IEEE

IEEE INGR

International Network Generations Roadmap

2023 Edition

Artificial Intelligence and Machine Learning

IEEE Future NETWORKS

An IEEE Future Networks Technology Roadmap

futurenetworks.ieee.org/roadmap

Table of Contents

Tables

Figures

ABSTRACT

In the evolution of artificial Intelligence (AI) and machine learning (ML); reasoning, knowledge representation, planning, learning, natural language processing, perception, and the ability to move and manipulate objects have been widely used. These features enable the creation of intelligent mechanisms for decision support to overcome the limits of human knowledge processing. In addition, ML algorithms enable applications to draw conclusions and make predictions based on existing data without human supervision, leading to quick near-optimal solutions even in problems with high dimensionality. Hence, autonomy is a key aspect of current and future AI/ML algorithms.

This chapter focuses on the development and implementation of AI/ML technologies for 5G and future networks. The objective is to illustrate how these technologies can be migrated into 5G systems to increase their performance and to decrease their cost. To that end, this chapter presents the drivers, needs, challenges, enablers, and potential solutions identified for the AI/ML field as applicable to future networks over three-, five-, and ten-year horizons.

AI/ML applications for 5G are wide and diverse. Some key areas described include networking, securing, cloud computing, and others. Over time, this paper will evolve to encompass even more areas where AI/ML technologies can improve future network performance objectives.

Key words:

AI, ML, DL, CNN, DNN, RNN, GAN, GPU, Cloud Computing, MEC

CONTRIBUTORS

Dr. Deepak Kataria	IP Junction, USA
Dr. Anwar Walid	Nokia Bell Labs, USA
Dr. Mahmoud Daneshmand	Stevens Institute of Technology, USA
Dr. Ashutosh Dutta	Johns Hopkins University Applied Physics Lab, USA
Dr. Michael A. Enright	Quantum Dimension, Inc., USA
Dr. Rentao Gu	BUPT, China
Alex Lackpour	Drexel University, USA
Prakash Ramachandran	Emerging Open Tech Foundation, USA
Dr. Honggang Wang	UMass Dartmouth, USA
Dr. Chi-Ming Chen	AT&T (Retired), USA
Baw Chng	BAWMAN LLC, USA
Dr. Frederica Darema	InfoSymbiotic Systems Society, USA
Pranab Das	Verizon, USA
T. K. Lala	ZcureZ, USA
Editors	
Baw Chng	BAWMAN LLC, USA
Ripal Ranpara	Atmiya University, India

Acknowledgements	
Brad Kloza	IEEE Future Networks Initiative
Matthew Borst	IEEE Future Networks Initiative
Craig Polk	IEEE Future Network Technical Community

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

1. INTRODUCTION

1.1. Working Group Vision

This chapter gives a broad summary of what to expect from the more in-depth roadmap effort being developed by the IEEE International Network Generations Roadmap (INGR) Initiative Artificial Intelligence and Machine Learning (AI/ML) working group. It describes a high-level perspective and projection of the AI/ML technology areas for 5G and future networks. The team has reviewed the research papers and forecasts in this 2023 edition of the IEEE INGR roadmap. The scope and stakeholders are summarized. Several expected linkages among the other INGR roadmap working groups are presented.

There are two understandings of artificial intelligence (AI) as a field of research. One approach takes a philosophical attitude and is interested in the possibilities of building artificial systems with intelligent behavior primarily as a quest for knowledge – to understand intelligence itself and test the abilities of computers. This attitude prevailed in the era of AI research starting in the 1950s.[1] The other attitude is aimed at practical applications and is interested in building computing systems that exhibit some form of intelligence. This roadmap is related to the latter, the practical approach in artificial intelligence.

Artificial Intelligence is a wide-ranging branch of computer science concerned with building machines capable of performing tasks that hitherto required human intelligence. The development of AI technologies opens an opportunity to use them for conventional applications (expert systems, intelligent databases[1], technical diagnostics[2],[3] etc.), as well as for automation of the manufacturing sector.

ML is the study of algorithms that improve automatically through experience[4]. ML, and its sub-area Deep Learning (DL), are part of AI. ML enables the capability of learning or sensing domain-specific data streams provided to it and helps build and/or improve the models for predicting, inferring, and adapting to guide or act as domain experts similar to human subject matter experts.

One area in which AI/ML can have a significant impact is Cloud Computing (CC). In recent years, CC has gained a lot of interest due to its ability to allow applications to provide infrastructure services to a large number of stakeholders with assorted and dynamically changing requirements. This capability is particularly important for 5G networks, which are described in the following sections. Technically, CC is composed of physical resources, including compute, memory, network, and storage resources. It uses the virtualization technique that allows for sharing a single physical instance of a resource or an application among various stakeholders and enterprises[4]. Virtualization, which is a key element of Network Function Virtualization (NFV) for cybersecurity needs, and CC are growing rapidly in today's technological ecosystem and the objectives are to include and optimize these technologies for future networks. This document demonstrates how AI/ML can be smoothly migrated to support current and future systems.

[1] *See, for example, A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.*

To support the aforementioned objectives, the goal of the AI/ML working group is to explore research in data science, AI/ML algorithms and their application to future networks, and to define a framework that uses open-source technology and commercial architecture to run AI/ML workloads. In addition, this working group will study solutions that may need standardization by IEEE International Networks Generations Roadmap (INGR) for edge computing, Internet of Things (IoT), 5G, and other efforts in coordination with other INGR and international working groups.

1.2. Scope of Working Group Effort

This working group roadmap is to undertake the following areas:

- Define the taxonomy of AI as to what is the domain and hierarchy of cognitive processing of network data and metadata — the paradigm that defines problems and solutions using the metadata and hierarchy of terms or the terminology. Ontological tools address the entity relation behavior of the data in the metadata. Knowledge represented by such data enables machines for specific domains to demonstrate the intelligence used by machines in cognitive computing.
- Identify the state of AI (sense, infer, and act like a human) and ML (detection, classification, segmentation, predictions, and recommendations) with respect to the state-of-the-art, e.g., at use-case level for efficient computing.
- Survey existing frameworks that support AI/ML workloads for domains: security for identity and privacy, health management, intelligent transportation, smart cities, connected cars, and others. All these need to specify new High-Performance Computing (HPC) architectures and identify a reference architecture to compare emerging protocol stacks and infrastructure elements.
- Provide the roadmap based on research and industry advancement to deliver the AI/ML vision beyond 5G. As technology matures, AI/ML will be used to deliver advanced capabilities. This will require tensor and parallel computations for use in fields like bioinformatics, blockchain, quantum computing, etc. The domain-specific life cycle of data collection, cleaning, batch, build, deploy, and test to improve the accuracy of predictions; depend on use cases and technology used. The goal of this is to review some of the prominent use cases to illustrate the emerging trends and gaps and contribute to new standards.

1.3. Linkages and Stakeholders

The proliferation of INGR working groups covering similar domains is to be expected as enabling technologies in AI/ML is expected to work with the following categories.

- ***Enablers & Users***: The AI/ML working group falls under this bundle of working groups: Security, Applications and Services, and Deployments for AI/ML. Therefore, this team will work with other teams to identify and collaborate on opportunities to implement and enhance the AI/ML architecture.
- ***Access***: The access groups are mostly tied to connecting through different mediums like radio nodes. As a result, we see their specific interactions with this category of working groups.
- ***Networking***: Edge Automation Platform (EAP), security and satellite working groups may bring critical data to edge and IoT analytics and processing, and thus need advanced AI/ML algorithms and technologies to process and optimize their systems. Identifying these specifics will be an important part of this collaboration.

Some areas to specifically probe for wireless and core networks may include:

- ***AI/ML for Wireless Networks***: management, security, policy, performance, optimization
- ***AI/ML for Energy Efficiency***: energy management for access networks, data centers, edge, core
- ***AI/ML for Spectrum Management***: 5G FR1 and FR2 frequency ranges, licensed / unlicensed / shared network and spectrum resources, cognitive radio
- ***Distributed AI/ML over Wireless Edge Networks***: vBBU, micro data center, energy optimization, vertical and horizontal network slicing
- ***AI/ML for Wireless Services***: Quality of Service (QoS) adaptation and Quality of Experience (QoE)
- ***Systems & Standards***: This category will require evaluating the infrastructure and systems standards to propose for AI/ML and HPC workloads. As a result, scope may include aspects of benchmarking used for processing AI/ML workloads. Emerging technologies in AI/ML can be used to optimize media like text, speech, voice, image, video, and associated platforms. It will be critical to utilize commercial components with existing interoperability standards, such as Open Neural Network eXchange (ONNX), for future platforms and applications.

The AI/ML working group will address the above categories to arrive at appropriate work items. The core work of identifying use cases dictates what possible platforms this group will either propose building or using.

1.4. 2023 Edition Update

The INGR AI/ML chapter's 2023 edition contains the following updates:

- New section on AI/ML use cases (see Section 5)
- New topical section on AI-NWDAF (Network Data Analytic Function) (see Sections 2.3)
- New topical section on AI/ML for Radio Access Network (RAN) (see Section 3.10)
- New topical section on Intelligent RAN (see Section 0)

2. TODAY'S LANDSCAPE

2.1. Types of Learning

Artificial intelligence (AI) gives machines the ability to exhibit human-like intelligence without being explicitly programmed with detailed and specific instructions. Machine learning (ML) uses statistical methods to enable machines to learn from data. Deep Learning (DL) uses neural networks (NN), referred to as a Deep Neural Network (DNN), with a combination of multiple layers of ML, with each layer learning from the previous one, to learn and infer non-linear relationships among elements of an input feature vector. This relationship is demonstrated via a Venn diagram[5] in Figure 1.

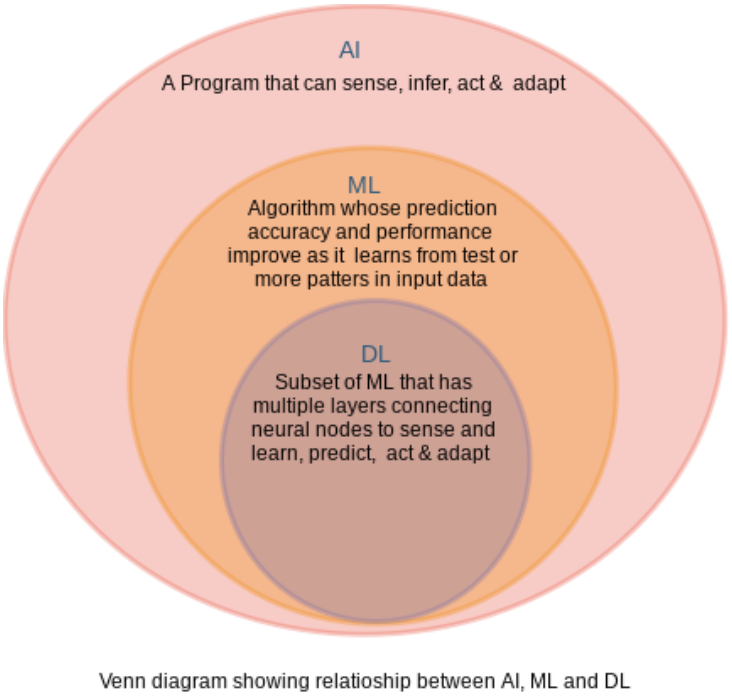

Figure 1: Artificial Intelligence and its Relationship to Machine Learning and Deep Learning

One of the most active areas of AI/ML is supervised learning with DNN. It has been increasingly adopted in many application areas due to:

- Advances in research and algorithms
- Wide availability of big data
- Cheaper computing power and storage

The motivation behind machine learning is to allow and enable the application to analyze and draw conclusions on the data. The algorithm is capable of doing this by learning the intricacies of the problem by attempting to solve it with data-driven predictions and decisions, progressively improving its performance in one predetermined task. In this area, the objective is to learn the problem (i.e., train the program), which can be done by feeding it historic input and output. In other words, what is the output (performance) X of solution Y, given the input Z. Given a sufficient number of tuples (X, Y, Z) and a method for evaluating the value of X, the program should be capable of drawing patterns and eventually creating efficient solutions.

This definition is very generic and broad and there are many variations of it, but the fundamental point is the same: ML technologies are capable of finding solutions that could not normally be devised (or sometimes even understood) by a traditional method. These solutions, while not necessarily optimal, are usually efficient enough. Moreover, and more important in this discussion, current convex optimization techniques rely on relaxation methods to deal with the high dimensionality of the realistic scenarios, i.e., they cannot be applied to the original problems[6], which can only be tackled by ML. Besides that, ML is also a very suitable option for dynamic scenarios, since ML models can find efficient solutions even with small changes in the problem where conventional approaches may need a new execution.

ML algorithms learn from data through pattern recognition and can then make predictions accordingly. Supervised learning is where example inputs and desired outputs are given, and the goal is to learn a rule that maps inputs to outputs. In unsupervised learning, no examples are given to the learning algorithm and the goal is to find structure or hidden patterns, as shown in Figure 2.

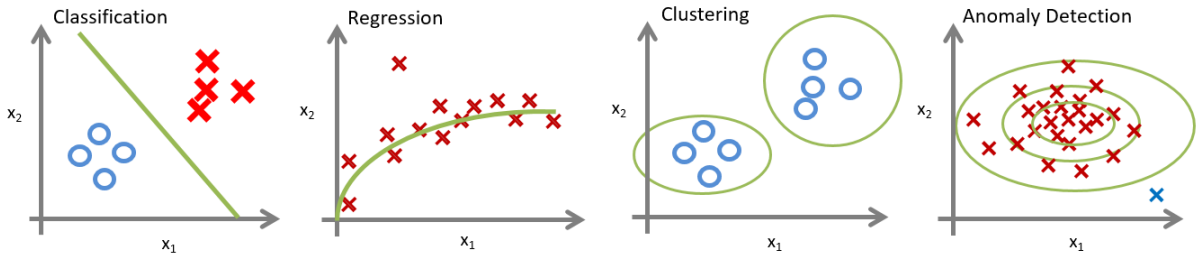

Figure 2: Classification, Regression, Clustering, and Anomaly Detection

With Deep Learning (DL), construction of very deep NN models with many levels is usually needed to meet accuracy requirements. Tens of layers with tens of nodes per layer may be required in autonomous driving, for example. A simple, single-class classifier is shown in Figure 3 and a more complicated multi-class classifier is shown in Figure 4.

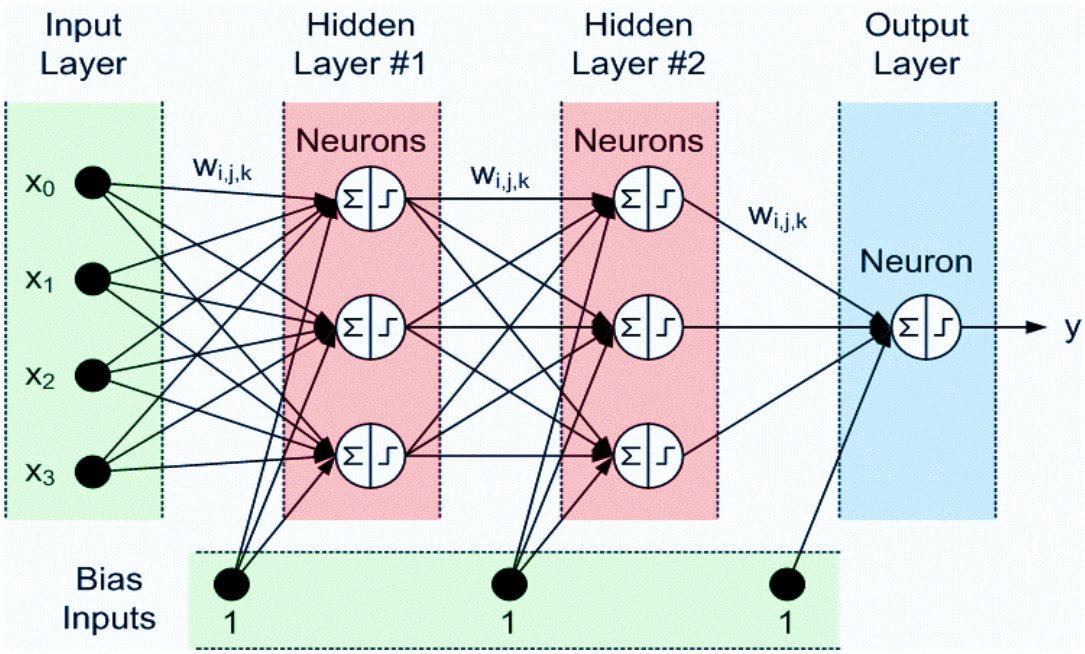

Figure 3: Machine Learning with Neural Networks

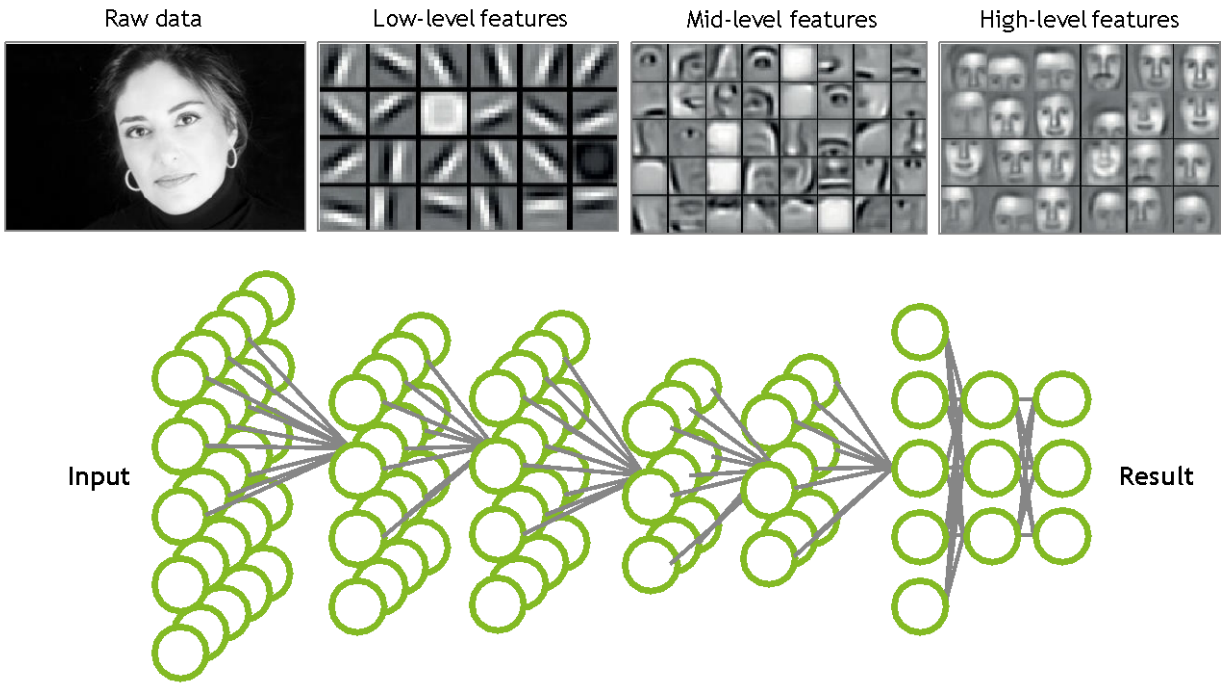

Figure 4: A Multi-Class Deep Neural Network

DL mainly focuses on the hypothesis or models that can be exposed to test data or can learn from input data patterns and arrive at a model to start inferring outcomes with higher and higher accuracy. This capability is very close to what a human can do or, in some cases, even surpasses humans in the speed of computing and/or communicating. It helps the computer to analyze and derive hidden insights without being explicitly programmed to do so. It has multiple uses in today's technology market in the areas of safety and security, such as handwriting recognition, face detection, face recognition, image classification, speech recognition, data science analytics for pandemic, antivirus, search, antispam, genetic, signal diagnosing, and weather forecast.

One type of learning that does not require training data is Reinforcement Learning (RL), illustrated in Figure 5. This type of learning approach has the following characteristics:

- System observes environment and takes action to maximize performance.
- Feedback is provided as reward function.
- System receives a positive reward if it is closer to desired state and a negative reward if it is far from desired state.
- System makes a new observation and the process continues until it reaches the desired state.

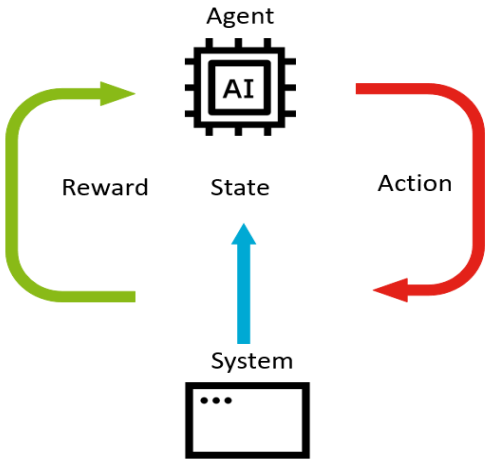

Figure 5: Reinforcement Learning Paradigm

In addition to the methods described above, there are many other types of ML algorithms that may be employed based on the application. Recurrent Neural Networks (RNN) are used when there is a time aspect to the data, for example with speech. Generative Adversarial Networks (GAN) can be used to generate data when there is a lack of training data. AI/ML algorithms and networks are plentiful and over time there will be others that will be beneficial to 5G and future networks.

2.2. Application of Learning to 5G and Future Networks

While the list of domains where AI/ML can be used is growing, this white paper describes some of the potential areas where AI/ML may be used to improve performance and support new applications and opportunities in 5G and Future Networks, which includes:

- Network Automation
- Network Slicing
- Network Digital Twins
- Cybersecurity
- Cloud Computing
- Multi-access Edge Computing
- Dynamic Spectrum Access
- Improve Network Performance and Efficiency

There are a number of international organizations developing standards, frameworks, and technologies related to 5G. Notable examples include:

- ***International Telecommunication Union (ITU)*** has the 5GML working group that developed ML frameworks, such as ITU-T Y.3172 "Architectural framework for machine learning in future networks including IMT-2020" to describe the integration of ML from a system perspective. This organization also created the AI/ML in 5G Challenge in 2020 for technologies related to communication and networking.

- ***European Telecommunications Standards Institute (ETSI)*** has developed technologies related to network function virtualization (NFV) and orchestration that can support advanced AI/ML algorithms and dissemination.
- ***3GPP Specifications and 5G PPP*** develop 5G specifications and provide funding, respectively, for technology prototypes through the European Union's Horizon 2020 and Horizon Europe.
- ***European Union Agency for Cybersecurity (ENISA)*** develops EU security standards, toolboxes and webinars related to 5G cybersecurity. In December 2020, they released the "*AI Cybersecurity Challenges*" and, in February 2010, ENISA published a report entitled "*Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving*".
- ***Open Radio Access and Core Network*** includes development organizations, such as O-RAN Alliance Tech Infra Project, Open RAN, and Open CORE, among others, that are focused on open networks that will eventually include advanced AI/ML algorithms and frameworks.

2.3. AI-NWDAF (Artificial Intelligence Network Data Analytic Function)

AI-NWDAF is a function being developed and standardized for 5G networks.[2] AI-NWDAF uses artificial intelligence and machine learning to analyze data from the network and to provide insights that can be used to improve the performance and efficiency of the network.

AI-NWDAF aims to provide the following benefits:

- Identify and resolve network issues more quickly to improve network performance for users.
- Automate tasks, such as network planning and optimization, to increase the efficiency of network operations.
- Identify and mitigate security threats to enhance network security.

Expected use cases for AI-NWDAF include:

- Analyzing data from the network to identify areas that need improvement; simulating new network configurations to predict performance.
- Analyzing data from the network to identify potential problems before they cause outages; recommending solutions to problems that have already occurred.
- Allocating network resources (e.g., bandwidth, power) more efficiently to improve network performance.
- Identifying and mitigating security threats to 5G networks to protect the network from unauthorized access and attacks.

AI-NWDAF has the potential to be a framework that facilitates the incorporation of AI/ML into the operations of 5G networks to improve network performance, efficiency, and security.

The development of AI/ML technologies for 5G and future networks is actively being undertaken by a number of organizations in addition to the technology roadmap chapters that are being developed by the INGR. This initial AI/ML roadmap is intended to describe some of the exciting future areas of AI/ML for 5G and future networks.

[2] *3GPP TS 29.520 provides a definition for NWDAF; 3GPP TR 23.791 outlines various use cases for NWDAF.*

3. FUTURE STATE

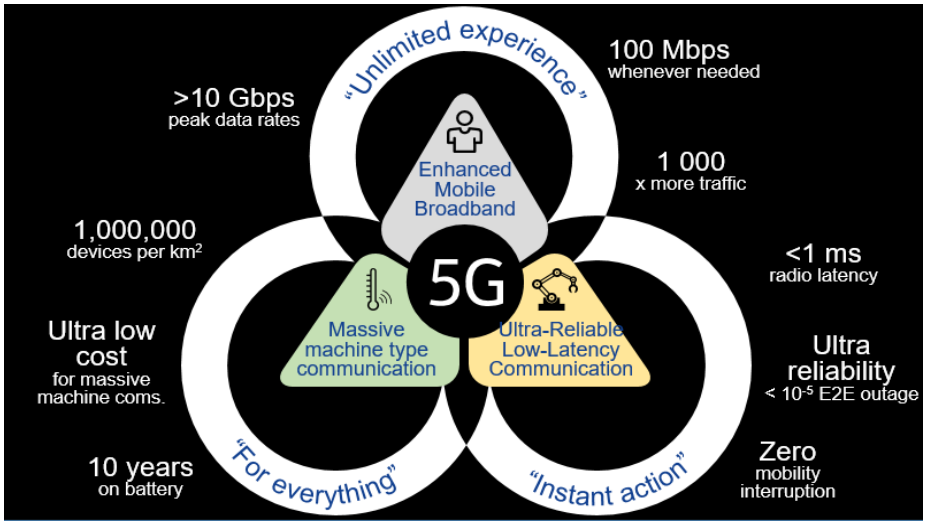

Figure 6: 5G Requirements and Market Verticals

5G brings scale, speed, and low latency while introducing a surge in complexity to support highly heterogeneous traffic carrying diverse and demanding applications for billions of devices, as shown in Figure 6. The use of AI/ML enables:

- Learning from complex patterns
- Holistic and adaptive control
- Proactive and autonomous operation
- New cost structure and revenue

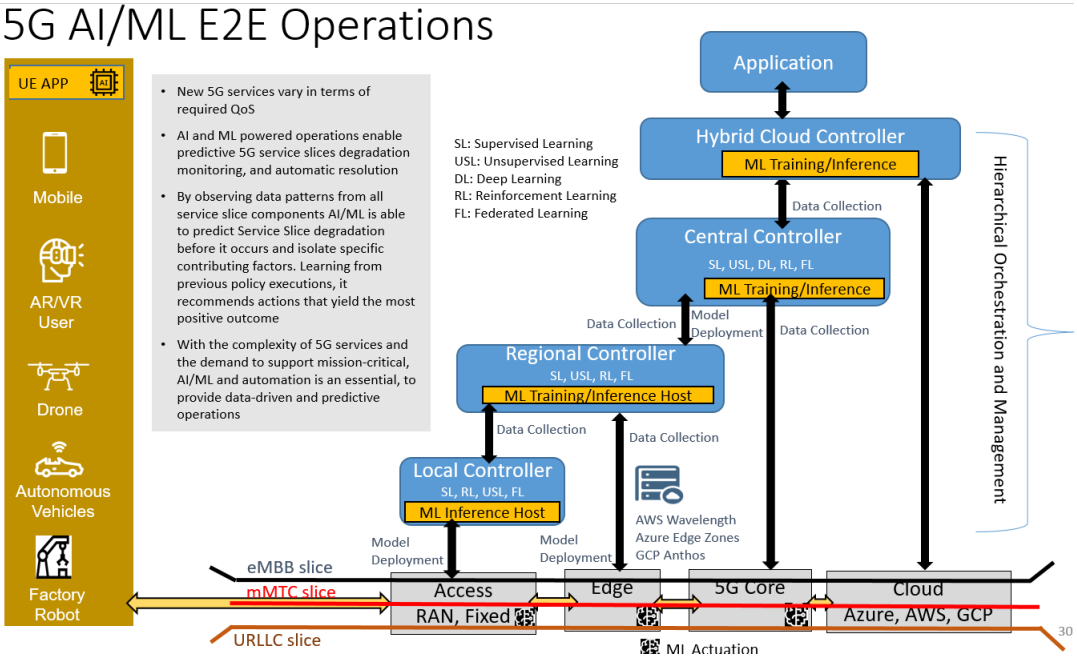

Figure 7: 5G AI/ML E2E Operations

One of the core tasks of the AI/ML working group is to develop an end-to-end (E2E) architecture in which the various components of the network that use AI/ML can integrate and coordinate. This is shown in Figure 7.

3.1. AI/ML for Network Automation

With machine reasoning and machine learning working together, the vision of "intent-based networks" can become a reality.

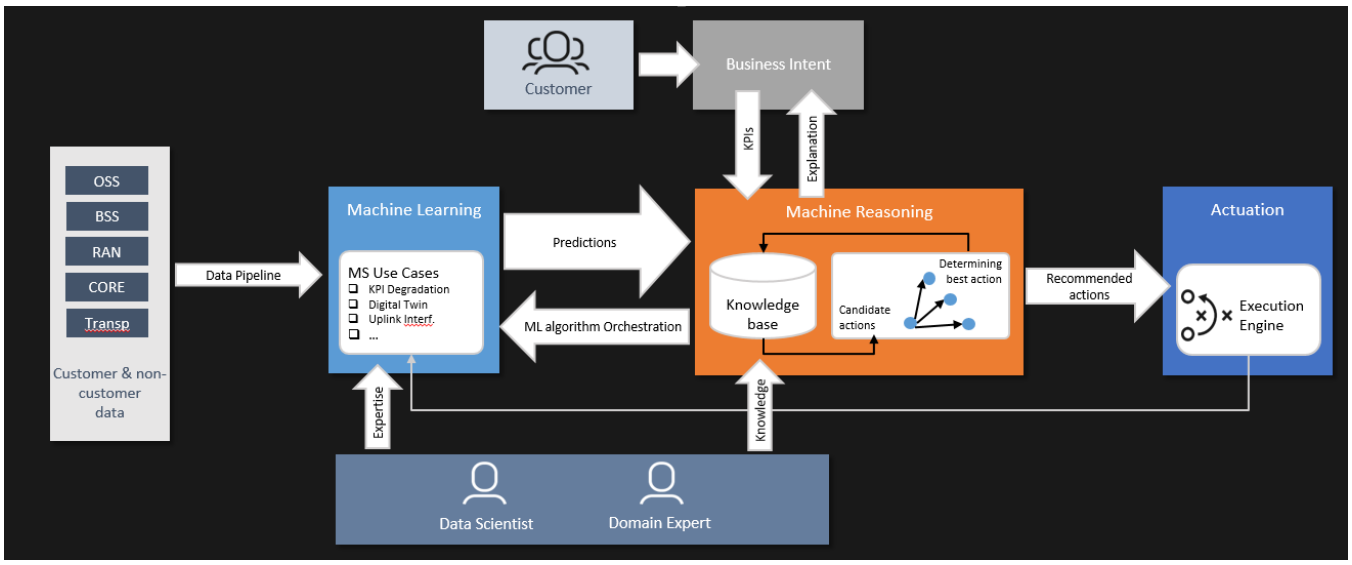

Figure 8: Machine Reasoning and Machine Learning to Realize Vision of Intent-Based Networks

ML relies on using collected data sets and finding patterns in the data. Machine reasoning focuses on understanding the relationships across data and deducing new information. ML systems can learn on their own, but only by recognizing patterns in large datasets and making decisions based on similar situations. ML is dependent on large amounts of data to be able to predict outcomes. If there are few or no structured inputs to extract patterns, ML systems cannot solve a new problem that has no apparent relation to prior knowledge.

Machine reasoning trains on and learns from available data, like ML systems, but tackles new problems with a deductive and inductive reasoning approach (manipulating previously acquired knowledge to answer a new question). To this end, a logical reasoner works on a knowledge base. The reasoner is comprised of logical rules that deduce logical consequences from previous antecedents that ultimately are facts given in an ontological knowledge base.

Business intent specifies the requirement for network slicing in cloud gaming (as an example). One can insert some KPIs and the network will respond if it is feasible. The feedback from the execution engine is evaluated against of a set of KPIs, say from the network operations of the slice (actuation) towards the ML engine, is then used to re-train and improve the models or algorithms.

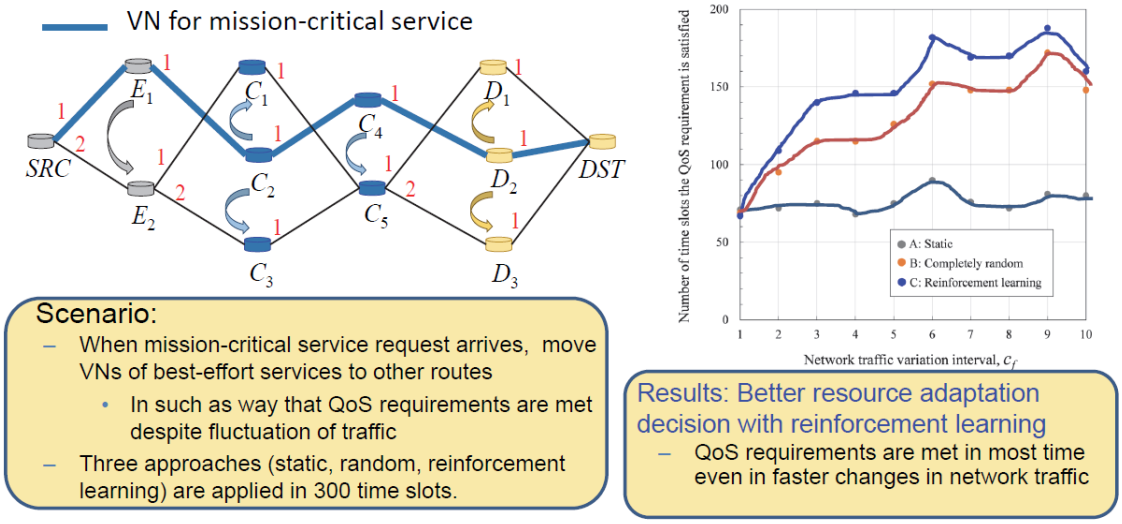

Figure 9: Network Resource Adaptation with Reinforcement Learning

With the development of the network, it is necessary to introduce AI/ML technology to achieve self-adjustment, self-optimization, and self-recovery of the network through the collection of huge amounts of data on the network state and machine learning, as shown in Figure 9.

Network automation is broad in 5G as it covers end-to-end networks. This involves User-to-Network Interface (UNI) and several Network-to-Network Interfaces (NNIs) before it connects to a data network, e.g., "links" or "hops" are Over-the-Air link (OTA) or radio link. The node that follows may be macro cell eNodeB (4G) or gNode (5G) followed by optical termination on Ethernet nodes for baseband processing (BBU) or vBBU. This is at the metro or edge, depending on the provider and the optical rings or the uses of a fixed network for a Wide Area Network (WAN).

To simplify the explanation, we will use standard ETSI 5G system architecture to show that, at the minimum, the network automation will involve interfaces N1, N2, N3, N4, N9, & N6 interfaces besides most of the software modules in the management and control plane. This has release 15 Non-Stand Alone (NSA) mode with 4G LTE eNodeB anchoring the access for packet sessions along with back-end packet gateway for IP access. However, in stand-alone 5G mode with NR, as in release 16, the gNodeB manages the new 5G Radio spectrum bands in sub-6 GHz or millimeter waves of 28 MHz for 100 MB or greater baseband.

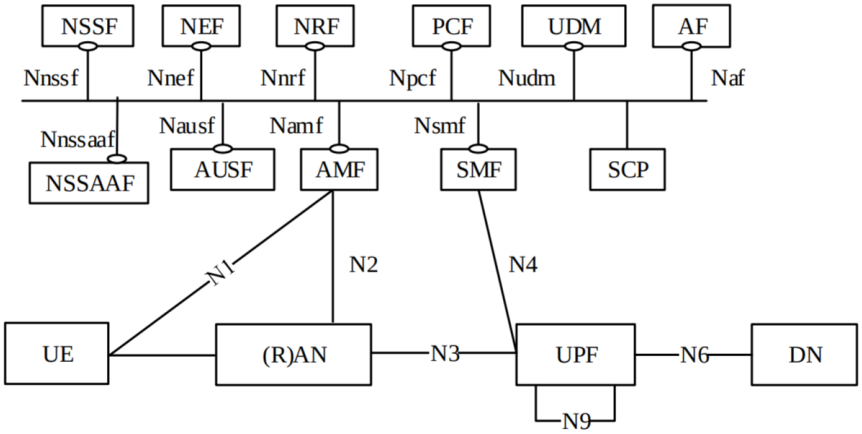

Figure 10: ETSI 5G System Architecture[18]

The carriers have a legacy network and ideally should have started a new radio with a green field network. However, the experience with 3G to 4G transitions helped them to agree to and deploy standards. ETSI Released 15 NSA to use a new radio with Evolved Packet Core (EPC) using eNodeB. Later it was switched to ETSI Release 16 SA. Currently, this automation in carriers is underway with several trials and actual market basis deployments globally. The Covid-19 Pandemic disrupted some of the work, but it is expected to see more and more applications of AI/ML once O-RAN and OpenRAN have been deployed. Significant amounts of AI/ML technology is in use in SU-MIMO and MU-MIMO in carrier aggregation and Sections 4.2 Network Slicing and Section 4.5 Dynamic Spectrum Access cover the use of AI/ML. As we are still scoping the areas to cover in Network Automation in this release, we will limit our discussion to what we have described herein for Network Automation.

3.2. AI/ML for Network Slicing

Under the trend of diversity of services, services with multiple Service Level Agreements (SLAs) should be held in a common underlying network. Thus, the network virtualization techniques are employed in current networks. To efficiently accommodate a wide range of services and ensure that the network can be flexibly tailored for distinct applications, the concept of network slicing has been proposed as a promising approach to fulfill the prominent paradigm shift from one-size-fits-all solution to a "softwarized" and virtualized design[7]. Network slicing offers an effective means to meet the diverse use case requirements and is a critical enabler for realizing industrial applications with stringent latency requirements on a shared infrastructure[8]. Network slices are logically isolated End-to-End (E2E) on virtualized networks operating on a shared physical infrastructure and can be controlled and managed independently to support flexible and efficient service provisioning.

Flexibility and efficiency are realized through the combination of two emerging technologies: Software Defined Networking (SDN) and Network Functions Virtualization (NFV)[9], which decompose traditional monolithic and proprietary network appliances into multiple small and software-based modular network capabilities called Virtual Network Functions (VNFs) running on standard servers. The services are usually modeled as a slice or a virtual network from the service providers and Infrastructure Providers (InP) will respond to the requests. A major challenge is how to allocate network resources more efficiently to several slices, termed slicing or Virtual Network Embedding (VNE), to meet user demands and optimize operator revenue.

3.3. AI/ML for Network Digital Twins

Recently, Digital Twin (DT) is a hot topic. It uses simulation and AI/ML techniques to simulate and evaluate the network status without real deployment. It allows cyber-physical integration by creating virtual replicas of physical objects to simulate their behaviors. DT is used to establish dynamic digital models of physical devices in digital space through monitoring real-time devices parameters, where intelligent comprehensive analysis, digital preview, and feedback control for physical equipment can be realized[10]. It is specialized in achieving low-cost and high-efficiency physical equipment dynamic modelling and controlling[11]. The management of networks can greatly benefit from a network DT. Firstly, the virtual representation of the physical network can be used to conduct various what-if scenarios and resource allocation approaches without affecting the physical network. Secondly, through interactions with the physical network, a network DT can generate and process its own data and predict the QoS performance after any configuration changes. A DT of network is important to achieve cost-efficient and performance-optimal slicing management and keep monitoring the performance under a

diverse set of operating conditions without impacting the physical network. However, unlike a DT in the industrial domain, a network DT needs to merge both the physical and virtual components in networks.

3.4. AI/ML for Security

The 5G system architecture is much more challenging from a security perspective than previous mobile systems. The system architecture is based upon a complex mix of different types of networks, e.g., open and cloud architectures, network slices, Internet of Things (IoT), Core Network (CN), and more. An example is illustrated in Figure 11, where the 5G security cloud that is meant to demonstrate the interconnections of the many elements of the 5G system. Airplanes, smart cars, mobile phones, and others are all interconnected; thus, they are all subject to attack. Furthermore, the architecture must consider current 5G market verticals that include ultra-reliable low-latency communications (uRLLC), enhanced mobile broadband (eMBB), massive machine type communications (mMTC), while also being extensible to future market verticals. As such, security must not be addressed from a piecemeal perspective; it should be looked at more from cloud perspective.

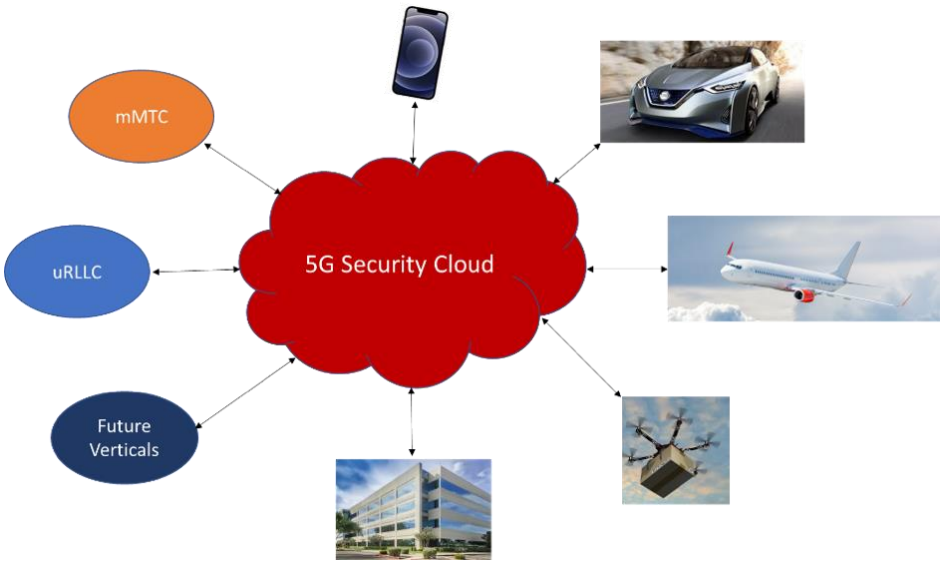

Figure 11: 5G Security Cloud

From a security perspective, attackers have a wide variety of targets and tools that they can use. In fact, the hacking process has been automated using software scripts that are available online. In many cases, hackers require only remedial knowledge to undertake an attack. With the addition of professional and government-sponsored hackers, network security of the future becomes an even greater challenge. To counter this, many companies provide tools and patches or updates that can be used to prevent known attack methods. MITRE, for example, provides a database of known attacks and resolutions[12]. However, as attackers become more and more sophisticated, newer methods will be needed to autonomously detect and mitigate network attacks rather than waiting weeks or months for a software patch to be applied.

As future networks become more complex, security challenges will become even greater. This is especially true when considering zero-day attacks that occur without prior knowledge or warning. Consequently, future networks will need security technology that can predict and mitigate such attacks.

AI and ML have shown the ability to detect and learn different types of features, such as imaging, Natural Language Processing (NLP), communication channel allocation, and more. This type of technology has great potential to address the security needs of 5G and future networks.

3.5. AI/ML for Dynamic Spectrum Access

AI/ML will play a critical role in enabling the operation of future Dynamic Spectrum Access (DSA) and spectrum sharing radio networks. As defined by the IEEE 1900.1 Standard for Definition and Concepts for DSA[13], a DSA radio network uses cognitive processing and knowledge of its RF environment to opportunistically access underutilized spectrum according to spectrum regulatory policy. As shown in Figure 12, the cognitive processing and control functions of a typical DSA radio are based on awareness of the RF spectrum environment, awareness of self and other networked nodes, a machine-readable set of operating goals for the network, and a machine-readable policy, or set of rules, for autonomously accessing spectrum. Furthermore, an intelligent DSA radio network uses AI/ML technologies to classify, predict, and experientially learn how to reconfigure operating behaviors and configurations to achieve network operating goals for a perceived environmental state. This concept of intelligent control builds on the described reinforcement learning agent architecture shown in Figure 5.

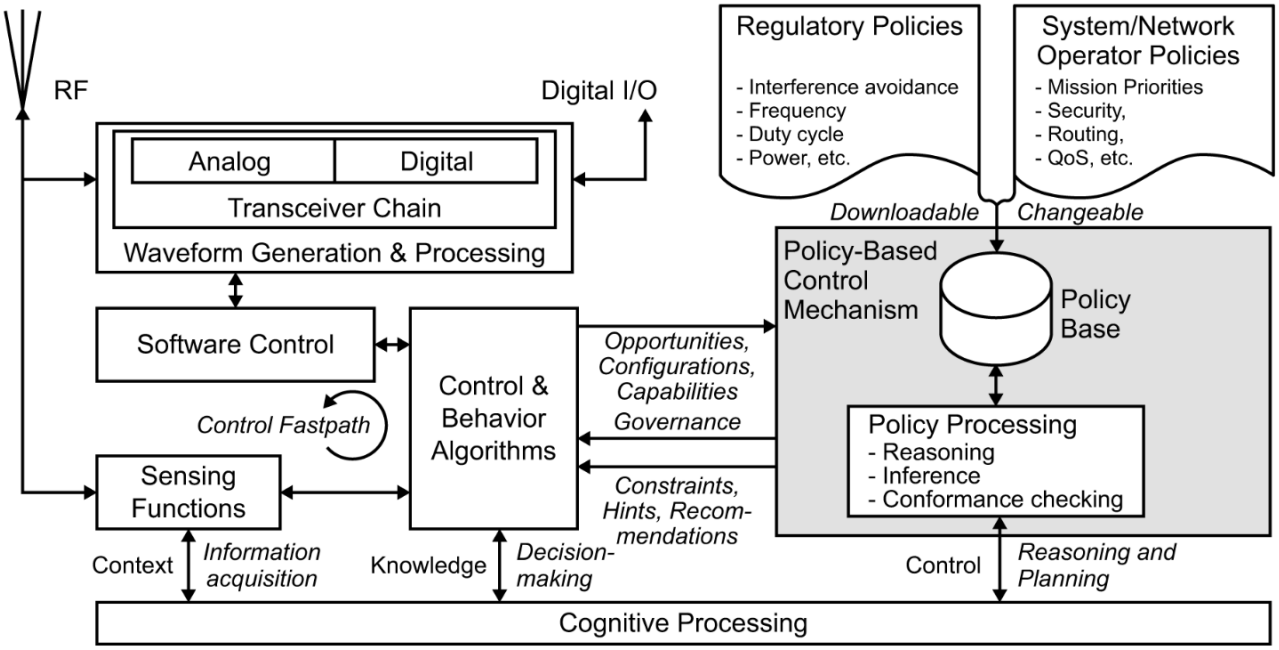

Figure 12: Architecture of a Dynamic Spectrum Access (DSA) Radio Node with Cognitive Processing [Ref-1900.1]

A recent pathfinder example of the potential for using AI/ML to create intelligent and autonomous DSA radio networks was demonstrated in DARPA's Spectrum Collaboration Challenge that concluded in 2019[14]. The competition encouraged teams to imbue their cognitive radio networking software with AI/ML capabilities to explore how future radio networks can autonomously operate in highly congested RF spectrum. The competition culminated in a final event in the third year of the program where five intelligent DSA radio networks with 10 nodes per network demonstrated that they could dynamically reconfigure their operating behavior to find opportunities to share spectrum as they were moved through an emulated radio network[15].

Another critical application of AI/ML for intelligent future radio networks is extracting enhanced RF spectrum awareness from sensed RF data. In this case, the sensed RF data can be raw in-phase / quadrature data, Received Signal Strength Indicator (RRSI), or time-frequency representation of the spectrum data as spectrograms. Once the spectrum data is obtained, trained ML models can be used to detect, classify, characterize, and/or identify radio signals and their signal emitters. This enhanced spectrum awareness can then be used to make decisions about how to reconfigure the intelligent DSA radio networks to share spectrum with primary incumbent spectrum dependent systems, such as military radars (e.g., Citizens Broadband Radio Service), as well as detecting, characterizing, and identifying malfunctioning and misconfigured radio devices.

3.6. AI/ML for Cloud Computing

Cloud Computing (CC) is considered the basis for providing unlimited resources to any digital application. AI/ML needs all kinds of resources in the cloud and is therefore a user of cloud resources. On the other hand, AI/ML has advanced so much in the last decade that, in many instances, the common practice of using "simple scripting" to install server nodes and clusters for the life cycle management of the infrastructure have started using AI/ML. The wheel of fortune has turned around and the discussion here is regarding the use of AI/ML for cloud computing.

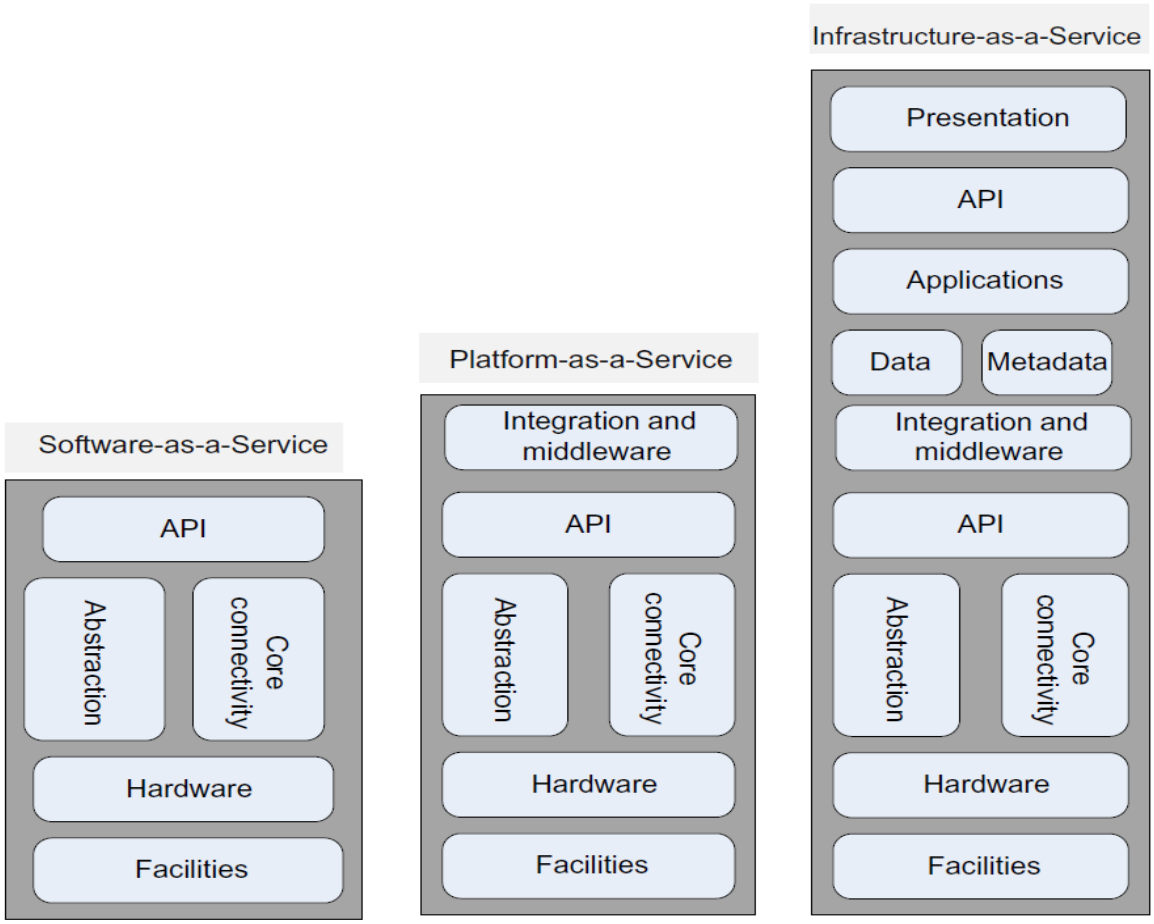

Figure 13: Cloud Delivery Models[16]

Cloud essentially connects massive data centers through private and public networks or the internet. The "network is computing decade" with client-server Application Service Provider (ASP) model, was overtaken by Amazon, Google, and Azure through standardization and offering on-demand managed servers and access through private Virtual Private Networks (VPN) to enterprise, government, and consumers. Cloud computing, with its massive data centers and core network connectivity along with access gateways, has always sought to reduce costs and serve consumers and enterprises by leveraging digital technology.

Technically speaking, the CC stack(s), with its three layers being Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS), as shown in Figure 13, are the cloud delivery model. The RESTful API level is now focused on objects (infrastructure resources, platform services, and domain-specific software) and functions as microservices to build, deploy, and manage them.

AI/ML mainly enhances the value[17] to cloud by enabling:

1. Adding value to self-managing the cloud using AI/ML services
    a. Streamlining the observability of workloads as cloud providers acquire and release pools for resources
    b. Enabling programmable infrastructure (network, compute, and storage)
    c. Flexible on-demand workload distribution across the network cloud
    d. Separation of concerns between undercloud and overcloud with dynamic policies & rules
2. Improving network utilization and reliability for users and tenants with dynamic policy to improve QoS and QoE
    a. Enabling multi-tenant network through role-based access control with quota allocations
    b. Enhancing NS with view to local and global network anomalies and the state of network through AI/ML
3. Adding value to data management, regulations, privacy, and security through AI/ML
    a. Defending against internal and external breaches protecting security and data privacy of users and groups
    b. Driving end-to-end orchestration and automation through regulatory rules for large e-commerce running over hybrid clouds, such as banking networks, stock markets, and logistics (airlines, railways, shipping)
4. Integrating multi-services to provide domain specific portals using AI/ML tools
    a. Examples are travel portal with airlines, hotel booking, railway and road transport, or Uber / Lyft last-mile additions with all possible global currency integrations
    b. Hospital portals to support online scheduling of appointments and rescheduling based on emergencies and travel and vagaries of weather and door delivery of prescriptions to patients based on AI/ML as drugs and vaccines evolve and support vaccinations

There are hundreds of use cases of cloud that lend themselves to the use of AI/ML and all major analysts (i.e. Gartner, IDC and others) predict multiplication by 3x to 5x of cloud revenues attaining trillions of US dollars by 2024.

3.7. AI/ML for Multi-access Edge Computing

Computing aware networks will introduce the computing resource and network resource together, which makes the problem complicated. It may need AI/ML to accelerate the optimization process. In particular, MEC enables:

- Processing closer to UE
- Ubiquitous thin clients
- Lower latency services
- Local breakout
- Reduced traffic on core network
- AI/ML processing

3.8. AI/ML for Optical Networks

Recently, with the rapid development of communication technologies (such as 5G, Internet of Things, and cloud computing) and emerging network services (such as VR / AR and 4K video), the data traffic in communication networks has been growing exponentially. It is reported that the global IP traffic will increase threefold from 2017 to 2022 and the number of devices connected to IP networks will be more than three times the global population by 2022. In response to the exponential growth of network traffic and increasing network operations complexity, it is imperative to enhance the automation and intelligence of underlying optical networks.

Figure 14 shows an architecture of AI-driven autonomous optical networks[18]. The AI module should obtain the network states to perform intelligence. The data processing in network controller is responsible for collecting the network states and preprocessing this state information to support the AI-driven modules for analysis and decision. The raw data that describe the network states, which may include traffic data, network topology, link status, and resource status, are collected from the underlying optical networks in data plane with telemetry techniques from southbound interface (SBI). The control plane receives these data and uploads to the data processing modules in perception function for data preprocessing. The data aggregation module deals with heterogeneous data from different vendors and ISPs with data alignment to obtain a uniform data. Then the data are desensitized in data desensitization module to leave out or mask private information for privacy protection. The desensitized data are labeled under specific rules and features are extracted with a feature engineering module. The features and corresponding labels are used as structural data for AI algorithm training. In addition, a database should be adopted to store the raw data from networks and instances for training.

The self-aware functions are realized through AI based modules to predict and estimate network parameters. The self-aware of network status includes AI-based network traffic prediction, AI-based optical performance monitoring, and AI-based quality of transmission (QoT) and estimation of light paths[19],[20]. Self-adaptive network control changes the traditional static management mode of the network and enables the operation of the network to adapt to changes in the network state through AI. AI-based self-adaptive includes AI-based nonlinear compensation of optical signals, AI-based EDFA control, and AI-based optical network resource allocation. The application of AI in network management mainly focuses on automatic fault management, including AI-based fault prediction, fault detection, and fault location. One of the typical examples is failure management, which includes failure

prediction, failure detection, and failure localization. The AI-based failure prediction module receives the information on network signals, links, and devices from perception functions and further predicts whether there will be failures in network with AI models or not. In the AI-based failure detection, classification models are used to identify the classes of failures that have already occurred. The AI-based failure localization faces the challenges that the large number of alarms appear in current optical networks and the true failure is not explicit. The AI methods are used to localize the real failures with the information from alarms. The predicted failures, failure type, and true failure localization are output to the failure management module in controller functions.

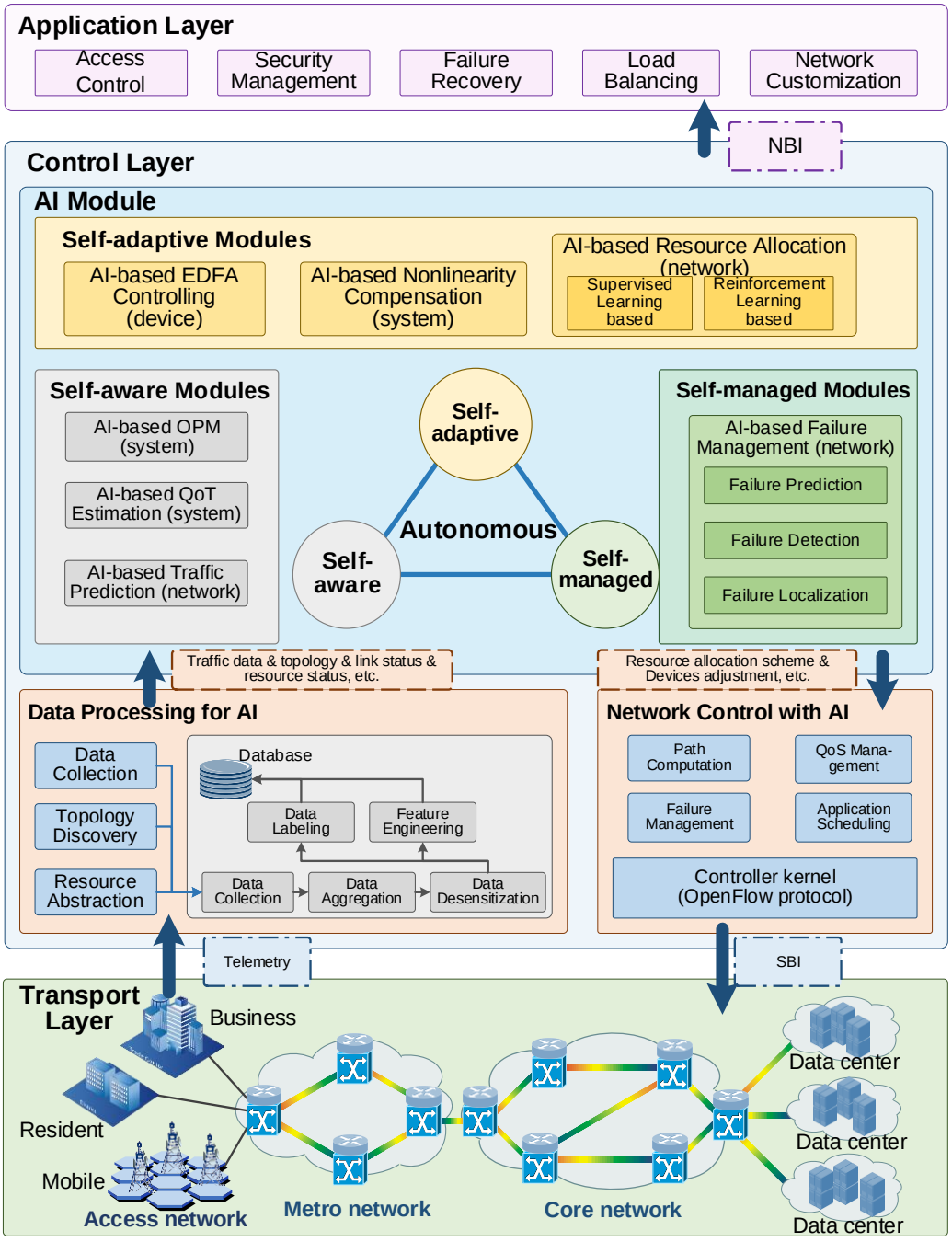

Figure 14: Reference Architecture of AI-Driven Autonomous Optical Networks[18]

3.9. AI/ML for Systems Analytics

Data analytics has emerged as a predominant methodology for understanding various natural phenomena and engineered systems. However, what we actually need are "Systems Analytics" not only "Data Analytics"[21], [22]. Actually, data analytics, per se, are not sufficient to address the challenges of the complex and dynamic systems that we increasingly deal with today and will do even more in the future.

Automation has been in play for a while; a notable early example being NASA's landing mission to the moon on July 20, 1969. Over time, increased computing capabilities through heterogeneous and distributed platforms gave rise to the notion of autonomic computing. Advances in networking and communications with ever more complex applications exploiting these advanced and heterogeneous computing and networking infrastructures. The need to manage such complex resources gave rise to the idea of autonomic computing and the more general term "autonomic" operations for many other engineered components and systems.

In tandem with autonomic notions, the 50-year-old concept of artificial intelligence (AI) reemerged, in various implementations, such as deep learning, and became conflated with Machine Learning (ML), a predominant method for data analytics. ML or DL and data analytics are successful for some applications – typically those dealing with "longitudinal analyses" that allow training of the ML parameters on a set of historical data and then apply the derived ML model to similar classes of data.

However, for systems that manifest complex and dynamic behaviors and when we need to address systems-of-systems (systems interoperating with other systems), ML and data analytics are not sufficient to address the ensuing complexities and dynamic behaviors encountered in such environments. There are many examples, such as electrical power grids comprising of a mix of resources (renewables, subject to changing environmental conditions) and operation demands by multiple classes of consumers with multiple levels of priorities as well as subject to disruptive conditions due to adverse weather or power grid component failures. ML methods alone cannot adequately represent such complex and dynamic conditions; nor can they provide the required real-time decision making support. The advent of the xG series of networks, the recent 5G, and the foreseen B5G present unprecedented capabilities but also challenges to address their multi-components and multilevel heterogeneity, dynamic resource availability and demand, and the needs to deliver QoS in energy efficient ways. Such complex dynamic systems present end-to-end and systems-of-systems support needs, which require more powerful methodologies than ML alone.

Methods such as DDDAS[21] provide the powerful methodology needed: comprehensive models and dynamic data inputs utilizing ML as an auxiliary tool to adequately represent the behaviors and manage adaptively the dynamic operational conditions. Moreover, DDDAS methods are not subject to the challenges of explainability, transparency, and interpretability manifested in AI/ML when the ML algorithm / model changes on the basis on new or additional training data. Namely, there are pitfalls in the reliability of the ML model as the model / algorithm evolves through training. Explainability refers to where the ML is treated as a black box and the ML-model outputs do not always correspond to a linear mapping of the inputs. Transparency is where it may not be fully understood how the ML-based model / algorithm changes itself based on training data. Interpretability is where it may not be fully understood whether or how adequately and accurately the changed model / algorithm still represents the system and whether the inferences for decision-making are still valid. It is not certain whether the self-modifying ML-based model / algorithm uses the appropriate data and/or whether a set of skewed data creates an inadequate or inaccurate algorithm, and thus the ML model / algorithm can go "rogue"[21].

These are challenges with ensuing implications on adequacy, safety, and security when using that ML-model / algorithm[22].

DDDAS differs from ML and is superior to ML in representing a system because DDDAS uses comprehensive models representing the characteristics and behaviors of system. DDDAS is not just a fitting method, it has been demonstrated that it can create more accurate models (representations of the system), more efficient models of systems, and "decision support systems with the accuracy of full-scale models". Endowing a DDDAS-based model with the additional data can be considered a learning mode for the model. However, because the DDDAS-based model is cognizant of the characteristics of the system it represents, e.g., is physics-based – cognizant of the physics of the natural or engineered system, and these physics constraints safeguard the DDDAS-based modeling from going rogue. In ML, the physics (characteristics of the system at hand) is not explicitly represented[22].

3.10. AI/ML for RAN

Operators need better efficiency, flexibility, security, and operating costs when operating large networks like 5G. The cost of guaranteeing QoS for different business slices, like eMBB, URLLC, and mMT, is high for diverse application scenarios.

AI can be applied to recommend solutions, upsell opportunities associated with particular pain points the customer may be experiencing, and suggest upgrade services. If the customer were experiencing bandwidth issues in their home location, the operator could upsell a different hardware or network slice to the customer to resolve this problem.

Proactive network repair based on AI improves customer experience. High-volume network issues are typically monitored with alarms and triggers for repair. These issues often mask minor long-tail network issues that impact users.

Artificial intelligence (AI) and machine learning (ML) are being used to automate and optimize Radio Access Network (RAN) operations. AI/ML can be used for a variety of tasks, including:

- In network planning and optimization, AI/ML can analyze data from existing networks to identify improvement areas. It can also simulate new network configurations to predict their performance.
- In fault detection and resolution, AI/ML can analyze data from RAN sensors to identify potential problems before they cause outages. It can also recommend solutions to problems that have .
- In resource allocation, AI/ML can allocate network resources, such as bandwidth and power, more efficiently, which can improve the network performance for all users.
- In security, AI/ML can be used to identify and mitigate security threats to RAN networks, which can help protect the network from unauthorized access and attacks.

AI/ML is a powerful tool for improving RAN network performance, efficiency, and security. Here are some specific examples of how AI/ML is being used in RAN:

- Ericsson is using AI/ML to automate the deployment and configuration of RAN equipment, reducing the time and resources required to deploy new RAN networks.[3]
- Nokia uses AI/ML to predict and prevent network outages, improving RAN network reliability and uptime.[4]
- Qualcomm is using AI/ML to improve the performance of RAN networks for users with high-bandwidth applications, such as video streaming, which provides a better user experience for these applications.[5]

These are just a few examples of how AI/ML is used in RAN. As AI/ML technology develops, more innovative ways to improve RAN networks may be expected.

[3] *Corcoran et al, "AI-enabled RAN Automation," Ericsson Technology Review, October 28, 2021. https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/ai-enabled-ran-automation.*
[4] *Nokia press release, "Nokia Launches Intelligent RAN Operations to Manage the Power of 5G with Machine Learning #MWC22", February 22, 2022. https://www.nokia.com/about-us/news/releases/2022/02/22/nokia-launches-intelligent-ran-operations-to-manage-the-power-of-5g-with-machine-learning-mwc22/.*
[5] *Qualcomm OnQ Blog Post, "How is Qualcomm driving 5G Advanced into new verticals and toward 6G?" February 26, 2023. https://www.qualcomm.com/news/onq/2023/02/mobile-world-congress-2023-qualcomm-demos-5g-advanced-to-6g.*

4. NEEDS, CHALLENGES, ENABLERS, AND POTENTIAL SOLUTIONS

4.1. Networking Slicing

4.1.1. Needs, Challenges, and Potential Solutions

Model Interpretability

In operation and maintenance tasks of optical networks, configuration actions should have clear reasons to be taken, and monitoring results should also be interpretable to operators. However, most AI/ML algorithms work in a black-box mode. Although their training processes are open and transparent, the trained models are uninterpretable. Without the interpretability, it may be difficult for network operators to troubleshoot the problems when network performance is not as good as expected.

There are possible solutions that include:

1. More interpretable models, such as logistic-regression and decision-tree models, should be employed with priority. It is undeniable that these models may be too simple to handle the complex problems in optical networks; however, it is still worthwhile to try these models first to find the tradeoff between interpretability and performance.
2. When the simple interpretable models cannot work well, Explainable Artificial Intelligence (XAI) in AI/ML discipline can be exploited in optical network domain.
3. Seek a balance between rule-based and data-driven methods. In optical networks, there are plenty of explicitly known relationships and expert knowledge, so it is possible to exploit these relationships as built-in rules to construct learning models in a top-down approach, in which the whole structure of problem-solving is based on human knowledge, and AI/ML only acts as submodule, such as value-fitting. With these built-in rules, the interpretability of an AI/ML model will be promoted.

Reality Gap between Simulation and Real Network

Most AI/ML applications in optical networks are demonstrated in a simulation environment or in a small-scale network; very few have been demonstrated in real optical networks. It is difficult to test ML methods in the real network due to security and privacy concerns. However, how many characteristics of a real network can be emulated and whether the model performance in an emulation network is the same as in a real network is difficult to evaluate. These reality gaps may hinder the well-trained ML model under a simulation platform being used directly in real networks.

There are possible solutions that include:

1. Take more characteristics of real networks (both physical parameters and network effects) into consideration when building simulation platforms to narrow the reality gap.
2. Although putting methods into practice is costly, more field tests are still worthwhile to evaluate the proposed methods in real networks.

The needs, challenges, and potential solutions for a working group in the area of ML for network slicing is summarized in Table 1.

Potential Conflict with Net Neutrality

Conceptually, slicing allows fractions of network resources to be reserved or otherwise designated for specific uses by specific users with specified service-level assurances. Net neutrality, on the other hand,

is doctrine to have all network traffic be treated equally. Some aspects of net neutrality have already been codified into laws and regulations (for example, see Article 3[23]). As such, slicing, much like differentiated service before it, is another tool with great potential to conflict with regulations inspired by net neutrality.

The legal landscape surrounding net neutrality is complex and still evolving, with different jurisdictions actively drafting or refining laws and regulations to codify various aspects of net neutrality. 5G slicing is similarly new with business cases still to be explored and actual deployment practices still to be developed. The interaction between slicing and net neutrality is therefore also an area of active study and is expected to be complex and ambiguous for years to come. (For example, read an exposition on the regulatory complexity of the interaction between slicing and net neutrality[24].)

Specific regulations surrounding net neutrality still need time to be established and the interpretations and practical implications need to be clarified through case laws that will take time to accumulate. Nonetheless, to the extent that AI/ML is expected to drive and enhance the deployment of slicing and the deployment of slicing is expected to be complex, safeguards should perhaps be incorporated into the AI/ML facilities that drive or monitor slicing to detect and/or prevent any deployment or provisioning of slices from running afoul of net neutrality regulations.

4.1.2. Roadmap Timeline Chart

Table 1: Network Slicing Needs, Challenges, and Enablers and Potential Solutions

Name	***Current State (2023)***	***3 years (2026)***	***5 years (2028)***	***Future State 10 years (2033)***
Need #1 Guarantee the safety of ML models	Human audit after the ML model	Using explainable models	Using unexplainable DL models with explanation module after each output	Using explainable DL models
Challenge(s) for Need 1	Huge human effort will be needed	The model accuracy is low	Interpretation ability from the ML model aspect	
Possible Solution for Challenge	Using explainable ML models	Ensemble methods, boosting methods	Ad-hoc and post-hoc explanation methods, such as LIME and Sharply method	
Need #2 Make up the gap between simulation and real network	Building domain specific platforms	Digital twins method		
Challenge(s) for Need 2	High cost			
Possible solution for challenge	Digital twins method			
Need #3 Prevent Conflict with Net Neutrality Regulations	Manually set strict, conservative limits (including business rules) on the practice of slicing	Gradual relaxation of the strict limits set for slicing; let AI/ML flag cases that require human review	Depending on clarity of regulations and volume of case laws, increase the latitude allowed for AI/ML to drive slicing	Continue to increase latitude for AI/ML to drive slicing, expect diminishing fraction of cases flagged for human review

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Challenges for Need 3	Net neutrality laws and regulations still evolving; lacking case laws to provide clarifications	Continue to monitor the development of regulations and case laws		
Possible solution for challenges	Manually, selectively curate slicing parameters	Depends on regulations and case law development		

4.2. Network Digital Twins

4.2.1. Needs, Challenges, and Potential Solutions

The large number of elements, devices, and upper services will pose challenges for developing the DT, since they change frequently and can be created or destroyed in real-time. The recent advances in AI/ML provides a promising means to fulfil the demands of network DT development. The emergence of data-driven ML techniques, especially DL, has gained popularity in networking areas and led to a new breed of models that learn from data, instead of being explicitly programmed. Researchers are using DNN to model complex network behaviors and develop decision-making strategies based on Deep Reinforcement Learning. Network DT can be established based on AI/ML methods to discover the complicated relationships and inter-dependency among network slices and elements, resource utilization, and physical infrastructure; and generate the E2E metrics prediction of each slice under diverse scenarios. As shown in Figure 15, multiple network monitoring techniques based on DNN, such as optical performance monitoring and quality of transmission estimation, can provide accurate network parameters to make optical network DT more accurate[25].

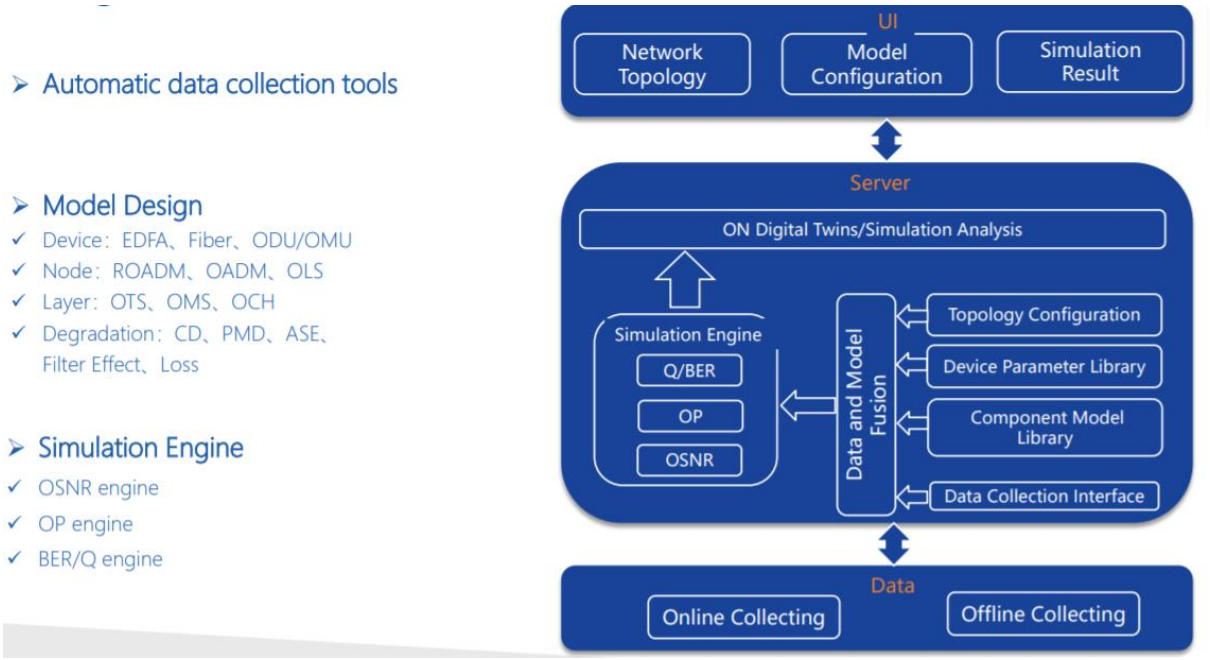

Figure 15: Example Optical Network DT with ML-Based Monitoring Techniques

Model Generalization Ability

Generalization ability is an important metric to evaluate a ML model and it is especially important when applying ML for intelligent optical networks with dynamic network environment. Because in the network problems, the trained ML model is tightly coupled with the network environment, which can only target one network structure or scenario. If the network environment changes, model retraining is necessary, which will be of great computational cost.

There are possible solutions that include:

1. Decouple the input features and network state in the feature engineering process. For example, training the model with features focused on the local network state instead of using features only focused on the total network.
2. Transfer learning-based methods so they can be adopted as powerful tools to enhance the adaptation of the model to different environments and datasets. Adopt algorithms that are capable of online learning, which can fine tune the model itself, instead of retraining when the environment changes.

System Security and Reliability

The adoption of ML will increase the flexibility and automaticity of optical network and reduce the necessity of manual operations. However, ML models often work in a best-effort way and do not provide performance guarantees, which may cause security and reliability issues. Security issues refer to inherent vulnerabilities in ML models and reliability issues refer to performance degradation and errors of trained ML models. The lack of security and reliability guarantees may hinder the practical use of ML in real networks.

There are possible solutions that include:

1. Establish periodical model effectiveness evaluation mechanism and model performance degradation alarm mechanism.
2. In the system design stage, ML-aided mode is suggested instead of ML-dominate mode and interfaces for manual intervention need to be reserved.

4.2.2. Roadmap Timeline Chart

Table 2: Network Digital Twins Needs, Challenges, and Enablers and Potential Solutions

Name	***Current State (2023)***	***3 years (2026)***	***5 years (2028)***	***Future State 10 years (2033)***
Need #1 Good generalization ability when environment varies	Retraining the ML models	Using transfer learning to modify the ML model when the environment changes		
Challenge(s) for Need 1	Huge computational cost			
Possible solution for challenge	Use transfer learning			
Need #2 system security and reliability	Human check	Model-driven framework with ML modules		
Challenge(s) for Need 2	High cost	Not an end-to-end solution; still needs human efforts		
Possible solution for challenge	Automated model monitoring	x		

4.3. Security

4.3.1. Needs, Challenges, and Potential Solutions

Figure 16 provides a detailed view of the security pillars across the 5G system. While 5G technologies provide various enablers, namely edge cloud, network function virtualization, software defined networking, network slicing, orchestration, and cloud RAN; these also contribute to additional security threats. While features such as MEC have been present since 4G, it is expected that this technology will become a much more significant part of future networks. In addition, newer technologies, such as cloud RAN and IoT, will add new capabilities; but they also increase the security challenges.

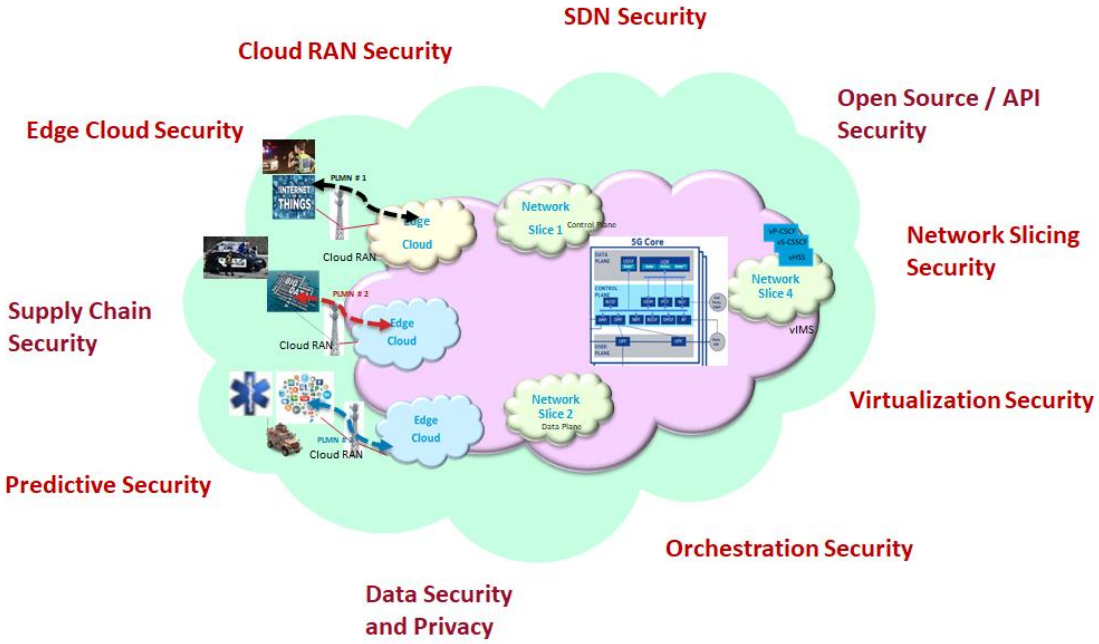

Figure 16: 5G Security Pillars

AI/ML technologies can be applied to mitigate the risks associated with these security risks. We provide a few examples of how AI/ML techniques can be used for this purpose. In addition to AI/ML, security applied at the endpoints is shown in Figure 16, an AI/ML-based security orchestration function is needed to manage the security ecosystem, which consists of the elements shown above. Furthermore, an open system is needed to allow the different components of the system to communicate necessary information and update the AI/ML security defenses in real-time. Below is an example of how AI/ML can be used to provide predictive security in a 5G network and system-level security for 5G networks[20].

Predictive Security

Traditionally, closed-loop-control and orchestration tools are used to detect and mitigate any kind of attacks in the signaling plane and data plane. Placement of these loops in the end-to-end network is important as it determines how quickly the detection and mitigation can be performed. Hence, new techniques augmented by AI/ML algorithms are needed to determine the exact location of these control loops. To detect the attacks proactively, AL / ML algorithms can analyze the behavioral pattern of the data. This way, the zero-day type of attacks can be averted completely. Figure 17 gives an example of a control loop and automation and its applicability in 5G type networks. It shows how a control loop is implemented using a combination of data analytics, orchestrator, SDN controller, and various virtual

security functions, namely virtual Distributed Denial-of-Service (DDoS), virtual Intrusion Detection System (IDS), and virtual Intrusion Prevention System (IPS) system. In case of proactive security, the orchestration and automation can be implemented before the attack takes place. Various AI/ML algorithms can be applied to correlate the past and present data and control traffic and determine the probability of future attacks. By applying the control loop, attacks can be prevented. Thus, AI/ML can help both reactive and predictive security. AI/ML algorithms can be designed to optimize the placement of control loops at different parts of the network based on the severity of the attacks. These control loops can be placed at the edge or core of the network based on the desired KPIs of the system, such as the ability to detect and mitigate the attacks in a relatively expedited manner.

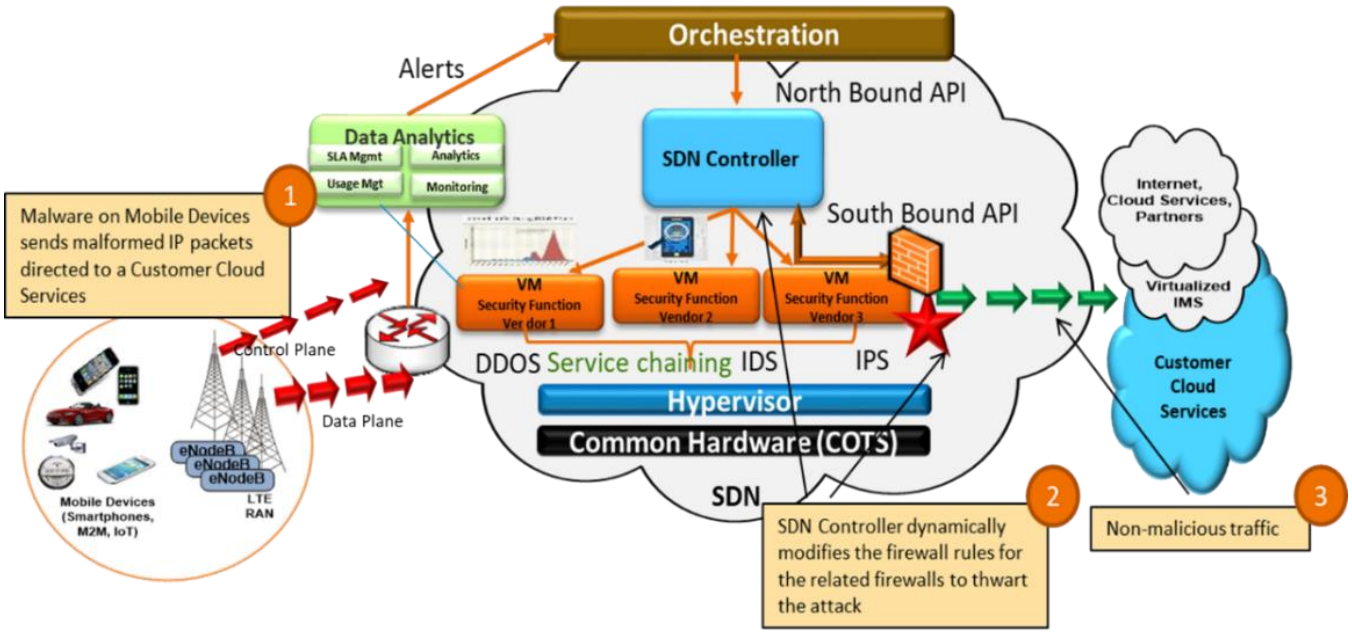

Figure 17: Application of Control Loop Algorithm for Predictive Security[26]

There are many potential opportunities where AI/ML-based security can be a key component of future networks. Both supervised and unsupervised AI/ML will form the backbone of future AI/ML techniques, which may include k-means, Convolutional Neural Networks (CNN), GANs, RL, and others. Integrating these techniques into the 5G architecture will allow much stronger and more resilient cybersecurity methods.

Table 3: Summary of Future 5G AI/ML Security Research Areas

Opportunity	***Description of Need***
Open software and interfaces	A modular, open architecture supports innovation, collaboration, and faster cyber event response times
Development of the security cloud	Like today's networks clouds, security clouds will quickly collect cyber event information, update models, and provide cyber defense
Cooperative and distributed AI/ML for cyber security	Cooperative and distributed AI/ML models and algorithms are needed for accuracy and response time improvements

Opportunity	*Description of Need*
Development of advanced models	Whereas many of the current AI/ML systems are targeted for video processing, Natural Language Processing (NLP) and others; models and systems need to be developed that consider the properties and statistics of cybersecurity data
Integrity of the 5G AI/ML security autonomous ecosystem	It is crucial that autonomous systems are built to recognize when model parameters have been hacked and altered – the AI/ML system must know that a model has been hacked, even if it came from a trusted source given that the source itself may have been hacked

In summary, 5G Security using AI/ML is an open area that will need important contributions in the future, which includes many areas, some of which are included in the Table 3.

4.3.2. Roadmap Timeline Chart

Table 4: Summary of Future 5G AI/ML Security Research Areas

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Need #1 An AI/ML-based security cloud is needed that cooperates with networks components to aid in the implementation of security policy, threat detection, and analysis. The security cloud must be open to allow for cooperative sharing between end users and security orchestration.	Currently, there is no security architecture that describes how to implement an AI/ML into 5G and Beyond systems.	Identify teams and architectural trade-offs. Examine potential technologies.	Define architecture and requirements for a subset of market verticals or use cases.	
Challenge(s) for Need #1	There are many stakeholders and methods to implement AI/ML, though less for security			
Possible Solution for Challenge	Specify requirements and interfaces that allows vendors to plug in solutions.			
Need #2 Open systems and interfaces are such that entities within the cloud can communicate information such as model parameters, threat data, etc.	In order to support many vendors and technologies, and allow for plug-and-play operation, open interfaces are needed for the AI/ML security cloud.	Organize teams to develop standard interfaces and requirements. Xxx Determine approach to develop open-source AI/ML security interface software.	Complete first version of open interface specification and develop initial version of open-source software.	
Challenge(s) for Need #2	Current AI/ML models are standalone and need to be developed to support distributed systems.			

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Possible Solution for Challenge	By defining standard interfaces, vendors can develop modular technologies that can be easily implemented, which accelerates innovation an implementation.			
Need #3 Development of advanced AI/ML models and techniques to support security functions, such as network intrusion detection and prevention, authentication, etc.	While some development has taken place in AI/ML security, models and techniques need to be specific to security.	Complete architectural studies and recommend approaches.	Complete interface requirement and develop standards.	
Challenge(s) for Need #3	Security focused AI/ML need to operate in a distributed and automated environment.			
Possible Solution for Challenge	Develop algorithms and models that are security centric, adaptable and can operate in a cooperative, distributed ecosystem.			
Need #4, AI/ML model design updates and exchanges must be secure and must detect “bad” models, even if they come from a trusted source.	AI/ML models and parameters need to be secure in both operation and updatability.	Specify approaches to secure the models and provide recommendations for self-recognition of the attack on the model.	Standardize security requirements for AI/ML systems and performance requirements related to self-recognition.	
Challenge(s) for Need #4	AI/ML systems, models, and parameters will be subject to attack. In addition, these systems must recognize when their updates have been compromised.			
Possible Solution for Challenge	Utilize traditional security technologies such as public key encryption in combination with technologies that can self-recognize compromised models.			

4.4. Dynamic Spectrum Access

4.4.1. Needs, Challenges, and Potential Solutions

Dynamic spectrum access technologies enable radio networks to autonomously find opportunities to share spectrum and increase their spectrum efficiency while meeting the network performance goals of end users in congested RF spectrum. DSA-enhanced radio networks benefit from enhanced environmental awareness and decision intelligence offered by AI/ML technologies. For example, an intelligent DSA radio network can use a ML model to detect and classify the emissions of nearby high priority RF transmitters. This observed information can then be used to make informed decisions about whether it should attempt to share a frequency channel with the higher priority RF system or not. An intelligent DSA radio network may also use reinforcement learning to experientially learn how to improve its operation by reconfiguring the radio for a perceived environmental state and maximizing its long-term networking goal function. AI/ML-based DSA radio technology enables dynamic spectrum sharing with non-DSA RF systems, such as radars and legacy radio networks. Figure 18 shows how the following functions of an AI/ML-enhanced DSA radio maps to radio network components:

- Collect RF spectrum data and network key performance indicators and goals
- Classify and characterize RF spectrum data and network state
- Plan for how to achieve network operating goals
- Adapt and reconfigure the radio node and network in order to achieve goals

The mapping of these DSA radio network functions to radio network components is important because it shows where interfaces need to be developed in future networks. DSA component interfaces and functions would benefit from standardization so that components from different vendors interoperate and the customer can avoid being locked into buying DSA components from a single vendor. An example standard that describes the interfaces between the components of a policy-based DSA radio system is the IEEE 1900.5-2020 standard for Policy Language for Dynamic Spectrum Access Systems[6].

[6] "IEEE STANDARD FOR POLICY LANGUAGE FOR DYNAMIC SPECTRUM ACCESS SYSTEMS," IN 1900.5.1-2020 , VOL., NO., PP.1-204, 26 FEB. 2021, DOI: 10.1109/IEEESTD.2021.9366651.

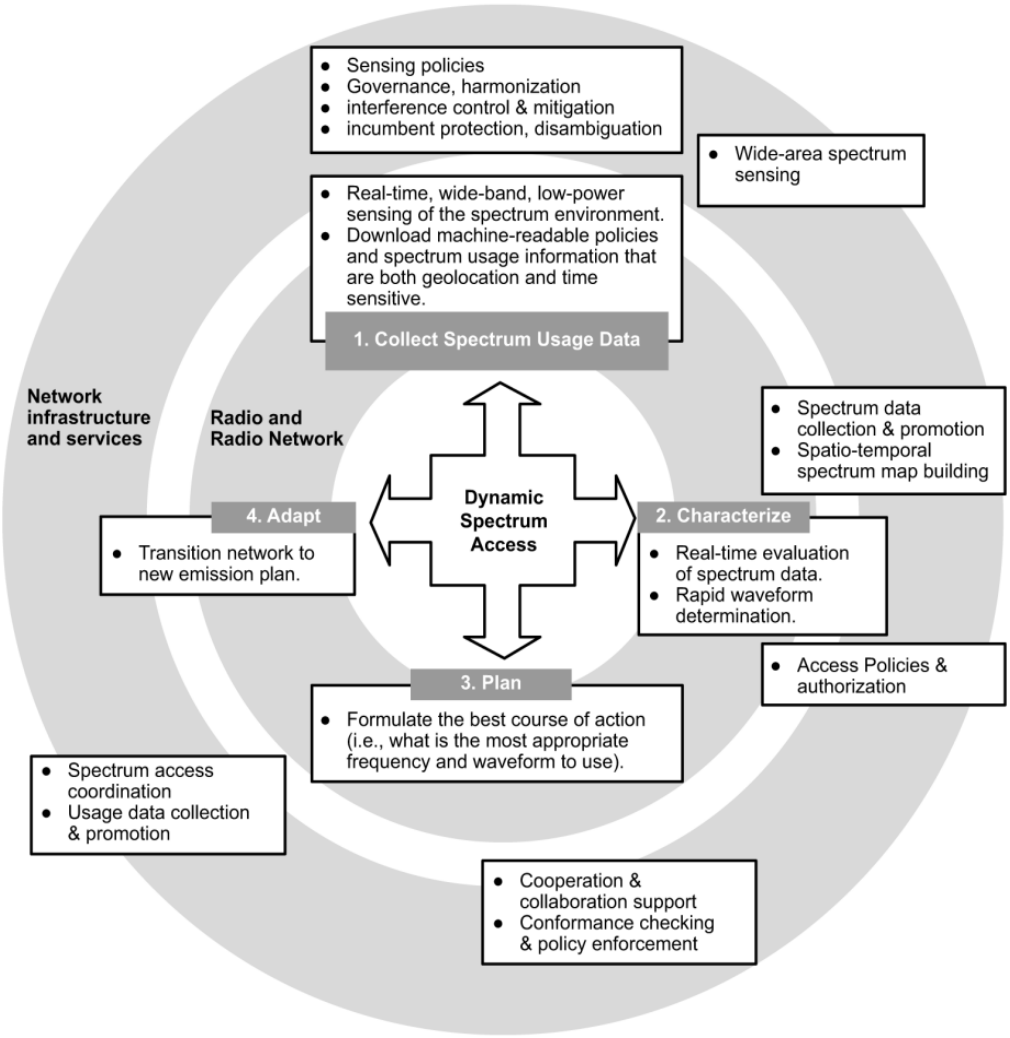

Figure 18: Key Functions of an Intelligent DSA Radio Mapped to Network Infrastructure Components

A key future growth area for applying ML to intelligent DSA radio networks is enhanced RF spectrum awareness. In the case of a spectrum-sharing scenario with an incumbent RF system that has priority access to the spectrum, future AI/ML-enhanced DSA radio networks must be able to detect the presence of the incumbent system to decide how to avoid degrading performance due to excessive RF interference. It is challenging to develop a robust incumbent waveform detector because the incumbent RF signals are often obscured by other overlapping RF signals and the sensed signal data often exhibits unknown RF multipath fading effects.

Another impediment to advancing the state-of-the-art in RF Machine Learning (RFML) spectrum awareness models is the lack of a standard approach and schema for generating high-quality ML model training and testing datasets. Existing RFML datasets suffer from data veracity issues and insufficient sample sizes to develop ML models that meet the performance requirements of real-world operating scenarios. These datasets also have insufficient representation of the variability of realistic RF environmental effects that include a range of RF mobile multipath fading channels and interference signals, greatly limiting the performance of RFML models once deployed in the real-world. RFML

model developers are forced to create their own RFML dataset generation approach and format for datasets, significantly increasing the risk of incurring unexpected costs, encountering hidden technical challenges, and creating models that are unable to meet the performance requirements of real-world operating scenarios. The IEEE Dynamic Spectrum Access Networks Standards Committee (DySPAN-SC) recently approved a Project Authorization Request (PAR) to develop a new IEEE standard for software interfaces between the pipeline stages for training a RFML model that extract spectrum awareness from RF data *(refer to Section 5 on Standards for more information).*

4.4.2. Roadmap Timeline Chart

Table 5: Dynamic Spectrum Access Needs, Challenges, Enablers, and Potential Solutions

Name	***Current State (2022)***	***3 years (2025)***	***5 years (2027)***	***Future State 10 years (2032)***
Need #1: Development of advanced Reinforcement Learning (RL) agent technologies that autonomously control radio networks with outcomes that meet operator / user goals.	Cognitive radio researchers independently develop bespoke RL agents that are non-standard and have limited compatibility with radios from different vendors.	Insufficient trust in how RL agents learn to behave.	RL agents are insufficiently constrained by the operating intent of radio network operator / users.	RL agents that can autonomously modify their own design and goal functions based on user / operator intentions.
Challenge(s) for Need #1	No existing standard for interfacing RL agents and radios. Vendor radio architectures and interfaces vary widely and requires additional software development work to connect RL agent software.	Limited explainability for why RL agent decides to take a particular action.	RL agent algorithms are designed to maximize the long-term reward of taking a radio control action for a given state of the environment - leading to potentially undesirable outcomes (e.g., Skynet).	
Possible solution for challenge	IEEE creates a new standard for interfacing RL agents to a variety of radio architectures.	New RL agent designs that emphasize explainability and traceability of their decision making to their input data and goal functions.	Introduce human-readable control policy into the behavioral control of DSA agents so that they are unable to learn undesired behaviors.	
Need #2: Development of advanced AI/ML models and techniques to support enhanced RF Machine Learning (RFML) spectrum awareness in the context of DSA and spectrum sharing.	RF spectrum sensing and monitoring system vendors independently develop bespoke RFML model development datasets and pipelines. The ML models are trained to detect, classify, and identify RF emitters under a limited set of circumstances.	RF sensor system vendors develop RFML models automatically collect new RF datasets as needed to improve their spectrum awareness.	Robust and resilient anti-spoofing RFML algorithms that mitigate zero-day adversarial learning attacks.	

Name	*Current State (2022)*	*3 years (2025)*	*5 years (2027)*	*Future State 10 years (2032)*
Challenge(s) for Need #2	Lack of common RFML spectrum awareness datasets prevents the community from advancing the state-of-the-art in RFML spectrum awareness model performance.			
Possible Solution for Challenge	IEEE P1900.8 WG will publish a standard on RFML spectrum awareness in the next couple of years. This IEEE standard will lead to the creation of openly shared RFML datasets and ML model development pipeline software.			

4.5. Cloud Computing

4.5.1. Needs, Challenges, and Potential Solutions

AI/ML has helped manage costs through data collection, processing of parameters, and resource usage outcomes via big data and an analytics approach to provide online metrics for cost models. Costs in cloud include location real estate cost, server infrastructure amortized over its projected lifetime, power consumption costs to keep them running reliably, energy costs of cooling, and skills and talents to manage the operations. The self-service model, server or service-on-demand moved from a co-location model to managed services and on to pay-per-use model. All this needed virtualization, which occurred in the period of 1998-2008. This virtualization period was followed by containerization with Docker and then Kubernetes. This built closed-loop automation similar to feedforward and feedback loops of AI/ML models for CNN / DNN and offered a modular approach to build AI/ML applications.

Containerized AI/ML modules have helped cloud to become more predictable and reliable with the ability to predict or locate anomalies in cloud networks and replace them before or just in time through live migration to improve reliability of cloud computing.

The early adopters of cloud were digital goods starting with books from Amazon. The industry never looked back with automation and AI/ML fed the frenzy with Yahoo, Google, Alibaba, and Facebook, along with other social networks providing data to promote retail recommendations based on various intelligence models. The advertising model to opt-in users has had the biggest impact for a decade because of the almost free technology for executing digital transactions and organizing logistics fulfillment via fast courier operators.

The first real users of AI/ML in cloud turned out to be e-commerce retailers. All of the brick-and-mortar players were forced to move to cloud because transaction costs were reduced to cents and supply chains got disrupted. More algorithms and models got into Wall Street and various marketplaces, including

commodity, trade, banking, and insurance. No one can deny that Robin Hood-like AI/ML-based trade can disrupt the hedge funds and shake the establishment.

There is a question of rules, ethics, and good governance and providing transparency while maintaining privacy. The disruption due to AI/ML in cloud is humongous. We saw how a combination of contact tracing with fast mobile tracking has helped isolate Covid-19 people during the pandemic and the ability to do computations with GPU / HPC cloud has helped identify vaccines.

Other big wins for AI/ML were in identity management in large volumes. For instance, India has developed a system called Aadhaar in which more than billion people are being recorded with fingerprint and retina images to recognize the individuals. VISA and MasterCard are working through cloud and crypto capability to allow ecommerce transactions and digital wallets.

It is important to note that cloud and AI/ML have brought human intelligence to computing, with neural nodes being planted all across the cloud, helping to accelerate automation and anomaly detection across all domains, geography, and applications. Cloud and AI/ML have also made cloud ubiquitous, starting with simple page indexing of the web to IoT device tapping for monitoring and delivering control. We will be working on a taxonomy and impact of AI/ML on cloud with quantitative figures on domain-specific use cases.

4.5.2. Roadmap Timeline Chart

Table 6: Cloud Computing Needs, Challenges, Enablers, and Potential Solutions

Name	***Current State (2022)***	***3 years (2025)***	***5 years (2027)***	***Future State 10 years (2032)***
Need #1 Cloud Resource Optimization	Understanding domain and their use cases for AI/ML.	Transfer learning to reuse models for different domains & resource types.	Multi-service Gov. cloud & edge emerging to support Sustainable Development Goals (SDG).	New modular resources appear due to qubits and revival of analogue & HPC.
Challenge(s) for Need 1 Cloud Descriptive Templating	Understanding AI/ML for Domain experts and vice versa.	Customizing modeling tools to improve regional & local delivery for efficiency.	More Global Collaboration & Competition as Political landscape changes, adding to Jurisdiction & governance issues.	Expect a new generation of quantum automation templates & AI/ML models.
Possible Solution for Challenge Use Labels & Annotate to Describe Resource Modules & Granularity	Training and learning tools and methodologies.	Moving from ML to DL for solutions.	Regionalization blocks (EU, NA, SA, Gulf & Middle East, Eurasia, ASEAN, BIMSTEC, China) comes into play instead of globalization or localization. The carriers may merge with edge / cloud providers.	Stronger regionalism will drive fragmentations and improve quality at cost that may go up. Solution will need more customizations using AL / ML.

Name	*Current State (2022)*	*3 years (2025)*	*5 years (2027)*	*Future State 10 years (2032)*
Need #2	Building domain specific platforms like health, education & finance.	Support for NLP & voice activation and translation will be a standard requirement.	New architectures needed for regions with RISC-V will emerge as competition to ARM & x86-64.	Quantum computing will need reimagining vertical solutions.
Challenge(s) for Need 2	Library management with different platforms for AI/ML along with cloud native services.	AI/ML tools and API will need to address NLP and Voice customizations. More dictionaries will be required in the cloud.	The very nature of Instruction Set Architectures (ISA) may need a relook.	Disrupt the existing crypto and domain-specific solutions.
Possible Solution for Challenge	Standardize languages, libraries, and run times	Faster storage & file systems will help alleviate some of the issues along with FPGA & DPUs.	New evolution of Quantum computing will spawn newer semi & optical computing elements.	As scale & speed break the barriers of Terabit, networks will come to provide possible new solutions.
Need #3	Building drivers for integrating emerging devices.	Drivers will need updates as xPUS & DPUS evolve.	Qubit drivers will need Analogue and Mixed Signal designs	Qubit & quantum drivers will need to integrate with hybrid solutions.
Challenge(s) for Need 3	Evolving accelerators drivers for emerging devices.	RDMA, GRPC, NVLink, & ROE will need updates as DPUs bring new functions as a service.	Skill set for analog computing expats will revive.	Digital vs. analogue vs. magnetics in merchant silicon & mems.
Possible Solution for Challenge	Testing accelerators in lab environment.	Harmonizing thread and multi-process will move to DPU managing CPU as well GPUs.	Reuse and updates of older Analog techniques will be possible solutions. AI/ML will switch to older Signals & systems, Control systems theory for solutions.	Magnetics, memes, and graphene may provide some options.
Need #4	Improving smart NICs, NVME, FPGAs, & GPUs for specific domains.	Software development & maintenance for offloading / acceleration will be a baseline.	Thinking serial & programming parallel will need new AIML models.	New media & materials will need newer fabs & facilities.
Challenge(s) for Need 4	Architecture for caching and shared memory for parallel computing.	More skill set and research to design new busses like next gen PCIE / USB etc.	Software to evolve for buffering, queuing, & caching for DPU combined with newer architectures.	Vacuum processing will emerge as the new challenge.
Possible Solution for Challenge	Evaluating right attributes for optimizing outcomes.	Extend PCIE & USB with newer alloys & tracks.		Some learnings going to vacuum tubes may provide pointers.

Name	*Current State (2022)*	*3 years (2025)*	*5 years (2027)*	*Future State 10 years (2032)*
Need #5	Innovation in AI across the world.	AI will start pushing towards AGI the generic & wider coverage.	AGI support in DPU will be needed for better throughput & latencies.	AGI will need Terabit networks and 6G following WiFi6 will be needed to address AGI with DPUs.
Challenge(s) for Need 5	To keep learning and trying newer ways to improve algorithms.	Skills will be needed to leverage transfer learning and other transforms.	AIML adaption will need optimization for different DPU architectures.	6G and DPU with AGI disruptions will need more specialized skills.
Possible Solution for Challenge	Applying and adopting new algorithms.	New kinds of mapping and transforms will evolve for the solution space.	Interop standards will evolve with portability for Qubits usage.	More labs and trainings by Gov. and research orgs. will emerge to meet the needs.

4.6. Multi-Access Edge Computing

4.6.1. Needs, Challenges, and Potential Solutions

There are two important use cases for MEC: (1) intelligent load balance across multiple sites (as illustrated in Figure 19) and (2) fault discovery and recovery (as shown in Figure 20).

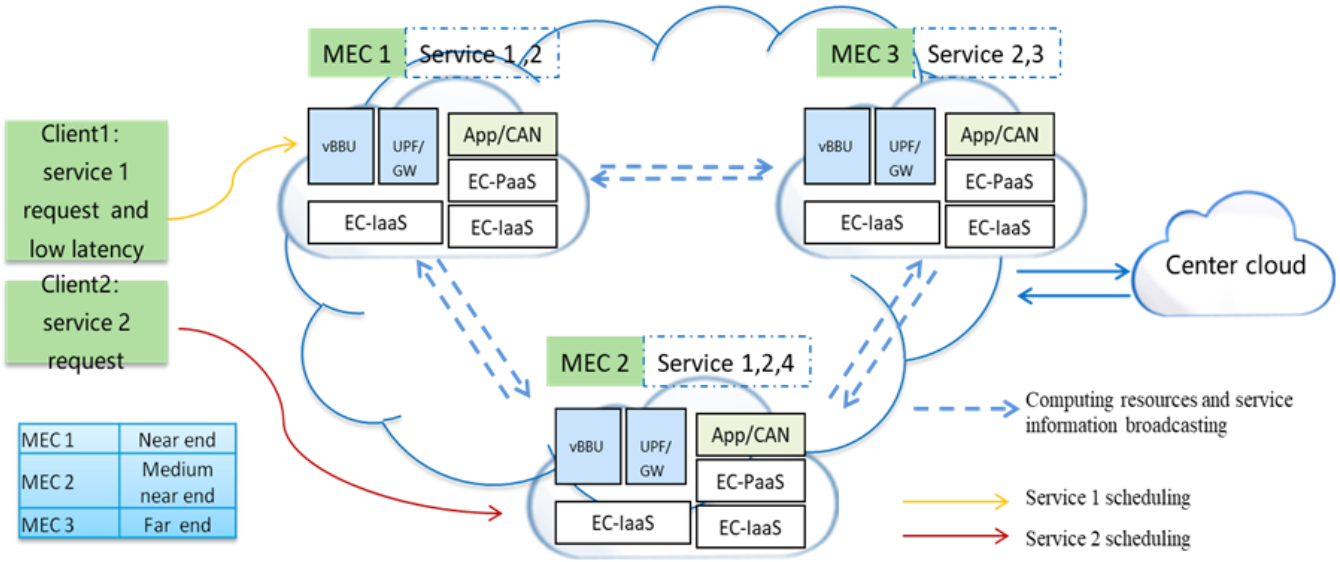

Figure 19: Intelligent Load Balancing

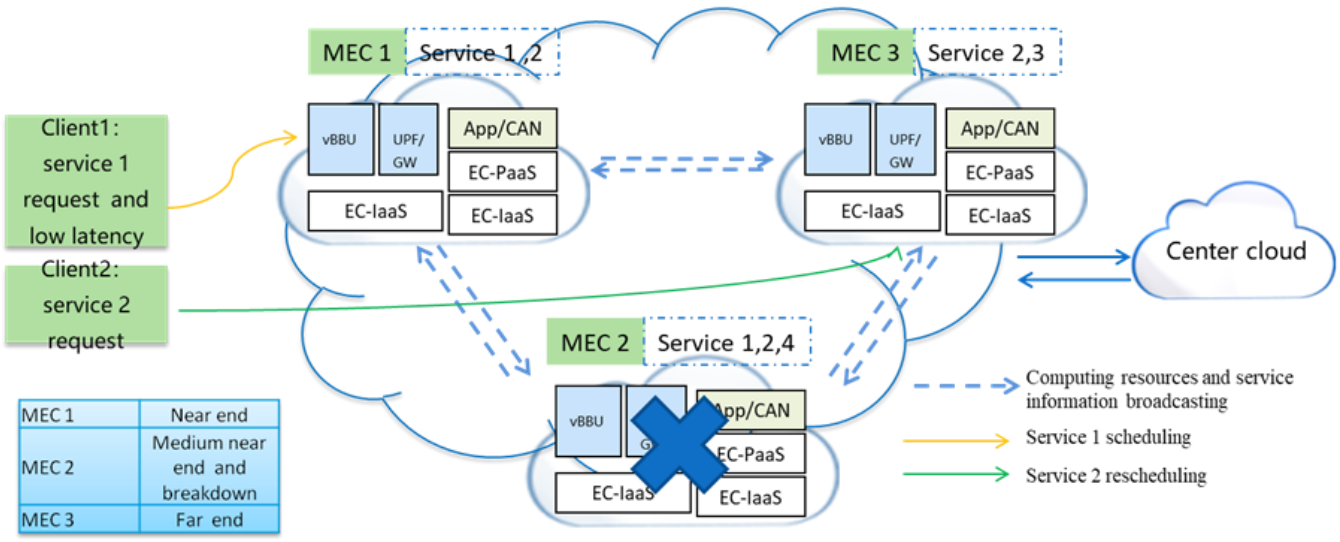

Figure 20: Fault Discovery and Recovery

4.6.2. Roadmap Timeline Chart

Table 7: MEC Needs, Challenges, Enablers, and Potential Solutions

Name	***Current State (2023)***	***3 years (2026)***	***5 years (2028)***	***Future State 10 years (2033)***
Need #1 Cloud Resource Optimization	Understanding domain and their use cases for AI/ML.	Transfer learning to reuse models.	Multi-service Gov. cloud & edge emerging to support Sustainable Development Goals (SDG).	New modular resources appear due to qubits and revival of analog & HPC.
Challenge(s) for Need 1 Cloud Descriptive Templating	Understanding AI/ML for domain experts and vice versa.	Customizing models tools to improve delivery efficiency.	More global collaboration & competition as political landscape changes, adding to jurisdiction & governance issues for data sovereignty.	Expect a new generation of Quantum automation templates & AI/ML models.
Possible Solution for Challenge Use labels & annotate to describe Resource modules & granularity	Training and learning tools and methodologies.	Moving from ML to DL for solutions.	Regionalization blocks (EU, NA, SA, Gulf & Middle East, Eurasia, ASEAN, BIMSTEC, China) comes to play instead of globalization or localization. The carriers may merge with edge / cloud providers.	Stronger regionalism will drive fragmentations and improve quality at cost that may go up. Solution will need more customizations using AL / ML.

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Need #2	Building domain specific platforms.	Support for NLP & voice-activated and translation will be a standard requirement.	New architectures needed for regions with RISC-V will emerge as competition to ARM & x86-64.	Quantum computing will need reimagining vertical solutions.
Challenge(s) for Need 2	Library management with different platforms for AI/ML.	AI/ML tools and API will need to address NLP and voice customizations. More dictionaries will be required in the cloud.	The very nature of Instruction Set Architectures (ISA) may need a relook.	Disrupt the existing crypto and domain-specific solutions.
Possible Solution for Challenge	Standardize languages, libraries and run times.	Faster storage & file systems will help alleviate some of the issues along with FPGA & DPUs.	New evolution of Quantum computing will spawn newer semi & optical computing elements.	As scale & speed break the barriers of Terabit, networks will come to provide possible new solutions.
Need #3	Building drivers for integrating emerging devices.	Drivers will need updates as xPUS & DPUS evolve.	Qubit drivers will need analogue and mixed signal designs.	Qubit & quantum drivers will need to integrate with hybrid solutions.
Challenge(s) for Need 3	Evolving accelerators drivers for emerging devices.	RDMA, GRPC, NVLink, ROE will need updates as DPUs bring new functions as a service.	Skill set for analog computing expats will revive.	Digital vs. analog vs. magnetics in merchant silicon & MEMs.
Possible Solution for Challenge	Testing accelerators in lab environment.	Harmonizing thread and multi-process will move to DPU managing CPU as well GPUs.	Reuse and updates of older Analogue techniques will be possible solutions. AI/ML will switch to older signals & systems, control systems theory for solutions.	Magnetics, memes and graphene may provide some options.
Need #4	Improving Smart NICs, NVME, FPGAs & GPUs for specific domains.	Software development & maintenance for Offloading / Acceleration will be a baseline.	Thinking Serial & programming parallel will need new AIML models.	New media & materials will need newer Fabs & Facilities.
Challenge(s) for Need 4	Architecture for caching and shared memory for parallel computing.	More skill set and research will be needed to design new busses like next generation PCIE / USB, etc.	Software to evolve for buffering, queuing, & caching for DPU combined with newer architectures.	Vacuum processing will emerge as the new challenge.
Possible Solution for Challenge	Evaluating right attributes for optimizing outcomes.	Extend PCIE & USB with newer alloys & tracks.	xxx	Some learnings going to vacuum tubes may provide pointers.

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Need #5	Innovation in AI across the world.	AI will start pushing towards AGI the generic & wider coverage.	AGI support in DPU will be needed for better throughput & latencies.	AGI will need Terabit networks and 6G following WiFi6 will be needed to address AGI with DPUs.
Challenge(s) for Need 5	To keep learning and try newer ways to improve algorithms.	Skills needed will be to leverage transfer learning and other transforms.	AIML adaption will need optimization for different DPU architectures.	6G and DPU with AGI disruptions will need more specialized skills.
Possible Solution for Challenge	Applying and adopting new algorithms.	New kinds of mapping and transforms will evolve for the solution space.	Interop standards will evolve with portability for Qubits usage.	More labs and trainings by Gov. and research orgs. will emerge to meet the needs.

4.7. Intelligent Optical Networks

4.7.1. Needs, Challenges, and Potential Solutions

In general, there are mainly three challenges need to be tackled for intelligent optical networks[27]:

- **Network complexity:** Firstly, the number and complexity of optical network devices increase with the scale of optical networks extend. The diversity of vendors of network devices and protocols makes the management of network complex. Additionally, the optical network acts as the bearer network in communication systems, which may carry multiple heterogeneous networks, such as 5G mobile networks, IoT, vehicle networking, and cloud computing. How to adapt the traffic from these different networks is becoming the first challenge for optical network operation and management.
- **Service complexity:** Different levels of Quality of Service (QoS) agreements require optical networks to provide differentiated services. The novel techniques and applications, such as network slicing, ask the service provision to be implemented in real time. Optical networks should furnish a flexible service-providing a mechanism to provide different levels of QoS to different services in the physical domain in a short period, which is difficult for a large network.
- **Resource management complexity:** The optical network is the bridge between upper-layer traffic and underlying physical layer resources. Thus, it is responsible for physical layer resource allocation for traffic provision. However, there are multiple dimensional physical resources to be allocated: fiber, wavelength, spectrum, modulation format, and time slots. The joint assignment of multiple resources is time consuming with high computational complexity.

With the increasing complexity of the optical network, traditional manual operation in optical networks requires too much time and may result in local optimization instead of global optimization. Conventional control and management approaches cannot satisfy the targets of low latency, scalability, and accuracy for future optical networks. To meet the demands of future optical networks, more intelligence should be introduced into the optical networks for monitoring, control, and management to minimize manual intervention as well as increase network flexibility and automation level[28]. ML algorithms can deal with complex problems by iteratively learning from input data and environment feedback. Deploying ML to optical networks is a promising way to introduce intelligence into optical

networks. Instead of manual-based, rule-based, and static programming-based network operations; intelligent optical networks aided by ML techniques can learn inner relationships from data and environment to perform more automated and flexible network operations.

There are still several challenges to be considered in the future:

(1) Open dataset access: Open datasets are crucial to applying ML because the performance of different methods can be compared based on the same dataset. In this way, an accessible open dataset will reduce the repetitive work and accelerate academic research progress. However, in the field of optical networks, there only a few open datasets available. Most work that previously surveyed use synthetic data for ML model training, which may lack credibility, and is difficult to be compared with other works. There are several hinderances in collecting and opening datasets of the real optical networks. Firstly, in real network operations, large resource margins will be reserved to maintain a good performance. Therefore, there are few negative samples in real networks and the datasets collected from the real environment may suffer data imbalance problems where the positive samples are much more than negative samples. Secondly, the real telecommunication data may face privacy issues, which make it difficult for Internet Service Providers (ISPs) to make datasets public.

Possible solutions: *i)* Standardize the process of data collecting, labeling, cleaning, and anonymizing to reduce the costs of data processing while maintaining data privacy. *ii)* Train the ML models with encryption mechanism, such as federated learning algorithm proposed in 2016[29], with which models can be trained without exact access to the data, and privacy can be guaranteed.

(2) Algorithm computational complexity

In real optical network scenario, there may be strict requirements for the timeliness of tasks. Therefore, it is important for ML algorithms to reduce their computational complexity. However, only a small set of works analyze ML feasibility based on time complexity. Most previous experiments demonstrate the effectiveness of the proposed methods without time restriction. These optimal results were obtained assuming there is enough time for computation. However, real environments are usually time-sensitive and a better comparison principle is to compare the suboptimal solutions of different algorithms under a given time threshold.

Possible solutions: *i)* In monitoring tasks, although many ML algorithms, such as deep neural networks, have the capability of automated feature extraction, the original data still should be preprocessed and expert knowledge should be introduced into feature engineering to simplify the model training, e.g., pre-test the data, and only input the indicators of tests into neural networks for training. *ii)* In decision-making tasks, suboptimal solutions should be allowed to balance the network performance and time consumption of action computation.

4.7.2. Roadmap Timeline Chart

Table 8: Intelligent Optical Network Needs, Challenges, Enablers, and Potential Solutions

Name	***Current State (2023)***	***3 years (2026)***	***5 years (2028)***	***Future State 10 years (2033)***
Need #1 Open Dataset Corresponding to Practical Networks	Old open dataset, synthetic data set	Open data set from different companies	Open data set platform for optical networks	
Challenge(s) for Need 1	No open dataset	Privacy preserving	Platform construction and maintenance	

Name	*Current State (2023)*	*3 years (2026)*	*5 years (2028)*	*Future State 10 years (2033)*
Possible Solution for Challenge	Using small dataset	Federated learning	Standardization	
Need #2 High Efficiency AI Methods	Huge computational complexity	Preprocess the raw data and expert knowledge should be introduced	Knowledge transfer learning	Knowledge distillation
Challenge(s) for Need 2	High computational cost	Huge human expense	Efficient transfer learning algorithms	Efficient neural network architecture
Possible Solution for Challenge	Transfer learning	Pretraining models for optical network	Open platform for knowledge transferring	Open platform for knowledge distillation from pretraining models

4.8. Intelligent Radio Access Networks (iRAN)

Intelligent RAN (iRAN) is a new approach to radio access network (RAN) design that leverages artificial intelligence (AI) and machine learning (ML) to improve network performance, efficiency, and security.

4.8.1. Needs, Challenges, and Potential Solutions

iRAN has the potential to address many challenges facing traditional RAN networks, including:

- RAN networks are becoming increasingly heterogeneous, with a mix of different radio access technologies (RATs) and vendor equipment. Different vendors and technologies can make managing and optimizing the network challenging.
- RAN networks are complex systems with many moving parts, making troubleshooting problems and identifying performance bottlenecks difficult.
- RAN networks are increasingly vulnerable to cyber attacks. iRAN can help to improve network security by using AI/ML to identify and mitigate threats.

iRAN also has the potential to offer many benefits, including:

- Improved performance: iRAN can help improve network performance by optimizing the use of resources and by predicting and preventing problems.
- Increased efficiency: iRAN can help increase network efficiency by automating tasks and reducing the need for human intervention.
- Enhanced security: iRAN can help enhance network security by identifying and mitigating threats.

However, several challenges need to be addressed before iRAN can be widely adopted, including:

- Data privacy: iRAN requires collecting and analyzing large amounts of data, raising concerns about privacy and security.
- Cost: iRAN can be expensive to implement and maintain, which may be a barrier for some operators.
- Standardization: iRAN is still a relatively new approach and there is no single standard for its implementation, which can make it difficult for operators to deploy and interoperate with other networks.

Potential solutions to the challenges of iRAN include:

- Address data privacy concerns using anonymized data and implementing strong security measures.
- Reduce the cost of iRAN by using open-source software and deploying iRAN in stages.
- Work with industry bodies to develop standards for iRAN to make it easier for operators to deploy iRAN and interoperate with other networks.

As part of the ecosystem, the Open Radio Access Network (O-RAN) may be a vehicle through which iRAN may be further developed. The O-RAN provides an open disaggregated architecture for RAN networks. The O-RAN architecture provides for software defined component called the RAN Intelligent Controller (RIC). The RIC is further divided into Near Real Time RIC and Non-Real Time RIC (Non-RT RIC). The Non-RT RIC operates on timescales of seconds or slower and may be used to perform such tasks as policy management, resource allocation, and fault detection and resolution.

AI/ML may be incorporated into the Non-RT RIC in use cases that may include:

- Developing models to predict network performance and identify potential problems. This can help to improve the performance of RAN networks.
- Automating tasks such as policy management and resource allocation to increase the efficiency of RAN networks.
- Developing models that can identify and mitigate security threats. This can help to enhance the security of RAN networks.

AI/ML will likely be used in the RAN for automation, optimization, and security. AI/ML may be used to automate tasks that humans currently perform, such as network planning and optimization, fault detection and resolution, and resource allocation. AI/ML may also be used to improve the security of RAN networks by identifying and mitigating threats. Examples of how AI/ML may be used in the future of RAN include:

- Automating the deployment and configuration of RAN equipment, reducing the time and resources required to deploy new RAN networks.
- Predicting and preventing network outages, improving RAN network reliability and uptime.
- Improving the performance of RAN networks for users with high-bandwidth applications, such as video streaming, which will provide a better user experience for these applications.
- Identifying and mitigating security threats to RAN networks, which will help protect the network from unauthorized access and attacks.

Attendant benefits of using AI/ML in RAN may include:

- Reducing the costs of operating and maintaining RAN networks through more automation.
- Improving customer experience by improving reliability and consistency of network performance; analyzing data from RAN sensors to identify potential problems before they cause outages.
- Enhancing network security by identifying and mitigating security threats.

4.8.2. Roadmap Timeline Chart

Table 9: iRAN Timeline Chart

Name	***Current State (2023)***	***3 years (2026)***	***5 years (2028)***	***Future State 10 years (2033)***
iRAN	Development of standards Pilot projects by a few operators	Availability of standards Larger trials by more operators	Widespread adoption of standards Widespread adoption by operators	

5. USE CASES

5.1. Training and Deployments of Inference Models at the Edge

AI/ML as a technology has tremendous impact on what, why, when, and where one applies the different forms of learnings, models, and predictions across the edge cluster. The top cluster topology in Figure 21 shows three clusters: training, inference, and analytic clusters.

Figure 21 shows the infrastructure that delivers streaming video content to the deployment edge may not have the computing resources to execute the training cycles beyond using runtime for inference. Thus training is offloaded to the cloud in the training cluster, leaving only the inference process in the deployment edge. The deployment edge may have other limitations, such as limited GPU instance support for Multi-Instance GPUs (MIG), and other computing, networking, and latency constraints.[7] To facilitate the exchange of ML models between the training cluster in the cloud and the inference cluster in the edge, an open-source format to represent AIML models, such as Open Neural Network eXchange (ONNX), may be used. Other use cases that may benefit from similar deployment model (inference at the edge, training in the cloud) include Autonomous Vehicle Systems (AVS) that leverage these techniques for object and computer vision for V2X and vehicle driving in traffic.

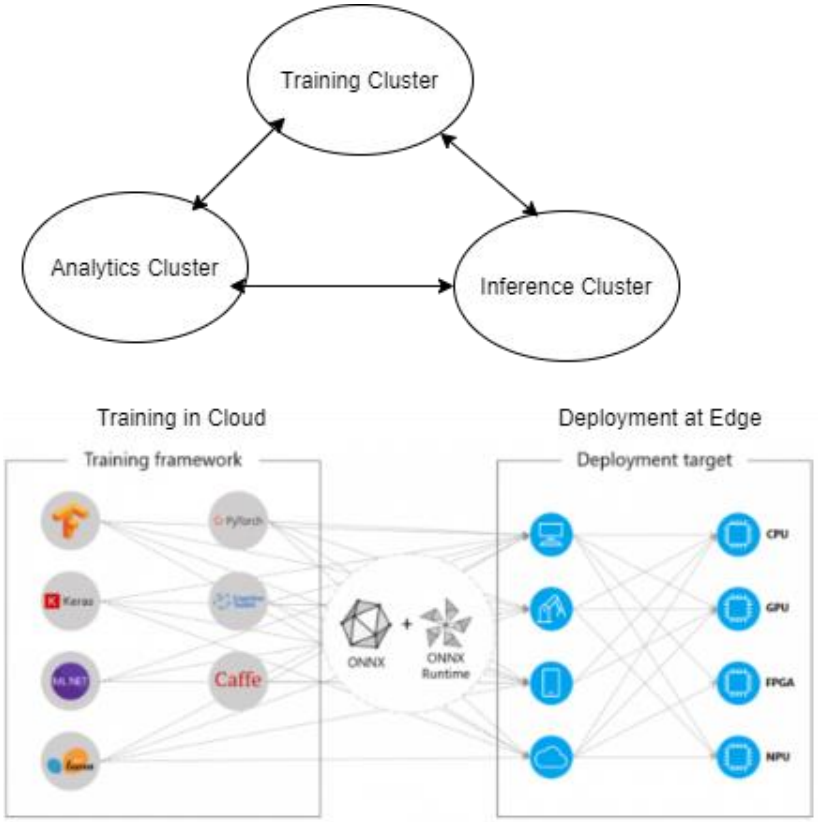

Figure 21: Training and Deployments of Inference Models at the Edge

[7] *For various techniques to address constraints concerning computation, networking, latency, etc., see "Machine Learning at Edge Cloud", Open Infrastructure Summit, October 19-23, 2020, panel discussion video published by the OpenInfra Foundation, https://youtu.be/TJIj8oMONPo*

5.1.1. Edge Infrastructure for Generative and Analytical AI

Traditional analytical AI models typically have a much smaller inference task compared to the training models requiring very large datasets and more powerful computational and storage platforms.

Though the generative AI modeling techniques (GANs, VAEs) might present challenges with the need for higher computational power, the inference tasks (with some optimization effort) might be permissive on smaller edge compute platforms. For instance, recent advances in generative AI modeling, such as the stable diffusion model, could be very large in scope for training. However, the inference task could be optimized to run even on a very small compute footprint.[8]

The edge infrastructure would allow offloading of latency-sensitive optimized inference tasks (except for the wider scope ones) from the cloud (e.g, Microsoft Azure[9] for OpenAI's ChatGPT[10]) to the edge resources scalable through interconnected local edge nodes.

It is not just the hardware, but the software architecture that will continue to be the core of such optimization. The AI architectural optimization along with the perpetual demand for near real-time actions will incentivize increased offloading of AI modeling tasks to the edge.

Latency-compliant resource allocation among interconnected edge nodes with multi-ownership and multi-operator jurisdictions are some of the key challenges for edge scalability. But since the limited use cases are much more specific and therefore narrower in scope, they would not require such high scalability, but higher performance as compared to a cloud.

[8] *ETSI GS MEC 003 V3.1.1 (2022-03), Multi-access Edge Computing (MEC); Framework and Reference Architecture.*
[9] *Microsoft Azure, https://azure.microsoft.com/.*
[10] *OpenAI's ChatGPT: https://chat.openai.com/.*

6. EXTERNAL OPPORTUNITIES

Identify and highlight potential external entities or activities going on in your field that may be applicable for collaboration or partnership.

7. AI/ML STANDARDS DEVELOPMENT

7.1. In Progress: IEEE P1900.8 Standard on Machine Learning for RF Spectrum Awareness in DSA and Sharing Systems

The IEEE DySPAN-SC recently created a new working group to develop the IEEE P1900.8 standard on the topic of ML for RF spectrum awareness. This standard is expected to be completed by 2023 and will lead to improved RFML spectrum awareness models through the creation of standard and reusable datasets and ML model pipeline interfaces. Once published and adopted, it will enable the creation of a dataset sharing community and a marketplace for RFML spectrum awareness model developers. Figure 22 depicts a variety of use cases being addressed in the informational annex of the standard. The P1900.8 working group participants are experts in these RFML spectrum awareness use cases and are planning to adopt the standard once it is approved.

Figure 22: Continuum of Spectrum Awareness Use Cases Based on RF ML Model Inferencing and Prediction Capabilities

The following text is copied from the PAR for the P1900.8 standard that was approved by the IEEE SA New Standards Committee in November of 2021:

Purpose Statement: This standard defines a common approach for creating datasets used to train and test machine learning models that detect, classify, characterize, and/or identify radio frequency signals and signal emitters. The standard also defines a criterion for evaluating the performance of the machine learned spectrum awareness models.

Need Statement: There is no existing standard for how to generate high-quality datasets to train and test Machine-Learned Spectrum Awareness (MLSA) models or how to evaluate their performance. Openly available MLSA datasets suffer from insufficient data veracity and sample sizes to develop models that meet the performance requirements of real-world scenarios. Available datasets also suffer from insufficient representation of the variability of realistic spectrum operations that include a range of RF propagation channel effects and interference signals – greatly limiting the performance of the MLSA models once deployed in the real-world.

Therefore, MLSA model developers are forced to create their own unique approach for generating datasets and evaluating model performance – greatly increasing the risk of incurring unexpected costs, encountering hidden technical challenges, and creating models that are unable to meet the performance requirements of real-world scenarios.

We expect that this standard will also be adopted by the research community that engages in the sharing and exchange of machine learning datasets. The open availability of high-quality MLSA datasets will greatly accelerate improvements to the state-of-the- art of MLSA model design, similar to how the computer vision research community rapidly improved the image classification performance of deep neural networks trained on shared image datasets.

Scope of the IEEE P1900.8 standard: This standard describes how to generate datasets to train and test Machine-Learned Spectrum Awareness (MLSA) models that detect, classify, characterize, and/or identify radio frequency.

RF signals and signal emitters. The scope of this standard includes:

- methods for creating training and test datasets for MLSA models that are representative of real-world Dynamic Spectrum Access (DSA) and spectrum sharing scenarios,
- methods for using data augmentation techniques to introduce sufficient sample variation so that the MLSA model can generalize to real-world scenarios,
- methods for enhancing the training dataset with RF propagation channels and interference sources that are representative of real-world scenarios,
- specifications for how to structure and store MLSA datasets,
- methods for creating secure and performant MLSA models that operate on resource-constrained RF sensors and processors, and
- criteria for evaluating the performance of MLSA models.

8. CONCLUSION

8.1. Summary of Conclusions

The IEEE Future Network Initiatives AI/ML working group has identified requirements for integration of AI/ML technologies into various elements of 5G and future networks according to a 3-, 5-, and 10-year timeline. In this edition, the foundation for AI/ML integration into future networks was laid in a limited set of areas. Initial technology roadmaps were set forth for network slicing, network digital twins, security, dynamic spectrum access, cloud computing, and multi-access edge computing. In addition, AI/ML for network automation and supplemental information on AI/ML workflow and security were provided.

The objective of the AI/ML working group for this version was to lay the groundwork for future development by first focusing on a core set of application areas so that subsequent versions could undertake the following two tasks: 1) Further refine the initial roadmaps and 2) add new roadmaps to support other technology areas. With regard to the latter, it is important for this working group to interface with others to allow integration of AI/ML approaches to edge automation platform, satellite, security, etc.

The application areas in this document were very diverse and some are more mature than others; the results of this working group demonstrate that. For example, the use of AI/ML for DSA has been studied for the better part of the last ten years, so it is much more mature. This is demonstrated by the fact the IEEE is currently working on standards. Other areas are not so mature, so more developmental work is needed in these areas. As an example, while security is mature in some ways related to AI/ML; in others, it is not. Spam filters that use AI/ML to classify emails are commonplace, but predictive network security is not, and neither is the orchestration process discussed here. There is no shortage of exciting AI/ML-based work to be undertaken for 5G and future networks in the future.

8.2. Working Group Recommendations

Based on our group's findings as described in the previous section, we recommend the following set of AI/ML-based activities:

- Interface with other FNTC working groups to determine which AI/ML algorithms or technologies can be used to augment their systems.
- Investigate additional 5G and future networks areas where technology gaps can be closed using AI/ML. One example is quantum computing and security.
- Set priorities for future development to include both technological advances and AI/ML developments that are being undertaken by other organizations, such as ETSI, 3GPP, and others.
- Develop a management and orchestration framework that addresses the operation of the AI/ML elements of the network.
- Define how open source and open architectures can be used and adopted with buy-in from equipment manufacturers. With regard to open RAN, for example, a joint effort could produce technologies for adoption by industry via the O-RAN Alliance and Telecom Infra Project.
- Develop and demonstrate use cases to show the value of AI/ML integration into 5G and future networks.

9. CONTRIBUTOR AND EDITOR BIOS

Ashutosh Dutta is currently Chief 5G Strategist and JHU / APL Sabbatical Fellow at Johns Hopkins University Applied Physics Labs (JHU / APL), USA. He also serves as Chair for Electrical and Computer Engineering for Engineering Professional Program at JHU. His career includes Director of Technology Security and Lead Member of Technical Staff at AT&T, CTO of Wireless at a Cybersecurity company NIKSUN, Inc., Senior Scientist in Telcordia Research, Director of Central Research Facility at Columbia University, adjunct faculty at NJIT, and Computer Engineer with TATA Motors. Ashutosh has authored more than 100 technical papers and has 31 issued patents. Ashutosh is co-author of the book titled, *Mobility Protocols and Handover Optimization: Design, Evaluation and Application* published by IEEE and John & Wiley. As a Technical Leader in 5G and security, Ashutosh has been serving as the founding Co-Chair for the IEEE Future Networks Initiative that focuses on 5G standardization, education, publications, testbed, and roadmap activities. Ashutosh was IEEE Communications Society's Distinguished Lecturer for 2017-2020 and an ACM Distinguished Speaker (2020-2022). Ashutosh currently serves as the founding co-chair for IEEE Future Networks Initiative and Member-At-Large for IEEE Communications Society. Ashutosh has served as the general Co-Chair for the IEEE STEM conference for the last ten years. Ashutosh served as the Director of Industry Outreach for IEEE Communications Society from 2014-2019. He was recipient of 2009 IEEE MGA leadership award and 2010 IEEE-USA professional leadership award. Ashutosh currently serves as Member-At-Large for IEEE Communications Society for 2020-2022.

Baw Chng is a technologist, inventor, and author who consults in the wireless networking and telecommunications industry. Baw Chng has been awarded multiple patents in such areas as network and system architecture, network planning, security and authentication, mobility, system selection, service provisioning, access control, user interface and user experience, and network management. Baw Chng has published papers on computer networking, computer architecture, and magnetic recording in peer reviewed publications. Through his consulting practice at BAWMAN LLC, Baw Chng offers patents and intellectual property consulting, technology consulting, standards consulting, and business strategy consulting services in the networking and communications industry. Baw Chng also offers expert consultation to the legal community and investors. Baw Chng has extensive experience working with large Tier-1 communications service providers, their strategic technology vendors, technology start-ups, and various standards-development organizations in the industry. Baw Chng has done pioneering work on femtocell and small-cell technologies and his body of work spans a broad array of critical technologies; including Wi-Fi (mesh and infrastructure), cellular communications, Internet-of-Things (IoT), low-power wide-area network (LPWAN), network function virtualization (NFV), software defined networking (SDN), self-organizing and self-optimizing networks (SON), cloud computing, plastic optical fiber (POF), optical networking, quality of service (QoS), and fair queueing. Baw Chng also serves as an editor for the 2023 edition of the AI/ML Chapter of the IEEE INGR.

Deepak Kataria has over 25 years' experience in data networking, cloud computing and telecom domains with the unique distinction of having worked in technical leadership and service delivery roles with telco operators (AT&T Bell Labs), system OEM vendors (Lucent Technologies, Fujitsu), silicon & software providers (Agere Systems, LSI), and system integrators (HCL America). He co-founded IPJunction, Inc. in 2009, consulting telco clients on new solution opportunities, target markets, key

differentiators, product management consulting, and creating ecosystem partnerships for the successful execution of identified opportunities. Currently, he serves as the Principal Solution Consultant at Ericsson and leads service delivery of complex multi-domain solutions covering cloud native, orchestration, networking, and RAN technologies for telco edge / core computing and emerging enterprise services use cases for AT&T and Microsoft projects. He holds ten US patents and has several others pending. He has published extensively in industry and IEEE publications. He serves as the Chair of IEEE Princeton Central Jersey Section (Region 1), has been the General Chair of IEEE Sarnoff Symposium since 2015, and leads the IEEE FNI working group on AI/ML as a Co-Chair. He holds a B.S. in Electronics and Communications Engineering and pursued M.S. and Ph.D. degrees in Electrical Engineering from Rutgers University, NJ. He has completed Harvard's Emerging Leader's professional program on virtual leadership covering strategy, customer focus, corporate governance, and innovation.

Anwar Walid (Fellow, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering from New York University and a Ph.D. degree from Columbia University. He was with Nokia Bell Labs as the Head of the Mathematics of System Research Department and the Director of University Research Partnerships. He is currently the Director of Network Intelligence and Distributed Systems Research and a Distinguished Member of the research staff with Nokia Bell Labs. He is also an adjunct professor with Electrical Engineering Department at Columbia University. He has more than 20 U.S. and international granted patents on various aspects of networking and computing. His research interests include the control and optimization of distributed systems, learning models and algorithms with applications to the Internet of Things (IoT), digital health, smart transportation, cloud computing, and software-defined networking. He is an elected member of the International Federation for Information Processing Working Group 7.3 and the Tau Beta Pi Engineering Honor Society. He was the recipient of awards from the IEEE and ACM, including the 2017 IEEE Communications Society William R. Bennett Prize and the ACM SIGMETRICS / IFIP Performance Best Paper Award. He was an Associate Editor for the IEEE / ACM IEEE *Transactions on Cloud Computing*, *IEEE Network Magazine*, and the IEEE / ACM *Transactions on Networking*. He was the Technical Program Chair for the IEEE INFOCOM, as the General Chair for the 2018 IEEE / ACM Conference on Connected Health (CHASE), and as a guest editor for the IEEE IoT Journal special issue *AI-Enabled Cognitive Communications and Networking for IoT*.

Dr. Frederica Darema is the President and CEO of the InfoSymbiotic Systems Society. In 2019, she retired as Senior Executive Service member and as Director of the Air Force Office of Scientific Research where she led the entire basic research investment for the AF on science and technology for future capabilities and transitioning the discoveries to defense industries & broader commercial sector. She also served as Research Director in the Air Force's Chief Data Office and as Associate Deputy Assistant Secretary the Air Force Office for Science, Technology, and Engineering. Prior career history includes: research staff positions at the University of Pittsburgh, Brookhaven National Laboratory, and Schlumberger-Doll; and management and executive-level positions at the T.J.Watson IBM Research Center, the IBM Corporate Strategy Group, the National Science Foundation, and the Defense Advanced Research Projects Agency. She was Director of the AFOSR Directorate for Information, Math, and Life Sciences. Dr. Darema, earned her PhD in Nuclear Physics, is a graduate of the University of Athens, the Illinois Institute of Technology, and University of California at Davis, and was a Fulbright Scholar and a Distinguished Scholar. She is a Life Fellow of the IEEE, among other professional recognitions. She has made seminal contributions in the supercomputing field, including SPMD, the predominant

computational model for parallel computing and methods for performance engineering parallel and distributed systems. She pioneered the DDDAS paradigm and has organized and led research initiatives, programs, workshops, conferences, and other forums to foster and promote DDDAS-based science and technology advances.

Dr. Mahmoud Daneshmand is Co-Founder and Professor of Department of Business Intelligence & Analytics as well as the Data Science Ph.D. Program, and Professor of Department of Computer Science at Stevens Institute of Technology. His industry and university experience includes Executive Director, Assistant Chief Scientist, professor, researcher, Distinguished Member of Technical Staff, technology leader, Founding Chair of Department, and Dean of School at: Bell Laboratories; AT&T Shannon Labs–Research; University of California, Berkeley; University of Texas, Austin; New York University; Sharif University of Technology; University of Tehran; and Stevens Institute of Technology. Dr. Daneshmand received his Ph.D. and M.S. degrees in Statistics from the University of California, Berkeley, M.S. and B.S. degrees in Mathematics from the University of Tehran.

He is a Data Scientist, expert in big data analytics, artificial intelligence, and machine learning with extensive industry experience including with the Bell Laboratories as well as the Info Lab of the AT&T Shannon Labs – Research. He has published more than 300 journal and conference papers; authored / co-authored three books, and graduated more than 2500 Ph.D. and M.S. students. He holds key leadership roles in IEEE journal publications, including *IEEE IoT Journal*, *IEEE Transaction on Big Data*; IEEE major conferences; industry-IEEE partnerships; IEEE future direction initiatives, and the IEEE AI/ML Initiative for Future Networks. He has served as General Chair, Keynote Chair, Panel Chair, Executive Program Chair, and Technical Program Chair of many IEEE major conferences. He has given many keynote speeches in major IEEE and international conferences.

Dr. Michael A. Enright is the CEO / President of Quantum Dimension, Inc. with experience in cybersecurity, artificial intelligence / machine learning, embedded and quantum computing, RF communication, and more. At Quantum Dimension, Michael has led engineers in the company's technology developments, which includes AI/ML, RF communication, and navigation using advanced embedded technologies. He has been an adjunct professor in the Electrical Engineering Department at the University of Southern California, where he taught both undergraduate and graduate courses in image and signal processing and wireless communication systems design.

Prior to founding Quantum Dimension, Michael was a researcher with the Signal and Image Processing Institute (SIPI) and the Integrated Media Systems Center (IMSC) at USC, where he worked on multimedia cross-layer communication techniques. He led the development of the video compression architecture for Boeing Digital Cinema, acted as an information security lead for network security, and was responsible for the design and development of a phased-array at Hughes Space and Communication. At Motorola Cellular, he designed the handset design for the Iridium satellite system. He began his career at AT&T Bell Laboratories where he developed DSP software for AT&T's 5ESS telephone switching systems.

Michael has a Ph.D. in Electrical Engineering from USC, an M.S. in Electrical Engineering from the Illinois Institute of Technology, an M.S. in Mechanical Engineering from the University of Missouri-Columbia and a B.S. in Aeronautical and Astronautical Engineering from the University of Illinois at Champaign-Urbana. Michael is a Senior Member of the IEEE.

Dr. Chi-Ming Chen is retired from AT&T Labs after years of working in the telecommunications industry. Chi-Ming received his Ph.D. in Computer and Information Science from the University of Pennsylvania; M.S. in Computer Science from the Pennsylvania State University; M.S. and B.S. in Physics from Tsing Hua University, Taiwan. Chi-Ming is a Life Senior Member of IEEE and Senior Member of the ACM. He is an Advisory Board Member of IEEE Communications Society (ComSoc) Communications Quality & Reliability Technical Committee (CQRTC). He was a voting member of the IEEE GLOBECOM & ICC Management & Strategy (GIMS) Standing Committee and served as the GLOBECOM and ICC Site Selection Chair from 2012 to 2017. He also served as the Executive Chair of ICC 2019, Shanghai, China. Dr. Chen has been co-chairing the Roadmap Working Group of IEEE Future Networks Initiative (was named as 5G Initiative initially) since 2016. The Working Group publishes annually the International Network Generations Roadmap (INGR).

Rentao Gu, Vice Dean and Professor in School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), China. He received the Ph.D. degree from BUPT in 2010. From 2008-2009, he studied in the networks and Learning Lab in Georgia Institute of Technology, USA, as a visiting scholar. Since 2010, he has joined BUPT as a faculty member. After that, as the project leader and principal researcher, he has successively undertaken more than ten national, ministry-level research projects related to the software defined optical networks, 5G transport network, and AI enabled networking. He has held more than 30 U.S. invention patents and Chinese invention patents and published many peer-reviewed journals and conferences papers. His inventions and technical contributions were awarded the National Science and Technology Progress Award of China and several other science and technical awards. Rentao served in the Organizing Committee / Technical Program Committee of various international conferences (such as GLOBECOM / ICC / NFV-SDN), was Young Professional Standing Committee Member of IEEE Communication Society from 2016 to 2017, and MDC member of IEEE Computer Society from 2012 to 2013. He is currently the Lead Coordinator of Region 10 Young Professional Committee and a Senior Member of the IEEE.

Honggang Wang is a professor of Electrical and Computer Engineering at UMass Dartmouth. His research interests include Internet of Things (IoT), wireless health, Body Area Networks (BAN), cyber and multimedia security, mobile multimedia and cloud, wireless networks and cyber-physical systems, and big data in mHealth. He has produced a body of high-quality publications in prestigious journals and conferences in his research areas, winning prestigious best paper awards six times, including Globecom'19 and WCNC'08. His research has been mainly supported by federal agencies such as NSF, NIH, and DoT. He is an alumnus of National Academic Engineering (NAE) Frontiers of Engineering program and was an invited participant by NAE for 2017 German-American Frontiers of Engineering Symposium. He serves as the steering committee and founding co-chair of the IEEE Conference on Connected Health (CHASE) and TPC co-chair of IEEE CHASE 2016, a leading international conference on connected health. He has also been serving as the Editor in Chief for *IEEE Internet of Things* journal since 2020 and Associate Editor for *IEEE Transactions on Big Data* and *IEEE Transactions on Circuits and Systems for Video Technology*. He was the past Chair (2018-2020) of IEEE Multimedia Communications Technical Committee and is the Chair of IEEE eHealth Committee (2020-2021). He is an IEEE Distinguished Lecturer (ComSoc, 2019-2020) and an IEEE Fellow for his contribution to low power wireless for IoT and multimedia applications.

Alex Lackpour is a Principal Investigator / Chief Scientist of Wireless at Peraton Labs. He has experience leading research projects as a contractor for the US Government. At Peraton Labs, he leads several projects that enable co-existence and spectrum compatibility for a variety of commercial and DoD spectrum-dependent systems. He is an established leader in the Dynamic Spectrum Access and Sharing (DSA / S) community with a long record of successful program capture, management, and execution. Prior to Peraton Labs, Mr. Lackpour was a Systems Engineering and Technical Assistance (SETA) contractor to the Deputy Director of NOAA's Radio Frequency Management Division (RFMD). In this role, he advised RFMD leadership on spectrum sharing and representing NOAA's severe weather surveillance mission within the multi-agency SENSR radar program. Previously, he was a senior member of the Engineering staff at Lockheed Martin Advanced Technologies Laboratories, where he developed autonomous AI/ML spectrum control and sharing technologies.

Mr. Lackpour currently chairs the IEEE P1900.8 working group on Standard for Training, Testing, and Evaluating Machine-Learned Spectrum Awareness Model. He is also Vice-Chair of the IEEE 1900.5 working group on Policy Language and Architectures for Managing Cognitive Radio for Dynamic Spectrum Access Applications. He is also Secretary of the IEEE Dynamic Spectrum Access Networks Standards Committee (DySPAN-SC).

Mr. Lackpour is a part-time Ph.D. candidate in the CSE Department at Drexel University (ABD). He earned his M.S. in Electrical Engineering from Pennsylvania State University and his B.S. in Electrical Engineering from The College of New Jersey.

Pranab Das is a technology professional with experience in IT, software engineering, E2E solutions architecture, infrastructure, IP networks, building innovative products, and services with best practices and positive ROIs. His research interests include cloud, 5G, ORAN, AR, VR, SDN, NFV, MEC, satellite, robotics, smart cities, SoC, UAV, distributed systems, IoT, AI/ML, blockchain, network architecture and automation, product manufacturing, and emerging technologies.

He has successfully solved complex architecture problems and delivered customer satisfaction savings in the millions annually at many enterprises. He has demonstrated expertise in enterprise architecture, technology strategy, technology consulting, systems integration, cloud technologies, security architecture, network administration, troubleshooting and maintenance, application management, telecommunication technologies, and technical product management.

He has a B.S. in electronics & telecommunications engineering and an M.S. in electrical engineering. He is a member of the IEEE Communications Society.

Prakash Ramachandran received his M.Tech from IIT Bombay. Prakash is the Executive Secretary of the eMerging Open Tech Foundation. He is a Senior Member of IEEE, CS, Comm. Co-Chair of IEEE FNTC Edge Services & Platform working group 2018-2023.

T.K. Lala is the founder of ZcureZ and on a mission to architect modern edge and hybrid computing solutions fueled by 5g +, virtualization, and AI/ML technologies. He serves as an executive advisor to a cyber security startup board and a Silicon Valley high tech incubator. TK led several Silicon Valley startups as CTO / VP of Engineering and held senior engineering management responsibilities at

Fortune 500 organizations: including AT&T Bell Labs, NEC, Hitachi, Fujitsu, 3Com. Motorola, Mitre Labs, Rockwell, and BAE.

T.K., a senior member of IEEE, has served as a technical editor of IEEE *Communication* magazine for about a decade, held executive responsibilities in MIT-affiliated Venture Lab, and is serving as a co-chair of IEEE INGR Edge Networking Roadmap working group and contributes regularly to the LF CNCF Security working group. He was a major contributor and member of many SDOs, including IEEE 802 working groups, ATM, ADSL ANSI, and IETF. T.K. is a certified PMP and recipient of several awards and has been certified as an internationally recognized Expert System Engineering Professional ESEP (INCOSE) and has published in IEEE and other publications. He has been featured as an invited speaker, panelist, and moderator in IEEE and SAE conferences.

In addition to a B.S.E.E., T.K. holds an M.S.E.E. from Queen's University, pursued a Ph.D in Electrical Engineering, holds patents in cybersecurity and networking, and is listed in *Who's Who in America*.

Reinhard Schrage is a senior consultant at SchrageConsult, Germany. He studied Pure Mathematics, Stochastics, and Theoretical Computer Science at USC, Los Angeles, and Leibniz University of Hannover, Germany, where he received his MSc in Mathematics. His work expertise includes the calculation of nuclear fallout areas for the German Army, customer-specific design, creation, and implementation of financial application software for the Treasury in New Zealand, customer-specific adaptation of LAN protocol software for New Zealand Parliament, support for strategic Northern Telecom customers (like Deutsche Telekom and Telefonica in Madrid, Spain), European Customer Service Manager at UK-based Cable and Wireless, planning and management responsibility for the global financial services network of British Telecom in London, Resident Nortel Networks Engineer at Volkswagen headquarters, and stochastic performance analysis of smart phone clusters. He is contributing member for several IEEE standardization workgroups, like IEEE 1903 Next Generation Service Overlay Network, vice chair for 1910.1 Standard for Meshed Tree Bridging with Loop Free Forwarding, vice chair for 1900.1 Standard for Definitions and Concepts for Dynamic Spectrum Access, editor of the 1900.5.1 Standard Policy Language for Dynamic Spectrum Access Systems, as well as contributing member to RFC 6349 Framework for TCP Throughput Testing.

Dr. Ripal Dilipbhai Ranpara is an accomplished educator, esteemed researcher, and prolific author with a focus on Artificial Intelligence (AI) and cyber security. With a wealth of knowledge and expertise, Dr. Ranpara has written over 14 books, contributing to the dissemination of valuable insights and advancements in the field. In addition to academic pursuits, Dr. Ranpara offers comprehensive curriculum design services, crafting engaging and innovative learning experiences for students. They also provide technology services, leveraging their expertise to support organizations in implementing cutting-edge solutions. Dr. Ranpara's exceptional contributions have been recognized through numerous prestigious awards, including the esteemed "Women in IT of the Year" and "Mentor of the Year" accolades. These honors highlight their dedication to empowering others and making a significant impact in the field. With a passion for excellence in education, research, and technology, Dr. Ranpara continues to inspire and drive positive change in the AI and cyber security domains. Dr. Ranpara serves as an editor for the 2023 edition of the AI/ML Chapter of the IEEE INGR.

10. REFERENCES

[1] *R. Eberhart. "Overview of computational intelligence [and biomedical engineering applications].", In Proc. 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1998, pp. 1125-1129.*

[2] *T. Hui, D. Brown, B. Haynes, W. Xinxian. "Embedded e-diagnostic for distributed industrial machinery", in IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications, 2003, pp. 156-161.*

[3] *M. Awadallah, M. Morcos, "Application of AI tools in fault diagnosis of electrical machines and drives-an overview.", IEEE Transactions on Energy Conversion, vol 18, Issue 2, 2003, pp. 245-251.*

[4] *A. Abdelaziz, M. Elhoseny, A. S. Salama, A.M.Riad, "A machine learning model for improving healthcare services on cloud computing environment". Measurement, Vol. 119, 2018, pp. 117-128.*

[5] *TowardsDataScience.com, "Cousins of Artificial Intelligence", https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55.*

[6] *C. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer-Verlag, 2006. [42] N. M. Nasrabadi, "Pattern recognition and machine learning," J. Electron. Imag., vol. 16, no. 4, pp. 1–2, Oct. 2007.*

[7] *A. Banchs, D. M. Gutierrez-Estevez, M. Fuentes, M. Boldi, and S. Provvedi, "A 5G Mobile Network Architecture to Support Vertical Industries," IEEE Commun. Mag., vol. 57, no. 12, pp. 38–44, 2019.*

[8] *S. Zhang, "An Overview of Network Slicing for 5G," IEEE Wireless Commun., vol. 26, no. 3, pp. 111–117, 2019.*

[9] *A. E. Kalor, R. Guillaume, J. J. Nielsen, A. Mueller, and P. Popovski, "Network Slicing in Industry 4.0 Applications: Abstraction Methods and End-to-End Analysis," IEEE Trans. Ind. Informat., vol. 14, no. 12, pp. 5419–5427, 2018.*

[10] *F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin driven product design, manufacturing and service with big data," Int. J. Adv. Manuf. Technol., vol. 94, nos. 9–12, pp. 3563–3576, Feb. 2018.*

[11] *N. Mohammadi and J. E. Taylor, "Smart city digital twins," in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Nov. 2017, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8285439.*

[12] *MITRE ATT&CK, https://attack.mitre.org/.*

[13] *"IEEE Standard for Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management," in IEEE Std 1900.1-2019 (Revision of IEEE Std 1900.1-2008) , vol., no., pp.1-78.*

[14] *DARPA Spectrum Collaboration Challenge (SC2). [Online]. Available: https://www.darpa.mil/about-us/timeline/spectrum-collaboration-challenge.*

[15] *DARPA's Grand Challenge Is Over—What's Next for AI-Enabled Spectrum Sharing Technology? [Online]. Available: https://spectrum.ieee.org/tech-talk/telecom/wireless/with-darpas-spectrum-collaboration-challenge-completed-whats-the-next-step-for-spectrum-shari.*

[16] *D. C. Marinescu. "Cloud computing: theory and practice". Elsevier, USA, 2013.*

[17] *Connected, "4 Ways AI is Improving Cloud Computing", https://community.connection.com/4-ways-ai-is-improving-cloud-computing/.*

[18] *Yuefeng Ji, Rentao Gu, Zeyuan Yang, Jin Li, Hui Li, and M. Zhang, "Artificial intelligence-driven autonomous optical networks: 3S architecture and key technologies," SCIENCE CHINA Information Sciences, vol. 63, no. 7, p.160301:1–160301:24, 2020.*

[19] *Danshi Wang, Mengyuan Wang, Min Zhang, Zhiguo Zhang, Hui Yang, Jin Li, Jianqiang Li, and Xue Chen, "Cost-effective and data size–adaptive OPM at intermediated node using convolutional neural network-based image processor," Opt. Express 27, 9403-9419 (2019.*

[20] *Che-Yu Liu , Xiaoliang Chen , Roberto Proietti and S. J. Ben Yoo, "Evol-TL: Evolutionary Transfer Learning for QoT Estimation in Multi-Domain Networks," 2020 Optical Fiber Communications Conference and Exhibition (OFC), 2020, pp. 1-3.*

[21] *F. Darema, in Panel on: DDDAS in the New Age of Data Analytics; INFORMS2020, Nov 7, (2020).*

[22] *F. Darema, E. Blasch, S. Ravela, A. Aved The Dynamic Data Driven Applications Systems (DDDAS) Paradigm and Emerging Directions; Handbook of Dynamic Data Driven Applications Systems, Volume 2, F. Darema, E. Blasch, S. Ravela, A. Aved (eds.), Springer 2022.*

[23] *European Parliament and the Council of the European Union (25 November 2015). "Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/.*

[24] *Yoo, Christopher S. and Lambert, Jesse, "5G and Net Neutrality" (2019). Faculty Scholarship at Penn Law. 2089. https://scholarship.law.upenn.edu/faculty_scholarship/2089.*

[25] *X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu and S. J. B. Yoo, "DeepRMSA: A Deep Reinforcement Learning Framework for Routing, Modulation and Spectrum Assignment in Elastic Optical Networks," in Journal of Lightwave Technology, vol. 37, no. 16, pp. 4155-416.*

[26] *5G Security Challenges and Opportunities: A System Approach, A. Dutta, E. Hammad, IEEE 5G World Forum 2021.*

[27] *Rentao Gu, Zeyuan Yang, and Yuefeng Ji, "Machine learning for intelligent optical networks: A comprehensive survey," Journal of Network and Computer Applications, vol. 157, p. 102576, 2020.*

[28] *X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu and S. J. B. Yoo, "DeepRMSA: A Deep Reinforcement Learning Framework for Routing, Modulation and Spectrum Assignment in Elastic Optical Networks," in Journal of Lightwave Technology, vol. 37, no. 16, pp. 4155-416.*

[29] *Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Aguera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR.*

[30] *Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective". IEEE Transactions on Knowledge and Data Engineering. Available at: https://arxiv.org/pdf/1811.03402.pdf.*

[31] *I. G. Terrizzano, P. M. Schwarz, M. Roth, and J. E. Colino, "Data wrangling: The challenging yourney from the wild to the lake," in CIDR, 2015.*

[32] *V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," in KDD, 2008, pp. 614–622.*

[33] *S. Subramanian, "Modern AI Stack & AI as a Service Consumption Models". [Online]. Available: https://medium.com/clouddon/modern-ai-stack-ai-service-consumption-models-f9957dce7b25.*

[34] *T. Schardl and S. Samsi, "TapirXLA: Embedding Fork-Join Parallelism into the XLA Compiler in TensorFlow Using Tapir", in Proceedings of the 2019 IEEE High Performance Extreme Computing Conference (HPEC), 24-26 Sept. 2019.*

[35] *Horovod. [Online]. Available: https://github.com/horovod/horovod.*

[36] *Ludwig. [Online]. Available: https://github.com/uber/ludwig.*

[37] *Adlink. [Online]. Available: https://www.adlinktech.com/en/AI_Training_Platform.*

[38] *Elastic Deep Learning. [Online]. Available: https://github.com/elasticdeeplearning/edl.*

[39] *ForestFlow. [Online]. Available: https://forestflow.ai/.*

[40] *NumPy. [Online]. Available: https://numpy.org/.*

[41] *OpenCV. [Online]. Available: https://opencv.org/.*

[42] *OpenNMT. [Online]. Available: https://opennmt.net/.*

[43] *Google AI Platform. [Online]. Available: https://cloud.google.com/ai-platform.*

[44] *Amazon ML. [Online]. Available: https://aws.amazon.com/machine-learning/.*

[45] *Google Cloud AutoML. [Online]. Available: https://cloud.google.com/automl.*

[46] *Amazon SageMaker. [Online]. Available: https://aws.amazon.com/sagemaker/.*

[47] *Microsoft Azure Machine Learning Studio. [Online]. Available: https://studio.azureml.net/.*

[48] *Amazon Alexa AI. [Online]. Available: https://www.amazon.jobs/en/teams/alexa-ai.*

[49] *University of New Brunswick, https://www.unb.ca/cic/datasets/nsl.html.*

[50] *S. Revathi and A. Malathi, "A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection," International Journal of Engineering Research & Technology (IJERT), vol. 2, no. 12, Dec. 2013.*

[51] *O. Ibitoye et. al., "Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks," https://arxiv.org/abs/1905.05137, May 2019.*

[52] *UNSW Canberra, https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iot.php.*

[53] *The Linux Foundation, https://adversarial-robustness-toolbox.org/.*

[54] *M. Shahriar et. al., "G-IDS: Generative adversarial networks assisted intrusion detection system," https://arxiv.org/abs/2006.00676, June 2020.*

[55] *A. Khraisat et. al., "Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class support vector machine," Electronics 2020, Jan. 2020.*

[56] *The Open Group, https://www.opengroup.org/face.*

11. ACRONYMS / ABBREVIATIONS

Term	*Definition*
3GPP	Third Generation Partnership Project
5G	Fifth Generation
6G	Sixth Generation
AI	Artificial Intelligence
API	Application Programming Interface
ASP	Application Service Provider
ASR	Automatic Speech Recognition
AutoML	Automatic Machine Learning
AVS	Autonomous Vehicle System
BBU	Baseband Unit
CC	Cloud Computing
CN	Core Network
CNN	Convolution Neural Networks
DDoS	Distributed Denial-of-Service
DL	Deep Learning
DN	Data Network
DNN	Deep Neural Network
DS	Data Science
DSA	Dynamic Spectrum Access
DT	Digital Twin
E2E	End-to-End
EAP	Edge Automation Platform
eMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
FPGA	Field Programmable Gate Arrays
GAN	Generative Adversarial Network (for Image generation)
GPU	Graphic Processing Units
HPC	High Performance Computing
IaaS	Infrastructure-as-a-Service
IDS	Intrusion Detection System
IEEE	Institute of Electrical and Electronics Engineers
INGR	International Network Generations Roadmap
InP	Infrastructure Provider

Term	*Definition*
IoT	Internet of Things
IP	Internet Protocol
IPS	Intrusion Prevention System
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KPI	Key Performance Indicator
LCM	Life Cycle Management
MAC	Medium Access Control
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MLSA	Machine-Learned Spectrum Awareness
mMTC	Massive Machine-Type Communication
NFV	Network Function Virtualization
NLP	Natural Language Processing
NN	Neural Network
NNI	Network-to-Network Interfaces
NR	New Radio
NS	Network Slicing
NSA	Non-Stand-Alone
NVME	Non-Volatile Memory (Storage) Express (accessed via PCIE interface)
ONNX	Open Neural Network eXchange format
PaaS	Platform-as-a-Service
PHY	Physical Layer
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio access network
RNN	Recurrent Neural Network
SaaS	Software-as-a-Service
SDN	Software Defined Networking
SLA	Service Level Agreements
Smart NIC	Smart Network Interface Card
UE	User Equipment
UNI	User-to-Network Interface
uRLLC	Ultra-Reliable Low-Latency Communications

Term	*Definition*
vBBU	Virtual Baseband Unit
V2X	Vehicle to Anything
VNE	Virtual Network Embedding
VNF	Virtual Network Functions
VPN	Virtual Private Network
WAN	Wide Area Network
WG	Working Group
XAI	Explainable Artificial Intelligence

12. APPENDIX A – SUPPLEMENTAL INFORMATION ON AI/ML WORKFLOW

Data Handling

Data handling is a major bottleneck in ML and an active research topic in multiple communities. As ML is becoming more widely used, it is well known that new applications evolve that do not necessarily have enough labeled data. Also, unlike traditional ML, DL techniques automatically generate features, which reduces feature engineering costs, but may require larger amounts of labeled data. A survey is given for data collection from a data management point of view[30]. Data collection consists of data acquisition, data labeling, and improvement of existing data or models. In the subsequent subsections, these features are explained.

Data Acquisition

The aim of data acquisition is to find datasets that can be used to train ML models. There are three main approaches: data discovery, data augmentation, and data generation, as listed in Table 10[30]).

- ***Data discovery*** is used for sharing or searching for new datasets. The generated data must be indexed and published for sharing. Collaborative analysis or the web can be utilized. People can search the datasets for their machine learning tasks. Data lake (where many datasets are generated internally[31]) and the web are remarkable candidates for searching methods.
- ***Data augmentation*** complements data discovery where existing datasets are enhanced by adding more external data. A common data augmentation is to derive latent semantics from data. In many cases, datasets are incomplete and need to be filled using data augmentation. Finally, data integration can also be considered data augmentation, especially if extending existing datasets with other acquired ones.
- ***Data generation*** can be used when there is no available external dataset, but it is possible to generate crowd sourced or synthetic datasets instead. For manual construction, crowd sourcing is the standard method where human individuals are given tasks to gather the necessary data that collectively becomes the generated dataset. Alternatively, automatic techniques can generate synthetic datasets.[30]

Table 10: Data Acquisition Techniques

Task	Approach	Techniques
Data Discovery	Sharing	Collaborative Analysis, Web
	Searching	Data Lake, Web
Data augmentation		Deriving Latent Semantics, Entity Augmentation, Data Integration
Data generation	Crowdsourcing	Gathering, Processing
	Synthetic Data	Generative Adversarial Networks, Policies, Image, Text

Data Labeling

When enough data has been acquired, the next step is to label individual examples. In many cases, data acquisition is done with data labeling. The different data labeling categories are listed in Table 11:[30]

- ***Use existing labels***: A well-known technique of data labeling is to exploit any labels that already exist. There is an extensive literature on semi-supervised learning where the idea is to learn from the labels to predict the rest of the labels.
- ***Crowd-based***: The next set of techniques are based on crowd sourcing. One approach is to label individual examples. A more advanced technique is to use active learning where questions to ask are more carefully selected. More recently, many crowd sourcing techniques have been proposed to help workers become more effective in labeling.
- ***Weak labels***: While it is desirable to generate correct labels, this process can be expensive. An alternative approach is to generate less than perfect labels (i.e., "weak" labels) in large quantities to compensate for the lower quality. Recently, the latter approach is gaining more popularity as labeled data is scarce in many new applications.[30]

Table 11: Data Labeling Categories

Category	Approach	Machine learning task	Data types
Use Existing Labels	Self-labeled	classification	all
		Regression	all
	Label propagation	Classification	graph
Crowd-based	Active learning	Classification	All
		regression	all
	Semi-supervised+Active learning	Classification	Text
			Image
			graph
	Crowdsourcing	Classification	All
		Regression	all
Weak supervision	Data programming	Classification	All
	Fact extraction	Classification	text

Using Existing Data and Models

An alternative approach to acquiring new data and labeling it is to improve the labeling of existing datasets or improving the model training. The different techniques that can be used for this are listed in Table 12[30] and described below:

- Data cleaning. A major problem in machine learning is that the data can be noisy and the labels incorrect. Data cleaning can be used for improving data quality. Another approach is to improve the quality of existing labels.[32]
- Improve the model. In addition to improving the data, there are also ways to improve the model training itself. Making the model training more robust against noise or bias is an active area of research. Moreover, transfer learning is a popular approach for training models when there is not enough training data or time to train from scratch.[30]

Table 12: A Classification of Techniques for Improving Existing Data and Models

Task	Techniques
Improve Data	Data Cleaning
	Re-labeling
Improve Model	Robust Against Noise
	Transfer Learning

AI/ML Stack

The AI/ML stack is shown in Figure 23. It consists of two components, namely the infrastructure and development components.[33] Next, these components are described.

Infrastructure Component

Infrastructure component are the tools, platforms, and techniques used to run store data, build, and train AI/ML algorithms, and the algorithms themselves. It consists of several layers, as shown in Figure 23.

More specifically, the compute layer concerns the computational resources required to run AI/ML algorithms. Several choices exist for the physical servers, virtual machines (VMs), containers, and specialized hardware (such as GPUs), cloud-based computational resources, including VMs, containers, and serverless computing.

The compilers layer translates AI/ML code written in one programming language into another language. Usually, the latter language is a lower-level language, such as the assembly language, object code, or machine code. Also, the compilers layer creates the executable AI/ML program. Some examples are the NNVM and TVM.[34]

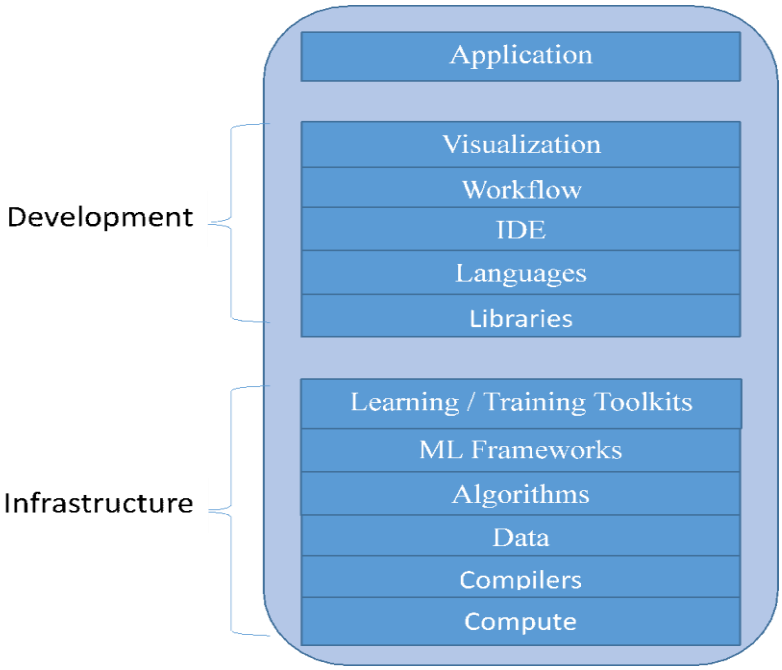

Figure 23: AI/ML Stack

Moreover, data, algorithms and ML frameworks layers comprise the techniques mentioned in Section 0. However, the data layer also comprise database technology for storing the needed data, such as SQL, noSQL, and so on.

Lastly, the learning / training toolkit layer uses all the underlying layers to orchestrate the AI/ML procedure. Some examples are the Horovod[35], Ludwig[36], Adlink[37], Elastic Deep Learning[38], and ForestFlow[39].

Development Component

Developer component refers to the tools that assist in developing code for implementing the AI/ML features. Its layers are depicted in Figure 23. Specifically, the libraries layer is comprised of the software libraries that can be used for developing several AI/ML features. A developer can choose whether to leverage advanced mathematical operations (NumPy)[40] or to add specific cognitive capability, such as computer vision (OpenCV)[41], language translation (OpenNMT)[42], etc.

Any programming language can be used, such as Python, LISP, Haskell and so on, as soon as it is compatible with the underlying technologies. The integrated development environment (IDE) layer makes the job easier for developers as it provides comprehensive facilities to computer programmers for software development. Examples are the PyCharm, Visual Studio Code, MATLAB, and so on. IDEs for AI/ML may not have the advanced debugging capabilities that are used for procedural or object-oriented programming languages.

The workflow layer is important as it makes sharing, collaboration, and automation much easier. As more developers start leveraging AI/ML capabilities, developer collaboration becomes more important. A variety of workflow tools are available, such as Jupyter, Anaconda, GitHub, VSTS, and so on.

The visualization layer plays an important role as it visualizes the functionality of the whole AI/ML stack. Visualization choices include MATLAB, Seaborn, Facets, or data analytics platforms like Tableau.

Migration Based on IaaS

This migration option is based on the IaaS cloud delivery model, shown in Figure 24.

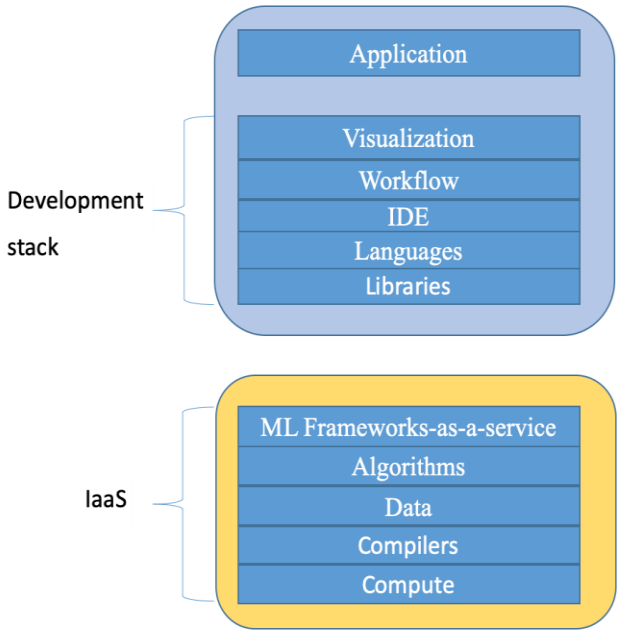

Figure 24: AI/ML Execution in Cloud Based on IaaS

Using this approach, the models, algorithms, types of data stores, compilers and computing resources layers of the infrastructure component are created and deployed in a public / private / hybrid cloud provider. There is no need for managing and maintaining these layers. Moreover, the learning / training toolkit and ML frameworks layers are substituted with the ML framework-as-a-service. This technology consists of all the necessary tools for running AI/ML in cloud (examples are the Google AI platform[43] and Amazon ML[44]).

Migration Based on Managed IaaS

This migration option is based on the managed IaaS, shown in Figure 25.

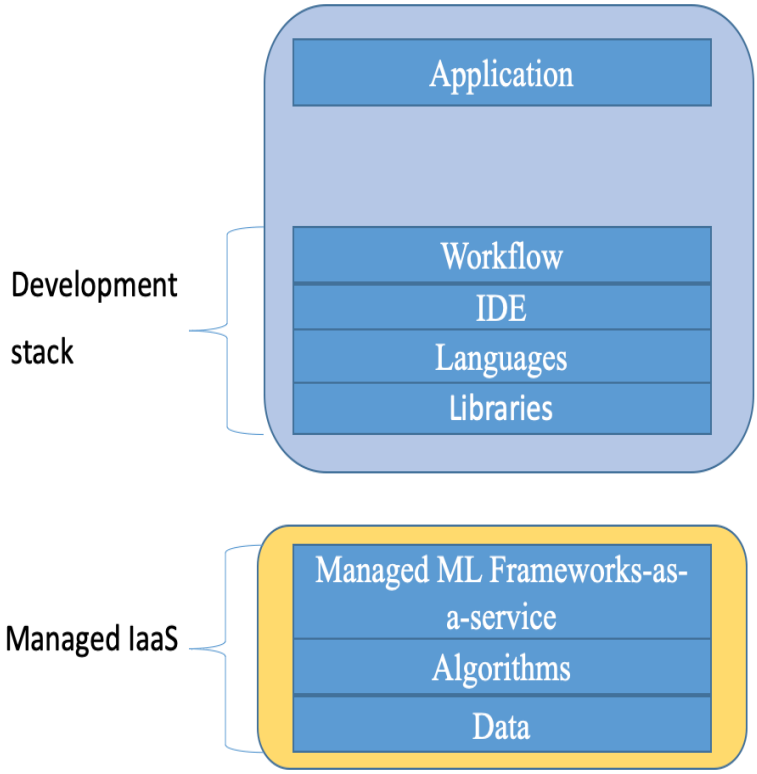

Figure 25: AI/ML Execution in Cloud-Based on Managed IaaS

Using this approach; the models, algorithms, types of data stores, compilers, and compute resources layers of the infrastructure are created by a managed IaaS, which runs on a public / private / hybrid cloud provider. The user does not need to be involved with creating and deploying data stores and compute resources, as described in Section 3.1. These features are handled by the Managed ML Framework-as-a-service. Users choose the technologies that best fit their use case and the appropriate functionality will be automatically created in a cloud infrastructure. Examples are the Google Cloud AutoML[45], the Amazon SageMaker[46], and the Microsoft Azure Machine Learning Studio[47].

Migration Based on Cognition-aaS

This migration option is based on the cognition-aaS, shown in Figure 26.

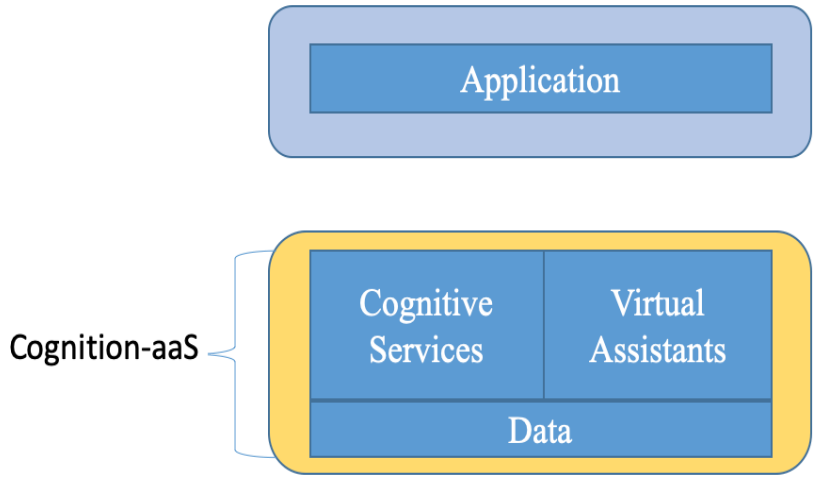

Figure 26: AI/ML Execution in Cloud-Based on Cognition-aaS

Using this approach, the user selects the required cognition-aaS; such as image, vision, speech, language, conversation, analysis, and so on. These services are offered by the cloud provider. Then, all the required development and infrastructure components are created and deployed automatically in the cloud infrastructure. Eventually, the user utilizes the services to accomplish specific tasks. An example is the Amazon Alexa service[48].

13. APPENDIX B — SUPPLEMENTAL INFORMATION ON AI/ML FOR SECURITY

Network security using AI/ML is not an entirely new topic. For years, companies have been implementing algorithms for email spam detection and filtering. However, the use of AI/ML for network intrusion detection is much more recent and its application to 5G even more so. Early research used different ML techniques for intrusion detection using the NSL-KDD dataset[49] with unsupervised Random Forest and Support Vector Machine (SVM) clustering methods[50] in 2013. Deep learning was used for intrusion detection in IoT networks[51] using the BoT-IoT dataset[52] from the University of New South Wales (UNSW) Canberra. The authors used the Adversarial Robustness Toolbox (ART)[53] to demonstrate that Adversarial can greatly reduce the IDS accuracy. A Generative Adversarial Network (GAN) was used to mitigate the effects of having a limited dataset for network intrusion detection[54]. The approach was to train a neural network to recognize attacks using data generated from a GAN. Ideally, the GAN would learn the attacks' model parameters, which would then be used to train the NN classifier to a higher degree of precision. The older, and more problematic dataset, KDD99, was employed. A hybrid intrusion detection system was developed[55] with the goal of recognizing both well-known and zero-day attacks. The authors tested their approach against multiple datasets that included the NSL-KDD and ADFA datasets.

The importance of dataset availability for AI/ML cannot be understated. As discussed above, an early dataset from 1999 formed the backbone of algorithmic verification of network intrusion detection in the 2010s. Fortunately, more attack data has come in and there are other datasets available. A prime example is the datasets from the Canadian Institute for cyber security, listed below. These datasets will allow the development and verification of advanced future AI/ML security algorithms.

- CCCS-CIC-AndMal2020
- DNS over HTTPS (CIRA-CIC-DoHBrw2020)
- CICMalDroid 2020
- Darknet 2020
- Investigation of the Android Malware (CIC-InvesAndMal2019)
- DDoS Evaluation Dataset (CIC-DDoS2019)
- IPS / IDS dataset on AWS (CSE-CIC-IDS2018)
- Intrusion Detection Evaluation Dataset (CIC-IDS2017)
- Android Malware Dataset (CIC-AndMal2017)
- Android Adware and General Malware Dataset (CIC-AAGM2017)
- DoS dataset (application-layer) 2017
- VPN-non-VPN traffic dataset (ISCXVPN2016)
- Tor-nonT or dataset (ISCXTor2016)
- URL dataset (ISCX-URL2016)
- ISCX Android Botnet dataset 2015
- ISCX Botnet dataset 2014

- ISCX Android Validation dataset 2014
- ISCX IDS dataset 2012
- ISCX NSL-KDD dataset 2009

Intrusion detection is not the only area where AI/ML can be used for 5G and future networks. Traditional areas such as authentication still apply, but the manner in which the algorithms and the associated data, are used will vary greatly. For example, monitoring the physical properties of a device, such as the parameters of the RF oscillator or other components, may be used for authentication as well as to mitigate man-in-the-middle attacks. Similarly, attacks against computing systems that include ransomware may be mitigated by monitoring the performance of the operating system. For example, understanding the nature of a ransomware attack, in addition to the underlying parameters that define the attack method, may lead to algorithms that can predict future or zero-day attacks. Hence, improved 5G cyber security using AI/ML is an open area that requires significant development in the future.

Future networks will need to be more sophisticated and adaptive than those in today's environment. For example, they tend to be single-point systems made to address a single solution in a particular environment and will not be able to deal with security problems in the future. Rather, a cooperative AI/ML-based security system is needed that can adaptively monitor its parameters in real-time and share information accordingly. These self-adapting systems will be able to mitigate threats more quickly than traditional systems.

In an AI/ML system, the large amount of data normally collected and models developed will not be sufficient for future networks. Instead, AI/ML models and techniques will have to be developed in real-time based on information collected from other AI/ML and non-AI/ML-based systems. In addition, AI/ML systems must be secure enough so model parameters cannot be easily changed by an attacker. In the event that a compromised source has updated a model or its parameters, the system should self-recognize the security breach.

A key component of the security cloud is open software and interfaces that allow disparate systems to share information. Using this type of architecture, a machine learning orchestration (MLO) can be utilized to distribute information to all components of the security cloud. Such open-source technology is common in other environments. One example, Future Airborne Capability Environment (FACE)[56] for aircraft multi-sensor integration by the United States Department of Defense, focuses on sensor fusion, but the same approach can be taken for AI/ML-based security orchestration.

Open software will allow different systems to collect and share information. In this way, each device can be seen as a security sensor. With distributed security monitoring, it is possible to detect cyberattacks in real-time, which will lead to improved cyber security situation awareness.

14. ANTITRUST STATEMENT

Generally speaking, most of the world prohibits agreements and certain other activities that unreasonably restrain trade. The IEEE Future Networks Initiative follows the Anti-trust and Competition policy set forth by the IEEE Standards Association (IEEE-SA). That policy can be found at: https://standards.ieee.org/wp-content/uploads/2022/02/antitrust.pdf