OMDIA

Using AI in the RAN: How AI addresses the challenges in the RAN

Utilizing data driven intelligence to automate RAN operations and management

Publication Date: 24 Apr 2020 | Product code: SPT001-000106

Adaora Okeleke

Summary

In brief

The top business challenge for communications service providers (CSPs) is in reducing operations cost and complexity. According to Omdia's 2020 ICT Enterprise Insights survey of over 400 senior ICT managers in telecoms, almost 80% of respondents ranked reducing operations cost and complexity as their most important or second most important business challenge to address in 2020. The radio access network (RAN) plays a critical role in CSPs' overall spend for the mobile network infrastructure (with RAN equipment accounting for approximately 20% of mobile operator's capex). Optimizing spend in managing and operating the RAN will contribute to CSPs' addressing their top business challenge. However, achieving this cost savings will require CSPs transforming how the RAN is managed.

This report investigates the role that AI will play to enable CSPs to reduce the costs and improve the operational efficiencies of the RAN. It looks at the current challenges faced in managing the RAN, the impact of these challenges, and how AI can address these challenges. We also provide a list of AI use cases for the RAN and an analysis of these use cases to answer critical questions CSPs will have as they look to implement AI technologies within the RAN.

Omdia view

- **The RAN is increasingly complex**. It has undergone several changes to improve spectral and power efficiency, quality of services, and reduce costs. CSPs however run legacy and modern RAN infrastructure together making it difficult to manage the RAN. Poor visibility into the network performance, increases capex and opex costs and inefficient use of spectrum and radio resources occur as a result of this complexity.
- **Humans alone can't manage the RAN**. Automation has become critical to operating the RAN as humans do not have the capacity to manage its increasing complexity. Self-organized network (SON), which is an attempt at automating operational functions in the RAN, has not effectively achieved its purpose as it is reactive and relies heavily on human operators to execute its functions.
- **AI presents an intelligence driven approach to automating the RAN**. AI's speed of processing, computation, and learning makes it suitable to address the issues currently faced in managing the RAN, enabling CSPs to improve RAN performance and experience delivered to customers.
- **Clarity on the characteristics of AI use cases will influence how they are deployed in the RAN**. CSPs should invest in using AI within the RAN. However, they need to fully understand the key characteristics of these use cases to identify how to prioritize and provide the necessary resources needed to implement them.

Recommendations

Recommendations for CSPs

- **Develop a clear strategy for managing legacy and next generation RAN environments**. Co-existing legacy and modern RAN equipment will remain for some time, and so will issues associated with such a complex environment. These issues particularly the access to data from proprietary RAN equipment needs to be addressed to facilitate RAN AI implementation.
- **Look beyond performance monitoring AI use cases**. While performance monitoring use cases present the easiest path to implementing AI in the RAN, the benefits are limited. Utilizing AI to perform closed loop automation presents a much bigger opportunity to transform RAN operations from being reactive to being more proactive. This change will take CSPs much closer to achieving their objectives of improved customer experience and operational efficiency and reduced costs.
- **Collaboration within RAN organization is critical**. To successfully implement centralized RAN AI use cases, current silos within the RAN organization need to be removed to provide greater visibility into operations performed across teams. This visibility will facilitate impact assessment of recommendations provided by AI before they are implemented. CSPs must transform to a more collaborative network organization.

Recommendations for vendors

- **Create and commit to a roadmap to use AI to enhance RAN and SON functionalities**. As vendors invest in developing 5G RAN equipment, they also need to prioritize the use of AI in this equipment (including upgrades to 4G). SON solution providers should fast track implementation of AI in the SON to fulfil the capabilities they were developed to achieve.
- **Collaborate with CSPs**. Vendors must collaborate with CSP customers to accelerate the implementation of AI within the RAN. Actively engage with them to identify how to address roadblocks that will impact the pace of adoption of AI including defining and simplifying access to required data sets. For example, vendors should aim to implement standard interfaces and APIs with their RAN equipment to ease access to data.
- **Be ready to perform exhaustive testing of RAN AI solutions.** The RAN accounts for about 20% of CSPs' mobile capex spend and so plays a critical role in enabling services they deliver. While AI comes with several benefits, CSPs will demand high performance levels from these solutions. Vendors must be ready carry out as much testing and trials to validate the functions on these AI based solutions. Vendors also need to prepare an effective maintenance plan to ensure AI models remain relevant to RAN operations.

Key Issues in the RAN

CSPs' RAN are excessively complicated to manage

The CSP RAN is a critical component of service providers' networks, which directly impacts on their ability to meet business objectives. The RAN as covered by Omdia, includes base stations (which could be macro base stations (BTSs) or small cells both indoor and outdoor), radio controllers, and small cell gateways. The RAN provides the radio coverage, and capacity, needed to deliver mobile based services to CSP subscribers. According to Omdia research, the RAN equipment accounts for approximately 20% of mobile operator's capex.

RAN architectures and technologies are changing to create new efficiencies and support new service types. These changes include:

- RAN architectures are slowly becoming more centralized for better site to site coordination to improve spectrum efficiencies and simplify network operations.
- Virtualization and disaggregation of the RAN software allowing for RAN vendor software to run on COTS. This will allow for greater network flexibility including the separation of the control plane from the user plane.
- Increasing the number of spectrum bands in use and the optimization of those bands for best performance.
- Growth in the number of sites and deployed radio equipment increases the effort and costs required to manage the RAN.
- Introduction of dynamic spectrum sharing (DSS). DSS allows for multiple radio access technologies (RATs) to be used in a single spectrum band, adding more computation resources in the BBU.
- Changes in traffic type due to changes in devices and service delivered over the network. The RAN needs to detect and distinguish traffic for a broader set of service and cater to their various requirements

CSPs' approach to co-existing legacy and next generation RAN infrastructure supporting different radio access technologies (RATs) has led to the RAN environment being convoluted and difficult to manage. 2G and 3G networks are still in operation today as they are relevant to services such as IoT. While CSPs have indicated the switch over of 3G to 4G to carry voice services, this switch-over is occurring at a much slower rate than planned. VoLTE also requires another level of optimization not needed for LTE data, as voice requires consistent performance throughout the network. Data traffic is not as sensitive. The increased complexity of the RAN has led to a conundrum of challenges summarized in Table 1.

Table 1: Overview of key challenges faced by CSPs following changes to the RAN

Key changes	Additional comments
Poor visibility into network and service performance	Due to the complex RAN, monitoring, analyzing, and optimizing the performance of the network has become difficult. This has led to other issues such as an inability to track and improve network quality, resulting in reduced quality of experience (QoE) for customers. Addressing security risks is challenging as limited visibility into network traffic hampers CSPs' ability to detect security threats and take appropriate actions.
Increasing power/energy consumption at network base stations	As CSP RAN footprints, wireless traffic, and demand for data services increase so do their energy requirements. AT&T illustrates this rise in data traffic by highlighting that it has seen a 470,000% increase since 2007 in the data flowing through its wireless network. These energy levels will continue to rise with 5G. The GSMA record that data traffic could increase up to 1,000 times or more with 5G, and the accompanying infrastructure could require two to three times as much energy. While 5G base stations will come with enhanced energy efficiency capabilities (reducing energy consumption per bit per hertz), the rise in data traffic and wider range of frequency bands will lead to increased energy levels. This will be further increased by the need for 5G infrastructure to work with legacy networks.
Challenging operations and management of the RAN network	As the number of BTS and traffic increases, even so will the complexity in optimizing and managing these networks. CSPs are finding it more difficult to optimize networks to reduce congestion, release capacity, and provide better coverage. They must find ways to address the optimization challenge as their ability to meet growing customer demands for higher QoE (for services, such as video, gaming and industrial IoT) will depend on their ability to optimize radio resources effectively and in real time.
Inefficient spectral efficiency during network design and planning phases of the network	Radio spectrum is a critical and expensive asset for operators. Sadly, the co-existence of multiple network technologies operating on different spectrum bands with differing characteristics makes achieving spectral efficiency difficult. As spectrum bands increase with the use of mmWave frequency bands to support 5G use cases with high latency requirements, the challenge in managing and maximizing spectral efficiencies will continue to increase.

Source: Omdia

Humans alone can't operate the RAN

Given how complex the RAN is, CSPs can no longer depend on humans alone to manage the exponential increase in the number of network elements and functions currently residing within the RAN. Thomas Hodi, Senior RAN Expert at A1 Group said: "I'm still working with the same team that deployed and manage 2G, 3G, and 4G networks."

“5G is just another G to add onto their skills. The team’s capacity has not increased to meet the new challenges coming with 5G. For these engineers to cope with the added workload the team must be able to automate the repetitive tasks on the legacy and 5G infrastructure.”

Automation will play a critical role in enabling CSPs meet their business objectives despite the intricate nature of the RAN. This demand led to the concept of Self-organizing network (SON). Unfortunately, SON has not been effective in addressing all the challenges for which it was developed; a view shared by CSPs like AT&T, A1 Telecom, BT, and Vodafone.

SON has not been effective at addressing RAN issues

The concept of SON was introduced by the Next Generation Mobile Network group (NGMN) and standardized by the 3GPP to utilize automation to improve performance of the RAN. The main objectives of SON are to simplify the configuration, optimization and healing procedures of the network by automating these tasks; thereby reducing CAPEX and OPEX.

While SON has fulfilled some of these objectives, it is wanting in other areas. SON functions, such as automatic neighbor relation (ANR) and plug and play (PnP), have been deployed by CSPs with success. However, other use cases that have failed as existing methods used in current SON deployments are limited in their ability to achieve self-configuration, self-optimization, and self-healing. These capabilities are desperately needed to manage the functions and health of a highly complex RAN environment. Figure 1 provides a summary of the limitations of today’s SON deployments.

Figure 1: Limitations of today’s SON deployments

Use of limited and non-real time data sets	• Data is collected, processed and aggregated at 15 mins to 1hour intervals. This cycle is too long for a CSP to detect any faults or degraded service occurring on the network.
Reactive operations	• With SON being slow to detect network issues, resolution becomes reactive. While being reactive does not impact use cases such as network planning, it does for use cases like network optimization that increasingly needs to operate in real time to fulfil latency, capacity, reliability, and other service quality requirements for services like video and gaming services .
Limited use of historic data	• Non storage of large and long-term historic data sets.
Depends on manual functions	• For example, analysis of measurement reports for issue detection, configuration, and adjustment of base station parameters for network optimization.
Uses rule-based approach for automation	• Where functions are automated, static rules are used which are difficult to maintain.
Reliance on humans	• To observe alarms and use monitoring software to perform self-healing.
Challenges with multi-vendor network environments	• Due to the proprietary interfaces that come with RAN equipment, SON solutions are limited to supporting networks with fewer RAN providers.

Source: Omdia

These limitations vary depending on how the SON solution is deployed. Centralized SON (C-SON) uses optimization algorithms that reside centrally in the network. Key limitation of C-SON is long cycle time for data collection and analysis; resulting in reactive performance of this deployment model. With distributed SON (D-SON), *SON* functions are distributed among the radios to rapidly configure the physical cell identity, transmission frequency, and power. The cycle time for analysis is shorter compared to C-SON but data used for analysis is limited to small quantities of local and neighbor data. Actions are therefore not optimal from a network-wide perspective.

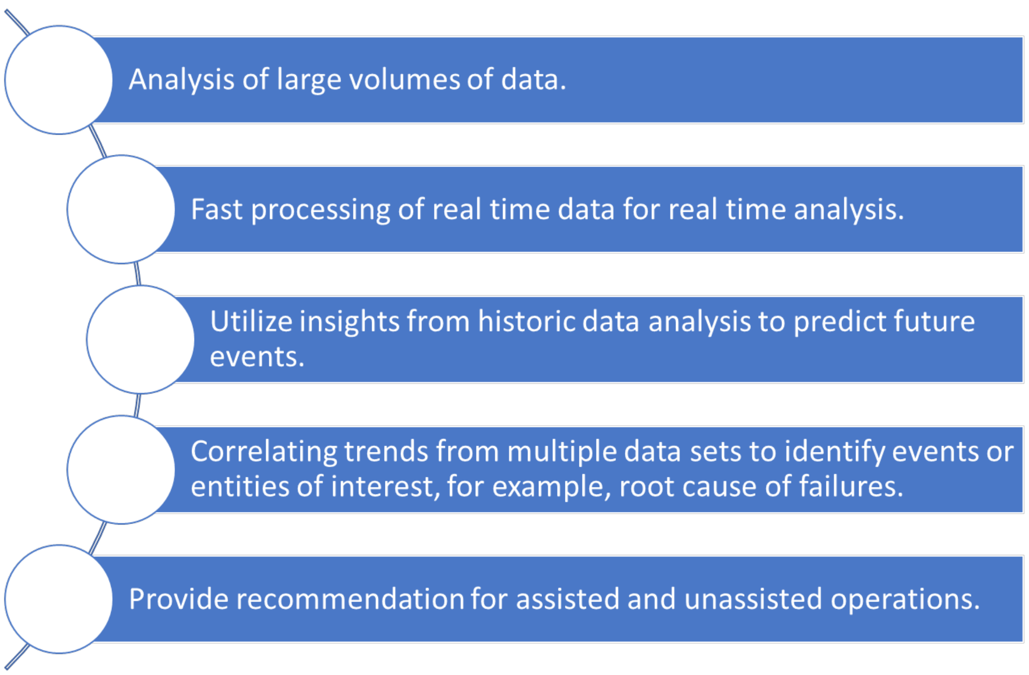
Both deployment models reflect an inefficient use of available RAN data, limiting the level of intelligence from the RAN that's generated and utilized to achieve automated functions in SON solutions. Some CSPs like A1 Group are utilizing software development kits (SDKs) from existing SON solutions to extend their SON use cases. However, scaling these use cases is impacted by factors such as access to data, quality of the data, speed of analysis, and interoperability between RAN equipment. These factors are direct consequence of the challenges of a multi-vendor environment. They also indicate CSPs require a new approach to automating their RAN environment, with the capability to function despite the challenges posed by the continuous evolution of the RAN.

AI as a key technology to address RAN issues

Key capabilities that AI brings to the RAN

CSPs require a combination of technology capabilities, including intelligence to address problems they face with the RAN. Intelligence from RAN and non-RAN related assets will enable CSPs gain better control and management over the complex RAN. AI is a key asset for CSPs that can provide the required intelligence for managing and controlling the RAN, as outlined in Figure 2 below:

Figure 2: Key capabilities of AI relevant to RAN operations and management



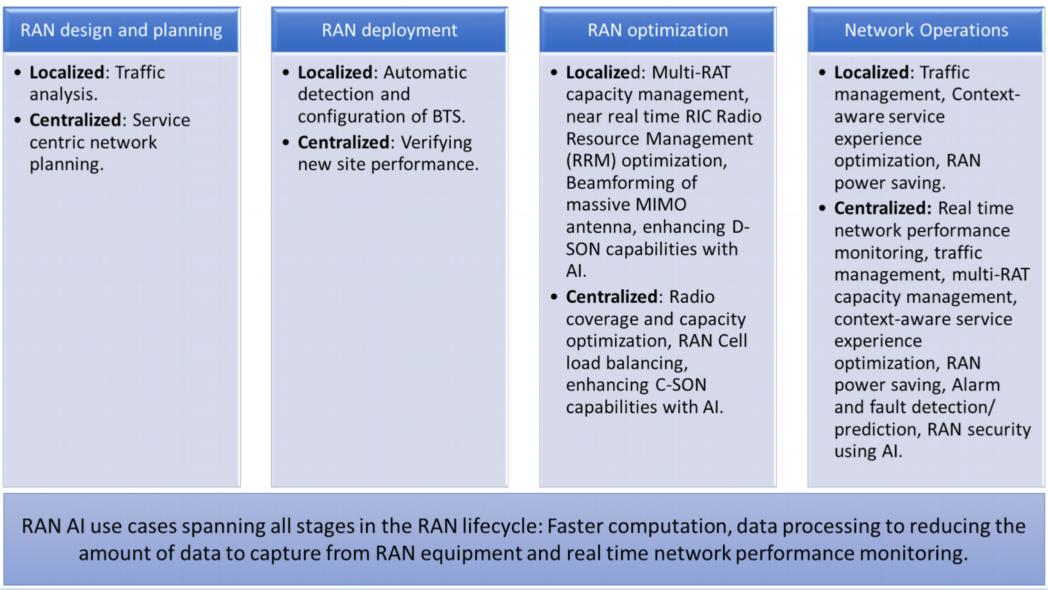
Source: Omdia

Previously, standard analytics tools were used to provide similar capabilities in the RAN. However, as the complexity of today's RAN network grows, more advanced capabilities are required to meet CSPs' and their customers' requirements. In Omdia's report, *Using AI to improve CSP Network Operations: Use Cases*, we highlighted key characteristics of the telco network environment that makes the implementation of AI suited to addressing key telco network challenges. These include generating massive amounts of data, with the data having time series distribution which changes continuously over time and consisting of multiple variables with non-linear relationships. The RAN environment presents similar characteristics.

As a result, AI can be applied to the RAN environment as it is difficult to find a deterministic function (as in the case of standard analytics tools) to model the changes related to a specific condition or state of the RAN network. AI systems can however be used to collect the relevant data sets, use one or more AI mechanisms to analyze the data, and then predict the relevant information or insights to address specific issues on the network.

CSPs as well as vendors have been investigating and developing use cases where these AI capabilities can be implemented. Examples of these use cases and where they are implemented within the RAN (local or central) and its lifecycle are included in Figure 3.

Figure 3: RAN AI use cases



Source: Omdia

A number of these solutions are in commercial deployment while others are still under development. Please refer to Table 2 in the Appendix for a list and explanation of examples of these use cases.

Deep dive into RAN AI use cases

Critical assessment of RAN AI use cases is required to facilitate adoption

Analyzing use cases for AI in the RAN reveals important information that will help CSPs and vendors gain answers to the following questions:

- What are the stages involved in developing these AI use cases?
- Where are these use cases likely to be deployed?
- Which use cases are likely to be deployed first?
- Would all AI use cases require the same data infrastructure to implement them?
- Who—the vendor or the CSP—is best placed to implement any specific use case?
- How can operators get access to these AI capabilities with their RAN infrastructure?

The answers to these questions are useful to understand how best to invest and implement AI within the RAN, bearing in mind the likely pitfalls and differences associated with the AI use cases.

Four key stages are involved in developing an AI use case for the RAN

There are four key stages involved in developing AI use cases; the data collection, processing, training and inference phases.

- **Data collection:** is performed using network probes or directly from the RAN equipment as it reports its telemetry data to an element management system.
- **Data processing:** filters andtransforms the raw data captured from the RAN into meaningful formats to train AI algorithms or for analysis by the trained AI model. In some cases, an analysis of the data is performed to identify and store data that's relevant to the training of the AI algorithms.
- **Training phase**: involves creating the AI model to identify events of interests, for example abrupt changes in RAN performance indicating a fault or dip in normal functions of the RAN.
- **Inference phase**: involves the process of deploying the AI model onto the RAN equipment, to then analyze incoming data to look for and identify events it was trained to recognize

The training phase is the only phase that usually occurs offline and outside of the RAN. The other phases can occur either within the RAN or outside of the RAN depending on if the use case occurs in real time or not.

When and where these stages occur is dependent on three key characteristics of the use case; where the required data sets will be sourced, the level and speed of computation needed to execute the use case. For example, use cases that need to occur in real time can have the data collection, processing and inference stages occur within the RAN, while non real time use cases can occur outside of the RAN.

AI use cases to be deployed first will depend on impact to the real time operations of the RAN

ETSI's Yue Wang indicated that there are two ways to categorize RAN AI use cases: performance monitoring AI use cases and performance enhancing AI use cases. The impact of these use cases on the real time performance of the RAN determines which would be deployed first. With CSPs' RAN accounting for a large portion of network CAPEX, CSPs will be very careful regarding how they implement AI within the RAN. We expect CSPs to implement first, AI use cases that present the least impact to their network. As confidence in the capabilities of these AI systems grows to enhance the performance of the RAN, we'll see CSPs accelerate the implementation of use cases that have direct impact on the network (for example closed loop automation).

Performance monitoring and analysis AI use cases

These use cases involve collecting and using pre-trained AI models to analyze the data sets to understand the performance of the RAN. They are quick to implement as they do not interrupt the current operations of the network. Data is collected and analyzed offline with results and recommendations assessed by an engineer before implemented within the network. These performance monitoring use cases often serve as the first step of every AI based network automation (or network closed loop automation use cases).

The immediate value of these performance monitoring use cases includes faster detection of network anomalies. As the number of variables or KPIs to be analyzed grows (as is the case with current RAN infrastructure), the more difficult it is for humans to detect issues quickly and respond to them. For example, Ericsson indicated that for a client in Japan, its machine learning (ML) based solution was able to analyze 80 KPIs from 200 cells in five minutes. Without the ML solution, an engineer could only analyze 8 KPIs from the 200 cells in 4,000 minutes. Ericsson took the same systems and applied it nationwide (192,000 cells) and it was completed in 15 minutes. With AI based solutions for the RAN, more KPIs can be correlated at the same time, anomalies and root causes detected and triggers or recommendations raised for an engineer to act on.

By utilizing real time telemetry data, performance monitoring AI use cases can provide deeper insights into network performance. These insights enable CSPs to be more effective when managing the RAN and deliver better experiences to customers. Figure 3 highlights how Uhana by VMWare (also known as Uhana, a company recently acquired by VMWare), utilizes AI to analyze a larger volume of data which is collected, processed and analyzed in real time.

Figure 3: Real time data insights for deep visibility at subscriber level of network performance

Source: Uhana

Uhana collects and processes at the same time the trace information that the element management system (EMS) instructs all BTS' to send to it. Thus, the vendor's AI system is not reliant on the aggregated data sets fed to the data lake, but on real time data sets collected from the BTS.

Performance Enhancing use cases

These use cases will involve implementing a change within the network and will have a direct impact on current RAN operations. This change or action will be based on the insights and recommendations provided by the AI system. The resulting impact of these actions will be reported to the OSS systems (network management and operations systems). Examples of performance enhancing AI use cases includes use cases where AI is embedded within the RAN to optimize its functions in real time, for example power saving and resource management use cases. SON use cases also fall within this category.

Combining both performance monitoring with performance enhancing use cases yields the most value to CSPs

Implementing both the performance monitoring and enhancing use cases presents the most benefits to CSPs especially when realizing closed loop automation. A quote from a field operations' engineer from a Tier 1 telco operator in the Asia-Pacific region summarizes this value nicely: the AI-based performance monitoring and enhancement solution "clearly improves the quality of our mobile services in the affected areas and saves millions of dollars currently spent annually to address this real problem (that is the radio frequency interference)."

However, due to the direct impact these solutions will have on the network's functionalities, they need to undergo extensive tests and multiple rounds of trials before they are implemented on commercial networks. Vendors and CSPs planning to implement performance enhancing use cases must be prepared to carry out these tests as any failure in the functions of the AI systems has the potential to adversely affect the network and impact customer services.

Latency requirement of RAN AI use cases determines their data infrastructure requirements

A big data lake (for example a Hadoop based big data platform) or fast data infrastructure can be used to execute the RAN AI use cases. However, the specific data infrastructure used should be determined by the latency and speed requirements of each use case.

Use cases deployed at the network core or outside of the RAN, for example an AI based RAN network planning or management tool will require a big data lake infrastructure. The results from these use cases aren't required in real time. The round-trip time needed to move data from the data source to the data lake and into the AI system is long relative to the time required to achieve the more real time use cases. Consequently, the more real time related use cases will rely on fast data infrastructure with real time processing and analysis capabilities.

Who implements a use case depends on how easily they can access data needed to implement the use cases

Both operators and RAN vendors can implement the RAN use cases. However, the ease of access to data and relative expertise required for each use case determines who is best placed to implement a specific AI use case.

CSPs are best placed to implement use cases running on centralized AI platform

Operators are in a better position to implement use cases that require big data (that is data from multiple data sources and RAN equipment) and can run on a centralized AI platform. They have access to all the data sets from the different RAN equipment deployed within the network. Due to the proprietary systems and interfaces that come with RAN equipment and the limitations to data sharing (due to GDPR guidelines) across regions, it is difficult for vendors to deliver these use cases.

Even when a vendor implements these use cases, they need to collaborate with the CSP to provide access to the data sets. Examples of these use cases include network planning, optimization and configuration use cases like SON use cases. A good example of an AI-based solution developed by a CSP is A1 Group's Superior Analytics of RAN (SARA). It is designed to utilize machine learning to ease network planning and management.

Furthermore, use cases that involve intercepting the actions of the network are best implemented by the operator or vendors working closely with the operator. Operators define the network policies across the network. They have an overall view of the network and a better understanding of how an action in one part of the network (within the RAN) will impact other parts of the network (other parts of the RAN or the transport or core network). However, the siloed approach to managing the network will make it challenging for a CSP to implement these centralized use cases. To successfully implement these use cases, CSPs must transform to a more collaborative network organization.

RAN vendors are best placed to implement AI locally within the RAN equipment

RAN vendors are best positioned to implement AI use cases that run directly within the RAN equipment. For example, the radio resource utilization use case, power saving use case (as it relates to the internal functions of the BBU) and antenna beam coordination are best deployed by the vendors. The RAN vendors have a better expertise in handling the internal functions of the RAN equipment. In addition, the data generated within the equipment (based on direct interactions with users) are sufficient to drive the functions of such use cases.

Access to RAN AI capabilities will be primarily through software upgrades

CSPs can get access to these AI capabilities through remote software updates. Vendors like Ericsson for example, have introduced AI capabilities within their RAN equipment for traffic management and carrier optimization. Customers that have installed the vendor's RAN Compute infrastructure can access these features through software updates. Other customers that have not deployed this type of RAN equipment will require site visits or additional equipment. There is also the option for CSPs with dated RAN equipment to get access to these AI capabilities by replacing existing equipment with more modern RAN infrastructure.

Benefits to implementing AI within the RAN

There are several benefits to enjoy by implementing AI within the RAN and these include

- **Informed CAPEX investment**. CSPs can rollout new sites and allocate radio and spectrum assets based on analytical insights.
- **Better experiences delivered to customers.** With the context aware service experience use case, CSPs can optimize resources at the RAN based on specific service requirements.
- **Proactive operations**: Leveraging the predictive capabilities of AI, CSPs can become more proactive and responsive to network issues.
- **Reduced OPEX costs**: By utilizing insights from AI to trigger automated workflows, CSPs can save OPEX costs incurred in the long term. For example, reducing costs in sending field engineers to radio sites to resolve issues on the network.

As CSPs consider investing in AI to improve RAN performance, they must be aware that there are several challenges ahead of them. They must be prepared to address these challenges to successfully implement AI technologies within the RAN. These challenges include access to high quality data sets, lack of interoperability within the RAN, and lack of skillsets to support and manage RAN AI development. Our report on *Using AI in the RAN*: *Identifying and addressing the roadblocks* will provide an overview of these challenges and ways CSPs may address them.

Table 2: Examples of RAN AI use cases

RAN AI use cases	Commentary notes
Network design and planning	
Service centric network planning	Given the broad range of services that run on CSP networks, more criteria need to be considered when planning the network; where the cells will be located, the coverage and capacity that will be provided, the number of people covered within the area and their current service consumption patterns. Building a view of the external environment and its impact on the radio environment will be required in addition to information on the buying power of customers. These insights are relevant to assure ROI of network assets planned within an area. CSPs can utilize the advanced analytics capabilities of AI to create network plans which take into consideration all of these variables. Small cell planning is another use case where machine learning will be useful in identifying precise locations for small cells
Network deployment	
Automatic detection and configuration of base station	By utilizing AI's learning capabilities, the RAN can automatically detect when new hardware devices are added to the network and configure them according to the preset hardware configuration policies. Human intervention in site deployment is minimized, inconsistent parameter configurations between different network elements will also be avoided.
Verifying new site performance	AI can be used for the pre and post deployment verification and acceptance of planned RAN sites. Following the planning of a new site, AI tools can be used by an operator to test network performance for different applications such as VR and online gaming. They can measure throughput, latency, and even stress test the network to make sure the design plans meet their service requirements before the site is deployed. Once deployed, AI can be used by the CSP to accept network from the RAN vendor without having to send workers to each site. This greatly decreases the time involved in the network verification process. For AI to perform these pre and post deployment verification tasks, a digital twin of the site needs to be in place. It is on the basis of the digital representation of these sites (provided by the digital twin) that AI can perform these use cases. Nokia and Huawei are examples of vendors currently utilizing this AI based approach to support their RAN deployment services
Network optimization	

Radio coverage and capacity optimization	CSPs' limited budgets for network expansion, forces them to find ways to optimize RAN assets to meet customer needs. AI can enable radio optimization by enabling the periodic change of RF parameters for cells/sites within its control in-line with the traffic and radio environment. An AI system (e.g. ETSI's ENI system) can leverage ML to analyze and learn what are the right combinations of RF parameters and other factors such as UE location, traffic load etc for each state of the RAN network. Based on the learnt model (which can be continuously optimized), the AI system can then instruct the BBU of the base station on how to adjust the RF parameters to optimize coverage and capacity
Real time network performance monitoring	This is a critical use case spanning all stages of the RAN life-cycle. With AI, CSPs can monitor in near real time the performance of key KPIs either at the radio, cell, site or RAN level. Based on this real time network monitoring, CSPs can predict likely network outages or failures and implement automated remediation to correct them. This approach is particularly relevant to the migration of voice services to VoLTE as CSPs need a consistent network performance to support voice service on 4G
Traffic management	This use case is particularly relevant in achieving spectral efficiencies as well as improving quality of experience. For example, real time intelligence that these AI systems can generate can be utilized to ensure traffic are allocated to the frequencies that provide the best QoE. It can also ensure that traffic is evenly distributed across spectrum bands to maximize this rather expensive resource
Multi-RAT capacity management	CSPs need to ensure that users are allocated to the right RAT that delivers optimal experience for each service they consume. Machine learning algorithms are being used to understand service usage patterns in the network and to detect when users aren't served with the right radio technology. For example, where a 5G user isn't connected to a 5G frequency carrier, the user is reallocated to the right frequency to ensure that QoE is maintained. Nokia indicated that they are deploying this use case with one their top tier CSPs. Ericsson's secondary carrier prediction use case is an example of how the multi-RAT capacity management use case is achieved within the RAN. It utilizes AI techniques to predict coverage on the 28GHz band based on measurements at the serving carrier (for example at 3.5GHz which may be used for 4G). This approach decreased the measurements on a secondary carrier, thus reducing the energy consumption and the delay for activating features like CA, inter-frequency handover and load balancing.
Context-aware service experience optimization	AI can be used to adjust the values for RAN parameters that are specific to the services they support. The VoLTE service RSRP of the RAN is a good example as it is currently configured to be static in the RAN. By adjusting the VoLTE RSRP (which is reliant on several factors, for example mobile terminal type and user location), the service quality can be adapted according to the VoLTE service information instead of dropping the call or experiencing excessive handover to

	legacy mobile network technologies.

RAN Cell load balancing	This is a type of optimization technique that involves the use of AI capabilities to redistribute the cell load evenly among cells, transfer part of the traffic from congested cell and offload users from one cell or carrier or RAT. This kind of optimization aims to increase system capacity and improve user experience
Network operations	
Reducing the amount of data to capture	AI can be used to reduce the noise in data being collected to focus on what most important. By using techniques such as ML, AI systems can enable CSPs focus on the right data sets that provide insights into understanding the state of the network or resolving network issues.
Supporting the shutdown of legacy network	AI can be used to detect 2G and 3G traffic and migrate then (steering traffic away from legacy infrastructure to 4G LTE and 5G RAN infrastructure
Power saving in the RAN	CSPs are very keen on running their networks reliably while reducing energy consumption levels as increased energy consumption drives up operational costs. Utilizing AI's real time analytical capabilities, it can determine cells/sites that aren't carrying traffic and recommend that they be powered off temporarily until traffic picks up. AI systems can predict when these base stations get switched off as it has established that they are inactive at those times. One of the top NEPs indicated that they have developed and embedded an AI solution for automated MIMO sleep in radio base stations that provides 14 percent energy savings when turned on. This use case will also be relevant as RAN gets virtualized. ETSI indicates that servers in the DCs consume about 70% of power in the DCs. When not carrying peak traffic, traffic can be moved across servers to optimize and bring time power levels.
Enhancing SON capabilities with AI	Utilizing AI to extend the capabilities of SON (including C-SON and D-SON deployment types) from being driven by predefined rules and policies that can't be changed in real time. AI can utilize its data driven approach to defining rules of operations to update rules in real time. P.I. Works and Nokia are examples of vendors that are utilizing AI to enhance their C-SON offerings
Alarm and fault detection/ prediction	A pre-trained AI model (adapted to specific cell behavior) can analyze traffic and other base station performance indicators in real time to determine/predict if an alarm or fault is eminent. If an alarm or fault is eminent, alerts are raised to engineers to resolve the issue or where possible automatically resolve the issue by sending recommendations to automation engines or policy platforms to take action. For example, switching served traffic to neighboring site to avoid service degradation.
Securing the RAN using ML	High proportion of internet traffic is encrypted, making it difficult for CSPs to detect malicious attacks targeted at the CSP network. By populating AI models with telemetry data

	from the RAN, such as sequence of packets, packet boundaries, nature of compute operations and memory access patterns, AI systems can perform real time intrusion detection, network isolation and preventive actions on encrypted traffic. This approach secures the RAN even with encrypted traffic traversing the network

AI use cases embedded within the RAN	
Faster compute at the radio level	The fast and robust computational capabilities of AI systems can be used to analyze the large and multi-dimensional data sets generated within the RAN in record time. For example, Ericsson indicated that for a client in Japan, their ML based solution was able to analyze 80 KPIs from 200 cells in 5 mins. Without the ML solution, an engineering could only analyze 8 KPIs from the 200 cells in 4000 mins. Ericsson took the same systems and applied it nationwide (192,000 cells) and it was completed in 15 mins
Embedding AI within the open source RAN	AI techniques such as machine learning are being used to achieve faster response and improved optimization of resources using closed loop automation. Leading CSPs like AT&T, China Mobile and KDDI are working with vendor partners like Nokia, NEC, Commscope, Samsung, Sterlite and Viavi to embed AI within O-RAN based RAN equipment. Examples include China Mobile and Bravocom have developed near real time Radio Intelligent Controller (RIC) Radio Resource Management (RRM) optimization, a video KQI prediction and RRM enforcement xAPP using machine learning. They also developed a non-real time RIC performing dynamic cell-splitting/merging based on the KQI/KPI status of the cell. China Mobile and Nokia also developed AI-empowered QoE assurance of Cloud VR with machine learning based radio available bandwidth and QoE prediction to enable AI-adaptive rate adaptation. Sterlite on the other hand developed a Non-Real Time Radio-Intelligent Controller managing a machine learning algorithm that can predict cell load. This load prediction model is used by a Mobility Load Balancing xAPP in the Near-Real Time RIC Controller to manage the cell load using performance metrics from the E2 nodes under varying mobility scenarios.
Beamforming of massive MIMO antenna	With the introduction of massive MIMO, the scheduler within the radio resource manager of the RAN's BBU can be trained to determine where users are located and concentrate the signal power where the users are. This way energy savings can be achieved as the radio does not need to scan continuously to detect where users are located. Huawei and ZTE are currently developing RAN solutions to utilize AI for beamforming and radio resource scheduling.
Antenna tilting	AI -based antenna tilting can be achieved by utilizing AI techniques to adjust base station antennas' electrical tilt. Using reinforcement learning (RL), an agent running in the core is trained to dynamically control the electrical tilt of multiple base stations. This process is performed jointly across multiple BTS to improve the signal quality of a cell and reduce the interference on neighboring cells in response to changes in the environment, such as traffic and mobility patterns. Coverage and capacity is enhanced, leading to an overall improvement of network performance and QoE for the users while reducing operational costs.

Efficient scheduling of radio resources	Radio resources can be allocated effectively using AI's advanced computation capabilities. With more parameters to consider when allocating a radio resource, it becomes important for the BBU to utilize AI to perform functions such as channels to allocate or reserve to serve customer needs effectively, particularly when they are in motion. Mavenir is an example of a vendor currenting developing AI-based solution to enhance functions of its virtualized RAN offering

Source:

Appendix

Methodology

Extensive primary and secondary research were conducted to support this paper. Interviews and discussions with key CSPs and vendor players were conducted to support this research. CSPs interviewed include AT&T, A1 Telecom, BT and Vodafone. Vendors include Ericsson, Nokia, Mavenir, VMWare, and Parallel Wireless. We also interviewed stakeholders at ETSI (standards organization) and the GSMA (trade organization for CSPs. Insights from Omdia's reports on the RAN and AI in telecoms and survey tools and databases were also used to the views shared in this report.

Further reading

The role of AI in addressing 5G challenges, SPT001-000092 (January 2020)

2020 Trends to Watch: Telecoms Operations and IT, SPT001-000095 (December 2019)

RAN Vendor Update 2019, SPT002-000266 (November 2019)

Using AI in CSP Network Operations: Challenges and Best Practices, SPT001-000052 (March 2019)

Using AI in CSP Network Operations: What Operators Have Deployed, SPT001-000031 (September 2018)

On the Radar: ParallelM MCenter 1.1 manages the lifecycle of ML in production, INT002-000111(May 2018)

Act on Your Intelligence: An AI Roadmap for Service Providers, ME0002-000769 (July 2017)

Author

Adaora Okeleke, Senior Analyst, Telecoms Operations and IT

Chris Silberberg, Research Analyst, Network Software and Infrastructure

askananalyst@omdia.com

Citation Policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia’s consulting team may be able to help you. For more information about Omdia’s consulting capabilities, please contact us directly at consulting@omdia.com.

CONTACT US

omdia.com

askananalyst@omdia.com