Augmented Intelligence

Next Level Network and Services Intelligence

A Technical Paper prepared for SCTE•ISBE by

Dr. Claudio Righetti
Chief Scientist
Telecom Argentina S.A.
Agüero 2392, Buenos Aires, Argentina
Phone: +5411 5530 4468
crighetti@teco.com.ar

Mariela Fiorenzo
Expert Tech Scientist
Telecom Argentina S.A.
mafiorenzo@teco.com.ar

Omar Hurtado
Tech Scientist
Telecom Argentina S.A.
ojhurtado@teco.com.ar

Gabriel Carro
Director Network Technolgy Architecture and Services
Telecom Argentina S.A.
gcarro@teco.com.ar

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

Abstract

Augmented Intelligence, also known as intelligence amplification, cognitive augmentation, decision support, machine augmented intelligence, and enhanced intelligence, is essentially Artificial Intelligence with a novel approach. While Artificial Intelligence is creating machines to work and react like humans, Augmented Intelligence is using those same machines with a different approach to improve human capabilities. In fact, Augmented Intelligence involves people and machines working together, leveraging their own strengths to achieve greater business value. In other words, the primary goal of Augmented Intelligence is to empower humans to work better and smarter. In this paper, we present the journey we are taking in Telecom Argentina, from data analytics, ML, AI to Augmented Intelligence.

Content

1. Introduction

We started applying data analysis technologies, then machine learning and Artificial Intelligence applied to our networks and services more than a decade ago in the former Cablevisión Argentina and continue after the 2018 merger, in Telecom Argentina. Telecom is in a process of digital transformation and as part of that process we understood that we should not only continue with the application of AI to our mobile, fixed and service networks, but take another step. Focus AI technology on humans.

Many times when people hear about algorithms, robots and AI, they imagine that such technologies compete with them. "*The technology is the easy part. The hard part is figuring out the social and institutional structures around the technology*" [REF John Seely Brown]

The application of Artificial Intelligence technology may not be successful if it is poorly adapted, designed or implemented. We must ensure that it is designed to help humans think better.

That's why we focus not only on AI technology, but also on human-machine collaboration, processes, and interfaces. This is Augmented Intelligence seeking to elevates human capabilities and experiences. By focusing AgI on serving people rather than replacing them. AgI can help them achieve their greatest potential.

1.1. Humans developing AI algorithms

It is well known that supervised ML algorithms are trained and tested with a large amount of data whose output variables are known (either numerical or categorical). These data consist of a large number of explicative variables or attributes that are in principle chosen at random.

Before "feeding" the machine the data are "worked", eliminating for example the missing data and outliers, standardizing, etc., all this so that no problems in learning the algorithm occur. With the algorithm written by humans, with the data that feeds the training also chosen by humans, we expect that the algorithm can predict a numerical or categorical variable with very little error when an unknown data is presented. The error is inherent in the system because this is an algorithm that does not want to "interpolate" the results but rather "approximate" them and therefore its goal is to minimize the error of its approximation. One way to fully understand this process is to use the PAC ("probably approximately correct") Learning framework [1] [ ] .

In a second phase, the human, can diminish the number of explicative variables with which he fed the system and the amount of the same ones, going from a big data to a small data. In short, we are moving towards a process of purification in terms of data quality. The human, expert, could be constantly testing new explicative variables, new sources of data, perhaps unconsciously pursued by the search for a causal law.

But we are not in a world of rules and laws, we are in a world where we are offered an innumerable amount of data with its attributes and we want to use them to decide or to predict at best. But sometimes what "is not" is as important as what "is". What we are not considering because of our own limitations or bias could be being left out of the data that train the algorithm.

And that should be a human task, to go out in search of the unknown, to make them present to us. As we see, human action is immersed in all aspects of AI. Defining the data, building the algorithms, which are constantly evolving and becoming more specialized, interpreting the results.

We could make an analogy with Kant's Copernican revolution, paraphrasing him and say that all knowledge has to "start" with data, but knowledge does not have its only origin in data. The constantly evolving algorithms used do not have their origin in data. Neither does the interpretation we give to the results according to our expertise.

In short, in this phase of our technology, a feedback process is needed between us, the humans and the machines we also create, constantly mediated by the data we know how to get.

As you can see, just like "Artificial Intelligence is not Intelligence, Machine Learning is not Learning" [ref Burkov, Andriy. The Hundred-Page Machine Learning Book (Páginaxvii). Andriy Burkov. Edición 2019].

2. What is Augmented Intelligence and why?

Our definition of Augmented Intelligence (AgI) is a powerful intelligence as the result of the collaboration between humans and machines, is the evolution of the Artificial Intelligence (AI) that enables humans to make better-informed decisions from complex data and Machine Learning (ML) algorithms. One aspect of AI that disturbs humanity is the possibility that people may be replaced by machines at their jobs. The goal of augmented intelligence is not to replace human beings or automate them out of existence but to enable them to make better decisions. From our perspective this collaboration between humans and machines will be the enabler for our industry to transition to the information revolution that is coming.

Gartner identifies this emerging technology as key for the design approach of new business solutions, balancing short-term automation with a mid/long-term approach that ensures improving quality not only by automation means, but also by amplifying human talent.

---

[1] In this framework, the learner receives samples and must select a generalization function (called the hypothesis) from a certain class of possible functions. The goal is that, with high probability (the "probably" part), the selected function will have low generalization error (the "approximately correct" part). The learner must be able to learn the concept given any arbitrary approximation ratio, probability of success, or distribution of the samples. The model was later extended to treat noise (misclassified samples). Source: Wikipedia

Figure 1 corresponds to the Hype Cycle for Emerging Technologies that Gartner publish every year, since 1995. According to the cycle, we can see that this technology will reach the plateau in 2 to 5 years [ref: gartner.com/SmarterWithGartner].

Figure 1 – Gartner Hype Cycle for Emerging Technologies, 2019 (Source: Gartner)

2.1. Some definitions

According to Merriam-Webster, information can be defined as the "communication or reception of knowledge or intelligence; knowledge obtained from investigation, study, or instruction". The future information revolution will be conducted by three main pillars: Free data and information, Small data and Augmented Intelligence (ref en Bell Labs). In this future, the goal will be to store the right amount of data that enables data scientist to discover knowledge and extract useful information by using new tools provided by the AgI.

We found that there is a confusion about the meaning of the terms data and information, it is believed that they both mean the same. These are the definitions we considered for this paper:

- *Data*: the actual captured observations.
- *Information*: the determination of relationships between data.
- *Knowledge*: the determination of models that describe the meaning of Information.
- *Technology*: a manner of accomplishing a task especially using technical processes, methods or knowledge (ref: Webster's Dictionary 2015).

Although data is easy to obtain, to establish the relationships between data is not. In many cases, it is possible to determine this relationship as the statistical correlation but it not necessary implies causation. Determining models and applications where this correlation makes sense is the most important component to succeed in discovering the meaning of information, that is, knowledge. We can define the five basic knowledge acquisition questions: *who, what, when, where* and *why?* These questions are an ideal starting point for assessing the relative value of information.

But the question *why?* is distinct from *who, what, when* and *where?* because forming an answer to *why?* implies some understanding or knowledge of a situation.

2.2. The goal of Augmented Intelligence

How this human-machine collaboration works? Both humans and machines have limitations doing their tasks. Humans cannot process or understand huge amount of data or complex data such as metadata, images, videos, etc. that machines can do and in a very fast way. On the other hand, machines cannot make business or economical decisions or understand the quality of data that they are processing or to choose the best strategy to achieve the industry goals.

Therefore, this collaboration surpasses humans and machines limitations by combining their intelligence and strengths. Humans are the ones that must use their knowledge and expertise to redesign the business process because the technology is not going to make such difficult decisions to transform the industry.

Augmented Intelligence will produce outcomes that neither humans nor machines could achieve alone (IBM). There are some steps to implement AgI:

- Decide whether to change the business process and task flow for human-machine cooperation[2].
- Select which tasks and decisions within the business process to automate.
- Determine the proper AI tools.
- Determine what data to acquire to better understand and model the business and customers.
- Build the data models.
- Test the results for reliability and accuracy.

Businesses powered by AgI systems have these four key attributes in common [ref: AgI eBook]:

- Discover and learn from hidden meaning in data and user interactions.
- Continuously learn, evolve, and improve with time.
- Build and test new user and process engagement capabilities.
- Drive new business processes and business model innovation.

In the last years, AgI was ranked in second place in AI technology rankings in terms of the value they create for businesses. However, Gartner predicts that “Decision support and AgI will surpass all other types of AI initiatives”, as it is shown in Figure 2, doubling virtual agents by 2025.

[2] We will see later in section 2.3 the importance of this point in the power of AgI

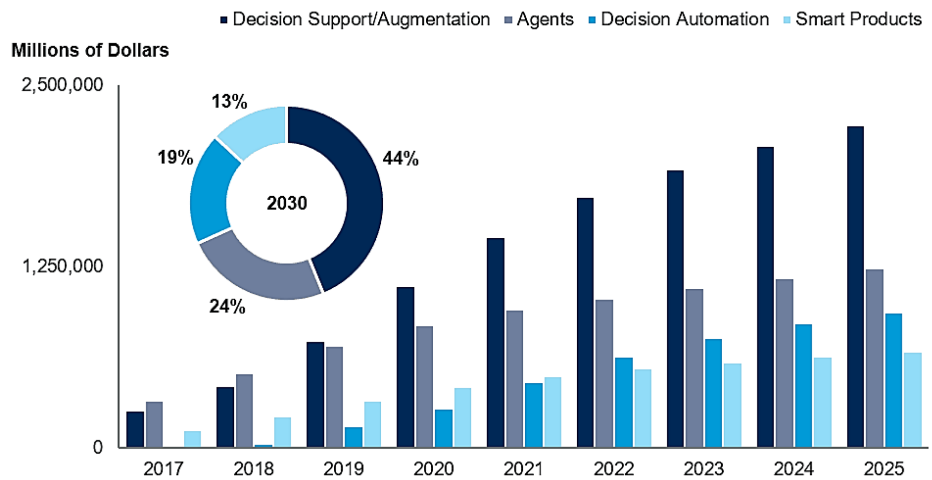

Figure 2 – Business Value Forecast by AI Type through 2025, expressed in millions of Dollars (Source: Gartner).

2.3. Historical Background of Augmented Intelligence

Throughout the history of humanity, man has created tools (language, writing elements, mathematics, etc.) that have allowed him to increase his intelligence. In the early 1960s Doug Engelbart [3]was researching how those tools shape our thoughts, founding the field of human–computer interaction. At the time, most of his colleagues only viewed computers to compute numbers and somehow man machine competition. However, he saw something deeper: He saw a way to increase the human mind [ref Engelbart, D. C. (1962, October). "Augmenting Human Intellect: A Conceptual Framework." Retrieved 10 July 2020 from https://www.dougengelbart.org/pubs/papers/scanned/Doug_EngelbartAugmentingHumanIntellect.pdf]

On February 10, 1996, a computer won a game of chess against a world champion for the first time. The computer was Deep Blue, a machine designed by IBM. An improved version of Deep Blue recorded its famous May 11, 1997 victory over world champion Kasparov, a milestone in artificial intelligence. Designed to understand high-power parallel processing, the "brute force" system could examine 200 million chess positions per second, beating Grandmaster 3.5-2.5. The story quickly centered around a Man vs. Machine narrative. For Kasparov it was a turning point. He may have asked himself : "Why want to compete against a machine when we could play with a machine?"

The next year, Garry Kasparov held the world's first game of "Centaur[4] Chess" [ref https://www.parc.com/blog/half-human-half-computer-meet-the-modern-centaur/] or, as it is more commonly known today, "freestyle chess" [5] (The concept of using computers to augment play had been around for a long time.) Humans can use input from chess programs to select their moves.

---

[3] Was an American engineer and inventor, and an early computer and Internet pioneer. He is best known for his work on founding the field of human–computer interaction, particularly while at his Augmentation Research Center Lab in SRI International, which resulted in creation of the computer mouse, and the development of hypertext, networked computers, and precursors to graphical user interfaces. These were demonstrated at The Mother of All Demos in 1968. Engelbart's law, the observation that the intrinsic rate of human performance is exponential, is named after him" https://www.dougengelbart.org/

[4] A centaur (/ˈsɛntɔːr/; Greek: κένταυρος, kéntauros, Latin: centaurus), or occasionally hippocentaur, is a creature from Greek mythology with the upper body of a human and the lower body and legs of a horse.

[5] In Freestyle Chess, human players are assisted by computers, software, and database tools.

In 2005 a freestyle chess tournament was organized, team called ZackS won by beating an opponent that included Vladimir Dobrov, a grandmaster, his highly rated teammate, and their computer programs.

The two members of the team were amateur chess players. They didn't play with the best hardware in the world. In fact, they had three different AI systems that run on mass-use computers. We had average players, with average computers, but a very good workflow. One of them was a soccer coach, the other was a database administrator. In this team the important thing was how they interacted and collaborated with the machines. That interface is what allowed it to succeed.

"Weak human + machine + superior process was greater than a strong computer and, remarkably, greater than a strong human + machine with an inferior process." [ref Garry Kasparov, How Life Imitates Chess: Making the Right Moves—from the Board to the Boardroom (New York: Bloomsbury, 2007), pp166]

By combining human intelligence with technological intelligence (Augmented Intelligence), these players tend to outdo anyone. In other words, centaurs can outperform humans and machines in the chess domain.

Summarizing centaurs combine the main characteristics of humans: INTUITION, JUDGMENT and FLEXIBILITY. With those of the machines: CONSISTENCY, PRECISION, SCALABILITY. Machines are for giving ANSWERS and humans are for asking QUESTIONS.

$$Centaur > Man\ or\ Machine$$

The "platforms" [6] have long since incorporated centaurs, today we say Augmented Intelligence, into their work teams.

2.4. Augmented Intelligence use cases

Although it is early days, AgI has already had a positive impact on many sectors and the nature of this technology, which learns more, adapts faster and continuously improves with time, means that early adopters do gain an advantage. But the main source of augmentation's business value will continue to be an improved human experience. AgI will deliver a level of personalization and it will also minimize errors, creating a higher standard of service.

Some of the following examples correspond to real applications and others are just possible use cases.

2.4.1. Healthcare

Augmented intelligence is transforming the industry, from detecting outbreaks to providing more customized care and explainable diagnoses. It has also been used in the fight against COVID-19.

2.4.2. Biotechnology

Biotech companies help augment doctors for radiology and clinical data. They use machine learning algorithms to extract key characteristics from radiology and pathology images and also organize the unstructured data of each patient's clinical notes, test results and health history. Doctors are provided with information to combine with their expertise in selecting treatment options for the patient.

[6] The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power

2.4.3. Financial services

Assisting financial planners to offer personalized services based on the customer's goals, capacity and risk appetite.

2.4.4. Retail

AgI helps increasing shopper engagement and conversion by enabling online shoppers to shop the way they think, using machine cognition of their declared, observed, and inferred behaviors.

2.4.5. Manufacturing

Aiding and accelerating the generative design process, whereby a human worker inputs the parameters and the machine finds countless ways of designing the object. The machine explores a plethora of options in record time and the human uses their expertise to select the best option, delivering value to customers and boosting efficiency.

2.4.6. Oil and gas

Optimizing precision drilling. The human worker can understand the environment they are operating in more accurately, leading to faster results and less wear, tear and damage to machinery. The list of possibilities is endless, but the common element is clearly to increase efficiency by heightening the worker's knowledge.

2.4.7. Geospatial images

Machine Learning is used to analyze geospatial imagery data. This provides real estate, energy and government agencies with information on land use, car and air traffic, and demographic trends so they can make better decisions.

2.4.8. Telecommunications

AgI uses information gathered from applications to maximize network configurations and simplify troubleshooting. This next level of network intelligence comes from AI, data analytics and ML, to enable better correlation among events on the network, user and device behavior.

3. The technology behind Augmented Intelligence

There is a convergence of technologies that have come together to lead the market towards the reality of augmented intelligence and in this section, we will explore them.

3.1. Types of data

Augmented Intelligence systems can work with all types of data (structured, unstructured, semi-structured and metadata) from many sources, across disparate and siloed systems.

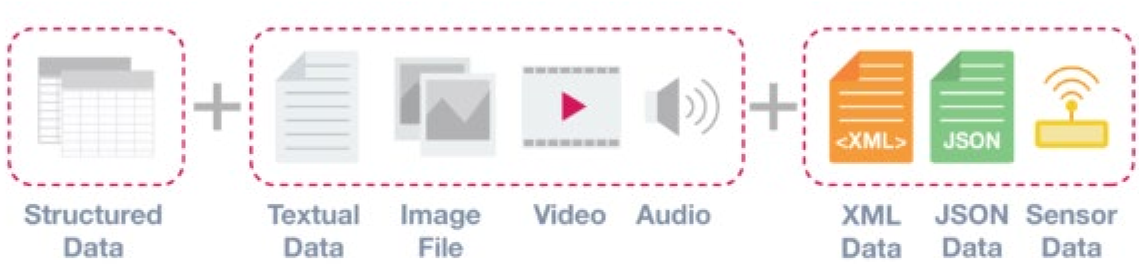

Figure 3 – Three types of data: Structured, Unstructured and Semi-structured Data

3.1.1. Structured data

It is considered the most 'traditional' form of data storage. It refers to information that has a defined length and format such as numbers, names or dates. This type of data is linked to a pre-defined data model and is therefore straightforward to analyze. They fit to a tabular format with relationship between the different rows and columns. This structure made it possible to create understandable answers to questions inside this data [ref: https://www.bigdataframework.org/data-types-structured-vs-unstructured-data/]

3.1.2. Unstructured data

It does not have a predefined data model nor is it organized in a predefined manner. Therefore, it is not surprising that most of the information in the world is unstructured, for example, videos, images, text documents. The ability to analyze unstructured data is especially relevant in the context of Big Data, since a large part of data in organizations is unstructured. Of course, there is inherent structure, but the difference is that humans must do the hard work to understand the hidden structure of the data.

3.1.3. Semi-structured data

It is a form of structured data that does not conform with the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contain tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure. JSON and XML are two examples of this type of data. They are considerably easier to analyze than unstructured data and many tools have the ability to 'read' and process them.

3.2. Big Data and Small Data

The terms Big Data and Small Data have become popular buzzwords over the last years and it is not always clear what either of these terms means or how or when to use each.

The acquisition of information requires the ability to capture data, compute something based on the data available and obtain a result.

Over the last decade, the promising concept of Big Data has generated huge expectations to industries. Therefore, companies have purchased and deployed scalable storage and processing systems with the intention of preserve every single byte of data obtained from their systems and customers. But actually, much of this data has no real information or valuable content. As storage costs began to increment, big data applications suffered a transformation into a new type of applications of small data where the value arises not from the volume of the data set, but from the ability to extract useful information and to make decisions based upon the smallest data set. The goal will not be to measure and store every byte; it will be to measure and store "just the right amount" of data [ref Bell Labs].

Next, we present some formal definitions with respect these concepts.

3.2.1. Big Data

Gartner proposed an early definition using three 'Vs' (Volume, Velocity, and Variety) to represent the key characteristics of Big Data. Certainty, it is still widely used.

"Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enables enhanced insight, decision making, and process automation."

More recent definitions include other "V-terms" into the Gartner model, adding Veracity to reflect data accuracy and Value to address the usefulness of the data.

Big data techniques are designed to manage a huge volume of disparate data at the right speed and within the right time frame. The goal is to enable near real-time analysis and action.

3.2.2. Small Data

A formal definition of small data has been proposed by Allen Bonde, former vice-president of Innovation at Actuatet:

"Small data connects people with timely, meaningful insights (derived from big data and/or "local" sources), organized and packaged – often visually – to be accessible, understandable, and actionable for everyday tasks".

Martin Lindstrom defines it as 'the tiny clues that uncover huge trends', based on observational data. It's also defined as data that is small enough size for human comprehension.

From the white paper [ref Philosophy of Small Data], "Just only one apple falls on Isaac Newton's head, not ten, not thousand".

3.3. Artificial Intelligence

Artificial Intelligence means getting a computer to mimic human behavior in some way. The goal is to get computers to perform tasks as human: things that required intelligence.

Despite recent hype around the technology, AI is not a new technology and not a product of this century's innovations. The beginnings of AI can be traced to the middle of the 20th century. During the second world war, an English mathematician and computer scientist, Alan Turing documented his ideas on creating an intelligent machine. He proposed a test for machine intelligence:

A human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboard and screen so the result would not depend on the machine's ability to render words as speech. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. The test results do not depend on the machine's ability to give correct answers to questions, only how closely its answers resemble those a human would give [ref Wikipedia Turing Test].

In short, if a machine can trick humans into thinking it is human, so then it has intelligence.

Marvin Minsky of MIT's Project Mac, who has made major contributions to Artificial Intelligence, in 1970 said: "In from three to eight years we will have a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable."

Figure 4 illustrates the AI Timeline. Despite the fall between 1966-1997, which was called the AI winter, the constant evolution of this technology is evident.

Today, AI refers to a range of technologies form automation to deep learning. It includes the subfields of natural language processing, vision, robotics, machine learning, and knowledge representation and reasoning.

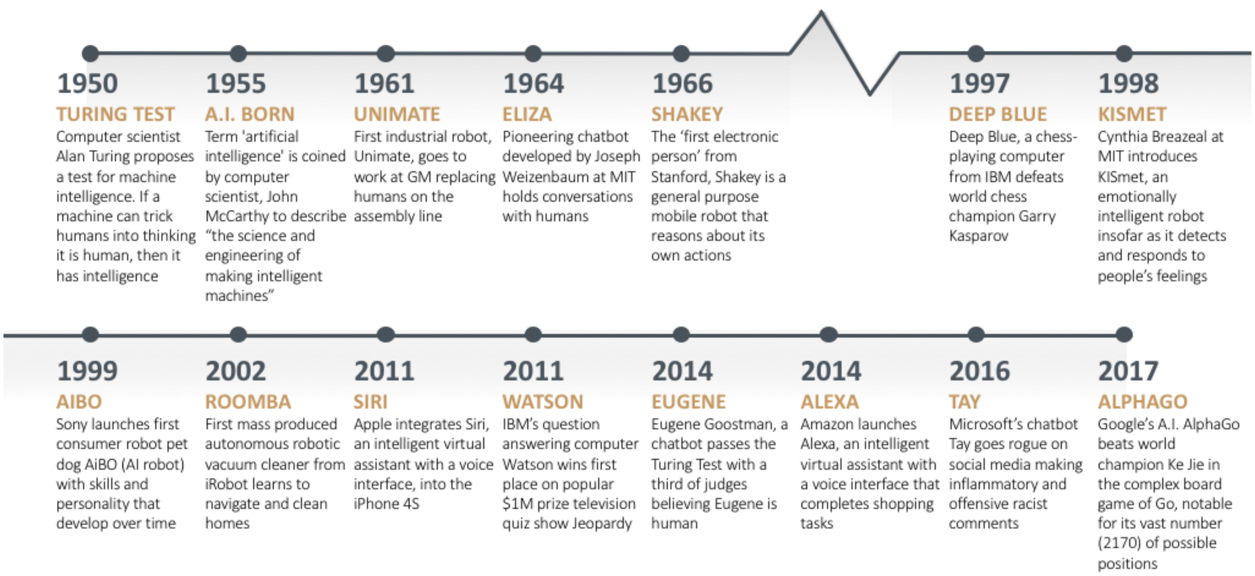

Figure 4 – Artificial Intelligence Timeline: It is not something new.

The question that arises is what the difference between Artificial Intelligence and Augmented Intelligence is. The answer can be found within its objectives; the goal of an artificial intelligence system is to simulate human cognitive capabilities in a system that can function independently of humans. In contrast, the goal of an augmented intelligence system is to enhance human intelligence by human–machine collaboration to get work done.

3.4. Machine Learning

Machine learning is a form of artificial intelligence that enables a system to learn from data rather than through explicit programming of a set of rules. It consists of a variety of types of algorithms, all of which learn from data. A machine learning model is the output generated when you train your machine learning algorithm with this data. After training, when you provide the model with new data input, its output will be providing new insights such as data classification or prediction.

The end-to-end machine learning process includes the following phases:

- Business Goal Identification
- ML Problem Framing
- Data Collection and Integration
- Data Preparation
- Data Visualization and Analytics
- Feature Engineering
- Model Training
- Model Evaluation
- Business Evaluation
- Production Deployment

Machine learning models can be online or offline, online models are constantly ingesting data and interacting with it in near real time mode improving the model outcomes. On the other hand, offline models once they are deployed, they only can be retrained manually with new data.

There are several approaches to machine learning that are relevant to the ability to create algorithms that support the industry problems, they are based on the type of the data. These approaches are divided into three main areas: supervised learning, unsupervised learning (it includes deep learning) and reinforcement learning.

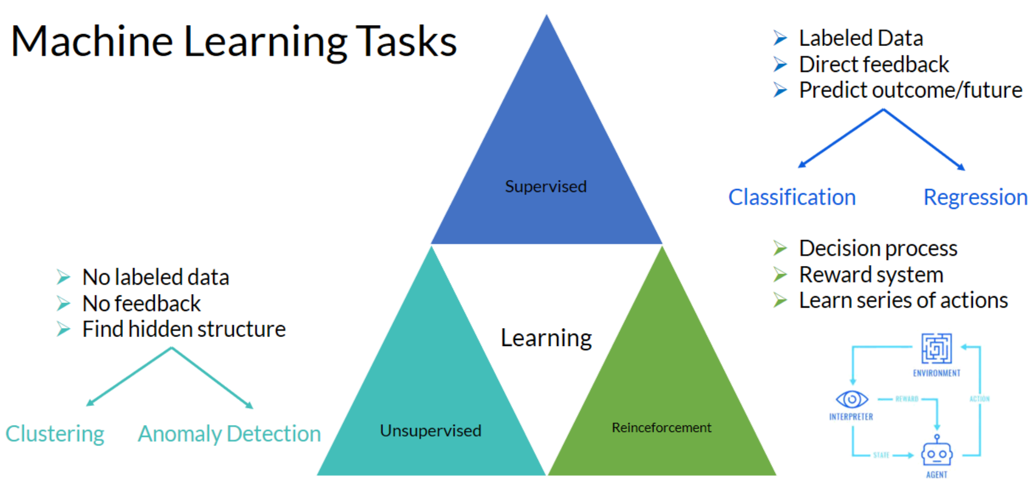

Figure 5 gives an explanation about the type of data, type of outcome and the algorithms methodology of each category. It also shows some of the most common techniques: regression, classification, anomaly detection, etc.

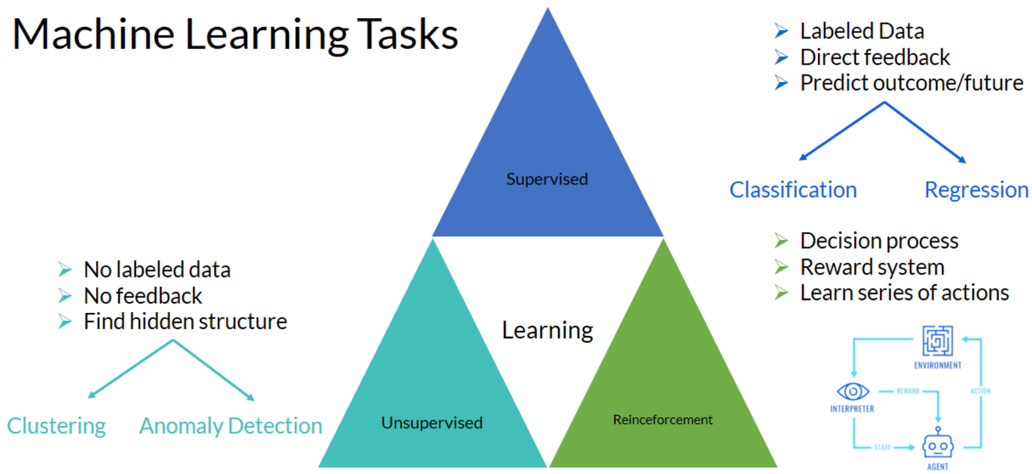

Figure 5 – Machine Learning tasks divided by the three main categories and most common algorithms.

3.4.1. Supervised Learning

Supervised machine learning algorithms are designed to learn by example. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, a supervised learning algorithm will take in new unseen inputs and will determine which label the new inputs will be classified as based on prior training data. The resulting model must be evaluated against test data to see how well it learned. If the model is fit to only represent the patterns that exist in the training set, overfitting[7] occurs. Using unseen data for the test set can help you evaluate the accuracy of the model in predicting outcomes.

3.4.2. Unsupervised Learning

Unsupervised learning is a set of statistical tools for scenarios in which there is only a set of features and no labels. With these techniques we are interested in finding discovering subgroups of similar observations. It tends to be more challenging, because there is no clear objective for the analysis. Besides, it is hard to evaluate if the obtained results are good, since there is no accepted mechanism for validating results on an independent dataset, because we do not know the true answer.

3.4.3. Reinforcement Learning

Reinforcement learning is a behavioral learning model. It is about taking suitable actions to maximize reward in a situation wherein an agent interacts with a new environment using actions and discovering errors or rewards. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation. In short, a reinforcement learning system learns through trial and error. Therefore, a sequence of successful decisions will result in the process being "reinforced" because it best solves the problem at hand. Gaming and robotics are the most common applications of this technique.

[7] The production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably [ref: Oxford Dictionaries.com]

3.4.4. Deep Learning

Deep learning is a specific method of machine learning that incorporates neural networks in successive layers in order to learn from data in an iterative manner. Deep learning is especially useful when you are trying to learn patterns from unstructured data.

A neural network consists of three or more layers: an input layer (ingested data), one or many hidden layers (which include weighted nodes), and an output layer (the outcome). The term deep learning is used when there are multiple hidden layers within a neural network.

Neural networks and deep learning are often used in applications where images, videos or speech are involved.

3.5. Automation

Within MSOs and MNOs, this means automation of processes that were previously carried out by people like configuration, management, operation and testing of physical and virtual devices within the network. With growing costs and the daily emergence of bandwidth-hungry applications, networks cannot be managed manually. Increased levels of network automation help to reduce complexity and are essential for businesses to keep up in the digital world. AI is an enabling technology that may (or may not) help with the process of automation. What it culminates is a network that is highly predictable and highly available improving the business outcomes.

3.6. Cognitive computing

Like AI, cognitive computing is based on the ability of machines to sense, reason, act and adapt based on learned experience. Cognitive computing refers to computing that focuses on reasoning and understanding at a higher level and in a manner that is analogous to human cognition, rationale, and judgement. Applications of cognitive computing include speech recognition, sentiment analysis, face detection, risk assessment and fraud detection. The difference between AI and cognitive computing lies in the way they approach the purpose of simplifying tasks. AI is used to augment human thinking and solve complex problems. Cognitive computing mimics human behavior and reasoning to solve complex problems similar to the way humans solve problems.

3.7. Intent-Based Networking

"Intent" is the keyword in this technology, which describes a network's business objective or an outcome.

Intent-based networking (IBN) is an emerging technology concept that aims to apply a deeper level of intelligence and intended state insights to networking. Ideally, these insights replace the manual processes of configuring networks and reacting to network issues. The goal is networking that uses machine learning and cognitive computing to enable more automation and less time spent on manual configuration and management. With intent-based networking, network administrators define the intent and the network's software finds how to achieve that goal using AI and ML by performing routine tasks, setting policies, responding to system events, and verifying that actions have been done.

These systems not only automate time-consuming tasks and provide real-time visibility into a network's activity to validate a given intent, they also predict potential deviations to that intent, and prescribe the action required to ensure it. This greater intelligence makes the network faster and more agile and reduces errors [ref https://www.vmware.com/topics/glossary/content/intent-based-networking#:~:text=Intent%2Dbased%20networking%20relies%20on,and%20actions%20have%20been%20achieved.].

4. Knowledge Plane, state of the art

The research community has considered in the past the application of AI techniques to control and operate networks. For example, in 2003 David Clark et. al propose the knowledge plane (KP) as a *pervasive system within the network that builds and maintains high level models of what the network is supposed to do, in order to provide services and advice to other elements of the network. The knowledge plane is novel in its reliance on the tools of AI and cognitive systems* (Clark, Partridge, Ramming, & Wroclawski, 2003).

The Knowledge Plane (KP) paradigm proposes the evolution to a cognitive network, where the devices learn, decide, and act to achieve end-to-end goals. This emerging paradigm is clarifying a set of new cognitive-based protocols and algorithms that optimize network's performance.

In (Mestres, y otros, 2017) progress is made in the definition of a new paradigm based on this plane. This is knowledge-defined network (KDN) operates by means of a control loop to provide automation, recommendation, optimization, validation and estimation.

4.1. Introduction

There are several organizations and working groups proposing frameworks, standards and how to apply the AI to the industry of communications, networks and services. Below we detail those in which Telecom Argentina is participating or is in consultation with:

- SCTE & CableLabs AI/ML Working Groups
- Telecom Infra Project - AI/ML Working Group
- TM Forum - AI & Data Analytics
- ITU-T Study Group 13 and ITU-T FG ML5G Studying network architectures, use cases, and data formats for the adoption of machine learning methods in 5G and future networks.
- ETSI ISG ENI (Experiential Network Intelligence). Defining a cognitive network management architecture based on AI methods and context-aware policies; five deliverables have already been released
- 3GPPP
- ONAP

AI began to be studied in most of the different Standards Developing Organizations (SDO). We provide in this item a summary of what they are proposing, how they are defining it and the specific use of this new technology in each branch of the networks.

As has been said repeatedly in various academic circles, AI makes use of another phenomenon that occurs: the massive growth of data, driven mainly by network technologies such as IoT, 5G and 10G project.

The introduction of AI and automation using AI, in short, seeks to ensure greater performance and efficiency of networks.

The paradigm shift that will be brought about by the introduction of these new technologies includes a substantial shift from a focus on network operations to a focus on the user experience.

We have conceptualized these AI and automation tools that interact with the different types of networks as **Knowledge Plane**, a "place" where the massive amount of data obtained from the network is processed with the different AI tools, either in real time or in a post-processing, and that, based on results, produces modifications in the network itself. This is called **closed-loop automation**.

Before we start it would be necessary to break down what we used to call “network” into a general form.

Mobile networks:

Split up into:

- Terminals (mobile phones, IoT device, etc.)
- Access (cells, RAT, Fronthaul/Midhaul/Backhaul)
- Packet core (EPC (4G), 5GCN, etc)

Fixed networks:

- XDSL (DSLAM/BNG)
- FTTH, GPON (ONT/OLT)
- HFC (CM/CMTS)
- IPBB (routers, DNS, CDN, etc.)

And in common with all of them, a growing trend towards virtualization.

4.2. Autonomic networks

4.2.1. ETSI (European Telecommunications Standards Institute)

In the current research on the ETSI recommendation we still do not see clearly the difference between the initiatives that are established within this SDO, as some of them promote doing exactly the same thing. This is further complicated by the fact that each of them does not refer to interaction with the others. Anyway, we consider it important to explain them.

4.2.1.1. ETSI-ENI (Experiential Networked Intelligence)

The concept of ENI refers to a working group developed at ETSI to improve the operator's experience that is active since 2017.

This working group has collaboration with all the SDO's in the industry.

It is based on the introduction of AI systems in the Network Management System and a control model that takes the following actions:

- Observe
- Guide
- Decide
- Act

The **knowledge plane**, in ETSI is the ENI, which is materialized in this recommendation as a layer that covers transversally all the layers of the MANO (NFV Management and Orchestration) architecture.

In this case the AI-based system will have to adjust the services offered according to the following possible changes:

- In the client's needs
- Under the ambient conditions
- In business objectives

As a challenge it is established:

- Determine which services comply with the SLA and which would be about to fail to comply depending on the changing context.
- Provide an experiential architecture (i.e. an architecture that uses the benefits of AI).
- Establish incident detection.
- Possessing the capacity for autonomous incident management.

4.2.1.2. ETSI-ZSM (Zero-touch service and network management)

In 2017 ETSI launched the Zero touch network and Service Management Industry Specification Group (ZSM ISG).

The aim of this group is to specify all processes and operational tasks as an example:

- Deployment
- Configuration
- Assurance
- Optimization

are executed automatically (agile, efficient management and automation).

4.2.1.3. ETSI-GANA (Generic Autonomic Networking Architecture)

It is a reference architecture model for the creation of autonomous networks, "cognitive" networks and self-management of networks and services in which the AI plays an important role in such autonomy.

It is designed to realize the AMC (Autonomic Management and Control) paradigm (closed-loop service instantiation and adaptive operations) of networks and services with reference to:

- Control architectures
- Management architectures

The relevance of each of these ETSI proposals needs to be analyzed to see if they replace or complement each other.

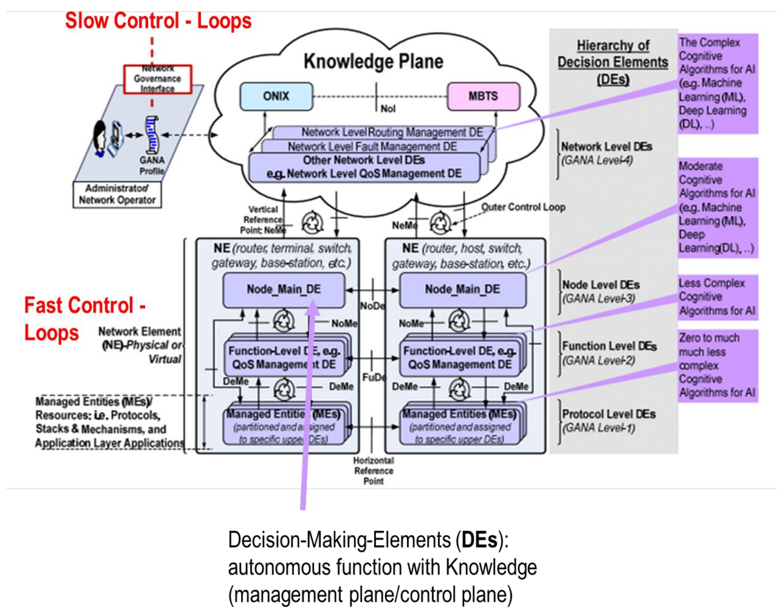

Figure 6 – GANA Architecture

4.2.2. ONAP (Open Networking Automation Platform)

ONAP provides a comprehensive platform for the real-time, policy-based orchestration and automation of physical (PNF: Physical Network Functions) and virtual (VNF: Virtual Network Functions) network functions that will enable software, network, IT and cloud providers and developers to rapidly automate new services and support full life-cycle management.

In short, ONAP is a platform that offers a complete set of tools to automate assurance processes in the field of network management.

Some operators have selected it to automate layers of:

- The MANO domain (except for the VIM functions (VNF life cycle management).
- The orchestration service for both PNFs and VNFs.

ONAP makes use of a functionality called DCAE (Data Collection, Analytics and Events Project) that is the general name for several components that collectively fulfill the role of Data Collection, Analysis and Event Generation.

The DCAE architecture aims at the deployment of components and the composition of flexible, pluggable, microservice-oriented and model-based services. DCAE also supports multi-site data collection and analysis that are essential for large ONAP deployments.

The DCAE is a place where analytics applications and AI/ML models could reside.

4.3. Mobile Networks

4.3.1. O-RAN Alliance (Open Radio Access Network)

This SDO, which is "openness", proposes exactly the same solutions as the SDOs in terms of the use of the AI, but specifically applied to the RAN.

It proposes that networks should:

- Be self-managed.
- Be able to take advantage of new learning-based technologies to automate the operational functions of the network and reduce OPEX.
- Leverage emerging deep learning techniques to integrate intelligence into every layer of the RAN architecture.

Embedded intelligence, applied at both the component (NE) and network levels, will enable dynamic local allocation of radio resources and will optimize the efficiency of the entire network in closed-loop automation using AI.

As it is possible to see, the O-RAN describes in the orchestration and automation layer a Non-Real Time RAN Intelligence Controller (RIC) with AI/ML capabilities.

It should be remembered that there are currently initiatives for the virtualization of the Non-Real Time BBU.

4.3.2. 3GPP (3rd Generation Partnership Project)

Service and System Aspects (SA) is the Technical Specifications Group of 3GPP where most work is currently being done on the use of AI, especially in

- SA WG2, Architecture
- SA WG5, Telecom Management

4.3.2.1. SA2 NWDAF (Network Data Analytics Function)

The NWDAF was first introduced into the 5G system architecture at the 3GPP SA2#119 meeting in February 2017.

The NWDAF (Network Data Analytics Function), as defined in TS 23.503, is used to collect data such as FCAPS (Fault Configuration Accounting Performance Security) events and Data events, and then perform the analysis of the data centrally. An NWDAF can be used for analysis of one or more network slices.

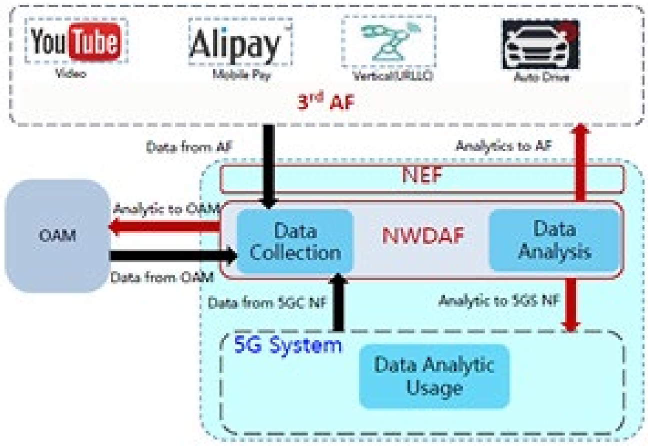

Figure 7 – NWDAF uses case

4.3.2.2. SA5 MDAF (Management data analytics function)

MDAF (management data analysis function) uses network management data collected from the same network, for example data related to

- Services
- Slicing functions
- Network functions
- Related slicing and networking functions

and perform the corresponding "analytics".

The MDAF can be deployed at different levels, for example, at the domain level:

- RAN Management Data Analytics Function
- Network Management Data Analytics function
- Network Slice Subnet Instance Management Data Analytics function

In order to establish a difference between NWDAF and MDAF we can say that the first one is a function in the Packet Core domain or for the management of use cases in the slice domain, while the last one is in the OAM layer supporting the assurance. Both relate to each other by exchanging information.

4.4. Use cases of AI in networks

4.4.1. General use cases:

The following cases are added to those already established in each standard:

- In the NOC (Network Operation Center):
  - In the preventive support to allow to identify and solve problems before they affect the performance of the network, avoiding critical cuts and providing stability.

    - Incident pattern recognition.
    - Incident based subscriber type clustering.
- Anomaly detection
    - Discovery of deviations from standard behavior.
    - Determination of outliers is multidimensional spaces.
    - AI Tools:
        - ARIMA (autoregressive integrated moving average)
        - Random Forest
        - Self-encoder
        - Principal Component Analysis
- In the RAN
    - Embedded Analytics components (incorporation of chips) in the same Radio Base Stations to perform:
        - Closed-loop automation (internal automation and localized training components).
        - Improve performance and spectral efficiency.
        - Improving Mobility Management.
        - Adaptation of links.
        - Energy saving in MIMO (Multiple Input Multiple Output) in its sleep state.
        - Reduce energy consumption.
        - Evolution of SON (Self-organizing network).
        - Interference diagnosis.
        - Real-time analysis in the baseband unit.
        - ML to improve the algorithms of the cell itself:
            - In the User Plane.
            - With QoE optimization.
            - In the management of radio resources.
            - In the scheduler (especially for MIMO).
- On terminals connected to the mobile network.
- IoT terminals.
- AI can improve network efficiency by being able to cluster and detect anomalies in the initial state of implementation of the ever-growing diversity of devices, in an environment where vendors freely interpret GSMA standards.
- Cellphones.
- The application of the AI tools as in the case of the RAN network elements, will allow them to
    - Reduce latency time and optimize spectrum management.
    - Applications are developed where they can be made:
        - Image and sound recognition and interpretation.
        - Development of Augmented Reality and Virtual Reality.
- Visual inspection of network equipment (deep learning).
- In Energy.
- In Planning:
    - Work is already underway with the mobile access management of Network and Service Planning using forecasting methods (e.g. ARIMA).

- On the WiFi:
  - Not yet determined by any organism.

4.4.2. Particular use cases with 3GPP SA2 NWDAF (Network Data Analytics Function)

The NWDAF can be used for the following applications:

- Assist in the provision of Quality of Service (QoS) profiles.
- Assist in the adjustment of the quality of service (QoS).
- Assist in policy determination.
- Collaborate with 5G Edge Computing.
- Improve performance and monitoring of mIoT (massive Internet of Things) terminals.
- Assist in load/balance balancing of NF (Network Function).
- Assist in areas of the network with instability (oscillating conditions).
- Improve performance and monitoring of mIoT terminals.
- Support for the exposure of the network status on the Northbound interface using APIs (application programming interface).

5. The Future of Augmented Intelligence

5.1. Augmented Intelligence

As we mention before Augmented Intelligence is still in the technology trigger phase in the hype cycle according to Gartner. And, in early stage of the evolution of Artificial Intelligence and Machine Learning. The main question that organizations is asking about the future of these technologies is: Will the human-machines collaboration result in fewer jobs for people?

What is clear is that the nature of work is already changing and will continue to change through the human-machine collaboration. Both, humans and machines will do what they do best. Machines will automate routine tasks that don't need human intelligence to let humans focus on handling exceptions. When humans handle exceptions, they must get an informative alert with context from the machine, often with a recommendation on how to proceed.

There will be many new jobs that do not exist today. With the evolution of augmented intelligence and its presence in more and different domains, there will be a greater need for regulatory frameworks. In the future, many jobs will be needed to manage augmented intelligence and handle the exceptions have not existed before. One of the greatest challenges for society will be the massive training of those who are displaced by intelligent systems to fill the new jobs that augmented intelligence enables.

Also, in the future, data and information will become free and freely available, big data will be replaced by small data to discover "knowledge", and new augmented intelligence tools will be developed that assist in the acquisition of knowledge (cognition) by enabling critical thinking from multiple perspectives.

5.2. Explainable AI

Another emerging field is Explainable AI (XAI). Whether when an expert must decide based on augmented intelligence tools or when they are implemented in our networks and services, they must understand how the results were achieved and the level of confidence that the model has.

The goal of enabling explain ability in ML, as stated by "is to ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms". S. Barocas, S. Friedler, M. Hardt, J. Kroll, S. Venka-Tasubramanian, and H. Wallach. The FAT-ML Workshop Series on Fairness, Accountability, and Transparency in Machine Learning. Accessed: July. 20, 2020. [Online]. Available: http://www.fatml.org

Explain ability sits at the intersection of transparency (consumers have the right to have decisions affecting them explained in understandable terms), causality (it is expected of the algorithms to provide not only inferences but also explanations), bias (the absence of bias should be guaranteed), fairness (it should be verified that decisions made by AI are fair) and safety (reliability of AI systems) (Hagras, 2018).

According to ((BSI), 2016), (Gasser & Almeida, 2017) we know that many machine learning algorithms have been labeled "black box" models because of their inscrutable inner-workings. What makes these models accurate is what makes their results difficult to interpret and understand they are very complex. So, even when some abstraction or transformation the models can be explainable, not always they are auditable. The discussion about audit AI is still open (Forum, 2019).

6. Augmented Intelligence at Telecom Argentina

More than a decade ago, in the former Cablevisión Argentina, we started working with Analytics to generate outcomes about our networks and services. Then, we began to innovate with other technologies related to Machine Learning and Artificial Intelligence and we continue after the 2018 merger, in Telecom Argentina.

Nowadays, as a STEM management our mission at Telecom Argentina is:

- Find new AI and ML based technologies to add value to the business.
- Define the technological strategy and emerging technologies.
- Develop innovative scientific tools to improve our infrastructure based on the demand for our services.
- Generate dimensioning models and performance parameters from statistical and mathematical analysis.

Our industry is undergoing a process of digital transformation and we, Telecom Argentina, are part of this process. Being part of this process also implies a transformation in our ways of working and in how we focus on them, but always oriented towards networks and services. So, we understood that we should jump to the next level.

We are not alone doing that; we belong to different Working Groups with the objective to define standards and best practices of these technologies. We also share our experiences and use cases.

They are:

- Artificial Intelligence and Machine Learning - SCTE
- Cross Industry AI/ML/Data Analytics Collaboration - CableLabs
- DOCSIS Data Analytics Work Group - CableLabs

Below, we detail some of the challenges we face from the emergence of this new AI that is human-machine collaboration.

6.1. VMAF: a tool for measuring video based on human perception

VMAF (Video Multimethod Assessment Fusion) is a model proposed by Netflix in 2016 and is a video quality metric that combines human vision modeling with ML in order to provide a great viewing experience to their members.

VMAF is a fusion of elementary metrics into a final metric using a Support Vector Machines regressor which assigns weights to each elementary metric (Visual Information Fidelity, Detail Loss Metric, Temporal Information). In this way, the final metric is able to preserve all the strengths of the individual metrics and deliver a more accurate final score.

The model is trained and tested using a dataset of several 10 seconds long clips of different video genders and content characteristics. Those videos are distorted using different resolutions and bitrates. Then a subjective experiment is performed where a focus group of non-expert viewers score a source clip and the same clip distorted. From this experiment, the Mean Opinion Score (MOS) is calculated and it is also used to train the model.

Last year we have been working in an adaptation of this algorithm to our own OTT video platform: FLOW. We designed our datasets based on FLOW catalog and produced the distortions. Then, we performed some focus groups in order to obtain the corresponding MOS [ref SCTE 2019]. After doing that, we train and test the model with the elementary metrics mentioned before but applied to our content, the obtained MOS from the focus group and the SVM parameters that fitted the best.

With this tool we can qualify videos in a scale from “bad” to “excellent”, not only from an objective perspective but considering human perception too, in an automatic way with the help of machines. Previously, this task was performed by video experts, consuming a lot of time and resources. One of the advantages of VMAF is that video experts can dedicate time to analyze other video aspects and making decisions that machines can’t do and letting them do the repetitive tasks. This is a clear example of human-machine collaboration.

6.2. STEM-ML: a tool for capacity planning

We developed a neural network-based machine learning tool for planning the capacity of the DOCSIS network [SCTE 2017], which was clearly superior to the tools used at the time.

However, it was very costly for us to adopt the tool by the teams responsible for capacity planning.

The business teams won't use the models because they do not know how they made their decisions and cannot be sure they work as advertised. It was difficult for us to explain to them that the decision to divide or not a node results from patterns in the data that the ML algorithm found and fixed into the model.

We finally understood that it was very risky for them to rely on models that cannot be explained.

From that moment on, we involve the teams from the moment they raise the need for an augmented intelligence tool. We focused on humans who for example carried out mobile network planning.

Although we have not yet used the formal tools that are being developed to explain the models, we have involved the team from the beginning of the development of STEM-ML for the planning of our 4G network. Applying agile methodologies for development. Explaining each model using and providing the planning team with the design criteria. Once an algorithmic model is up and running, the team must test the model carefully to see that it is operating in a reasonable way.

We were also developing with the team how to make the comparison between the results of the model and the methodology they had been using before.

We are defining with the team different levels of testing before putting the tool into production.

6.3. Telecom Argentina Knowledge Plane

We at Telecom Argentina adopt the knowledge plane paradigm for our Telco Cloud project [paper 2019]. During the proofs of concept that we will begin to deploy with next year, we will not only advance on AI / ML technologies, but it is essential to understand this abstraction as an AgI tool that will allow us to operate our future networks and services.

At the beginning, and in order to have a common language, a framework was defined in which the preponderance of the knowledge plane in the different layers of the network, both virtual and physical, is highlighted.

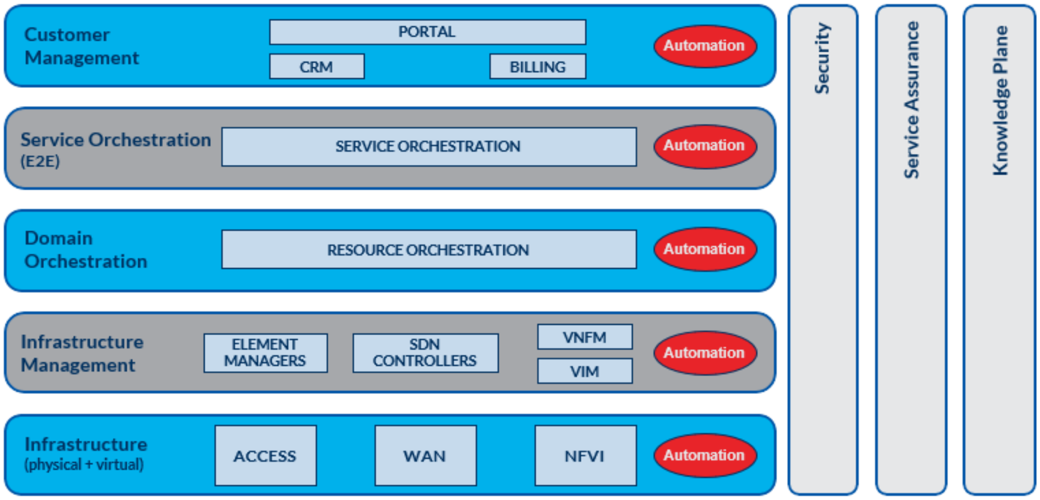

Figure 8 – Telecom Argentina Knowledge Plane

There is no doubt that in the course of time this general vision will not only be modified but also specified. A result that only experience can provide.

7. Conclusion

Humans and machines working in collaboration can have a powerful impact on the effectiveness of business processes. Augmented Intelligence overcomes the limitations of isolating human understanding from the massive amounts of available data to analyze complexity in record time.

Telecom Argentina has begun the journey to AgI. In a jungle of recommendations from the different SDOs for each technology, the proof of concept with vendors and developers will introduce us to the best practices of human-machine collaboration, human beings contributing with interpretation, critical thinking and multiple perspectives or looks to the solutions that AI provides us.

Abbreviations

3GPP	third generation partnership project
5G	fifth generation of mobile networks
5GCN	5G core network
AgI	augmented intelligence
AI	artificial intelligence
AMC	autonomic management and control
ARIMA	autoregressive integrated moving average
BBU	band base unit
BNG	broadband network gateways
CAPEX	capital expenditure
CDN	content delivery network
CLI	command line interface

CM	cable modem
CMTS	cable modem termination system
DCAE	data collection, analytics and events project
DE	decision element
DNS	domain name servers
DOCSIS	data over cable service interface specification
DSLAM	digital subscriber line access multiplexer
ENI	experiential networked intelligence
EPC	evolved packet core
ETSI	European Telecommunications Standards Institute
FCAPS	fault configuration accounting performance security
FM	failure management
FTTH	fiber to the home
GANA	generic autonomic networking architecture
GPON	gigabit-capable passive optical network
GSMA	global system for mobile communications association
HFC	hybrid fiber-coaxial
IBN	intent-based networking
IoT	internet of things
IPBB	ip backbone
IT	information technology
KDN	knowledge-defined network
KP	knowledge plane
MANO	management and orchestration
MDAF	management data analytics function
MIMO	multiple input multiple output
mIoT	massive internet of things
ML	machine learning
MNO	mobile network operator
MOS	mean opinion score
MSO	multi system operator
NaaS	network as a service
NE	network element
NFV	network functions virtualization
NOC	network operation center
NSSF	network slice selection function
NSSI-MDAF	network slice subnet instance management data analytics function
NWDAF	network data analytics function
NW-MADF	network management data analytics function
OLT	optical line termination
ONAP	open networking automation platform
ONT	optical network terminal
OPEX	operating expenditure
O-RAN	open radio access network
PAC	probably approximately correct
PCF	policy control function
PLMN	public land mobile network
PM	performance management

PNF	physical network functions
QoE	quality of experience
QoS	quality of service
RAN	radio access network
RAN-MDAF	radio access network management data analytics function
RAT	radio access technology
RIC	ran intelligence controller
SA	service and system aspects
SCTE	Society of Cable Telecommunications Engineers
SND	software defined network
SDO	standards developing organizations
SLA	service level agreement
SOC	service operation center
SON	self-organizing network
STEM	science, technology, engineering and mathematics
SVM	support vector machines
VIM	virtualized infrastructure manager
VMAF	video multimethod assessment fusion
VNF	virtual network functions
WG	working group
XAI	explained artificial intelligence
ZSM	zero-touch service and network management