply.

This article explores strategies, software tools, and technologies for electricity management and efficiency without neglecting the analysis of the impacts on the grid, electricity production, and overall electrical system. With the correct strategy and by facilitating the work of the electrical system logistics operator, consumers themselves will help stabilize and lower electricity production management prices.

## Supporting strategies for dynamic electricity management

This section presents concepts that help to improve the effectiveness of managing electricity, as our world becomes more reliant on it while facing the

challenge of effectively balancing energy needs and supply. In other words, electricity management is an essential element, not only for the consumer, who will get advantages in electricity price, but also for the grid operator, which has the responsibility of maintaining the system balanced. By using smart tactics to manage energy consumption dynamically, electricity usage efficiency can be enhanced, expenses decreased, and our impact on the environment reduced. The subjects covered in this chapter will delve into ways to boost energy efficiency, encourage sustainability efforts, and strengthen the stability of the grid for a more resilient system [5].

Electricity price is crucial to stimulate electricity management, not only for consumers, but also for system operators, producers, and utility companies. This means that electricity price is probably the most important aspect that affects the way consumers use electricity. Considering that electricity prices are variable throughout the day, if consumers are aware of those variations and if they have the right tools, they will make smart decisions and reduce their expenses. By making these decisions, consumers will impact the electricity system (e.g., grid, production) as their consumption habits change, and the system continues to need balance. However, these changes are not the only aspect that impacts the system. In fact, several other factors contribute to increasing complexity in electricity management, such as the weather conditions, the renewal sources' unpredictability, and the higher demand for electric vehicles.

As explained above, electricity price is a major factor in influencing consumption. In that sense, there are already systems that adapt tariff prices based on supply and demand throughout the day, a concept that is known as "Dynamic Pricing". This approach offers benefits, such as improving energy efficiency and promoting the use of green energy sources. In Portugal, the adoption of this strategy is revolutionizing how people control their energy usage, thanks to dedicated technology and specialized software tools.

Dynamic pricing encompasses strategies like Time of Use (TOU) pricing, Real-Time pricing, and

Critical Peak pricing. Each model presents incentives and challenges for both consumers and utility companies. For instance, TOU pricing encourages consumers to use electricity during off-peak hours, while Real-Time pricing adjusts prices based on market conditions. TOU pricing segments the day into periods with different rates for each period – including peak, mid-peak, and off-peak times. On the other hand, Real-Time pricing reflects the wholesale cost of electricity at any given moment, leading to fluctuations throughout the day. Finally, Critical Peak pricing involves high rates during times of demand, often accompanied by alerts to encourage reduced consumption.

Some examples of tactics to optimize energy usage are described below:

1. **Energy Management Systems (EMS):** these systems enable consumers to continuously monitor and regulate their energy usage in real time. They use dynamically managed data to optimize energy consumption based on pricing. EMS platforms also offer insights into energy consumption, allowing users to make quick adjustments. These solutions are essential for adapting consumption to pricing models and for optimizing energy costs. Advanced features of EMS platforms include analytics that forecast energy usage and expenses based on historical data. These systems can anticipate future energy demands by analyzing historical data and planned activities, thereby optimizing, in real time, energy usage accordingly. Machine learning algorithms can also be used to improve the accuracy of these forecasts by recognizing patterns and trends in energy usage [6];
2. **Bidirectional Charging Platforms:** as explained above, the use of batteries to store electricity is stalled by challenges such as high costs, among other difficulties. Vehicle to Grid (V2G) technologies enable vehicles not only to charge, but also to supply energy back to the grid, which means that vehicle batteries may be used to store energy for other purposes than the vehicle itself. These platforms manage the flow of energy, assisting in balancing grid loads and providing support during demand

periods. When necessary, the vehicle can give back energy to the grid, helping to lower demand and stabilize the grid. The integration process involves both hardware, like chargers, and software, such as energy management systems. This technique also benefits the consumer by lowering expenses. Research indicates that V2G can cut electricity expenses by leveraging off-peak rates and offering grid services during high-demand periods. Consumers can also earn money by joining demand response programs;

3. **Smart Energy Communities:** to complement individual energy production, such as solar panels, this strategy aims to unite multiple individual energy producers and consumers within close proximity to households (neighborhood). Energy communities are a form of aggregated self-consumption, allowing neighborhoods to share energy locally before drawing from the public grid. Also known as collective self-consumption, these communities enable greater investment in renewable energy and improve access to renewable sources. By integrating Internet of Things (IoT) systems, they facilitate the sharing and management of energy resources among members. IoT devices, such as meters and sensors, provide real-time data and control for more efficient energy management, quickly identifying system issues to enhance reliability [7].

## Software tools and technology for energy optimization

This section explores methods and tools that turn these strategies into solutions for consumers and network operators. With tools and creativity, there are solutions that enable the control of energy needs, reduce costs, and have more sustainable and intelligent energy consumption.

### 1. Energy Consumption Simulation Software:

these software tools enable the modeling and simulation of energy consumption scenarios, assisting consumers and energy managers in forecasting and optimizing energy usage based on fluctuating tariffs and prices. Algorithms, like Monte Carlo simulations and agent-based modeling, are commonly employed. These programs consider factors such as weather patterns, past usage data, and planned activities for accurate forecasts. Real-world instances of these tools showcase how simulation software has been effectively utilized to enhance energy efficiency. For instance, a case study demonstrates how a manufacturing facility leveraged simulation software to significantly reduce its energy expenses, optimizing through detailed analysis and forecasting tools [8]. The detailed analysis covers the actions taken, the obstacles faced, and the ultimate results.

### 2. Electric Vehicle (EV) Charging Management Systems:

dedicated software solutions are available for managing electric vehicle charging stations. These systems optimize charging, but also discharging schedules to capitalize on tariffs. An examination of charging strategies and their financial implications is crucial. Strategies like scheduled charging demand response and V2G integration can significantly influence costs and grid stability. Scheduled charging involves setting times for charging based on tariff schedules, while demand response adjusts charging times according to real-time grid conditions. The discussion also explores the infrastructure requirements for adopting EV charging management systems. This encompasses communication networks, standardized protocols, optimal placement of charging stations, local grid capacity considerations, as well as integration with renewable energy sources. Several sites in Portugal are currently running trial initiatives to test out the combination of V2G technologies with management. These initiatives showcase how electric vehicles can serve as mobile energy storage units, feeding energy back into the grid during high-demand periods.



**3. Peer-to-Peer Energy Trading Apps:** these apps empower users to trade energy between themselves, considering private producers and communities. They use contracts between consumers to ensure efficient transactions, adjusting to price fluctuations in dynamic management. These smart contracts can execute transactions automatically when specific conditions are met, reducing the reliance on intermediaries. For instance, a smart contract could autonomously sell surplus energy from one user to another, under certain conditions. An easy-to-use interface is vital for the acceptance of these platforms, and some features may include real-time pricing updates, transaction history, and usage insights [7].

## Implementation examples

Let us take the example of a Consumer who has a contract with their energy operator at a fixed rate. This is the reality of the majority of the population, especially the domestic one. On average, in Portugal, a domestic fixed electricity contract costs €90/MWh (€0.09/kWh) in the regulated market, excluding network access fees and taxes. This Consumer has an annual electricity consumption of around 4.000 kWh per year, which is the average consumption of a household. Having a fixed tariff, this Consumer will spend around €360 per year on electricity, excluding fees and taxes.

Suppose that this Consumer chooses a dynamic tariff plan, meaning that electricity's price will vary throughout the day.

The graphic below represents the Iberian wholesale market (OMIE - <https://www.omie.es/>) on a day when there are enough renewable sources to cover needs during solar hours (lower price during solar hours).

By analyzing electricity price variation during the day, it can be observed that there is a significant

range in wholesale price. At the cheapest period, electricity costs €20.50/MWh, and at the most expensive period, it costs €157.66/MWh (Figure 1).



FIGURE 1 – OMIE market data for 09/23/2024

Allow us to assume that the operator supplying electricity to this Consumer (or the Consumer itself) is concerned about its energy efficiency and suggests switching to the dynamic tariff, by providing a forecasting software for the consumer. Only with this consumption forecasting software, and by understanding the market, the consumer will strive to place their energy needs at the cheapest period of the day, which in this case would be between 12 pm and 6 pm. If they are successful, they will be able to consume a large part of their energy needs at around €25/MWh.

In case a fixed tariff is used, the Consumer would have to pay:

- Fixed tariff for 100% of consumption: €0.09/kWh × 20 kWh = €1.8/day

If the consumption is around 20 kWh per day, in winter, and 80% of usage is during the cheapest period, then by using a Dynamic tariff, the Consumer would have to pay:

- 80% of consumption at low price: €0.025/kWh × 20 kWh × 80% = €0.4/day

- 20% of consumption at high price:  $\text{€}0.157/\text{kWh} \times 20 \text{ kWh} \times 20\% = \text{€}0.628/\text{day}$

Using this combination, the Consumer will spend  $\text{€}1.028/\text{day}$ , instead of  $\text{€}1.8/\text{day}$  in a fixed tariff.

In conclusion, only by switching to a dynamic tariff, with a prediction tool and with “manual” work to guide the consumption to the cheapest hours, an incredible saving in relation to the tariff fixed is achieved. By making an investment in sensors and actuators (IoT devices) to improve and automate the efficiency of their home, the savings can even be higher.

Taking into consideration that the number of electric vehicles is increasing, the above example may even achieve higher savings, if a V2G technology is used. Statistically, a car is parked for more than 80% of its useful life. Realizing this, a V2G approach may increase energy efficiency by only charging the vehicle at the cheapest periods of the day, assuring mobility for day-to-day life, with at least 20% of the car's battery reserved to power a household at the most expensive periods of the

day. With this approach, the car battery will be used as a power source during peak hours.

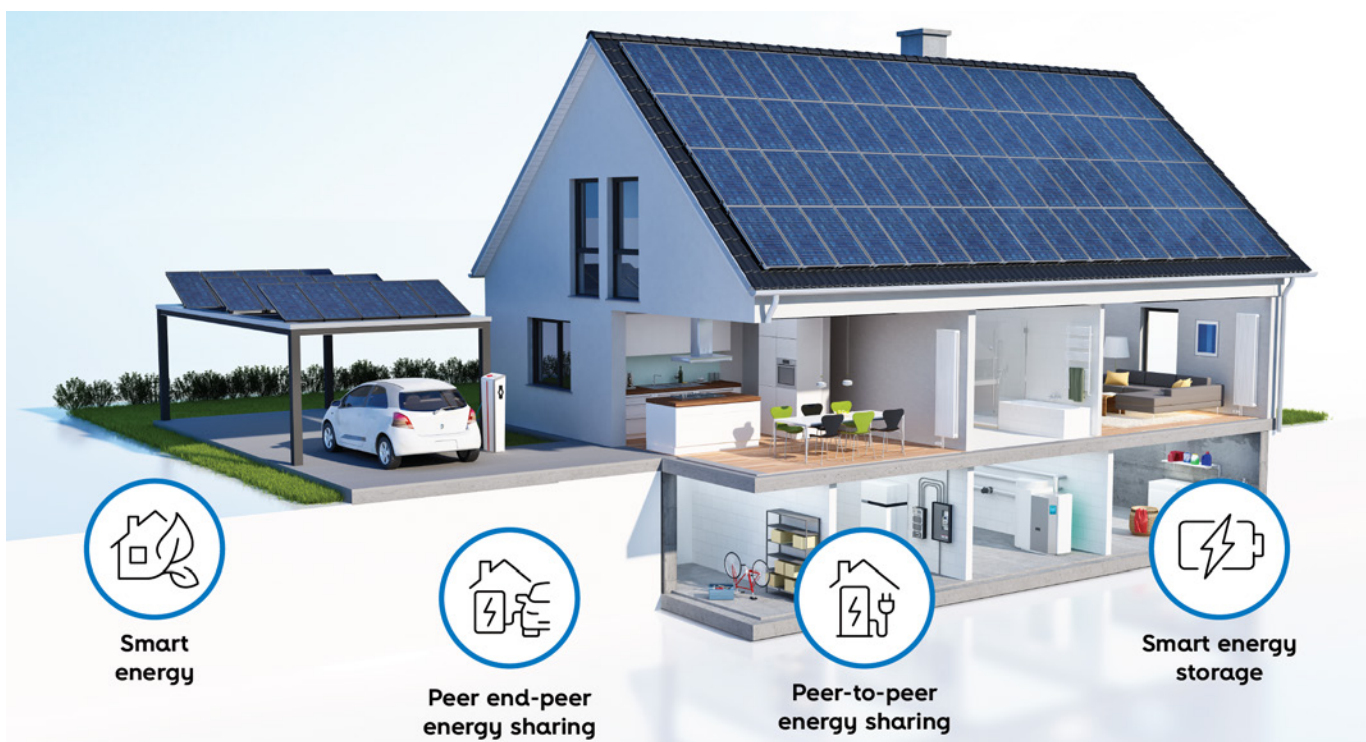
With this solution, it is possible to use energy from the grid most of the time at the cheapest price of the day. In that optimal scenario:

- $\text{€}0.025/\text{kWh} \times 20 \text{ kWh} \times 100\% = \text{€}0.5/\text{day}$

The Consumer will save  $\text{€}1.3$  per day ( $\text{€}1.8 - \text{€}0.5$ ).

After all these changes, this Consumer realizes that the weight of taxes in the monthly bill is very high, as the price of electricity is already optimized. These taxes are due to pay the grid and the management of the energy system. However, a new strategy may be implemented to reduce even more the costs. An investment in solar panels and joining a Smart Energy Community are the next possible steps.

By adding solar panels (**Figure 2**) to the household, this Consumer will not pay taxes when using energy from the panels and can optimize even more by charging the vehicle as much as possible during solar hours at no cost.



**FIGURE 2** - Illustration of a house that shares electricity via p2p

With an investment in solar panels, simulation software, forecasting, and charging and discharging control software for the electric vehicle, the consumer is almost self-sufficient (off-grid). Unfortunately, daily routine sometimes does not allow for the vehicle to be fully charged thanks to the solar panels, and not all the solar production is enough for all the equipment at home, forcing the consumption from the grid. Even if electricity is at its cheapest period, the network access charges (taxes) make the price quite expensive.

That said, joining or creating a Smart Energy Community (Figure 3 and 4) is a possible strategy to follow. By that, all neighbors joining the community may benefit from each other at a low cost and, by European legislation, without paying taxes because each neighbor may consume surplus energy from others (decentralization of production). Energy exchanges are established on a peer-to-peer platform that takes care of negotiations between neighbors and takes care of payments and charges for shared energy.

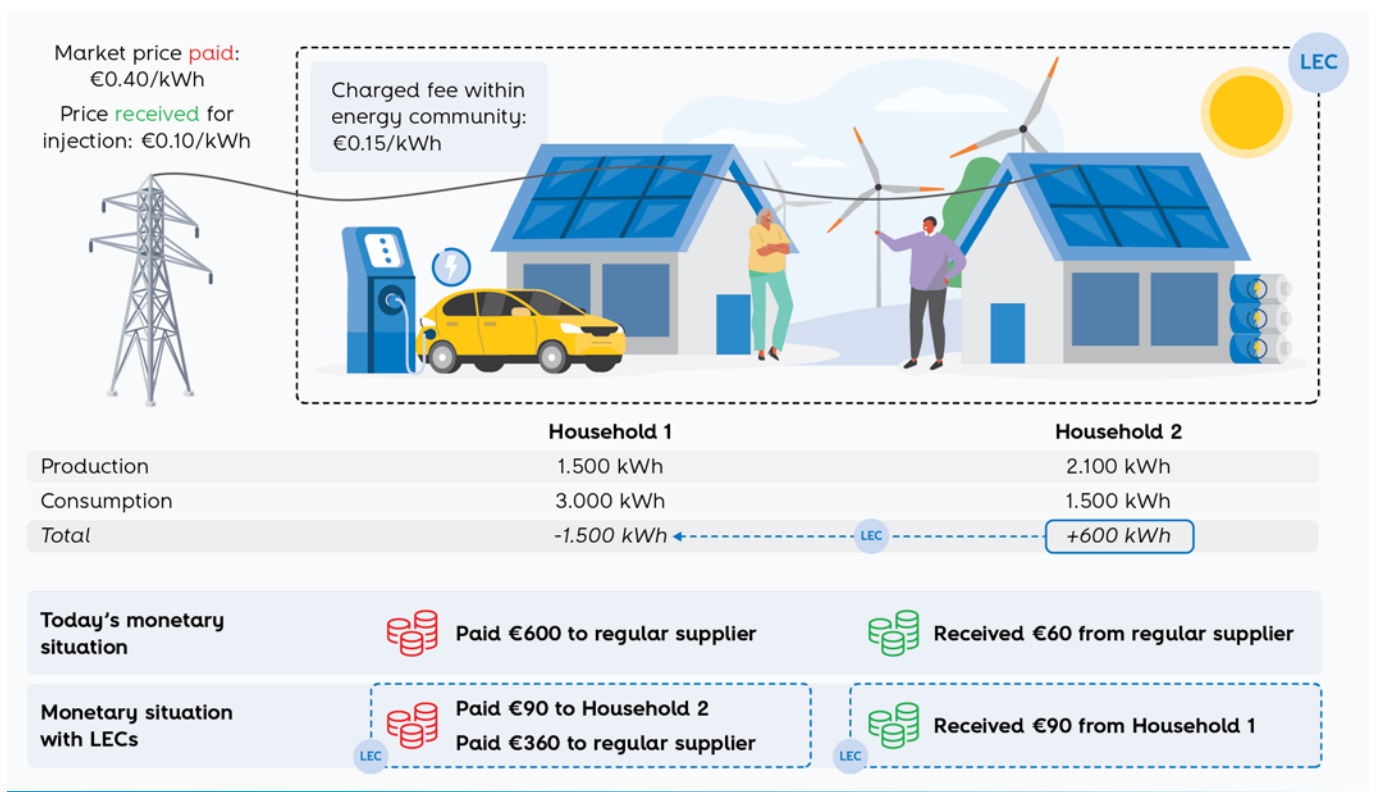


FIGURE 3 - Simulation of an energy sharing situation

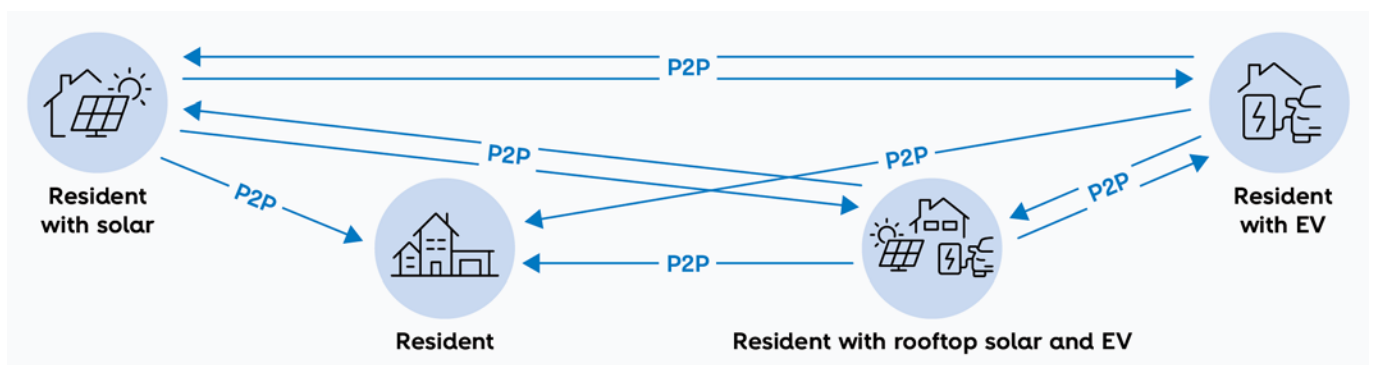


FIGURE 4 - Example of how electricity can be shared

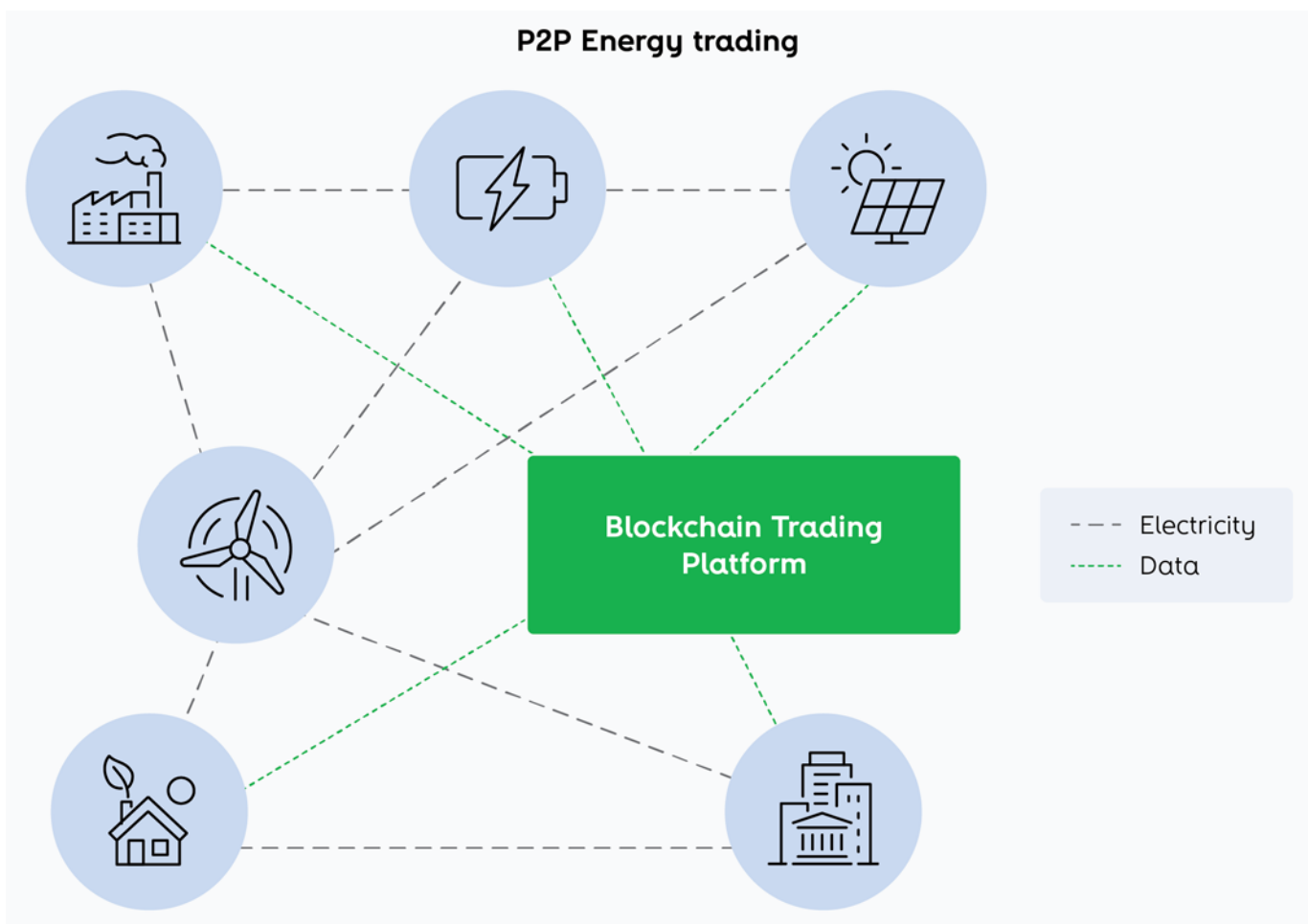
## Future disruptive evolutions

The tools and techniques showcased above facilitate notable changes in the energy industry by decentralizing power generation and enhancing the efficiency and affordability of energy consumption, while also supporting the shift towards cleaner energy sources.

These significant changes play a role in establishing a more robust and eco-friendly energy system that can effectively tackle the issues arising from the fluctuations in renewable energy sources and the increased need for clean power supply solutions. By distributing energy production across locations, instead of relying solely on big power stations,

self-sufficiency within neighborhood settings is enhanced. Furthermore, optimizing how efficiently energy is used results in decreased expenses for individuals and diminished environmental harm.

The importance of advancements like employing Artificial Intelligence (AI) for predicting consumption patterns and utilizing smart contracts through blockchain technology is essential (**Figure 5**). The incorporation of these innovations enables an accurate and dependable management of energy resources that supports the integration of renewable energy sources and boosts the resilience and flexibility of the electricity network. This not only leads to enhancements in energy efficiency and financial savings, but also impels the shift towards a more environmentally friendly and sustainable future for energy consumption.



**FIGURE 5** – Example of how electricity can be shared, stored and traded within a blockchain platform

Looking towards the future, several technologies stand ready to bring changes in the domain of dynamic electricity control. For instance, advanced computing has the potential to greatly improve real-time energy management and optimization. This could lead to adjustments in energy pricing and consumption patterns. Moreover, progress in AI and machine learning could pave the way for energy systems that are able to adjust consumption and production levels on their own, maximizing efficiency and seamlessly integrating with intelligent grids. The emergence of advanced computing holds the promise of tackling optimization challenges that are currently beyond the capabilities of traditional computers. AI algorithms, for instance, could optimize the distribution of energy resources in real time by taking a multitude of variables and constraints into account. Advancements in AI and machine learning will provide systems that can learn from data to forecast consumption patterns, allowing real-time energy optimization based on these predictions.

Another fascinating development lies in the applications of Decentralized Autonomous Organizations (DAOs) within energy management. These organizations could function independently from utility companies, using contracts and blockchain technology to oversee energy resources and transactions. This model could establish a transparent energy ecosystem capable of adapting to fluctuations in supply and demand with minimal external supervision. DAOs have the potential to enable peer-to-peer energy trading on scale, empowering

individuals and communities to manage their energy resources. This innovative approach could decrease dependency on grids. Using contracts, DAOs can streamline energy transactions, ensuring effectiveness while cutting down on expenses.

## Conclusion

To conclude, innovative electricity management practices are revolutionizing the energy sector. By integrating these systems with cutting-edge technologies and specialized software tools, like energy management platforms, bidirectional charging stations, energy communities, and peer-to-peer applications, electricity management and consumption are experiencing transformations. These advancements not only boost energy efficiency and sustainability, but also empower consumers to save costs, while contributing to a more environmentally friendly future. By decentralizing power sources, these tactics will also contribute to a more stable energy system.

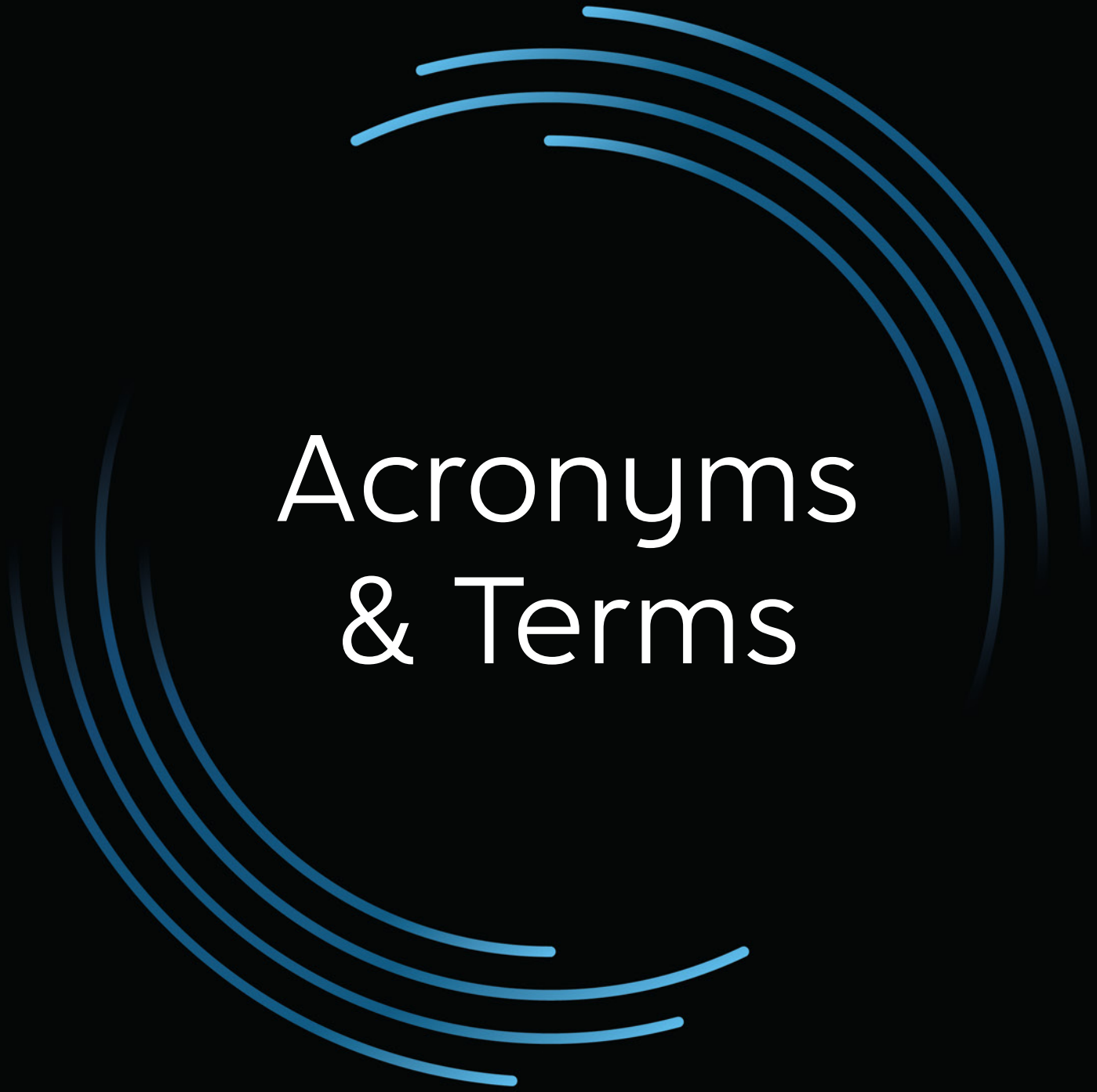
The incorporation of technologies like advanced computing, AI, and DAOs holds the potential to drive other substantial transformations, shaping a more effective and robust energy environment. As these technologies progress and evolve, possibilities for creativity and enhancements within the energy industry will be open. 🔗



---

## References

- [1] A. A. Bayod-Rújula, "Future development of the electricity systems with distributed generation," *Energy*, vol. 34, no. 3, pp. 377-383, Mar. 2009, doi: 10.1016/j.energy.2008.12.001.
- [2] M. Z. Jacobson et al., "100% Clean and Renewable Wind, Water, and Sunlight All-Sector Energy Roadmaps for 139 Countries of the World," *Joule*, vol. 1, no. 1, pp. 108-121, Sep. 2017, doi: 10.1016/j.joule.2017.07.005.
- [3] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18-28, Jan.-Feb. 2010, doi: 10.1109/MPE.2009.934876.
- [4] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381-388, Aug. 2011, doi: 10.1109/TII.2011.2158841.
- [5] G. A. Pagani and M. Aiello, "The Power Grid as a complex network: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688-2700, Jun. 2013, doi: 10.1016/j.physa.2013.01.023.
- [6] S. Bahramirad, W. Reder, and A. Khodaei, "Reliability-Constrained Optimal Sizing of Energy Storage System in a Microgrid," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2056-2062, Dec. 2012, doi: 10.1109/TSG.2012.2217991.
- [7] J. K. Kok, "The PowerMatcher: Smart Coordination for the Smart Electricity Grid," Ph.D. thesis, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, 2013. [Online]. Available: <https://repository.tudelft.nl>.
- [8] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91-99, Jun. 2016, doi: 10.1016/j.segan.2016.02.005.

A decorative graphic consisting of several concentric, light blue arcs that form a circular shape, centered behind the main text.

# Acronyms & Terms

---

<b>2</b>	<b>2FA</b>	Two-Factor Authentication
----------	------------	---------------------------

---

<b>3</b>	<b>3G</b>	Third Generation
	<b>3GPP</b>	3 <sup>rd</sup> Generation Partnership Project

---

<b>5</b>	<b>5G</b>	Fifth Generation
----------	-----------	------------------

---

<b>6</b>	<b>6G</b>	Sixth Generation
----------	-----------	------------------

---

<b>A</b>	<b>ACL</b>	Access Control List
	<b>Act</b>	Activation Functions
	<b>ADCs</b>	Analog to Digital Converters
	<b>AI</b>	Artificial Intelligence
	<b>AIM</b>	Automated Intelligent Management
	<b>AIM-DE</b>	Automated Intelligent Management Decision Elements
	<b>API</b>	Application Programming Interface

---

<b>B</b>	<b>BAA</b>	Broadband Access Abstraction
	<b>BB84</b>	Bennet-Brassard 1984
	<b>BBF</b>	Broadband Forum
	<b>BERT</b>	Bidirectional Encoder Representations from Transformers
	<b>BtB</b>	Back-to-back
	<b>BW</b>	Bandwidth

---

<b>C</b>	<b>CaC</b>	Configuration as Code
	<b>CapEX</b>	Capital Expenditure
	<b>CCODO</b>	CCO Domain Orchestrator
	<b>Cex</b>	Coexistence Element
	<b>CFR</b>	Crest Factor Reduction
	<b>CI/CD</b>	Continuous Integration/Continuous Delivery
	<b>CIR</b>	Committed Information Rate
	<b>CloudCO</b>	Cloud Central Office
	<b>CNAPP</b>	Cloud Native Application Protection Platforms
	<b>CNF</b>	Cloud-Native Network Function
	<b>CNN</b>	Convolutional Neural Networks
	<b>CORD</b>	Central Office Re-Architected as a Datacenter
	<b>COTS</b>	Commercial Off The Shelf
	<b>CP</b>	Cyclic Prefix
	<b>CPE</b>	Customer Premise Equipment
<b>CPU</b>	Central Processing Unit	

<b>CRFs</b>	Conditional Random Fields
<b>CSP</b>	Communication Service Providers
<b>CU</b>	Central Unit
<b>CV-QKD</b>	Continuous Variable Quantum Key Distribution

---

<b>D</b>	<b>DACs</b>	Digital to Analog Converters
	<b>DANs</b>	Deep Averaging Networks
	<b>DAOs</b>	Decentralized Autonomous Organizations
	<b>DAST</b>	Dynamic Application Security Testing
	<b>DBA</b>	Dynamic Bandwidth Allocation
	<b>DBS</b>	Double-Sideband
	<b>DCCs</b>	Digital Down Converters
	<b>DFE</b>	Digital Front End
	<b>DFT</b>	Discrete Fourier Transform
	<b>DHCP</b>	Dynamic Host Configuration Protocol
	<b>DHCP RA</b>	DHCP Relay Agent
	<b>DL</b>	Downlink
	<b>DLT</b>	Distributed Ledger Technology
	<b>DMCNN</b>	Deep Multi-Scale CNN
	<b>D-OLT</b>	Disaggregated OLT
	<b>DPD</b>	Digital Pre-Distortion
	<b>DSL</b>	Digital Subscriber Line
	<b>DU</b>	Distributed Unit
	<b>DUCs</b>	Digital Up Converters
	<b>DV-QKD</b>	Discrete Variable Quantum Key Distribution

---

<b>E</b>	<b>E2ESO</b>	End-to-End Service Orchestrator
	<b>ECL</b>	External Cavity Laser
	<b>ECP</b>	Extended Cyclic Prefix
	<b>EIR</b>	Excess Information Rate
	<b>EML</b>	External Modulated Lasers
	<b>eOMCI</b>	Embedded OMCI
	<b>ER</b>	Extinction Ratio
	<b>EU</b>	European Union
	<b>EV</b>	Electric Vehicle

---

<b>F</b>	<b>FANS</b>	Fixed Access Network Sharing
	<b>FCAPS</b>	Fault, Configuration, Accounting, Performance, and Security
	<b>FFT</b>	Fast Fourier Transform
	<b>FFTH</b>	Fiber To The Home
	<b>FIR</b>	Finite Impulse Response
	<b>FPGA</b>	Field Programmable Gate Array

**FTTH** Fiber-to-the-Home**FWM** Four-Wave Mixing

---

<b>G</b>	<b>GA</b>	Genetic Algorithm
	<b>GBT</b>	Gradient Boosting Trees
	<b>Gen AI</b>	Generative AI
	<b>GPON</b>	Gigabit Passive Optical Network
	<b>GPT</b>	Generative Pre-trained Transformer

---

**H** **H-QoS** Hierarchical Quality of Service

---

<b>I</b>	<b>IaC</b>	Infrastructure as Code
	<b>IAST</b>	Interactive Application Security Testing
	<b>IBN</b>	Intent-Based Networking
	<b>IDFT</b>	Inverse Discrete Fourier Transform
	<b>IDS</b>	Intrusion Detection Systems
	<b>IFFT</b>	Inverse Fast Fourier Transform
	<b>IFFTSHIFT</b>	Inverse Zero-frequency Shift
	<b>IFT</b>	Inverse Fourier Transforms
	<b>ILP</b>	Integer Linear Programming
	<b>InP</b>	Infrastructure Provider
	<b>IoT</b>	Internet of Things
	<b>IQBER</b>	Intrinsic Quantum Bit Error Rate
	<b>ISI</b>	Inter-Channel Interference
	<b>ISI</b>	Intersymbol Interference
	<b>ISP</b>	Internet Service Provider
	<b>IT</b>	Information Technology
	<b>IV</b>	Information Value
<b>IVR</b>	Interactive Voice Response	

---

**J** **JQL** JIRA Query Language

---

<b>L</b>	<b>LF</b>	Linux Foundation
	<b>LLMs</b>	Large Language Models
	<b>LNA</b>	Low Noise Amplifier
	<b>LSTM</b>	Long Short-Term Memory

---

<b>M</b>	<b>M&amp;A</b>	Mergers & Acquisitions
	<b>MAC</b>	Medium Access Control
	<b>MAE</b>	Mean Absolute Error
	<b>MDUs</b>	Multi-Dwelling Units
	<b>MI</b>	Mutual Information
	<b>MILP</b>	Mixed Integer Linear Programming
	<b>ML</b>	Machine Learning
	<b>MLP</b>	Multilayer Perceptron

---

**MPLS** Multiprotocol Label Switching**MSE** Maximum Spectral Excursion**MTE** Maximum Tuning Error

---

<b>N</b>	<b>N</b>	Neurons
	<b>NBI</b>	Northbound Interface
	<b>NCO</b>	Numerically Controlled Oscillator
	<b>NCP</b>	Normal Cyclic Prefix
	<b>NER</b>	Named Entity Recognition
	<b>NFE</b>	Network Federated Ecosystem
	<b>NFVI</b>	NFV Infrastructure
	<b>NLP</b>	Natural Language Processing
	<b>NMS</b>	Network Management System
	<b>NN</b>	Neural Network
<b>NR</b>	New Radio	

---

<b>O</b>	<b>ODN</b>	Optical Distribution Network
	<b>ODP</b>	Optical Distribution Point
	<b>O-DU</b>	O-RAN Distributed Unit
	<b>OECD</b>	Organization for Economic Co-operation and Development
	<b>OFDM</b>	Orthogonal Frequency-Division Multiplexing
	<b>OLT</b>	Optical Line Termination
	<b>OLT</b>	Optical Line Terminator
	<b>OLT</b>	Optical Line Terminal
	<b>OMCI</b>	ONU Management Control Interface
	<b>ONF</b>	Open Networking Foundation
	<b>ONOS</b>	Open Network Operating System
	<b>ONT</b>	Optical Network Terminal
	<b>ONU</b>	Optical Network Unit
	<b>ONU Auth</b>	ONU Authentication
	<b>ONU Mgmt</b>	ONU Management
	<b>O-RAN</b>	Open Radio Access Network
	<b>OSA</b>	Optical Spectrum Analyzer
	<b>OSJ</b>	Optical Splitter Joint
	<b>OSS</b>	Operation Support Systems
	<b>OTB</b>	Optical Terminal Box

---

<b>P</b>	<b>P2P</b>	Peer-to-Peer
	<b>PA</b>	Power Amplifier
	<b>PAPR</b>	Peak to Average Power Ratio
	<b>PAs</b>	Power Amplifiers
	<b>PHY</b>	Physical
	<b>PM</b>	Power Meter
	<b>PoC</b>	Proof of Concept
	<b>pOLT</b>	Physical OLT

<b>PON</b>	Passive Optical Network
<b>PoS</b>	Proof-of-Stake
<b>PPG</b>	Pattern Generator
<b>PPPoE</b>	Point-to-Point Protocol over Ethernet
<b>PRACH</b>	Physical Random Access Channel
<b>PRB</b>	Physical Resource Blocks
<b>PTP</b>	Precision Time Protocol
<b>PUSCH</b>	Physical UL Shared Channel

---

<b>Q</b>	<b>QA</b>	Quality Assurance
	<b>QAM</b>	Quadrature Amplitude Modulation
	<b>QKD</b>	Quantum Key Distribution
	<b>QoS</b>	Quality of Service

---

<b>R</b>	<b>RAG</b>	Retrieval-Augmented Generation
	<b>RAN</b>	Radio Access Network
	<b>RASP</b>	Runtime Application Self-Protection
	<b>RBs</b>	Resource Blocks
	<b>R-CNN</b>	Region-based Convolutional Neural Networks
	<b>RDD</b>	Resilient Distributed Datasets
	<b>REs</b>	Resource Elements
	<b>RF</b>	Radio Frequency
	<b>RF</b>	Random Forest
	<b>RFM</b>	Recency, Frequency, and Monetary Value
	<b>RLHF</b>	Reinforcement Learning from Human Feedback
	<b>RMS</b>	Root Mean Square
	<b>RoE</b>	Radio over Ethernet
	<b>ROP</b>	Received Optical Power
	<b>RPI5</b>	Raspberry PI 5
	<b>RSA</b>	Rivest-Shamir-Adleman
	<b>RU</b>	Radio Unit

---

<b>S</b>	<b>SA</b>	Simulated Annealing
	<b>SAST</b>	Static Application Security Testing
	<b>SBI</b>	Southbound Interface
	<b>SBOM</b>	Software Bill of Materials
	<b>SCs</b>	Subcarriers
	<b>SDLC</b>	Systems Development Life Cycle
	<b>SDN</b>	Software Defined Networks
	<b>SDN M&amp;C</b>	SDN Manager & Controller/ SND Management and Control
	<b>SDO</b>	Standards Development

		Organization
	<b>SDUs</b>	Single-Dwelling Units
	<b>SEBA</b>	Software Enabled Broadband Access
	<b>SEM</b>	Element Management System
	<b>SEM</b>	Energy Management Systems
	<b>SHAP</b>	SHapley Additive exPlanations
	<b>SIEM</b>	Security Information and Event Management
	<b>SKR</b>	Secure Key Rate
	<b>SLA</b>	Service Level Agreements
	<b>SMOTE</b>	Synthetic Minority Oversampling Technique
	<b>SOA</b>	Semiconductor Optical Amplifier
	<b>SOM</b>	Service Order Management
	<b>SOP</b>	State of Polarization
	<b>SPD</b>	Single Photon Detector
	<b>SpRS</b>	Spontaneous Raman Scattering
	<b>SSMF</b>	Standard Single-Mode Fiber
	<b>SVDT</b>	Systems Vulnerability Detection Tool
	<b>SVMs</b>	Support Vector Machines
	<b>SVN</b>	Apache Subversion
	<b>SW</b>	Software

---

<b>T</b>	<b>TaaS</b>	Testing-as-a-Service
	<b>TDM</b>	Time Division Multiplexing
	<b>TELCO</b>	Telecommunication Company
	<b>TM</b>	Telecom Management
	<b>TOU</b>	Time of Use
	<b>TS</b>	Tabu Search
	<b>TWDM</b>	Time and Wavelength Division Multiplexing

---

<b>U</b>	<b>UE</b>	User Equipment
	<b>UL</b>	Uplink
	<b>USE</b>	Universal Sentence Encoder

---

<b>V</b>	<b>V2G</b>	Vehicle to Grid
	<b>VC</b>	Verifiable Credential
	<b>VHDL</b>	Very High-Speed Integrated Circuit Hardware Description Language
	<b>VNO</b>	Virtual Network Operator
	<b>VOA</b>	Variable Optical Attenuator
	<b>VOLTHA</b>	Virtual OLT Hardware Abstraction
	<b>vOMCI</b>	Virtual OMCI
	<b>VON</b>	Verifiable Organizations Network

**VSA** Vector Signal Analyzer  
**VSG** Vector Signal Generator

---

**W**      **WAF** Web Application Firewalls  
          **WDM** Wavelength Division Multiplexing  
          **WLAN** Wireless Local-Area Network  
          **WM** Wavelength Multiplexer

---

**x**      **XDR** Extended Detection and Response  
          **XGS-PON** 10 Gigabit Symmetric Passive  
                    Optical Network

---

**Y**      **YOLO** You Only Look Once

---

**Z**      **ZTP** Zero Touch Provisioning

# Acknowledgments

Year after year, InnovAction is the result of powerful teamwork. Several individuals from Altice Labs get involved and, along with external partners, contribute with 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 contributions and for sharing the knowledge resulting from their research and work.

We also want to leave a note of gratitude to all technical and editorial reviewers who did a meticulous job and contributed to improving all articles' quality and excellence and, therefore, the entire magazine.

Finally, a special word and acknowledgment to our designers for the excellent job of granting an appealing and coherent design across the entire magazine.

**Edition**

December 2024

The background features a series of concentric, semi-circular blue arcs that originate from the left side and curve towards the right. The arcs vary in length and thickness, creating a sense of depth and movement. The overall effect is reminiscent of a stylized fingerprint or a series of ripples in water.

InnovAction  
is more  
than just  
technology

# 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 and reinforce 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.





[www.alticelabs.com](http://www.alticelabs.com)