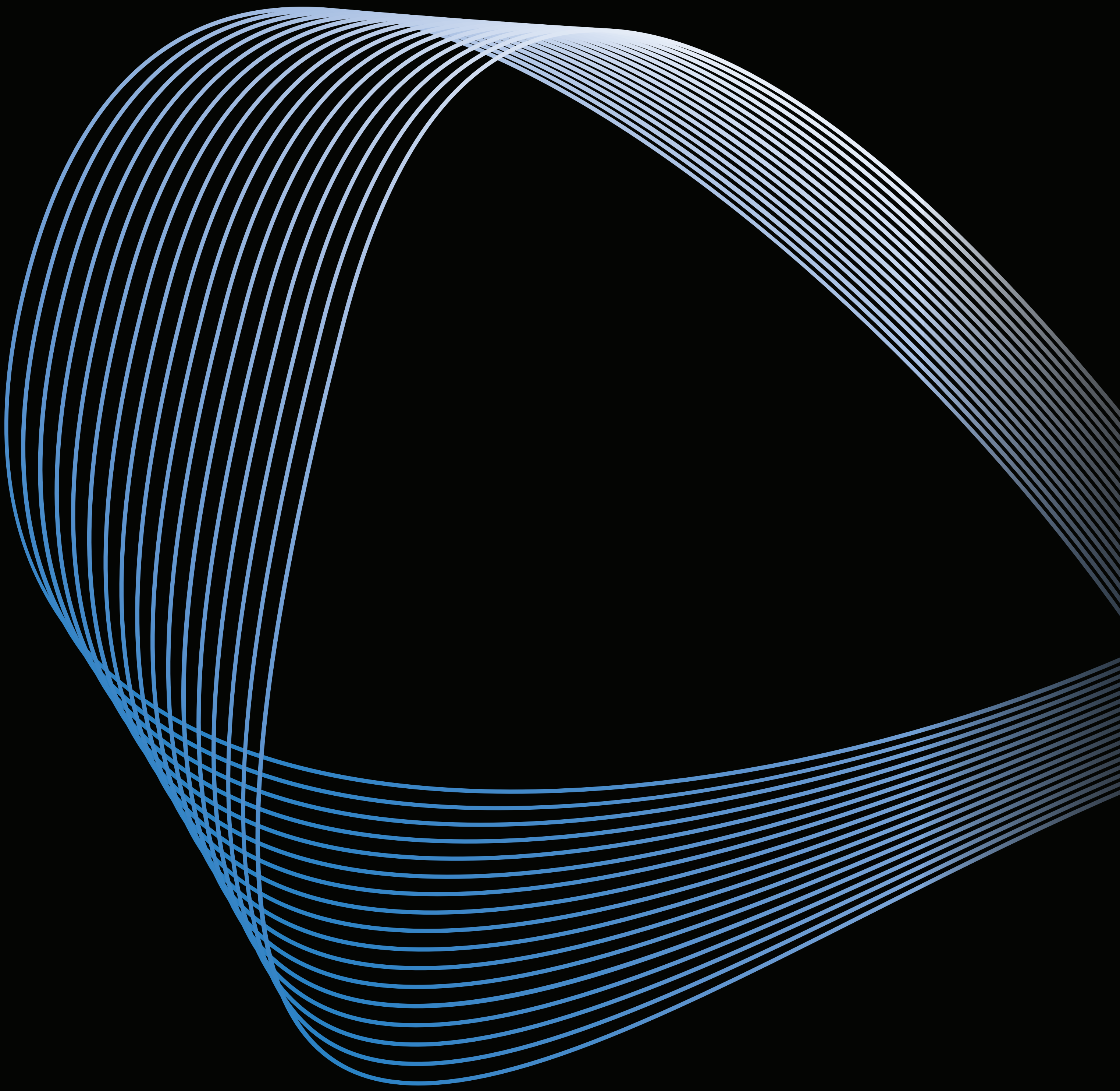
altice labs

InnovAction

#8 | December 2023



InnovAction

#8 | December 2023

Title
Innovaction
Altice Labs journal

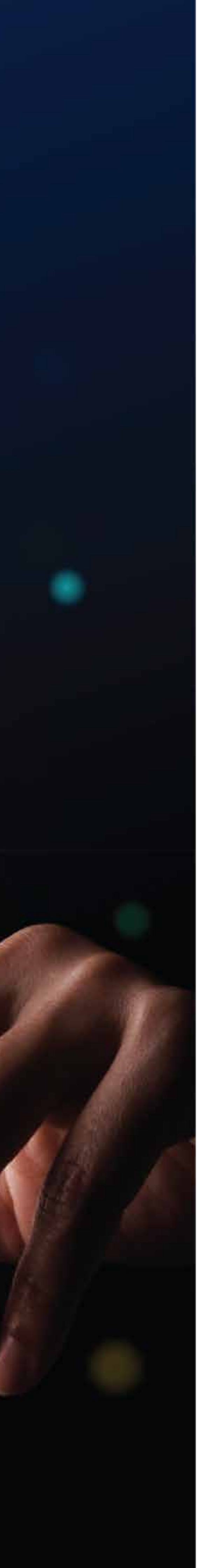
Edition
2023

Design
Altice Labs

Legal deposit
512167/23

ISSN | DOI
2975-8459 | 10.57650/INNOVACTION

altice
labs



Opening note

Welcome to the eighth edition of InnovAction, Altice Labs' magazine, which provides the latest insights into our knowledge, expertise, technological developments, and Product Innovations.

As we reflect on another year gone by, we turn our attention to the complexities of both the ongoing wars and global conflicts and the reshaping of the world's innovation and technological ecosystems. At the technological level, advancements in Artificial Intelligence continue to shape the landscape of innovation, uncovering the transformative power of generative AI and the ever-rising significance of data. Several new challenges are also now posed by decentralized and virtualized businesses, particularly when considering the increasing remoteness of individuals and the pivotal role of sustainability. Furthermore, we are witnessing the rise of cybersecurity to a military-grade status, the dawn of quantum technologies, and the profound influence of the so-called MAGA corporations on operators' businesses.

To address all these challenging topics, Altice Labs is committed to pursuing its goal of exploring the realms of high-speed access networks and their diverse applications. This eighth edition of InnovAction highlights the most important recent innovations of the year 2023, which was notably marked by a tribute to the incomparable contribution of Alcino Lavrador. It's an honor to be writing the first institutional message as the General Manager of Altice Labs, after so many years of leadership and innovation by Alcino, who was responsible for launching this publication, which has become a reference in the telecommunications and technology sectors, and for leading Altice Labs to new heights of excellence and recognition. He left a legacy of knowledge and action, vision and passion, and collaboration and creativity that inspires all of us who are part of this team.

On behalf of all Altice Labs employees, I thank Alcino for his dedication, enthusiasm, and contribution to the development of the innovation ecosystem in Portugal and around the world.

João Paulo Firmeza
General Manager of Altice Labs
PFirmeza@alticelabs.com

Editorial note

We live in challenging and changing times, facing an uncertain future. Combining efforts to build a fairer, safer, and more inclusive world is necessary to correct this path. Altice Labs is proud to work daily towards a society of all and for all, investing every day in innovation to overcome obstacles and ensure the transition to digital, where access to information is available to everyone. We work continuously with academia, startups, and companies, sharing experiences and knowledge to build solutions that respond to actual needs.

InnovAction magazine is a showcase where we present our firm commitment to innovation. This edition celebrates the achievements reached in advanced connectivity, highlighting optical networks and the 6G communications path; looks at security issues and explores the new capabilities that quantum communications can offer; presents the successes achieved in operations support systems in mitigating possible software and hardware problems; and finally, exposes different cutting-edge solutions, spotlighting the immersive experiences that extended reality allows, as well as solutions based on various aspects of artificial intelligence, always with sustainability in mind.

- **The path from 5G to 6G** outlines the next steps of this evolution process following two distinct approaches: in the short and medium term, evolving and enriching 5G technologies toward the so-called "5G Advanced" or "Beyond-5G" while in a longer-term perspective, starting the research towards the definition of technological foundations of 6G.
- **Evolving invisible 5G small-cells to 6G smart radio environments** highlights efforts to manage the densification of cells, considering the seamless integration with urban infrastructure while designing the future 6G innovative radio environments, providing opportunities for widespread adoption in the short term.
- **50G TDM PON and 50G TWDM-PON, challenges on the migration from current PON technologies** discusses and analyzes the key issues in developing the upcoming PON technologies and the careful planning and coordination efforts required to ensure seamless integration with the existing G-PON and XG(S)-PON systems.
- **Automated testing in telecommunications: enhancing OLT software quality assurance** to enable the prompt detection of defects, minimize manual labour, and improve overall testing efficiency while keeping high motivation levels among testers.
- **Quantum Key Distribution for secure communications** highlights the disruptive importance of Quantum Computing and Communications and

its transformative impact on the security of numerous widely employed cryptographic systems.

- **The use of an integrated maturity model for cybersecurity on software development teams** to evaluate the existing cybersecurity-focused processes aiming at delivering reliable, consistent, and transparent software.

- **The integrated vision of sustainable energy solutions in the Connected Home ecosystem** explores enhanced energy possibilities in the context of smart homes, employing a case study approach to demonstrate practical applications and benefits.

- **VR2Care: enabling multiuser immersive experiences for rehabilitation, physical activity, and socialization** addresses functional and technical insights into the VR2Care digital ecosystem to create age-friendly virtual environments that foster interactive technologies to promote physical activity and social interaction.

- **Glaucoma diagnosis automation** introduces the cognitive challenge and technical approach to implementing a model to automate glaucoma diagnosis using artificial intelligence methodologies without neglecting explainability methods to make comprehension by healthcare professionals easy.

- **Data governance: understanding, assessing, and evolving** explores developing a data governance maturity model to answer the needs of data-driven organizations requiring robust data governance practices.

- **Speech and language analytics** explore natural language process techniques to enable automatic performance evaluation, subject categorization, relevant information extraction, and even conversation summarization, allowing organizations to gain business competitiveness.

- **Generative AI – unleashing knowledge from information** presents the latest evolution in artificial intelligence, where machines have started to produce content, opening the doors to a green field of new opportunities.

Come and join us on this transformative journey!

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

Pedro Carvalho
PCarv@alticelabs.com

Filipe Cabral Pinto
filipe-c-pinto@alticelabs.com

Contents

DOI: 10.57650/INNOVACTION/2023.01

01 The path from 5G to 6G

Since the first generation in 1980, a new mobile network generation has been released approximately every 10 years. Now that the deployment of 5G is well underway across multiple geographies, it is time to prepare for the next generations of this evolution process following two distinct but complementary approaches: in the short and medium term, evolving and enriching 5G technologies toward the so-called "5G Advanced" or "Beyond-5G", while, on a longer-term perspective, starting the research toward the definition of the key 6G technological building blocks. Following the same pattern, the 6G commercial release is expected to happen around 2030 and initial work on its research and requirements identification has started. As in previous mobile network generations, Europe will play a key role in this process and will be one of the leading forces behind the development of the technology, through initiatives such as the 6G Smart Networks and Services Industry Association (6G-IA) and One6G, both of which have Altice Labs as an active member.

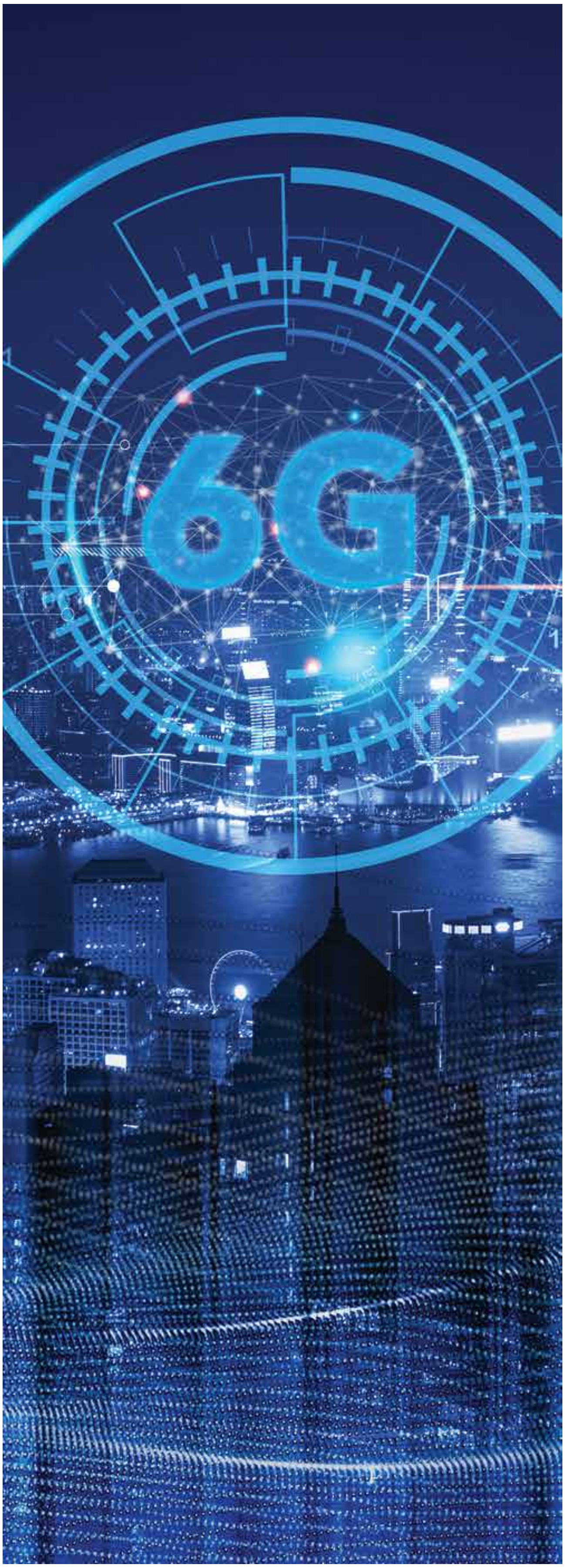
Authors

Jorge Carapinha, Altice Labs, jorgec@alticelabs.com

Francisco Fontes, Altice Labs, fontes@alticelabs.com

Keywords

6G; 5G; Key Performance Indicators; Key Value Indicators; 6G Smart Networks and Services Industry Association; one6G



Introduction

Cellular radio communications acquired a fundamental role in voice communications in the 1990s, with the introduction of the 2G radio system. Since then, other radio data communication networks have emerged, and wireless connectivity became omnipresent in our daily activities. Many technologies are available today, like WLAN, LoRa, ZigBee, and Bluetooth, addressing different application areas, based on their coverage, cost, power consumption, and bitrates, but also with significant overlaps. Together they impact the way humans and things communicate in a common environment.

Cellular networks evolved from 2G to 5G, expanding from the consumer (B2C) into the business market (B2B), making it an integral part of Industry 4.0 and the overall society digitalization process. Thanks to 5G and other wireless technologies, machines and all kinds of objects are getting interconnected. Technology evolutions are allowing high-definition video transmission to spread, 360° videos are emerging, as well as augmented and virtual reality based on edge computing.

All of this demands better performance and more complex functionalities. Thus, while in the coming years, 5G will fuel the business use cases it was designed for, in the future it will present limitations in fulfilling the requirements of next-generation services and applications. For end systems, the pursuit includes achieving more bandwidth, lower latency, higher reliability, stronger security, and the ability to connect everything everywhere, both indoors and outdoors. For infrastructure owners, more efficiency, easy deployment, and automation in many processes will be strong requirements in multiple application domains: industry, traffic management, and public safety, among others.

Now that 5G phases 1 and 2 (based on 3GPP Rel 15 to 17 technical specifications) is completed, the technology is evolving toward 5G Advanced (3GPP Rel 18 to 20), enhancing and expanding functionality and performance. 5G is being widely deployed around the world, even if just a few Public Land Mobile Networks (PLMN) have been deployed in pure, native configurations (Standalone/SA deployments), with many using the 5G New Radio (5G-NR) technology to enhance existing 4G commercial operations (Non-Standalone/NSA deployments).

In this context, efforts from industry and academia have started to look beyond 5G and prepare for the journey toward the next generation, 6G. 6G is in the process of being defined, providing the means for the next wave of services and applications, based on technological evolutions. As with the previous mobile generation, 6G is expected to incorporate disruptive innovations. Areas such as Artificial Intelligence (AI) and sensing, which traditionally have been seen as out of the scope of communication technologies, are expected to be integrated as core components of 6G in several areas.

The ambitions of 6G

6G use cases

6G is expected to enhance existing use cases by targeting multiple vertical sectors and markets, as well as to enable new use cases. The evolution Beyond 5G will allow a range of new services relying on features like indoor location and higher capacity while leveraging the benefits of IoT and data analytics. New lightweight devices (RedCap) and wearables will emerge relying on distributed computing, intelligent computing surfaces, and storage enabled via edge cloud. Examples of emerging services and use cases include:

- Holographic teleportation: enabling virtual teleportation in real-time, based on 3D video;
- Extended reality: making augmented reality (AR) and virtual reality (VR) commonplace;
- Ubiquitous connectivity;
- Unmanned Aerial Vehicle (UAV) services;
- Autonomous services: including autonomous driving, autonomous delivery of goods, autonomous agriculture in remote areas, public-safety network connectivity, autonomous harbors;
- Internet-of-everything: involving billions of devices/sensors with Internet connectivity interacting directly, providing interactive intelligence, and creating a surrounding ambience.

In North America, the Next G Alliance is exploring new opportunities that are expected to arise from 6G applications, in four fundamental use case areas [1]: **Living** (improving the quality of everyday living), **Experience** (improving the Quality of Experience in areas such as entertainment, learning, and healthcare), **Critical** (improving the quality of critical roles in sectors such as healthcare, manufacturing, agriculture, transportation, and public safety), and **Societal** (achieving or improving on high-level societal goals).

Objectives and target KPIs and KVIs

Every new mobile generation is expected to significantly improve the performance of previous generations, and 6G is no exception. Technology performance is usually quantified through Key Performance Indicators (KPI), a set of measurable and quantifiable metrics. 6G aims at improving and expanding on top of 5G functionality and KPIs, addressing the anticipated requirements of the connected society of 2030 and beyond.

Despite the remarkable performance enhancements accomplished by 5G, the ever-growing dependency on wireless connectivity, and the search for innovative devices, services, and applications, anticipate the limitations of 5G to handle challenges ahead. 6G use cases and requirements are still being identified and KPIs are not fully defined, but some predictions can be made. 5G-PPP produced one of the first attempts to define B5G and 6G KPIs, as well as the respective measurement approaches [2].

Often the new KPIs result from new requirements posed by emerging use cases in several application domains in relation to which 5G performance is not fully satisfactory [2]. **Table 1** (adapted from [3]) identifies some of the limitations of 5G, especially in view of specific application domains, and the respective challenges to be faced by 6G.

In relation to a few selected KPIs, **Figure 1** illustrates the ambitions of 6G, using the corresponding 5G reference numbers as a baseline.

While technology-driven KPIs are useful to measure performance from a strictly technical point of view, the assessment of the societal and economic value, the contribution to satisfy concrete citizens' needs, and the impact on key values (KV) require a different approach. The 6G-IA [4] identified a set of fundamental KVs to which 6G is supposed to contribute [4]: environmental sustainability, societal sustainability, economic sustainability and innovation, democracy, cultural connection, knowledge, privacy and confidentiality, simplified life, digital inclusion, personal freedom, personal health and protection from harm, and trust.

To measure the contribution to those KV, Key Value Indicators (KVI) are a more effective tool by emphasizing societal, economic, and environmental drivers. As the role of mobile technologies becomes more and more diversified through an increasing number of vertical sectors and application domains, the importance of KVIs becomes more significant. Examples of relevant KVIs could be the reduction of accidents or injuries in urban traffic, in the domain of urban mobility, or the reduction of person-hours required to execute a specific activity in agriculture or industry.

Precisely assessing KVIs is challenging, especially in technologies with low maturity levels, such as 6G. The quantification of KVIs often requires deployed technologies and well-established use cases, which is only possible in advanced development stages. Nevertheless, through interviews, questionnaires, and trials, preliminary target values can be estimated. General guidelines on methodologies to define 6G KVIs have been produced by the 6G Infrastructure Association (6G-IA) [5].

	Application area	5G KPI	6G challenge
Ultra-low latency	Automatic drive, high precision industrial production	Delay < 1 ms at static and low speeds, but cannot be reached at high speeds	< 1 ms in high speed mobile environments
High precision positioning	Unmanned vehicle positioning and navigation, indoor precise positioning	Outdoor ~ 10 m Indoor ~ 1m	Outdoor meter level, indoor cm level positioning
Ultra-high data rate transmission	Ultra high-definition video, holographic image	Transmission rate < 20 Gbps User experienced data rate ~ 100 Mbps	Peak data rate at Tbps level User experienced data rate: 1-10 Gbps
Ultra-dense connection	Shopping malls, stations, fully automatic production lines	10^6 devices/km^2	10^8 devices/km^2
Ultra-reliable/ safe	V2X, telemedicine	99.999%	99.99999%
Global coverage	Marine and satellite communication	Ocean coverage: 5% Land coverage: 20%	Global coverage: space-air-ground-sea
High energy efficiency	Internet of bio-nano-things	Network energy efficiency: 10^7 bit/J	Network energy efficiency: 10^9 bit/J

TABLE 1 – 6G challenges

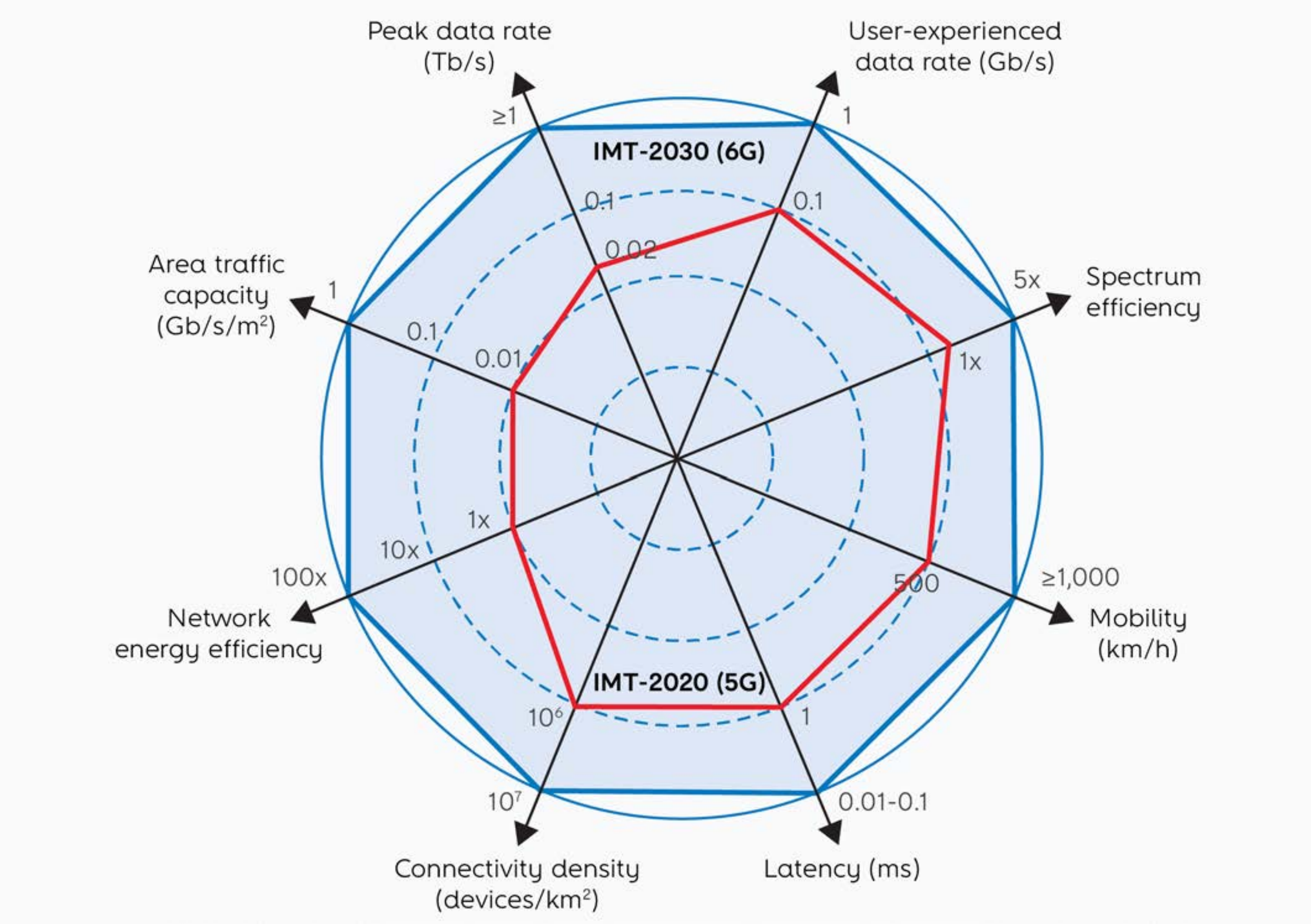


FIGURE 1 – 5G/6G reference KPIs compared. Source:[6]

A new architectural approach for 6G

Given the profound and forward-looking changes introduced by 3GPP with the 5G architecture, one may wonder why new architectural principles will be needed for 6G. In fact, some of the key transformations expected in 6G and the incorporation of new technology enablers will certainly require a revision of fundamental architectural principles. The 5G-PPP Architecture Working Group identified several areas in relation to which enhancements are required for 6G, eventually leading to the creation of a new architectural framework [7]:

- **Enhanced Service-Based Architecture:** enhancements to the 5G Service-Based Architecture (SBA) may include better cross-plane and cross-domain interactions, particularly regarding data collection for analytics and AI/ML needs, greater flexibility to accommodate new types of end-user devices, and access network topologies;

- **AI:** 6G networks will be distributed AI platforms. AI will play a key role in optimizing several aspects of the 6G radio interface: physical layer configuration, QoS assurance, mobility, and resource management. Network operations will also benefit from AI to automate and streamline operational workflows and handle increasing complexity;

- **Programmability:** 6G service management is expected to cope with a much larger set of heterogeneous services and infrastructure complexity, reinforced by the ongoing trend toward *softwarisation*, *edgification*, and *cloudification*; programmability will be extended to the data plane for additional flexibility and efficiency;

- **Continuum Orchestration:** multiple orchestration domains that traditionally are separately handled (core, RAN, transport network, edge, extreme edge) require seamless integration toward the continuum of orchestration domains.

ITU-T FG-NET2030 *Focus Group on Technologies for Network 2030* [8] defined a number of basic architectural principles to be followed in the specification of the next generation of public networks targeted at the year 2030 and beyond, including native programmability, backward compatibility, heterogeneity, native slicing, intrinsic anonymity, and security support for all network operations, resilience and network determinism. Although some of these principles have already been associated with 5G, new use cases and functionalities are expected to impose new architectural approaches.

Technology enablers

As with 5G, multiple technologies will compete to define 6G. Some of these technologies are intimately linked to specific 6G developments, whereas in other cases new technological advances from different areas will be adopted to strengthen and diversify 6G capabilities. This section provides a non-exhaustive list of promising 6G technology enablers expected to improve 6G KPIs.

Radio Access Network (RAN): the ITU-R has identified a number of candidate evolutions for the radio interface, such as advanced modulation, coding and multiple access schemes, advanced antenna technologies (e.g., evolutions of massive MIMO), in-band full-duplex communications, reconfigurable intelligent surfaces, holographic radio, orbital angular momentum, THz communications, technologies to support ultra-high accuracy positioning [9]. In addition, enablers to enhance the radio network have also been identified, including RAN slicing, resilient and soft networks for guaranteed QoS, a new RAN architecture, interconnection with non-terrestrial networks, ultra-dense radio network deployments, or enhanced RAN infrastructure sharing.

Artificial Intelligent (AI) & Machine Learning (ML): traditional network design, deployment, and operation are not efficient enough to cope with increasingly complex resource management in mobile networks. In this scenario, AI/ML tools will be key to automating complex tasks by using large amounts of data collected by the system and driving optimization at multiple levels, in the data, control, operation, and management planes. AI/ML-based solutions can be especially effective in handling complex challenges of the RAN involving a wide range of performance requirements, such as coverage, throughput, latency, reliability, and user experience.

Integrated Sensing and Communication (ISC): the high-frequency bands used for communication and sensing gradually approach or even overlap with each other. In 6G, sensing and communication functions are expected to coexist and be fully integrated, by exploiting the sensing capabilities of radio frequency signals in the millimetric and THz bands. Localization with centimeter accuracy is expected to be achieved. Sensing capabilities could also be used to enhance network performance by providing optimization input for network steering – for example, by detecting objects that obstruct the direct propagation path between a transmission node and a device [10].

RAN virtualization, cloudification, and decentralization: in recent years, the Open RAN (ORAN) Alliance [11] has accelerated the trend toward virtual RAN architectures by creating open interfaces between RAN elements. The trend of RAN decentralization and virtualization, initiated in 5G, is expected to accelerate in 6G. New challenges will be faced in this process, for example, the optimization of the overall network performance, but ORAN will certainly be a catalyst for innovation in the RAN, especially through the integration of AI/ML capabilities.

Non-Terrestrial Networks (NTN): historically, terrestrial and non-terrestrial networks (NTN) communication systems have evolved separately. In general, solutions tailored for terrestrial networks cannot be directly applied to NTNs due to specific characteristics (e.g., highly dynamic network topologies, specific channel models, and longer latencies). Already introduced in 5G, with 6G NTN technologies will be natively considered from the beginning of the design stage, which will permit the evolution of technologies for terrestrial and non-terrestrial networks to be aligned and consistent. 6G will natively incorporate an NTN component to enable better KPIs in terms of bandwidth, latency, reliability, coverage, and connectivity density in unserved and underserved areas.

Digital twins: a digital twin is the virtual representation of elements and dynamics of a physical system, to monitor, design, simulate, analyze, optimize, or predict its behavior. The relationship between digital twins and 6G can be mutually beneficial. Digital twins impose strict requirements in terms of throughput, reliability, resilience, and latency that go beyond what 5G is currently able to offer, thus the widespread use of digital twins may become an important driver for the deployment of 6G [12].

Edge and ubiquitous computing: an important trend in the last few years, expected to accelerate in the future, is the explosive growth of interconnected edge devices, supporting a wide range of applicational areas, from transportation systems to surveillance and smart cities. At the same time, the so-called Cognitive Cloud [13] is set to provide an AI-enabled Cloud-Edge framework that is able to adapt automatically to the growing data volume and complexity, by integrating seamlessly and securely multiple computing and data environments, spanning from core cloud to edge, thus becoming a continuum of edge zones. 6G will represent a catalyst and a facilitator for this evolution.

Blockchain: the key characteristics of Distributed Ledger Technologies (DLTs), which enable services among distributed actors, are transparency, immutability, non-repudiation, proof of provenance, integrity, and pseudonymity. DLT has the potential to provide distributed trust among different stakeholders in a multitenant/multi-domain environment. Blockchain technology, a specific type of DLT, which has been used in many fields as a safe and reliable decentralized distributed network to guarantee the privacy and safety of legitimate users, is expected to become a key 6G technology, mainly to enable secure collaboration and coordination in transparent, trustless, and distributed environments [14]. Blockchain technology provides a solution to enhance spectrum-sharing technology in a decentralized way.

Quantum: important synergies are expected from the combination of 6G and quantum technologies, with a direct impact on how we communicate, engage, and trade information [15]. Quantum technologies are developing at a fast pace and enable promising applications for mobile networks. Quantum computing can provide increased computation capacity of the network, whereas quantum communications and even quantum sensing are likely to have an important role in 6G. The ability of quantum communication to enable unconditional security will surely represent a key asset to be exploited in 6G. For further information on Quantum technologies, refer to the article "*Quantum Key Distribution for Secure Communications*".

The journey toward 6G

6G standardization plan

As with previous mobile technologies, beginning with 3G in 1998, the standardization of 6G will be conducted by two main actors: the ITU-R (the ITU sector for radio communications) and the 3rd Generation Partnership Project (3GPP).

The Technical Specifications (TS) for cellular technologies, specified by the 3GPP, are structured in 'Releases', with cycles of 15 to 24 months, going through Stages 1 to 3 [16]. The first set of 5G standards, known as 5G phase 1, was provided by

3GPP Release 15 and completed in 2019. 5G phase 2 corresponds to Releases 16 and 17 (completed in March 2022), which further evolved and expanded Release 15.

The outcome of this effort provides a strong baseline and a rich set of enablers for the development of the second phase of 5G standardization, known as 5G-Advanced, kicked off with the launch of Release 18 in 2022. 5G-Advanced corresponds to 3GPP Releases 18, 19, and 20, expected to provide technology expansions and improvements, and can be seen as the first step of the journey toward 6G. Release 18 introduces a number of significant enhancements, namely improved network performance, more flexible and efficient spectrum use, new types of devices, e.g., reduced capability (RedCap) devices, evolved network topologies (e.g., device-to-device communication, non-terrestrial networks), widened scope of AI/ML-based solutions (e.g., to enhance the performance of NR air interface) [17].

The preparatory work on 6G is set to start with Release 19, while 3GPP 6G studies are likely to start in 2025 under Release 20 and the first technical specifications are expected to be closed with Release 21, in 2028. Technical Specifications for first industry implementations are expected in 3GPP Release 22 (2028-2029), just in time to go to market in 2029-2030. A tentative plan for 3GPP standardization of 6G, as foreseen by Ericsson [21], is provided in **Figure 2**.

ITU-R is responsible for the definition of the overall research and development directions, as well as the technical performance requirements of 6G. As of October 2023, ITU-R is working on the definition of the International Mobile Telecommunications (IMT) for 2030, which will set the targets to be accomplished by 6G. ITU-R Recommendation "Framework and overall objectives of the future development of IMT for 2030 and beyond", to be approved by the end of 2023, will define guidelines on the framework and overall objectives of the future development of IMT-2030, commonly known as 6G.

The ITU-R report M.2516-0 "Future technology trends of terrestrial International Mobile Telecommunications systems toward 2030 and beyond" [9], published in November 2022, provides an overview of the most relevant technology trends considering the 2030 timeframe and beyond. Based on ITU-R output, 3GPP will be then challenged to develop detailed technical specifications for candidate technologies, which will be later submitted back to ITU-R for evaluation and eventually approval as 6G ITU-R Recommendations.

European roadmap toward 6G, SNS JU

The European Commission is firmly committed to the European industrial leadership in 5G, 6G, and next-generation network technologies. Following a

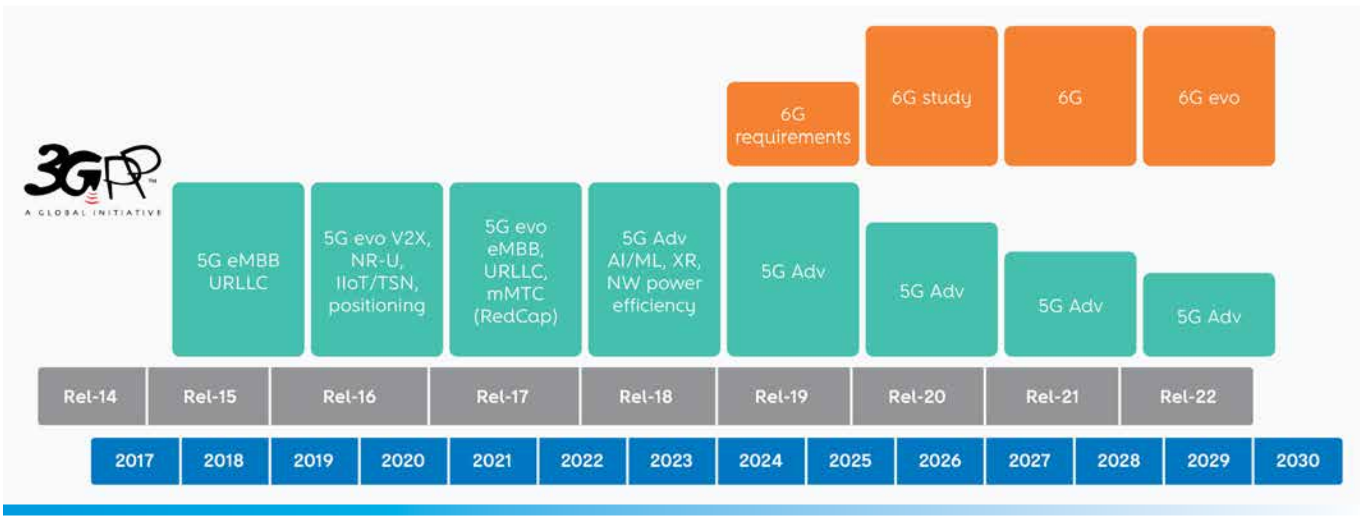


FIGURE 2 – 3GPP plan for 6G standardization (source Ericsson)

similar approach to the one adopted with 5G, the Smart Networks and Services Joint Undertaking (SNS JU) was established in November 2021 to prepare the transition to 6G [18].

The mission of SNS JU is twofold. In the short term it is aimed at boosting the deployment of 5G in Europe across multiple vertical sectors, thus enabling the digital and green transition of the economy and society; in a long-term perspective, the goal is to implement the research and innovation program leading to the conception of 6G and respective standardization around 2025 [18]. The roadmap of SNS JU is aligned with the relevant standardization efforts, especially ITU-R and 3GPP, as shown in **Figure 3**.

SNS JU is jointly funded by the European industry and the EU. The 6G Smart Networks and Services Industry Association (6G-IA) represents the private side of SNS JU. Altice Labs is a full member of 6G-IA since the beginning of 2022.

SNS JU launched its first portfolio of 35 projects in January 2023. The program will run until 2030, covering all steps, starting with the definition of fundamental technology concepts, specification of requirements, standardization, up to technology validation and demonstration. The workplan, illustrated in **Figure 4**, is divided into three phases [16]: 1. Evolutionary 5G, 6G exploration, concepts, definitions; 2. 6G detailed design, system optimization; 3. Pre-commercial 6G systems.

SNS JU activities are structured in four streams, including projects tracing an evolutionary path through further enhancements of 5G technologies (Stream A), or a more disruptive path through promising technological enablers (Stream B), developing experimental infrastructures for technology validation and proofs of concept (Stream C) and executing large-scale trials and pilots in multiple business and industrial vertical sectors (Stream D).

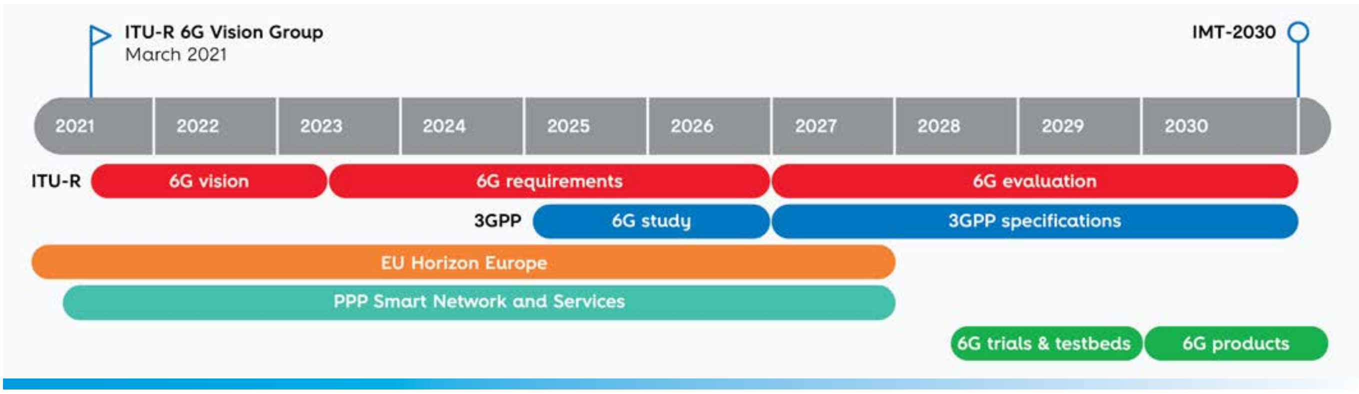


FIGURE 3 – SNS JU alignment with ITU-R and 3GPP (source 6G-IA)

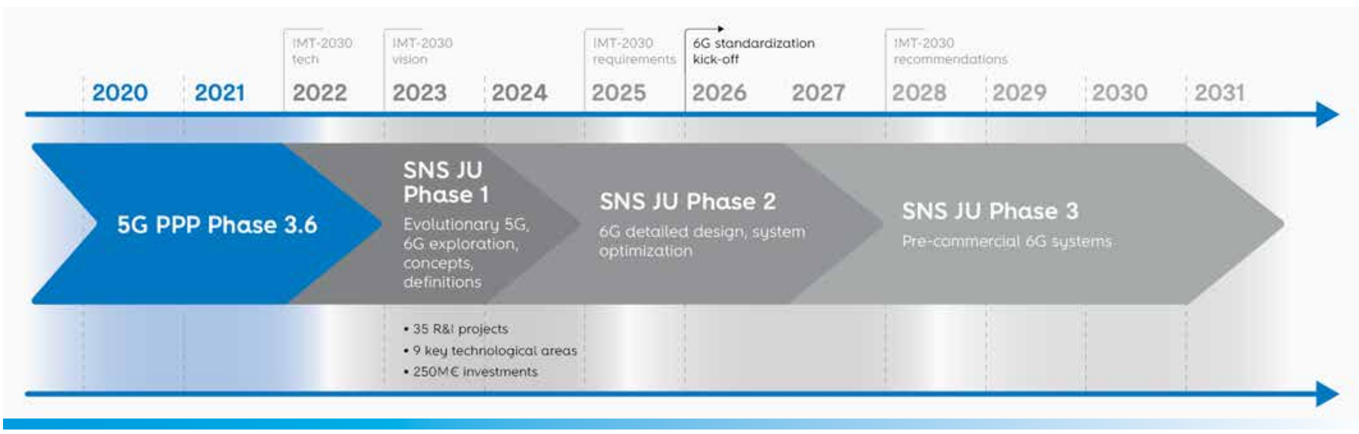


FIGURE 4 – SNS-JU roadmap [18]

As with previous mobile generations, Altice Labs is following the development of 6G since its early stages. In SNS JU phase 1, Altice Labs has been involved in the SNS JU program through the IMAGINE-B5G project [19]. The focus of IMAGINE-B5G is the validation and demonstration of 5G-Advanced technologies, an initial step toward 6G. The project aims to implement an advanced, accessible, secure, and programmable end-to-end 5G platform for large-scale trials and pilots by leveraging "beyond 5G" features.

SNS JU phase 2 was launched in 2023 with a new call for proposals, especially focused on key technological topics complementing the phase 1 retained projects and the support of the upcoming 6G standardization phase. The selected phase 2 projects are planned to start in 2024, including 6G-PATH with the participation of Altice Labs.

One6G

One6G [20] is an initiative promoted by the European industry and research sectors, focused on the transition and roadmap toward 6G. Objectives of one6G include the elaboration of roadmap evolution strategies, the execution of joint innovation and development activities, the definition and harmonization of use cases and business cases, as well as the contribution to regulatory policies. Presently, one6G has more than 100 members, bringing together stakeholders from several research, innovation, and technological areas and vertical domains. The activities of one6G are organized in working groups, covering a wide range of topics including use cases, KPIs, business scenarios, enabling technologies, system architecture, testbeds, and pilots. Altice Labs has been a member of one6G since the beginning of 2023.

Conclusions

6G is expected to be ready for initial deployments around 2030, ten years after the same happened with 5G. Following the process adopted with previous generations of mobile cellular communications, 6G is now being defined, starting with the identification of use cases and KPIs to be fulfilled, in the scope of expected societal and industry demands. 6G shall follow those use case requirements, adopting current research in several enabling technologies, as presented above (see "*Technology Enablers*"). It is a multidisciplinary activity, resulting in an even more complex and software-based architecture. Automation via a wide adoption of Artificial Intelligence will address that increased complexity.

The next six years will observe a crescendo in 6G standardization activity with the first Technical Specifications being expected by 2028, with 3GPP Rel-22. Besides ITU-R and 3GPP, other organizations have been established to contribute to the 6G definition, as is the case of 6G-IA and one6G. Altice Labs is actively following those efforts by being an active member of those organizations and by contributing to 6G research via the work that is being done in the scope of national and international collaborative projects.

References

[1] Next G Alliance, "Roadmap to 6G," 2022. Accessed: Nov. 13, 2023. [Online]. Available: https://www.nextgalliance.org/white_papers/roadmap-to-6g/.

[2] 5G PPP, "5G Public Private Partnership Test, Measurement and KPIs Validation Working Group Whitepaper Beyond 5G/6G KPIs and Target Values", doi: 10.5281/zenodo.6577506.

[3] C. X. Wang *et al.*, "On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 2, pp. 905–974, 2023, doi: 10.1109/COMST.2023.3249835.

[4] 6G-IA, "6G Smart Networks and Services Industry Association (6G-IA)." Accessed: Nov. 28, 2023. [Online]. Available: https://6g-ia.eu/.

[5] 6G Infrastructure Association (6G-IA), "What societal values will 6G address? Societal Key Values and Key Value Indicators analysed through 6G use cases," 2022, doi: 10.5281/zenodo.6557534.

[6] Z. Zhang *et al.*, "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Vehicular Technology Magazine, vol. 14*, no. 3, pp. 28–41, Sep. 2019, doi: 10.1109/MVT.2019.2921208.

[7] "The 6G Architecture Landscape European Perspective", doi: 10.5281/zenodo.7313232.

[8] ITU-T FG-NET2030-Focus Group on Technologies for Network 2030, "Network 2030 Architecture Framework".

[9] ITU-R, "Report M.2516-0 Future technology trends of terrestrial International Mobile Telecommunications systems towards 2030 and beyond," 2022. Accessed: Nov. 2, 2023. [Online]. Available: http://www.itu.int/ITU-R/go/patents/en.

[10] H. Andersson, "Joint communication and sensing in 6G networks." Accessed: Oct. 26, 2023. [Online]. Available: https://www.ericsson.com/en/blog/2021/10/joint-sensing-and-communication-6g.

[11] O-RAN Alliance, "O-RAN Alliance." Accessed: Nov. 13, 2023. [Online]. Available: https://www.o-ran.org/.

[12] H. Ahmadi, A. Nag, Z. Khan, K. Sayrafian, and S. Rahadrja, "Networked Twins and Twins of Networks: an Overview on the Relationship Between Digital Twins and 6G," Aug. 2021. Accessed: Nov. 2, 2023. [Online]. Available: http://arxiv.org/abs/2108.05781.

[13] S. Moreschini *et al.*, "Cognitive Cloud: The Definition." Accessed: Nov. 2, 2023. [Online]. Available: https://www.researchgate.net/publication/361903749.

[14] A. H. Khan *et al.*, "Blockchain and 6G: The Future of Secure and Ubiquitous Communication," Jun. 2021. Accessed: Nov. 2, 2023. [Online]. Available: http://arxiv.org/abs/2106.05673.

[15] G. G. Rozenman *et al.*, "The quantum internet: A synergy of quantum information technologies and 6G networks," *IET Quantum Communication*. John Wiley and Sons Inc, 2023. doi: 10.1049/qtc2.12069.

[16] 3GPP, "3GPP Releases." Accessed: Oct. 26, 2023. [Online]. Available: https://www.3gpp.org/specifications-technologies/releases.

[17] W. Chen *et al.* "5G-Advanced Towards 6G: Past, Present, and Future," Mar. 2023. Accessed: Nov. 2, 2023. [Online]. Available: http://arxiv.org/abs/2303.07456.

[18] "HOME - SNS JU." Accessed: Aug. 4, 2023. [Online]. Available: https://smart-networks.europa.eu/.

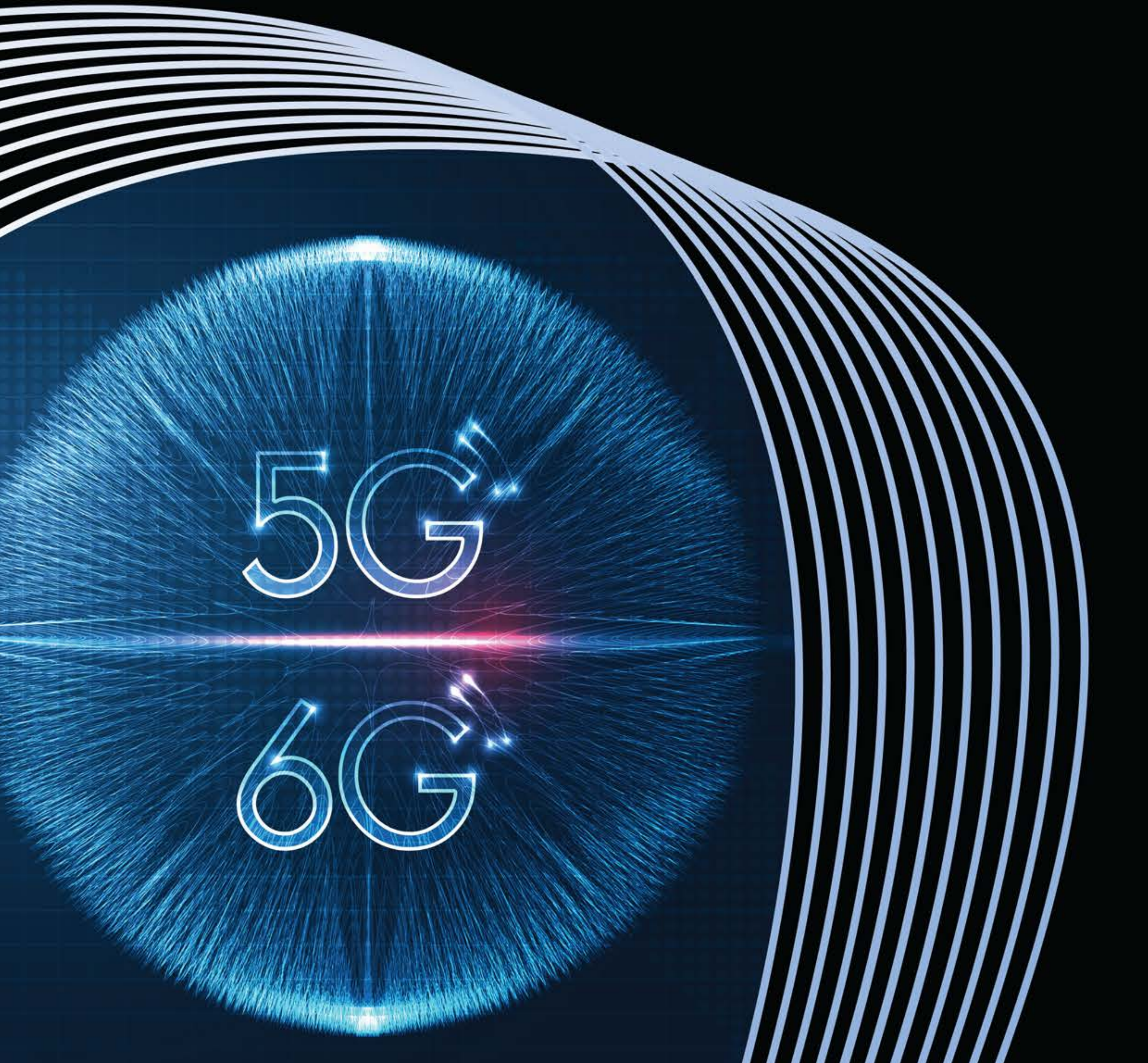
[19] IMAGINE-B5G, "Imagine Beyond 5G," 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://imagineb5g.eu/.

[20] one6G, One6G Vision and Mission. Accessed: Nov. 2, 2023. [Online]. Available: https://one6g.org/.

[21] I. Rahman et al, "5G Advanced: Evolution towards 6G", Jun. 2022, Ericsson White Paper.

DOI: 10.57650/INNOVACTION/2023.02

02 Evolving invisible 5G small-cells to 6G smart radio environments



During the "Invisible 5G" project, it was possible to address the challenges of 5G network densification, with a particular focus on site-specific requirements in urban areas. Altice Labs has been pursuing 5G densification efforts and is laying the groundwork for 6G smart radio environments. The project introduced a novel concept of "invisible" 5G small cells that are designed to seamlessly integrate with existing urban infrastructure, such as lampposts and utility poles, rather than being traditionally placed on building rooftops. These small cells consist of two key modules: the Radio Unit (RU) and the antenna. One of the project's innovations involved the incorporation of Reconfigurable Intelligent Surfaces (RIS) to address issues like power consumption and interference. This makes the small cells highly versatile, and capable of advanced electronic beam steering (T-RIS – Transmit-RIS) to serve micro and macro-coverage areas. This unique hybrid solution, combined with urban integration, represents a significant paradigm shift in mobile communications. Moreover, this innovative approach not only enhances urban landscapes but also offers the potential for short-term adoption in a vast market. This can lead to increased exports in the mobile broadband sector, positioning Portugal as a product supplier alongside its service offerings. The project's evolution from 5G small cells to 6G smart radio environments, powered by RIS, holds great promise for meeting the increasing demands of mobile network deployments.

Authors

Rafael Caldeirinha, Politécnico de Leiria – Instituto de Telecomunicações, rafael.caldeirinha@ipleiria.pt

João Reis, Politécnico de Leiria – Instituto de Telecomunicações, joao.reis@ipleiria.pt

Tiago Oliveira, Politécnico de Leiria – Instituto de Telecomunicações, tiago.s.oliveira@ipleiria.pt

Renato Bispo, Politécnico de Leiria – Laborabory in Design and Arts, renato.bispo@ipleiria.pt

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

José Salgado, Altice Labs, JSalgado@alticelabs.com

Keywords

Invisible 5G; 6G; Smart radio environments; Small cells; Reconfigurable intelligent surfaces; T-RIS; R-RIS; Urban furniture

Introduction

The "Invisible 5G" project aimed to address the challenges associated with 5G network densification as a key driver for enabling 5G. It particularly focused on the challenges arising from site-specific requirements that bring unique challenges to the design and deployment of small cells in urban scenarios at a scale required for future 5G (and 6G) coverage and services. During the ramp-up phase of Portugal's 5G rollout, denser localized coverage will allow 5G networks to move closer to the end-user to support evolving communications technologies, which ultimately will help build smarter and more technologically advanced communities and infrastructures.

It is proposed a state-of-the-art, low-power, small-form-factor, highly concealed (or invisible), ruggedized, weather-proof, and easy-to-deploy 5G small cell. Invisible 5G small cells are, therefore, sought to be world-class and premium compact 5G small cells that are highly concealed and have high levels of integration with existing urban furniture, such as utility poles, lampposts, or even technical cabinets, reducing the need to deploy base stations on building rooftops and facades.

The evolution of 5G small cells into a 6G smart radio environment (which is powered by reconfigurable intelligent surfaces, or RIS) is a promising solution to meet the demand for increasing throughput in 5G and 6G mobile network deployments. Power consumption and two-tier interference issues are bottlenecks that hinder further developments in typical small-cell deployment. Thus, the rationale is to evolve invisible 5G small cells by integrating reconfigurable intelligent surfaces as either transmit (T-RIS) or reflective (R-RIS). In the former, T-RIS will enable advanced electronic and agile 2D beam steering, serving multiple users as a small base station while assisting the macro-coverage users in enhancing their radio coverage by anomalous reflections in an R-RIS configuration. The initial results of the proof-of-concept will also be presented. Such a hybrid T-RIS and R-RIS solution, combined with the high levels of urban integration in the street landscape, makes the evolved invisible 5G small cells a novel paradigm in the mobile communications sector and delivers a unique solution that is timely and topical. Furthermore, it has the potential for short-term adoption in an enormous market and, thus, to further leverage Portuguese exports in the mobile broadband arena as a product supplier other than services.

The paper is structured into seven sections, with an introduction followed by a focus on the need for 5G densification in urban areas and the specific challenges faced in Portugal. It then introduces the invisible 5G small cells and delves into Altice Labs' role in the densification race. The paper concludes by looking ahead to the future, exploring the transition to 6G technology and the integration of Reconfigurable Intelligent Surfaces in separate sections.

The need for 5G densification

The importance of 5G densification in urban areas

Sitting on the cusp of a new decade that is expected to be profoundly impacted by 5G, the fifth-generation wireless network standard is here. The national priority for many countries is to secure a 5G rollout as its spread is pivotal for the connected world. For the first time in mobile communications history, 5G will serve as the backbone of the fourth industrial revolution and, thus, the global pacesetter for the convergence of all connected technologies and bring this technology transformation to fruition in an economic and efficient manner.

It's important to mention that all benefits and capabilities that have been promised will be possible on 5G SA (Standalone) and not on the 5G NSA (Non-Standalone), that generally the telecom operators started to make available. It will still take some time before 5G SA becomes mainstream and achieve wider coverage areas.

Site-specific requirements and challenges

One of the fundamental changes that 5G introduces in the radio environment, is the need for cell densification in urban scenarios. This could be achieved from the combinations of frequency bands (a blend of low- and high-frequency spectrum) that 5G must use to yield its extremely high data rate (throughput), low latency, and ubiquitous coverage. In densely built-up areas where propagation through obstacles, such as buildings and trees, is an issue, operators need to densify their mobile networks with small cells. In the 5G age, it is less a "nice to have" and more of an operational "must-have" to achieve the quality of service (QoS) that consumers and enterprise customers expect.

Despite the slow take of both 5G regulatory issues and spectrum auctions, mobile network operators (MNO) are now referring to their radio usage heat maps to evaluate the geographical areas of high 4G usage to plan their 5G small-cell strategies. Deployment places like busy city centers, motorways and train routes, airports, and shopping malls are all candidates. Outdoor small cells come in many different shapes, sizes, and configurations. For instances, MNOs may want to do their recent investments on 4G (LTE) spectrum acquisition by using carrier aggregation to increase capacity. This implies using three, four, or five different licensed bands at the same time, and they may use multiple-input-multiple-output (MIMO) antenna technology for additional capacity. These requirements multiply the amount of RF (Radio Frequency) hardware at a site. A small cell is defined as a single geographic site and can be made up of radios, antennas, and other equipment. Outdoor small cells can differ from city to city, even street corner to street corner, depending on the requirements of the site, municipal jurisdiction, MNO, or subscriber population and mobility phenomena in the area. The site-specific requirements bring unique challenges to the design and deployment of small cells at the scale required for future 5G networks.

Ramping up 5G rollout in Portugal

The challenge of location means that small cells should be put in the available space, which may not be ideal for placing, e.g., MIMO antennae. Street furniture, used to define objects in public spaces that house small cells in boxes and are considered visually commonplace and acceptable to the public, such as utility poles, bus-stop enclosures, or any other street-level infrastructure that can house wireless equipment, offers highly effective tools to MNO in expanding their networks using small cells, distributed antenna systems (DAS), backhaul and other means to transmit and increase radio coverage to augment network densification. These small cells should be a miniature version of the traditional macro-cell, in the sense that radios and antennas should be somehow compressed into a low-power, easy-to-deploy radio device. Typically, small cells have a range from ten to a few hundred meters and can be incorporated into street furniture, as long as this is equipped with a power source for the radio equipment to operate, and can accommodate power, antenna, and associated fiber and other cabling equipment. This imposes stringent design requirements, and thus, good engineering is crucial to a successful small-cell deployment.

Additionally, the co-existence with e.g., traffic control, on the same traffic light utility pole or on a lamppost, limits the physical space available for small-cell radio equipment, more so as this is typically regulated by local municipality. Also, neighborhood residents will not accept an eyesore to get better radio service. So, pleasing concealment is of uttermost importance.

The concept of invisible 5G small cells

Small-cell concept

The primary objective of the "Invisible 5G" project was to tackle the challenges associated with the densification of 5G networks. This involved the development of compact 5G small-cell base stations designed to be virtually invisible, boasting high levels of concealment for easy deployment on street furniture. The core concept revolved around introducing innovative design solutions capable of seamlessly integrating the radio unit, antennas, and associated fiber and cabling equipment in a camouflaged fashion.

To achieve this, the project relied heavily on effective design and engineering strategies. These strategies encompassed the creation of custom protective casings for antennas, radio units, and infrastructure. The purpose of these casings was multi-faceted: to safeguard internal components from the rigors of weather and vandalism, ensure optimal signal transmission, simplify maintenance procedures, and seamlessly blend with their environment, rendering them inconspicuous.

Key features

According to the project requirements, the small cells in the "Invisible 5G" project were designed with the following key features in mind:

- **Concealed design:** the paramount objective was to achieve a high level of concealment. This was accomplished by seamlessly integrating the radiating systems, radio unit, and associated wiring within customized enclosures or boxes. This design allowed for a harmonious visual integration, ensuring that the small cells could blend into their environment, essentially going unnoticed once installed;

- **Easy deployment:** the small cells were engineered for easy and non-intrusive deployment. They could be quickly attached to existing infrastructure, such as public lighting poles, with minimal mechanical intervention. The solutions presented in the project were designed to be easily reversible, thereby protecting public assets while facilitating installation;

- **Small-form-factor:** the small cells were compact in design, achieved through the use of Altice Lab's Radio Unit, which had dimensions of approximately 352x235x120 mm³. This unit incorporated all the necessary radio frequency electronics and hardware for connecting to the core of the 5G network. Custom antenna designs with reduced footprints and high modularity were also employed, making them adaptable to various deployment scenarios;

- **Heat dissipation:** in addition to their compact and low-power design, the small cells addressed heat dissipation concerns. The proposed solutions included techniques to increase the exposure of hardware components to outdoor air and an arrangement that facilitated natural air flow through convection, thereby managing heat effectively;

- **Ecologic solutions:** environmental sustainability was a crucial consideration in the development of these small cells. The project emphasized an environmentally responsible approach throughout the entire product lifecycle, from material sourcing to disposal. Recycled materials, specifically recycled tire-based compounds, were utilized for producing structural parts and some customizable casings. These materials were employed through additive manufacturing, reducing waste, and providing eco-friendly alternatives to traditional composite materials. The parts were mostly produced using a polymeric compound made of shredded end-of-life tyres and polypropylene (75%/25%). This approach aligns with the project's commitment to addressing environmental challenges in a creative and sustainable manner.

Overall, the "Invisible 5G" project aimed to deliver small cells that were highly concealed, easy to deploy, compact, efficient in heat dissipation, and environmentally friendly, making them a noteworthy advancement in the evolution of 5G network infrastructure.

Integration with urban furniture

The design project was structured around three distinct case studies, each characterized by unique visual and production requirements. This approach ensured that the outer casing could be easily adapted to the specific visual characteristics of each deployment location. Three primary integration and customization strategies were developed within the project:

A - The Standard casing

The Standard casing (**Figure 1**) was developed with the objective of being a compact, orientable, and economical solution for telecommunication poles or urban areas where the visual integration needs are less demanding.

A strategy of embedding the antenna into the technical access cover of the RU was used, making the final product as compact as possible and minimizing the wiring required between the RU and the antenna. This solution abuts the antenna in front of the RU, requiring the use of a unidirectional antenna, which was developed in the project, and which takes advantage of this characteristic to increase its efficiency from the point of view of signal emission.

The main parameter of urban integration in this scenario is the choice of the pole to be used and its location relative to the area where you want to have the best signal. At the same time, the integration of the antennas depends essentially on the choice of the most appropriate place for them to go unnoticed, intending with the adjustment and orientation system proposed to avoid the installation in points of greater visibility.

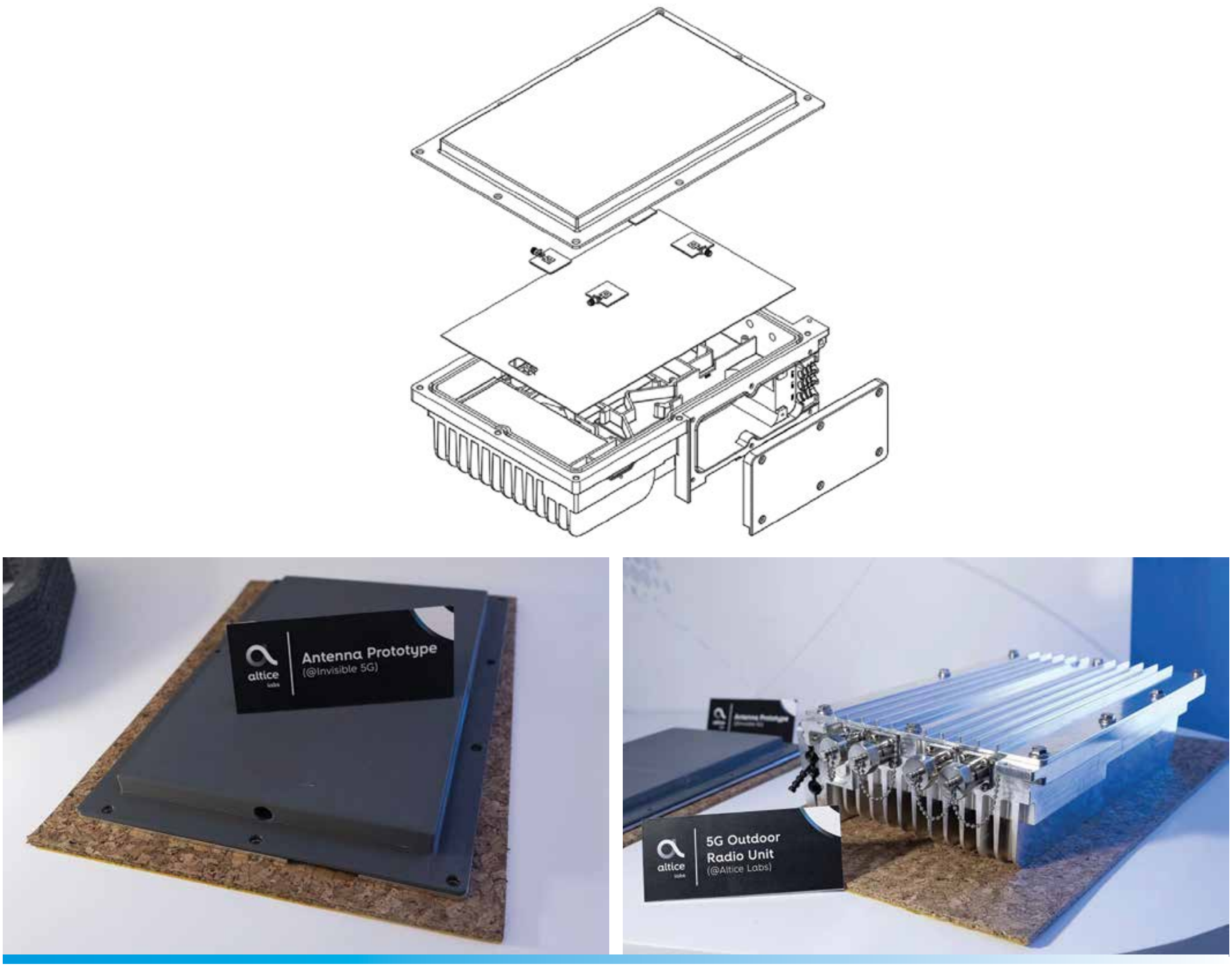


FIGURE 1 – Standard casing solution with embedded antennas

B - The Small informative casing

The second model developed was a Small informative casing (**Figure 2**) that hides the RU and the antenna through a camouflage strategy based on an informative function, which can be used either as an identification, directional or regulatory sign. A functional diagram of this mode is depicted in **Figure 2**.

The relatively small area of the informational support that hides the RU has a diameter of 50 cm, which brings it close to the scale of a traffic sign and allows its use as a support for pictographic language. In this case, the unidirectional antenna attached to the radio unit can also be used or, as alternative, it can be detached from the RU and be positioned on the opposite side of the pole, which allows the use of both unidirectional and omni-directional antennas. The cabling is completely hidden, being carried to the RU from the inside of the pole and connecting to the antenna through the horizontal metal support that connects the two elements to the lamppost.

The customization parameters in this model are essentially the graphic or visual content that may be placed in the Small informative casing, which allows an easy and low-cost customization. This solution is intended for urban environments where concealment is considered pertinent, but where it is necessary to install a solution of moderate cost and easy customization.

FIGURE 2 – Prototype of the Small informative casing

C - The Informative panel

The third model developed (**Figure 3**) consists of an information panel for urban areas with more demanding integration, which allows total concealment of all the elements that make up the system (RU, antenna, and electrical and optical fiber cabling).

This system allows a completely reversible installation, without the need to drill holes in the lampposts, allowing for its installation in public streetlamps with historical value or integrated into urban regeneration processes where there is a greater consistency in the urban equipment installed. The proposed system is based on the use of a customizable information panel, both in terms of shape and content, and a RU protection box designed by scanning the base of the existing poles, produced in recycled composite material through additive manufacturing. This solution presents a higher production complexity and is justified in scenarios where invisibility is considered as particularly relevant, such as historical centers or monuments.

The integration and concealment of the system can be optimized through the customization of the information panel, which although it always has a higher edge at about 2.5 meters from the ground (for the safety and efficiency of the antenna) can be adjusted in its width, to be more suitable to the location, facilitating its placement in circulation areas, squares, or gardens. A snap-in system has been designed to allow easy exchange of the panel.

The safety of the technical components (RU, antenna, and cabling) must be ensured by their integration in the information support, with the antenna and respective RU connection cabling integrated into the information panel and the RU in the base, protected by an external box.

FIGURE 3 – Rendering of the Informative panel (left) and custom protective case (right)

The final demonstration

The project was showcased in two public demonstration events, as depicted in **Figure 4**.

FIGURE 4 – Public demonstrations of the "Invisible 5G" project: (a) 7th Anniversary Event at Altice Labs; (b) Public demonstration at the ESTG/IPL campus

The first event occurred at the Altice Labs facilities in Aveiro during their 7th Anniversary celebration. This event was characterized by a targeted demonstration of a 5G small cell seamlessly integrated into a lighting pole, complete with an information panel. These detailed mock-ups were complemented by a promotional video that effectively highlighted the innovative solutions and outcomes of the project, providing a comprehensive and visually engaging demonstration of the project's achievements.

The second event was held at the Politécnico de Leiria - Escola Superior de Tecnologia e Gestão – Campus 2, and will continue to be operational even after the project's completion. This ongoing deployment serves as a data collection and network metrics platform, offering opportunities for potential network optimization. Furthermore, the sustained deployment at the Politécnico de Leiria offers a valuable opportunity to engage the local industry in their digital transformation efforts. It empowers production plants by providing 5G connectivity, allowing them to conduct testing and experimentation before making substantial investments in their digital infrastructure. This dual role not only supports network optimization but also promotes the integration of 5G technology into local industries, fostering innovation and growth.

Altice Labs: from 5G densification to 6G

Altice Labs has taken an active role researching and developing products and solutions for 5G densification. The company's contributions and vision are pivotal to the ongoing transformation of wireless communication networks, and they are shaping the path toward 6G smart radio environments (**Figure 5**).

Densification, or the strategic placement of small cells and base stations in high-demand areas, is a critical aspect of ensuring robust and reliable 5G connectivity, especially in densely populated regions. Altice Labs' efforts in this arena are multifaceted:

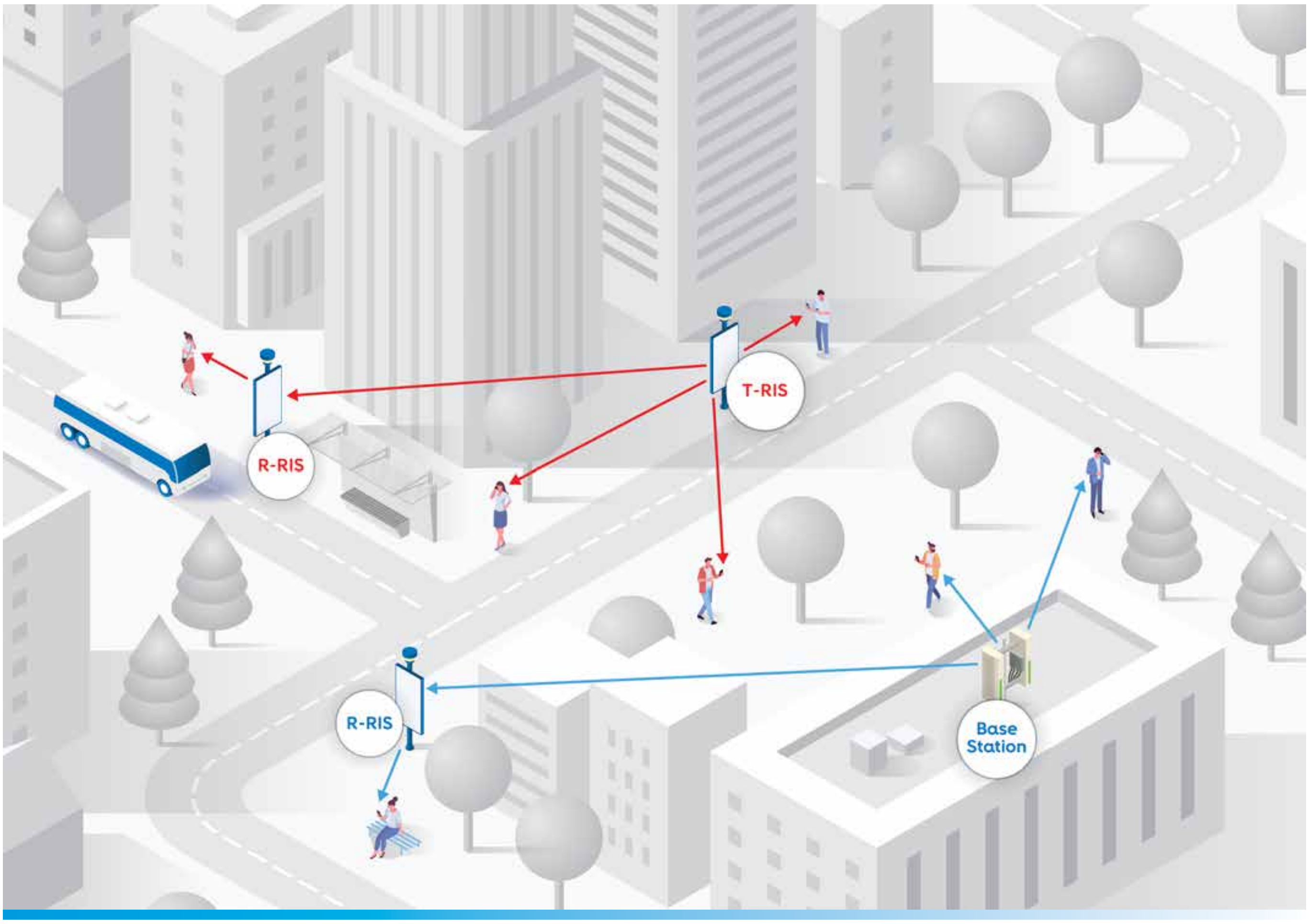


FIGURE 5 – Smart-radio environment scenario using T-RIS/R-RIS, for 6G and beyond

- **Technology innovation:** Altice Labs has invested significantly in research and development to create state-of-the-art, low-power, and highly concealed 5G small cells. These small cells are designed to seamlessly integrate with existing urban infrastructure, such as utility poles and lampposts, while delivering the performance required to meet the demands of 5G and beyond;
- **Urban integration:** the emphasis on integrating 5G infrastructure into the urban landscape sets a new standard for unobtrusive and aesthetically pleasing deployments. This approach not only addresses the technical challenges of densification but also ensures minimal visual impact on cities and communities;
- **6G preparedness:** the commitment to innovation extends to the future with a focus on 6G smart radio environments. By actively exploring the integration of reconfigurable intelligent surfaces (RIS), Altice Labs is laying the foundation for next-generation wireless networks that can deliver even higher throughputs and superior network quality.

The integration of RIS technology holds immense promise. RIS technology can dynamically manipulate electromagnetic waves, enabling advanced beamforming and interference management. This translates to substantially increased data rates and network performance, vital for future applications like augmented reality, holography, and immersive communication. Furthermore, in terms of energy efficiency, RIS can be used to optimize energy consumption, addressing one of the key challenges in wireless network deployments. The energy-efficient operation aligns with sustainability goals and contributes to reduced operational costs.

Evolving to 6G with Reconfigurable Intelligent Surfaces

The 5th generation of mobile network (5G) is being currently deployed worldwide. As natural evolution, the research community is now looking beyond 5G [1-3] and seeking for the next-generation of communications. Some authors [1-5] are already suggesting that 6G will go way further than the typical exploration of more spectrum at high-frequency bands, but it will rather be a convergence of upcoming technological trends driven by exciting, underlying services [5]. Some key enabling technologies for 6G are already being identified [1-3], with RIS being highlighted as a key technology for future wireless networks [3-5].

Within a radio channel and if deployed on a large scale, RIS structures would allow for the modification, on-the-fly, of the properties of the propagation channel (even though it remains physically unaltered). By doing so, it is possible to add, constructively or destructively, the propagating electromagnetic waves and suppress multi-path within the radio channel [10-12]. Like so, the radio channel will play, for the first time, an active role in wireless communications, and the performance of a communication system will not only depend on the intelligence of the transmitter and the receiver but also the intelligence of the radio path [10-12].

To date, plenty of examples of transmit and receive antenna arrays can be easily found in the literature [6-9]. However, most are typically demonstrated on a small scale, limited to antenna applications, with control performed locally. T-RIS and R-RIS, on the contrary, presuppose the use of a large-scale controlling mechanism that not only controls each local surface but optimally configures the most appropriate operation of several spatially distributed RIS, using artificial intelligence and machine learning algorithms to cope with real-time complex network planning [10-13].

Project TERRAMETA [14] joins key players in the design and development to investigate new technologies for 6G and demonstrate the feasibility of Terahertz in RIS for ultra-high-data-rate wireless communications networks. Novel high-performance hardware is being developed, and advanced network analysis/optimizations techniques will be developed.

T-RIS and R-RIS integration

Smart radio environments are revolutionizing the way we connect and communicate, and one of the most promising developments within this realm is the integration of T-RIS and R-RIS. These technologies promise to usher in a new era of wireless communication, ensuring seamless connectivity, improving coverage, and optimizing network efficiency.

Understanding T-RIS and R-RIS configurations

T-RIS (Transmitting – Reconfigurable Intelligent Surfaces): T-RIS is a critical component of smart radio environments. It comprises an array of reconfigurable intelligent surfaces that enable wireless devices to transmit signals more effectively. These surfaces consist of numerous small and controllable elements, which can adjust their properties to optimize signal transmission. T-RIS enhances communication by manipulating the phase and amplitude of transmitted signals, effectively steering them toward their intended targets, improving:

- **Micro coverage:** by customizing signal transmission, T-RIS allows for precise targeting of signals toward individual devices, including smartphones, IoT sensors, and other small, low-power devices. This enables micro-coverage users to enjoy enhanced connectivity and extended battery life;
- **Macro coverage:** T-RIS also improves macro coverage. By optimizing signal transmission,

it can mitigate the effects of obstacles such as buildings, trees, and terrain irregularities. This results in more comprehensive network coverage and reduced dead zones, ultimately enhancing the user experience for a broader audience.

R-RIS (Reflecting – Reconfigurable Intelligent Surfaces): in contrast, R-RIS is designed to enhance the reflection of incoming signals. This technology uses reconfigurable intelligent surfaces that consist of specially designed materials or structures. R-RIS intelligently reflects incoming signals toward their intended destinations, playing a pivotal role in extending macro coverage in smart radio environments:

- **Reflection of signals:** R-RIS intelligently reflects incoming signals, effectively redirecting them toward their intended destinations. This technology is particularly valuable in urban areas with tall buildings and complex layouts, where signals may be blocked or weakened due to obstacles;
- **Eliminating dead zones:** by strategically reflecting signals, R-RIS can fill in coverage gaps, eliminating dead zones within a network. This ensures a more consistent user experience, especially for macro users who may be on the move;
- **Improved penetration:** R-RIS is also capable of improving signal penetration through buildings and structures. This is a critical benefit in scenarios where users need connectivity in indoor environments, such as shopping malls, airports, and office buildings;
- **Reducing interference:** R-RIS not only extends coverage but also helps reduce interference caused by multipath signals and signal reflections. This results in a cleaner and more stable connection;
- **Network load balancing:** R-RIS can be dynamically configured to balance network load by selectively reflecting signals to alleviate congestion in high-traffic areas, thereby improving the overall network's performance.

In summary, this innovation brings us closer to achieving the goal of ubiquitous and reliable wireless communication in diverse settings. Due to their typical planar architecture and associated low power consumptions, such devices can be easily adapted, e.g., to the informative panel developed in the "Invisible 5G" project, as illustrated in **Figure 6**.

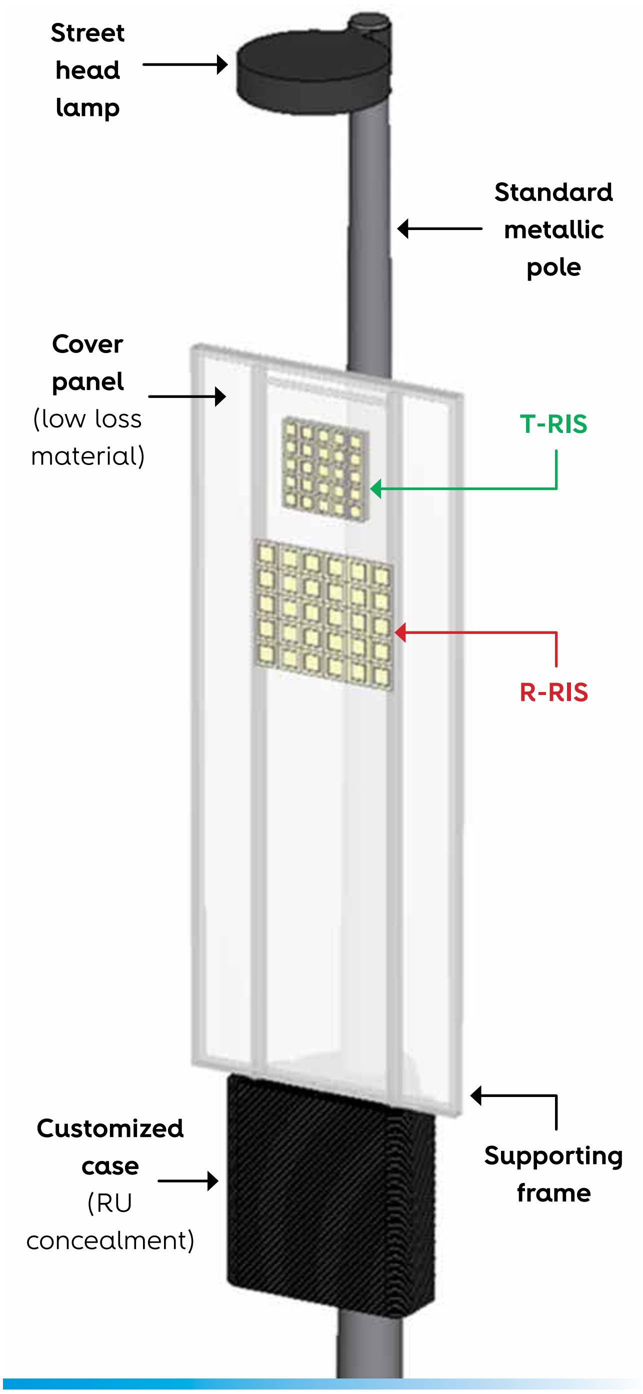


FIGURE 6 – Informative panel of project "Invisible 5G" with T-RIS and R-RIS solutions

Conclusions and future prospects

In conclusion, the concept of densification emerges as a pivotal driver for the successful implementation of 5G and the anticipated arrival of 6G networks. Our work, focused on the specific challenges presented by the need for small-cell deployment in urban environments, signifies a fundamental shift in the evolution of mobile communications technologies in Portugal and beyond.

The drive towards densification is essential for bringing 5G and the future 6G closer to end-users, offering unprecedented opportunities for technological advancements in communities and infrastructures. The proposal for state-of-the-art, invisible 5G small cells was poised to revolutionize traditional base station deployment, promising low-power, small-form-factor, ruggedized, and highly concealed solutions that seamlessly integrate with urban furniture.

Altice Labs' research and development work in the 5G densification race and the introduction of reconfigurable intelligent surfaces (RIS) represent a significant leap forward. This ground-breaking solution not only advances the mobile communications sector but also offers export potential, positioning Portugal as a product supplier on a global scale.

The journey from 5G densification to 6G smart radio environments, coupled with the innovative RIS integration, propels the mobile broadband arena into a new era. This visionary approach promises to redefine connectivity in urban landscapes, underlining its significance in the continuous evolution of wireless communication technologies.

Through collaborative partnerships with academic institutions and research laboratories, the journey toward densification and the evolution of invisible 5G small cells has been enriched by cutting-edge research and continuous learning and adaptation. Central to these pioneering advancements in 5G densification, 6G smart radio environments, and the integration of reconfigurable intelligent surfaces has been a steadfast commitment to collaboration with academia and research labs. This visionary approach recognizes the invaluable contribution of knowledge sharing and cross-disciplinary efforts to drive innovation in the realm of telecommunications.

References

[1] M. Di Renzo *et al.*, "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead," in *IEEE Journal on Selected Areas in Communications*, Nov. 2020, doi: 10.1109/JSAC.2020.3007211.

[2] S. Gong *et al.*, "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," in *IEEE Communications Surveys & Tutorials*, Fourthquarter 2020, doi:10.1109/COMST.2020.3004197.

[3] Z. Zhang *et al.*, "Active RIS vs. Passive RIS: Which Will Prevail in 6G?," in *IEEE Transactions on Communications*, March 2023, doi: 10.1109/TCOMM.2022.3231893.

[4] Y. Wang, W. Zhang, Y. Chen, C. -X. Wang and J. Sun, "Novel Multiple RIS-Assisted Communications for 6G Networks," in *IEEE Communications Letters*, June 2022, doi: 10.1109/LCOMM.2022.3163587.

[5] B. Rana, S. -S. Cho and I. -P. Hong, "Review Paper on Hardware of Reconfigurable Intelligent Surfaces," in *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3261547.

[6] M. H. Dahri, M. I. Abbasi, M. H. Jamaluddin and M. R. Kamarudin, "A Review of High Gain and High Efficiency Reflectarrays for 5G Communications," in *IEEE Access*, 2018, doi:10.1109/ACCESS.2017.2786862.

[7] P. I. Theoharis, R. Raad, F. Tubbal, M. U. A. Khan and A. Jamalipour, "Wideband Reflectarrays for 5G/6G: A Survey," in *IEEE Open Journal of Antennas and Propagation*, 2022, doi: 10.1109/OJAP.2022.3195665.

[8] J. R. Reis, M. Vala and R. F. S. Caldeirinha, "Review Paper on Transmitarray Antennas," in *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2924293.

[9] Reis, J.R.; Vala, M.; Oliveira, T.E.; Fernandes, T.R.; Caldeirinha, R.F.S. "Metamaterial-Inspired Flat Beamsteering Antenna for 5G Base Stations at 3.6 GHz". *Sensors 2021*. doi: 10.3390/s21238116.

[10] Renzo, M.D., Debbah, M., Phan-Huy, DT. *et al.* "Smart radio environments empowered by reconfigurable AI metasurfaces: an idea whose time has come". *J Wireless Com Network 2019*, 129 (2019), doi: 10.1186/s13638-019-1438-9.

[11] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini and R. Zhang, "Wireless Communications Through Reconfigurable Intelligent Surfaces," in *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2935192.

[12] A. Taufique, M. Jaber, A. Imran, Z. Dawy and E. Yacoub, "Planning Wireless Cellular Networks of Future: Outlook, Challenges and Opportunities," in *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2680318.

[13] L. Dai *et al.*, "Reconfigurable Intelligent Surface-Based Wireless Communications: Antenna Design, Prototyping, and Experimental Results," in *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2977772.

[14] Project TERRAMETA, Terahertz Reconfigurable Metasurfaces for ultra-high rate wireless communications, HORIZON-JU-SNS-2022-STREAM-B-01-02. Accessed: Nov. 2, 2023. [Online]. Available: https://www.it.pt/Projects/Index/4833.

[15] 5G Invisible: Virtually Invisible Small-cells for 5G Densification. Accessed: Sep. 12, 2023. [Online]. Available: https://www.alticelabs.com/innovation/rdi-projects/5g-invisible/.

DOI: 10.57650/INNOVACTION/2023.03

03 50G TDM PON and 50G TWDM-PON, challenges on the migration from current PON technologies

As the demand for higher bandwidth and increased traffic requirements in access networks rises, the development of next-generation passive optical networks becomes crucial. Two potential solutions currently being explored are 50G TDM PON and 50G TWDM PON. However, the migration from current PON technologies to these upcoming architectures presents several challenges. One of the main challenges is the coexistence of the upcoming PON technologies with the existing G-PON and XG(S)-PON systems. This requires careful planning and coordination to ensure seamless integration between the upcoming PON technologies and the existing G-PON and XG(S)-PON systems. This paper will discuss and analyze the challenges in developing 50G TDM PON and 50G TWDM PON.

Authors

Cláudio Rodrigues, Altice Labs, claudio-e-rodrigues@alticelabs.com

Ana Tavares, PicAdvanced, ana@picadvanced.com

Francisco Campos, Altice Labs, francisco-j-campos@alticelabs.com

João Ventuzelos, Altice Labs, joao-p-ventuzelos@alticelabs.com

Duarte Viveiros, Altice Labs, carlos-d-viveiros@alticelabs.com

Francisco Rodrigues, PicAdvanced, francisco@picadvanced.com

António Teixeira, PicAdvanced, teixeira@picadvanced.com

Keywords

50G TDM PON; 50G TWDM-PON; 25GS-PON; XGS-PON; GPON; Wi-Fi 6; Wi-Fi 7; OLT; ONU; PON technologies

Beyond 10G PON - 25GS-PON & 50G-PON

The ITU-T released the ITU-T G.9804.1 at the end of 2019, which defines the requirements for High-Speed Passive Optical Networks (HSPONs) [1].

The basic architectures of higher-speed PON (HSP) systems can be separated into TDM/TDMA-based systems and PtP-based systems. Connecting multiple OLT (Optical Line Terminals) CTs (Channel Terminations) via a Wavelength Multiplexer (WM) in a high-speed multi-channel PON system, such as 50G TWDM PON (Time- and Wavelength-Division Multiplexer PON), is possible. An OLT in a 50G TDM PON (Time-Division Multiplexer PON) is a higher-speed PON system with only one channel in each direction [1].

The higher-speed PON scenarios now support a new PON-based 5G Mobile FrontHaul (PON-MFH) service category. The transportation between the control unit (CU) and a remote unit (RU) is provided by a network of OLTs and ONUs (Optical Network Units). With the elimination of the quiet window and the cooperative DBA (Dynamic Bandwidth Allocation) function, PON has been enhanced to offer ultra-low latency [1].

The 50G TWDM PON is a multi-wavelength network whereby each channel pair can be shared with multiple ONUs using the TDM/TDMA scheme. Each OLT CT has a channel pair with a downstream wavelength and a different upstream wavelength, in a point-to-multipoint configuration and using TDM/TDMA mechanisms. Its architecture implies the usage of a Wavelength Multiplexer at the OLT, which can either be integrated into multiple CTs or externally used. The total network capacity in this type of architecture depends on the number of OLT CTs supported, and typically each CT/ONU has a similar capacity as a 50G TDM PON. However, it is possible to take advantage of the multiwavelength feature and increase its capacity by channel bonding. The 50G TDM PON can be seen as a special case of a 50G TWDM PON with a single wavelength channel, thus not needing any WM device. It should be capable of supporting a nominal line rate per fiber of approximately 50 Gbit/s in the downstream direction and approximately 50 Gbit/s in the upstream direction. Meanwhile, a 50G TDM PON ONU must support a maximum service rate of approximately 40 Gbit/s. Additionally, selecting an asymmetric nominal line rate combination per wavelength channel should be possible, such as 25 Gbit/s upstream, 12.5 Gbit/s, or 10 Gbit/s upstream [1]. For an HSP system to coexist successfully, it should be able to reuse existing legacy ODN (Optical Distribution Network) PON and operate in a spectrum that is not already occupied by legacy PONs in a particular deployment. However, the new 50G-PON could re-use the spectrum allocated to legacy PON systems if it does not coexist with those legacy PON systems on the fiber. Some approaches include multi-rate receivers to facilitate spectrum reuse by using an ordinary wavelength band. Coexistence over the entire end-to-end ODN, including the feeder fiber, should be allowed by HSP systems through the use of a Coexistence Element (CEx) or equivalent Wavelength Division Multiplexer (WDM). Additionally, 50G TDM PON and other higher-speed TDMA systems can integrate a multi-PON module into the OLT PON port while coexisting with a legacy PON system [1].

HSP systems must enable technology migration on existing infrastructure without prolonged service interruptions. They should also be capable of upgrading individual customers' services on-demand to maintain service operations.

To achieve a migration path, three options are available. These options differ in their level of flexibility:

- A straight two-step full migration to HSP refers to a migration process involving two steps. The first step involves migrating from GPON to XG(S)-PON, which is followed by the second step of upgrading to HSP. Before starting with the HSP upgrade, a full migration from GPON to XG(S)-PON is necessary. This can be achieved by removing all the GPON from the ODN and re-using the GPON wavelength windows to enable HSP technology to coexist with XG(S)-PON. This scenario allows for the coexistence of two PON technologies simultaneously;

- Direct migration to HSP involves transitioning from GPON to 50G TDM-PON, and requires an HSP system that can work alongside GPON through a double PON technology coexistence;
- All-embracing migration to HSP involves the coexistence of GPON, XG(S)-PON, and HSP, **Figure 1**, which is quite encompassing. It poses a significant challenge due to the limited optical spectrum available and the reduced inter-band guard band between the three PON technologies. Essentially, a triple coexistence is required, which can only be effectively managed through proper support systems, technician tools, and increased OLT port and ONU-type inventory [1].

It is imperative that legacy ONUs and OLTs remain unchanged during any migration, including coexistence, and should not require additional wavelength filters to protect them from HSP signals. To avoid truck rolls to many locations of the ONU [1], additional filtering should ideally be carried out at the OLT, where access may be more accessible, rather than at the ONU [1].

In October 2020, a 25G symmetric PON multi-source agreement (25GS-PON MSA) was signed to promote and accelerate the development of 25GS-PON. The MSA Group established the 25GS PON specification to bridge the gap between 10G XGS-PON and 50G PON in the ITU-T. The 25GS-PON MSA Group created a specification for 25GS-PON that incorporates optical specifications based on the IEEE 802.3ca 25G EPON standard, adapting it to be compatible with ITU-T networks, along with a Transmission Convergence (TC) layer that is an extension of XGS-PON [2].

A 25G TDM PON ONU is expected to have the capability to accommodate a downstream service rate of up to 25 Gbit/s and an upstream nominal line rate that is also symmetrically set at 25 Gbit/s. It will also be possible to support an asymmetric 10 Gbit/s in the upstream.

The coexistence scenario is becoming increasingly difficult, and spectral scarcity has become a reality with the proliferation of multiple generations of PON technologies.

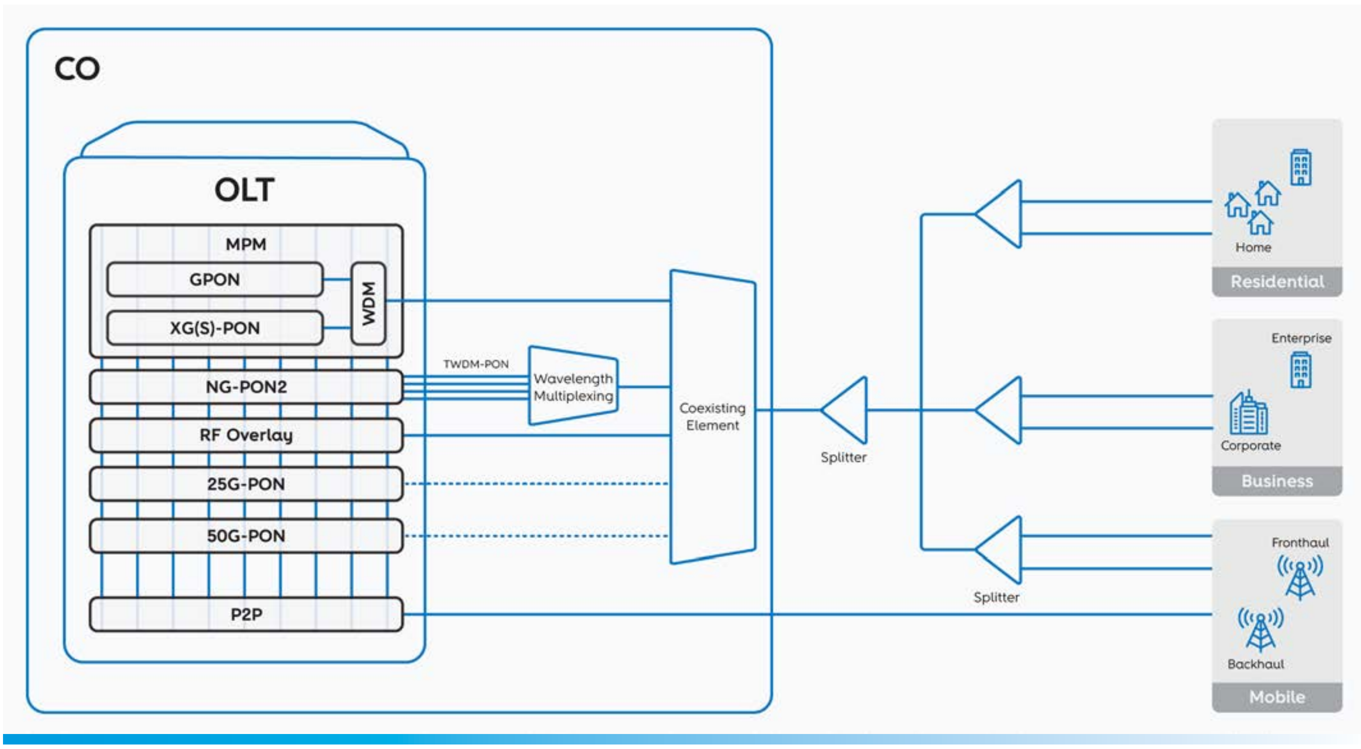


FIGURE 1 – Coexistence scenario, G-PON, XG(S)-PON and HSP (CO – Central Office)

Applications driving higher bandwidth demand

The main driver of Digital Transformation is connectivity, but as the transformation progresses, additional bandwidth is needed as more services are supported. To accommodate the existing use cases, including online gaming, education, e-health, and the persistent dependency on cloud applications in enterprise environments, the networks within residential and commercial buildings, as well as the fixed access network and transport network, will be enhanced to provide sufficient bandwidth and quality of service. Furthermore, the enhancement aims to support emerging, high-bandwidth applications, such as cloud-based virtual reality (VR) and augmented reality (AR), and uncompressed 4K and 8K videos (**Figure 2**). In addition, a fundamental principle of future networks is that passive fiber connections and active electronics will lead to a sustainable network that consumes less energy and reduces the carbon footprint of network operators. A 50G PON not only meets the bandwidth requirements of consumer applications, including virtual reality, augmented reality, online gaming, videoconferencing, and 8K video, but also allows operators to serve business customers. It allows a provider to divide four 10G Ethernet connections among a variety of businesses using 50G PON. Furthermore, 50G PON technology is ideal for deployments of Passive Optical LAN (POL), where fiber can provide desktop connectivity with less power consumption, rack space, and cooling than traditional point-to-point Ethernet architectures.

Wi-Fi 6E and Wi-Fi 7 are expected to increase the demand for bandwidth, so 50G PON will also be used in backhaul for public Wi-Fi hotspots and private wireless LANs. With Wi-Fi 6, an individual user can reach a maximum speed of 9.6 gigabits per second, while with Wi-Fi 7, that speed increases to almost 40 gigabits per second. The evolving 50G PON standard also supports very low latency and jitter levels similar to those of Wi-Wire 7. There is an actual demand for gigabit plans in different markets, demonstrating the value of rolling out XGS-PON. With increasing traffic, XGS-PON becomes a more valuable option for consumers, as GPON networks can become overloaded and fail to deliver 1 Gbit/s of quick bandwidth for speed tests. The in-home connectivity problems associated with

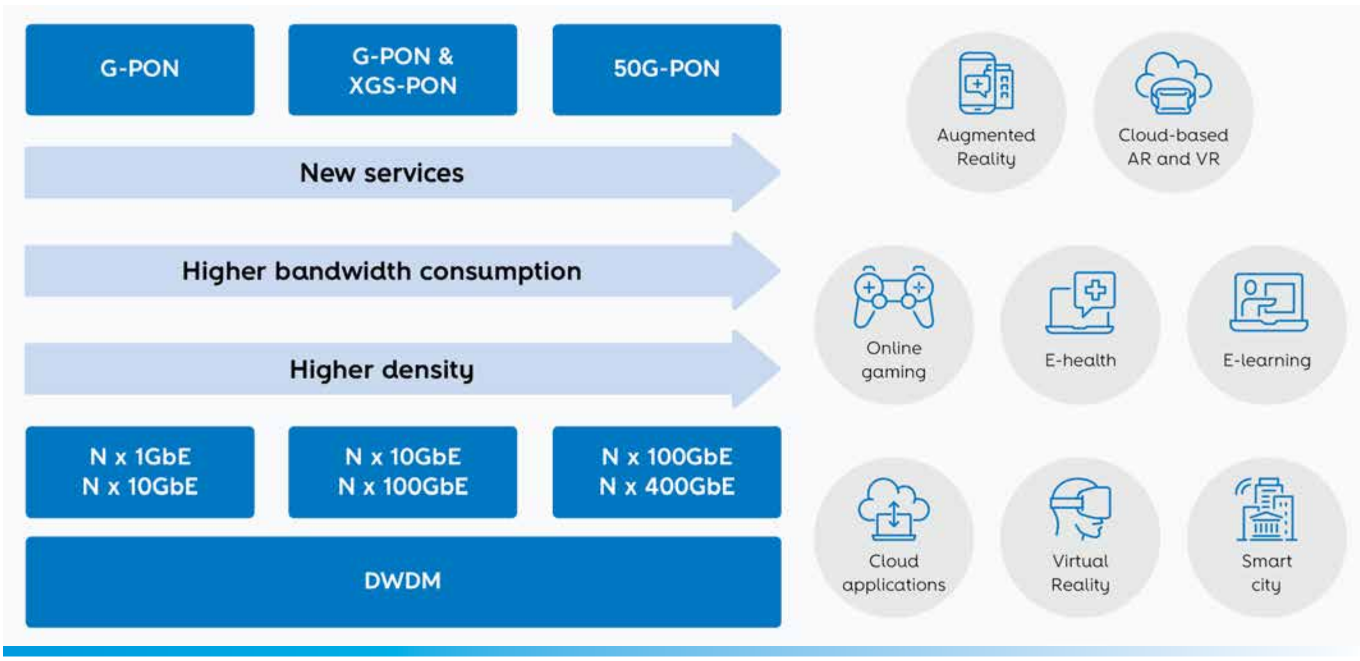


FIGURE 2 – Application driving high bandwidth on Optical Access Networks

XGS-PON require new technologies, such as Wi-Fi 7, to achieve multigigabit speeds. The introduction of Wi-Fi 7 and the rollout of XGS-PON and 50G-PON technologies can enable gigabit speeds for many devices in the home [3].

As compared to Wi-Fi 6, Wi-Fi 7 provides higher data density (4K QAM – Quadrature Amplitude Modulation) and higher channel bandwidth (320 MHz) than previous Wi-Fi generations. As a result of 4K QAM, each symbol is capable of carrying 12 bits rather than 10. This results in a 20% improvement in throughput [3]. Standard 8K video streaming is not likely to make consumers want the higher speeds that XGS-PON and 50G-PON can offer. First of all, 8K TV sets are not very common and will probably stay that way, at least for a while. The bandwidth needed for 8K video, even with multiple streams, is also lower than the gigabit and multi-gigabit speeds that XGS-PON and even 50G-PON can provide. An 8K stream with today's encoding technologies only needs about 50 Mbit/s.

Consumer demand for more bandwidth may be driven by naked-eye 3D videos or 3D videos that do not require viewing glasses. 8K naked-eye 3D video would require 160 Mbps for a single view, which is still below the gigabit and multigigabit speeds of XGS-PON. Multi-view 3D video (i.e., when 3D video can be seen from different viewpoints) increases bandwidth requirements. For example, 8K naked-eye 3D video with nine viewpoints may require approximately 1.1 Gbit/s bandwidth [3].

The following factors may contribute to higher bandwidth requirements in virtual reality (VR) applications [3]:

- VR video requires high-resolution video since it is near the eye. Cloud gaming also requires high-resolution video;
- VR video should have a higher resolution than what the viewer sees. For example, 4K VR video may require 8K video from the network;
- The frame rate for VR video must be at least 60 frames per second to reduce nausea and increase interactivity, which would also increase bandwidth requirements. High frame rates may also be advantageous for cloud gaming;
- VR that requires more interactivity may require more than a gigabit of throughput, which 50G-PON and XGS-PON can provide;
- Gaming in the cloud requires less bandwidth. For example, GeForce NOW requires 15 Mbps for 720p at 60 frames per second and 45 Mbps for 4K at 120 frames per second;
- VR and cloud gaming also require low latency. Extreme VR requires less than 10 milliseconds of streaming latency, which can be achieved through XGS-PON. Cloud gaming requires less than 40 milliseconds of latency to NVIDIA's data center, which can be achieved through GPON networks.

Forecasting the 25G-PON and 50G-PON[1]

Technological diffusion model

Early models of technology diffusion used the metaphor of epidemic spread to represent the process of diffusion [4], [5] [6]–[8]. As technology has developed in the marketplace, it has been compared to the spread of disease in the population. Innovations have historically diffused in a certain way, regardless of their nature. Based on the logistic model, the rate of change of the process is proportional to both the fraction of the market that has been penetrated by the technology

1 This section of the article is based on the following conference article: "Cláudio Rodrigues, Francisco Rodrigues, Cátia Pinho, Nuno Bento, Marlene Amorim, António Teixeira, " 25G and 50G Optical Access Network Deployment Forecasts using Bi-Logistic curves", Proceedings Optical Fiber Communication Conference (OFC) 2021, 6–11 June 2021, Washington, DC, United States"[11]

and the fraction that remains to be penetrated [9]. This process is analogous to the diffusion of a new and technologically advanced product. In the logistic model, the initial sales of a new technology are difficult despite its potential and the size of the potential market. During the early stages, there may be institutional barriers, and the survival of the technology and its supplier is uncertain. In addition to significant risks associated with adoption, as applications grow, so do knowledge and understanding of technology as well as its benefits. Support infrastructure is also improved. From this point, penetration rapidly grows. At some point, after 50% penetration, further market diffusion becomes increasingly difficult, thus causing the rate of penetration to decrease. In the GPON market, this is observed as the uncertain evolution of the legal and regulatory framework of the FTTH network, as well as the general knowledge of fiber, GPON technology, and uncertainties in the optical market, presented significant risks at the beginning of the network. The penetration of GPON and FTTH infrastructure increases as knowledge of these infrastructures grows. Recently, the aim of model formulation has centered on modeling both the growth processes and the random environmental effects associated with them, i.e., the market uncertainties, as they can occur on the 25G and 50G-PON networks. Many dynamic real phenomena show a sigmoidal type pattern in their behavior, Sigmoidal curves that have multiple inflection points [10]. As part of the analysis of the Global GPON ONT/ONU market, a bi-logistic model is presented that describes the two phases of its logistic development. Meyer P. describes this system as "Bi-logistical" in [11] because it has two logistic growth pulses that exist simultaneously or sequentially. An analysis of time-series data sets with two well-defined Logistic growth curves can be divided into two, and each set will be modeled using a separate Logistic function in the case of a system with two well-defined Logistic growth curves.

The OMDIA market share 2Q20 [15] presents the units of ONT/ONU - 2.5G GPON on a global level from 2008 to 2019. The data was modeled using a nonlinear regression to form a Bi-Logistic curve [11], **Figure 3**.

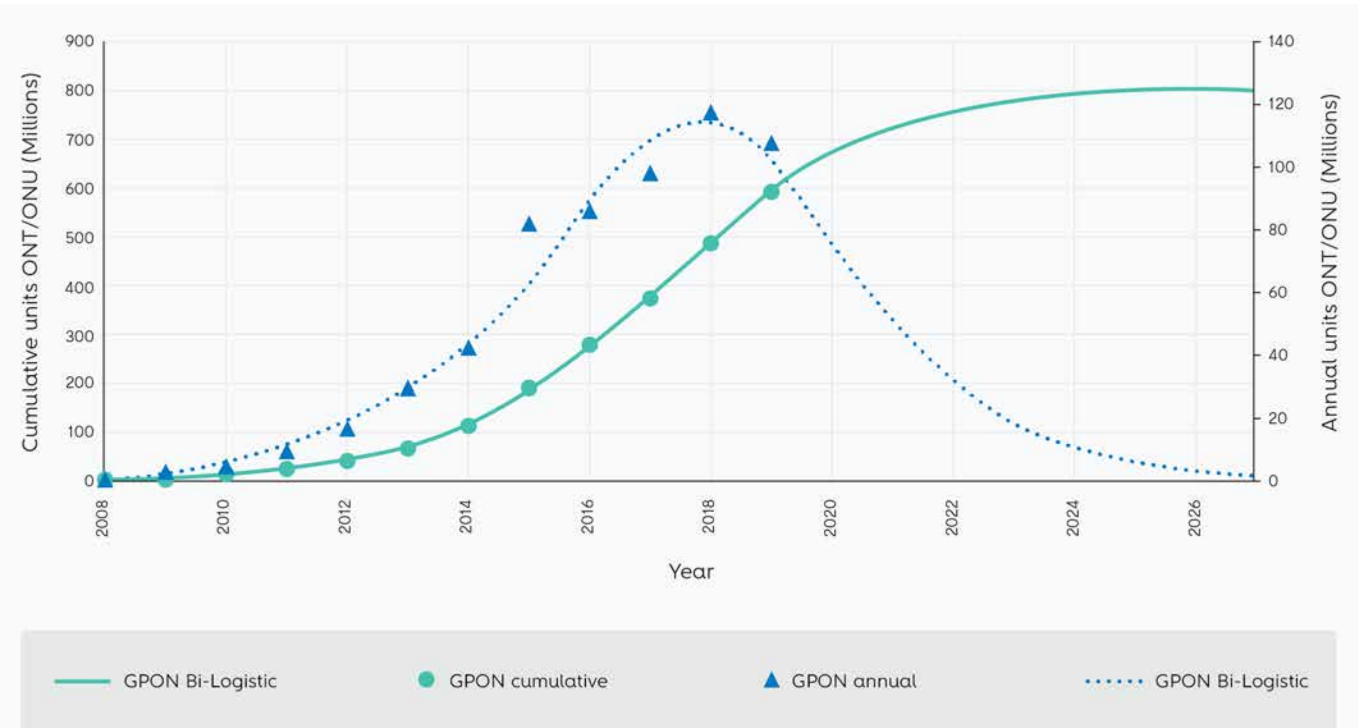


FIGURE 3 – Forecast for the Global Vendor Units ONT/ONU – 25GS-PON & 50G-PON using Bi-Logistic

In summary, the GPON data was modeled to a bi-logistic function that resulted in an r-squared value of 0.999. This model allowed the identification of two phases of the market. The first logistic curve showed a market penetration of 37 million units, with a duration of 5.8 years to grow from 10% to 90%, a growth rate of 0.748, and an inflection point in the middle of 2012 (4.6 years). The second component of the function exhibited a market penetration of 765 million units, with a characteristic duration of 7 years, a growth rate of 0.62, and an inflection point at the end of the first quarter of 2017 (9.3 years). Thus, the total GPON market is forecasted to be 802 million units.

Assuming that the future 25GS-PON and 50G-PON ONT/ONU markets will behave similarly to the historical market diffusion of GPON, different forecast scenarios were defined and presented in **Table 1**.

Some companies have already presented commercial solutions for 25GS-PON, while 50GS-PON is expected to be deployed in 2025 [10].

The forecast scenarios (FS) presented in **Table 1** were defined based on the assumptions that the 10 Gbit/s market is widely expanding beyond 2021. **Figure 4** (see on the next page) illustrates the forecast for the Bi-Logistic for 25G-PONs as well as 50G-PONs. The adoption of 25GS-PON technology to support 5G wireless networks was expected to grow significantly after 2022. However, a market split with the 50G-PON after 2025 will result in a decrease in market share for the 25GS-PON. Considering that GPON is primarily used for residential applications and that this technology is widely used worldwide and that the 10 Gbit/s market has taken hold, it is reasonable to estimate that the 25GS-PON market potential will be 5% and the 50G-PON market will be 10% in the first phase.

The 25GS-PON will be competitive in the first phase against the 10 Gbit/s market, despite having a significant disadvantage in terms of cost compared to the 10 Gbit/s optics market. Since the 25G-PON technology is leveraging existing technologies, it can be viewed as a middle step toward

<table>
<tr><th colspan="2" rowspan="2">Forecast scenario</th><th colspan="3">1st curve</th><th colspan="3">2nd curve</th></tr>
<tr><th>Market potential</th><th>Characteristic duration</th><th>Inflection time</th><th>Market potential</th><th>Characteristic duration</th><th>Inflection time</th></tr>
<tr><td rowspan="2">50G-PON</td><td>FS1</td><td rowspan="2">10% of the total GPON</td><td colspan="2">Same as found for GPON market</td><td rowspan="2">20% of the total GPON</td><td colspan="2">Same as found for GPON market</td></tr>
<tr><td>FS2</td><td>20% higher growth, 4,63 years</td><td>Same as GPON market</td><td>20% higher growth, 5,35 years</td><td>Same as GPON market</td></tr>
<tr><td rowspan="2">25GS-PON</td><td>FS3</td><td rowspan="2">5% of total GPON</td><td colspan="2">Same as found for GPON</td><td rowspan="2">10% of the total GPON</td><td colspan="2">Same as found for GPON</td></tr>
<tr><td>FS4</td><td>20% higher growth, 4,63 years</td><td>Reaches inflection time one year earlier</td><td>20% higher growth, 5,35 years</td><td>Reaches inflection time one year earlier</td></tr>
</table>

TABLE 1 – Forecast scenarios for the global market for the Bi-Logistic 25GS-PON and 50G-PON ONT/ONU

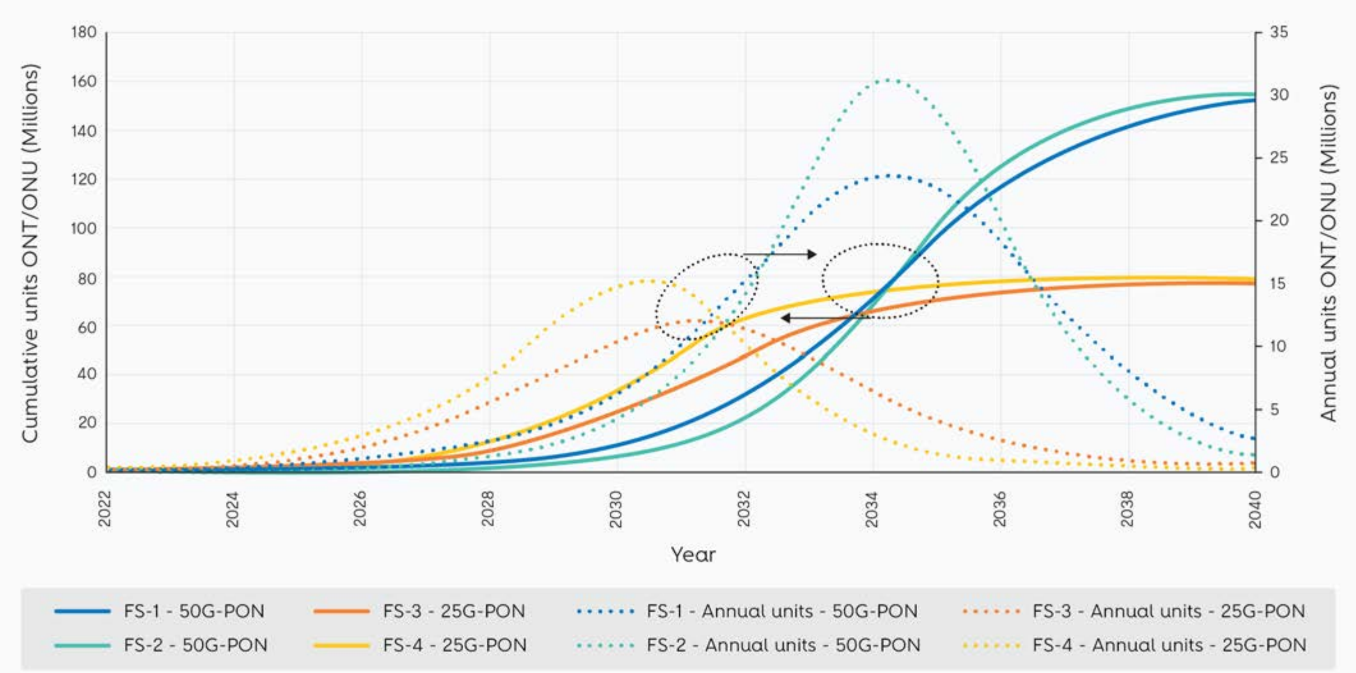


FIGURE 4 – Forecast for the Global Vendor Units ONT/ONU – 25GS-PON & 50G-PON using Bi-Logistic

a higher-speed PON. The delay and squeeze of 5G investment plans by several operators will benefit the 50G-PON technology. Because the second phase of the 25GS-PON occurs in the first phase of the 50G-PON, the 50G-PON will be adversely affected by the first phase of the high-cost optical system. The second phase should therefore be assessed in terms of a 20% market potential, given the decrease in optics costs, as well as the widespread use of fiber in cell sites and the support provided by this standard for PON-MFH.

The 25GS-PON and 50G-PON technologies are not intended for residential customers but for X-Haul connections and enterprises. This results in limited market potential for these technologies. However, since they can benefit from the existing FTTH/B infrastructures, the growth is expected to happen faster, resulting in a time inflection. Compared to the 25GS-PON market, the 50G-PON market forecast is larger because the 25GS-PON is competing against the 10 Gbit/s market, and PON-MFH supports 50G-PON. Because 25G Ethernet optical technology is already available, 25GS-PON can reach the market much faster and earlier, but with a smaller market potential, mostly in EMEA and North America.

Conclusions

A significant part of the evolution of PON technologies has been its ability to adapt to support residential services, Wi-Fi, high resiliency services, high bit rate services, and RAN applications. Thus, the most adaptable telecom equipment, including OLTs and ONUs, is the most widely distributed technology and the most adaptable telecom equipment. There is a clear need for different types of OLTs and ONUs with different capabilities, transport speeds, and support services. It is noticeable that, depending on the circumstances, PON technology can meet the requirements.

PON technology can provide a range of services, such as data, voice, and video, at a low cost. It also has the potential to provide higher bandwidth, improved reliability, and low latency. Furthermore, PON technology is highly scalable and can be easily upgraded to meet future needs.

References

[1] ITU-T, "ITU-T Recommendation G.9804.1- Higher speed passive optical networks – Requirements," 2019.

[2] 25GS-PON MSA Group, "25GS-PON MSA Group," 2020. Accessed: Dec. 20, 2020. [Online]. Available: https://www.25gspon-msa.org/.

[3] S. Wilson, "XGS-PON : Maximizing the Value in the Consumer Segment," 2023.

[4] R. Pearl and L. J. Reed, "A Further Note on the Mathematical Theory of Population Growth," *Proc Natl Acad Sci*, vol. 8, no. 12, pp. 365–368, 1922, doi: 10.1073/pnas.8.12.365.

[5] M. Pearl, "On the Summation of Logistic Curves Author (s): Lowell J . Reed and Raymond Pearl Source : Journal of the Royal Statistical Society , Vol . 90 , No . 4 (1927), pp . 729-746 Published by : Blackwell Publishing for the Royal Statistical Society Stable U," vol. 90, no. 4, pp. 729–746, 2011.

[6] Z. Griliches, "Hybrid Corn : An Exploration in the Economics of Technological Change A," *Econometria*, vol. 25, no. 4, pp. 501–522, 1957.

[7] F. M.Bass, "a New Product Growth Durables," *Manage Sci*, vol. 15, no. 175, pp. 215–227, 1969.

[8] P. A. Geroski, "Models of technology diffusion," *Res Policy*, vol. 29, no. 4–5, pp. 603–625, 2000, doi: 10.1016/S0048-7333(99)00092-X.

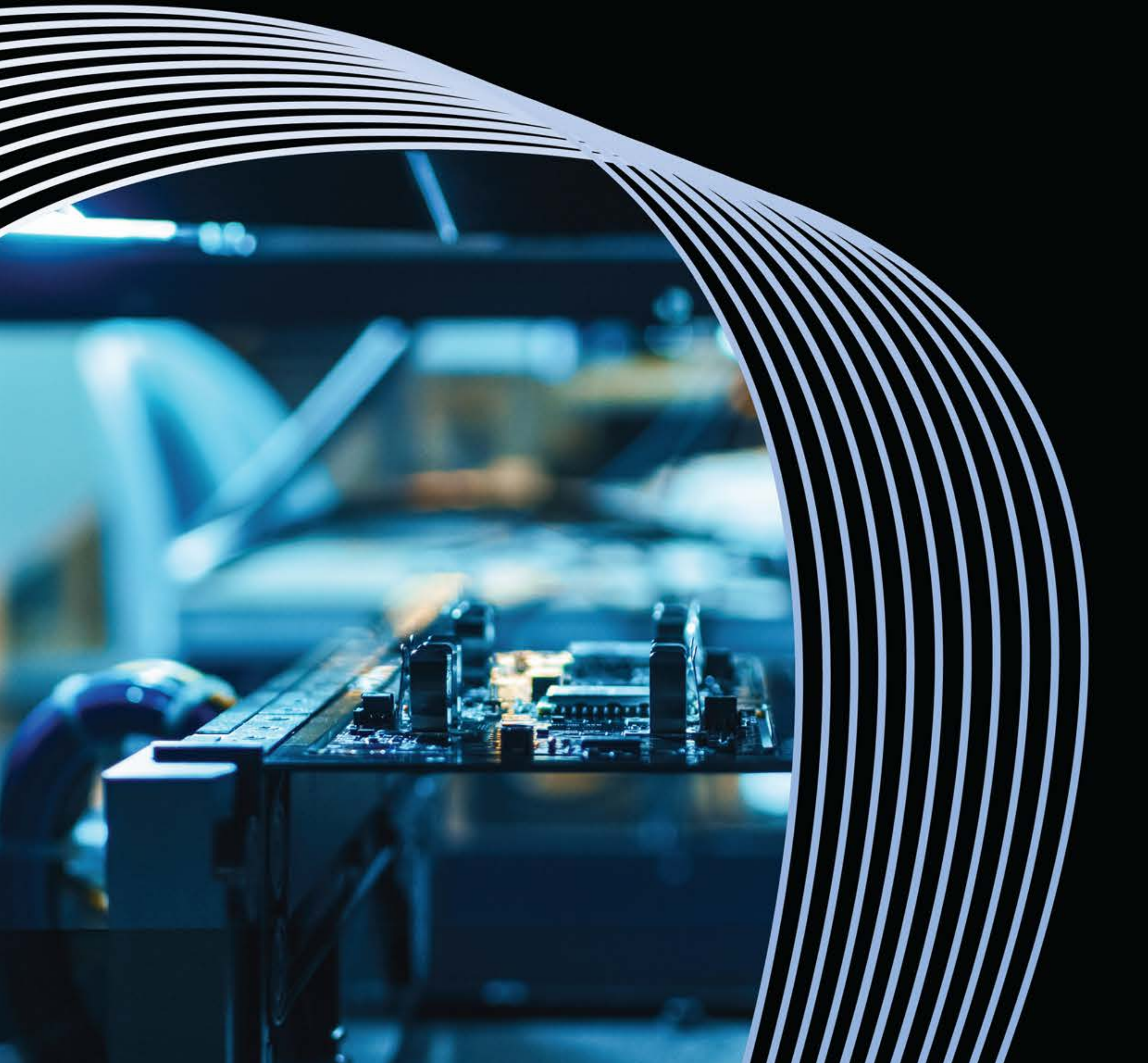
[9] Alan L. Porter, A. Thomas Roper, Thomas W. Mason, Frederick A. Rossini, and Jerry Banks, *Forecasting and Management of Technology*. New YorK: John Willey & Sons, inc., 1991.

[10] D. Zhang, D. Liu, X. Wu, and D. Nesset, "Progress of ITU-T higher speed passive optical network (50G-PON) standardization," *J Opt Commun Netw*, vol. 12, no. 10, p. D99, 2020, doi: 10.1364/jocn.391830.

[11] C. Rodrigues, F. Rodrigues, C. Pinho, N. Bento, and M. Amorim, "25 and 50G Optical Access Network Deployment Forecasts using Bi-Logistic curves," in *Optical Fiber Communication Conference (OFC) 2021*, 2021, pp. 10–12.

DOI: 10.57650/INNOVACTION/2023.04

04 Automated testing in telecommunications: enhancing OLT software quality assurance



Nowadays, telecommunications play a central role in people's lives. Their failure has the same or higher impact than other critical services, such as energy or water supply.
The complexity of the telecommunications equipment, namely the Optical Line Terminal (OLT), has increased over the past years with the launch of new technology standards (100/400G Ethernet, XGS-PON, 10/25/50G xPON). Introducing these standards while maintaining compatibility with older ones (1/10G Ethernet, GPON) and increasing network security concerns have led to system testing challenges to ensure OLT equipment and network robustness and resilience.
Automated testing has emerged as a powerful technique to enhance software and hardware quality assurance processes, as it allows the earlier detection of defects, reduces manual effort, and improves overall testing efficiency while keeping the testers motivated.
This article presents an overview of the OLT's software test automation process, the main challenges, and the strategies and tools used to overcome them.

Authors

Celso Covelo, Altice Labs, celso-c-covelo@alticelabs.com

Tiago Cantão, Altice Labs, tiago-j-cantao@alticelabs.com

João Carreira, Altice Labs, joao-c-carreira@alticelabs.com

Nuno Duarte, Altice Labs, nuno-m-duarte@alticelabs.com

Robert Vidal, Altice Labs, robert-h-vidal@alticelabs.com

Gracinda Rodrigues, Altice Labs, gracinda-l-rodrigues@alticelabs.com

Luís Santos, Dellent, luis-santos@alticelabs.com

Jerónimo Costa, Dellent, jeronimo-m-costa@alticelabs.com

Keywords

Quality assurance; Test automation; Optical Line Terminal; Continuous integration; Continuous delivery; xPON

Introduction

In today's rapidly evolving and interconnected world, telecommunications stand as one of the essential lifelines of modern society. Their integral role in our daily lives has made their smooth operation crucial, and any disruption can have profound implications comparable to critical services such as energy or water supply.

Within the intricate networks of modern telecommunications, PON access networks assume the role of delivering Internet Service Providers (ISP) services to end customers' doorstep, whether homes, businesses, or as backhaul infrastructure. In this model, as depicted in **Figure 1**, each Optical Line Terminal (OLT) equipment assumes a crucial role by providing Fiber to the x (FTTx) access to up to tens of thousands of clients while also providing functionalities such as DHCP relay or IGMP proxy, QoS, among others.

Over the past years, the xPON telecommunications landscape has witnessed a surge in complexity, largely driven by the introduction of new technology standards such as 100/400G Ethernet, and 10/25/50G xPON, while ensuring compatibility with older standards like 1/10G Ethernet and GPON. Furthermore, growing concerns regarding network security have intensified the need for additional testing procedures to ensure the resilience and robustness of the OLT equipment.

Automated testing has proven to be a powerful tool, elevating the Quality Assurance (QA) processes for both software and hardware. Its main advantages are early defect detection, reduced manual effort, and enhanced testing efficiency. However, the growing complexity and scale of modern OLT systems (see **Figure 2**) have sparked the need for more advanced testing methodologies.

In response to these evolving challenges, one of our current efforts is introducing Artificial Intelligence (AI) into our test automation process. AI-driven

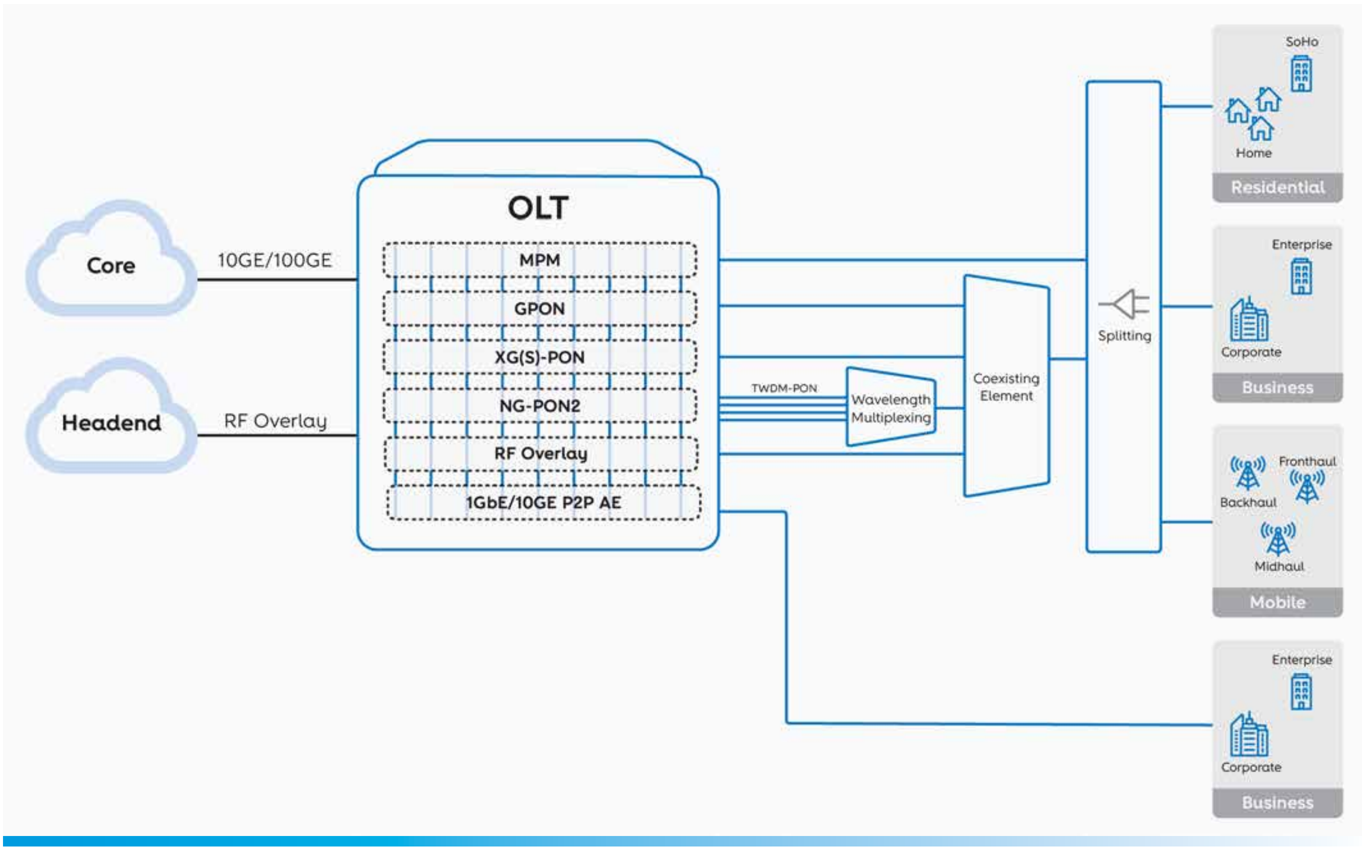


FIGURE 1 – xPON architecture

FIGURE 2 – Altice Labs OLT equipment portfolio

testing holds immense potential in optimizing test coverage, identifying intricate patterns, and even predicting potential system vulnerabilities. By harnessing AI's capabilities, we aim to conduct more comprehensive testing, enhance fault detection, and seamlessly adapt to the rapid evolution of new technology standards.

This article presents a comprehensive overview of the OLT's software test automation process, which sheds light on the main difficulties faced and the strategies and tools we use to ensure robustness and resilience. We delve into the strategies employed to overcome these challenges and explore our ongoing journey toward integrating AI into the testing process. The potential benefits of AI-driven testing and our anticipated challenges will also be discussed.

Challenges in OLT testing

The OLT is the backbone of passive optical networks. Internet Service Providers rely on it to handle access needs in terms of Fixed, Mobile, and Convergent networks, supporting Video (IPTV, OTT TV, and RF Overlay), Data (High-Speed Internet - HSI), and Voice (VoIP) services [1] with high availability, reliability, and density. For this reason, ensuring proper testing of such an essential network component faces many challenges that will be summarized next.

Functionalities complexity

Seamless internet services rely on complex multi-layer protocols. The OLT is a platform where these different protocols at different layers work together to ensure fast and reliable data transmission. Testing for protocol conformance requires a comprehensive understanding of intricate standards to enable the implementation of rigorous procedures that evaluate the correct operation of different service models, message formats, and error-handling mechanisms. Examples of this include

VLAN management, QoS management, multicast management, security management, dynamic bandwidth allocation, and DHCP or PPPoE relay, among other functionalities.

Performance and user experience

As our daily lives are more dependent on and craving internet access, we expect fast, reliable, and uninterrupted connectivity. For this, OLT performance and user experience tests play a crucial role. To assess performance, multiple equipment and 3rd party testing tools indicators are monitored, such as data throughput, packet loss or errors, latency, protocol statistics, and different system resources. Additionally, complex mechanisms of traffic management and bandwidth allocation must be validated to ensure proper and predictable service quality for different applications and end clients. All the above must be evaluated in many scenarios, including load, long-duration, and different client environments, to ensure the correct handling of traffic flow variations and network conditions.

New standards and retro compatibility

1.25/2.5GPON networks that were initially built based on ITU-T G.984.x GPON recommendation have evolved to accommodate higher bandwidth demand, and new 10G PON architectures were introduced defined by ITU-T G.987.x (XG-PON), ITU-T G.9807.1 (XGS-PON), ITU-T G.989.x (NG-PON2) and ITU-T G.9804.x (Higher speed passive optical networks) recommendations [1]. On the Network-to-Network Interface (NNI) side, where the standard was 1G/10G Ethernet connections, speeds of 100/400G are now becoming customary. The initial product line of Altice Labs OLT was designed based on GPON recommendation but is now 10G PON capable. On the other hand, the second-generation OLT line was designed with the new standards much in mind, but they must be fully backward-compatible to replace the old OLT if the client so desires.

Interoperability

Altice Labs OLT products support interoperability with other vendors [2], meaning that the network operator can install the Altice Labs OLT in its already implemented network while keeping the customer premises equipment (CPE). This brings an added challenge to the QA team because it greatly increases the test scenarios that need to be verified since the OLT is tested not only with Altice Labs own Optical Network Unit (ONU) product line [3] but also with another vendor's ONU.

Even if we only consider Altice Labs ONU products, as shown in **Figure 3**, there are a myriad of different hardware models and some client-related software specificities that also increase the number of test scenarios that need to be addressed.

FIGURE 3 – Altice Labs ONU equipment portfolio

Security

Over the last few years, there has been an increasing concern about network security as the number of attacks has only grown larger with novel resources available to perpetrators to identify and exploit network attack vectors. New security features have been implemented to our OLT and are validated thoroughly, such as improved client authentication, IP and MAC spoofing mechanisms, and DoS event detection. Additionally, new testing tools have been introduced to help identify unauthorized access and mitigate the risks associated with port vulnerabilities.

To explore these concerns, Altice Labs has an ongoing partnership with the Instituto Politécnico de Bragança that has already proven fruitful in identifying points of improvement in both the development and testing phases.

Risk management

With all the points explained above, it can be concluded that it is not feasible to test every single scenario combination, especially if there are time constraints on the product roadmap timeline. There needs to be a thorough thought process in predicting an issue's likelihood to occur and its impact, in order to decide on the key scenarios to be tested [4]. This should be done with the development teams' collaboration and previous experience while considering the defined timelines and constraints such as test environment availability.

Testing strategies

Continuous integration/ continuous delivery (CI/CD)

The growing complexity and scale of modern OLT systems, the continuous development of new features as well as the maintenance of older features, presents the QA team with a challenge: how can we test the OLT's features quickly and efficiently to deliver the best quality software as soon as possible? Test automation and CI/CD emerge as a natural answer. A flow diagram of the test automation process can be seen in **Figure 4**.

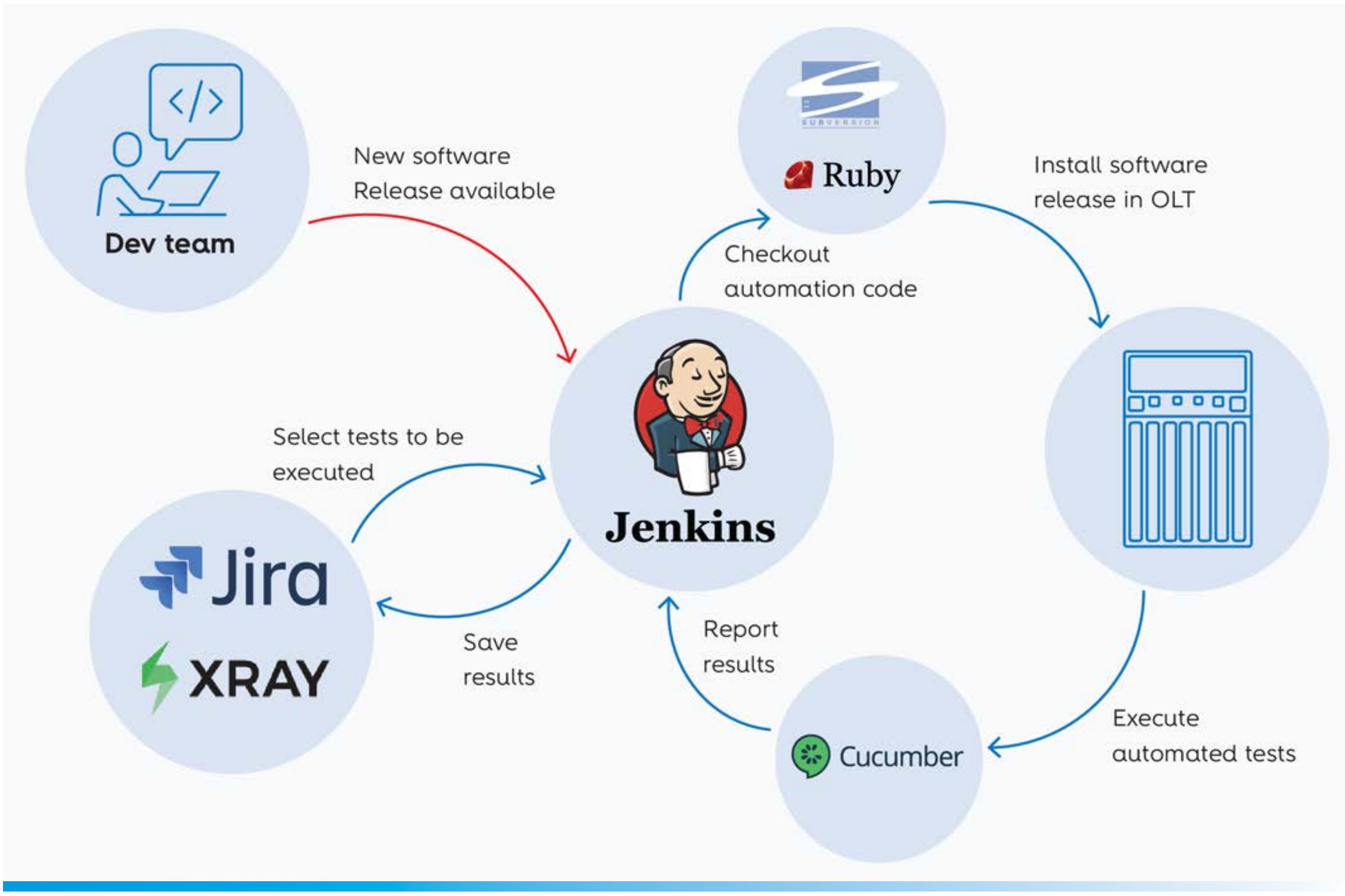


FIGURE 4 – OLT test automation process

The automated tests are created in Xray Test Management [5], a plugin for JIRA software [6], using a Behavior-Driven Development (BDD) approach, where even a non-programming tester can describe the automated test steps in a more natural language that can be easily interpreted by developers, testers, project managers and other stakeholders.

The tester starts the process by analyzing the requirements and understanding how the feature is intended to work. Then, the test steps are written with Gherkin syntax [7], which are then interpreted by Cucumber [8], a software tool designed to execute BDD automated tests. At the same time as new features are implemented, new Gherkin steps can be created and older ones reused. The backbone of the Gherkin steps and the software created to support the Cucumber test execution is implemented in the Ruby programming language [9].

Following the CI/CD approach, the testing scenario represented in **Figure 5** should be a constant throughout the process, making the software version currently being tested the only changing variable between the CI/CD loops. In this regard, the testing scenario is defined during the first phase of the test implementation and includes external equipment used to assist the testing process, namely:

- Optical network units to simulate the user side of the xPON network [3];
- A hardware and software tool for communication protocol emulation and traffic generation to simulate network traffic;
- Optical switches to manage the shared resources among other scenarios and to simulate network conditions, such as fiber cut/restoration;
- Security software tools to perform security tests, vulnerability assessments, and network scanning.

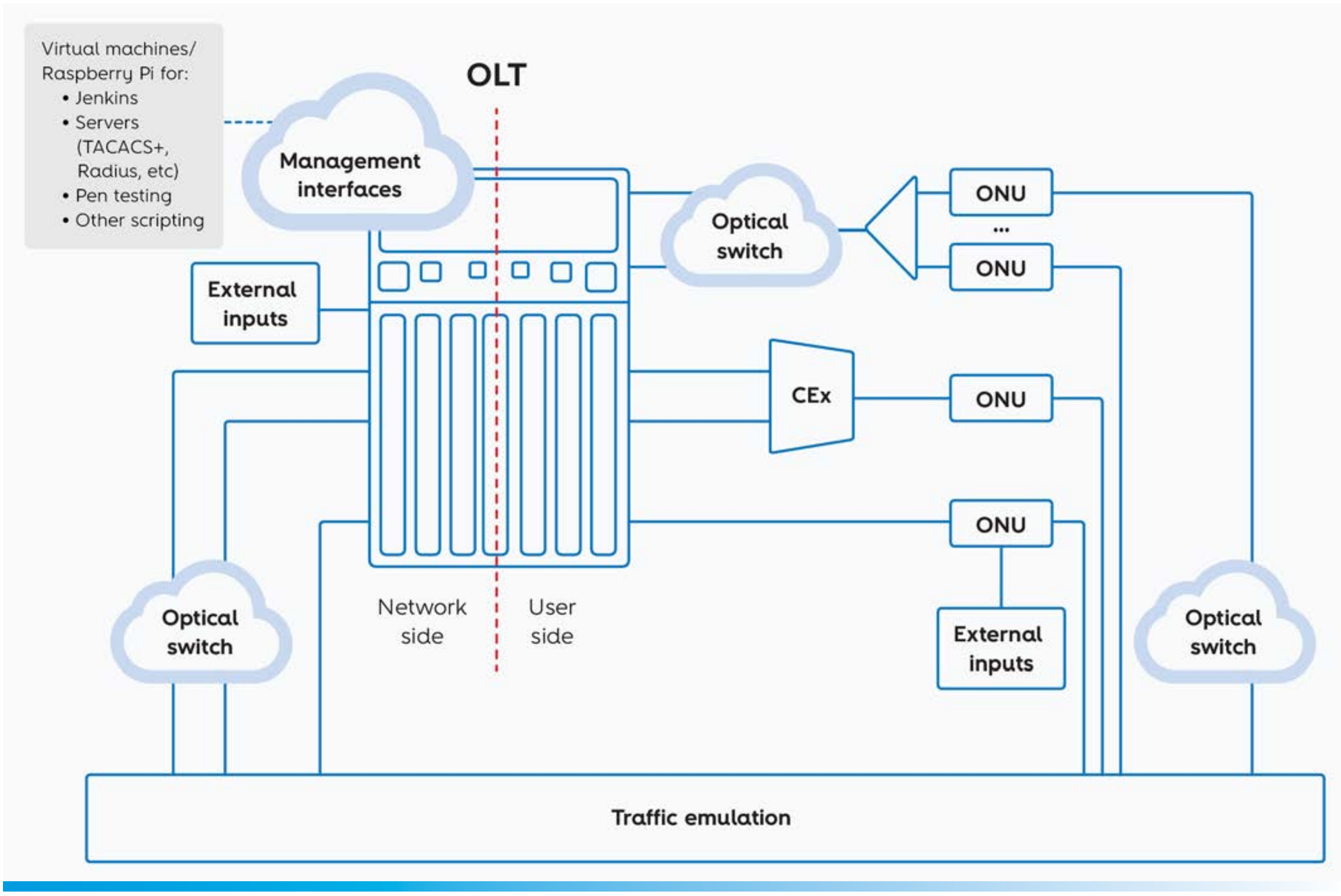


FIGURE 5 – OLT test automation scenario

To interact with the OLT's command-line interface (CLI) and the auxiliary hardware and software tools, Ruby gems were created in-house, giving harmony to the code implemented in the Gherkin steps and providing an easier interface to manage and configure those tools.

The Gherkin steps and the support Ruby code are then committed to SVN [10] and the newly automated tests are ready to be added to the CI/CD pipeline.

Jenkins [11] is an open-source automation server and a key player in our CI/CD workflow. With its extensive plugin support, this server is responsible for creating the OLT software version and for executing the automated tests.

Jenkins can automatically start the process of creating a new software version and, as soon as the version is available, it is automatically installed in an OLT that belongs to the testing scenario defined by the QA team. If the installation is successful, a set of automated tests are executed and the results are imported to JIRA, where all the stakeholders involved in the process can access the test execution history.

Currently, more than 10,000 automated tests exist for the Altice Labs' new generation OLT (OLT2T4, OLT2T2, OLT2T0, OLT2T0E). If they were to be executed in a serialized manner (only one test at a time), the execution time could take a full week. To improve the execution time, several test scenarios exist, each one associated with one Jenkins job. Jenkins can then execute those jobs concurrently, thus reducing the execution time to about 48 hours.

Although the total execution time was greatly reduced, it is still not possible to execute all the automated tests daily, which brings us to the question: which automated tests should be executed? To address this challenge, a sorting algorithm was implemented to choose tests based on their priority, fault history, and the last time they were executed. This process usually runs after working hours and the test report is available when the QA team arrives at the office. Then, the report is analyzed and, if there are any failing tests, the defects are created in JIRA and linked to those tests. In order not to negatively affect future test executions and report analysis time, the failing tests that are linked to unresolved defects are not executed again until the corrections have been made. The defects are reported to the development team which has the responsibility of implementing the necessary software corrections. This leads to the creation of a new software version, starting the process all over again and thus feeding a new iteration of the CI/CD pipeline.

When a release candidate software version is created, the CI/CD pipeline iterations stop and all automated tests are executed, with the final goal of approving the software version for its official release.

Load scenarios

Load testing plays a crucial role in system testing, by assessing the system's performance under various levels of stress, demand, and load. They simulate real-world usage scenarios to ensure that the network can handle the expected traffic volume without significant degradation in service quality. In our load tests, more than 4,000 physical ONUs of up to 30 different models and technologies (GPON and XGS-PON), as shown in **Figure 6**, are used to simulate, as closely as possible, the different conditions the OLT will be subjected to when installed on the customer's network.

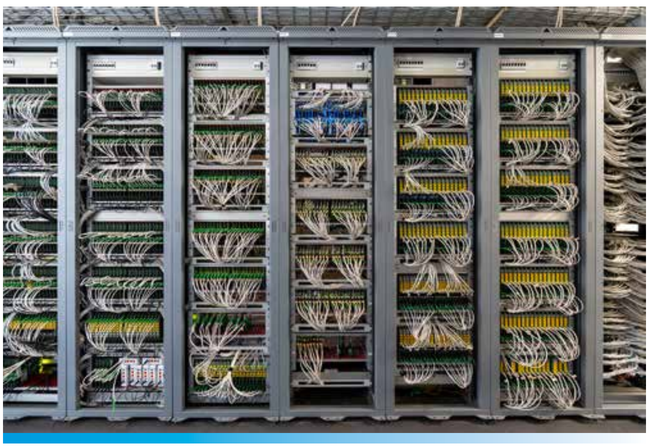


FIGURE 6 – Rack cabinets with several types of ONU used in load tests

The load and automation scenarios are independent of each other, therefore the load tests can be executed simultaneously with the automated tests in the CI/CD pipeline. This ensures that defects are identified early and changes in the software versions do not inadvertently degrade performance.

The OLT load tests are supported in Ruby scripts that control the key elements of the load test scenario, such as OLT configurations and operations, ONU/OLT power sources, traffic generator and protocols emulator, optical switches for fiber connectivity, and optical fiber coil insertion/removal.

Ruby scripts are also used to collect and analyze the most significant metrics obtained from the load test scenarios (under different load conditions), such as:

- CPU and Memory usage per operation;
- Service recovery time after OLT upgrade, OLT power down/up, OLT board reboot, fiber cut/restore, OLT reboot, ONU massive reboot, and ONU massive power down/up;
- OLT response time using the different OLT management interfaces (CLI, HTTP/HTTPS, and SNMP).

Client & long-term scenarios

In the long-term scenarios, we test the system's behavior over long periods, under operating conditions like those of the customer's environment (see **Figure 7**). For this purpose, production versions of hardware and software are used, together with the configurations and service models of each client.

To achieve this goal, two types of scenarios are implemented:

- The first one is connected to a Provider Edge Router through a VPN connection, thus guaranteeing the same type of traffic used in the customer network. The equipment under test are installed in outdoor cabinets exposed to the elements, therefore approaching the customer's real conditions as closely as possible. In this type of scenario, continuity, and quality of service are tested. Monitoring parameters are also collected, such as the systems' processing percentage, memory occupation, temperature, and uptime, amongst others;
- The second type of scenario completes the one described above. It replicates the previous scenario, but it is connected to a traffic generator instead of being connected to a VPN. In this scenario, customer operations and user experience tests are carried out while also performing exploratory and/or negative tests.

FIGURE 7 – Customer and long-term test scenarios

As a complement to the previous scenarios, we use equipment to test the physical insertion and removal of OLT cards. In **Figure 8** such equipment can be seen. This allows us to test the mechanical solution robustness, as well as the boot process and the start-up time.

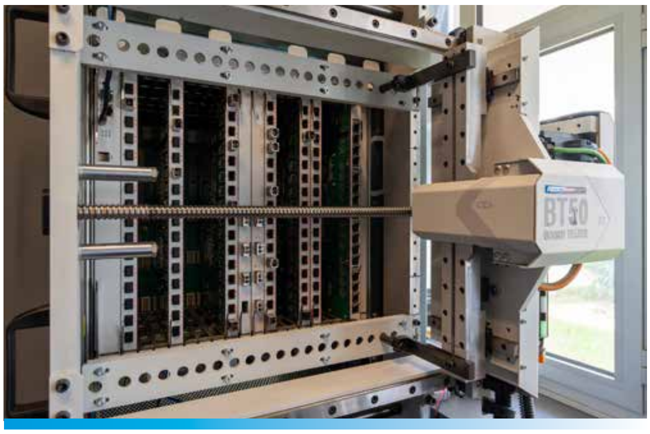


FIGURE 8 – Automated test equipment used for physical insertion and removal of OLT cards

Manual testing

Given the accelerated pace of software releases promoted by an Agile development methodology, automated testing is the preferred testing technique, but one cannot forget the importance of manual testing.

This type of testing is usually done in an early stage of the feature implementation when the tester is learning about the feature while designing the automated tests that will be included in the CI/CD pipeline. It is a crucial phase to detect early defects, either in the requirements definition or already in the implementation.

Exploratory testing, a type of manual testing technique, is also relevant and performed as a complement to the automated tests executed in the CI/CD pipeline. This type of manual testing is not structured and relies on the experience and knowledge of the tester to find edge cases that are missing in the structured tests executed automatically. When a new defect is found in exploratory testing, a new automated test can be created to replicate the issue, thus increasing the test coverage in the CI/CD cycle [12].

Artificial Intelligence (AI)

The complexity and large number of the OLT's software requirements imply a great number of automated test cases that need to be executed to properly validate a given software version. In an Agile development process that focuses on fast software deliveries, time constraints dictate that not all test cases can be executed to validate the daily software versions. This presents the QA team with the following challenge: which software testing activities can be improved by applying AI techniques? The answer to this question is given in several articles [13] and includes the following activities: test case generation, test oracle generation, test execution, test data generation, test results reporting, test repair, test case selection, flaky test prediction, and test order generation. Based on this and considering the specific reality of the tests carried out within the scope of OLT validation, Altice Labs has an ongoing partnership with the Universidade de Coimbra with the following objectives:

Build AI models, based on meta-heuristic algorithms, to improve the set of tests that will be executed. These models must obey a set of pre-conditions, such as time execution constraints and/or maximum test coverage.

Use data analysis techniques on anonymized client data to detect configuration patterns with the goal of selecting the best set of tests that are focused on the most frequent operations performed by clients.

Use mutation testing techniques to create faults in the test scenario and expose that faulty scenario to the current set of tests. The objective is to evaluate and improve the quality of the existing automated tests.

This project is proving to be a great challenge for both the Altice Labs OLT testers, due to the lack of expertise in AI technology, and for the Universidade de Coimbra party, due to the lack of knowledge of the OLT features and functionalities. At the moment, we don't have yet detailed results that can be shared with a broader audience. Nevertheless, we expect that this partnership could bring new perspectives that would be valuable to optimize the test generation, execution, and analysis processes, thus resulting in improved product quality.

Conclusions

Testing an OLT is a challenging task that requires extensive know-how on multiple concepts and adaptive problem-solving capabilities. The ever-evolving world of telecommunications also forces us to accept changes and constantly develop, upgrade, or adapt our testing procedures and scenarios. We tried to shed some light on the complexity, range, and reach of the work we do, the difficulties we face, and the solutions we have applied to overcome them.

Automated software testing has proven to be an essential practice to improve the OLT software quality while allowing testers to attend to other less repetitive, more valued, and motivating tasks that include requirement refinement, manual or exploratory testing, code, and scenario improvement, among others.

Currently, one of the tasks at hand is to evaluate the possibility of introducing new testing and metrics-gathering methods based on AI. We are taking our first steps and some constraints may arise in this journey, but we hope to achieve meaningful results so that the OLT software quality can be improved even further.

References

[1] Altice Labs, " Ultra-Fast Broadband," Nov. 2022.

[2] Altice Labs, "Central Office – Altice Labs." Accessed: Sep. 18, 2023. [Online]. Available: https://www.alticelabs.com/products/central-office/.

[3] Altice Labs, "Customer Premises Equipments – Altice Labs." Accessed: Sep. 18, 2023. [Online]. Available: https://www.alticelabs.com/products/customer-premises-equipments/.

[4] ISTQB Foundation, "ISTQB Foundation certification: 5.5 Risk and Testing (K2)." Accessed: Sep. 18, 2023. [Online]. Available: http://istqbfoundation.blogspot.com/p/55-risk-and-testing-k2.html.

[5] Xray, "Xray - Native Test Management for Jira." Accessed: Sep. 17, 2023. [Online]. Available: https://www.getxray.app/.

[6] Atlassian, "Jira | Issue & Project Tracking Software | Atlassian." Accessed: Sep. 17, 2023. [Online]. Available: https://www.atlassian.com/software/jira.

[7] SmartBear, "Gherkin Reference - Cucumber Documentation." Accessed: Sep. 18, 2023. [Online]. Available: https://cucumber.io/docs/gherkin/reference/.

[8] SmartBear, "BDD Testing & Collaboration Tools for Teams | Cucumber." Accessed: Sep. 18, 2023. [Online]. Available: https://cucumber.io/.

[9] Ruby Community, "Ruby Programming Language." Accessed: Sep. 18, 2023. [Online]. Available: https://www.ruby-lang.org/en/.

[10] Apache Software Foundation, "Apache Subversion." Accessed: Sep. 18, 2023. [Online]. Available: https://subversion.apache.org/.

[11] Jenkins, "Jenkins." Accessed: Sep. 18, 2023. [Online]. Available: https://www.jenkins.io/.

[12] D. Parmar, "Exploratory Testing | Atlassian." Accessed: Sep. 18, 2023. [Online]. Available: https://www.atlassian.com/continuous-delivery/software-testing/exploratory-testing.

[13] A. Trudova, M. Dolezel, and A. Buchalcevova, "Artificial intelligence in software test automation: A systematic literature review," in *ENASE 2020 - Proceedings of the 15th International Conference on Evaluation of Novel Approaches to Software Engineering*, SciTePress, 2020, pp. 181–192. doi: 10.5220/0009417801810192.

DOI: 10.57650/INNOVACTION/2023.05

05 Quantum Key Distribution for secure communications

The emergence of Quantum Computing is expected to disrupt the security of many widely used cryptographic systems, but quantum mechanics can also be the basis for new approaches to data security.
Quantum Key Distribution (QKD) uses quantum principles to securely distribute cryptographic keys across a network. Software Defined Networks (SDN) provide the architecture for the integration of QKD in existing and future networks. Combining QKD with alternative and complementary approaches like Post-Quantum Cryptography algorithms will pave the way for richer, more secure solutions.
This article makes a high-level introduction to the reasons for QKD. It explores how SDN is a decisive enabler, shares the insights of Altice Portugal, and highlights the common participation of Altice Labs, Deimos, Instituto de Telecomunicações, Instituto Superior Técnico and Adyta in relevant projects in the area, exposing an aligned view for the exploration of quantum communications.

Authors

Armando Nolasco Pinto, Departamento de Eletrónica e Telecomunicações, Universidade de Aveiro and Instituto de Telecomunicações, anp@ua.pt

Catarina Bastos, Deimos, catarina.bastos@deimos.com.pt

Gerardo Carrilho Ângelo, Altice Portugal, gerardo.c.angelo@altice.pt

Luís Maia, Adyta, lmaia@adyta.pt

Ricardo Chaves, INESC-ID/IST, ULisboa, ricardo.chaves@tecnico.ulisboa.pt

Rui Calé, Altice Labs, cale@alticelabs.com

Keywords

Security; Quantum communications; QKD; SDN; Cryptography; EuroQCI; PTQCI; DISCRETION

Introduction

Quantum computing is bound to disrupt the security of many present cryptographic mechanisms, but quantum mechanics can also be the basis for creating new and safer cryptography solutions to overcome this challenge. One of these solutions is Quantum Key Distribution (QKD), which uses quantum principles for the distribution of cryptographic keys in a way that is physically guaranteed to be secure, even against quantum computer attacks.

Modern cryptographic services are based on Kerckhoffs' principle [1], which states that the security of cryptographic systems should rely only on the secrecy of cryptographic keys and not on the secrecy of the cryptographic algorithms. This openness of the used cipher encourages open design and deeper analysis of cryptographic algorithms, increasing systems' trust and security.

Ciphers are based on complex computational problems that are difficult to invert, in which complexity grows exponentially with the size of the used keys. Therefore, assuming an eavesdropper with a certain amount of computational power and taking advantage of this exponential complexity growth, it is always possible to define a minimum key length that makes any attempt to break the cipher worthless.

We can also divide cryptographic services into symmetric and asymmetric services depending on the type of keys used. Symmetric cryptographic algorithms use secret but shared keys to ensure the secrecy of our communications. Asymmetric cryptographic algorithms, also known as public-key cryptography, use pairs of keys that comprise a public and a private key. Public cryptographic algorithms are often used to generate and exchange secret symmetric keys between two or more entities.

The development of quantum computers threatens our cryptographic systems because they substantially increase computational power. However, the threat is not equal for all systems:

- A quantum computer using the *Grover algorithm* [2] can reduce the complexity of symmetric cryptography from an exponential to a sub-exponential problem. However, we can mitigate this increase in computational power by increasing the key size length. Indeed, doubling the key size will make the number of operations required to break an ideal hash function equal in a classical and in a quantum computer, and no evidence indicates that a quantum computer can further reduce the complexity of the problems used in symmetric cryptography from exponential to polynomial. Therefore, quantum computers are not considered a significant threat to the security of symmetric cryptography;
- A quantum computer running the *Shor algorithm* can reduce the complexity of the factoring and the log discrete problem from exponential to polynomial. These are the mathematical problems that support the RSA (Rivest–Shamir–Adleman) and all DDH (Decision Diffie–Hellman) based public cryptographic algorithms used nowadays [2].

Therefore, a quantum computer jeopardizes present public-key cryptography, attacking one fundamental block of our cryptographic systems: the key distribution services.

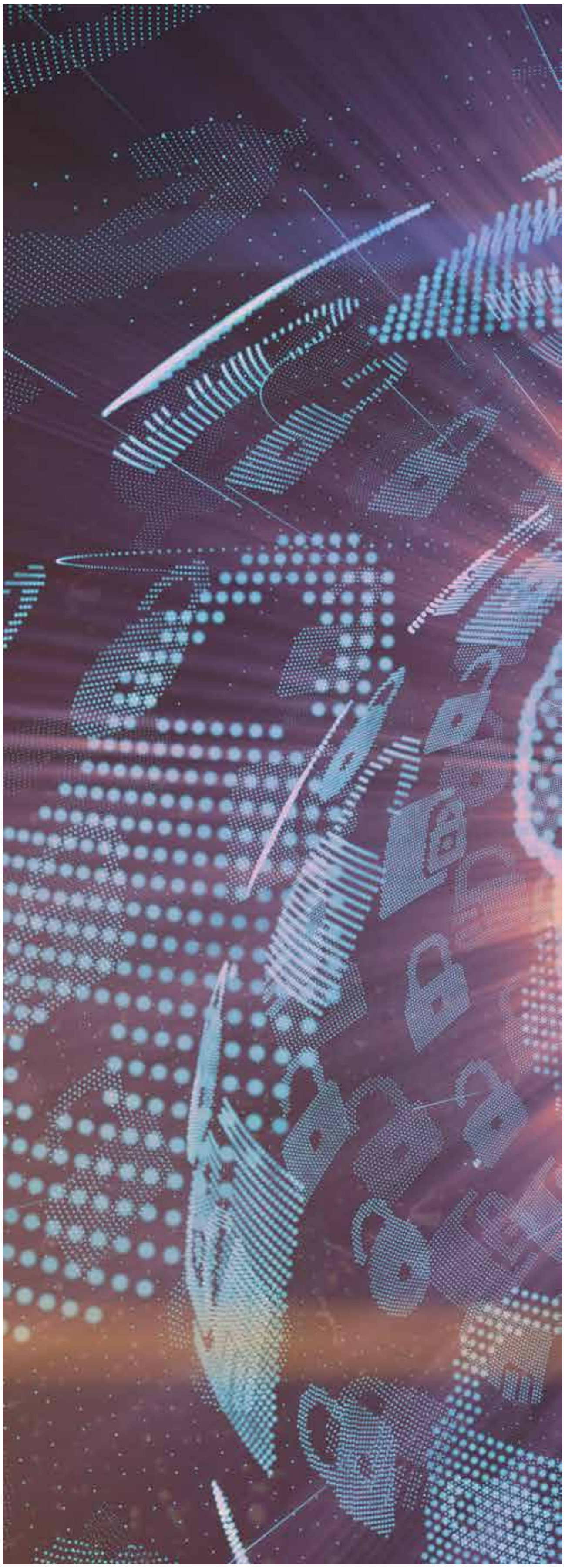
Apart from key generation, key storage and key usage, the key distribution process is vital for our cryptographic systems. If we cannot assure it anymore using algorithms based on factoring and the log discrete problem, how can we keep distributing cryptographic keys? This is a problem that needs an urgent solution because, even if a large enough quantum computer seems to be unavailable now, it is already possible to store the encrypted data and decrypt it later, as soon as we have access to a quantum computer.

Three different approaches have been used to address this problem:

- **Distribution using human intervention:** this requires the definition of a strict protocol, and it is a process largely used by the military and governments worldwide. The major drawback of this approach is that it is not scalable. It is feasible if the communication is between a limited set of users/entities. Nevertheless, it is virtually impossible in the civil area due to the large number of communication entities. Besides, it is a high-cost solution with all the limitations that arise due to human intervention;

- **Post-Quantum Cryptography (PQC):** in this approach, new algorithms are used to implement public cryptographic protocols. The major advantage of this approach is that it can be implemented by replacing only the ciphers, i.e., keeping all the remaining components and processes of the public key infrastructures. The major drawback is that there is neither proof nor very strong evidence suggesting that these other algorithms cannot be attacked by quantum computers soon. Besides that, post-quantum algorithms tend to be more complex than the public cryptographic algorithms used nowadays, which can preclude their usage in some applications that demand a large number of cryptographic keys, or where the available computational power is limited;
- **Quantum Key Distribution (QKD):** in this approach, quantum theory, namely the non-cloning theorem and entanglement, have been used to build systems in which it is possible to precisely quantify the amount of information that is lost in the communication channel. By doing that, we can obtain the upper bound for the maximum amount of information that an eavesdropper may have gained from that channel, and by using privacy amplification techniques, at a certain point, all leaked information can be eliminated, and keys can be exchanged securely. If the eavesdropper can gain such an amount of information that does not allow their complete elimination, the channel should be considered not trusted and cannot be used for key exchange.

In this article, we will address QKD, which can be considered the first quantum-based communications technology to be commercially available.



Just another network?

QKD's basic purpose is to distribute cryptographic keys in a manner that guarantees ITS (Information-Theoretic Security) for the process, relying on the laws of physics, instead of computation complexity.

QKD protocols rely on quantum properties like non-cloning and entanglement. The first protocol was proposed in 1984 and named BB84 [BB84]; many others have been proposed since. This article does not describe any of them. A vast literature can be found on these protocols, and a good practical summary is present in [3] and [4].

There are two main classes of QKD protocols, which require different types of hardware components:

- **Discrete Variable (DV-QKD):** BB84 and its variants (e.g., *Decoy State* to avoid *Photon Number Splitting* attacks) and E91 are DV-QKD protocols. Require single-photon detectors. The most commercially available systems are DV-QKD;
- **Continuous Variable (CV-QKD):** this approach uses homodyne detection, widely applied in classic telecom systems and easy to implement, instead of single-photon detection; however, it requires much higher processing capacity.

QKD protocols in any of these classes can be implemented in two different ways, according to the quantum physical properties on which the QKD protocol is based:

- **Prepare-and-measure:** commercially available today; based on measurement of unknown quantum states;
- **Entanglement-based:** implementation is still being developed; it requires sources of entangled photon pairs and is technologically more complex but could allow the implementation of device-independent protocols.

A basic QKD system comprises a pair of QKD transceivers (sender/receiver), bound by a QKD connection that includes two kinds of channels: a quantum channel, usually an optical fiber or a free space optical connection, and a classical channel.

- The quantum channel is used for transmitting quantum information, making use of quantum properties and effects. It is typically implemented in an optical fiber (or a wavelength in a fiber) or a free-space optical connection, such as a satellite link, and based on the transmission of single photons;
- The classical channel is used for transmitting end user data encrypted with the symmetric keys distributed by the QKD mechanism and is typically supported on IP connectivity. The classical channel does not need to be secret; it only needs to be authenticated.

Figure 1 illustrates a basic QKD system. This system produces a flow of symmetric keys at both ends of the connection, which feed a Key Management System in charge of providing keys to the applications that request them.

As for any kind of network, a single point-to-point connection is the simplest scenario. In fact, many of today's QKD solutions look like this and are used for key distribution between two locations, usually to replace the physical delivery of symmetric keys, but complexity is building up at a fast pace:

- Connection length limitations (typically ~100km for fiber) imply having intermediate secure "Trusted Nodes" that can end a connection and start a new one to build a segmented connection, which means controlling several network segments and their associations;
- Multipath protection and redundancy require a complex network setup;
- Using optical wavelengths instead of purpose-specific fibers is a possibility, which implies the existence of complex optical switched networks underneath;
- QKD networks will be partially based on multiple existing infrastructures (e.g., dark fiber or

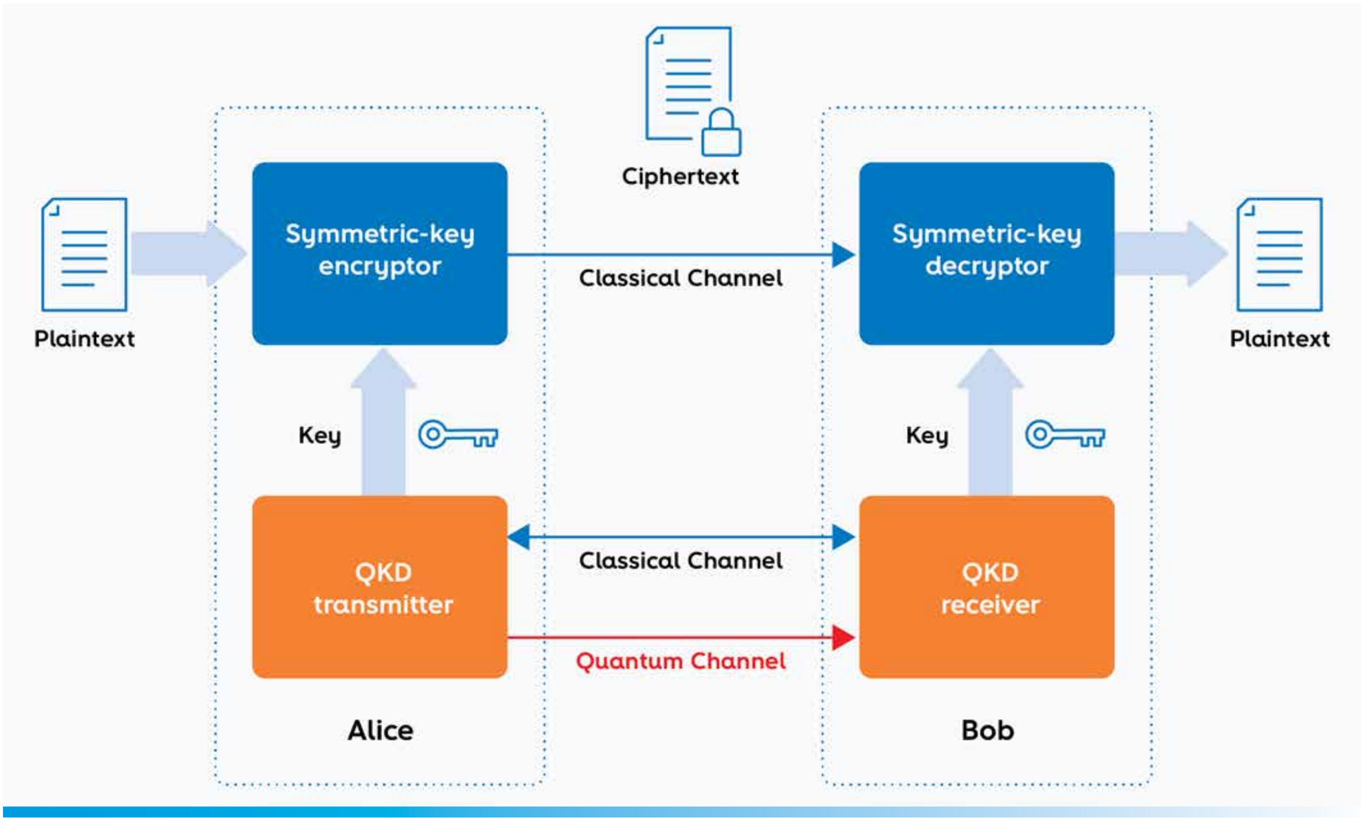


FIGURE 1 – Basic QKD system

fiber wavelengths), combined with new ones, like free-space optical interconnections;

- The exploration of QKD services will require that Key Management Systems interact with a network entity that can provide the necessary paths and resources.

In addition, QKD networks (and quantum communications in general) are not only made of quantum connections. As previously mentioned, they also rely on "classical" channels for data and for control and management of these networks (**Figure 1**).

For the reasons above, as QKD becomes operational, integrating it with existing infrastructure, and supporting it on flexible, evolution-oriented networks makes a Software Defined Networks (SDN) approach absolutely necessary. An SDN can receive requests from a Key Management System (KMS) for the distribution of keys between Alice and Bob and then coordinate the necessary QKD connections, trusted nodes, optical paths and data connections to guarantee that a keystream is actually established between the two actors. This need has already been acknowledged by the Standards Definition Organizations (SDO) [5] [6] [7].

Figure 2 illustrates the layered structure that extends from key production at the Key Generation level to key consumption by applications.

It should be noted that QKD and SDN have a mutually beneficial relationship. On the one hand, the flexibility of the SDN approach allows the integration of QKD technologies into a network, a feature much more complex and costly to have when following past network control paradigms. On the other hand, QKD can be used to provide secure control and data connections to the SDN.

But looking at QKD just as a new network is not enough: in the Quantum Key Distribution approach and for the civil sector, where it is assumed that the final user cannot be equipped with quantum transmitters and receivers, keys should be provided as a service by an operator of the Quantum Key Distribution network. Therefore, as in any telecommunication network, there are several issues that must be addressed: the installation, operation, and maintenance of the network; the network territorial coverage; the access points for end users; network survivability; and the interconnection with other Quantum Key Distribution networks run by other operators.

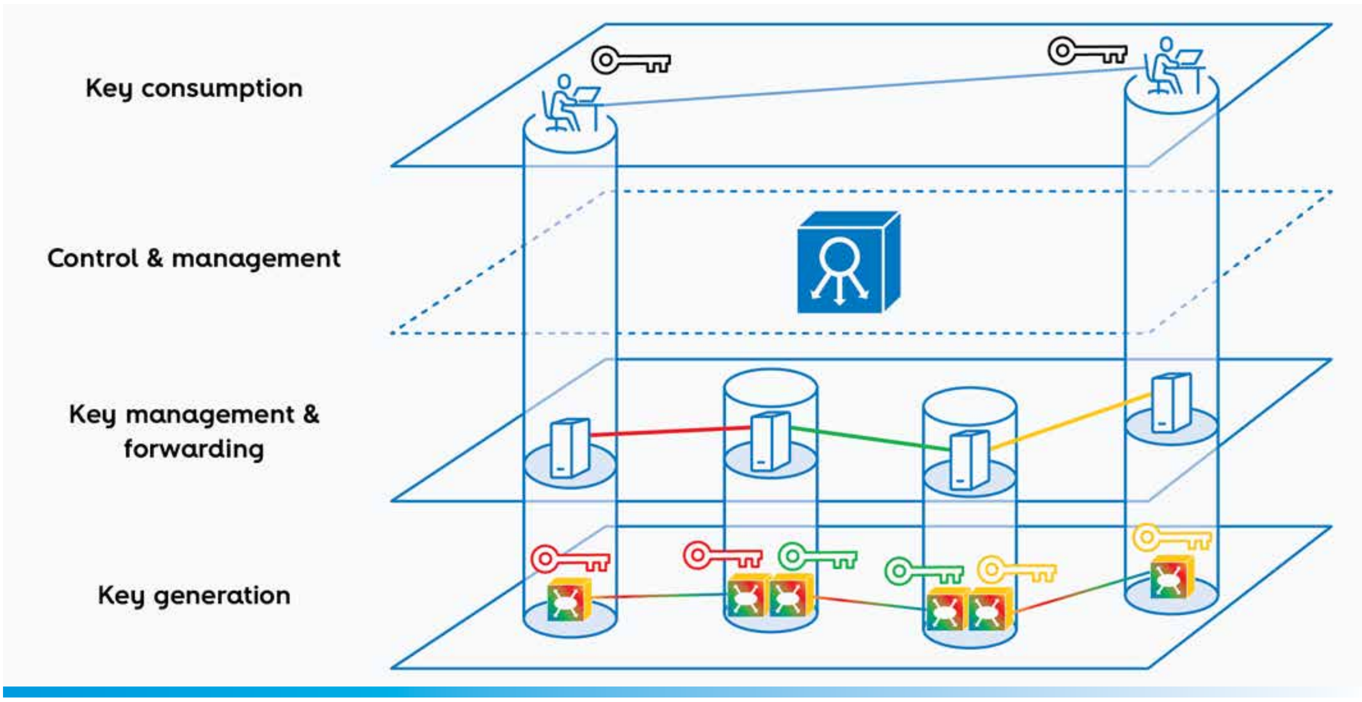


FIGURE 2 – Layered view of a QKD platform

Business perspective

Altice Portugal view

Telecom operators play a key role in cybersecurity and constantly need to advance their capabilities in this area, as threats rapidly evolve and become more sophisticated. At some point in the future, quantum computers capable of breaking public-key cryptography mechanisms are likely to become available and, as a preventive measure, communication networks need to become quantum-safe.

Altice Portugal, as a telecom operator, needs to look at cybersecurity from all perspectives. We need to secure data both at rest and in transit while ensuring all its basic properties: confidentiality, integrity, availability, authenticity, and non-repudiation. And all this must be achieved end-to-end, for each service provided to our customers, and not only for some building blocks of the ecosystem. These are key aspects to consider when planning the evolution to quantum-safe networks.

Despite its current limitations, we recognize the potential of quantum technologies and have already taken steps to evaluate how they can be used in our ecosystem and assess their business perspectives.

In March 2022, we conducted a QKD live demo between our Data Center in Covilhã and a corporate customer located in Fundão.

In February 2023, Altice Portugal publicly announced quantum communications as one of its technological priorities in its strategic vision of digital transformation and innovation for 2030.

We envisage two types of use cases in our journey toward quantum-safe communications:

- Evolve our own telecom network to become quantum-safe;
- Provide B2B customers with ultra-secure keys or quantum-safe connectivity between their endpoints (e.g., *QKD-as-a-service*).

QKD is a promising technology but still needs to evolve and mature from point-to-point proofs of concept toward carrier-grade, large-scale implementations. Key aspects to address include distance limitations, provable security of the complete system, integration of all components into a resilient and trustable system, standardization, certification, and cost.

Current QKD prepare-and-measure implementations are constrained by distance limitations. They require the in-path insertion of trusted nodes, creating communication segments in the range of hundreds of kilometers, or even for shorter distances if high key rates are required. This inevitably results in some misalignment with the telecom operator's long-haul optical transport topology and much higher upfront investment. While metro area QKD implementations are the natural starting point, nationwide and European-wide networks are required at some point to adequately address relevant use cases. Furthermore, the need for trusted nodes and the challenge of their proper physical implementation somewhat detracts from the ideal purpose of building a theoretically secure system.

In addition, a QKD system based on satellite will be necessary to extend QKD service to geographical areas that depend on subsea cables, such as the Autonomous Regions of Açores and Madeira, since a QKD solution for repeated subsea cables is not likely to be developed in the foreseeable future.

The quantum layer is only one component of a wider ecosystem that is expected to achieve the best possible level of security. A key management layer, as proposed in the functional architecture model of QKD networks specified by ITU-T [8] , is necessary to transform individual point-to-point links into a mesh topology that can serve multiple endpoints with resilience. Other notable building blocks in this framework are the control layer and the management layer. Currently Network Management Systems are rightly regarded as critical assets themselves, from a security perspective, so the design of QKD systems must ensure that these do not introduce vulnerabilities. Automation is also a requirement to operate at scale, and it shall encompass a secure method for initial key distribution to authenticate the classical channel (this initial step is not addressed by QKD protocols). The evolution of legacy Public Key Infrastructures (PKIs) and their adequate integration with QKD systems are also necessary. Even if key generation and distribution are to be addressed by a QKD system, an evolved PKI is required to manage digital certificates. All these components need to be stitched together to build a resilient and trustable QKD offer.

The provable security of the entire QKD system shall be preserved as we add layers of complexity. Complete security proofs for practical QKD protocols are still a matter of research, and apart from the robustness of the protocol itself, any practical system implementation needs to be secure against different types of attacks.

More than full standardization of all components, certification of QKD products by European authorities such as ENISA [9] will be critical for its commercial success. So far, ENISA and national authorities of some European countries recommend Post-Quantum Cryptography (PQC) instead of QKD [10] [11] [12] [13], which somehow contradicts the efforts being pursued under the Quantum Flagship program [14].

PQC, which is non-quantum, might provide an acceptable degree of protection against quantum computing threats, at least for some use cases. PQC algorithms are much easier to implement than QKD systems, but have not been fully standardized yet, and because of their own underlying mathematical construction, cannot be considered totally future proof. In fact, several candidate algorithms have been broken during the selection process [15]. However, PQC seems to be the only practical solution for securing end user devices such as smartphones, for which free-space QKD would not be viable.

Given the state of the art for both types of solutions, telecom operators and their customers face a dilemma. Hybrid approaches with two encryption layers, one based on QKD and the other on PQC, have been advocated by some industry players as a measure to reduce the overall risk. However, in many practical cases, hybrid deployments are likely to be combinations of two PQC algorithms of different families (e.g., structured lattices and structured codes), or combinations of PQC and classical algorithms, rather than a mix of QKD and PQC, given the cost/security trade-off.

Relative costs are a determinant factor in this discussion. Nevertheless, it is expected that QKD gradually becomes more cost-effective and can expand its addressable market, but it will never match the lower costs of implementing PQC algorithms. As a result, the value of the data to be protected will determine the demand for QKD solutions, either standalone or hybrid. Institutional and corporate customers who need to protect intellectual property or highly confidential information, such as that related to sovereignty matters, for long periods of time, could opt for a quantum-based approach to offset the inherent risk of PQC, assuming that practical QKD systems are proven to be more secure than PQC-based systems, as discussed above, or implement hybrid QKD/PQC solutions.

The next decades hold the promise of exciting new quantum technologies for secure communications besides QKD. The concept of a Quantum Internet [16] as a global infrastructure to distribute a resource – quantum entanglement – is a long-term vision that requires new building blocks such as quantum memories and quantum repeaters, still being researched in the labs, with many ongoing initiatives at European level, in the scope of the Quantum Flagship program. From a business perspective, it's obviously appealing, as it will enable several disruptive use cases. Apart from entanglement-based QKD, these include *Secure Distributed Quantum Computing*, *Quantum Digital Signatures*, *Quantum Data Locking* (theoretically secure alternative to one-time pad encryption with much shorter key sizes), one-time software programs, certified deletion of data and the appealing concept of *Quantum Money* (unforgeable, verifiable tokens). Other use cases likely to be of interest in a telecom ecosystem include secure clock synchronization and quantum position verification (proof of geographic position).

Altice Labs involvement

As a network solutions provider, Altice Labs is interested in exploring quantum communications, and QKD in particular, in two different dimensions:

- To improve the security of its network solutions, either by integrating QKD as a secure key distribution process or by providing the appropriate interfaces to quantum-enabled key management systems;
- To explore the use of QKD as the underlying technology for secure network products of the future.

As of today, Altice Labs' lines of work in this area are covered by the participation in DISCRETION [DISCRETIONSITE] and PTQCI [PTQCISITE], further described in the "Projects and Collaboration" section below. This participation is focused on leveraging Altice Labs' network expertise, namely on Software Defined Networks, as an enabler for QKD networks as a large step forward in security, but also on QKD as a technology that can help tackle security vulnerabilities on SDN, namely on the communication between controllers and between the control plane and the user plane.

Where we stand

Worldwide deployment

QKD is not something new. In recent years, many proof-of-concept trials, demonstrations, and pilots have been taking place around the world, with different approaches and at various scales; and commercial services using QKD to provide a secure means for symmetric key distribution are already a reality. **Figure 3** gives a 2022 view of that scenario [4]. For their relevance in the scope of this article, we can highlight some of those initiatives:

- MadQCI pilot [17]: a metropolitan Quantum Communication Infrastructure (QCI) pilot, implemented on the commercial fiber network of Telefonica in Madrid, featuring a SDN approach;
- London Quantum-Secured Metro Network [18]: a commercial QKD deployment, integrated into the BT network;
- KOREN [19]: a South Korean Advanced Research network, hosting many cutting-edge

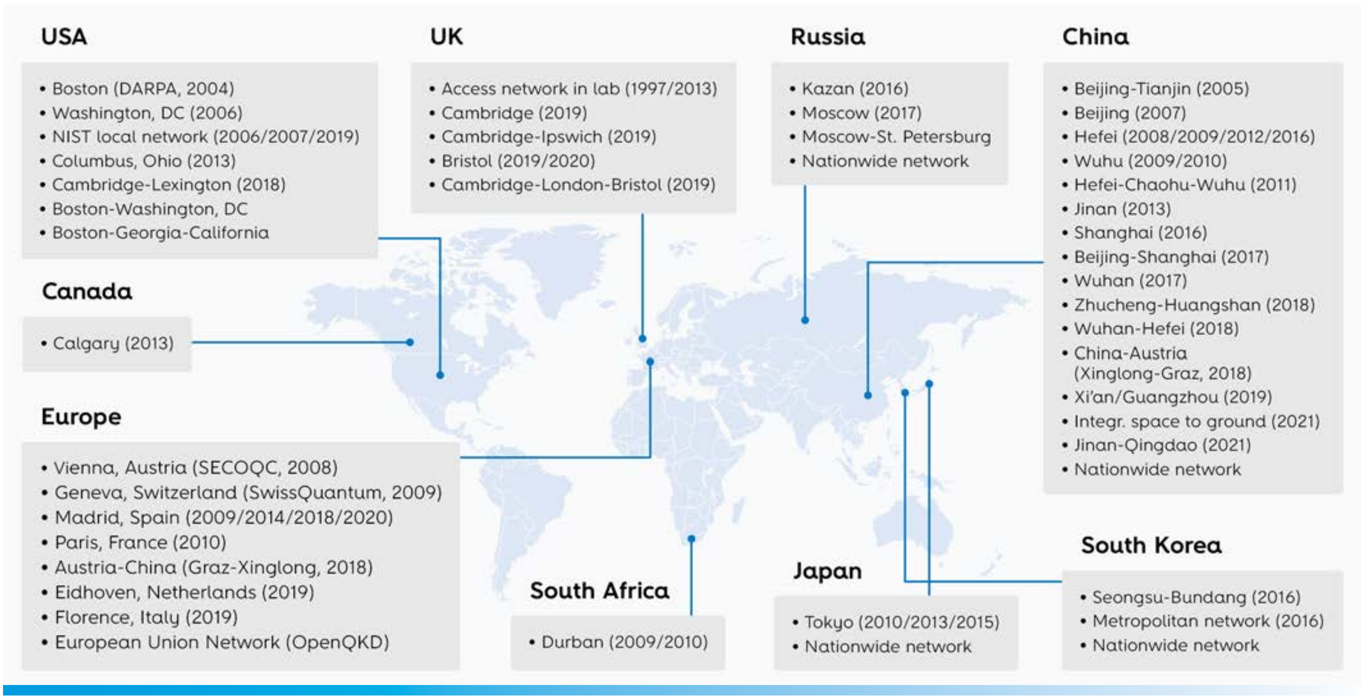


FIGURE 3 – QKD network testbeds around the world (2022) [YUANCAO]

QKD applications, namely for operators, e.g., providing secure communication for the midhaul of 5G networks (communication between the *Distributed Units* and the *Central Unit*).

Despite all the attention it gathers, QKD is not yet seen as a mainstream technology. This has much to do with its limitations and cost (as highlighted in the previous section), which hinder commercial approaches. Nevertheless, throughout the world, states recognize the importance of providing the mechanisms to keep information transmission secure and are promoting and funding national programs toward the implementation of Quantum Communications Infrastructures. Europe has its own QCI program, EuroQCI (next section), and, as a final and compelling example, China already has a working QKD fiber backbone, as well as several satellite links covering more than 10.000km [4], and uses it extensively to protect communications nationwide.

Standards

Technical standardization of QKD is underway, albeit some gaps still need to be addressed [20].

Multiple Standards Definition Organizations (SDO) are involved in creating standards for QKD, and the first regulations are already being adopted by various governments.

In ITU-T, several Study Groups (SG) are producing technical reports, standards, and specifications, based on a systematic top-down approach for a QKD architecture. Some of the recommendations that have already been published are:

- ITU-T SG13 - *Future Networks and Emerging Network Technologies*: Y.3800 to Y.3818;
- ITU-T SG17 - *Security*: X.1702 to X.1715.

In Europe, ETSI QKD Industry Standards Group (ISG) is using a bottom-up approach to define standards as their need becomes evident and, has already published several technical specifications and reports.

ISO and IEC keep workgroup 14 of the Joint Technical Committee, JTC 1/27, named Quantum Information Technology. CEN and CENELEC also collaborate on Joint Technical Committee 22, named Quantum Technologies. The IRTF has a Quantum Internet Research Group.

Like in many other areas, the alignment between all these (and other) organizations is not complete and many gaps remain to be closed, but the first standards for some relevant interfaces are already published. Many others will follow.

Projects and collaboration

Projects

In Portugal, the implementation of an infrastructure enabled by Quantum Key Distribution (QKD) is being done based on a successful trilogy: industrial partners, research and academic institutions, and public institutions as main users. It all started with the pioneer project for European Defense, DISCRETION [21], led by the Portuguese Ministry of Defense and the Portuguese National Security Authority, Gabinete Nacional de Segurança (GNS) and funded by the European Defense Fund, that aims to leverage on QKD and SDN to obtain a highly secure, scalable, and resilient QKD network to provide key material to a network of nationally developed and manufactured cipher machines.

Encryption of information will be done using cipher machines specially prepared to use keys generated with QKD systems, integrated with SDN, and a Key Management Service (KMS) developed in DISCRETION. The QKD nodes and the cipher machines are major outcomes of DISCRETION and are being developed fully by Portuguese entities. This is a European project that leverages not only on the successful support of public institutions in Portugal, but also on the experience of different partners in Spain, Austria, and Italy.

DISCRETION is also the stepping stone for the Portuguese Quantum Communication Infrastructure, PTQCI [22], a Digital Europe-funded project that encompasses the deployment of a highly secure, scalable, and resilient network based on QKD, between different public authorities in Lisbon, as well as a testbed network involving academic and private stakeholders, plus the design of its expansion to farther sites in Portugal, to Spain and the connection to space assets. PTQCI will produce an infrastructure that will be the Portuguese segment of the European Quantum Communication Infrastructure, EuroQCI [23].

EuroQCI will be a large quantum infrastructure in Europe, composed of land and space segments, to safeguard sensitive data and critical infrastructures by integrating quantum-based systems into existing communication infrastructures throughout Europe, providing an additional security layer based on quantum physics. Its main purpose is to reinforce the protection of Europe's governmental institutions, their data centers, hospitals, energy grids, and more critical infrastructures, becoming one of the main pillars of the EU's cybersecurity strategy for the coming decades.

Figure 4 illustrates the relation between DISCRETION, PTQCI and EuroQCI.

Putting it to work

As mentioned above, DISCRETION and PTQCI projects involve not only the quantum-based distribution of cryptographic keys, but also their usage by a network of cipher machines. To put this entire system to work, several implementation details need to be considered.

One of the main aspects is related to the rate at which the QKD can establish keys between endpoints. While one-time padding is considered the only theoretically fully secure encryption algorithm, XORing one bit of secret key with one bit of data, current QKD systems can generate at most a few Mbit of key per second. Since networks require bandwidths orders of magnitude higher, the key bits generated by the QKD need to be expanded. The way to do this is with symmetrical encryption algorithms considered secure, such as the AES algorithm. These algorithms receive 128 to 256-bit keys (called session keys) and use them to encrypt an arbitrarily large amount of data. The amount of data encrypted by each key must be balanced between the QKD key generation rate, the amount of data to be transmitted, and the targeted security level of the network.

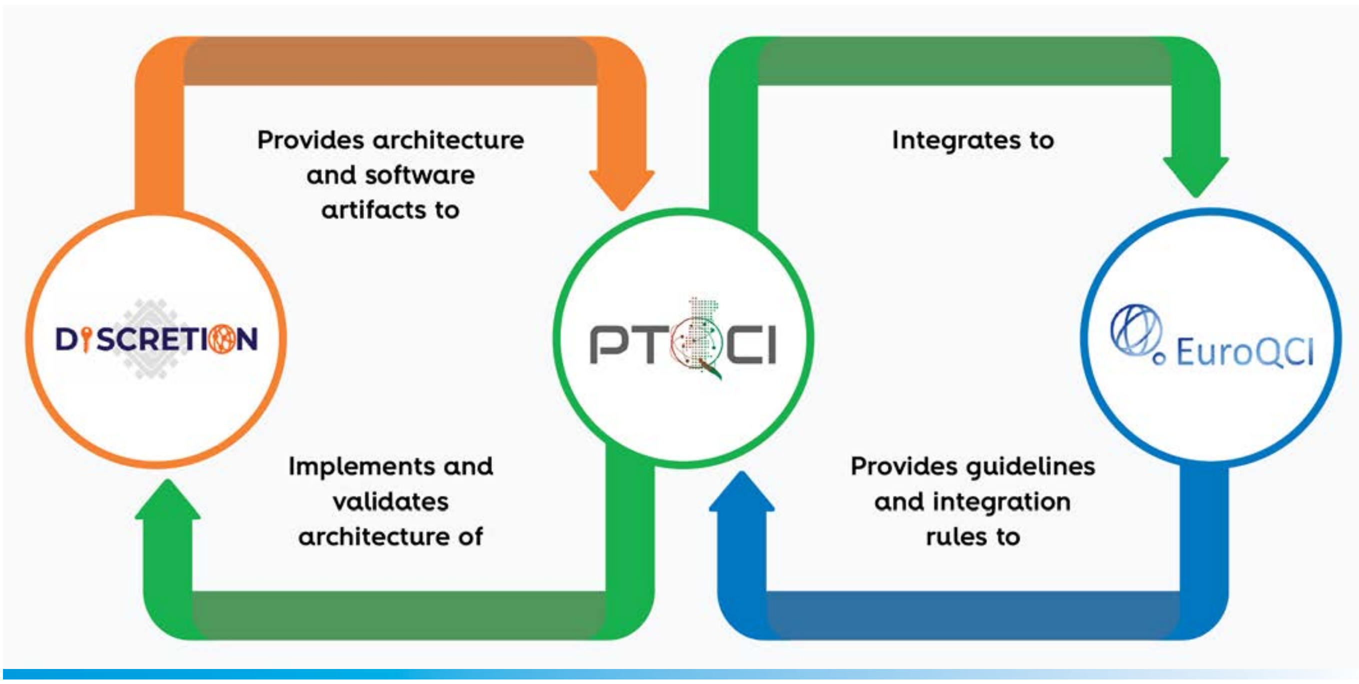


FIGURE 4 – Relationships between DISCRETION, PTQCI, and EuroQCI

Another very important aspect of the network implementation and its protection is the segregation between a secure domain where information is not ciphered, and an "external" domain, where all information has to be ciphered, which in the military paradigm of network segregation is called the Red-Black separation, being the Red the private side of the network. Unlike civilian networks, military networks require a strict segregation between the Red and the Black sides, meaning that no unencrypted data can flow between these two sides. To achieve this, ciphered data tunnels are implemented using dedicated Cipher-Machines. These devices are supported by customized Hardware Security Modules (HSM) designed to be tamper-resistant both at the logical level and at the physical level, making them extremely reliable and secure. In fact, the added security provided by the HSM is also exploited in the key management process, since the cryptographic session keys used to implement the secure network channels can only be accessed within the HSM. This is achieved by a technique called key wrapping, where the session keys are always transported between endpoints encrypted by other keys, called wrapping keys. These wrapping keys are only known by the session key generators (such as the QKD) and the HSM, which consume these keys to implement the secure data channels.

Since most interconnections are IP-based, the network encryption is based on the IPSec protocol, providing tunnel-based encryption to all IP datagrams transversing the Red-Black network borders. It provides confidentiality, authentication, integrity, and freshness to all the data, including the origin and destination IP addresses. With this, attackers on the Black/Public network only see encrypted data flowing from one Black end-node to another Black end-node. The provided cryptographic operations are performed within the HSM itself. With this and assuring that all data exiting or entering the Red network go through the Cipher-Machines, and consequently through the HSM, a full Red-Black segregation is assured. All data is encrypted within the HSM and using session keys known only by the HSM and its generator. These HSM can also use and contain bootstrap keys to authenticate and initialize the QKD operation. Note that QKD only allows the establishment of point-to-point keys but not multiparty key distribution. As such, each endpoint communicates with several endpoints using individually generated keys. This means that the Cipher-Machine must manage several secure channels and the associated session key, establishing one security channel for each connection.

Finally, the developed system also takes into account the old and new, meaning that it supports and uses both classically distributed session keys and Quantum distributed keys, merging them within the HSM. By combining the two keys, a high confidence can be achieved between the well-established key distribution approaches, currently used in the defense infrastructure, and the novel and groundbreaking QKD technologies.

This work also deals with the national sovereignty issue by providing national authorities the control of the key generation process, the associated technologies, and the cryptographic equipment itself, both in terms of the Cipher machines and QKD infrastructure.

Conclusions

The development of quantum computers threatens current public-key cryptography systems. As a preventive measure, communication networks need to become quantum-safe.

Quantum technologies can be used to secure data communications with a long-term perspective and thus become an essential pillar of national digital sovereignty. Several disruptive quantum technologies are expected to develop and mature over the next decades, enabling many ground-breaking applications and paving the way toward a *Quantum Internet*, thus creating significant business opportunities.

QKD is a promising technology that can be used for this purpose. It is commercially available today and expected to progressively become more cost-effective, by addressing critical aspects such as the distance constraints. Provable security of practical QKD implementations is also an essential requirement.

A key management layer and an SDN-based control layer can be combined with a QKD layer to form the building blocks of carrier-grade large scale quantum-safe networks, evolving from metro to nationwide and European-wide implementations and enabling a key-as-a-service delivery model to end customers. This will require integration of all these components into a resilient and trustable system that can be certified by European institutions, but a business model for its operation still needs to be defined.

PQC, which is non-quantum, is an alternative approach that might provide an acceptable degree of protection against quantum computing threats. PQC algorithms are much easier to implement than QKD systems, but have not yet been fully standardized, and because of their own underlying mathematical construction, cannot be considered totally future proof.

Both QKD and PQC can be deployed standalone or as part of hybrid implementations that can also include classical public-key mechanisms, allowing service providers and customers to offset the inherent risk of each individual technology and algorithm. This is an important research area, namely in the domain of mobile networks, toward 6G.

Projects like DISCRETION and PTQCI gather national and European organizations around the theme of quantum communications in general and QKD in particular, creating an important competence core from which other projects will emerge, hopefully creating the critical mass for other initiatives, targeting a market that craves for security.

References

[1] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, Vols. 28-4, pp. 656-715, 1949.

[2] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. Shamsul Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi and P. Wallden, "Advances in quantum cryptography," *Advances in Optics and Photonics*, vol. 12, no. 4, p. 1012, 12 2020.

[3] O. Amer, V. Garg and W. O. Krawec, "An *Introduction to Practical Quantum Key Distribution," IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 3, pp. 30-55, 3 2021.

[4] Y. Cao, Y. Zhao, Q. Wang, J. Zhang, S. X. Ng and L. Hanzo, "The Evolution of Quantum Key Distribution Networks: On the Road to the Qinternet," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 2, pp. 839-894, 2022.

[5] ETSI, "GS QKD 015 - V2.1.1 - Quantum Key Distribution (QKD); Control Interface for Software Defined Networks," ETSI, 2022.

[6] ETSI, "GS QKD 018 - V1.1.1 - Quantum Key Distribution (QKD); Orchestration Interface for Software Defined Networks," 2022.

[7] ITU-T, "ITU-T Rec. Y.3805 Quantum key distribution networks – Software-defined networking control," 2021.

[8] ITU-T, "Rec. ITU-T Y.3802 Quantum key distribution networks – Functional architecture," 2021.

[9] European Union Agency for Cybersecurity, "ENISA". Accessed: Oct., 2023. [Online]. Available: https://www.enisa.europa.eu/.

[10] ENISA, "POST-QUANTUM CRYPTOGRAPHY Current state and quantum mitigation, V2," 2021. Accessed: Oct., 2023. [Online]. Available: https://www.enisa.europa.eu/publications/post-quantum-cryptography-current-state-and-quantum-mitigation/@@download/fullReport.

[11] ANSSI, "ANSSI views on the Post-Quantum Cryptography transition," Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), 2022. Accessed: Oct., 2023. [Online]. Available: https://cyber.gouv.fr/en/publications/anssi-views-post-quantum-cryptography-transition.

[12] ANSSI, "Should Quantum Key Distribution be Used for Secure Communications?," Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), 2020. Accessed: Oct., 2023. [Online]. Available: https://www.ssi.gouv.fr/publication/should-quantum-key-distribution-be-used-for-secure-communications/.

[13] S. Ehlen, H. Hagemeier, T. Hemmert, S. Kousidis, M. Lochter, S. Reinhardt and T. Wunderer, "Quantum-safe cryptography – fundamentals, current developments and recommendations," Federal Office for Information Security, Germany, 2021. Accessed: Oct., 2023. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Brochure/quantum-safe-cryptography.pdf?__blob=publicationFile&v=4.

[14] Quantum Flagship, "Homepage of Quantum Flagship | Quantum Flagship". Accessed: Oct., 2023. [Online]. Available: https://qt.eu/.

[15] G. Alagic, D. Apon, D. Cooper, Q. Dang, T. Dang, J. Kelsey, J. Lichtinger, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, R. Perlner, A. Robinson and D. Smith-Tone, "Status report on the third round of the NIST Post-Quantum Cryptography Standardization process," 9 2022. Accessed: Oct., 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8413-upd1.pdf.

[16] S. Wehner, D. Elkouss and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, 10 2018.

[17] MADQ-CM, "Madrid Quantum – Desbloqueando el Poder de las Comunicaciones Cuánticas para un Mundo más Seguro". Accessed: Oct., 2023. [Online]. Available: https://madqci.es/.

[18] HSBC, "HSBC becomes first bank to join the UK's pioneering commercial quantum secure metro network". Accessed: Oct., 2023. [Online]. Available: https://www.hsbc.com/news-and-views/news/media-releases/2023/hsbc-becomes-first-bank-to-join-the-uks-pioneering-commercial-quantum-secure-metro-network.

[19] SKT, "Press Release: SKT Develops World's First Technology for Integration of Heterogeneous Quantum Cryptography Communication Networks," 4 2023. Accessed: Oct., 2023. [Online]. Available: https://www.sktelecom.com/en/press/press_detail.do?idx=1563.

[20] M. Loeffler, C. Goroncy, T. Länger, A. Poppe, I. Khan Max Planck, M. Lucamarini, A. Shields, E. Spigone and M. Ward Vicente Martin, "Current Standardisation Landscape and existing Gaps in the Area of Quantum Key Distribution," OpenQKD, 2021.

[21] DISCRETION Project, "DISCRETION". Accessed: Oct., 2023. [Online]. Available: https://discretion-eu.com/.

[22] PTQCI Project, "Portuguese Quantum Communications Infrastructure (PTQCI)". Accessed: Oct., 2023. [Online]. Available:https://ptqci.pt/.

[23] European Commission, "The European Quantum Communication Infrastructure (EuroQCI) Initiative | Shaping Europe's digital future". Accessed: Oct., 2023. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/european-quantum-communication-infrastructure-euroqci.

DOI: 10.57650/INNOVACTION/2023.06

06 The use of an integrated maturity model for cybersecurity on software development teams

Software has become present in almost any successful organization, becoming an essential asset for achieving business objectives. Nowadays, the advent of Artificial Intelligence (AI) is making software even more critical for supporting businesses. AI algorithms can automate repetitive tasks, analyze large amounts of data to identify trends and patterns, and make predictions to help make better decisions. However, with increased reliance on software and AI comes an increased risk of cybersecurity threats. As businesses collect and store more data, they become a more attractive target for cybercriminals. Additionally, for software to deliver value, it needs to be reliable, consistent, and transparent. Consequently, our vision is that to address those challenges, organizations building software must use a new set of methods in their Software Development Lifecycle. Specifically, a new maturity model can help diagnose the existing situation and define the desired improved state. This integrated maturity model should include DevOps as a foundational set of practices and additional capabilities to address the challenges of cybersecurity and AI use. We believe this is a large challenge to be addressed, so we decided to divide it into two parts. The first one focuses only on cybersecurity, and the second one on AI. Therefore, the work reported in this article is only related to the cybersecurity component. The definition of such a model has been part of the POWER project executed by Altice Labs and the Universidade de Coimbra. In this context, we applied two cybersecurity maturity models – BSIMM and OpenSAMM – to assess the existing cybersecurity-focused processes at Altice Labs and therefore build a baseline. Now that we have the first version of the integrated maturity model, we are repeating the assessment with the same teams.

Authors

Nuno Seixas, Universidade de Coimbra, naseixas@dei.uc.pt

Mafalda Guimarães Nunes, Altice Labs, mafalda-g-nunes@alticelabs.com

Paulo Rupino da Cunha, Universidade de Coimbra, rupino@dei.uc.pt

Miguel Biscaia, Altice Labs, miguel-b-pias@alticelabs.com

Marco Vieira, Universidade de Coimbra, mvieira@dei.uc.pt

Keywords

Cybersecurity; Software development; POWER project; Data

Introduction

Software is now considered an essential tool for organizations to operate effectively and efficiently, allowing companies to increase productivity and enhance customer experiences [1].

The advent of Artificial Intelligence (AI) is making software even more critical for supporting businesses. AI algorithms can automate repetitive tasks, analyze large amounts of data to identify trends and patterns, and make predictions that can help businesses make better decisions [2]. However, with increased reliance on software and AI comes an increased risk of cybersecurity threats. As businesses collect and store more data, they become a more attractive target for cybercriminals [3]. For software to produce value, it needs to be reliable, consistent, and transparent. Therefore, these two trends need to be built on top of a strong foundation, using DevOps as the underlying set of principles for software development [4].

Looking at the challenges posed by AI and cybersecurity over DevOps, our initial research showed that a new, integrated software development approach is needed. Any organization that produces software will need to consider the necessary skills, methods, and approaches for AI and cybersecurity as an interconnected network of capabilities and not as two different, isolated sets of practices. Only then can those organizations maximize the probability of success while decreasing the risks of failure when infusing their software with AI, as stated in [2]. As we have discovered in our research, such an integrated model does not exist, and so we set ourselves the objective of designing an integrated maturity model able to address the interconnected set of capabilities that arise from applying both AI techniques and cybersecurity principles over a DevOps foundation. This proposed model also needs to address some of the critiques that have been assigned to maturity models, such as being static, linear, and focused on a single definition of the right performance and improvement path [4], [5].

Altice Labs is one of those organizations that are looking to build a portfolio based on cloud and AI methods and techniques while still considering the growing importance of cybersecurity. As such, it was evident that this new proposed model would be a good fit for Altice Labs' business goals, incorporated in the POWER project, developed by Altice Labs in collaboration with the Universidade de Coimbra (UC), the Instituto de Telecomunicações de Aveiro (IT Aveiro), and the Instituto Pedro Nunes (IPN).

In this article, we will present the context of the POWER project, its goals, and its main subprojects. We will focus more on the DevSecOps line of work, given its relationship with the proposed maturity model. We will then present the state of the art on the topic of cybersecurity maturity models, which will be the basis for the following section, where we will present the application of a new maturity model and the obtained results at Altice Labs. Finally, we will discuss the conclusions and what the next steps are for this work.

The POWER project

Telecommunications operators, as well as all the other stakeholders operating in that ecosystem, are now at the epicenter of a digital revolution. The vast amount of data generated and processed over the past years calls for new approaches and tools to enhance the operational efficiency of companies and the creation of new and improved products and services, with a focus on user experience.

The project "POWER - Empowering a digital future" [6], [7] aims to create an innovative portfolio of products and services, mostly based on cloud and cognitive technologies, through a strong research and development effort aligned around four strong transformation technology vectors: 5G networks, edge/cloud computing continuum, data-driven technologies and business models, and Artificial Intelligence.

To this end, the project is structured into five technical subprojects (SP): New Technology Integration (SP1); Future Networks (SP2); Future Operations (SP3); Future Services (SP4); and Data Business and 360 Monetization (SP5). While SP1 has a horizontal nature, aiming to create cloud and AI adoption methodologies and respective infrastructure frameworks that can leverage the research and development of new products and services, the remaining technical SPs have a vertical nature, taking advantage of the knowledge created by SP1 to build new solutions at the forefront of the sector. This combination will allow Altice Labs and national-based technology to penetrate and conquer new markets while benefiting from the new business models that these new technologies will stimulate.

This article is written in the context of an SP1 work thread, DevSecOps. The adoption of the DevSecOps methodology has been gaining prominence in the emerging technological scenario due to the urgent need to integrate security into the software development life cycle. Recent cyber-attacks (such as those on MOVEit [8] and AT&T [9]) and the approval of more strict regulations (like the European GDPR [10] and the NIS2 Directive [11]) put tremendous pressure on the need for various industries to ascertain the security of their products and services. Through DevSecOps, security is treated as an inherent element at all stages of the software development life cycle and not just as an afterthought. This early and ongoing focus on security results in greater robustness against threats and vulnerabilities, a more efficient process, and a higher-quality end product. Ignoring security during the software development life cycle can result in significant risks, including data breaches, substantial financial losses, and ultimately, reputation damage.

The transition to DevSecOps calls for well-established processes and practices, supported by tools designed for modern technologies and working practices. A DevSecOps maturity model enables organizations to establish where they are on their journey to DevSecOps, assess their progress toward the ultimate goal, and identify the next steps to achieve their objectives. Since every organization is unique and modern multi-cloud environments are so complex, existing standard maturity models can be limited. Every DevSecOps maturity model has its own particular requirements. The main goal of the SP1 DevSecOps thread is to define the base maturity model to develop the next generation of Altice Labs software products. This model should respond to the current challenges concerning DevOps, Security, AI, and governance (regulation and compliance). Additionally, the defined model should be flexible and sustainable, so it can be adapted to different realities and integrated with different dimensions in the future.

State of the art on Cybersecurity Maturity Models

One of the most well-known maturity models in Information Technology (IT) [12] is the Software Engineering Institute's (SEI) Capability Maturity Model (CMM) and its evolution, CMMI [13], currently in version 3.0 under the Information Systems Audit and Control Association's (ISACA) [14] steering, which has been used to guide and evaluate different paths of improvement taken by different organizations. Although maturity models have historically been considered an important tool for process improvement [12], there is a growing critique of them [4], [5]. They are considered static, based on one single end goal and a unique improvement path. All those discussions on why maturity models are failing confirm the need for an improvement tool that is context-aware, multidimensional, and dynamic.

To understand the state-of-the-art maturity models that could connect cybersecurity and AI on top of DevOps and Governance, we conducted a multivocal systematic literature review [15].

None of the obtained results present maturity models for software development that simultaneously address concerns related to cybersecurity and AI infusion. Our findings fall into one of two groups: Cybersecurity Maturity Models and Artificial Intelligence Maturity Models. As this article focuses on Cybersecurity, we will only present the results in this domain.

Different authors present definitions for cybersecurity, but in this work, we'll be using the one from Hoang and Le [16], who state that "*cyber security can be considered as a collection of systems, tools, processes, practices, concepts and strategies that are used to prevent and protect the cyber space from unintended interaction and unauthorized access and to preserve the confidentiality, integrity, availability, authenticity, accountability (CIAAA) and other properties of the space and its resources*". This definition accounts for different factors – systems, tools, processes, practices, concepts, and strategies – to achieve not just one specific gain but an interconnected set of gains over confidentiality, integrity, availability, authenticity, and accountability.

The literature shows that cybersecurity is an important topic for two main reasons. First, a growing business complexity with a high degree of volatility and the use of heterogeneous systems [3]. Second, software is now part of critical infrastructures, and these are increasingly vulnerable to cyber-attacks, as stated in [17].

A relevant conclusion for our work stems from Frijns et al. [3], who state that, although cybersecurity is important and vulnerabilities represent a high risk of disrupting business, current software engineering methods still do not address it explicitly. In fact, looking at different Agile methods and DevOps, the same author confirms that there is not an explicit call to security practices or safeguards that address the increasing risk of cybersecurity problems. Also, different works confirm that there is a need for a new Maturity Model that explicitly addresses both DevOps and Security, therefore being able to handle a more complete set of threats [18], [19].

In this literature review, we identified two groups of security maturity models: one focused on organizational practices that can be used by any technology-related organization, and another focused on the software development lifecycle. In the first group, we find examples such as C2M2 [20], NICE [21], NIST Cybersecurity Framework [22], and CMMC [23]. In the second group, focused on the software development lifecycle, we identified two instances: BSIMM [24] and SAMM [25].

After analyzing the literature related to cybersecurity, we have identified some challenges. First, the existing models, focused on organizational practices, do not address how to include them in the software development lifecycle. Second, confirming the critiques of maturity models, all those that we identified have a closed architecture and do not allow for integrating additional perspectives or contexts. Third, there is a lack of a common language, making it very difficult to integrate different models with different levels and different quantification schemas.

Proposed maturity model and its results

As shown in the previous chapter, current maturity models are considered static, based on one single end goal and a unique improvement path [4], [5]. Nonetheless, we believe maturity models are still useful to help software engineering organizations understand their current maturity level. After that initial assessment, they allow the identification of which level is the best one for their business needs and guide the improvement path from the initial state to the desired state. So, as a starting point for this work, we considered a maturity model as a needed tool but also one that must be adaptable to their current business context and be sustainable, i.e., have an open architecture built in such a way that further domains can be integrated without having to change the model structure.

Our vision for the way this new model would be used in different business contexts is depicted in **Figure 1**.

As part of the work planned for the POWER project SP1, we took the two cybersecurity maturity models more focused on software development practices, BSIMM [24] and SAMM [25], and executed a self-assessment using both. By applying these publicly available models, we intended to have a baseline for comparing the applicability of the proposed model, which would integrate the practices from these other models.

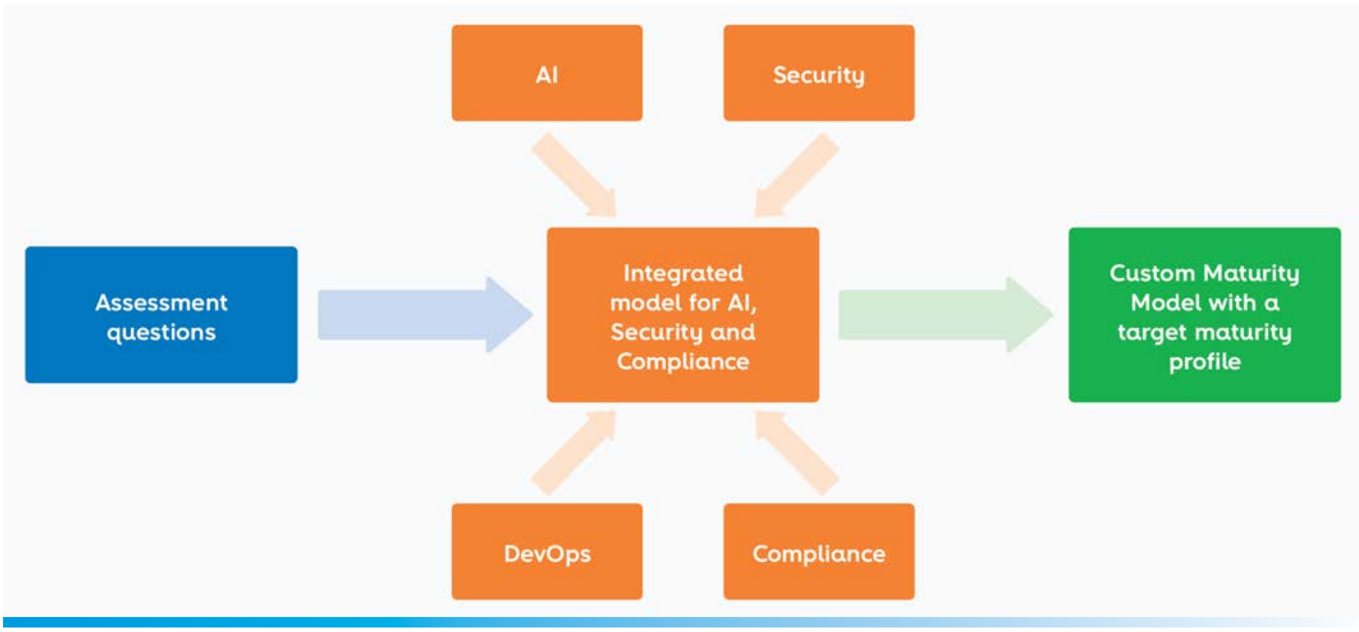


FIGURE 1 – Adapting the maturity model to each business context

After executing the self-assessment based on those two models, we analyzed the results, looking for differences. Despite their different structures, both models returned similar practices that need to be implemented.

So, the next step would be to evaluate the proposed new model and check how it compares with the two public maturity models. We carried out several working sessions with the same internal team from Altice Labs that executed the self-assessment, and the most important results are presented in the next paragraphs:

Full coverage of practices

Since the proposed model was built as an integration of the different practices from the publicly available maturity models, it was expected to contain all the important ones. Nonetheless, this was confirmed by the Altice Labs team, which obtained the same results from using this new model in terms of improvement needs. The team identified the same improvement gaps, and all the important topics were analyzed.

More clear characterization of the levels

In the proposed maturity model, the characteristics of the maturity levels were defined, and their combination was used to characterize each one of the maturity levels. Both publicly available models had a more vague and subjective characterization of each level. The team identified this difference as a major advantage for the new proposed maturity model.

More clear integration in the Software Development Lifecycle (SDLC)

The proposed model uses a language that makes it easier to understand and, therefore, easier to translate into processes that implement those practices. Although the two public models are focused on integrating cybersecurity into the SDLC, the team believes this new proposal has a language more appropriate for practitioners.

Faster application for assessment

According to the team that had to use the previous two models and also the proposed one as a baseline for the self-assessment, the latter allows for faster outcomes, therefore being able to be used more frequently. This advantage turns the proposed model into a continuous improvement tool that can be used even by the software development teams and not just a tool for cybersecurity specialists to use.

Conclusions and next steps

For organizations that build digital products and provide digital services, the growth of their portfolio to integrate AI techniques and methods and a cybersecurity-focused approach is now a must-have for their strategy. Altice Labs, being one such company, identified these needs as important and launched the POWER project with the objective of evolving its portfolio to address those changes. Regarding cybersecurity, all the new practices need to be built on a foundation of DevOps principles so that the teams can be more efficient while addressing those challenges. The way to identify and implement the new cybersecurity-focused practices must follow an integrated approach where different skills, knowledge, and constraints are considered. That's why we have decided to use a tool that has been successful in implementing such changes: maturity models. But the current business context is different, much more focused on dealing with change, requesting flexibility, and above all, adaptability. That's why we consider the need for creating a more adaptable maturity model that can be used in different contexts with the same success rate.

The first version of this new maturity model, focused on cybersecurity practices, was proposed and tested by Altice Labs with the intention of being used by practitioners as an improvement tool.

This initial proposal was compared against two other publicly available cybersecurity maturity models, BSIMM [24] and SAMM [25], and the results show that we were able to produce a new improvement tool that maintains the needed practices but introduces more clarity in the maturity level definition and allows for a quicker and simpler understanding of the different practices, therefore making the translation into the teams' processes also quicker and simpler.

The obtained results confirm this is the right path for having a maturity model adapted to the current business context, able to deal with the uncertainty and volatility that exists. As for the future, the next steps will be to define a set of improvement plans for Altice Labs to implement in its processes all the practices identified by the maturity model, and then re-execute the self-assessment to produce a follow-up analysis on its outcomes. Simultaneously, the model itself will continue its evolution with the integration of AI-focused practices, namely, evaluating its relationships with the cybersecurity ones that are already part of it. Topics such as integrity in the data used for training machine learning models, protection against data poisoning, or even making sure the data bias is under control are some examples of the ones that need to be analyzed and evaluated by the following versions of the maturity model.

References

[1] Marc Andreessen, "Why Software Is Eating the World," Andreessen Horowitz. Accessed: May 13, 2022. [Online]. Available: https://a16z.com/2011/08/20/why-software-is-eating-the-world/.

[2] S. Ransbotham, S. Khodabandeh, R. Fehling, B. LaFountain, and D. Kiron, "Winning with AI," *MIT Sloan Manag Rev*, vol. 61180, 2019.

[3] P. Frijns, R. Bierwolf, and T. Zijderhand, "Reframing Security in Contemporary Software Development Life Cycle," *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), Technology Management, Operations and Decisions (ICTMOD), 2018 IEEE International Conference on*, pp. 230–236, May 2018.

[4] N. Forsgren, J. Humble, and G. Kim, Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations. *IT Revolution Press*, 2018. Accessed: Sep. 27, 2023. [Online]. Available: https://books.google.pt/books?id=Kax-DwAAQBAJ.

[5] B. O'Reilly, "Why Maturity Models Don't Work." Accessed: Jul. 12, 2022. [Online]. Available: https://barryoreilly.com/explore/blog/why-maturity-models-dont-work/.

[6] IPN, "Project POWER - Empowering a digital future." Accessed: Sep. 27, 2023. [Online]. Available: https://www.ipn.pt/projecto/214.

[7] CISUC, "Project POWER - Empowering a digital future." Accessed: Sep. 27, 2023. [Online]. Available: https://www.cisuc.uc.pt/en/projects/power.

[8] C. Tayar, "MOVEit Supply Chain Attack Campaign August Update," Cyberint. Accessed: Sep. 27, 2023. [Online]. Available: https://cyberint.com/blog/research/moveit-supply-chain-attack/.

[9] A. Hope, "AT&T Confirms Third-Party Data Breach Exposing 9 Million Customer Accounts," CPO Magazine. Accessed: Sep. 27, 2023. [Online]. Available: https://www.cpomagazine.com/cyber-security/att-confirms-third-party-data-breach-exposing-9-million-customer-accounts/.

[10] "General Data Protection Regulation (GDPR) Compliance Guidelines." Accessed: Aug. 29, 2022. [Online]. Available: https://gdpr.eu/.

[11] European Union, Directive on measures for a high common level of cybersecurity across the Union (NIS2 Directive). *Official Journal of the European Union*, 2022.

[12] J. Becker, R. Knackstedt, and J. Pöppelbuß, "Developing Maturity Models for IT Management," *Business & Information Systems Engineering*, vol. 1, no. 3, pp. 213–222, Jun. 2009, doi: 10.1007/s12599-009-0044-5.

[13] CMMI Institute, "CMMI." Accessed: May 13, 2022. [Online]. Available: https://cmmiinstitute.com/.

[14] "ISACA," https://www.isaca.org/. Accessed: Sep. 23, 2023. [Online]. Available: https://www.isaca.org/.

[15] V. Garousi, M. Felderer, and M. v. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," *Inf Softw Technol*, vol. 106, pp. 101–121, Feb. 2019, doi: 10.1016/J.INFSOF.2018.09.006.

[16] N. T. Le and D. B. Hoang, "Can maturity models support cyber security?," *2016 IEEE 35th International Performance Computing and Communications Conference, IPCCC 2016*, Jan. 2017, doi: 10.1109/PCCC.2016.7820663.

[17] R. de Bruin and S. H. von Solms, "Cybersecurity Governance: How can we measure it?," in *2016 IST-Africa Week Conference*, IEEE, May 2016, pp. 1–9. doi: 10.1109/ISTAFRICA.2016.7530578.

[18] M. Carcary, E. Doherty, and G. Conway, "A Capability Approach to Managing Organisational Information Security," *Proceedings of the European Conference on Cyber Warfare & Security*, pp. 97–105, Jul. 2019.

[19] R. Kour, R. Karim, and A. Thaduri, "Cybersecurity for railways – A maturity model," *Proc Inst Mech Eng F J Rail Rapid Transit*, vol. 234, no. 10, pp. 1129–1148, Nov. 2020, doi: 10.1177/0954409719881849.

[20] US Department of Energy, "Cybersecurity Capability Maturity Model (C2M2)," Jun. 2022. Accessed: Aug. 23, 2022. [Online]. Available: https://www.energy.gov/ceser/cybersecurity-capability-maturity-model-c2m2.

[21] Rodney Petersen, Danielle Santos, Matthew C. Smith, Karen A. Wetzel, and Greg Witte, "Workforce Framework for Cybersecurity (NICE Framework)," Nov. 2020.

[22] National Institute of Standards and Technology, "Cybersecurity Framework | NIST," Apr. 2018, doi: 10.6028/NIST.CSWP.04162018.

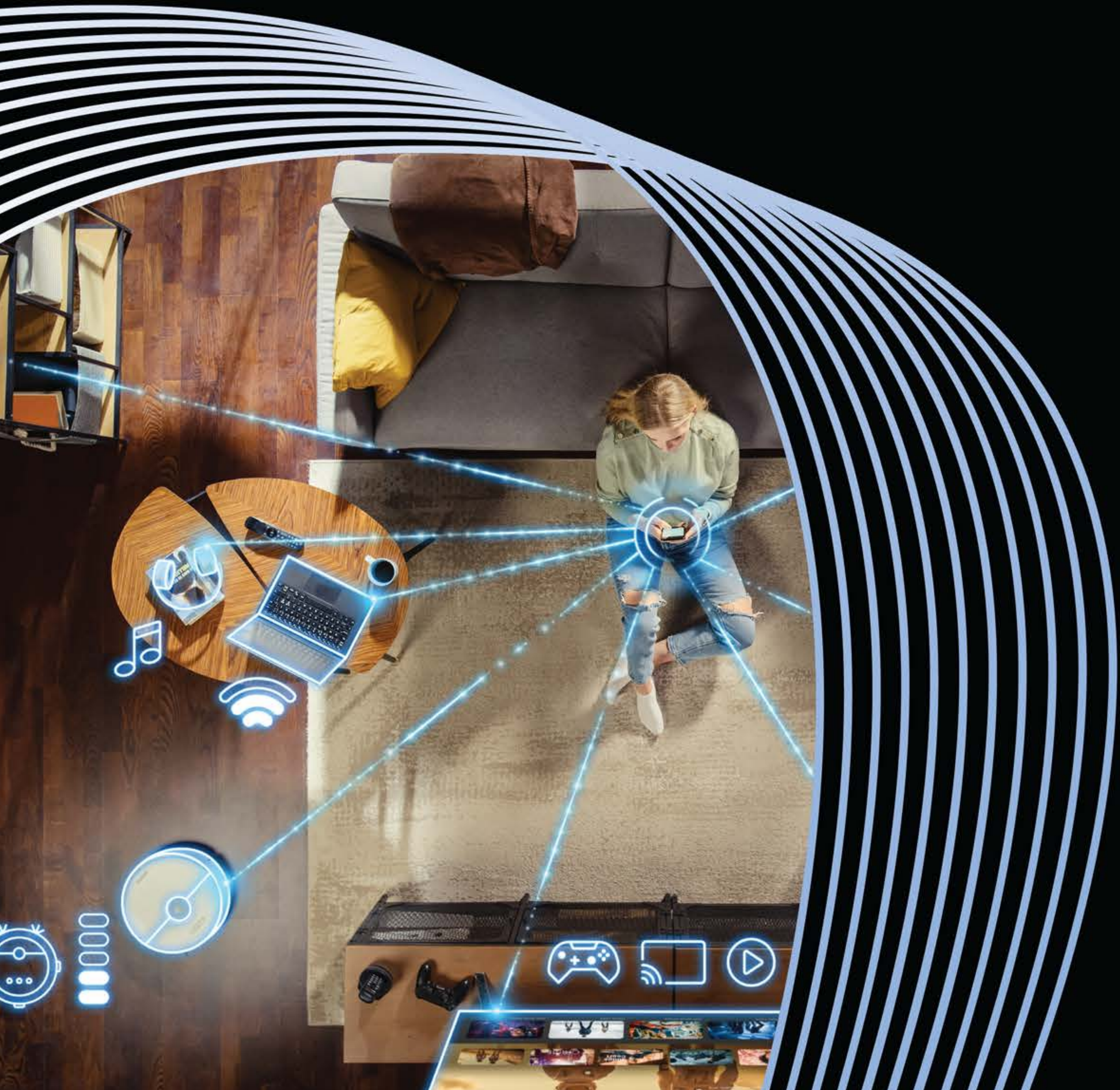
[23] Department of Defense, "Securing the Defense Industrial Base - CMMC 2.0." Accessed: Feb. 7, 2022. [Online]. Available: https://www.acq.osd.mil/cmmc/index.html.

[24] BSIMM, "Building Security In Maturity Model (BSIMM)." Accessed: Feb. 7, 2022. [Online]. Available: https://www.bsimm.com.

[25] OWASP Project, "Software Assurance Maturity Model." Accessed: May 15, 2022. [Online]. Available: https://www.opensamm.org/.

 DOI: 10.57650/INNOVACTION/2023.07

07 The integrated vision of sustainable energy solutions in the Connected Home ecosystem



This article intends to present a comprehensive exploration of sustainable energy solutions within the context of smart homes, employing a case study approach to illustrate practical applications and benefits.
One of the most important dimensions of a company is its social responsibility to offer efficient and sustainable solutions for every household, especially in an era where addressing urgent environmental challenges is paramount. Altice is committed to tackling these concerns through the Greener Altice initiative, which not only underscores the company's dedication to environmental responsibility but also catalyzes tangible change. The development of cutting-edge technologies, exemplified by the first eco-design gateway and Connected Home as a facilitator, promises to make sustainable living more accessible and achievable for individuals and families.
As the world faces rising global environmental issues, the Connected Home emerges as a game-changer in reshaping the future of sustainable living.

Authors

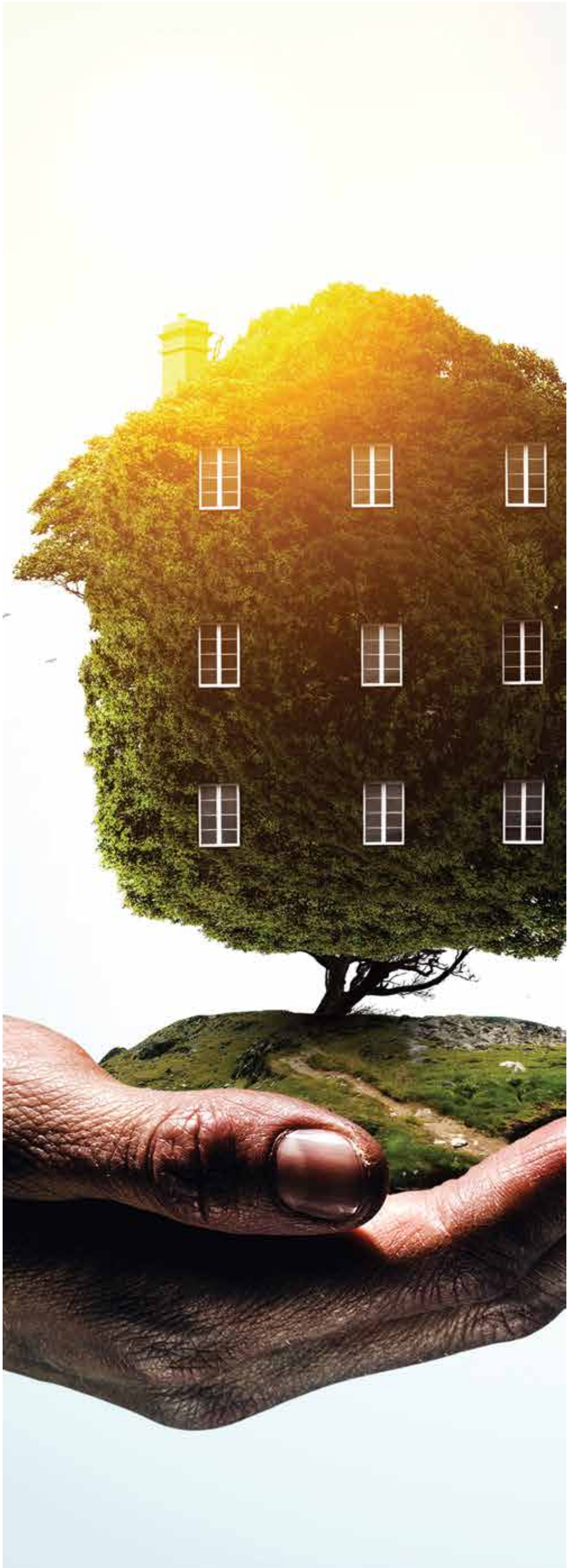
Paulo Chaínho, Altice Labs, paulo-g-chainho@alticelabs.com

Sara Frias, Altice Labs, sara-i-almeida@alticelabs.com

Elodie Breton, Altice, elodie.betron@altice.group

Keywords

Connected Home; Energy; Sustainability; User Experience; Internet of Things



Introduction

The Connected Home program presents a vision aligned with the path toward a greener Altice, incorporating devices like a gateway with a focus on adopting eco-design principles. With all the technological progress, it is essential to innovate through digital products and services to meet the emerging needs of consumers and the sustainability of the planet.

In this article, we will delve into the synergy between Altice's dedication to sustainability, the eco-design of devices, and the business strategy for connected homes. This forward-thinking approach not only represents a significant step towards a greener future but also highlights how technology and ecological responsibility can coexist, resulting in a more harmonious relationship between humans and their digital environments.

The path to a greener Altice

The Greener Altice initiative: a holistic approach to sustainability

Launched in November 2022, the Greener Altice initiative is an all-encompassing program involving all of Altice's operating companies and suppliers. Its primary objective is to implement a sustainable strategy for terminals, including Customer Premises Equipment (CPE), smartphones, and accessories. Altice is dedicated to continual improvement, ensuring that environmental considerations are seamlessly integrated into our business practices and product development processes.

Guiding principles: eco-design for a better tomorrow

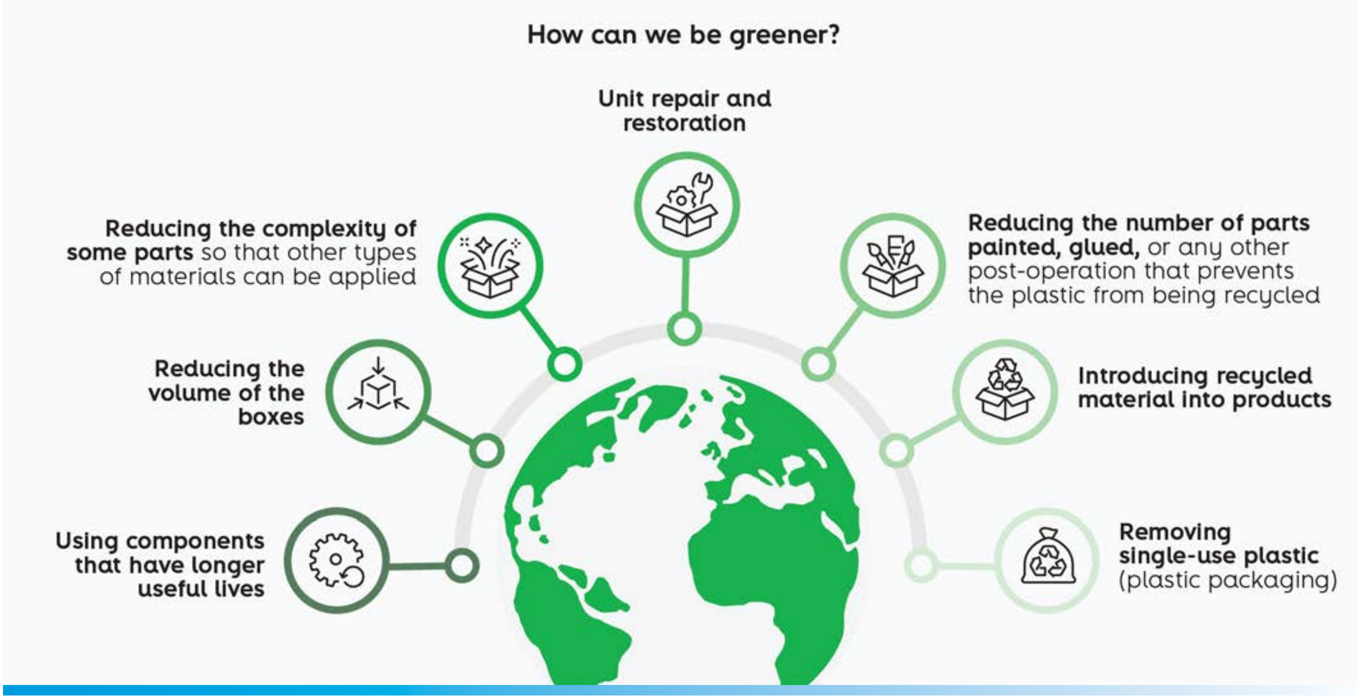


FIGURE 1 – Eco-design for a better tomorrow

The initiative is guided by the principles and methodology of eco-design. We use this systematic approach integrating environmental considerations into the product design and development process. Our goal is to reduce the environmental impact of products at each stage of their life cycle, from conception to use and eventual disposal, by minimizing natural resource consumption, reducing pollutant emissions, and promoting material recycling and reuse, among others, as illustrated in **Figure 1**. We aim to build a more sustainable portfolio without compromising quality and performance.

Mandatory Greener Altice specifications: driving innovation and differentiation

We have developed specific Greener Altice guidelines, which are now mandatory for all new products. These guidelines govern every phase of the product life cycle, providing a robust framework for sustainable innovation. We have incorporated stringent criteria, demanding increased durability, ease of recyclability, and low-consumption energy.

Embracing eco-design not only fulfills our environmental obligations but also serves as a powerful catalyst for innovation. By rethinking our offerings to meet environmental requirements, we seize opportunities to differentiate ourselves in the market, setting Altice as a leader in sustainable technology solutions.

Traditionally, marketing positioning involves the 4Ps: Product, Price, Promotion, and Place. At Altice, we add a fifth P—Planet—marking our dedication to environmental sustainability – **Figure 2**.

FIGURE 2 – 5Ps of marketing mix

Investing in green materials: the first eco-design gateway

Our vision for next generation gateway

The gateway holds a central position within households, serving as the core of the digital ecosystem. It is through the gateway that connectivity flows, enabling the functionality of an array of Internet of Things (IoT) and electronic devices. In an era where digital technology is indispensable, operators bear the responsibility of guiding customers with a robust, reliable, and intelligent network. At the same time, sustainability has become a paramount concern for customers. Customers seek products and brands that address their concerns and align with their values, and a product's environmental footprint is now considered a measure of performance and innovation.

The next generation of our gateway will usher in a new era in our product lineup as the first gateway designed with the environment in mind from conception, covering its entire life cycle (see **Figure 3**). The principles of eco-design have become pivotal for our product positioning, such as:

- **Eco-friendly hardware design:** we are dedicated to reducing non-renewable material usage and increasing the incorporation of recycled materials to minimize our environmental impact and actively contribute to circular economies. We will go further by eliminating harmful, unnecessary chemicals, ensuring user safety, and taking environmental responsibility.

  Moreover, our focus extends to easy dismantling, enabling efficient repair and refurbishment. Embracing future-proof technologies, we design products for upgradability, fostering extended life cycles, and minimizing electronic waste.

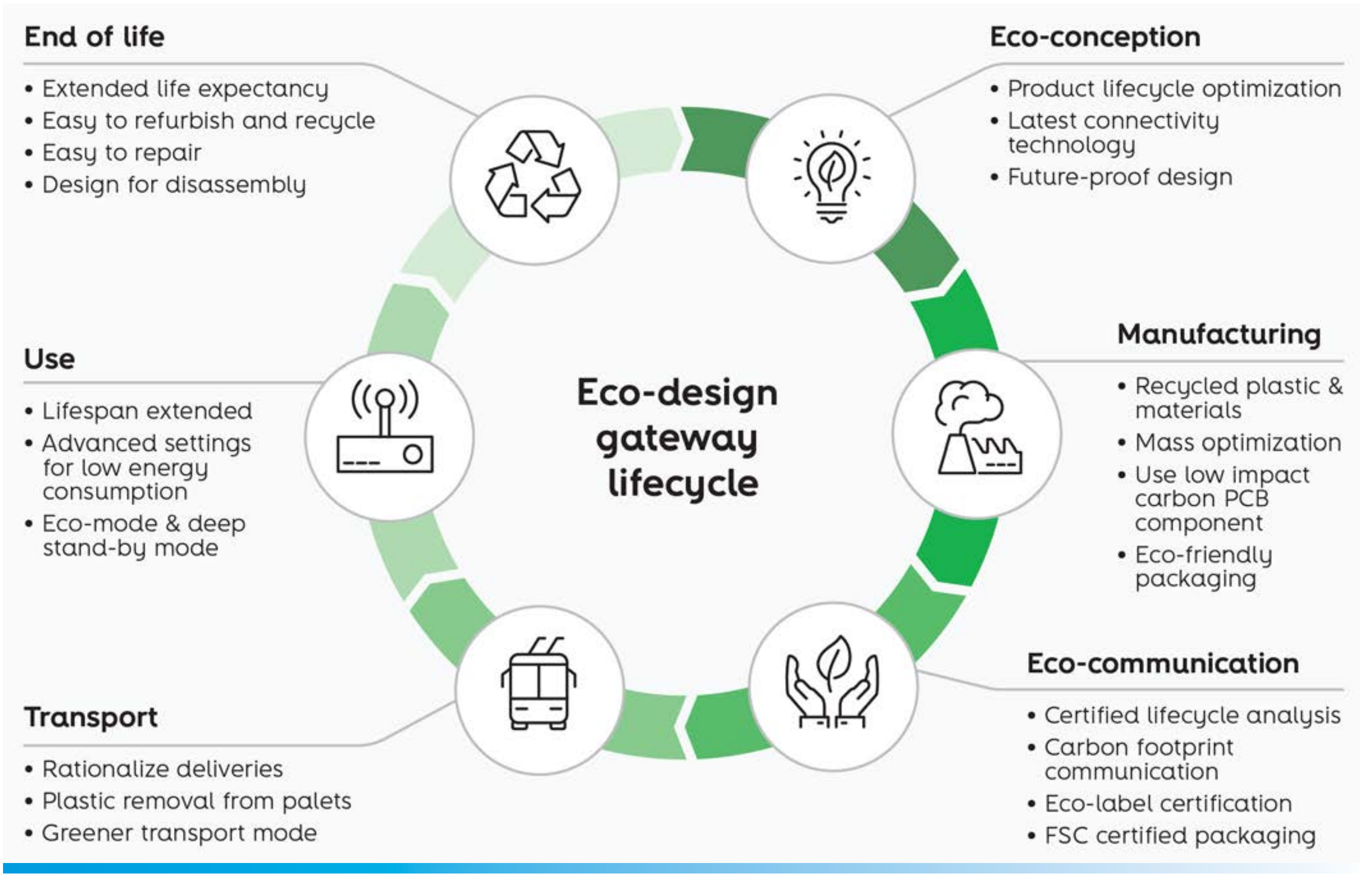


FIGURE 3 – Eco-design gateway life cycle

Our approach emphasizes prolonging a product's lifespan, fostering inherent quality, and promoting longevity.

- **Energy consumption: a vital environmental challenge:** energy consumption stands as a paramount concern in the realm of environmental impact. Even during periods when gateways are not actively streaming videos or browsing the internet, they continue to draw significant power. Home gateways consume 90% of their full power requirement during idle moments. Surprisingly, the energy difference between peak usage (when all services are active) and non-active periods is only reduced by 17%.

  To tackle this issue head-on, the next generation of products must be intelligent and capable of adapting their energy consumption based on user behavior. This requires a new wave of chipsets designed to support low-energy features. Managing connection peaks is pivotal, not just for enhancing the customer experience but also for reducing overall energy consumption.

- **Promoting user engagement for sustainability:** at the heart of our endeavors lies our unwavering focus on the customer. Transparency and communication with our clients are paramount, not only to showcase our eco-design efforts but also to actively involve them in our journey. We will conduct a life cycle analysis for all new products, communicating the findings openly. Transparency is key; it enables us to highlight improvements made between product generations.

  Engaging with our customers is a means of raising awareness, and we are committed to empowering them with tangible actions. New functionalities will be communicated clearly and practically, empowering users to make a difference. Moreover, we strive to recognize and support engaged customers. We are exploring the possibility of offering incentives, whether by highlighting their energy savings on their utility bills, expressing these savings in simple and relatable terms, or providing them with exclusive offers and rates. We aim to actively involve our customers, enabling them to be an integral part of our sustainable journey.

To summarize, our journey toward a sustainable home, supported by a sustainable portfolio, is not merely driven by external pressures or regulations. It's a collaborative process that we are just beginning, aiming for an environmentally sustainable and greener future where we meet our goals without compromising customer satisfaction. This endeavor is both exciting and rewarding.

The business perspective for sustainable energy solutions

Main business drivers

The value of digital services significantly relies on how they align with consumer behavior and expectations. Consumers nowadays are prone to choosing products and services that resonate with their environmental values, driven by a growing awareness of the urgency to combat climate change, reduce carbon emissions, and preserve our finite natural resources. This environmental consciousness increasingly influences their product choices. One of the main drivers for sustainable energy solutions is the substantial environmental benefit they offer by reducing the carbon footprint linked to technology use. By optimizing energy efficiency and embracing low-carbon energy solutions, digital products and services can markedly decrease their environmental impact, contributing to the broader goal of addressing climate change. Additionally, economic incentives play a significant role in driving both businesses and consumers toward sustainable energy solutions for digital services. These solutions often translate into substantial cost savings through reduced energy consumption, making them a more attractive choice for consumers. To effectively take advantage of these drivers, digital service providers must proactively boost consumer awareness of sustainable energy technologies as environmentally friendly options. Providing information about the energy savings potential, which consumers often consider when choosing products, becomes essential. Furthermore, digital service providers should also promote the adoption of eco-friendly practices and technologies, including renewable energy sources like solar and wind power. Thus, digital service providers actively contribute to reducing their customers' carbon footprints, conserving precious natural resources, and mitigating pollution. These initiatives foster a more sustainable way of life and position themselves as responsible protectors of the environment, aligning their goals with economic, environmental, and social dimensions of sustainability [1][2].

Setting the scene for a sustainable smart home

The vision of an eco-friendly smart home solution is firmly rooted in the foundational principles that underpin sustainable living, combining technological innovation with environmental responsibility [3]. As we progress towards a more sustainable future, the smart home is increasingly being seen as a "green" household that champions energy efficiency, conscious consumption, and environmental protection. This vision is built on top of the following foundational principles:

- **Smart home as a green household:** the smart home is envisioned as more than just a collection of advanced technologies; it is a conscious and green household. From heating and cooling to lighting and entertainment, smart homes are conceived to not only meet the needs of their residents but also minimize waste and maximize efficiency;
- **Energy certification for smart homes:** to ensure that smart homes align with the vision of being eco-friendly, energy certification becomes paramount. Much like energy efficiency ratings for appliances and vehicles, smart homes should have a certification that reflects their commitment to sustainable practices;
- **Carbon footprint measuring:** provides the ability to measure and minimize its carbon footprint by using integrated sensors and monitoring systems to track carbon emissions associated with energy use, transportation, and resource consumption;
- **Money savings measurement:** includes an emphasis on quantifiable economic benefits. With this principle, homeowners can track and measure their financial savings;
- **Gamification for eco-friendly behaviors:** gamification techniques are employed to encourage and reward eco-friendly behaviors. For example, homeowners can earn points or rewards for reducing their energy consumption, optimizing their resource use, or actively participating in sustainability initiatives.

The Connected Home example

The Connected Home [4] aspires to evolve into a sustainable energy solution by becoming the central hub for service delivery in the Residential and Small Office/Home Office (SOHO/SME) segments. Its core design principles revolve around agility, adaptability to market demands, business opportunities, and emerging technologies. The Connected Home leverages Telco's key assets to stand out from other smart home service providers, notably by eliminating the need for dedicated smart home hubs, offering robust customer support, and earning trust for privacy and personal data management.

In this transformative journey, Connected Home is looking to address the following major use cases that empower users to actively contribute to energy savings and reduce their carbon footprint:

Energy savings

- **Device-specific management:** Connected Home enables users to monitor and manage the energy consumption of specific devices. This feature allows users to identify energy-hungry devices and take action to optimize their usage;

- **Power usage measurement:** users can access real-time information on the power usage of their connected devices. This data helps them make informed decisions about when and how to use appliances and electronics more efficiently;
- **Smart scheduling:** Connected Home automation scenarios offer the ability to create customized schedules for devices, allowing users to automatically turn them off during periods of inactivity or when they are away from home;
- **Optimize the consumption of self-generated energy:** homeowners can reduce their reliance on the grid and minimize energy costs by effectively managing how energy produced through sources like solar panels or wind turbines is used. The solution can intelligently distribute surplus energy to power household appliances or electric vehicle charging, effectively storing excess energy for later use. This not only reduces the environmental impact but also maximizes the financial benefits, as consuming self-produced energy is more cost-effective than selling it back to the grid;
- **Away and night modes:** Connected Home can automatically switch devices to low-power or "away" mode when users are not at home or during nighttime hours. This ensures that energy is not wasted when it's not needed;
- **Seamless track of costs and earnings:** this technology provides real-time insights into electricity usage, allowing users to monitor and manage their energy expenses more effectively. Simultaneously, it keeps tabs on the surplus energy fed back into the grid, ensuring accurate compensation for excess energy generated. By offering this transparent and data-driven approach, the solution empowers homeowners to make informed decisions, save on energy costs, and contribute to an economically efficient energy ecosystem.

Reduce carbon footprint

- **Carbon footprint analysis:** dashboards offer a comprehensive analysis of the user's carbon footprint. By monitoring energy consumption and factoring in the source of energy (e.g., renewables or fossil fuels), users gain a clear understanding of their carbon emissions;
- **CO2 reduction planning:** users can set goals to reduce their carbon emissions and track their progress over time. Connected Home can provide insights and suggestions on how to achieve these reduction goals;
- **Climate care awareness:** Connected Home can provide information on the broader implications of climate change and how individual actions, such as reducing energy consumption and carbon emissions, contribute to a healthier planet. This creates a sense of responsibility and environmental protection.

Enhancing the user experience in the Connected Home ecosystem

The foundation of the Connected Home ecosystem is a turnkey solution that makes houses for consumers an extension of their way of life through simplicity of use and integration [4].

The Connected Home ecosystem is based on six key pillars: comfort, security, connectivity, entertainment, well-being, and energy and sustainability [4]. This paper focuses on the latter, on the design of the integrated experience of these pillars, which will be embodied in a mobile application for the end user.

To design experiences for this type of product and service, it is important to put the user at the heart of the strategy for this ecosystem and to consider the fundamental principles of human-centered design (HCD).

Human-centered design principles applied to digital products and services

There are 4 main principles in HCD (see **Figure 4**): people-centered, solve the right problem, everything is a system, and small and simple interventions.

FIGURE 4 – The four main principles for human-centered design [5]

The "people-centered" principle places individuals at the forefront of the design process of a product or service. It recognizes that the end-users are the ultimate judges of the product's success. In this approach, it's important to deeply understand the needs, preferences, behaviors, and aspirations of the people [6]. Engaging with the user community through research, interviews, and observations to gain insights that inform decisions is mandatory for successful solutions.

Understanding the human element in "energy and sustainability" is key. Data and statistics are not enough; it's necessary to connect with the individuals who are directly impacted by energy in their daily routines. This implies listening actively to their concerns, learning about their daily patterns, and identifying their goals. For example, families are trying to keep up with the rising cost of living as a result of the rapid increase in energy costs caused by social turmoil such as pandemics or war [7].

Empathizing with these families will be important not only to understand their perceptions and concerns but also to understand what their goals are in terms of sustainability, energy efficiency, or the adoption of renewable energies. These examples illustrate another fundamental principle of HCD - the importance of "solving the right problem" - the importance of correctly defining the problem before looking for a solution [6].

Another principle of HCD taken into account is "everything is a system", recognizing that challenges and solutions are rarely isolated entities. In the context of energy and sustainability, it's essential to understand the issues individuals contend with, notably the rising living expenses and the increasingly urgent environmental concerns. People are genuinely worried about these matters yet frequently lack clear guidance on how to tackle them, and there exists a significant deficiency in understanding and using energy-related data, which impedes informed decision-making.

In this type of ecosystem involving complex problems, the principle of "small and simple interventions" must be considered. This principle encourages a gradual and manageable approach [8]. Whenever possible it is relevant to iterate on solutions, collect feedback from users, and iterate again so that problems can be solved correctly.

Undercover user motivations and concerns for sustainable energy solutions

"There are many reasons consumers have smart homes, ranging from security and convenience to comfort and sustainability. On top of this, consumers increasingly want more details about their houses and how to manage various aspects of the home." [9]

Drawing from the principles of HCD, Altice Labs' Product Design and User Experience team conducted user interviews with a group (15 participants) of smart home, security, energy, and sustainability enthusiasts. These enthusiasts were selected because they have an interest in smart solutions and/or use, or have already used, these types of solutions for their domestic needs, such as comfort, security, energy, and sustainability.

These semi-structured interviews were crafted to probe into the primary motivations and barriers underpinning the adoption of smart energy and sustainability solutions.

In this pool of interviewees (see **Figures 5 and 6**), gender representation stood at 37.5% female and 62.5% male, spanning an age range from 18 to 65 years, and all lived on mainland Portugal. Most interviewees (87.5%) reported prior experience with intelligent solutions, including a spectrum of devices such as light bulbs, plugs, security systems, alarms, cameras, motion sensors, humidity sensors, and intelligent robots. Significantly, half of the respondents acknowledged prior use or engagement with energy consumption monitoring solutions in their homes, such as myenergi, Shelly, eSolar Air, Qcells, Solax, and E-redes.

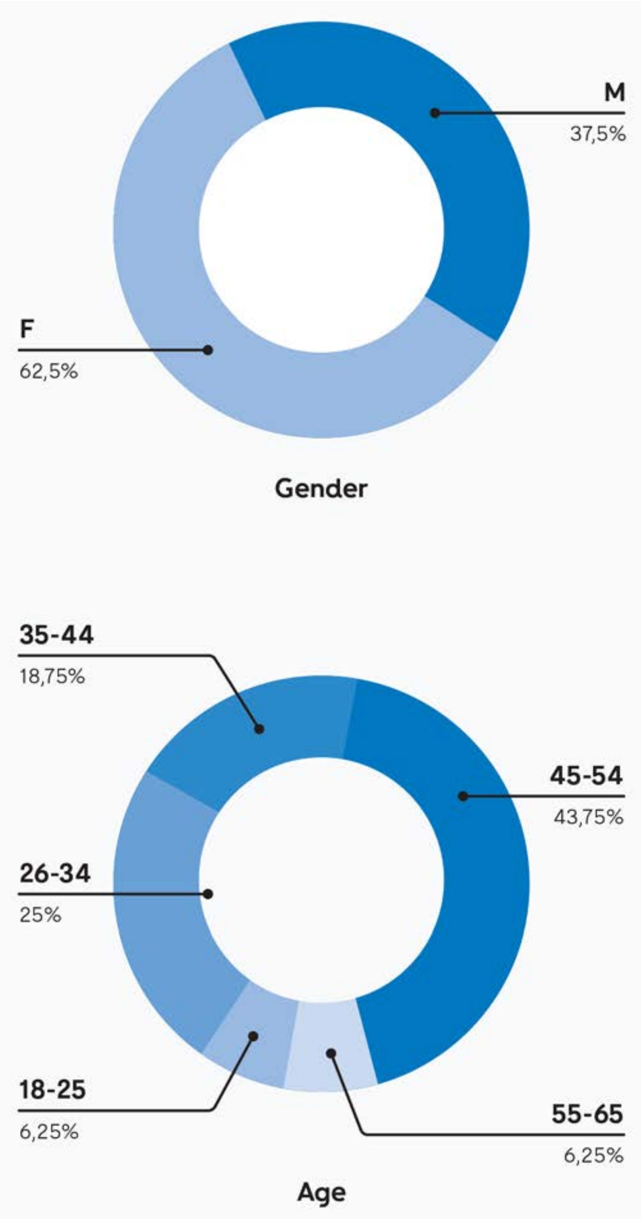


FIGURE 5 – Interviewees' gender and age

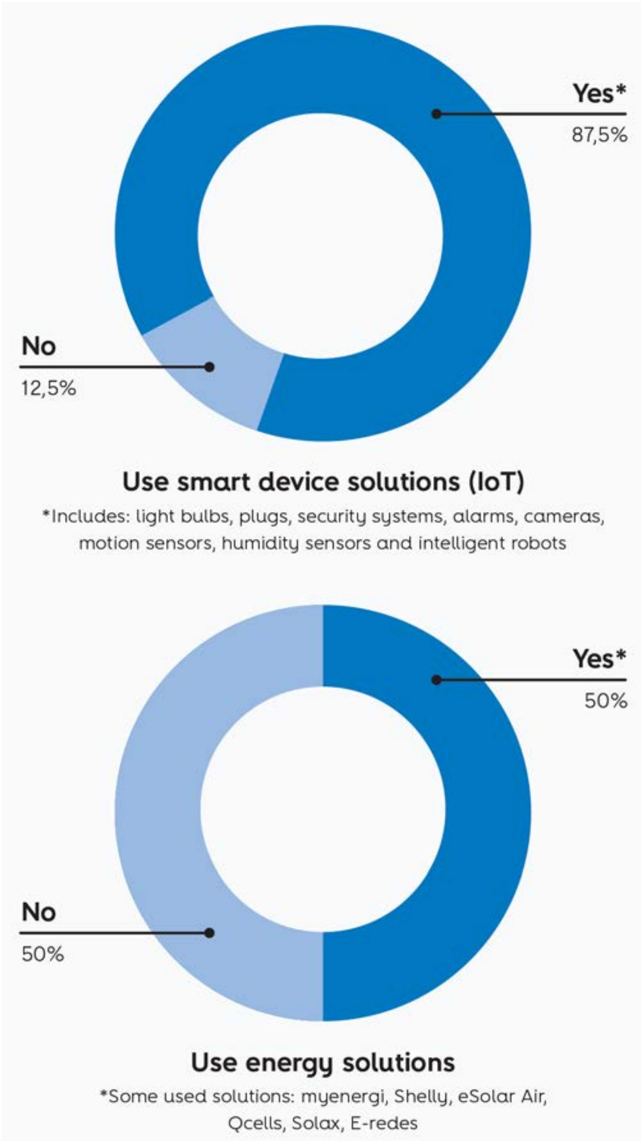


FIGURE 6 – Use of smart home and energy monitoring solutions

The central pivot of motivations that resonated across these interviews was the desire to optimize energy-related costs (**Figure 7**). This was underscored by a commitment to vigilant management and monitoring of energy consumption, regarding electricity and gas. Additionally, the concerns extended to encompass other critical resources within their households, namely water consumption.

FIGURE 7 – Interviewees' cloud of motivations

Notably, an overarching motivation among interviewees revolved around the desire to take greater control over their home energy consumption. They recognized the potential of this data to furnish valuable insights into device consumption patterns, including instances of unexpectedly high levels.

Throughout these interviews, major concerns were raised about future generations and how they will be able to live on this planet. The interviewees were willing to adopt sustainability-oriented solutions as a way of aligning their behavior with environmentally friendly practices.

However, the road to widespread adoption is not without its obstacles, as you can see in **Figure 8.** For almost 70% of interviewees, one source of frustration was the fragmented nature of the available solutions, spread across disparate services and platforms. This complexity was compounded by the challenges associated with initial configuration and costs, making the installation of intelligent systems also typically a daunting process.

FIGURE 8 – Interviewees' cloud of constraints

An intriguing aspect raised by the interviewees' concerns was the perceived lack of intelligence on these platforms. They expressed the expectation that, given the richness of data collected from smart devices, these platforms should inherently possess the ability to propose more automated actions in line with their daily routines. They regretted that the intelligent solutions they had interacted with so far, had often fallen short of expectations.

Empower users through smart home solutions

Smart home appliances are becoming more and more popular among consumers trying to reduce their energy costs as gas and electricity prices keep rising. In 2023, consumers are anticipated to boost their investment in cost-saving products like smart thermostats and smart lights, which also happen to be quite popular with environmentally aware consumers looking for more energy-efficient products for their homes [10].

Through smart home solutions, customers can be empowered to manage their energy consumption while reducing the cost of their energy bills [11]. These studies prove to be in line with what was observed through the interviews with the group of enthusiasts.

The Connected Home ecosystem has the potential to address these main expectations, offering features such as automation scenarios that can leverage data intersection to create mechanisms to make everyday life easier. Optimizing energy management, such as automating the start and stop of energy-intensive activities such as laundry, dishwashing, and charging electric vehicles, according to the most convenient and economical time of day, will be a core requirement [4].

Due to the planet's needs as well as people's increasing demands, smart home solutions and technological advances are transforming energy management and sustainability within an ecosystem of connected homes. This convergence promises a more efficient, cost-effective, and environmentally sustainable future.

Conclusions

The business perspective for sustainable energy solutions is an evolving landscape, deeply intertwined with consumer behavior and expectations. As the world increasingly acknowledges the urgency of addressing climate change and the finite nature of natural resources, consumers are making choices that align with their environmental values. This conscious shift in consumer behavior is a driving force behind the adoption of energy solutions in the digital services sector.

Altice's commitment to environmental responsibility and sustainability isn't just a response to external pressures; it's a dynamic and ongoing journey. We're moving towards a future where we can achieve our goals while enhancing customer satisfaction. This endeavor, while challenging, is inherently rewarding, leading us to a greener, more sustainable life where our products and services make a positive impact on the environment, our customers, and society at large.

References

[1] B. Basarir-Ozel, H. B. Turker, and V. A. Nasir, "Identifying the Key Drivers and Barriers of Smart Home Adoption: A Thematic Analysis from the Business Perspective," *Sustainability (Switzerland)*, vol. 14, no. 15, Aug. 2022, doi: 10.3390/su14159053.

[2] L. Ferreira, T. Oliveira, and C. Neves, "Consumer's intention to use and recommend smart home technologies: The role of environmental awareness," *Energy*, vol. 263, Jan. 2023, doi: 10.1016/j.energy.2022.125814.

[3] Home Assistant, "Home Assistant Energy Management." Accessed: Oct. 14, 2023. [Online]. Available: https://www.home-assistant.io/home-energy-management.

[4] P. Chainho and F. Cabral Pinto, "Connected life: the path to a smarter ecosystem," 2022.

[5] Rikke Friis Dam, "Human-Centered Design: How to Focus on People When You Solve Complex Global Challenges." Accessed: Sep. 6, 2023. [Online]. Available: https://www.interaction-design.org/literature/article/human-centered-design-how-to-focus-on-people-when-you-solve-complex-global-challenges.

[6] D. Norman, "THE DESIGN OF EVERYDAY THINGS," 2013.

[7] P. Gray and J. Hong, "IFA Trends Report-2022," 2022.

[8] Interaction Design Foudantion, "Human-Centered Design." Accessed: Sep. 6, 2023. [Online]. Available: https://www.interaction-design.org/literature/topics/human-centered-design.

[9] B. Kozak, "Smart Home Devices Database – Update 1 Analysis," 2022.

[10] M. Zamoszczyk, "2023 Trends to Watch: Smart Home," 2022.

DOI: 10.57650/INNOVACTION/2023.08

08 VR2Care: enabling multiuser immersive experiences for rehabilitation, physical activity, and socialization

VR2Care project is a Horizon 2020 innovation action that aims to create age-friendly virtual environments, fostering the use of interactive technologies to promote physical activity and social interaction. It enables professional monitoring and guidance of physical activity, allowing a multiplicity of activities, that go from functional training to rehabilitation. Physical activity is performed in a metaverse context of realistic multiuser virtual reality (VR), where users are represented by real human forms (avatars), with animations synchronized with natural movements (embodiment) and multimodal interaction combining voice, gestures, and body movement.
The VR2Care concept surfaced from a need identified by the home care market in which state-of-the-art technology could be applied, and the project aims to leverage the multiuser VR technology from the lab to the market, piloting the prototypes with the involvement of professionals and end-users.
This article addresses functional and technical insights into the VR2Care digital ecosystem. The core is brought to the project by technology partners with their products, which are being adapted and integrated to meet the requirements identified for the specific project pilot scenarios. This vision leverages a system-of-systems approach, a set of independent, useful systems integrated into larger systems that deliver unique capabilities, where individual components are regarded as self-contained operational systems, complemented by additional properties and features from the assemblage of components and specific requirements of the pilots.

Authors

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

Telma Mota, Altice Labs, telma@alticelabs.com

Hugo Paredes, INESC TEC, hugo.paredes@inesctec.pt

Vito Nitti, imaginary, vito.nitti@i-maginary.it

Michael Brandstötter, cogvis, brandstoetter@cogvis.ai

Keywords

Smart living; Active aging; VR; Metaverse; eHealth; Horizon 2020

Context and objectives

"VR2Care - 3D Community Aware Virtual Spaces as Smart Living Environments for Physical Activity and Rehabilitation" is an ongoing two-year innovation project co-funded by the Horizon 2020 Framework Programme of the European Union, bringing together a consortium including ten academic, industrial, and health professional partners from Austria, Italy, the Netherlands, and Portugal [1].

The VR2Care ecosystem is a technological response to the factors that constrain the practice of physical exercise, designed using co-creation methodologies, with three main objectives:

1. To integrate technological, clinical, and social perspectives of using interactive technologies, ensuring that the VR2Care platform is innovative while also extending our scientific understanding and practice-based experiments of engaging the community for active and healthy aging with physical activity and exercise;
2. To develop, implement, and validate the VR2Care platform that leverages cutting-edge approaches including a multiuser virtual reality environment, natural interaction, pervasive monitoring of physical activity, and smart virtual assistants, to produce a platform that is usable, adaptable, extendable, and sustainable;
3. To explore and measure the level of engagement, effectiveness, and impact that the VR2Care platform has on older adults in areas that range from monitoring physical activity and rehabilitation to social interaction, validating this through small-scale pilots involving a community of stakeholders in Europe and providing an exploitation and business plan for the platform adoption.

VR2Care breaks the current virtual reality paradigm in smart living environments by enabling a multiuser mixed reality service in a metaverse context, available for embodied exercising groups in different physical locations simultaneously and with expert exercise supervision. Motivation is empowered in VR2Care by a social commitment: a social experience of several people sharing difficulties and fighting isolation, replaces the paradigm "the patient/user and the application", and the gamified entertainment experiences.

This project leverages the technology by running four demonstrators that support pilots with real users in Italy, the Netherlands, and Portugal. VR2Care goes beyond the common requirements since it is co-designed with the active participation of older adults, caregivers, therapists, community, and clinicians to provide an immersive multiuser experience [1]. The consortium collected knowledge and experience from primary and secondary users, conducting focus groups in the four pilot sites. The results of this work were the key to designing the system architecture [2].

VR2Care platform

The proposed system directly addresses the Horizon 2020 ICT-55-2020 challenge of "developing richer virtual environments, new user interfaces and improved immersion maximizing the feeling of presence" [3]. By combining the requirements and existing solutions, the VR2Care architecture is a system-of-systems that preserves the individual products of the technological partners and provides an integration layer for the VR2Care digital ecosystem [4].

VR2Care high-level architecture main modules are present in **Figure 1**:

- REHABILITY and 3D Multiuser Environment are the platform components providing the VR Environment, with natural interaction and multiuser embodiment functionality enabling multiple VR scenarios;
- SmartAL Telemonitoring allows caregivers and patients to define and execute daily monitoring plans. It includes platform-wide user

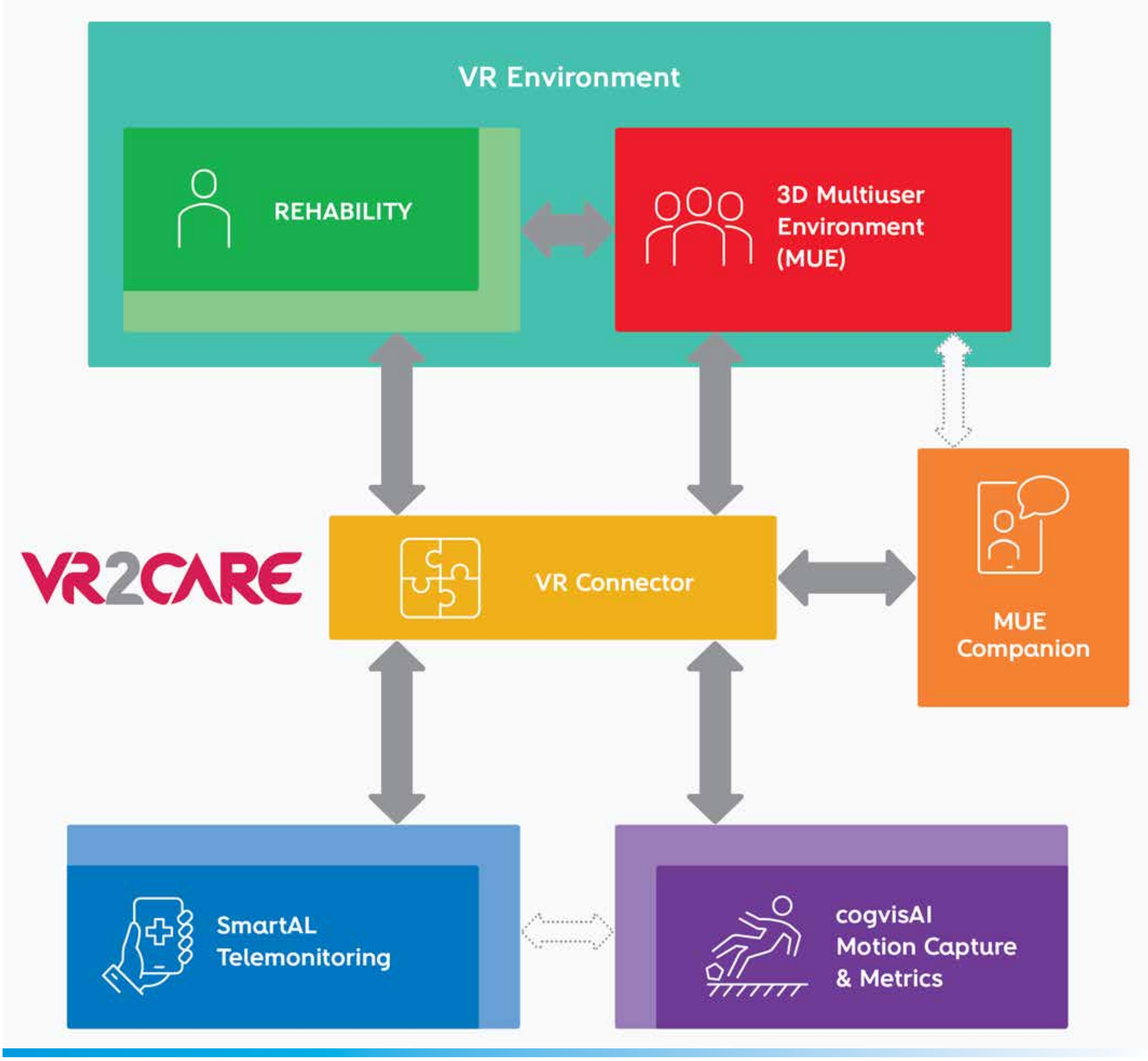


FIGURE 1 – VR2Care high-level architecture

management as well as interfaces with the Identity Provider for registration, authentication, and authorization services;

- cogvisAI is the Motion Capture and Metrics module that detects falls, obstacles, motion, and position while preserving patient's privacy;
- VR connector module is being developed to ensure interoperability and functional independence of the architecture modules while offering communication through well-defined API;
- The MUE Companion is a web app running on a tablet device, conveying coach/trainer functionality required for the multiuser sessions.

Beyond the needed adaptations to support new features and the required new API in the existing products, additional third-party technology is being integrated for specific functionality, including Keycloak identity provider [2], Dissonance voice chat [3], and RabbitMQ message broker [4].

Natural interaction and multiuser embodiment

With the rise of powerful and affordable metaverse technology, VR is often associated with the use of 3D glasses or head-mounted displays. Nevertheless, VR2Care relies on a semi-immersive VR approach [5], which uses a TV screen as a display and computer vision-based human motion capture (mocap) to enable natural interaction. It is a more cost-effective, simpler, and safer option, adequate for the use cases addressed by our scenarios, as shown in **Figure 2**.

The VR Environment is based on the Unity [5] engine running locally on the Home Kit: an Orbbec depth (RGBD) camera [6] connected to a LattePanda [7] pocket-sized Ubuntu-based [8] computer with Nuitrack middleware [9].

FIGURE 2 – Semi-immersive VR, natural interaction

Multimodal natural interaction ensures the usability of the system for end users. In the primary users' environments for physical exercise and rehabilitation, body motion and gestures are the most common input, with big-screen visualization (TV) and audio output. The semi-immersive VR multiuser environment creates users' embodiment using a pose estimation approach based on the depth camera sensor and representing the users in the environment by their avatars. Other interfaces, for secondary and tertiary users, use conventional web and mobile interfaces.

Scheduling and monitoring

The professional guidance and supervision of the user tasks to be performed within the virtual reality environment are a foundational aspect of the VR2Care value proposition. These activities are grouped as telemonitoring user plans that may include the prescription of asynchronous single-user rehabilitation physical and cognitive games, synchronous multiuser physical exercise sessions, and simpler user actions such as vital signs measurement.

After the caregiver defines a telemonitoring plan for each user, he/she will access it through an application and perform the assigned daily tasks. Simple activities such as taking prescribed medication, measuring blood pressure daily, walking three hours weekly, or practicing physical exercise by attending an online gym session, will be organized and scheduled with little effort. Reminders will be sent timely, and user compliance will be easily tracked. Virtual communication with the caregivers and other patients will also be possible, bringing a feeling of belonging to a group, eluding isolation, and encouraging participation.

The remote monitoring solution brought to the VR2Care project by Altice Labs (Portugal) is SmartAL (Smart Assisted Living) [10]. It's available as a mobile app, TV app, and web app and it consists of a technological ecosystem aiming to simplify people's daily lives from both a health and a social point of view. SmartAL has been designed as a flexible solution to address multiple assisted living business cases, which exceed specific VR2Care requirements. From seniors to children, this product helps people with care needs in their daily activities, as well as their families, professionals, and health/social care organizations [6].

The Telemonitoring module interfaces with the Identity Provider and enables patients and professionals user/profile management, as well as scheduling and tracking of user tasks and plans. Furthermore, it allows the connection of biometric devices and lifestyle wearables, including automatically collected measures and data from clinical (e.g., oximeter) and non-clinical devices (e.g., personal band), as well as from other data sources such as questionnaires and clinical survey instruments. The collected data may be compared with thresholds and normative data previously set by the physicians and if the values are outside the limits, notifications are immediately sent, so that actions can be taken by the health professionals according to their severity (e.g., schedule a teleconsultation, consult a specialist, prescribe medication, call emergency services).

New developments include a set of platform-wide gamification features introduced to support aspects of socialization and user engagement, such as awards/badges associated with progress and achievements in assigned tasks.

Virtual reality environment

The VR Environment module provides the assets for the semi-immersive virtual reality to enable access to a set of exercises for both rehabilitation and well-being. This VR container is the main frontend to the users, both in single and multiuser mode and to the infrastructure, enabling natural interaction and multiuser embodiment. On top of it, different environments have been created, maintaining a unified experience in terms of user interface and natural interaction.

A single-user base is used for rehabilitation, as the process requires fine control by the therapist and compliance with the proposed plan. A virtual gym application allows for one-to-one training or one-to-many sessions, where the participants are represented by avatars . Attendants perform physical exercises under the coach/trainer's supervision, with all avatars sharing a common virtual space in a metaverse context [7]. Real-time voice interaction with the trainer is provided, and the scene synchronization is done via the (remote) session server where each of the participants is transparently connected. Furthermore, caregivers and professionals have access to a customized VR module providing specialized functions.

The VR Environment module matches different requirements, both technical and functional, which have been elicited during the system co-design phase. Leveraging the existing building blocks (REHABILITY, SmartAL, and the previous embodiment results from INESC TEC research work), an integrated prototype is being developed to tackle all the technically feasible stakeholders' requests.

Game-based single-user physiotherapy/physical exercise

VR2Care single-user scenario is based on the serious games provided by REHABILITY [11]. This is the solution brought to the VR2Care project by imaginary srl (Italy). It is the result of different EU research projects and the direct investment of imaginary in the subsequent years to attain an industrialized and CE marked product [8], [9].

REHABILITY allows patients to take part in personalized rehabilitation therapy both within a specialist facility, through a web-based dashboard for therapy planning and a PC connected to an RGBD camera, and from home, using a Home Kit, for a semi-immersive 3D VR environment with remote medical support.

The REHABILITY system combines physical with cognitive stimulation offering a range of functional exercises that patients can easily lead back to realistic tasks so that older people and fragile patients understand the usefulness of what they are doing while enjoying the gamified setting. Professionals can quickly personalize the exercises for each single patient in each phase of the therapy/training in a very fine-grained, yet simple way, because games are parametric, thus a very easy-to-use but sophisticated solution for personalized therapy [10].

Game instructions are provided both in text and speech to the user in their native language, and the professionals can configure the system to provide postural guidance during games, deciding the level of interaction during game time (repeating a message more or less frequently). The system is capable of monitoring the following postures:

- Standing/sitting;
- Shoulders alignment (if one shoulder is higher than the other one);
- Spine (check if the spine is straight);
- Hand opening (check if the hand is open);
- Elbow flexion (check if the arm is completely extended).

Some adaptations and new developments have been undertaken to integrate REHABILITY into the VR2Care platform. These include integration work with SmartAL through the VR Connector, synchronization with the multiuser environment sessions, and development of an ANT+ plugin [12] for the REHABILITY suite, to allow automatic pairing/connection when a compatible controller/chest band is detected and the real-time monitoring of heart rate while the user is involved in VR exercises.

The VR2Care REHABILITY version meets most of the requested requirements. The system allows for the setting and assessment of a completely personalized therapy for each user in terms of speed, time, and a wide range of parameters for each single game. Also, the solution has been integrated with the SmartAL user interface and profile management, as shown in **Figure 3**.

3D Multiuser Environment for physical exercise

The 3D Multiuser Environment (MUE) for physical exercise is built on the research experience of INESC TEC (Portugal) on immersive environments and imaginary's expertise in VR.

The first proof of concept of a similar system was developed in 2013 during the research project Online-Gym with Altice Labs (former PT Inovação) [11]. The goal was to demonstrate the use of virtual worlds for the practice of physical exercise, synchronously in small groups, using natural interaction. Online-Gym targeted online multiuser playing by bringing the exergame experience to virtual worlds, with users being represented by

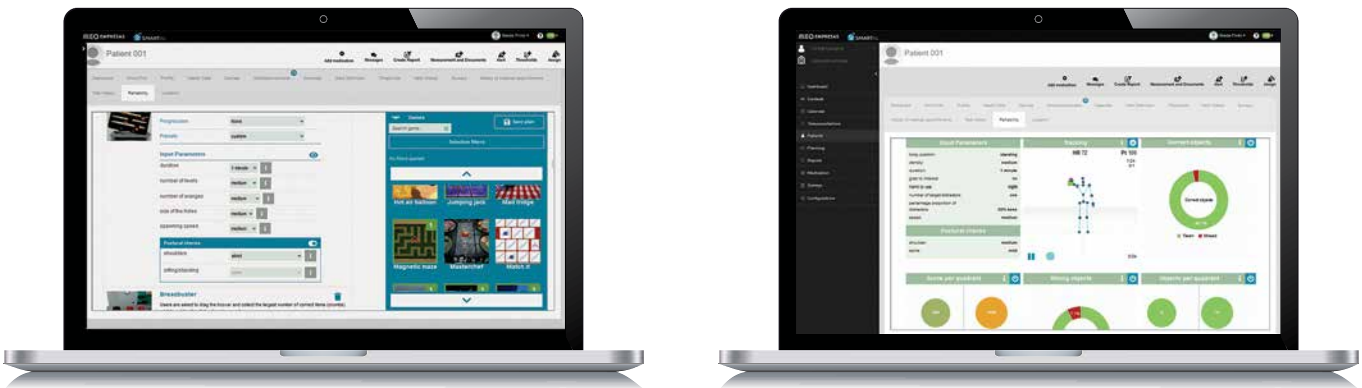


FIGURE 3 – REHABILITY backend via SmartAL user management

avatars, animated using Microsoft Kinect camera pose estimation, ensuring the embodiment, immersion, and co-presence in an OpenSimulator environment (open-source version of Second Life) [13]. Despite users' great interest in the system, the low level of technological readiness (TRL 3) [12] and the complexity of the user interface did not allow the advancement of the solution to a commercial product: a new technological approach was needed to make it more attractive for end users.

The 3D MUE provides two types of sessions that can be scheduled through SmartAL:

- **One-to-one:** this type of session is specific for private and more controlled physical exercise between the patient and the trainer. The session and its environment are designed to provide a one-to-one experience privileging the use of real identities and personalized interaction;
- **One-to-many:** normally used for groups to practice physical exercise with the support of a coach/trainer. One-to-many sessions promote interaction among the participants sharing the same virtual space (**Figure 4**).

FIGURE 4 – 3D MUE one-to-many session

Users are represented in the 3D multiuser environment either as hyper-realistic or game avatars, created with third-party technology (e.g., Didimo [14], Ready Player Me [15]), associated to the user profile, and uploaded to the system through web interfaces.

The immersive environment allows multimodal natural interaction, including voice, gestures, and body commands. Moreover, the interaction can be enriched with the addition of virtual objects that may be shared among the session participants. Despite these multiple interaction options, the avatar cannot move outside the spots specified for practicing exercise. Establishing boundaries for avatar movement in the immersive environment aims to reduce movement in the real world, protecting the user from potential hazards while interacting with the system.

The system provides different virtual spaces that can be selected by the trainer when scheduling a session and the MUE interface adheres to the simplicity of REHABILITY, to ensure a common user experience within the VR2Care ecosystem. Not much information is provided to the user during the interaction in order to draw attention to the trainer's instructions for the exercises to be performed. Guidance can be provided by the trainer as normally happens in a gym, exemplifying and verbally explaining the exercise.

In the interface provided to primary users, all session participants are embodied in their avatars. However, this view is not the best for the trainer, as the focus must be on each user's performance during the session. Therefore, secondary users (trainers) have a different interface where session participants are represented by 3D stickman avatars: simple skeletal figures, built with the joint points of estimated human pose. With this perspective, the trainer can have a clear view of each participant's movement, carrying out the necessary supervision of the exercise.

To ensure a simple and immersive user interface for all users and avoid overloading the interface with information, the MUE Companion application can be used by the trainer on a simple device (smartphone/tablet/laptop) to ensure necessary session control and supervision actions, such as open a private voice channel with a user, mute/unmute the ambient audio channel, turn ambient music on/off, or check correction for each user. Additionally, when available, supervisory information about users, including real-time physiological data, can be provided to the trainer via the companion application.

Motion capture and sensing

The Motion Capture and Metrics module is called cogvisAI, the product brought to VR2Care by cogvis (Austria) [16]. It is built on a 3D smart sensor mounted on the wall or ceiling overviewing the scene, to provide a pose detection system, scene metrics, and safety features.

This module provides extended monitoring capabilities that go beyond the duration of a game or VR session. It may stay operational 24/7, in-game as well as out-game, and its AI-based algorithms can detect several poses directly on the device, such as a person's upright position, sitting, squatting, on the floor, and more. Especially, the detection of falls is an important feature, usable 24/7 to ensure immediate action when it happens [13].

As shown in **Figure 5**, RGB images are not used, thus user privacy is preserved [14]. The module acts passively, so no interaction from the user is necessary for it to work. To use affordable hardware, the system is designed for saving computational resources, and idling until motion is detected.

In addition to API integration work, cogvisAI has been enhanced to dynamically detect obstacles within range of the user's bounding box (**Figure 6**). This is a relevant feature for VR2Care scenarios, e.g., when a pet enters the activity area unexpectedly for the unaware user, focused on the TV screen during a game or gym session.

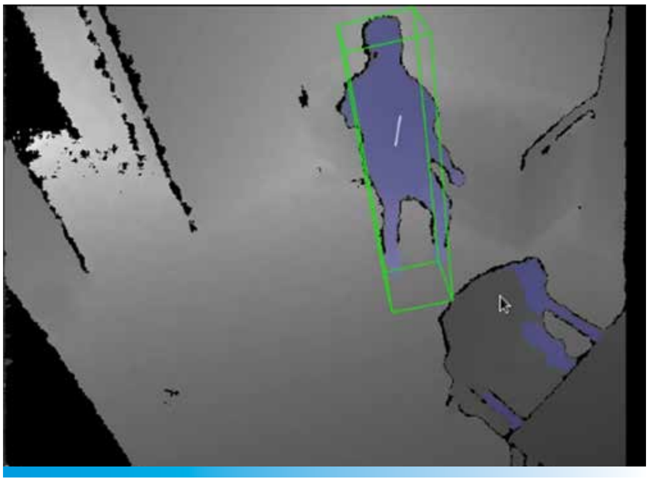


FIGURE 6 – cogvisAI bounding box

Outlook

Throughout the development and integration phases, the VR2Care platform components were iteratively tested by pilot sites, run by Università degli Studi di Napoli Federico II (Italy), Cooperativa Sociale COOSS (Italy), tanteLouise (Netherlands), and Venerável Ordem de São Francisco do Porto (Portugal). This task provided training to professionals to use the system and prepare for the pilot phase, as well as refinement of the requirements produced by co-design activities, coordinated by Smart Homes (Netherlands). Furthermore, the project carried out an analysis of the readiness level of the four pilot sites for the prescription of physical activity and the ability to use an ICT-based environment by all stakeholders (doctors, nurses, gym trainers, patients, and caregivers), and provided educational as well as training sessions to fill the gaps.

FIGURE 5 – cogvisAI lying behavior

A study protocol to assess the extent to which VR2Care improves the management of rehabilitation and adapted physical activity of aged people is agreed upon and carried out by all pilot sites, describing all possible situations to be tested through the pilots, as well as the user enrollment strategy. The piloting phase is ongoing and is expected to help professionals devise how to improve current workflows based on new features available, as well as novel approaches to enhance user engagement in achieving health, wellness, and active aging goals.

From a business perspective, VR2Care opens up a global marketplace to deliver services and products to customers, where distances between provider and customer disappear. The business opportunities go beyond technology and technical companies, broadening the challenges for the Active Assisted Living (AAL) ecosystem, including companies such as broadcasters, architects, therapists, and training and education institutions to offer training, advice, and design anywhere. Ongoing activities coordinated by AFEdemy (Netherlands), explore business models in the emerging market strengthened by the rapid growth of the metaverse.

The scenarios addressed by VR2Care encompass rapidly expanding topics and areas, including assisted living, personal health, metaverse, VR, and AI, so choosing centralized monolithic approaches is neither desirable nor plausible. VR2Care makes evident that approaching the market with services based on technological partnerships and integrating specialized complementary products is feasible and expands its reach. The evolution of this highly fragmented sector could trigger the introduction of standardization that will rationalize and bring flexibility to the tailoring of solutions, fitting effectively to the many aspects of living well and healthily in a smart digital world.

Acknowledgments

"VR2Care - 3D Community Aware Virtual Spaces as Smart Living Environments for Physical Activity and Rehabilitation" is a project co-funded by the Horizon 2020 European Union's Research and Innovation Programme under grant agreement No 951978. Project partners: INESC TEC (Portugal, coordinator), AFEdemy (Netherlands), Altice Labs (Portugal), Università degli Studi di Napoli Federico II (Italy), cogvis (Austria), Cooperativa Sociale COOSS (Italy), imaginary srl (Italy), Smart Homes (Netherlands), tanteLouise (Netherlands) and Venerável Ordem de São Francisco do Porto (Portugal).

References

[1] M. Qbilat *et al.*, "VR2CARE: An Age-Friendly Ecosystem for Physical Activity, Rehabilitation, and Social Interaction," *2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA)*, pp. 1–5, Sep. 2023, doi: 10.1109/IMETA59369.2023.10294571.

[2] M. Qbilat *et al.*, "Co-designing Interactive Technologies for Community Aware Physical Training and Rehabilitation in the VR2Care Digital Ecosystem," *Workgroup on Interactive Systems in Healthcare (WISH) at CHI 2023, April 28, 2023, Hamburg, Germany*, vol. 1, 2023, Accessed: Oct. 16, 2023. [Online]. Available: https://www.vr2care.eu/wp-content/uploads/2023/03/wip_paredes.pdf

[3] "ICT-55-2020 Interactive Technologies." Accessed: Oct. 10, 2023. [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-55-2020

[4] J. Klein and H. van Vliet, "A systematic review of system-of-systems architecture research," *QoSA 2013 - Proceedings of the 9th International ACM Sigsoft Conference on the Quality of Software Architectures*, pp. 13–21, 2013, doi: 10.1145/2465478.2465490.

[5] A. Salatino *et al.*, "Virtual reality rehabilitation for unilateral spatial neglect: A systematic review of immersive, semi-immersive and non-immersive techniques," *Neurosci Biobehav Rev*, vol. 152, p. 105248, Sep. 2023, doi: 10.1016/J.NEUBIOREV.2023.105248.

[6] T. Mota, F. de Carvalho, H. Paredes, and L. Morgado, "Simple VR for better living," *InnovAction*, vol. 2021, no. 6, pp. 16–25, Dec. 2021, Accessed: Oct. 12, 2023. [Online]. Available: https://repositorioaberto.uab.pt/handle/10400.2/11527

[7] J. J. Cummings and J. N. Bailenson, "How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence," *Media Psychol*, vol. 19, no. 2, pp. 272–309, Apr. 2016, doi: 10.1080/15213269.2015.1015740.

[8] "Design and Evaluation of the living lab to assess vCare system 'virtual coaching activities for rehabilitation in elderly' in Parkinson's disease patients - MDS Abstracts." Accessed: Oct. 12, 2023. [Online]. Available: https://www.mdsabstracts.org/abstract/design-and-evaluation-of-the-living-lab-to-assess-vcare-system-virtual-coaching-activities-for-rehabilitation-in-elderly-in-parkinsons-disease-patients/

[9] J. Jonsdottir *et al.*, "Response to commentary of Shirani and Okuda regarding 'Unilateral arm rehabilitation for persons with multiple sclerosis using serious games in a virtual reality approach: Bilateral treatment effect?,'" *Mult Scler Relat Disord*, vol. 37, Jan. 2020, doi: 10.1016/j.msard.2019.101464.

[10] L. Pannese, D. Wortley, and A. Ascolese, "Gamified wellbeing for all ages – How technology and gamification can support physical and mental wellbeing in the ageing society," *IFMBE Proc*, vol. 57, pp. 1281–1285, 2016, doi: 10.1007/978-3-319-32703-7_246/COVER.

[11] F. Cassola, L. Morgado, F. de Carvalho, H. Paredes, B. Fonseca, and P. Martins, "Online-Gym: A 3D Virtual Gymnasium Using Kinect Interaction," *Procedia Technology*, vol. 13, pp. 130–138, Jan. 2014, doi: 10.1016/J.PROTCY.2014.02.017.

[12] "Technology readiness levels." Accessed: Oct. 10, 2023. [Online]. Available: https://www.nasa.gov/wp-content/uploads/2017/12/458490main_trl_definitions.pdf

[13] C. Pramerdorfer, R. Planinc, M. van Loock, D. Fankhauser, M. Kampel, and M. Brandstötter, "Fall detection based on depth-data in practice," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9914 LNCS, pp. 195–208, 2016, doi: 10.1007/978-3-319-48881-3_14/FIGURES/5.

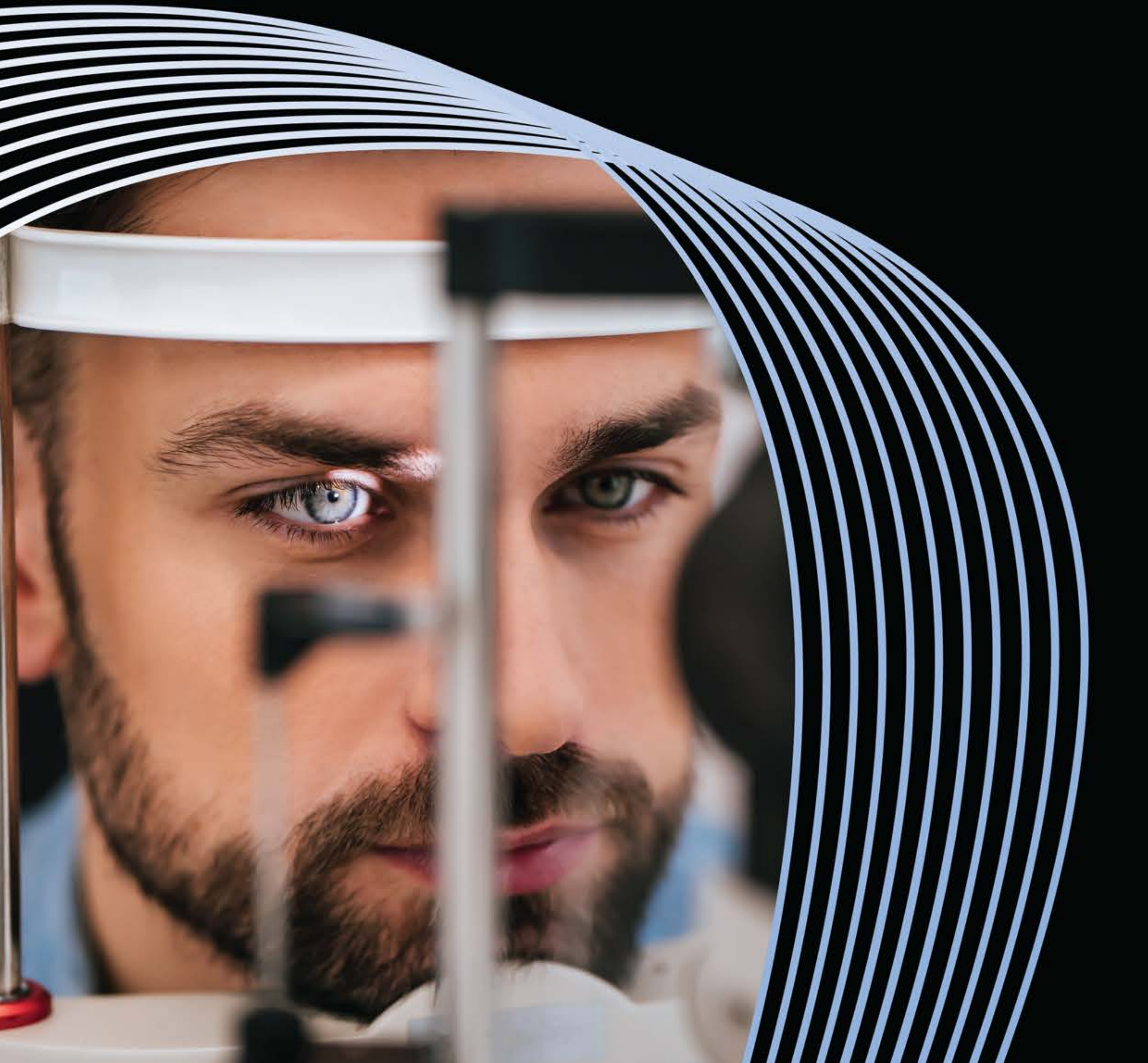
[14] M. Brandstötter, J. Lumetzberger, M. Kampel, and R. Planinc, "Privacy by Design Solution for Robust Fall Detection," *Stud Health Technol Inform*, vol. 306, pp. 113–119, Aug. 2023, doi: 10.3233/SHTI230604.

Endnotes

1 https://www.vr2care.eu/
2 https://www.keycloak.org/
3 https://placeholder-software.co.uk/dissonance/
4 https://www.rabbitmq.com/
5 https://unity.com/
6 https://www.orbbec.com/products/structured-light-camera/astra-series/
7 https://www.lattepanda.com/lattepanda-sigma
8 https://ubuntu.com/download/desktop
9 https://nuitrack.com/
10 https://www.alticelabs.com/products/ehealth-smart-assisted-living/
11 http://www.rehability.me/
12 https://www.thisisant.com/
13 http://opensimulator.org
14 https://www.didimo.co/
15 https://readyplayer.me/
16 https://cogvis.ai/

DOI: 10.57650/INNOVACTION/2023.09

09 Glaucoma diagnosis automation



Glaucoma is one of the leading causes of irreversible blindness worldwide. It encompasses a group of conditions that share a common characteristic: optic nerve damage. The optic nerve, responsible for transmitting visual information from the retina to the brain, undergoes structural and functional changes, resulting in a gradual, often unnoticed, loss of vision. These changes are primarily attributed to elevated intraocular pressure (IOP), stressing the optic nerve fibers, impairing their functioning and causing sudden degeneration in the case of closed-angle glaucoma, or a more progressive loss of vision in the case of open-angle glaucoma, which is the focus of this paper.

The clinical diagnosis of glaucoma involves a combination of qualitative and quantitative assessments, such as visual acuity, visual field testing, and optic nerve head evaluation using imaging techniques like optical coherence tomography (OCT). However, although these conventional diagnostic methods are helpful, they have inherent limitations, such as variability in human interpretation, subjectivity, and reliance on the clinician's expertise. The integration of Machine Learning (ML) approaches has emerged as a promising technique to aid in glaucoma detection, either by directing the professional's attention to riskier cases, or simply by acting as a second opinion during the final diagnosis.

This paper describes the cognitive challenge and technical approach to implement a model to automate glaucoma diagnosis. It describes the AI/ML process applied as well as the associated complementary explainability methods as, in real-life situations, it is valuable for healthcare professionals to comprehend what factors lead to a specific decision to determine whether to support it. Finally, some insights on the outcomes achieved and how to move on for an in-field evaluation are provided.

Authors

Rodrigo Ferreira, Altice Labs, rodrigo-m-ferreira@alticelabs.com

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

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

António Travassos, Centro Cirúrgico de Coimbra, oftalexames@ccci.pt

Robert van Velze, Centro Cirúrgico de Coimbra, oftalexames@ccci.pt

Keywords

Machine learning; Data science; Glaucoma

Introduction

Glaucoma, a group of progressive optic pathologies, remains a significant global health concern affecting millions worldwide. This sight-threatening condition is characterized by damage to the optic nerve, which leads to irreversible vision loss when left undiagnosed and untreated. As the leading cause of irreversible blindness worldwide, glaucoma presents an urgent need for accurate and timely detection to mitigate its detrimental effects on affected individuals.

This pathology constitutes a considerable burden on public health, with its prevalence steadily rising as the population ages. It is estimated that over 80 million people globally are affected by glaucoma and is projected to reach about 112 million by 2040. The condition is particularly prevalent among older individuals (60+ years old) and varies among different ethnicities and geographical regions.

The disease's progressive nature often leads to a gradual loss of peripheral and central vision, referred to as visual field loss. When left untreated, glaucoma can progress to complete irreversible blindness, impairing a person's ability to perform daily activities, navigate their environment, and engage in social interactions.

One of the primary challenges in managing glaucoma lies in its early stages' silent and asymptomatic nature. This characteristic makes a timely diagnosis challenging, as individuals may remain unaware of their condition until it is too late and irreversible damage has been done. Additionally, the diagnostic process involves several complexities, including the reliance on subjective assessment, observer variability, and the need for frequent and extensive monitoring. Traditional diagnostic techniques, such as intraocular pressure measurement and visual field testing, generate data with inherent limitations in terms of reliability. Furthermore, the scarcity of skilled ophthalmologists and the increasing global glaucoma burden requires the development of innovative and accessible diagnostic tools.

In terms of the exams, those that are most relied upon when checking for glaucoma are:

- **Optical Coherence Tomography (OCT):** non-invasive imaging technology used to get a detailed mapping of the inside of the eye, by following the logic of an ultrasound, but using invisible light instead of sound waves;
- **Visual Field Test (VFT):** non-invasive exam to assess the state of the patient's peripheral vision by asking him to focus on a central spot and press a button whenever he notices lights flicker around the fixation light. The exam is performed one eye at a time, with the other eye covered.

Figure 1 presents examples of the exams, highlighting the most relevant features.

The result of the OCT is a set of high-resolution images of different parts of the eye. These images are often summarized using metrics that are aggregated and presented in reports, relative to the specific part of the eye being evaluated. Examples of those reports include:

- **Biometry:** captures several measurements to understand different characteristics such as the axial length, corneal curvatures, and angle and thickness of various parts of the eye;
- **Anterion:** like biometry, it takes several OCT measurements of the patient's eye, with a focus on the frontal part. It also measures the angle of the anterior chamber;
- **Retinal Nerve Fiber Layer (RNFL):** measures the thickness of the nerve layers, providing a quantitative and qualitative assessment, useful for obtaining a reliable estimate of the damage to the optic nerve if it is found to be decreasing in thickness;
- **Glaucoma Overview (GO):** measures retina thickness, asymmetry, and ganglion cell layer thickness;
- **Deviation Map Single Report (DMSR):** quantitative and qualitative evaluations of the macular retina and macular ganglion cell layer.

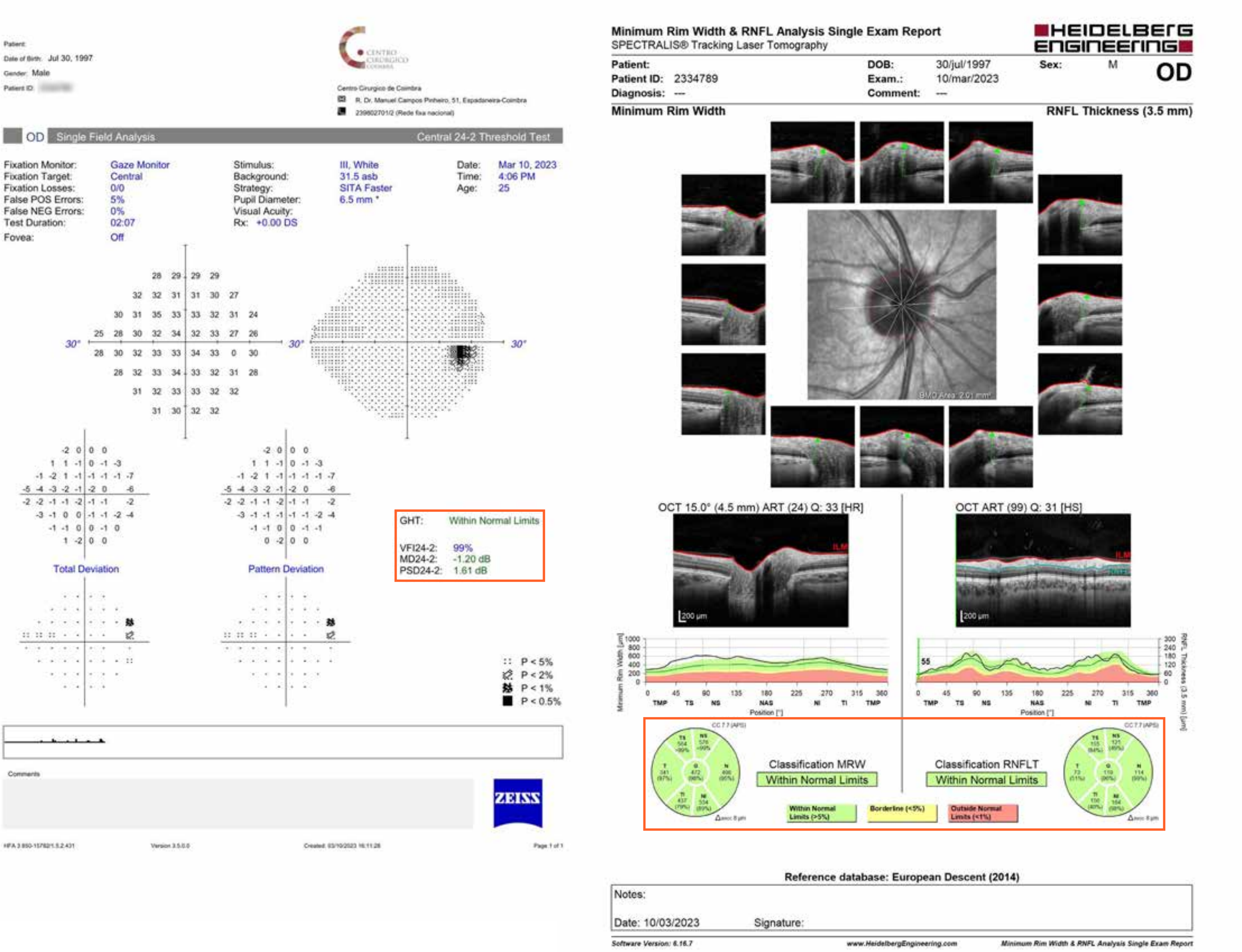


FIGURE 1 – Example of the VFT and RNFL exams, highlighting the most relevant features

As far as the VFT is concerned, the report contains a map of the eye's blind spots as well as other measurements that compare the state of the eye with itself in the past (follow-up), as well as with the general population. It is worth noting that the VFT is prone to errors since it is somewhat subjective and exposed to many outside factors that can influence the results directly, such as false positives (when the patient presses the button for some reason despite no light being shown) and false negatives (when the patient doesn't press the button when the light is turned on although it was expected to have pressed, based upon earlier responses).

In recent years, advancements in ML techniques have shown promising potential for improving the detection and monitoring of various health conditions, namely glaucoma. By leveraging large datasets, ML algorithms can extract intricate patterns and features from clinical data, enabling more accurate and efficient identification of possible cases of glaucoma.

With this work, we hope to contribute to the ongoing efforts to develop robust and reliable ML solutions that can aid clinicians in the early detection, monitoring, and management of glaucoma. Ultimately, the integration of ML techniques into clinical practice presents immense potential to revolutionize glaucoma detection, leading to better outcomes regarding the patients' ocular health in the future.

Relevant works

The use of ML for glaucoma diagnosis in the literature has been gaining popularity in the last few decades and has been following two main approaches: one that uses features already extracted from medical exams in a tabular approach, while the other uses high-quality medical scans of the eye and Deep Learning (DL) to derive its own features. The feature engineering and the tabular nature of the data used in the first approach usually lead to more explainable models from which we can extract relevant insights. On the other hand, DL often leads to better classification results at the cost of interpretability. **Table 1** presents a summary of most relevant works.

Maetchke *et al.* [1] used RNFL thickness values, rim area, and disc area features, among others, to predict glaucoma in a dataset with 263 healthy and 847 glaucomatous eyes, achieving Area Under the Curve (AUC) scores between 0.82 and 0.89, while a DL approach using a Convolutional Neural Network (CNN) to predict glaucoma from fundus images [1] achieved a better AUC score of 0.94. Although important, these results significantly depend on the size of the dataset used to compute them.

1 Fundus imaging is a noninvasive technique used to capture a detailed image of every aspect of the retina, including the size, shape, and color of significant regions such as the optic disc (OD), optic cup (OC), blood vessels, neuro-retinal rim, and fovea, that is used to indicate glaucoma presence.

Most of the published results on this subject rely on significantly smaller datasets, especially when it comes to the glaucoma class. Kim SJ *et al.* relied on tabular data from multiple exams to predict glaucoma on a dataset with 178 glaucomatous and 164 healthy eyes with classic ML algorithms, achieving high accuracy scores around 0.97.

Muhammad *et al.* [2] similarly worked with both summary metrics from OCT scans and the raw exams by using a CNN to extract features from the scans. They worked with 57 and 45 glaucoma and non-glaucoma (healthy or suspected glaucoma cases), respectively, obtaining significantly better results - 93% accuracy - when using the features extracted using CNN, compared to the 64% accuracy using OCT summary metrics and visual field test features, following a classic ML approach.

Asaoka *et al.* [3] relied on the visual field exam alone to distinguish between 53 glaucomatous and 108 healthy eyes comparing the use of DL and classic ML algorithms, where Deep Neural Networks (DNN) achieved a top AUC score of 0.926 while the alternatives all had scores below 0.80.

In this use case, the more standard approach with DL uses other types of imageology exams, mainly fundus and sometimes raw OCT scans (instead of summary metrics), rather than visual field tests. A meta-analysis by Jo Hsuan Wu *et al.* [4] revealed that approaches relying on fundus achieved an average of 0.97 AUC, while OCT's AUC was 0.96.

Paper	Features	Data	Results
Maetchke *et al.*	RNFL	Glaucoma: 847 Non-Glaucoma: 263	CNN: 0.94 AUC ML: 0.82 – 0.89 AUC
Muhammad *et al.*	OCT (summary and raw)	Glaucoma: 57 Non-Glaucoma: 45	CNN: 0.93 Accuracy ML: 0.64 Accuracy
Asaoka *et al.*	Visual field exams	Glaucoma: 53 Non-Glaucoma: 108	DNN: 0.926 AUC ML: <0.80 AUC

TABLE 1 – Summary of relevant works

Methods

Data sources and ingestion

The work described in this paper has been developed in partnership between Altice Labs and the Centro Cirúrgico de Coimbra (CCC), which provided different kinds of exams for healthy, suspect, and glaucomatous eyes. They were fundamental with the necessary expertise for the selection of the most relevant and reliable expected outcomes, which was of paramount importance for its validation.

Besides the anonymized exams' data, a diagnosis was also provided, indicating whether the case had been diagnosed as normal (non-glaucoma), suspect (of glaucoma), or glaucoma. Information regarding additional pathologies the patient might suffer, which is useful in ensuring that the model distinguishes eyes with glaucoma vs. non-glaucoma, instead of healthy vs. unhealthy was also included. It was decided to keep the glaucoma label as pure as possible by filtering out eyes with glaucoma and other pathologies simultaneously during training (although the test set isn't subjected to this filtering).

The exams were sent in an image format, and scripts had to be created, specifically tailored for each exam type, to extract the relevant information using OCR technology, particularly Tesseract [5]. Given the nature of the exams/reports, we opted for the tabular and classic ML approaches previously mentioned to process the data.

Features and dataset creation

For the development of our dataset, besides the features directly extracted from the exams, we took also into consideration features derived from them. The final feature selection process evolved as an iterative approach, and considered:

- The feedback from CCC, which provided insights into the reliability of the exams as well as the priority features to look at when making a diagnosis;
- Correlation studies among the features themselves and between the features and the target, along with statistical analysis;
- The results of the feature importance assessments conducted during the testing with various models.

Training

ML models

Throughout our experiments, we relied mostly on classic ML models such as Logistic Regression (LR), K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM), with the most out-of-the-ordinary models being the Light Gradient Boosting Model (LGBM) [6] and the Explainable Boosting Machine (EBM) from the InterpretML library [7]. The choice of the first four algorithms (LR, KNN, RF, and SVM) was primarily based on their common occurrence in the relevant literature, while LGBM was chosen since this type of gradient-boosted trees tend to achieve state-of-the-art performance on classification tasks with tabular datasets and it did perform the best. As for EBM, it was chosen because it allows for better explainability, but we found that our explainability tool (ExplainerDashboard [8]) was enough to satisfy our needs in a model-agnostic manner. All this considered, LGBM was selected as the final model to apply.

Setup

In preparing our machine learning model, one of the initial steps involved addressing missing values within the dataset. These missing values resulted from either the absence of the corresponding exam or the absence of the whole exam itself. To manage this, we opted to:

- Exclude the cases where an entire exam was missing;
- Use the KNN imputer method, which works by finding the nearest neighbors to the entry

with a missing value and imputing the missing value as a weighted average of its neighbors; this imputer was fitted on the train set to avoid information leakage.

The resulting class distribution of the number of eyes was the following (**Figure 2**):

- Normal: 520 (58.3%)
- Suspect: 169 (18.9%)
- Glaucoma: 203 (22.8%)

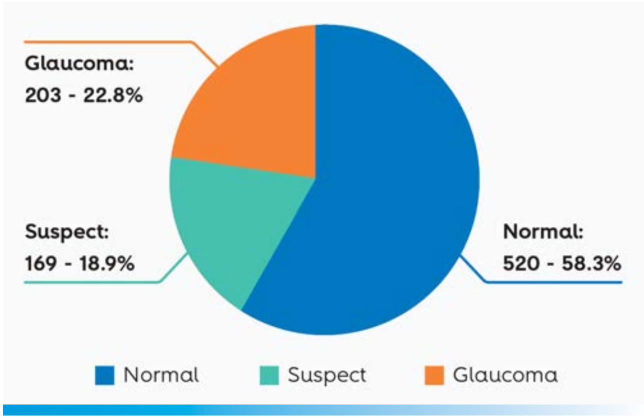


FIGURE 2 – Class distribution of the full dataset

We divided this dataset into train and test subsets (75%/25%), ensuring a similar class distribution in both sets. To prevent information leakage and maintain independence, we split the data at the patient level rather than the individual eye level. This approach guarantees that no patient has one eye in the training set and the other in the testing set.

Regarding the targets, we opted for a multi-class scenario to make full use of the diagnosis information available. In early experiments, we adopted a binary classification scenario (labeling suspicion of glaucoma as non-glaucoma) but we found that according to the resulting metrics, including a separate class for the suspects helped in differentiating glaucoma from non-glaucoma.

Our final feature selection method relied on finding the best feature sets, according to different feature selection methods, and then comparing each in a 4-fold cross-validation of the training set, picking the one that presents better results. The current feature set includes features extracted from the RNFL, DMSR, and GO exams, meaning that to make a prediction, the model evaluates the thickness of the retinal nerve fibers, the macular retina, and the macular ganglion cell layer, as well as the asymmetry found in thicknesses between the top and bottom hemispheres of the eye.

Regarding the hyperparameter tuning, we followed different approaches throughout the stages of the work, from grid search to randomized search and to genetic algorithms in the latest iterations. The main goal of this tuning was to avoid overfitting, by adapting the model's hyperparameters, since this was a problem identified during earlier iterations.

It is worth mentioning that we've also experimented with cascade learning approaches, as referred in the literature, where we divided the problem into different sub-problems, but the results were mostly subpar and did not justify the added complexity of implementing multiple models manually.

Results

We've incorporated diverse classification metrics, including Accuracy, Precision, Recall, F1-Score, and ROC-AUC (see **Table 2**), along with a thorough analysis of confusion matrices, to comprehensively assess model performance [10]. Given the imbalance in class distribution, special importance was given to the F1-score metric, especially of the positive class (Glaucoma), so, when deciding on certain parameters or features, this is what we tried to maximize.

Class	Glaucoma	Suspect	Normal
Precision	0.87	0.41	0.77
Recall	0.77	0.17	0.92
F1-Score	0.82	0.24	0.84
ROC-AUC	0.91	0.74	0.86
Accuracy	0.92	0.80	0.80

TABLE 2 – Classification metrics of the model

The F1-scores and confusion matrices (**Figure 3**) obtained from the final model suggest that it is capable of distinguishing eyes with glaucoma from those of the normal class while being consistently worse at classifying "suspect" ones. This might be explained by the distributions of the feature values in the normal and suspect classes, which tend to be more similar to each other than to the glaucoma class feature value distribution. However, it is also limited by the CCC's criteria for what constitutes a suspect eye, as it lies in a "gray zone". The results achieved with our evaluation metrics seem to be in line with those found in similar works when compared with the metrics that are most commonly reported (ROC-AUC), despite our focus lying on optimizing F1-scores, which were also good for the non-glaucoma and glaucoma classes. However, one must consider that a comparison with the relevant works isn't as simple as looking at the metrics, since this isn't a binary problem, as is the case in most of the papers we found. Despite bringing up some classification difficulties, we still consider the inclusion of the suspect class to be the correct approach since it is valuable information that otherwise would be wasted, and it is useful to distinguish sure cases from those more borderline in a real-life scenario.

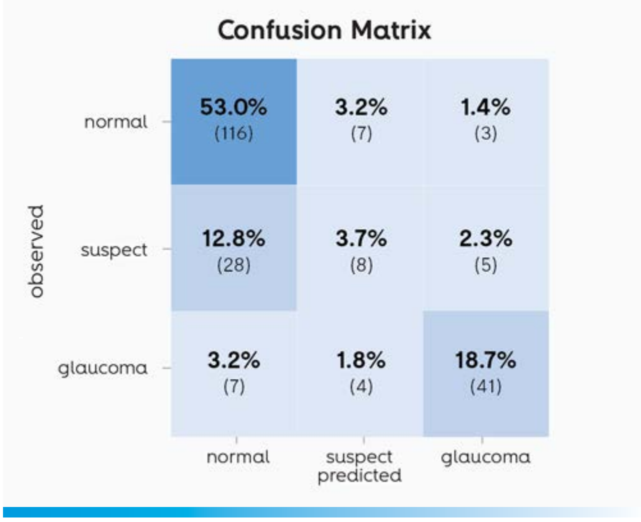


FIGURE 3 – Confusion matrix of the final model

In this scenario, the most critical errors involve misclassifying eyes, with a suspicion or actual glaucoma being labeled as normal, potentially leading to patients not receiving the necessary care. Other errors, though undesirable, are more tolerable as they still flag normal patients for monitoring.

To ensure that the approach chosen is robust, we've decided to experiment with training the same model using 100 different train/test splits. The distribution of F1-scores per class across all runs is shown in **Figure 4**, where we can see that

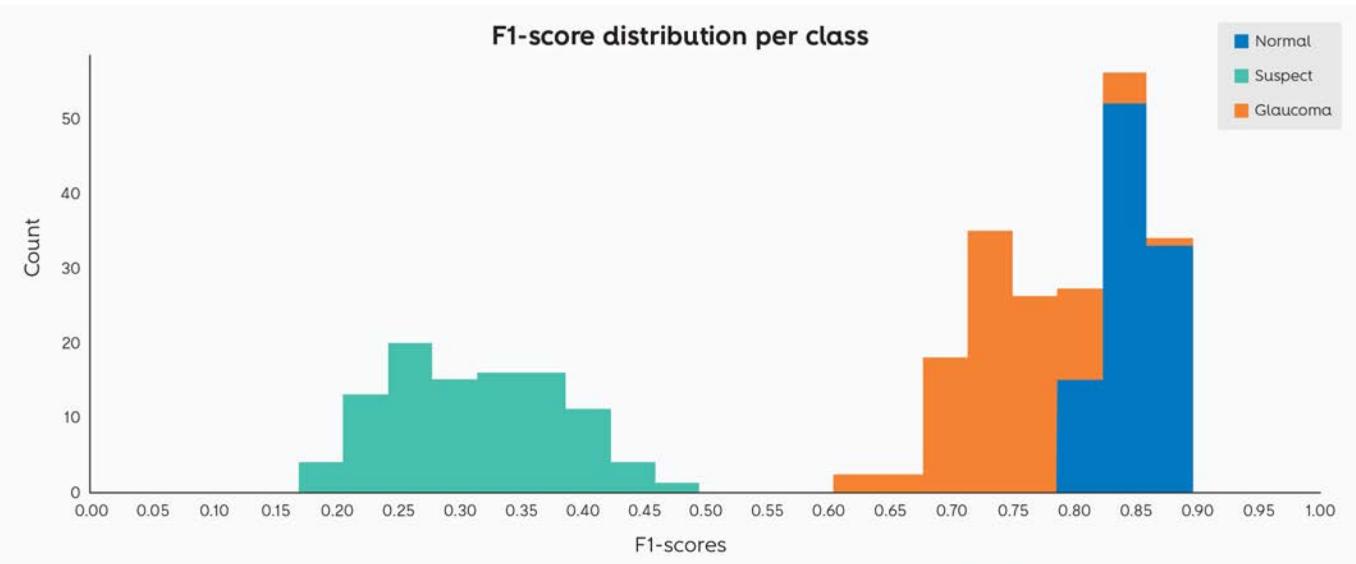


FIGURE 4 – Distribution of F1-scores per class when training and testing the final model with 100 different data splits

the suspect class is consistently the hardest to classify, while the model tends to predict normal or non-glaucoma the best, with glaucoma showing good results across different splits.

Explainability

Regarding explainability, we relied on a tool called Explainerdashboard [8], which could be of great use in a real-life scenario since it generates visualizations that allow the healthcare professionals to assess how relevant (and toward which diagnosis) each feature was to the final result for a particular case (a way in which it could be used in a real scenario since professionals focus their attention on one patient at a time). This tool allows the clinician to measure how the diagnosis would change if some particular values were altered (by letting the clinician customize some parameters) (see **Figure 5**).

The Explainerdashboard uses primarily SHAP [9] values to help visualize how each feature influences the model's final decision. SHAP values indicate how much and in which direction a feature's value influences the model's prediction. For example, a large positive value means that a given feature has a significant impact in making the model predict the positive class (glaucoma in this case). Likewise, a large negative value skews the model toward a non-glaucoma diagnosis, while a small value has a minor impact either way.

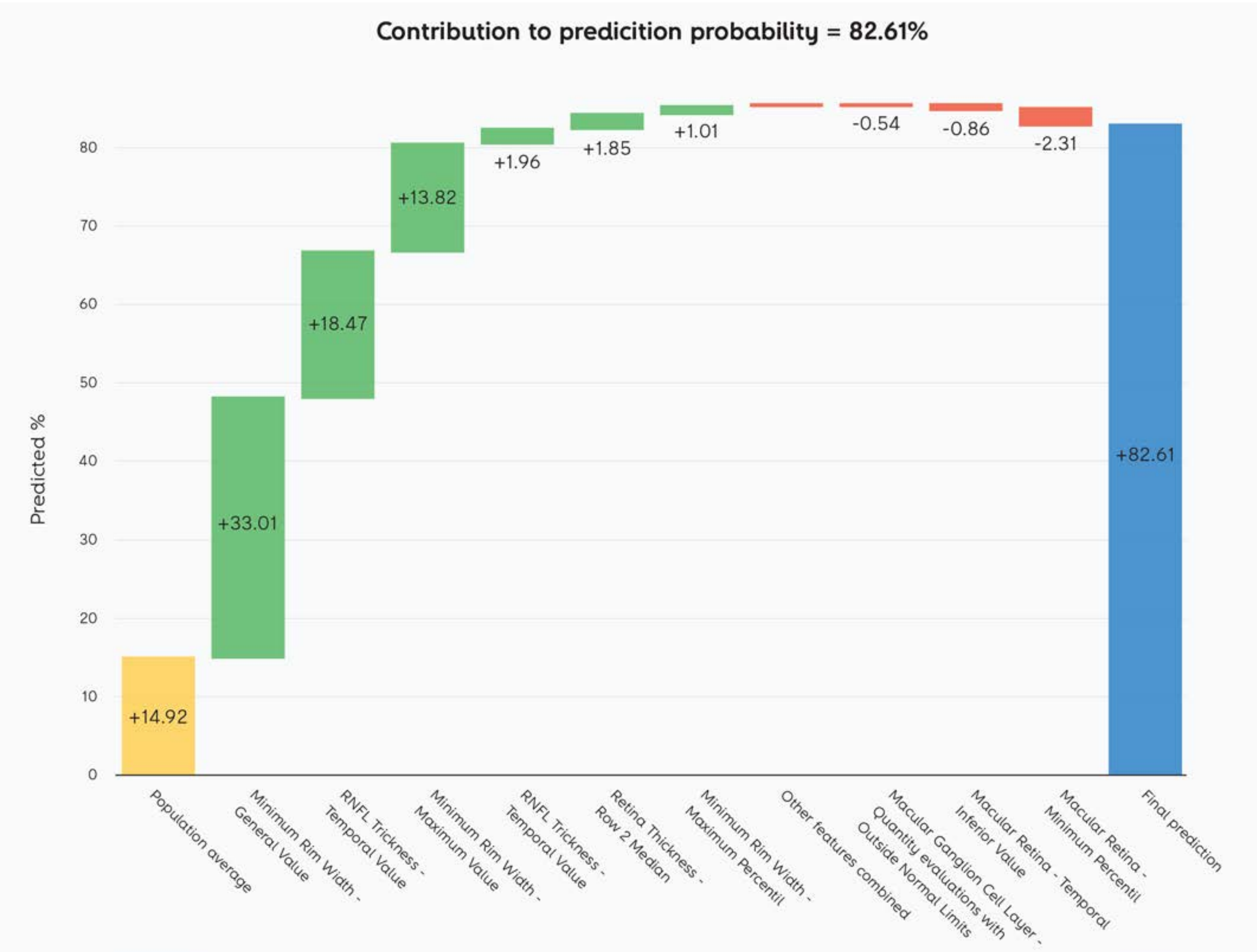


FIGURE 5 – Example of the use of Explainerdashboard to interpret the decision of one case in particular (an eye with glaucoma that was correctly classified as such in this example)

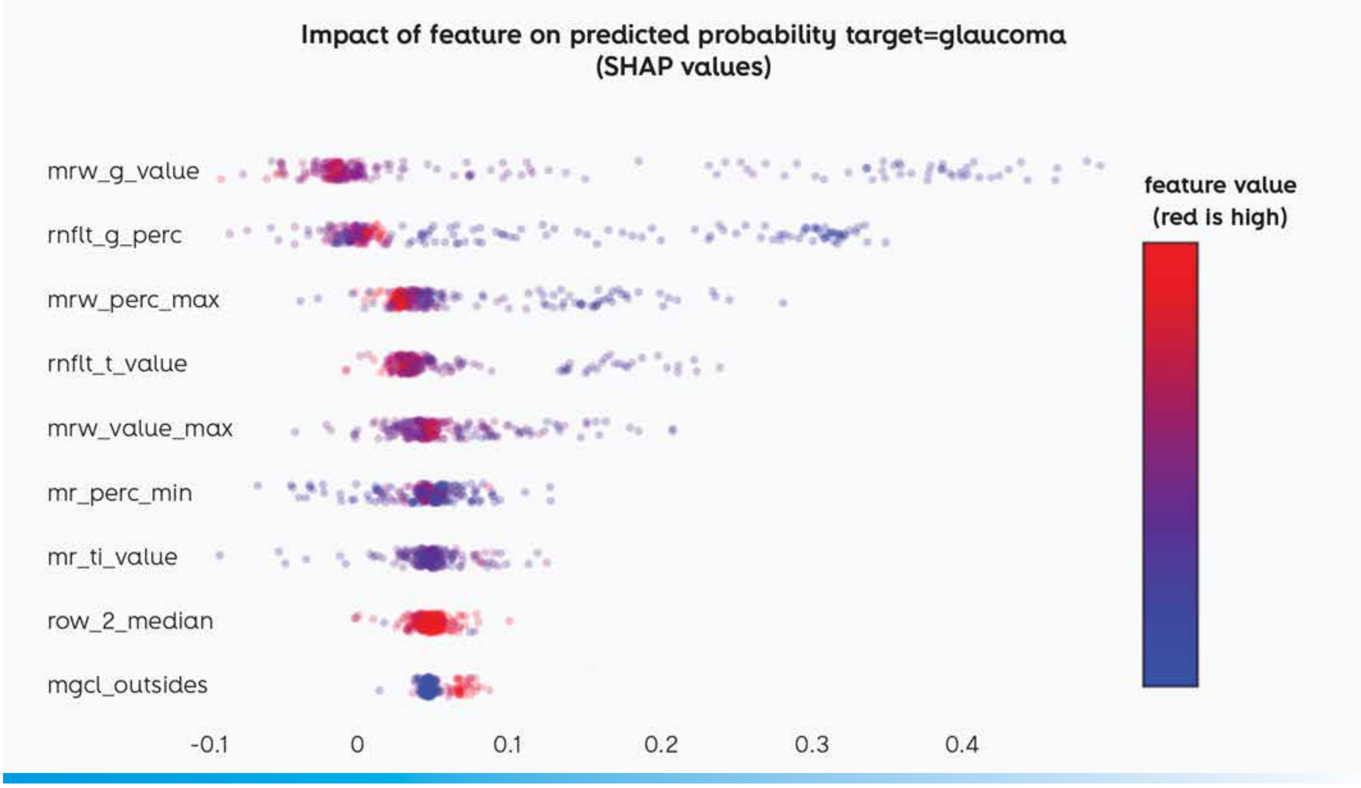


FIGURE 6 – Analysis (SHAP) of the impact of each feature in the glaucoma diagnosis (ordered from most to least impactful)

After analyzing these visualizations, namely **Figure 6**, we came to some conclusions that seem to be corroborated by the medical community, regarding glaucoma diagnosis.

Each line of the graph represents a feature, ranked by impact, and each point represents an eye. Points leaning toward blue denote lower feature values, while those leaning toward red signify higher feature values. The point's position indicates its impact on the model's decision: centered points had no impact, rightward points leaned toward a glaucoma diagnosis, and leftward points leaned toward a non-glaucoma diagnosis.

Overall, one can see the relevance of the RNFL exam, once its features top the feature importance charts throughout the various iterations, including in the presented iteration, where the top 5 features all belong to the RNFL exam. In many cases, RNFL features would be enough to provide a decently accurate diagnosis by themselves, without relying on any other exams at all.

Conclusions

In conclusion, the results obtained with this work look promising in terms of the ability to detect glaucoma, one of the most widespread and challenging ocular pathologies to diagnose. The results are in line with the best results of the relevant work identified in the reference papers and provide a clear path to the accomplishment of the proposed objectives. It was shown that each iteration offers better results than the previous ones, clearly benefitting from the access to more samples, despite still some model instability, which will probably soften with even more samples.

One of the challenges faced during the development of this project, perhaps the major one, was the unreliability of the OCR process, although it could be mitigated by extracting raw values directly from the medical equipment. In addition, one of the main issues in most ML projects is how well the data science team can cope with the specificities of the area under study and, therefore , create domain knowledge – relevant exam fields, non-evident correlations, etc. - to support the model evolution. Close involvement with CCC experts – ophthalmologists, biomedical engineers, and exam technicians – has proven crucial in the process and should be shepherded and intensified along the way.

The main goal is to introduce these processes into daily clinical practice. For now, we are envisaging two different approaches, which would require a slightly different model tuning. If the goal is to screen a vast array of patients and to remove those that pose a low risk of developing glaucoma, or point out those at high risk, we should tune it with a particular focus on the precision of the non-glaucoma class or the recall of the other two classes. If it is the case that we are looking to provide a professional who has already evaluated the patient's exams with a second opinion, the current model, along with the explainability tools, should be able to provide that assistance.

A natural evolution will be to move from diagnosis to predicting the risk of progression to pathology – to allow patients to receive proper care at the first signs of possible glaucoma development – which will require some additional history of tests and different technical approaches, given that we will be dealing with a use case with different characteristics. This work is already underway.

Aside from that, there are a few main options that could be further explored. One is the inclusion of high-definition imageology exams to apply DL along with the classic ML approach. Another one is the use of Neural Networks (NNs) as classification models. Still, a lot of high-quality samples would be needed, which does not seem realistic at this point without significant data augmentation (which could potentially introduce its fair share of problems as well). Finally, another direction would be to follow a similar methodology to address different ocular pathologies.

References

[1] S. Maetschke *et al.*"A feature agnostic approach for glaucoma detection in OCT volumes," *PLOS ONE*, vol. 14, no. 7, p. e0219126, Jul. 2019, doi: 10.1371/journal.pone.0219126.

[2] H. Muhammad *et al.*, "Hybrid Deep Learning on Single Wide-field Optical Coherence tomography Scans Accurately Classifies Glaucoma Suspects," *J Glaucoma*, vol. 26, pp. 1086–1094, 2017. Accessed: Nov. 2, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:205683952.

[3] R. Asaoka, H. Murata, A. Iwase, and M. Araie, "Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier," *Ophthalmology*, vol. 123, no. 9, pp. 1974–1980, 2016, doi: 10.1016/j.ophtha.2016.05.029.

[4] JH Wu, T Nishida, RN Weinreb, JW Lin, "Performances of Machine Learning in Detecting Glaucoma Using Fundus and Retinal Optical Coherence Tomography Images: A Meta-Analysis" *Am J Ophthalmol*, vol. 237, pp. 1–12, 2022, doi: 10.1016/j.ajo.2021.12.008.

[5] Tesseract, "Tesseract doumentation", 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://tesseract-ocr.github.io/.

[6] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. Accessed: Nov. 2, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[7] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A Unified Framework for Machine Learning Interpretability," *arXiv preprint* arXiv:1909.09223, 2019.

[8] O. Dijk *et al.*, "oegedijk/explainerdashboard: explainerdashboard 0.4.2: dtreeviz v2 compatiblity." *Zenodo*, Feb. 2023. doi: 10.5281/zenodo.7633294.

[9] S. M. Lundberg and SI. Lee, "A unified approach to interpreting model predictions," CoRR, vol. abs/1705.07874, 2017. Accessed: Nov. 2, 2023. [Online]. Available: http://arxiv.org/abs/1705.07874.

[10] Wikipedia, "Evaluation of binary classifiers", 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Evaluation_of_binary_classifiers.

DOI: 10.57650/INNOVACTION/2023.10

10 Data governance: understanding, assessing, and evolving

Data is becoming increasingly central for businesses, requiring robust governance practices. On the other hand, collaborative business networks are also expanding, aiming to combine resources, create synergies, and develop capacities to quickly adapt to changes in the market. These partnerships often lead to the creation of innovative data-based products and services.

Data governance aims to capitalize on data as a key corporate asset and address data-related risks. Generally, a data governance framework defines the decision rights and accountability for the organization's data assets. Furthermore, it defines the organizational structures and the policies, processes, and standards that guide all the data-related activities of the organization.

With organizations shifting from centralized to decentralized scenarios, there is a need to develop new data governance mechanisms. In addition, the increasing collaboration between the parties requires deploying data-sharing mechanisms and the technical structures that support them. Moreover, there is a need to clarify data ownership and accountability over data-related operations.

One of the most critical steps is the assessment of the data governance maturity, which aims at overseeing, diagnosing, and assessing the organization's data governance capacities. This process requires evaluating how the organization performs in various data governance dimensions (e.g., data quality, data security, and data architecture) based on a set of criteria defined by the maturity model. Based on the results of this assessment, it is possible to define a roadmap for the improvement of the existing practices.

In this article, we explore the first steps toward developing a data governance maturity model designed with the specificities of an organization like Altice Labs in mind, but also considering its integration into a broader telecommunications group.

Authors

Vítor Ribeiro, Universidade de Coimbra – CISUC – DEI, vhribeiro@dei.uc.pt

João Barata, Universidade de Coimbra – CISUC – DEI, barata@dei.uc.pt

Paulo Rupino da Cunha, Universidade de Coimbra – CISUC – DEI, rupino@dei.uc.pt

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

Keywords

Data governance; Maturity assessment; Maturity model

Introduction

Data is a key strategic asset to compete in nowadays markets, requiring new governance practices that "*specifies decision rights and accountabilities for an organization's decision-making about its data*" [1]. Therefore, it is a priority for organizations to develop and implement a data governance framework to handle the existing challenges. In 2021, about 79 zettabytes of data were created worldwide and this number is expected to double by 2025 [10]. Other difficulties include the need to deal with different sources that may cause inconsistencies [1], the impact of regulatory compliance (e.g., General Data Protection Regulations (GDPR), Artificial Intelligence Act) [1], the fragmented enterprise architectures [24], and the need to align the business and corporate goals of the organization with their data resources and operations [30]. Additionally, data governance is a foundation of trustworthy Artificial Intelligence (AI) systems [16]. In a setup such as telecommunications (the case of Altice Labs (ALB)), data governance is required to handle the increasing volume, rate, and complexity of big data (e.g., voice calls, message data, Internet usage) that organizations are dealing with [30]. In addition, the constant but inevitable evolution toward the concept of Autonomous Networks will have a major impact on the need for efficient and intelligent processes, which will enable the integration and correlation of data from multiple and diverse sources.

The expanding collaborative networks are now bringing new challenges in developing data governance [15]. With organizations exploiting the exchange of data across several partners in data ecosystems [20], they are losing control over their data assets [18], facing difficulties in defining data ownership and accountability [27], and in managing the partner's rights [17].

One of the most important steps at this stage is to conduct a data governance maturity assessment [38]. The goal is to evaluate/diagnose how the organization is performing in the several data governance dimensions (e.g., metadata management, data security, data strategy, data architecture) based on a set of predefined assessment criteria [9]. However, existing maturity models are focused on a single enterprise scenario, lacking the requirements to conduct an assessment of inter-organizational setups [20]. Moreover, the current models do not include data literacy, ethics, and data transmission dimensions. Lastly, they do not consider mechanisms required to handle datasets for developing Machine Learning (ML) models and AI systems. Therefore, we identified the need to create a maturity model that is suitable for both intra and inter-organizational scenarios in the context of ALB.

We started our research by searching for data governance maturity models. At this stage, we identified and described a set of data governance maturity dimensions. Lastly, we performed some initial steps on the maturity assessment of ALB, focusing on the antecedents of data governance in the organization.

The remainder of this paper is structured as follows: The basic concepts of data governance and data governance maturity models are described. Next, we present a set of data governance maturity dimensions for ALB. Then, the preliminary results of the first steps of the data governance maturity assessment are detailed. Lastly, the paper closes with the conclusions and the next steps of this research.

Data governance

Data governance is the exercise of authority and control over data management to maximize data value while minimizing costs and related risks [1]. It is a top priority for organizations worldwide to establish a cross-organizational data agenda, define mechanisms to capitalize on business opportunities, manage data-related risks, comply with regulations, deal with decentralized data scenarios, and improve data quality [1]. A recent data governance framework addresses the challenges of newer regulations (e.g., GDPR), aiming at data consistency, trustworthiness, and decision-making accountability. Nevertheless, the long-term vision of data value is challenging. In recent years, the establishment and deployment of data governance faced several challenges:

- The amount of data produced worldwide; it is expected to reach 158 zettabytes by 2025 [10];
- The increasing data volumes from different sources may cause inconsistencies that must be identified and corrected [1];
- The rising number of operations depending on data analytics and service reporting raises the need to create a shared vision of data in organizations [1];
- The recent and upcoming compliance and regulations (e.g., the European Union's regulation toward AI use) impact how companies should manage their data [1];
- The fragmented enterprise architectures and legacy systems require companies to deploy measures to handle decentralized data [24];
- The need for mechanisms to ensure the definition of the decision rights and accountability for data operations [3];
- The necessity to define mechanisms to align the business and enterprise goals of the organization with their data resources – data as a key organizational asset [37];
- The difficulties in minimizing the data-related costs and risks and increasing the data value for the organization [1];
- New technologies, such as AI and Machine Learning (ML), deeply rely on the input data (e.g., the datasets, the sources, and the quality of data) [19]. Additionally, AI algorithms can produce models or new data that require proper data governance mechanisms to be handled [19].

Figure 1 introduces an overall scheme for data governance concepts, adapted from the proposal of Abraham *et al.* [1] for a data governance framework supported in the three scopes: (1) data scope, (2) organizational scope, and (3) domain scope.

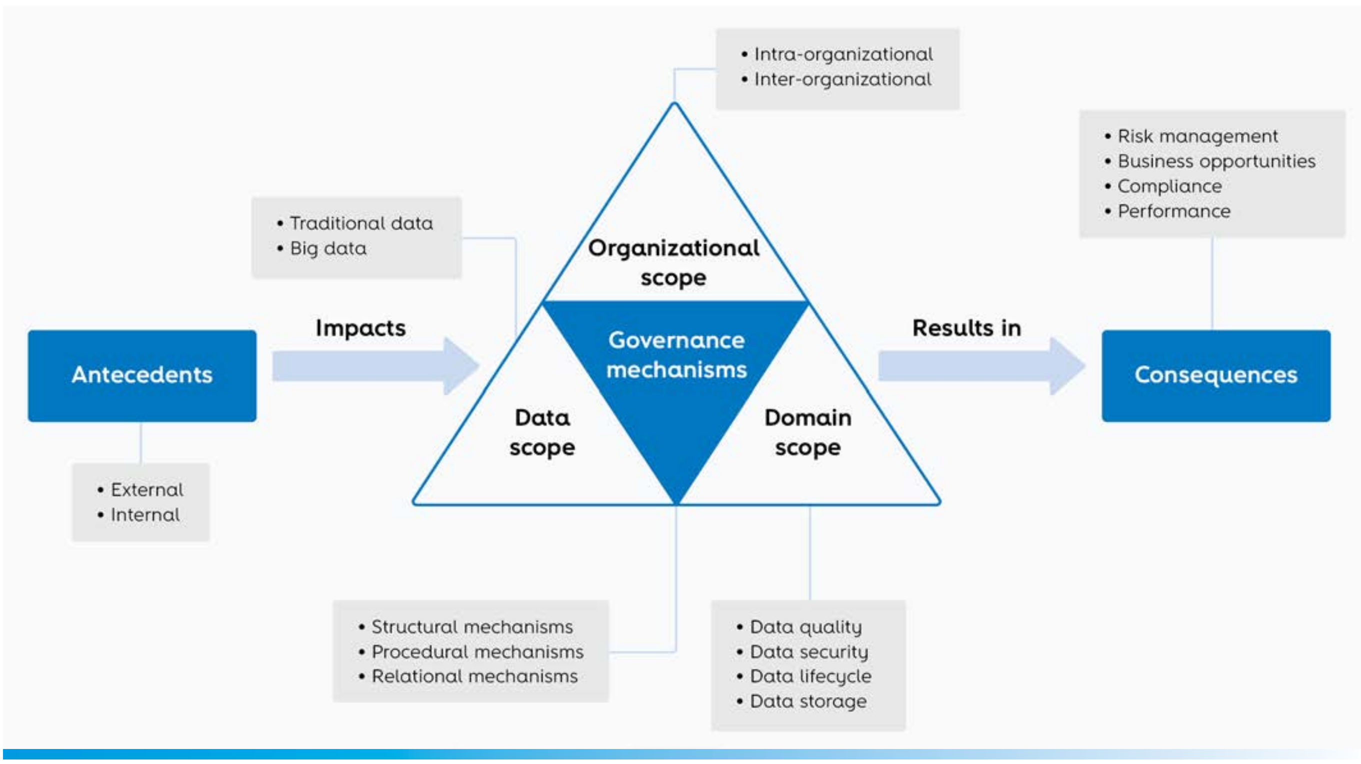


FIGURE 1 – Main concepts of data governance (based on [1])

First (on the left of **Figure 1**), the internal and external antecedents must be identified [35]. These antecedents are the contingency factors (e.g., the location of the organization, regulations) that impact the adoption and implementation of data governance. For the internal antecedents, it is necessary to understand the organizational culture, the Information Technology (IT) architecture, and the organizational strategy [35]. The external antecedents include the market dynamics [25], the type of industry [25], and regulations [35].

The scope of data governance is equally important [1]. Three essential scopes should be addressed with proper governance mechanisms: (1) data scope, (2) organizational scope, and (3) domain scope. Two main categories can be considered in the data scope: traditional data and big data [1]. The former typically includes master data (e.g., finances, clients, employees), transactional data (e.g., orders, bills), or reference data (e.g., product codes) [26]. Big data encompasses highly dynamic and substantial datasets (e.g., biometric data, social network data) [1]. As for the organizational scope, it concerns the scale of data governance [1], in which intra-organizational data governance scenarios (a single organization) are more focused on promoting internal alignment of the organization's business objectives and data and the management of the data quality and integrity across the organization's operations [25, 40]. Conversely, inter-organizational data governance scenarios should be more aimed at exploiting environmental opportunities and may raise difficulties such as the loss of control over data [28]. Finally, data governance includes the domain scope that addresses quality, security, lifecycle, metadata, and storage and infrastructure [1]. For example, organizations should focus on the definition of processes for quality management, the communication of the data quality strategy to all stakeholders, and the definition of roles and responsibilities (e.g., appointing a data architect or data manager)[26].

The data governance mechanisms can be categorized into three main groups: (1) structural mechanisms, (2) procedural mechanisms, and (3) relational mechanisms. They "*comprise formal structures connecting business, IT, and data management functions, formal processes and procedures for decision-making and monitoring, and practices supporting the active participation of and collaboration*" [1]. The structural governance mechanisms establish the governance entities, accountability of actions, and reporting entities, focusing on setting roles (e.g., data owners) and responsibilities as well as the allocation of decision-making (e.g., the hierarchy of the organization) [1, 4]. Procedural mechanisms ensure data accuracy, correctness, security, and efficiency [4]. These also include the definition of (1) the data strategy, (2) policies, (3) standards, (4) processes, (5) procedures, (6) contractual agreements, (7) performance measurement, (8) compliance monitoring, and (9) issue management [1]. Finally, the relational governance mechanisms define measures to promote collaboration among the stakeholders, communication (e.g., communication channels), training (e.g., promoting new knowledge), and the coordination of decisions (e.g., hierarchy) [1, 4].

Ultimately, organizations that implement data governance expect positive outcomes, including new business opportunities, compliance requirements fulfillment, internal data-related process improvement, and risk management [1].

Data governance for TELCO

Data governance is of utmost importance for TELCO companies, especially given the increasing amount and complexity of big data that they are handling [30]. Therefore, TELCO companies have used deep data analysis mechanisms to understand consumer behavior, interests, network load patterns, network issues, and usage patterns [30]. As an example, a company aiming to provide Internet Protocol Television (IPTV) on a broad scale, needs to have available both customer data (e.g., address data) and network infrastructure data (e.g., the bandwidth available in specific regions), completely and accurately [25]. The scenario is even more complex for TELCO platform developers, with a data prosumer profile: constantly producing data and using it to improve their digital offer.

Domains such as data security, data architecture management, and data privacy are also vital concerns for TELCO companies [25, 30]. TELCO is a highly regulated sector including legislation for data privacy, security, and accessibility. Data governance can assist TELCO companies in improving these aspects [25, 30], such as the interoperability of systems and hardware and complying with regulations [38].

Data governance maturity models

Maturity models are used to describe current scenarios of the organizations and depict improvement measures [5, 36]. These models have been used for decades in the information systems (IS) field [31], being first developed and used in the fields of quality management [12]. Recently, they have also been applied to IT companies' processes, being deployed for organizational assessment and strategic planning.

With organizations facing pressures to gain and maintain competitive advantages in fierce markets, identifying cost-reduction measures, improving quality, decreasing time to market, and so on, are becoming increasingly important [5]. Maturity models have been developed to assist organizations in this endeavor [5].

Maturity models can be positioned in between models and methods or roadmaps [21]. They encompass a set of paths (or levels) on which the organizations can diagnose their status and define the target state to which they want to evolve (e.g., 0-inexistent to 5-optimized). Moreover, the maturity model can support organizations in developing a roadmap to reach that desired state. Lastly, they can also be used by organizations to compare their status with specific benchmarks and best practices in industry and academia [12].

Data governance maturity models can be used to perform the continuous assessment of the status of a data governance framework through time [38]. This is only possible considering the definition of a set of metrics and key performance indicators (KPI) that facilitate the deployment of the framework, guaranteeing its sustainability and promoting organizational transformation. The development of data governance maturity models is essential to influence the strategic direction of the data governance framework [32].

Our research revealed that in the last 15 years, several data governance maturity models have been proposed in both, academic and grey (non-academic or technical) literature. **Figure 2** summarizes the findings of our comprehensive literature review on data governance maturity models and the chronological order of the development of such models.

Figure 2 introduces the chronological order of the data governance maturity models that were identified in our comprehensive literature review. We identified contributions coming from both academia and industry. The research in academia revealed that the current contributions are more focused on developing models for specific sectors and paradigms, such as the case of cloud computing [2, 6], industry 4.0 [41], data spaces [7], and micro-financial organizations [29]. The grey/technical literature has the highest number of works on data governance maturity models, with most of them being developed by companies (e.g., IBM [13], Gartner [23], and Oracle [33]). At this stage, we identified promising maturity models for the context of ALB. Firstly, there is the maturity model developed by the TM Forum [38], a platform that develops important contributions to the field of telecommunications. Secondly, there is the Data Management Maturity Model (DMM) [41], developed by the CMMI Institute [14].

The existing data governance maturity models and assessment solutions do not cover the full scope of data governance required for ALB. Therefore, new mechanisms are necessary to support the organization in increasing its data governance maturity,

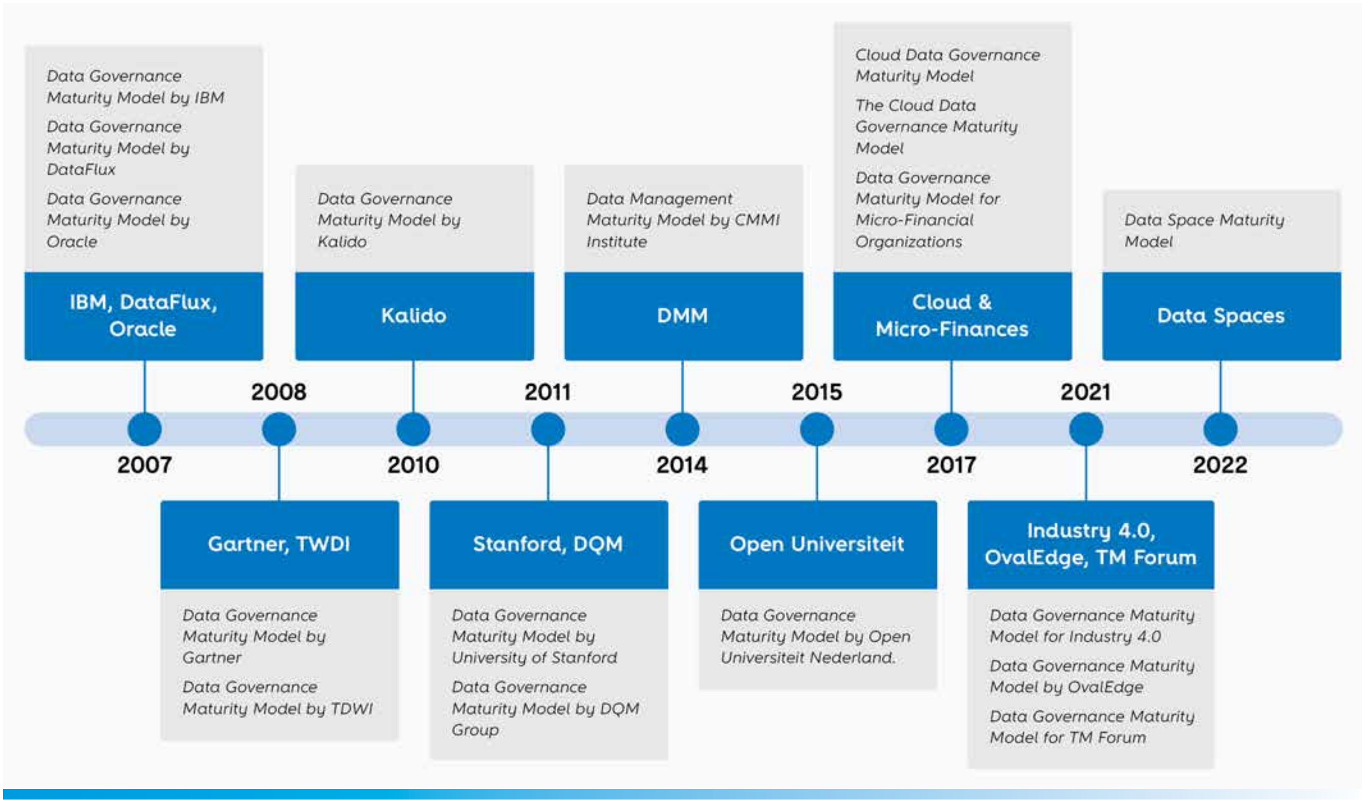


FIGURE 2 – Chronological order of the data governance maturity models development

which includes the assessment and implementation of data literacy, decentralized data operations, and decisions, data-based decision-making, ethics on the use of data, the specification of procedures during data science activities and data lifecycle, and mechanisms for the accountability of data and decisions based on data.

Data governance dimensions

The previous sections revealed important data governance maturity models developed by academia and industry. These maturity models include several dimensions with similarities and differences. After analyzing the maturity models, we identified a set of dimensions that must be considered when assessing the data governance maturity of an organization. The following list describes d the data governance maturity dimensions.

- **Data governance organization:** the formal definition of the level of mutual responsibility between the business and IT, and the recognition of the fiduciary responsibility to govern data at different levels of management [13]. Moreover, it includes the existence of data governance structures, that conduct the implementation and monitor the framework. It also covers the formal mechanisms that concern communication within the organization (e.g., decision-making authorities), staff training (e.g., ethics, regulations), leadership support, and organizational culture [22];

- **Data strategy:** the organization should define a data strategy that includes identifying its business goals, a vision statement toward the importance of the data, a set of guiding principles, a business case, and organizational objectives [22];

- **Data policies:** this dimension covers the existence of general guidelines for the use of data in the organization, which are influenced by the regulatory framework for establishing and managing policy guidelines, standards, and strategies under which data management is governed [29];

- **Data literacy:** data literacy refers to the ability of people (e.g., practitioners, researchers, and customers) to read, work with, analyze, and argue with data [8]. Organizations should develop a set of activities that contribute toward achieving these objectives;

- **Data lifecycle:** the several stages data goes through (e.g., collection, analysis, deletion) in the several departments of the organization [29]. The organization should define mechanisms to optimize data management throughout its lifecycle, making it efficient and contributing to the organization's needs [29];

- **Data quality:** the existence of defined mechanisms for planning, implementing, and maintaining techniques and processes that make data fit for use and ensure its high quality is essential for the organization according to its sector [9];

- **Metadata:** mechanisms for describing data and IT resources by linking important business and technical information. Moreover, this dimension includes a description of the data that facilitates a consistent understanding of its properties and usage;

- **Data storage:** the data storage dimension focuses on the storage, retention, deletion, and archival of all the data assets that are used by the organization [39]. Moreover, it covers the design, implementation, and operational support of data [39];

- **Data architecture:** the development and maintenance of an overall data architecture of the organization, which can include mechanisms for data integration, access, and control [9]. It covers the architectural design of structured and unstructured data systems and applications that enable data availability and distribution to the allowed users [9];

- **Data security:** the organization's practices and controls to mitigate risk and protect data assets [9]. The existence of formal mechanisms and structures that contribute to data accessibility, control, integrity, confidentiality, and availability [29]. Finally, it addresses the mechanisms for risk identification and data security planning, data security roles and responsibilities, access to network management, and internal audits [29];

- **Data stewardship:** this dimension targets the formal mechanisms that guarantee accountability for the description, utilization, and standard of quality of specific data assets within a defined organizational context [9];

- **Data risk management & compliance:** this dimension covers the formal mechanisms that are deployed in the organization toward risk management. It usually includes the methodologies by which risks are identified, qualified, quantified, avoided, accepted, and mitigated. For compliance, the organization should have defined mechanisms and structures targeted to track and enforce legal requirements and compliance with company policies, standards, and procedures [13];

- **Data privacy:** data privacy calls for mechanisms that contribute toward mitigating data leak situations, controlling data access and use, and retaining data [1]. These mechanisms are necessary to comply with a set of regulations (e.g., GDPR) [38].

Preliminary maturity assessment operations

The research of Abraham *et al.* [1] proposes a data governance framework aiming to deal with the challenges of recent data-related regulations (e.g., the GDPR), ensuring data consistency, promoting trustworthiness, and defining decision-making accountability. In the initial stage, the authors suggest that organizations should consider the internal and external antecedents that may impact the implementation, level of adoption, and planning of data governance mechanisms [35].

The internal antecedents include the definition of the strategy, the organizational structure, the culture of the organization, and the IT infrastructure [1]. Moreover, it includes the organization's culture, focusing on senior management support, active leadership participation, the promotion of a clear business vision of the value of data for the organization, and the role of data in the organization [35].

Our first step was to analyze the antecedents of data governance for ALB to obtain some initial insights on the organization and feedback for the selection of the data governance maturity model to be used in the assessment. We started with the identification and analysis of the internal antecedents of ALB. **Table 1** describes our findings.

Table 1 includes our findings for the internal antecedents of data governance in Altice Labs. Altice Labs considers data a key enterprise asset for the development of new technological solutions and obtaining competitive advantages in the market. The C-level officer and management officers support and understand the importance of the data governance framework for the organization.

Next, we moved to the identification and analysis of the external antecedents of data governance of Altice Labs. The external antecedents cover the factors originating outside the organization that may affect governance procedures. It includes market dynamics, geographical location, and type of industry [25]. Furthermore, organizations need to address regulations (e.g., GDPR [11]) that impose limits and restrictions on the use of data (e.g., collection, storage, retention, quality, privacy, and business use) [35]. These regulations can vary according to the geographical location (e.g., country-specific laws) [34], the type of industry (e.g., quality requirements in the production industry), and the market in which the organization operates [1]. **Table 2** describes our findings.

Internal antecedent	Findings
IT strategy & infrastructure	ALB has defined a set of standardized procedures for IT structures and processes across the departments, using tools for data analysis, storage, and manipulation. Some legacy systems are still used in data-related operations. Soon, the group leader will promote a transformation of the existing IT infrastructures across several members.
Organization culture	ALB considers data a key strategic asset that can be used to develop new solutions and improve existing ones. Moreover, data can be used for the improvement of internal operations and decisions. Senior management supports the implementation of data governance and is currently promoting the improvement of the existing mechanisms. Lastly, the organization has a business vision for data governance.
Organization strategy	ALB integrates an international group that includes TELCO operators and technology solutions providers. The group is growth-oriented, and the organization is profit-oriented, focusing on the development of new technological solutions for the TELCO field. The organization can be considered a data prosumer (data producer + data consumer), meaning that the organization produces data and uses it for internal consumption. Moreover, it also consumes data from other companies in the group. Essentially, the data prosumer plays the role of both data consumer and producer, focusing on data acquisition and extraction, storage and archiving, processing and analysis, visualization, publication, sharing, and use of data.

TABLE 1 – Data governance internal antecedents for Altice Labs

External antecedent	Findings
Country	ALB operates in Portugal, which is integrated into the European Union. The TELCO group in which Altice Labs is integrated operates in other countries (e.g., United States of America, Israel, France). Therefore, the organization must identify and analyze all the regulations on which their solutions will be operated (e.g., GDPR in European Union).
Industry	ALB operates in the TELCO sector, which requires the deployment of quality standards for data (e.g., data must be precise to understand existing anomalies). These organizations must deal with increasing amounts and dynamics of big data datasets, mainly used for deep data analysis mechanisms, to understand consumer behavior, interests, network load patterns, network issues, and usage patterns. Domains such as data security, data architecture management, and data privacy are also vital concerns for TELCO companies.
Legal and regulatory requirements	TELCO is a highly regulated sector that includes legislation for data privacy, security, and accessibility. Regulations such as the GDPR have had an enormous impact on organizations, considering that it requires organizations to describe to their users the way they handle data, including how it is stored, retrieved, analyzed, and deleted [11].
Market context	The TELCO market is highly dynamic and demanding, with many other companies developing solutions for the market. Despite this, Altice Labs is one of the oldest and most respected technology providers in Portugal, also developing solutions for other organizations besides the ones in the group.

TABLE 2 – Data governance external antecedents for Altice Labs

Target maturity score for Altice Labs

The antecedents' analysis is an important step toward identifying the target maturity levels for the several data governance maturity dimensions. Based on the obtained results, we moved toward the creation of a target scorecard for data governance maturity in ALB.

The development of a target scorecard can contribute to identifying the priorities to improve the data governance framework. Based on the results of the maturity assessment, it is possible to establish a degree of comparison between the current score and the desired status. This comparison will allow us to fit the improvement measures based on the specific needs of ALB and propose a final roadmap for data governance implementation. **Figure 3** introduces the target scorecard for the data governance maturity of ALB.

Figure 3 introduced the data governance target maturity scorecard for ALB. We identified as main priorities the development of mechanisms for data transmission, data lifecycle, data governance, data privacy, data policies, data strategy, data architecture, data quality, and data risk management.

Data governance maturity assessment planning for Altice Labs

Figure 4 presents a high-level process for the data governance implementation at ALB.

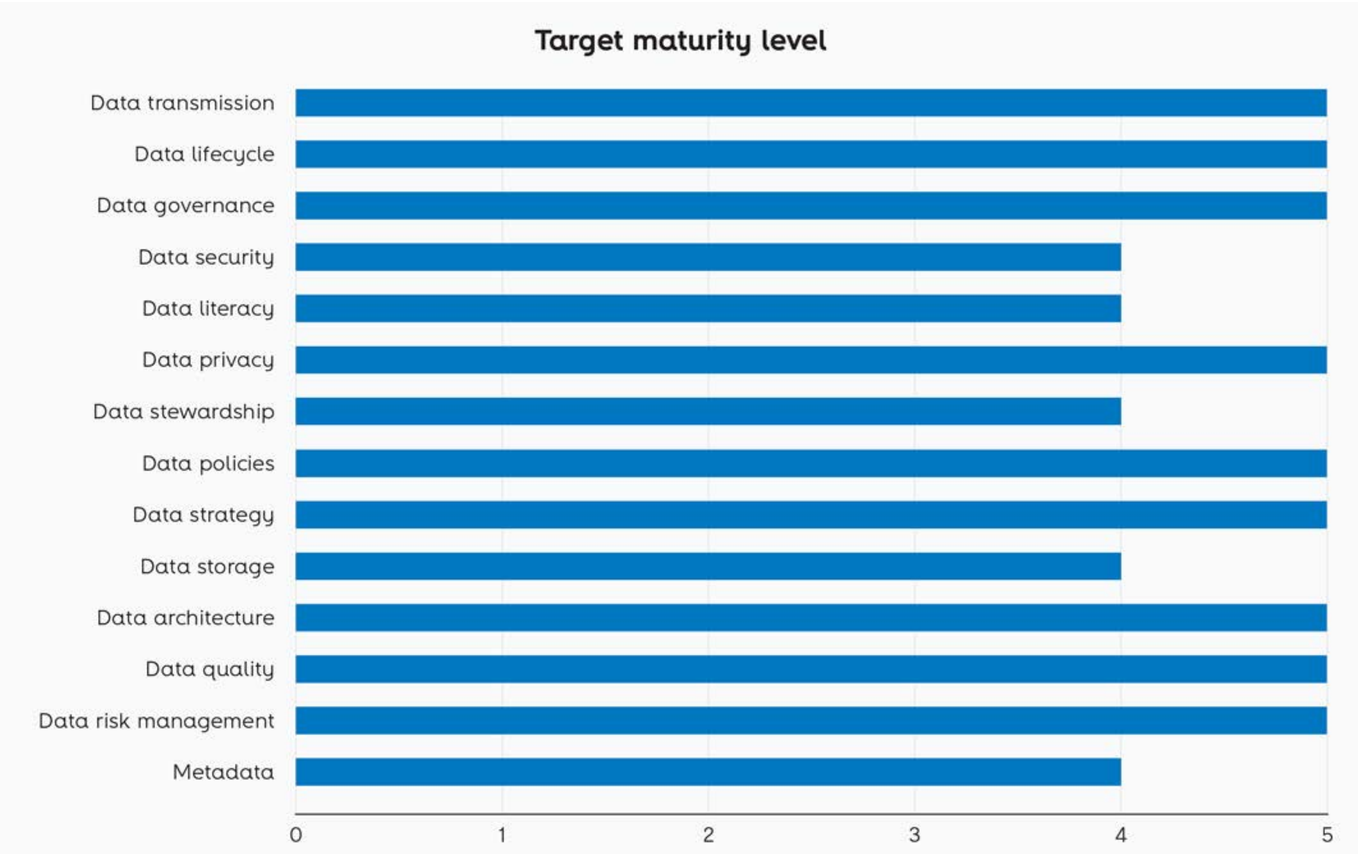


FIGURE 3 – Data governance target maturity scorecard for Altice Labs

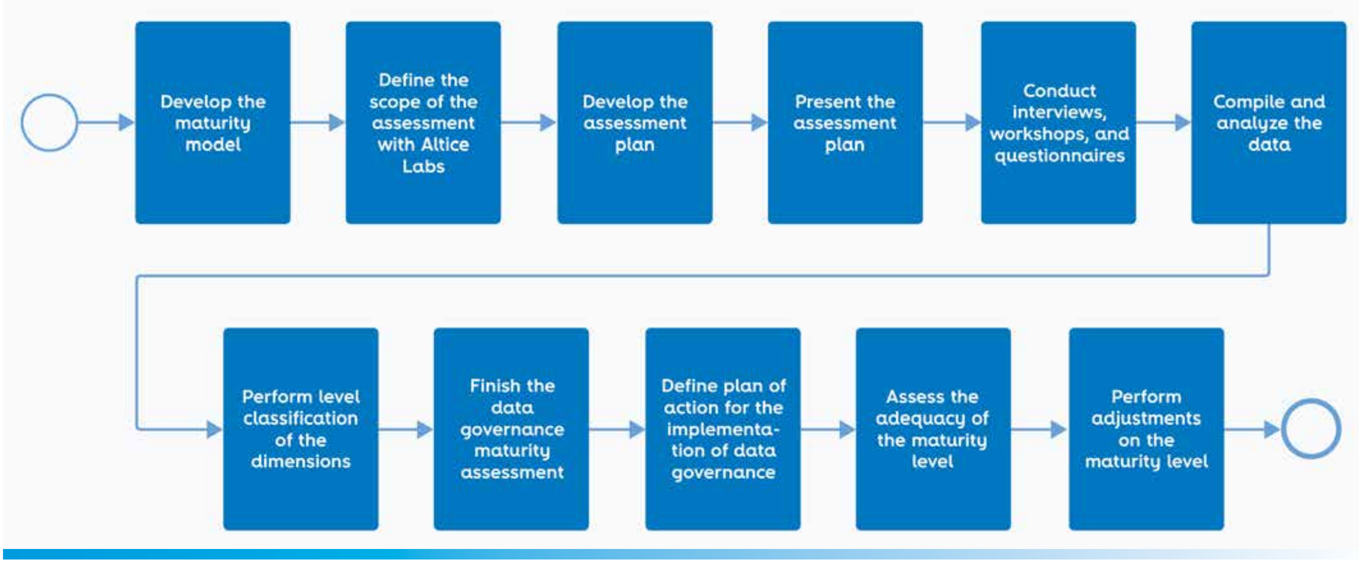


FIGURE 4 – Roadmap for data governance implementation at Altice Labs

Figure 4 summarizes the suggested data governance implementation process for ALB. The process started with the design and development of the maturity model that will be used in the assessment, inspired on the proposals of TM Forum Data Maturity Model [38, 39], the DMM/CMMI [41], and the Stanford Data Governance Maturity Model [32]. Moreover, we aim to include data literacy, the decentralized data governance mechanisms, and the AI governance requirements for input data and the outcome of the models. For this task, we have developed an initial diagnostic that supports the organizations in prioritizing the decentralized data governance maturity models based on their context. Next, we have defined the scope of the data governance maturity assessment. ALB identified one specific department/team to participate in the assessment, following the plan with the company.

Afterward, we created the plan for the data governance plan maturity assessment. Dates will be defined to conduct the necessary presentations, interviews, and workshops with the team that will be the target of the maturity assessment. An initial presentation will focus on describing and explaining the goal of the maturity assessment, the basic concepts of data governance maturity assessment and maturity models, and the role that each party will have in the process. At this stage, we will also investigate the existence of any kind of assessment mechanisms for specific dimensions of data governance (e.g., roadmaps, data management assessment). Finally, this plan will be documented and communicated to the stakeholders involved in the maturity assessment process.

Next, the data governance maturity assessment will be conducted. All the planned presentations, interviews, questionnaires, and workshops will be carried out during this time, to retrieve all the necessary information to fulfill the requirements of the data governance maturity assessment. All the retrieved information will be compiled and analyzed to classify each of the data governance dimensions that are included in the maturity model. The first conclusions will be obtained after this stage, based on which we will be able to propose a roadmap for the implementation and improvement of the data governance framework, according to its priorities.

The results will also be important to assess the adequacy of the model in the context of ALB. During this process, we will obtain feedback and insights on the performance of the model. This information will be used to conduct the necessary adjustments and corrections (e.g., include more details on the level of a specific dimension) on the data governance maturity model that is proposed for the assessment of a team of ALB.

Conclusions and next steps

This paper reports the early stages of the maturity assessment to be conducted in ALB. This research included a background analysis of data governance foundations, data governance in the TELCO context, and data governance maturity models. Furthermore, it includes a comprehensive literature review on data governance maturity. Next, it introduces the initial steps toward the data governance maturity assessment at ALB by depicting an external and internal antecedents' analysis. Lastly, target maturity scores for the several data governance maturity dimensions are proposed for the context of ALB, based on the results of the initial diagnosis of the antecedents.

The next stages will be focused on the data governance maturity assessment of ALB. We will start by executing the plan for the data governance maturity assessment. At this stage, we will retrieve all the data that is necessary to conduct the maturity assessment model. Then, the retrieved data will be analyzed and applied to classify the organization using the model. Based on the classification, we will propose improvement measures for the data governance framework at ALB.

The results will be also important to assess the adequacy of the model in the context of ALB. During this process, we will obtain feedback and insights on the performance of the model. This information will be used to conduct the necessary adjustments and corrections (e.g., include more details on the level of a specific dimension, include a new data maturity dimension) on the data governance maturity model that is proposed for the assessment of the department of ALB.

Acknowledgments

This work is funded by the project POWER (grant number POCI-01-0247-FEDER-070365), co-financed by the European Regional Development Fund (FEDER), through Portugal 2020 (PT2020), and by the Competitiveness and Internationalization Operational Programme (COMPETE 2020). It is also co-funded by the FCT - Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit - UIDB/00326/2020 or project code UIDP/00326/2020.

References

[1] Abraham, R., Schneider, J., vom Brocke, J.: Data governance: A conceptual framework, structured review, and research agenda. *Int. J. Inf. Manage.* 49 (January), 424–438 (2019).

[2] Al-Ruithe, M., Benkhelifa, E.: Cloud data governance maturity model. *ACM Int. Conf. Proceeding Ser.* (2017).

[3] Benfeldt, O., Persson, J.S., Madsen, S.: Data Governance as a Collective Action Problem. *Inf. Syst. Front.* 22 (2), 299–313 (2020).

[4] Borgman, H., Heier, H., Bahli, B., Boekamp, T.: Dotting the i and crossing (out) the T in IT governance: New challenges for information governance. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. pp. 4901–4909. (2016).

[5] de Bruin, T., Rosemann, M., Freeze, R., Kulkarni, U.: Understanding the main phases of developing a maturity assessment model. In: *ACIS 2005 Proceedings*. pp. 8–19. (2005).

[6] Cheng, G., Li, Y., Gao, Z., Liu, X.: Cloud data governance maturity model. In: Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, *ICSESS*. pp. 517–520. (2018).

[7] Curry, E., Tuikka, T.: An Organizational Maturity Model forData Spaces: A Data Sharing Wheel Approach. In: *Data Spaces*. pp. 21–42. Springer International Publishing Cham (2022).

[8] D'Ignazio, C., Bhargava, R.: Approaches to Building Big Data Literacy. In: *Bloomberg Data for Good Exchange Conference*. p. 6. (2015).

[9] Data Management Association: DAMA-DMBOK : data management body of knowledge. *Technics Publications*, LLC (2017).

[10] Djuraskovic, O.: Big Data Statistics (2022) - Amount of Data Generated in The World. Accessed: Mar. 27, 2023. [Online]. Available https://firstsiteguide.com/big-data-stats/.

[11] European Parliament: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Accessed: Mar. 27, 2023. [Online]. Available https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

[12] Hamel, F., Herz, T.P., Uebernickel, F., Brenner, W.: IT evaluation in business groups: A maturity model. In: *Proceedings of the ACM Symposium on Applied Computing*. pp. 1410–1417. (2013).

[13] IBM: The IBM Data Governance Council Maturity Model : Building a roadmap for effective data governance. *Gov. An Int. J. Policy Adm*. (October), 1–16 (2007).

[14] ISACA: CMMI Institute. Accessed: Jul. 29, 2022, (2023). [Online]. Available: https://cmmiinstitute.com/.

[15] Jagals, M., Karger, E.: Inter-organisational data governance: A literature review. *Twenty-Ninth Eur. Conf. Inf. Syst*. (June), 1–19 (2021).

[16] Janssen, M., Brous, P., Estevez, E., Barbosa, L.S., Janowski, T.: Data governance: Organizing data for trustworthy Artificial Intelligence. *Gov. Inf. Q*. 37 (3), 101493 (2020).

[17] Lee, S.U., Zhu, L., Jeffery, R.: Data governance decisions for platform ecosystems. *Proc. Annu. Hawaii Int. Conf. Syst. Sci*. 2019-Janua 6377–6386 (2019).

[18] Lis, D., Otto, B.: Data governance in data ecosystems - Insights from organizations. *26th Am. Conf. Inf. Syst*. AMCIS 2020. 0–10 (2020).

[19] Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M.: Defining organizational AI governance. *AI Ethics*. (2022).

[20] Marcelo, M.I., Barros Lima, G. de F., Farias Lóscio, B.: Investigations into Data Ecosystems: a systematic mapping study. *Knowl. Inf. Syst*. 61 (2), 589–630 (2019).

[21] Menon, K., Kärkkäinen, H., Lasrado, L.A.: Towards a maturity modeling approach for the implementation of industrial internet. In: *Pacific Asia Conference on Information Systems, PACIS 2016 - Proceedings*. p. 38. (2016).

[22] Merkus, J., Helms, R., Kusters, R.: Data Governance Capabilities: Maturity Model Design with Generic Capabilities Reference Model. (November), 102–109 (2021).

[23] Newman, D., Logan, D.: Gartner Introduces the EIM Maturity Model. *Gart. Res.* ID Number G00160425. (December), 1–8 (2008).

[24] Nielsen, O.B., Persson, J.S., Madsen, S.: Why governing data is difficult: Findings from Danish local government. In: *IFIP Advances in Information and Communication Technology*. pp. 15–29. (2019).

[25] Otto, B.: Organizing Data Governance: Findings from the telecommunications industry and consequences for large service providers. *Commun. Assoc. Inf. Syst.* 29 (1), 45–66 (2011).

[26] Otto, B., Wende, K., Schmidt, A., Osl, P.: Towards a framework for corporate data quality management. ACIS 2007 Proc. - 18th Australas. Conf. Inf. Syst. 916–926 (2007).

[27] De Prieelle, F., De Reuver, M., Rezaei, J.: The Role of Ecosystem Data Governance in Adoption of Data Platforms by Internet-of-Things Data Providers: Case of Dutch Horticulture *Industry. IEEE Trans. Eng. Manag.* 69 (4), 940–950 (2022).

[28] Rasouli, M.R., Trienekens, J.J.M., Kusters, R.J., Grefen, P.W.P.J.: Information governance requirements in dynamic business networking. *Ind. Manag. Data Syst.* 116 (7), 1356–1379 (2016).

[29] Rivera, S., Loarte, N., Raymundo, C., Dominguez, F.: Data governance maturity model for micro financial organizations in Peru. *ICEIS 2017 Proc.* 3 203–214 (2017).

[30] Sajjad, R., Dhary, A.-R.: Using big data in telecommunication companies: A case study. *African J. Bus. Manag.* 14 (7), 209–216 (2020).

[31] Selander, L., Henfridsson, O., Svahn, F.: Capability search and redeem across digital ecosystems. J. *Inf. Technol.* 28 (3), 183–197 (2013).

[32] Stanford University: Stanford Data Governance Maturity Model. Accessed: Jun. 3, 2022. [Online]. Available: https://www.lightsondata.com/data-governance-maturity-models-stanford/.

[33] Sun, H., Corporation, O.: Enterprise Information Management: Best Practices in Data Governance. An Oracle White Pap. *Entrep. Archit.* 20 (2011).

[34] Tallon, P.P.: Corporate governance of big data: Perspectives on value, risk, and cost. *Computer* (Long. Beach. Calif). 46 (6), 32–38 (2013).

[35] Tallon, P.P., Ramirez, R. V., Short, J.E.: The information artifact in IT governance: Toward a theory of information governance. *J. Manag. Inf. Syst.* 30 (3), 141–178 (2013).

[36] Team, C.P.: CMMI for Development, version 1.2. (2006).

[37] Thompson, N., Ravindran, R., Nicosia, S.: Government data does not mean data governance: Lessons learned from a public sector application audit. *Gov. Inf. Q.* 32 (3), 316–322 (2015).

[38] TM Forum: TM Forum Data Governance Guidebook. Accessed: Jul. 11, 2022. [Online], Available: https://www.tmforum.org/resources/standard/gb1023-data-governance-guide-book-v3-0-0/.

[39] TM Forum: TM Forum Data Governance Maturity Model. Accessed: Dec. 5, 2022. [Online], Available: https://www.tmforum.org/resources/standard/gb1025-data-governance-maturity-model-v3-0-0/.

[40] Weber, K., Otto, B., Österle, H.: Ones size does not fit all -A contingency approach to data governance. *J. Data Inf. Qual.* 1 (1), 1–27 (2009).

[41] Zorrilla, M., Yebenes, J.: A reference framework for the implementation of data governance systems for industry 4.0. *Comput. Stand. Interfaces.* 81 103595 (2022).

[42] Zwolenski, M., Weatherill, L.: The Digital Universe Rich Data and the Increasing Value of the Internet of Things. *Aust. J. Telecommun. Digit. Econ.* 2 (3), 13–19 (2014).

DOI: 10.57650/INNOVACTION/2023.11

11 Speech and language analytics

In a context of growing business competitiveness, where the differentiation and quality of the service provided have asserted themselves in the face of other factors, companies need help understanding what their customers think or say.
AI model evolution related to human understanding, such as speech-to-text, entity extraction, intent classification, or summarization, is starting to be used to understand the customer's voice. These activities would otherwise be performed manually and solely when needed due to the high cost.
Customer support centers, marketing operations, or emergency services (from voice or email to social networks) are now valuable sources of information that allow automatic performance evaluation, subject categorization, relevant information extraction, and even conversation summarization. When evaluated together, they offer an overview, providing metrics, alarms, and predictions about consumer patterns and trends or the ability to detect anomalies early, which can help companies react more quickly.
The BOTSchool Speech and Language Analytics feature empowers organizations that handle extensive volumes of communications, including recorded calls and emails, with a tool that takes data analysis to another level, offering new business insights and providing a comprehensive solution to transforming neglected communication data into a valuable knowledge base for decision-making, either business or operational.

Authors

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

José Miguel Freitas, Altice Labs, jose-m-freitas@alticelabs.com

André Miguel Ferro, Altice Labs, andre-m-ferro@alticelabs.com

Keywords

Speech and Language Analytics; BOTSchool; Generative AI; Large Language Models

Introduction

As customers, we constantly want to contact organizations to receive quick support or answers. Typically, companies delegate this relationship with customers to Contact Center systems and teams, trying to pursue positive interaction experiences and create differentiation in the service provided compared to the competition. Customers demand contextualized information tailored for them, but above all, they demand value-added service [1]. For companies, "having" the customer's voice is considered critical. Knowing the customer's problems, needs, and sentiments is becoming a key factor and one of the key challenges to achieving a Customer Experience of excellence. These insights, typically tough and complex to obtain, opened a new range of facilities for organizations with the rise of AI. The emergence of Large Language Models (LLMs) and Generative AI created an opportunity to put companies in touch with the voice of their customers and effectively be able to understand and act in real time [4].

Speech and Language Analytics powered by Generative AI technology aims to revolutionize how organizations gain deeper insights from customer interactions and is increasingly being regarded as an indispensable tool for businesses.

A data-driven approach to managing customer interactions automates the process of identifying key phrases, topics, and trends within conversations, making it more efficient than manual analysis and enabling many other use cases that otherwise would be impractical. Next, we highlight the most significant application domains of this technology [2] [3]:

1. **Enhanced customer experience:** evaluation of customer sentiments, preferences, and pain points, so organizations can tailor their services and provide on-the-spot guidance, leading to higher customer satisfaction and loyalty;

2. **Voice of the customer:** the emotional tone and sentiment of speakers can provide a nuanced

understanding of customers' emotions during interactions, enabling personalized responses and interventions, as well as insight to identify unsatisfied customers or mentions of competitor brands, helping companies to proactively align their offerings with customer expectations;

3. **Operational efficiency:** automation of routine processes, identification of performance gaps, and predictive insights help to optimize workforce management and free valuable human resources for more critical tasks, leading to lower costs and improved productivity. Real-time assistance to agents during customer calls is also a key asset;
4. **Customer revenue growth:** understanding customer buying signals and preferences to develop more effective sales strategies and revenue generation;
5. **Sales optimization:** gather tactics and behaviors from previous calls with booming sales and use them for training agents and increasing sales call effectiveness;
6. **Quality assurance and training:** monitor and score agent-customer interactions, enabling targeted coaching and training programs for service consistency and continuous improvement;
7. **Compliance and risk mitigation:** aids in compliance monitoring, reducing legal risks by ensuring agents adhere to regulatory guidelines and policies;
8. **Social listening and content moderation:** in real-time, language analytics can monitor social media trends to respond promptly or for historical data analysis. Sentiment analysis and text classification are useful for providing content recommendations and enabling personalization by understanding user preferences and behavior;
9. **Data-driven decision-making:** businesses can leverage the insights from speech analytics to make informed decisions about product development, marketing strategies, and service enhancements.

Speech and Language processing combined with AI, also known as Interaction Analytics, has emerged as a transformative technology in the realm of customer service that leverages artificial intelligence to understand, process, and analyze human speech. It can be used to assess call or meeting recordings and transcripts from digital channels such as chat and text messages. The fact that speech analytics software can analyze 100% of contacts 24/7 means companies can be more proactive and have a more accurate view of what is happening during customer interactions [4].

By converting spoken words into structured data points, it examines customer interactions to extract meaningful insights and customer sentiment. By leveraging natural language processing, machine learning, and artificial intelligence (AI), it uncovers customer preferences, behaviors, and emotions from customer conversations.

Altice Labs approach

BOTSchool was designed for companies and non-technical individuals to seamlessly implement virtual assistants through a low-code UI interface with multi-channel capabilities. The platform's primary goal remains to streamline the integration of conversational tools and be the communication hub between companies and customers.

In this new era of AI that leverages LLM (Large Language Models) and Generative models, new possibilities arise. Companies can seamlessly integrate their internal source documentation into an actual Knowledge Database, empowering ChatGPT with the capacity to gather insights from this information. This revolutionary understanding capability can then be used to serve as the voice of customers and enhance relationships with companies. The BOTSchool new generation introduces a new area of Interaction Analytics, specifically by exploring Speech and Language Analytics. This new feature aims to respond to the challenge of knowing how to answer questions such as what clients are asking or saying to me, how they feel about my service, or what they think about the quality that I am providing.

With BOTSchool Speech Analytics, companies can take their customer relationships to another level. It provides an easy and simple UI interface where audio or texts can be uploaded for analysis and an API interface for full integration and automation with existing information systems.

The Speech Analytics component of BOTSchool works based on a pipeline architecture that takes audio files or text and consumes them to extract and obtain information present in the audio stream (**Figure 1**).

The process starts with the upload of a single or multiple files to the platform. The files are then inserted in a cloud storage object, more specifically in an Amazon Web Services (AWS) Simple Storage Service (S3) bucket, a cloud storage, under a specific path in that same bucket. After the file is inserted, the workflow Machine Learning engine orchestrates all the tasks and jobs currently supported by the pipeline. The workflow engine offers the possibility to define DAGs (Directed-Acyclic Graphs) that support parallel execution of tasks, which allows for better time and processing power performance. Moreover, configuring a new job in the workflow is simple, which results in an effortless process of adding new features and tasks to the Speech Analytics product. An additional capacity is also used to automate the complete workflow process – event handling. This event-driven approach allows for an easy way to trigger the designed workflows, making the pipeline a completely automated process. An event is triggered when a new file is inserted in the S3 bucket, which results in the start of the first phase of the pipeline.

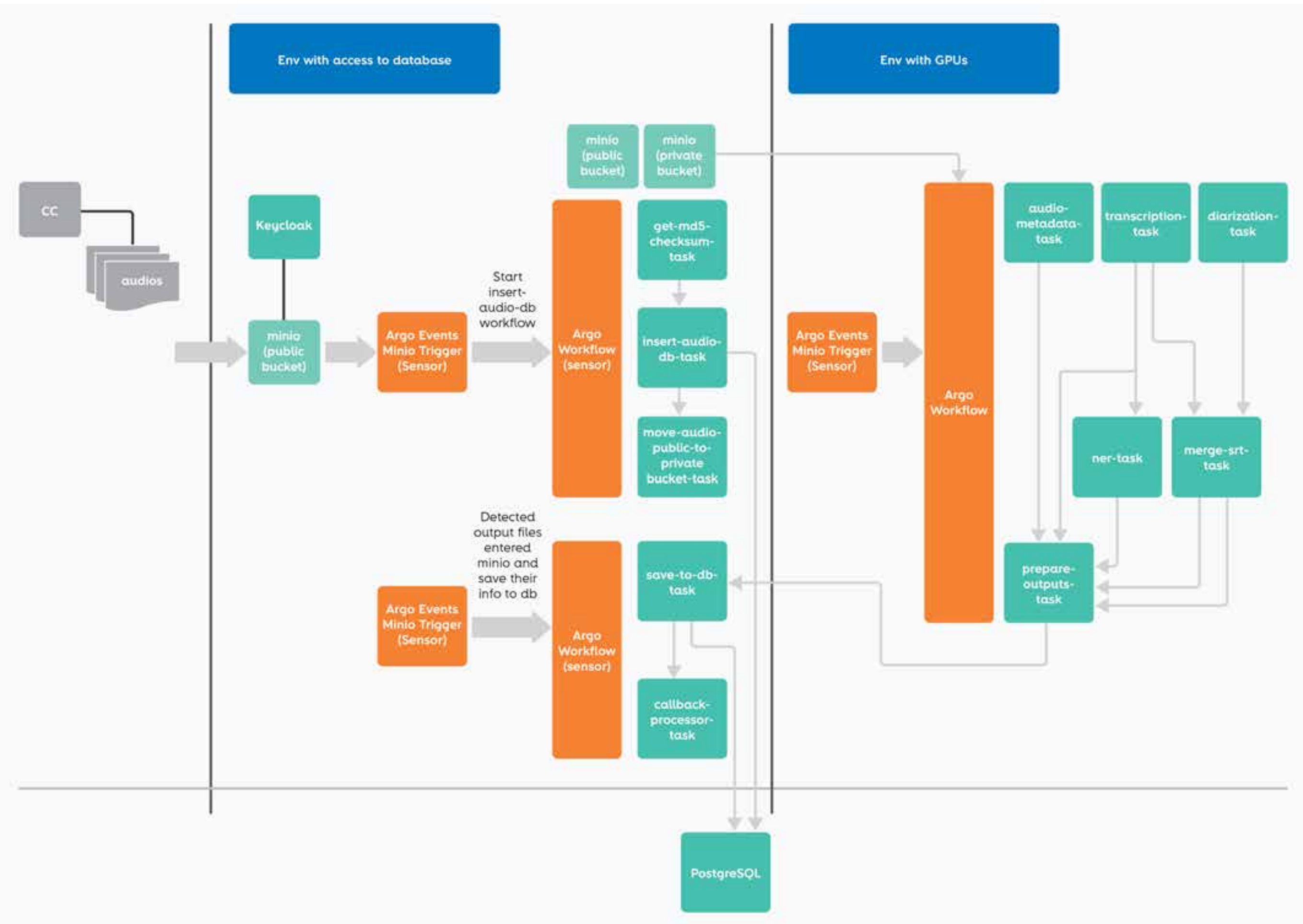


FIGURE 1 – BOTSchool Speech and Language Analytics architecture

Following the insertion in the bucket, the pipeline's second phase is started – workflow execution (**Figure 2**). The workflow contains all tasks related to the extraction of the information present in the submitted file. For example, in the case of an audio file, the initial workflow step obtains the metadata, which is then followed by the transcription task, which consists of transcribing the audio into text through the Whisper model from OpenAI. This step is the backbone of the complete process since various tasks use the transcription to extract information, such as named entity recognition, summarization, and extraction of text embeddings, which are all supported in steps following the conclusion of the transcription task. The workflow also contains a diarization step to partition the audio file into various segments, where each segment is associated with a specific speaker. This task allows the extraction of information from speakers' speaking duration and silence duration throughout the entire audio. After the transcription and diarization tasks are completed, another step is started, which consists of grouping each speaker's speaking time with the corresponding text obtained in the transcription. As a final step in the workflow, all the tasks' outputs are stored, once more, in the S3 storage.

The pipeline's final phase consists of taking the information that was stored in the S3, and storing it in the internal database, allowing advanced analytics to be performed.

The BOTSchool solution allows complete control of the tasks applied. It offers a data visualization layer on top of the information retrieval, enabling rich data exploration and the extraction of specific insights the company could require (**Figure 3**).

FIGURE 3 – BOTSchool Speech and Language Analytics dashboard

The solution also provides disruptive functionality, allowing companies to take information analysis and flexibility to another level. Taking advantage of the Generative capabilities supported by BOTSchool, companies can build prompts themselves in natural language based on the information they want answered, for example: "How satisfied is the customer?", "Was the customer satisfied with the service?", "Did my agent ask if there was

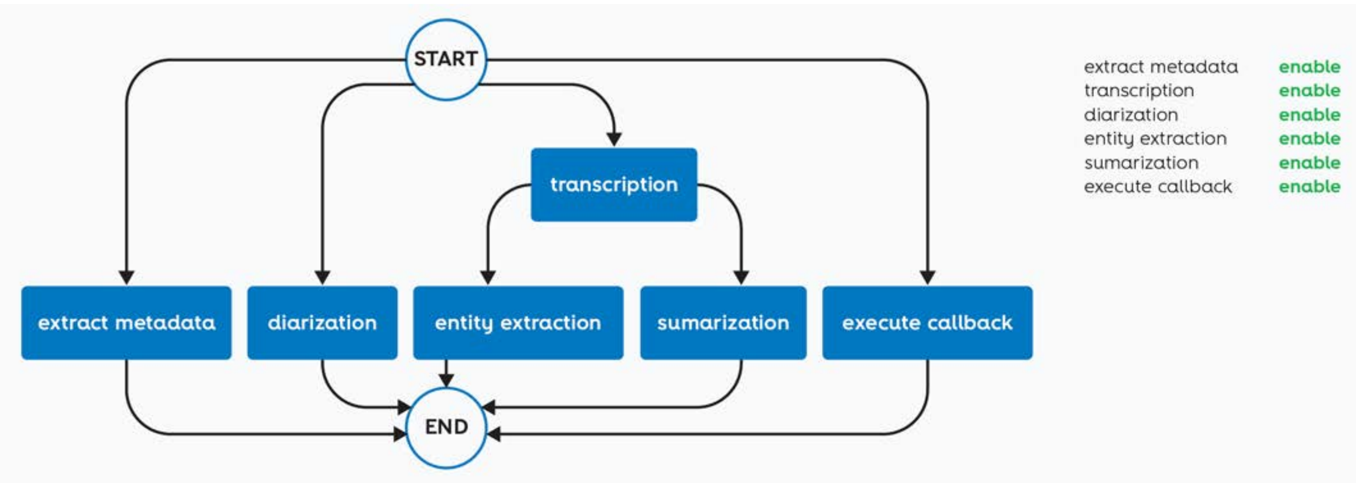


FIGURE 2 – BOTSchool Speech and Language Analytics data extraction workflow

another matter that wanted to be answered?", or "Was any campaign proposed?".

BOTSchool Speech and Language Analytics will boost companies' customer information and productivity with Augmented Analytics powered by GPT models.

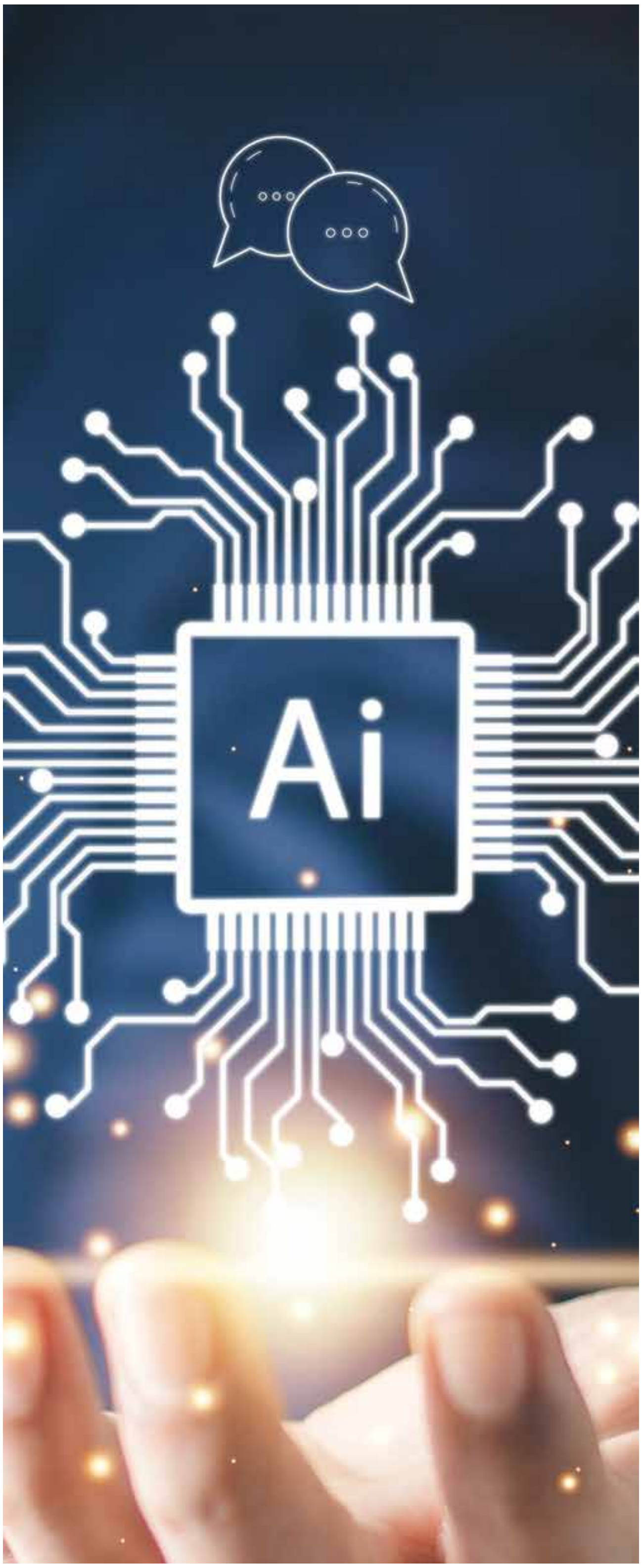
Conclusions

Speech and Language Analytics has emerged as a transformative technology in the realm of call centers and contact management automation, revolutionizing how organizations interact with their customers. In an increasingly data-driven world, adopting this technology is no longer an option but a real need.

By harnessing the latest trends in Interaction Analytics and leveraging its advantages, organizations can significantly enhance customer experience, streamline operations, and stay ahead in a competitive marketplace. With such insights, businesses can tailor their services and identify needs for self-service options, thus improving agent assistance, customer satisfaction, and loyalty.

By combining Analytics and Data Visualization with the latest Generative AI capabilities, the BOTSchool solution envisions a future where customer interactions are not just managed but also deeply understood, making the retrieval of such insights very intuitive and powerful.

The Speech and Language Analytics solution provides companies with the ultimate tool that allows them to hear the voice of their customers, helping them better understand how they perceive the services offered, or even identify new trends (problems or new subjects like new products/offers).



References

[1] Gartner, "How Customer Service Leaders Should Leverage Speech Analytics," 2020. Accessed: Nov. 2, 2023. [Online]. Available: https://www.gartner.com/en/customer-service-support/trends/customer-service-speech-analytics.

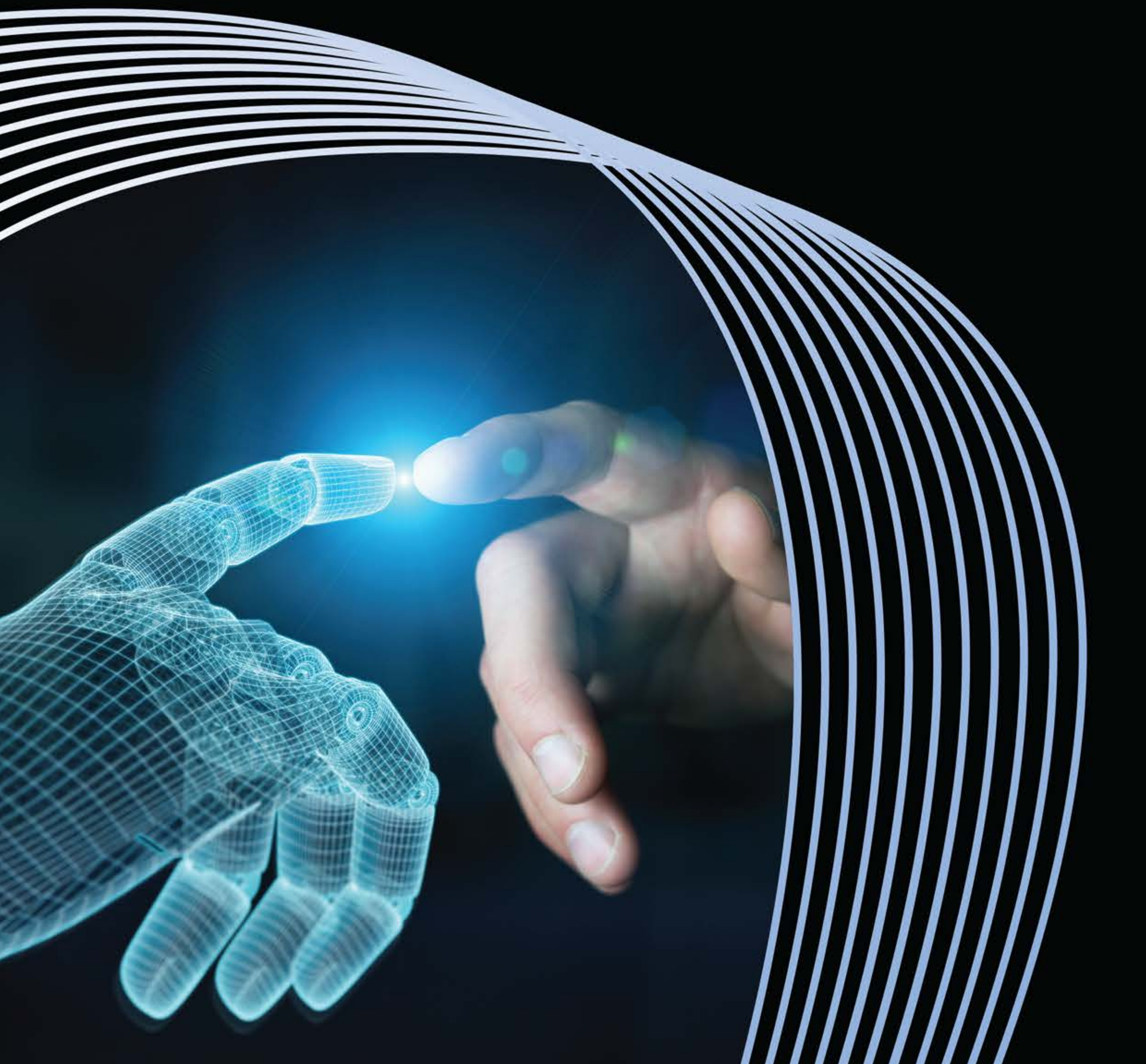
[2] CX Today, "The Impact of Speech Analytics on the Customer Experience," 2021. Accessed: Nov. 2, 2023. [Online]. Available: https://www.cxtoday.com/speech-analytics/the-impact-of-speech-analytics-on-the-customer-experience/.

[3] CX Today, "5 Use Cases for Generative AI In Conversational Analytics," 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://www.cxtoday.com/speech-analytics/5-use-cases-for-generative-ai-in-conversational-analytics-assemblyai/.

[4] PwC, "Driving customer service to the next level with data analytics and AI," 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-customer-experience-data-strategy.html.

DOI: 10.57650/INNOVACTION/2023.12

12 Generative AI – unleashing knowledge from information



The capacity to retrieve knowledge and contextual information from these increasingly larger digital stores is becoming critical, but it is also getting more complex due to the sheer volume and heterogeneity of information. In answer to the need for information access, Artificial Intelligence (AI) techniques are constantly evolving to deliver accurate responses to companies' needs for handling their information. The AI learning process, although too long and demanding, has been considered one of the most important processes for getting the right knowledge and insights from existing information.

The traditional AI learning process uses the supervised learning approach. However, this requires hard work in the creation of datasets that are used to train the model, datasets that quickly become outdated due to the emergence of new information sources. With the vast amount of information and processes that companies encounter daily, the traditional method is no longer sustainable. The arrival of large language models (LLM) with the Generative AI capacity, which allows content creation, has revolutionized the learning process, which is now faster, easier, and more intelligent to retrieve knowledge from heterogeneous information sources.

Authors

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

Sara Santos, Altice Labs, sara-r-santos@alticelabs.com

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

Keywords

Generative AI; Artificial Intelligence; Knowledge base; Large Language Models; Customer service

Introduction

Artificial Intelligence has made significant advancements and is becoming now part of our daily lives in almost every area, from social media, face ID and image recognition, apps (like online banking or driving assistance), and even shopping or leisure. An increasing number of business tasks are solved using tools that incorporate artificial intelligence into their capabilities.

This branch of computer science is quite complex and has more ramifications than one might imagine. Technologies such as Generative AI, which have accompanied AI since its inception, have been gaining significance. The ability to generate new content from an existing knowledge base or to navigate existing information more easily and objectively brings greater flexibility to businesses in decision-making and facilitates how they can reach their audience.

Given the substantial investments in the development of this area, it is anticipated that the resulting novel solutions, algorithms, or models to enhance capabilities and applicability will surpass current methodologies. This is expected to deeply impact multiple societal aspects, including the economy, legal system, politics, regulations, and businesses [1]. Therefore, it is imperative to engage in a comprehensive discussion to foresee potential consequences and mitigate the most significant adverse effects.

Changes driven by Artificial Intelligence

In the context of technological evolution, the concepts of threat and opportunity often intertwine. The opportunity for business growth can introduce risks to various factors, including the workforce, the environment, or even competitors and complementary industries. The line between these concepts is subtle, resulting in a certain degree of skepticism.

The ongoing transformative changes within the workforce present simultaneously an opportunity and a risk. The trend of transitioning specific job functions from humans to machines is gaining traction, generating concerns among the currently employed. An analysis by Challenger, Gray & Christmas, Inc. reveals that in May 2023, approximately 5% of job layoffs in the United States of America (USA) occurred due to the adoption of Artificial Intelligence for the automation of a range of tasks – such as creative work, writing, and administrative and clerical duties [2].

However, it is important to consider the broader perspective, as AI's adoption can offload certain tasks, allowing the human workforce to focus on roles that demand different skill sets. Additionally, this raises the need for creating new employment opportunities that call for specialized technical expertise. According to PwC [3], seven million existing jobs will be replaced by AI in the United Kingdom (UK) from 2017-2037, but 7.2 million jobs could be created. This uncertainty and the changes in the way some will make a living could be challenging. The noticeable shifts in how work is done and in the composition of the workforce pose challenges for both current professionals and those aspiring to build careers in areas where Artificial Intelligence is becoming increasingly relevant.

From another perspective, the improvement in businesses and company processes using Artificial Intelligence needs to be considered as an opportunity. A major example is the healthcare sector which holds the potential to experience significant transformations within its operational processes. Artificial Intelligence can reduce operational costs, leading to substantial savings, all while upholding the quality of healthcare facilities and medical organizations. According to estimates by McKinsey, the integration of big data in medicine and pharmaceuticals could generate up to $100 billion in value annually across the healthcare system in the USA [4].

The real impact, however, lies in the patient's journey. The integration of AI in medical procedures has the potential to enhance personalized treatment, simplify access to medical care, improve coordination among medical teams, and save time by streamlining information retrieval. Furthermore, AI serves as a powerful tool in disseminating information to the community during health crises, helping individuals comprehend regulations and procedures.

The risks and opportunities associated with societal changes resulting from the integration of AI in multiple contexts extend to new horizons with the creation of ChatGPT, developed by OpenAI. This super-intelligent chatbot is trained on a dataset of conversational text, processing the capability to understand and generate human-like text as well as perform tasks such as text completion, question answering, and language translating. Furthermore, the recent rise of creative generation through intelligent mechanisms has just begun to garner attention. Generative AI stands as a truly transformative force that merits exploration.

Generative AI

Generative AI, a branch of Artificial Intelligence, arises as an alternative to revolutionize the way to acquire, process, and transform data and information into new content, whether text, images, videos, code, or three-dimensional (3D) models, among others. Contrary to custom Natural Language Processing (NLP) models built for specific domains, Generative AI models excel in processing vast amounts of data in a short timeframe [5].

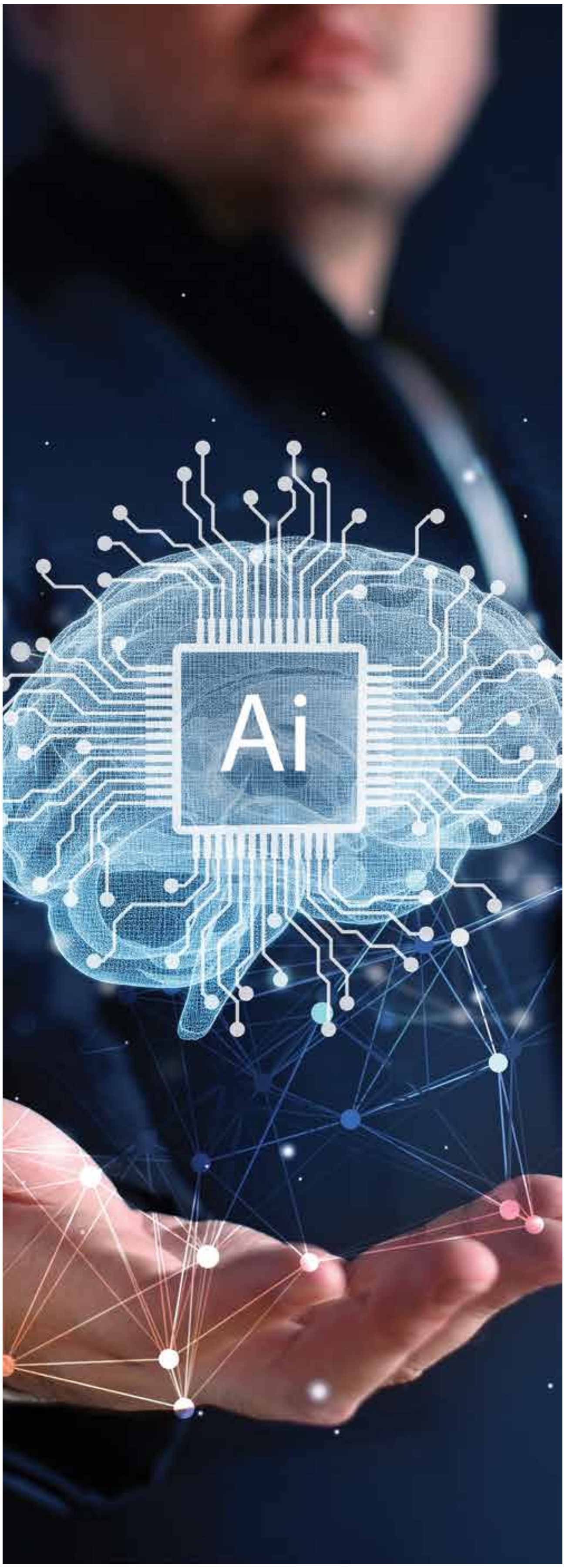
The concept of Generative AI is not new; it dates to the early days of Artificial Intelligence around 1950. Back in those years, when scientists and researchers introduced machine learning, the concept of using algorithms to create new data was also explored. However, as this field has evolved, Generative AI has gained new possibilities. The launch of ChatGPT in November 2022 coined the concept and brought it into the spotlight, making it imperative for smart tools to integrate this new capability [6].

Numerous technological companies have embarked on exploring the potential of Generative AI capabilities, with GitHub Copilot, Stable Diffusion, and other similar platforms capturing global attention. These platforms now offer solutions for generating and refining text, providing specific insights across diverse domains encompassing images, video, audio, and code. It is now possible to carry out a series of tasks on a single platform without the need to switch for each specific task. The current challenge lies in integrating the most advanced Generative AI models into existing platforms to provide the most distinctive and comprehensive solution possible.

OpenAI's recent introduction of GPT-4 (Generative Pre-trained Transformer) serves as an illustrative example. This substantial language model possesses enhanced data processing capabilities, enabling it to grasp complex contexts and interpret formulas and images with heightened accuracy. Furthermore, GPT-4 addresses issues present in prior models, like ChatGPT, which were launched merely four months earlier.

Anthropic, as an AI research company, created Claude, a Generative AI Platform, and has achieved another milestone in this progression. Its ability to process 100,000 tokens of text, equivalent to approximately 75,000 words in a minute, far surpasses its March 2023 iteration, which handled roughly 9,000 tokens. In May 2023, Google also stepped into the arena, announcing a host of new features driven by Generative AI, including the Search Generative Experience and the introduction of a new Large Language Model named Pathways Language Model – 2 (PaLM-2), set to power their Bard chatbot and other Google products [7].

New platforms are emerging continuously. According to McKinsey, the conversation around Generative AI has evolved significantly in a short period. With the rise of ChatGPT in 2022, business leaders were trying to understand Generative AI and distinguish between hype and reality. Now, nearly one year later, business leaders are having more sophisticated conversations, showing an opening to the viability of Generative AI in businesses.



Inside Generative AI models

Generative AI models use neural networks to generate original content by identifying patterns and structures within training data.

The term neural networks refers to computer systems inspired by the human brain, which can learn to distinguish patterns in data by themselves, dismissing human intervention. Unlike other machine learning approaches, Generative AI models can utilize unsupervised or semi-supervised learning methods during training. This allows organizations to more efficiently leverage a large amount of unlabeled data to create foundational models. These models serve as the basis for AI systems capable of performing a wide range of tasks.

It's important to highlight LLM, which is a type of neural network trained solely on text content to generate text. Generative AI text tools utilize LLM in their mechanisms since they are designed to predict the next word in a sequence of words, giving support to natural language processing tasks. They add value to intelligent conversational tools by providing sophisticated text-processing mechanisms.

The training methodology for LLM is deceptively simple, involving the pre-training of auto-regressive transformers on vast amounts of self-supervised data, followed by alignment with human preferences via techniques like Reinforcement Learning with Human Feedback (RLHF). However, despite the simplicity, the high computational requirements have restricted the development of LLM to only a select few players.

There have been public releases of pre-trained LLM, such as BLOOM, Large Language Model Meta AI – 1 (LLaMa-1), and Falcon, that match the performance of closed pre-trained competitors like GPT-3 and Chinchilla, however none of these public models are suitable substitutes for close "product" LLM, as ChatGPT, BARD, and Claude. The close products mentioned are extensively adjusted to align with human preferences, thereby greatly enhancing their usability and safety [8].

Examples like GPT-3 and Stable Diffusion offer users the ability to leverage the power of language and Generative AI in general. For example, popular applications like ChatGPT, which draws from GPT-3, enable users to generate essays based on a short text request. On the other hand, Stable Diffusion allows users to generate photorealistic images given a textual input.

Related to generative models, and by combining their best attributes, there are powerful tools to give answers to multiple market needs [9]:

1. **Diffusion models:** also known as denoising diffusion probabilistic models (DDPM), are a type of generative model that undergoes a two-step training process. This process involves a forward diffusion process that adds random noise to the data and after a reverse process undoes the noise to rebuild the data sample. This type of model offers greater performance, with a large processing scale and very good results for generalized use cases;

2. **Variational autoencoders (VAE):** VAE consists of two neural networks typically referred to as the encoder and decoder. When given an input, the encoder transforms it into a smaller, denser representation of the data, keeping only the relevant information for the decoder to reconstruct the original data while discarding any irrelevant information. These two components work in tandem to capture a simple and efficient representation of the data, which enables the model to map information and generate a fresh output. This model can be used to generate images quickly;

3. **Generative adversarial networks (GAN):** GANs were discovered in 2014 and were a widely used methodology before recent advances in this field. GANs work by pitting two neural networks against each other: a generator that generates new examples and a discriminator that learns to distinguish the generated content as real (from the domain) or false (generated). These networks are trained together and work symbiotically – they become more intelligent as the generator produces better content and the discriminator improves in identifying the generated content. The goal is that with each interaction, the content created is more redefined and indistinguishable from the original content. They are the most suitable model for generating domain-specific data;

4. **Transformer network:** transformers have revolutionized the creation of generative models by enabling the processing of sequential data in a non-sequential way. The transformer network consists of multiple layers capable of creating content, including the self-attention layer, feed-forward layer, normalization layer, and positional encoding, which work together to predict input streams. It is highly recommended for text-based generative applications, as the mechanisms used in the transformers focus on how words relate to each other and help represent the order of input words.

To determine the suitability of a Generative AI model for each particular use case, it is important to specify and consider certain key requirements, that are interconnected, as represented in **Figure 1**:

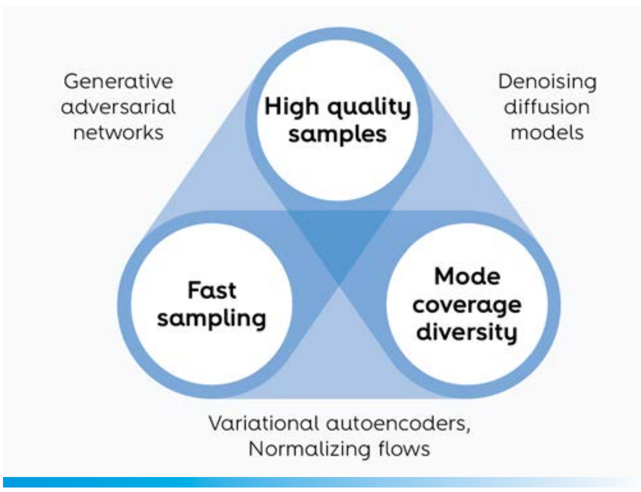


FIGURE 1 – Key requirements to variable Generative AI models, [9]

1. **Quality:** having high-quality generation outputs is crucial, particularly for applications that interact directly with users. As an example, in speech generation, poor speech quality can make it difficult for users to understand;

2. **Diversity:** a good generative model should capture the full range of data distribution without introducing biases;

3. **Speed:** in content creation workflows, there is often a need for real-time image editing, which requires fast generation. Many interactive applications serve this purpose.

As evidenced, the application of Large Language Models and Generative AI models have shown great promise as highly capable AI assistants. They excel in complex reasoning tasks demanding expert knowledge across various fields, including in specialized domains like programming and creative writing. They enable interaction with humans through intuitive chat interfaces. Among the most relevant and well-known interfaces, we can highlight:

- **ChatGPT and GPT Models:** developed by OpenAI with sponsorship from Microsoft, GPT models (currently in version GPT-4) were training in multiple phases, leveraging vast amounts of internet data, using deep learning techniques that give them extensive capabilities to almost anything from the simplest to the most complex problems;

- **Bard and Language Model for Dialogue Application (LaMDA):** Bard, created by Google to rival ChatGPT, uses LaMDA as a language model and places a strong emphasis on pre-training with diverse data sources. Bard aspires to become a more comprehensive model, embodying transparency in the AI process;

- **DALL-E:** DALL-E, created by OpenAI, arises as the most complete tool to generate and manipulate images. Currently, DALL-E 2 appropriately goes by user requests and has received more instructions on rejecting improper inputs to prevent inadequate outputs. With this tool, it is possible to use natural language inputs to produce high-quality images, manipulate and edit existing ones, and enhance pictures.

Applicability

The enthusiasm surrounding the potential of this innovation revolves around its seamless capacity to adeptly accomplish a broad spectrum of tasks:

- **Text generation:** Generative AI can easily create text content, including summaries, descriptions, essays, and various other forms of content for different platforms. The text can be flexibly adjusted to match the tone and language required for specific contexts. This versatility finds extensive applications in natural language processing, chatbots, and content creation;

- **Question answering:** Generative AI models, when applied to conversational tools, can be useful for answering questions, whether in a societal context, as seen in the case of ChatGPT, or in a business context;

- **Image generation:** these models can be trained to create or edit images, including real-world images or innovative compositions with various elements. Applications range from increasing machine learning model performance to creating art and generating product images, logos, or new types of visuals;

- **Code generation:** Generative AI models can be trained on a large dataset of code examples, and then, using the fine-tuned network, generate code that is similar in structure and function to the examples it has been subject to.

Advantages and disadvantages

A recent Gartner study polled over 2,500 executives on why they would adopt a Generative AI tool. The results are shown in **Figure 2**. The applications of Generative AI, in the right use cases with well-defined goals, can retain customers, and improve

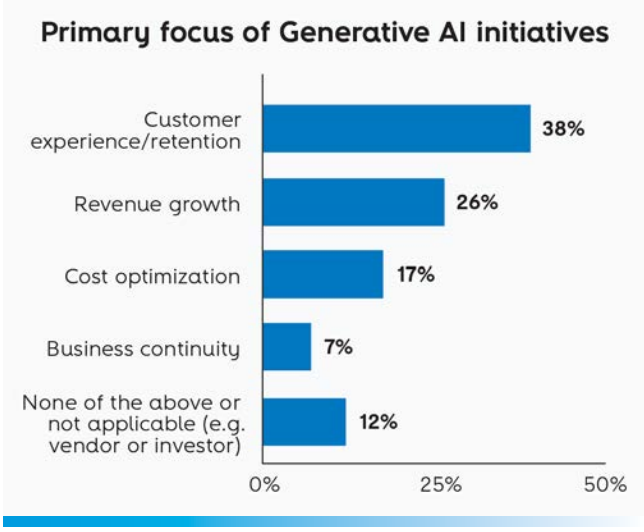


FIGURE 2 – Purpose of investments in Generative AI, [5]

their experience by promoting simpler and faster purchasing experiences and interactions with the company. On the employees' side, change can also be observed, as their day-to-day activities become more automated. Consequently, tasks that were previously repetitive and time-consuming, such as registering complaints and claims in a call center or recommending the right service/product to a customer can now be simplified with the application of Generative AI to tailor the needs of customers with contextual answers and personalized recommendations made based on the personal preferences of each customer.

The transformative impact on society is evident, Generative AI offers clear advantages in the business sector, where companies increasingly seek to streamline operations and reduce expenses. Overall, Generative AI is the next-generation technology that offers cost-optimization for companies, enabling them to thrive and remain competitive in today's dynamic business landscape.

However, it is important to acknowledge the real risks associated with this technology, which need careful consideration before companies and institutions move forward with the various use cases presented. The most prominent concern among the professionals from technology is data security. Data shared through these public models becomes public information, and since these companies are not forced to follow the General Data Protection Regulations (GDPR) or other copyright laws, information is at risk of being disclosed indiscriminately. The legal ambiguity in this field fails to safeguard companies in case of confidential informational leaks.

The ability to provide Generative AI capabilities on a platform also requires security measures for shared data. Therefore, it is imperative for platforms, including intelligent conversational platforms, to ensure that shared data will be handled transparently and in compliance with data protection laws.

Another challenge lies in the training and optimization of Generative AI models. The substantial volume of data required to train a model can lead to certain issues, such as mode collapse and overfitting. Mode collapse occurs when a model generates limited, repetitive outputs, failing to capture the full diversity of the data. Overfitting is when the model becomes overly tailored to the training data, making it less adaptable to new inputs. Finding the right balance between exploring new patterns and exploiting known patterns is crucial to producing high-quality generative tools.

Generative AI in a business context

But how can companies use and take advantage of this technological revolution in a business context? The answer is to embark on this technological revolution by leveraging their private information sources to ensure operational enhancements across different domains:

- **Customer experience:** the development of chatbots and virtual assistants can now be streamlined with the help of Generative AI models, enabling them to acquire essential knowledge and provide responses with richer context and a more human-like touch;

- **Reporting and data analytics:** business intelligence and performance reporting can benefit from Generative AI to process unstructured and qualitative data to extract insights for informed decision-making. Analysis enhances the efficiency of working with vast amounts of text and data;
- **Product development:** capacity to generate code facilitates the creation, development, and even design of new products.

The adoption of AI, particularly generative models, without a solid plan, presents a series of risks to companies. Training generative models is extremely expensive, requiring substantial investment. The necessary know-how and the corresponding learning curve are significant. The main options on the market are undergoing quick evolution, making it unclear which model is best to choose. Additionally, managing the ingestion of information sources to maintain uniqueness and privacy for companies is a critical point.

Altice Labs' BOTSchool Conversation AI Platform, thought and designed to help in day-to-day business activities, eliminating adoption difficulties, allows anyone without technical skills in a low code environment to create intelligent virtual assistants. With BOTSchool, companies can harness this technological revolution by leveraging their private information sources to ensure operational enhancements across different domains.

BOTSchool offers an information retrieval system with Generative AI capabilities that have the potential to transform the way companies harness knowledge. This powerful system leverages knowledge bases, which are repositories of various data sources, including PDF documents, text snippets, URLs, CSV files, and MS Excel sheets. These knowledge bases are intelligently converted into embeddings and stored in a dedicated vector database to facilitate future semantic searches. Integrated seamlessly with a robust language model, this system not only comprehends user inquiries but also produces thoughtful and perceptive replies. When a user submits a question, the system expertly analyzes it, transforms it into an embedding vector, and conducts a semantic search within the vector database, ensuring proximity-based results. It then assembles the most relevant knowledge, constructing a prompt for the large language model that includes system and context instructions. In multi-hop conversations, it maintains the memory of prior interactions, culminating in the delivery of an answer aligned with the knowledge base.

The BOTSchool platform with Generative AI integration allows companies/users to navigate information more quickly and easily, allowing them to perform tasks such as:

1. Define and manage as many different sources of information as needed, from information sources such as PDF, MS Word, and URL, keeping this information private and unique;
2. Have different information connectors that with a simple click, enable the ingestion of data from different data sources or formats into a simple document representation (text with metadata);
3. Select the most suitable Generative AI model for the intended use case, with the flexibility to choose from models that range from highly prescriptive to highly creative;
4. Allow no technological lock-in, and on the fly, the company can switch from different models: as an example, from GPT to LaMDA or LLaMa;
5. Process model fine-tuning through a set of parameters, like minimum similarity, top rank, platform, model, prompt, or temperature; Generative AI has the capacity, in a multichannel environment, to help companies, from customer service use cases to internal support.

BOTSchool helps companies build Generative AI-powered applications, like Questions & Answers (Q&A), chats, agents, and other internal and private custom information sources. BOTSchool combines the Retrieval Augmented Generation (RAG), as represented in **Figure 3**, which is a paradigm for augmenting LLM with companies' custom data that typically consists of the following stages:

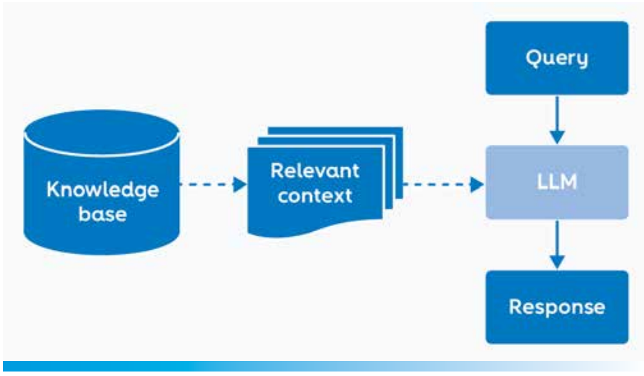


FIGURE 3 – BOTSchool – high-level RAG architecture

1. **Indexing stage:** preparing a knowledge base and indexing information in vector database (DB);
2. **Querying stage:** retrieving relevant context from the knowledge to assist the LLM/Generative AI in responding to a question.

The RAG [10] is a seq2seq (sequence-to-sequence) model that encapsulates two core components: a question encoder and a generator. During a forward pass, the input is encoded with the question encoder and passed to the retriever to extract relevant context documents. The documents are then prepended to the input.

Structure the query response

A query's response provides the input to the LLM prompt, so the quality of your search results is critical to success. To have higher accuracy, BOTSchool uses semantic search through cosine similarity. The composition or structure of the results depends on:

1. User question;
2. The last questions made by the user. Adding the last questions interactions, allows the system to support contextual questions e.g. first question: I want to subscribe to the internet service; second question: what's the price?;
3. Embedding removal is applied to questions and previous questions. Semantic search will return text segments matching the similarity of search embeddings to previously indexed embeddings (documents).

The matches to the query, ranked by relevance, similarity, or both are returned. BOTSchool permits the control of the number of segments returned and the minimum similarity value that will be "accepted" between the similarities of the search embeddings and the indexed ones. This value varies from zero to one, where zero indicates nothing similar and one is completely similar. For example, if the similarity result is three results, despite the top value being higher (for example, five), only three results will be returned.

With BOTSchool, companies can have the true benefits of this revolution, enabling new, more advanced, and intelligent interactions. The advancements are reflected in some of the following points:

1. Companies to have a true knowledge Information DB, where information can be quickly accessed in a smart way, as shown in **Figure 4**;
2. Companies can benefit from the Generative AI power, helping with Customer Service or Support interactions in all channels intended, quite effortless;
3. Companies can rely on a platform that easily maintains different information sources, making it possible to manage all these independently;
4. Companies can rely on the platform to keep the information private and specific, avoiding technological lock-in.

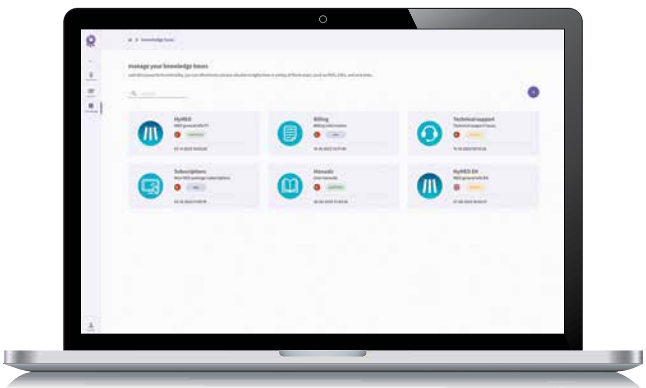


FIGURE 4 – Overview of the knowledge base in BOTSchool

Key takeaways

- Artificial Intelligence is revolutionizing business scenarios and opening new opportunities to streamline processes with AI-powered tools. The workforce will undergo significant changes with the advent of AI, presenting both risks and opportunities that need to be carefully studied to make the best use of this new technology;

- Generative AI is a field of Artificial Intelligence that has brought about innovative technology aimed at changing the way new content is created. With the help of tools like ChatGPT, Bard, StableDiffusion, and DALL-E, it is now possible to create and modify various types of content including text, images, videos, audio, code, and much more;

- The interfaces used today operate based on complex mechanisms that have been trained on vast amounts of data. As these models become increasingly more sophisticated, companies are constantly updating and creating new ways to improve them. This trend is followed by the rest of the world, leading to more and better tools being developed;

- BOTSchool priority is to make Generative AI technology accessible to everyone. It will enable the creation of even more intelligent chatbots and virtual assistants to help companies provide better customer support, and engagement, and improve internal processes. The main goal of BOTSchool's evolution with the introduction of Generative AI Capability is to offer a complete platform that delivers a seamless conversational experience for all users;

- The new capability added in BOTSchool with the integration of Generative AI will protect a company's knowledge base by allowing them to teach a model using their own database. Users can upload their database to the platform and the model will learn from it and become capable of completing a series of tasks ranging from answering questions, summarizing text, extracting entities, and building conversational tools more efficiently.

References

[1] B. Marr, "Bernard Marr & Co.". Accessed: Nov. 2, 2023. [Online]. Available: https://bernardmarr.com/what-is-the-impact-of-artificial-intelligence-ai-on-society/.

[2] E. Napolitano, "CBS News," 02 06 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://www.cbsnews.com/news/ai-job-losses-artificial-intelligence-challenger-report/.

[3] "PwC," 16 07 2018. Accessed: Nov.2, 2023. [Online]. Available: https://www.pwc.co.uk/press-room/press-releases/AI-will-create-as-many-jobs-as-it-displaces-by-boosting-economic-growth.html.

[4] J. Cattell, S. Chilukuri and M. Levy, "McKinsey & Company". Accessed: Nov. 2, 2023. [Online]. Available: https://www.mckinsey.com/industries/life-sciences/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d.

[5] "Gartner". Accessed: Nov. 2, 2023. [Online]. Available: https://www.gartner.com/en/topics/generative-ai.

[6] B. Marr, "Bernard Marr & Co.," 16 06 2023. Accessed: Nov. 2, 2023. [Online]. Available: https://bernardmarr.com/a-simple-guide-to-the-history-of-generative-ai/.

[7] M. &. Company, "The economic potential of generative AI: The next productivity frontier".

[8] Meta, "Llama 2: Open Foundation and Fine-Tuned Chat Models," 2023.

[9] "Nvidia," [Online]. Accessed: Nov. 2, 2023. [Online]. Available: https://www.nvidia.com/en-us/glossary/data-science/generative-ai/.

[10] "Huggin Face," [Online]. Accessed: Nov. 2, 2023. [Online]. Available: https://huggingface.co/.

Acronyms & Terms

1	**10G-PON**	10 Gigabit PON
2	**2D**	Two-dimensional
	2G	2nd Generation mobile networks
	25G/50G EPON	IEEE 802.3ca Standard
	25GS-PON	25G Symmetric PON
3	**3D**	Three-dimensional
	3GPP	3rd Generation Partnership Project
4	**4G**	Fourth generation mobile networks
	4K	3840 x 2160 pixels resolution
5	**50G-PON**	50 Gigabit PON
	5G	Fifth generation mobile networks
	5GPP	5G Infrastructure Public Private Partnership
6	**6G**	Sixth generation mobile networks
	6G-IA	6G Smart Networks and Services Industry Association
8	**8K**	7680 × 4320 pixels resolution
A	**AAL**	Active Assisted Living
	AI	Artificial Intelligence
	ALB	Altice Labs
	API	Application Programming Interface
	AR	Augmented Reality
	AUC	Area Under the Curve
B	**B2B**	Business to Business
	B2C	Business to Consumer
	B5G	Beyond 5G
	BDD	Behavior-Driven Development
	BLOOM	BigScience Large Open-Science Open-access Multiligual Language Model
C	**CCC**	Centro Cirúrgico de Coimbra
	CD	Continuous Delivery
	CE	Conformité Européene
	CEN	Comité Europeu de Normalização
	CENELEC	Comité Europeu de Normalização Eletrotécnica
	Cex	Coexistence Element
	CI	Continuous Integration
	CIAAA	Confidentiality, Integrity, Availability, Authenticity, Accountability
	CLI	Command-line interface
	CMM	Capability Maturity Model
	CNN	Convolutional Neural Networks
	CO2	Carbon Dioxide
	CPE	Customer Premises Equipment
	CPU	Central Processing Unit
	CT	Channel Termination
	CU	Control Unit
	CV	Continuous Variable
D	**DAG**	Directed-Acylic Graph
	DALL-E	A model for generating images from textual descriptions
	DAS	Distributed Antenna System
	DBA	Dynamic Bandwidth Allocation
	DDH	Decision Diffie-Hellman
	DevOps	Software development methodology that combines software development with information technology operations
	DevSecOps	Software development methodology that combines software development with information technology and security operations
	DHCP	Dynamic Host Configuration Protocol
	DISCRE-TION	Disruptive SDN Secure Communications for European Defence; pioneer project that aims at architecting a multi-level, multi-domain SDN
	DL	Deep Learning
	DLT	Distributed Ledger Technology
	DMM	Data Management Maturity Model
	DMSR	Deviation Map Single Report
	DNN	Deep Neural Networks
	DoS	Denial of Service
	DS	Data Science
	DV	Discrete Variable

E	**EBM**	Explainable Boosting Machine
	EU	European Union
	EuroQCI	A cross-boundary QCI that will span across the EU and its overseas territories
F	**FS**	Forecast Scenarios
	FTTB	Fiber-to-the-Building
	FTTH	Fiber-to-the-Home
	FTTx	Fiber-to-the-x
G	**GAN**	Generative Adversial Networks
	GDPR	General Data Protection Regulation
	GNS	Gabinete Nacional de Saúde
	GO	Glaucoma Overview
	GPON	Gigabit Passive Optical Network
	GPT	Generative Pre-trained Transformer
H	**HCD**	Human-Centered Design
	HIS	High-Speed Internet
	HSM	Hardware Security Model
	HSPON	High-Speed Passive Optical Network
	HTTP	HyperText Transfer Protocol
	HTTPS	HyperText Transfer Protocol Secure
I	**ICT**	Information and Communications Technology
	IEC	International Electrotechnical Commision
	IGMP	Internet Group Management Protocol
	IMT	International Mobile Telecommunications
	INESC TEC	Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
	IOP	Intraocular Pressure
	IoT	Internet of Things
	IP	Internet Protocol
	IPN	Instituto Pedro Nunes
	IPTV	Internet Protocol Television
	IRTF	Internet Research Task Force
	IS	Information Systems
	ISACA	Information Systems Audit and Control Association
	ISC	Integrated Sensing and Communications
	ISG	Industry Standards Group
	ISO	International Organization for Standardization
	ISP	Internet Service Providers
	IT	Information Technology
	IT Aveiro	Instituto de Telecomunicações de Aveiro
	ITS	Information-Theoretic Security
	ITU	International Telecommunication Union
	ITU-R	ITU Radiocommunication Sector
	ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
	ITU-T FG-NET2030	Focus Group on Technologies for Network 2030
	ITU-T G.984.x	Specific ITU-T's recommendation number that pertains to the GPON standard; the ".x" indicates that there might be multiple versions of the G.984 standard
J	**J**	Joule
	JTC	Joint Technical Committee
K	**Km**	Kilometer
	KMS	Key Management System
	KNN	K-Nearest Neighbors
	KPI	Key Performance Indicators
	KV	Key Value
	KVI	Key Value Indicator
L	**LaMDA**	Language Model for Dialogue Applications
	LAN	Local Area Network
	LGBM	Light Gradient Boosting Model
	LLM	Large Language Models
	LoRa	Long Range
	LR	Logistics Regression
	LTE	Long Term Evolution
M	**MAC**	Medium Access Control
	MAS	Multi-Source Agreement
	MFH	Mobile FrontHaul

	MIMO	Multiple-input-multiple-output/ Massive Input Massive Output
	ML	Machine Learning
	mmWave	Millimetric waves
	MNO	Mobile Network Operators
	MOCAP	Motion Capture
	ms	millisecond
	MUE	Multiuser Environment
N	**NLP**	Natural Language Processing
	NN	Neural Networks
	NNI	Network-to-Network Interface
	NSA	Non-Standalone
	NTN	Non-Terrestrial Network
O	**OCR**	Optical Character Recognition
	OCT	Optical Coherence Tomography
	OLT	Optical Line Terminal
	ONU	Optical Network Unit
	O-RAN	Open RAN
	OTT	Over-the-Top
P	**PC**	Personal Computer
	PKI	Public Key Infrastructure
	PLMN	Public Land Mobile Network
	POL	Passive Optical LAN
	PON	Passive Optical Network
	PoP	Point of Presence
	PPPoE	Point-to-Point Protocol over Ethernet
	PQC	Post-Quantum Cryptography
	PTQCI	The Portuguese portion of EuroQCI
Q	**QA**	Quality Assurance
	QAM	Quadrature Amplitude Modulation
	QCI	Quantum Communication Infrastructure
	QKD	Quantum Key Distribution
	QoS	Quality of Service
R	**RAG**	Retrieval Augmented Generation
	RAN	Radio Access Network
	rApp	Non-RT RIC App
	RedCap	Reduced Capability
	Rel	Release
	RF	Radio Frequency
	RF	Random Forest
	RGB	Red-Green-Blue
	RGBD	Red Green Blue-Depth
	RIC	RAN Intelligent Controller
	RIS	Reconfigurable Intelligent Surface
	RLHF	Reinforcement Learning with Human Feedback
	RNFL	Retinal Nerve Fiber Layer
	ROC	Receiver Operating Characteristic
	R-RIS	Reflecting-Reconfigurable Intelligent Surface
	RSA	Rivest-Shamir-Adleman
	RU	Radio Unit
	RU	Remote Unit
	RU	Radio Unit
S	**SA**	Standalone
	SBA	Service-Based Architecture
	SDLC	Software Development LifeCycle
	SDN	Software Defined Networks
	SDO	Standards Definition Organizations
	SEI	Software Engineering Insitute
	SG	Study Groups
	SHAP	SHapley Additive exPlanations
	SmartAL	Smart Assisted Living
	SME	Small and Midsize Enterprise
	SNMP	Simple Network Management Protocol
	SNS JU	Smart Networks and Services Joint Undertaking
	SOHO	Small Office/Home Ofice
	SP	Subproject
	SVM	Support Vector Machine
	SVN	Subversion
T	**Tbps**	Tera bits per second
	TC	Transmission Convergence
	TDM	Time-Division Multiplexer
	TDMA	Time-Division Multiple Access
	TELCO	Telecommunications
	THz	Tera Hertz
	T-RIS	Transmitting-Reconfigurable Intelligent Surface
	TRL	Technology Readiness Level
	TS	Technical Specifications
	TV	Television
	TWDM	Time and Wavelength Division Multiplexer

U	**UAV**	Unmanned Aerial Vehicle
	UC	University of Coimbra
	UI	User Interface
V	**V2X**	Vehicle to Everything
	VAE	Variational Autoencoders
	VFT	Visual Field Test
	VLAN	Virtual Local Area Network
	VoIP	Voice over IP
	VPN	Virtual Private Network
	VR	Virtual Reality
W	**WDM**	Wavelength Division Multiplexer
	Wi-Fi	IEEE 802.11x - Wireless Network (Wi-Fi Alliance)
	Wi-Fi 6	IEEE 802.11ax -Builds and improves on the 802.11ac Wi-Fi standard
	Wi-Fi 7	IEEE 802.11be -Builds and improves on the 802.11ax Wi-Fi standard
	WLAN	Wireless LAN
	WM	Wavelength Multiplexer
X	**xAPP**	eXtended Application
	XGS-PON	10-Gigabit-capable Symmetrical PON
	xPON	Designation for several PON technologies

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 2023

www.alticelabs.com