

# Accelerating the adoption of AI in programmable 5G networks

# Content

Introduction	3
Industry overview	5
Challenges for the adoption of AI in networks	10
Accelerating AI adoption	14
Conclusion	20
Glossary	22
References	23
Suggested reading	24
Authors	25

# Introduction

Communications service providers (CSPs) strive for relentless efficiency, business agility to address new revenue opportunities, and to meet or exceed customer expectations through a superior experience. This continues with the introduction of [5G programmable networks, \[1\]](#) which enable new revenue-creating opportunities through both enhanced user experience as well as the tailoring of telecommunications networks to provide differential services for both existing and new types of enterprise customers (e.g Industry 4.0, Automotives, Fixed Wireless etc) . The introduction of new technologies as well as additional services for customers, the densification of networks to support macro and micro coverage, and the need to ensure services with differing requirements significantly increases the complexity.

Artificial intelligence (AI) technologies have already matured to the point where CSPs have been applying them to their networks, often starting with non-time-critical processes, and are now applying them to the sensitive parts of their networks that directly impact user experience. The increased complexity of networks due to more services, new network technologies, and massive network densification further necessitates the application of AI in telecommunications networks as operations become more complex.

AI technologies can make many CSPs' system functions more capable as well as enable new system functions and approaches. Some example applications include:

- improving network performance through better radio scheduling, paging and so on
- improving assurance of offered services and resources, moving from reactive to proactive — even in the face of increasing network complexity and heterogeneity
- improving optimization and use of existing resources, such as spectrum, transport, cloud infrastructure and network functionality
- improving experience management through both increased customer understanding as well as increased tailoring of the offered experience
- improving product and service definition, design, planning and offerings
- improving network and performance planning (such as radio, data center location and transport)

The maturing capabilities of AI have resulted in increased attention within standardization and open source communities, both from a purely technology evolution perspective as well as from an architecture definition perspective. While open source and standardization are enablers for increased AI adoption, the fragmentation which occurs in the early phases of

industry specification can hinder adoption due to the uncertainty it creates, which occurs between different industry bodies as well as in different groups within industry bodies.

Consequently, CSPs are facing a number of challenges today regarding which standards to follow, which aspects of open source should be utilized directly or via vendors, how to increase industry alignment for scale while simultaneously allowing for differentiation, how to leverage the scale of public cloud providers, how to collect and manage data, and how to support the Life Cycle Management (LCM) of AI models.

# Industry overview

There are many AI-related activities taking place in the industry, such as those in the IT domain, among standardization bodies, and on the open source front. Both technology and the ecosystem are evolving rapidly. In order to accelerate the adoption of AI, it is important to have an overall view of the industry and establish an understanding of the driving organizations, including the challenges facing them.

## Organization and industry initiatives

Many organization and industry initiatives are relevant for AI and machine learning (ML) in telecommunications. Below is a table outlining such initiatives, ranging from open source to standard development organizations (SDOs).

### 3GPP RAN3

Has studied RAN-centric data collection and utilization. The goal of the study was to examine big data wireless acquisitions and applications for network automation and intelligence, including the definition of the wireless side use case, as well as the process and information interaction required by different use cases.

### 3GPP SA2

Has specified the network data analytics function (NWDAF), which may be distributed or centrally located. The NWDAF collects data from 5G core network functions, applications and the operations support system (OSS) in order to produce insights. NWDAF insights are mainly applied to 5G core networks to enhance their functionality. Optimized data collection and storage has been specified, together with training and ML model retrieval.

### 3GPP SA5

Has defined the management data analytics function (MDAF), which specifies that there can be centralized (end-to-end, PLMN-wide) MDAFs (such as slice assurance) and domain-specific MDAFs (such as those for core management or RAN management domains). SA5 has also specified closed-loop assurance, outlining the monitoring of data, analysis and decisions, and execution of actions.

### ETSI ENI

The European Telecommunications Standards Institute Experiential Networked Intelligence (ETSI ENI) aims to define a cognitive network management architecture using AI with an approach of adding intelligence on top of legacy systems without having to modify them. ENI aims for a cognitive layer for the telco industry by mainly focusing on adding ML, but it does not address intelligence in general.

#### ETSI ZSM

European Telecommunication Standards Institute Zero-Touch Service Management (ETSI ZSM) studies an architecture to support zero-touch (fully automated) management and operations. It recognizes different management domains (representing a separation of concerns) and then describes the services from these domains and integration fabrics (including requested services from the domains). One special domain is the end-to-end service domain. It also specifies closed-loop control (collect, analysis, decide, and act).

#### ITU-T SG13 ML5G

Has studied an architectural framework for machine learning in future networks produced by IMT-2020. This architecture mainly focuses on the abstract machine learning workflow (ML pipeline), including the required functionality to support it, such as the management orchestration unit machine learning function orchestrator (MLFO). The work of ITU-T SG13 is meant to be an overlay to (for example) the 3GPP architecture.

#### LF AI (Acumos)

Is a Linux Foundation AI project with the primary purpose of providing a streamlined environment for AI model sharing and training across different data sets. With the Acumos data collection, analytics, and events (DCAE) adaptor, it is possible to transform Acumos ML models into ONAP-compatible DCAE microservices.

#### LF Networking (ONAP)

The Open Network Automation Platform (ONAP) provides a reference architecture as well as a technology source. The ONAP subsystem DCAE provides a framework for the development of analytics. DCAE is designed for scalability and to be deployed hierarchically, which may support distributed machine learning principles like federated learning.

#### NGMN

Next-Generation Mobile Networks Alliance (NGMN) has recently published a 5G end-to-end architecture framework. It depicts a high-level architecture, including cognitive awareness in closed loops.

#### O-RAN Alliance

Open- RAN Alliance (O-RAN) is working to evolve 3GPP access with principles of openness and intelligence. It describes two environments for where to place intelligence — the non-RealTime RIC (RAN intelligent controller) and a near-RealTime RIC. It has both specifications (O-RAN Alliance) as well as having open source activities (O-RAN Open Source Community) and addressing some model LCM.

#### TMF

TeleManagement Forum (TM Forum), which concentrates on the higher layers of business and service operations, has recently published a technical architecture on autonomous networks. It defines a high-level architecture using intents as a means to control the different layers of the architecture.

Each activity covers a specific part of the intelligence network problem space, with some complementing and some overlapping. It is important to obtain an overall view here which takes consideration of all of the above activities when seeking to establish a holistic and complete architecture. The below figure shows the main responsibilities for the industry activities.

Fundamentally, it is positive that the industry has an interest in standardization; however, the challenge is to avoid inconsistency, double work and fragmentation.

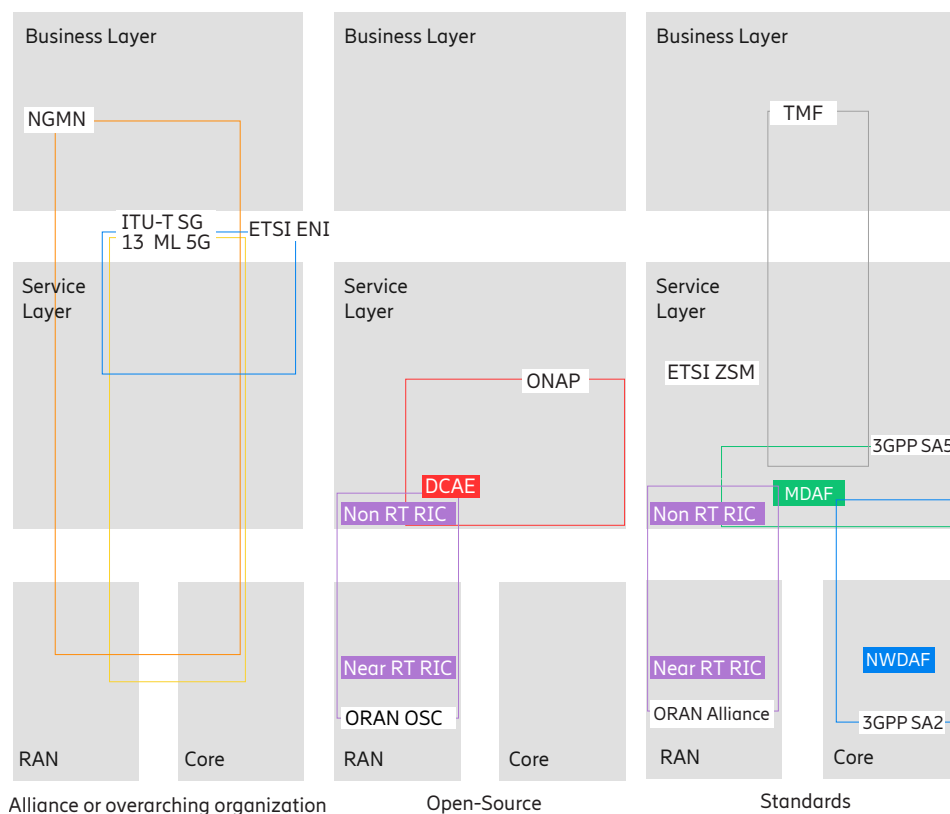


Figure 1. The scope of each standardization organization mapped to the data-driven architecture.

Beyond the industry initiatives represented in table 1 above, there is an active technology evolution for AI, covering both commercial and open source assets. These include, for example, PyTorch and TensorFlow. By adopting and engaging with selected IT open source AI technologies, the telecom industry can leverage the associated IT industry investments.

## Cloud provider initiatives

Public cloud and AI/ML-provider companies are investing in generic AI portfolios offered as platforms as a service, which include services like data labeling, model designing, model training, and execution capabilities, including the frameworks and tools. Due to their large-scale cloud infrastructure and generic AI capabilities and tools, cloud providers are appearing as important partners to CSPs in the ecosystem. An initiative has been undertaken recently by certain CSPs to partner with public cloud providers specifically on AI; however, cloud providers have their own purpose-built LCM processes and do not always follow the telecom best practices.

## Present status of AI in telecom

Table 1 - Industry Overview on page 6 describes the general trends from an industry body perspective. To complement this with a communications service provider perspective, the trends outlined below can be observed.

AI/ML has successfully been applied in a proprietary manner both within existing functions to enhance their performance and capabilities (as in RAN and core networks) as well as within parts of networks that are less specified and tend to rapidly adopt IT technologies, such as operations and business support systems (OSS/BSS) and cloud infrastructure. As telecom service providers industrialize AI technology, the tendency is now growing towards an increased industry specification both in standards and open source for common AI/ML functional architecture for training, inference data management and data collection.

BSS has seen early adoption of analytics in the customer, partner and product business processes — this is in the process of being enriched with AI technologies for use cases such as service-level agreement (SLA) management, customer care, product performance, prediction and subscriber churn, and so on.

Likewise, OSS has seen early adoption of analytics in such areas as network performance, assurance, and experience management, including performance management, fault management and predictive maintenance. A clear trend for horizontal automation platforms with multivendor and multidomain support is emerging that can support common access to data and advanced, real-time AI and ML capabilities.

Traditionally, core networks have adopted AI/ML within their products as well as for the proprietary management and streaming of data, leading to a probe and analytics ecosystem with a tight probe vendor lock-in. Recently, however, there is a clear trend toward increasing the specification of event and management data from core network nodes as well as increased specification of AI core use cases.

As for RAN, the early AI/ML-based software is now running inside the network functions and RAN management systems. Embedded in the network functions, the AI/ML models replace and outperform rule-based software in selective critical subtasks, such as choosing channel coding schemes and beamforming control. In RAN management systems, AI/ML software is used to detect RAN incidents, provide optimization insights and reconfigure proposals. The recent generation of AI/ML software promises more disruptive improvements with a higher degree of network automation and intent-based management, which differs from today's configuration parameter-based network operation.



Cloud infrastructure has relied on largely de facto standards, which is expected in areas where the IT and cloud native ecosystem is strong. There are, however, small steps being taken towards industry alignment among communications service providers, though still independently from cloud service providers.

# Challenges for the adoption of AI in networks

There are various organizational challenges facing the adoption of AI in telecommunications, and while we acknowledge such challenges, the focus in the section below will center on covering the functional aspects of networks. To learn more about the organizational challenges, please refer to the [Adopting AI in Organizations \[2\]](#) report.

## Overall challenge for AI

Beyond the open source and standards industry discussions, the application of AI/ML is being driven by real needs. Hence, both telecommunications service providers and vendors are already including AI/ML capabilities in their portfolios and networks; however, the adoption of AI/ML is at an early stage, and it is, therefore, worth reflecting on the barriers to the rapid adoption of AI/ML. Below are a few examples of these:

- The LCM of AI/ML models introduces new aspects beyond traditional software LCM processes.
- There is a lack of access to data and the management of access to real data due to regulations regarding privacy and ownership. (Relevant data is required to develop and train AI/ML models.)
- Fragmentation and overlap in different standards and open source initiatives continues to diffuse industry focus and create hesitation.
- It takes time to build trust in automation technologies, as some of the conclusions are difficult to explain. A gradual introduction with the appropriate guardrails to allow human oversight and control is required.
- Cloud service providers typically provide their own (different) tools and interfaces, which creates a lock-in effect and challenges CSPs' desire for openness, ultimately slowing down deployment and adoption.
- There is a lack of use cases qualifying returns on investment in the short term.

## AI/ML Life Cycle Management

Compared to traditional software, AI/ML technology introduces the elements of training, model concept drift, federated learning, and a stronger need for access to data. This adds new requirements to the industry Life Cycle Management (LCM) processes for telecom software (meaning developing, validating, delivering, operating and finally retiring software), as shown in Figure 2. The LCM processes set the roles of suppliers, integrators and CSPs in terms of responsibility, accountability, and deliverables between the stakeholders' organizations, stipulating, in essence, who is responsible for what and who sells what to whom. Over the past 20 years, the telecom industry has matured into adoption of a well-working LCM process for traditional licensed software with global acceptance and little fragmentation. As an industry, we must adjust the LCM software to include AI/ML-based technology to reach its potential as it evolves, maintaining a clear separation of concern and, with a minimum of variants, to avoid industry fragmentation. A very high-level AI/ML LCM process is captured in the Figure 2 below.

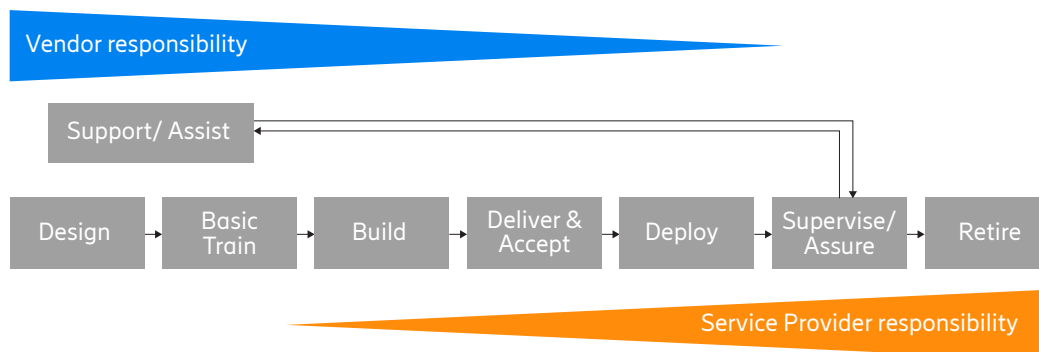


Figure 2. High-level AI/ML LCM process

## Challenges concerning access to data

Access to the relevant data at the right time is key for any analytics system and for developing and training AI/ML models. This requires an infrastructure towards a variety of data points and compute power to process. Unnecessary transfers should also be avoided, as the amount of data can be massive. Filtering and preprocessing close to the data points can greatly reduce the amount of data being transferred through the network.

Initial AI/ML model training is being done by the vendor. This requires access to relevant data. The AI/ML model may need to be re-trained with local data to improve the prediction quality in the target network. Questions related to cost of data, ownership and privacy are important to be agreed between CSP's and vendors and are part of a data ecosystem.

The technical solution must comply with regulatory rules, trustworthiness and CSP policies and the system functions need to support a wide range of flexibility to comply to differences in different countries.

Table 2 outlines where key functionality in an open AI/ML ecosystem need to be available.

	CSP domain	Vendor domain
Data preprocessing capabilities	yes	yes
Real-time data access capabilities	yes	no
ML model development capabilities	yes	yes
ML model training capabilities	yes	yes
ML model inference capabilities	yes	yes
Regulatory compliance capabilities (e.g. data anonymization, privacy, secure access, secure storage)	yes	yes

Table 2: Functional distribution in an open AI/ML ecosystem

## Fragmentation among standardization bodies

The industry significance of AI/ML is reflected in the strong interest that most standardization and open source communities have demonstrated in exploring how to apply AI/ML to their particular scopes and, in addition, their work to claim the lead on certain aspects of the architecture. As described above, there are specification efforts in at least ITU-T, ETSI ENI, ETSI ZSM, 3GPP, ONAP and ORAN. While much of the work is complementary, there is also fragmentation. Fragmentation diffuses focus and creates hesitancy for the adopters (both the network vendors and the CSPs). This hesitancy is driven by the risk of inconsistent standards and the inefficiency of duplicated effort.

One aspect of this fragmentation is visible in the use cases being described as well as in the resulting specified insights created by AI functions described by the SDO/open source organizations. Examples of this include network load and slicing load, which are studied in SA2 with the NWDAF, in SA5 in the MDAF, and in ORAN around the RICs for the same problem. Aside from the inefficiency of specifying the work twice, this also fosters uncertainty for service providers when deciding which approach to adopt.

Another aspect of fragmentation is visible in the efforts that go into describing the different components of AI/ML-enabled functions, such as the inference functionality, training functionality, and data storage functionality. This is present in a number of standardization bodies. While alignment around the basic architecture and concepts is useful to the industry, overspecification can inhibit innovation, and many different specifications slow down adoption.

A further hindering aspect of fragmentation can be found in data collection and management, which refers to the ability to support AI/ML applications to request, collect and receive data (see Ref [3] for more information). This is being studied within 3GPP SA2, 3GPP SA5, ONAP and ORAN and has potential for alignment.

# Accelerating AI adoption

## Taking a use case–driven approach

As seen above, there are fragmented standardization functions proposing overlapping use cases, and, at the same time, CSPs have invested in AI infrastructure over the years. In order to identify how to apply or adopt standards in their networks, it is recommended that CSPs take a value and use case–driven approach (where compelling use cases can be evaluated first), and then how to deliver those use cases from an end-to-end network (contextual) perspective can be studied. An example would be the explanation of how use cases connect from ORAN (rApps, xApps) to SA5 (MDAF, MDAS) to SA2 (NWDAF) specifications. Hence, CSPs might need to engage with partners who take an end-to-end (contextual) approach to analytics to establish a better understanding of domains, data, models, interworking, open source modules and communities.

When it comes to compelling business-driven use cases, analytics use cases can be categorized into three areas, where the primary area is reduction of operational expenses (opex) and capital expenses (capex) and increased efficiency. New technologies require networks to be operated in an efficient manner, and this cannot be possible without utilization of AI. The second area is enhanced customer experience, where CSPs want to differentiate themselves through a better customer experience in their network services. The third area is new revenues, where CSPs offer new capabilities to enterprises or consumers, resulting in new business. The table below provides a few examples of the first wave of use cases (covered in the standards) for CSPs to start evaluating across categories.

Use case	Description	Category
Anomaly detection	Anomaly detections can address a wide range of use cases related to network stability, customer experience or network optimization. The advantage of AI-powered anomaly detection is that it can detect unknown patterns by combining data from different sources and addressing different business scenarios. IoT device anomaly or quality of experience (QoE) functions are examples where AI-enabled anomaly detection can be of great benefit.	Relentless efficiency (opex)
Encrypted video QoE	Advanced traffic analysis to provide proactive assurance for encrypted traffic that would otherwise be difficult to understand. This is based on carefully selected KPIs reflecting the quality of services experienced by the subscribers.	Customer experience
Intelligent RAN Automation (ORAN)	Using ML algorithms, the aim is to automate LCM operations that have a high degree of complexity. Based on customer insights and utilizing RAN data, operations are de-risked, and continuous modifications to networks, upgrades, and the addition of capabilities are automated. More details can be found <a href="#">here</a> .	Relentless efficiency (opex/capex)
Paging optimization	Paging optimization is one example of mobility predictions. Paging is one of the most frequent signaling occurrences in a network. Reducing the number of paging attempts can greatly contribute to better compute resource utilization. Experience has shown that mobility predictions can reduce paging signaling by up to 60 percent.	Relentless efficiency and Customer experience
Predicted mobility	NWDAF mobility predictions provide information on the route a device will take through a network. The most common case is knowledge of the next cell to be entered. There are a number of scenarios where this can be turned into an enhanced customer experience or network optimization. Examples include load reductions (such as paging optimizations) and QoE (such as slice management).	Relentless efficiency and Customer experience
Slice service-level agreement (SLA) assurance	To ensure that SLAs are kept, the monitoring of overall network performance and quality of service on the slice or even user level is needed. When an SLS breach is imminent, automatic enforcements are needed to circumvent the breach and, at the same time, ensure that overall network performance efficiency is high. Detailed monitoring of KPIs on the slice or user level provides an overall view of best enforcements in a network. End-to-end analytics (including RAN, transport, core, and infrastructure) provide the assurance.	New revenue stream

It is key to engage with partners who have an end-to-end analytics view, have domain competence, and can provide guidance and a framework for feasible business-driven use cases.

## Making relevant adjustments to existing LCM processes and avoiding industry fragmentation

De facto—standardized telecom processes save our industry USD billions annually, as both suppliers and CSPs are able to avoid the cost of vendor/customer-specific software LCM tracks and deliverables. As AI/ML technology continues to be added, the industry should continue to avoid fragmentation and strive for de facto standards for LCM processes.

A reasonable augmentation of LCM processes is illustrated in the Figure 3 below.

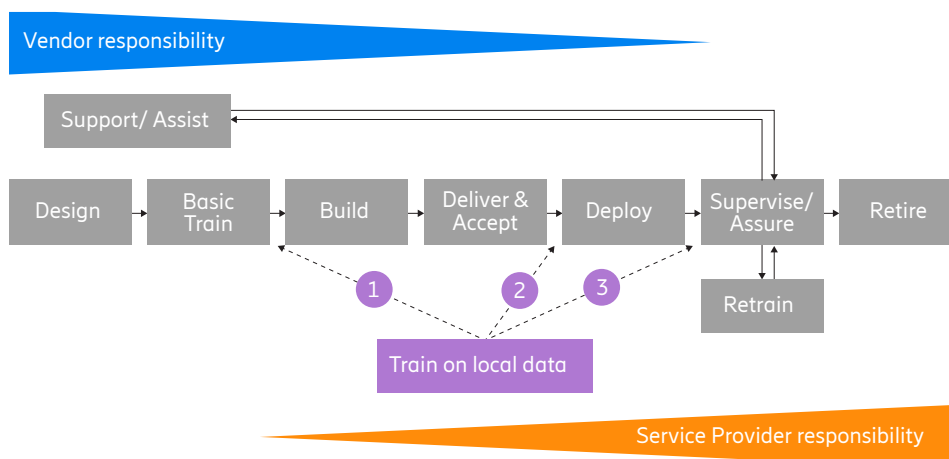


Figure 3. Industry LCM process for vendor-provided software, including AI/ML-based software, AI/ML model supervision, and three alternatives for local training.

The basic training of an AI/ML model is done using a combination of network and simulated data as part of the vendor R&D basic training. Though this may be sufficient for some models, the majority of AI/ML models benefit from further training on local data from the CSP network.

One viable alternative that maintains full vendor accountability is to do training on the local CSP data before delivering the ready-to-deploy software to that CSP (alternative 1 in Figure 3). This can be done either by transferring CSP network data from the CSP domain to the vendor R&D domain or by taking the AI/ML model to the CSP cloud domain, where the CSP network data is available. Regardless of domain, the training on the CSP network data in alternative 1 is done by the vendor.

Another alternative is to deliver the globally trained model to the CSP (alternative 2 in Figure 3), where it is the responsibility of the CSP to do the training on CSP network data in the CSP's domain. As the behaviour of the model changes as a result of the CSP-controlled training, the accountability for the final model performance is split between the vendor (that delivered the base model) and the CSP (that altered the behaviour of the model with training on the CSP's network data). This has implications for support levels and performance responsibility splits between the vendor, integrator and CSP, and the associated accountability principles must be sorted out before this approach is commercialized.

A final option is to do training on network data embedded in the run time (alternative 3 in Figure 3). In this LCM alternative, the globally trained model is part of a larger software deliverable. Once delivered and deployed following existing LCM processes, the system autonomously uses local run-time network data to train the AI/ML model.

We believe that the three alternatives in Figure 3 will mature and find application in the LCM of 5G networks. Alternative 1 gives clear accountability and support agreements but requires either CSP data in the vendor domain or vendor access to the CSP domain for local training. Alternative 2 avoids some of the data domain issues but requires CSPs to invest in AI/ML model training technology and competence and leads to an unclear CSP–vendor performance accountability split. The embedded approach avoids those difficulties but is not suitable for all use cases.

As the industry is working on the above three alternatives to modify the LCM processes to include AI/ML local training, it would be a costly industry mistake to introduce additional LCM fragmentation. Rather, it is beneficial to agree on the vital few LCM alternatives and take the resulting LCM processes as the baseline for AI/ML-related standardization and open source development.

## Optimizing the use of standards and open source

In the field of telecommunications, standards have had a strong role in creating the industry and the ecosystem by defining the functionalities and inter-CSP and multivendor interfaces. These standards have provided long-term guidance to the industry that is independent of the technology and, hence, can survive technology changes. At the same time, open source has moved from creating technology that can be used to build networks according to standards to create ecosystems around default interfaces. This is particularly useful where the technology is evolving rapidly.

One parallel is in the area of the cloud native ecosystem that has evolved around the cloud-native computing foundation (CNCF), which has created de facto interfaces around Kubernetes and is more efficient if standard organizations (such as ETSI NFV) can simply refer to or adopt these interfaces as de facto standards.

For AI/ML, standards have a strong role in and should be promoted for the following cases:

- specifying insights required for certain function scopes to support multivendor consumption of insights (such as those to enhance packet core functionalities in 3GPP SA2 or management insights in 3GPP SA5)
- specifying interfaces for a common way to collect and manage data (3GPP SA5)
- creating a common reference for the AI/ML architecture and LCM while avoiding over-specifying interfaces which will not benefit from multivendor deployments (such as training function or inference function interfaces) due to tight technology dependencies and the rapid pace of technology change



Open source has a strong role and should be promoted for:

- AI/ML technology, platforms and tools
- rapid innovation for different inference and training techniques
- technology for data storage and data storage interfaces
- reference realization of standards interfaces

**Leading standardization and open source bodies**

With the above in mind, some of the leading standardization bodies for the telecommunications industry are:

- 3GPP SA2 for 5G core event data collection and 5G core–related insights for user- and session-related use cases
- 3GPP RAN 3 and ORAN for RAN data collection and use cases
- 3GPP SA5 for domain management and end-to-end use cases (such as slice assurance)t, and management data collection
- TMF for Intent driven management services
- ETSI ZSM for closed loop control

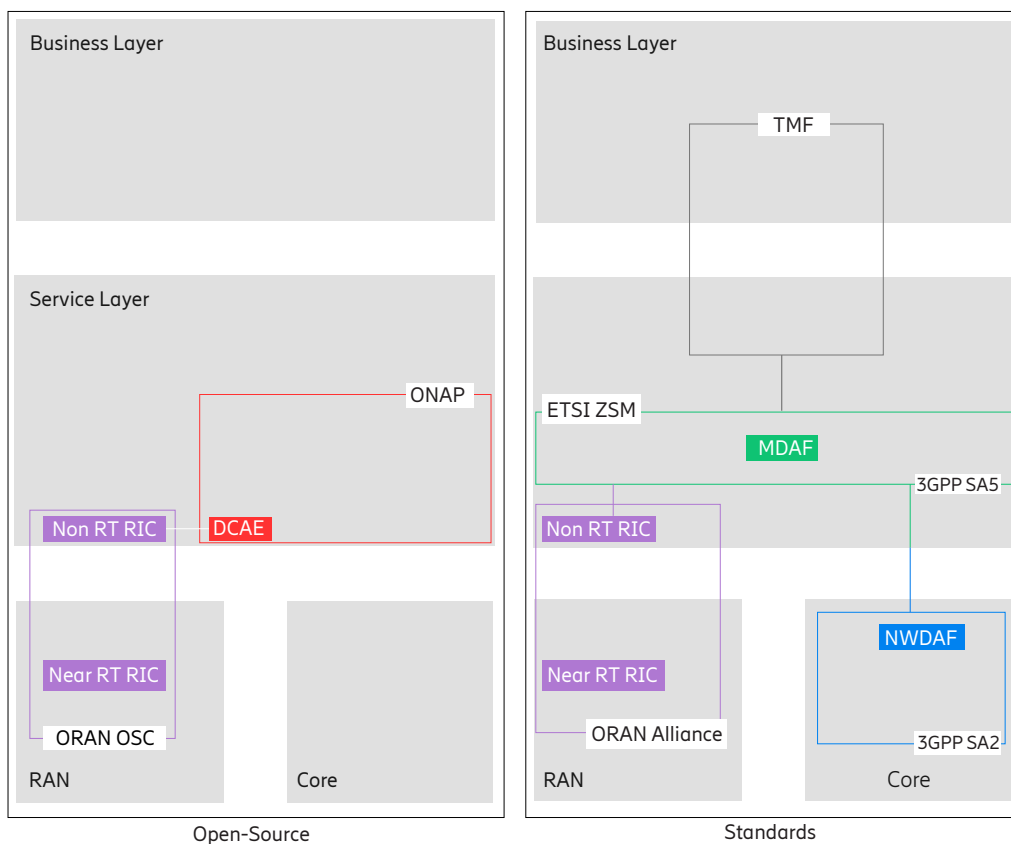


Figure 4. Wanted focus of each organization’s work within AI/ML mapped to the domains

3GPP SA5, ORAN and ONAP form an ecosystem for complementary alignment.

## Focus on a multi-cloud strategy for AI/ML

Public cloud providers and companies addressing AI/ML have developed a strong heritage in generic AI capabilities. The portfolios of all cloud providers today include machine learning services, engines, and frameworks for the conversational, vision, language, and knowledge areas. It also includes life cycle management of AI from training, deployment to monitoring & frameworks.

The challenge comes from the lack of standardization of cloud service providers and the risk of a lock-in; hence, the recommendation is to have a multi-cloud strategy with the help of trusted partners.

## Explainability and trustworthiness

Explainability and trustworthiness are key in AI systems in order to establish trust with the consumers of a system.

Trustworthy AI can be categorized in multiple dimensions, such as maintaining transparency on how an AI system uses AI, clarifying the consideration of different types of biases and ethics in model training, satisfying legal aspects, maintaining security and privacy, clarifying the data inputs and quality of data, and finally explainability of the AI methods used and the decisions made by them.

The need for explainability and trustworthiness of AI-based systems is well expressed by regional regulatory bodies [4] [5] and a global perspective of importance of explainable AI including the potential uses of explainable AI applications has been outlined [here](#). However, there should be requirements or associated studies in 3GPP on explainability and trustworthiness in AI-based systems.

## Data sharing in an open data ecosystem

An open and trusted data ecosystem is key for data sharing to accommodate the complexity of data exchange between CSPs and vendors. Already today, data exchange is common for network optimization and root cause analysis purpose. With AI/ML, its becoming a key resource for data-driven developments creating AI/ML software. Some important considerations in a data ecosystem are trustworthiness, data privacy, secure access and secure storage.

There is also the notion of public and non-public data. Public data is being made available from a product or service supplied by a vendor to individuals or entities for the purpose of product operations and/or service delivery. Non-public data, on the other hand, is data containing sensitive information relating to IPR or strategic business importance and is used by the vendor for innovation, product and/or service development, verification and deployment.

A data ecosystem also needs to support preprocessing close to the data points to avoid

unnecessary transfers and network load. Massive data volumes can be a burden if not addressed in the architecture. Different data points need to be included in the ecosystem to allow for network-wide data access, such as RAN and core, to create better AI/ML software and to address the specifics from different network domains.

3GPP has defined an architecture for network analytics. Recognizing these efforts, we at Ericsson have been working on an architecture for an open data sharing environment where data can be exchanged between CSPs and vendors to fasten AI/ML development and innovation. The image below shows the Ericsson Data Collection (EDC) architecture. [3]

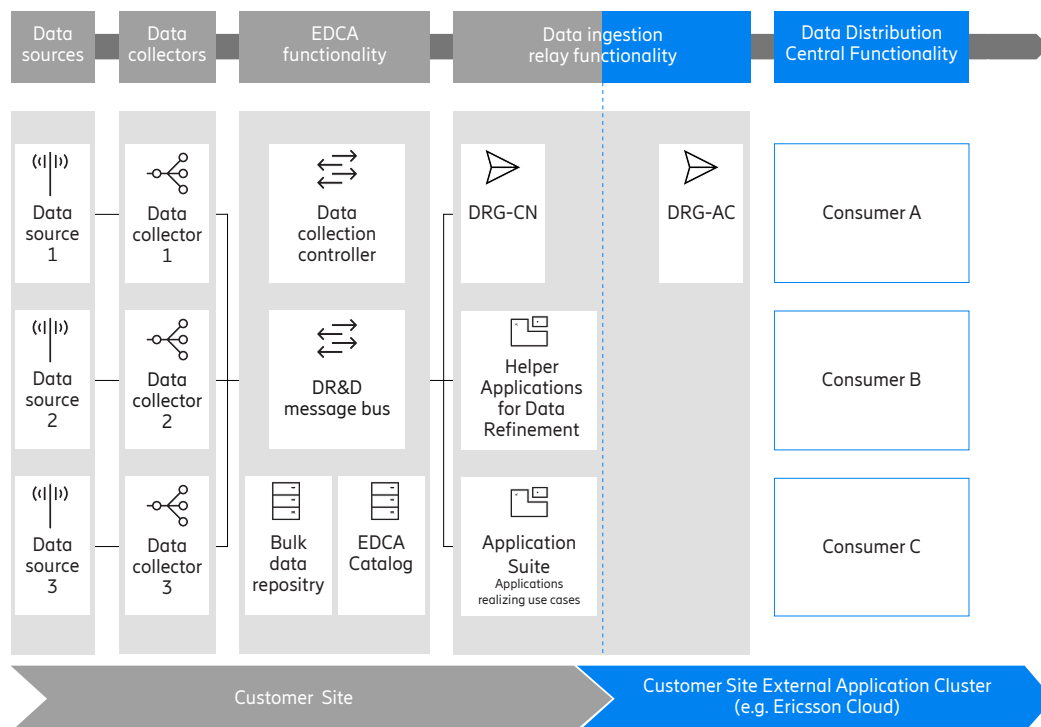


Figure 5. Ericsson Data Collection Architecture

- DR&D                      Data Routing and Distribution
- DRG-CN                 Data Relay Gateway - Customer Network
- DRG-AC                 Data Relay Gateway - Application Cluster

# Conclusion

Telecommunications service providers have an urgent need to reduce operational costs while supporting the rapid introduction of new services and products and identifying and leveraging monetization opportunities. AI/ML has emerged as a powerful technology that can support these needs.

While the journey of the application of AI/ML technologies in telecommunications networks has already begun, it has involved disparate and isolated approaches and has been applied within the current industry definition only as an afterthought. The step towards mass adoption and industrialization is yet to come and can be accelerated with the right level of industry alignment, supporting a multivendor ecosystem while still encouraging innovation enabled by the adoption of rapidly evolving technologies.

The industry has recognized that in order to transition to an industrialization phase and enable mass adoption of AI/ML, industry alignment is required. This results in all the major industry bodies trying to work out how they can leverage the technologies and claim their stake in the AI/ML landscape, leading to multiple and somewhat diverging directions being taken. To accelerate the coming industrialization phase and mass adoption, the industry must choose which guidance to follow.

AI/ML introduces new considerations to LCM and does so at a time when the industry is moving towards evolution of its software LCM with continuous deployment and integration. There are different approaches that can be taken, and taking an approach that maximizes end-to-end accountability is essential to accelerating adoption.

AI/ML should be adopted at all levels of a network architecture.

While enabling movement towards an aligned platform approach, service providers can benefit from a business-driven and use case-driven approach to the deployment of AI, covering required data, required insights and required actions.

It is recommended that telecommunications service providers and vendors do the following:

- pursue a use case-driven approach to prioritize network introduction
- enable the leveraging of rapidly evolving technologies from the IT and cloud industries to accelerate adoption by avoiding standardizing all the technology aspects that are rapidly evolving, such as model descriptions and data preparation (which are best covered as de facto technologies)
- focus the 3GPP-SA2 data collection on data collection using the Service Based Interfaces (SBI) events

- focus on aligning management data collection through 3GPP SA5 as well as the ORAN and ONAP ecosystem
- align 3GPP SA2, SA5, ORAN and ONAP perspectives of the functional architecture for AI/ML functions and LCM
- focus network analytics specifications in 3GPP and ensure alignment with ONAP and ORAN while enabling the optimization for different domains as described by ETSI ZSM

# Glossary

- AI Artificial Intelligence (including sub areas such as machine learning, deep learning, reinforced learning, deep reinforcement learning and machine reasoning)
- CSP Communication Service Provider
- LCM Lifecycle Management
- RAN Radio Access Network
- PC Packet Core
- 5G 5th Generation (of mobile networks)
- 3GPP 3rd Generation Partnership Project - Standardization body for cellular networks including 5G
- TMF Telecom Management Forum
- ETSI European Telecommunication Standards Institute
- ITU International Telecommunications Union
- QoE Quality of Experience
- Paging Process of locating a mobile device for communication
- SLA Service Level Agreement
- EDCA Ericsson Data Collection Architecture

# References

1. 5G programmable networks  
<https://www.ericsson.com/en/core-network/5g-core/forms/5g-core-programmability-underestimated-opportunity>
2. Adopting AI in Organizations  
<https://www.ericsson.com/en/reports-and-papers/industrylab/reports/adopting-ai-in-organizations>
3. Ericsson Technology Review for Data ingestion architecture for telecom applications.  
<https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/data-ingestion-architecture-for-telecom>
4. European Commission whitepaper – On Artificial Intelligence – A European approach to excellence and trust  
[https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificialintelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificialintelligence-feb2020_en.pdf)
5. European Commission Assessment list for Trustworthy Artificial Intelligence  
<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

# Suggested reading

- Adopting AI in organizations - Ericsson. 2021. Adopting AI in organizations - Ericsson. Available at: <https://www.ericsson.com/en/reports-and-papers/industrylab/reports/adopting-ai-in-organizations>
- Data-driven network architecture: An introduction - Ericsson. 2021. Data-driven network architecture: An introduction - Ericsson. Available at: <https://www.ericsson.com/en/blog/2020/10/data-driven-network-architecture#:~:text=The%20fundamental%20components%20of%20a,%2C%20and%20AI%20FML%20environments.&text=This%20solution%20can%20be%20used,be%20any%20network%20analytics%20function>
- Technology Review Magazine 2021 - AI special edition - Ericsson. 2021. Technology Review Magazine 2021 - AI special edition - Ericsson. Available at: <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/ai-special-edition>



# Authors



**Jitendra Manocha** is a Senior Portfolio Manager responsible for AI/ML portfolio strategy in Business Area Digital Services. In his 19 years of experience, he has held various leading positions in product management, R&D, and services. In recent years he has worked with AI/ML strategy, 5G analytics, 5G network exposure, and Edge computing. He holds an M.Sc. from KTH Royal Institute of Technology in Stockholm, Sweden.



**Stephen Terrill** is a Senior Expert in automation and management, with more than 20 years of experience working with telecommunications architecture, implementation, and industry engagement. His work has included both architecture definition and posts within standardization organizations such as ETSI, the 3GPP, ITU-T (ITU Telecommunication Standardization Sector) and IETF (Internet Engineering Task Force). In recent years, his work has focused on the automation and evolution of operations support systems, and he has been engaged in open source on ONAP's Technical Steering Committee and as ONAP architecture chair. Stephen holds an M.Sc., a B.E. (Hons.), and a B.Sc. from the University of Melbourne, Australia



**Ulf Mattsson** is a Technical Area Leader within AI/ML in Business Area Digital Services. He has more than 20 years of experience working with telecommunications, covering four generations of mobile systems. His work has included development, architecture definition, and standardization for networks and mobile phones. In recent years, his work has focused on architecture for AI/ML. Ulf holds an M.Sc. from Chalmers University of Technology, Gothenburg.



**Zlatko Filipovic** is a Senior Strategic Product Manager responsible for Business Unit Networks, AI/ML portfolio strategy. In 20+ years of professional career in the ICT industry, he has worked with strategy management, product & solution management, relationship sales, and cutting edge research & development.



**Erik Westerberg** is Chief Network Architect and Senior Expert at Business Unit Networks. Erik has experience from 25 years of development and standardization of four generations of mobile systems. He holds 50+ patents and is presently working with architecture, automation, and AI/ML in 5G systems and beyond.



**Dirk Kopplin** is a Senior Specialist for Core Networks. With more than 20 years of experience in telecommunication, he has a great deal of knowledge in mobile systems. During the last years, his focus was on network automation and AI/ML driving product development and standardization. Dirk holds a B.Sc. (Hons) from the University of Applied Sciences of Berlin, Germany.

#### **Acknowledgement**

Ali Sharaf, Christer Carlsson, Mika Mantynen, Dinand Roland