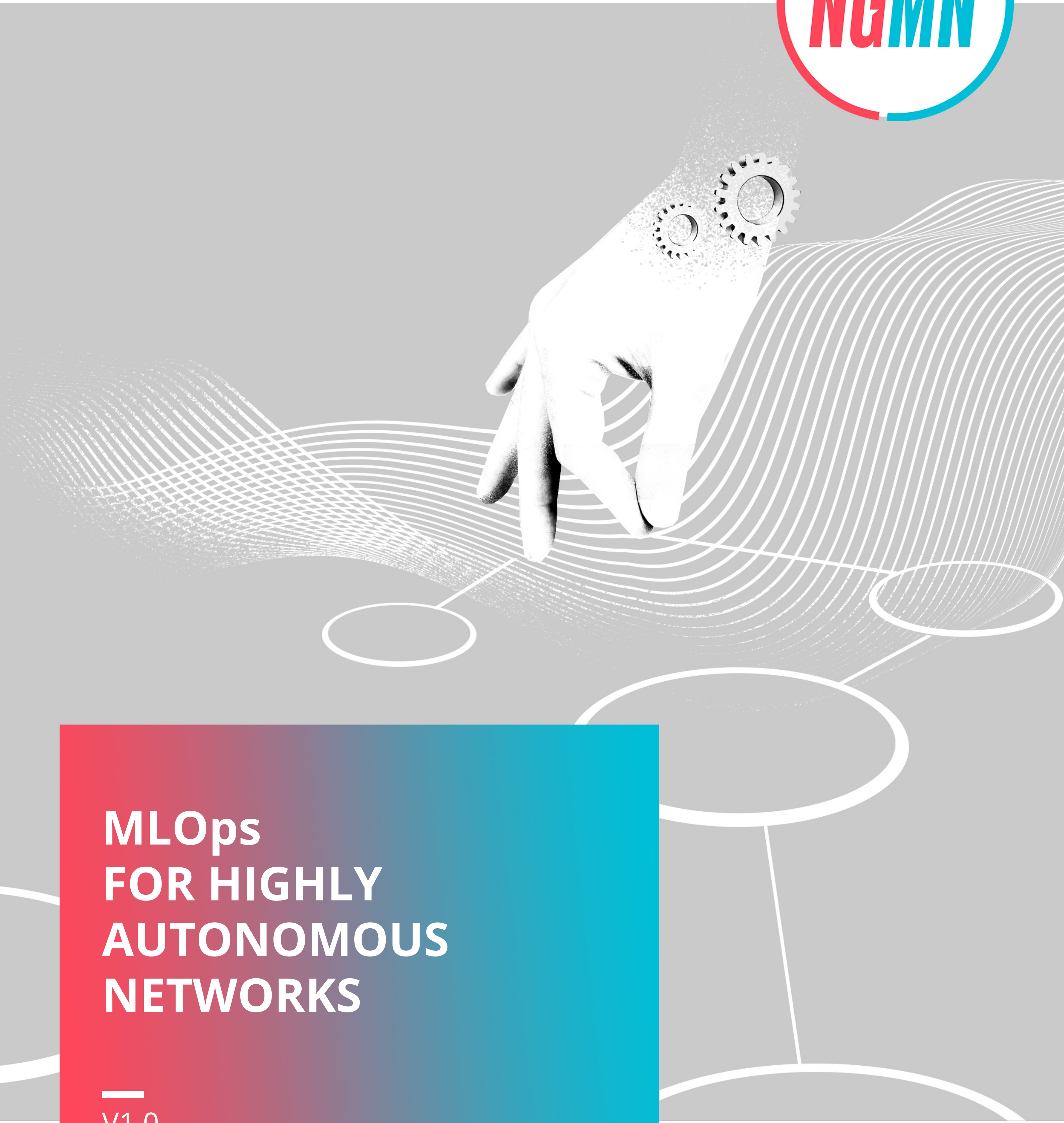

MLOps FOR HIGHLY AUTONOMOUS NETWORKS

V1.0

ngmn.org

WE MAKE BETTER CONNECTIONS

MLOps FOR HIGHLY AUTONOMOUS NETWORKS

by NGMN Alliance

Version:	1.0
Date:	18 February 2025
Document Type:	Final Deliverable
Confidentiality Class:	Public
Project:	Network Automation and Autonomy based on AI
Project Lead and Editor:	Lingli Deng/China Mobile, Sebastian Zechlin/Deutsche Telekom
Project Co-Leads:	Sebastian Thalanany/UScellular
Active Participants:	Yuhan Zhang/CMCC, Sheng Gao/CMCC, Cheng Cheng Feng/CMCC, Jean Paul Pallois/Huawei
NGMN Programme Manager:	Sparsh Singhal
Approved by / Date:	NGMN Alliance Board, 13 February 2025

The information contained in this document represents the current view held by NGMN Alliance e.V. on the issues discussed as of the date of publication. This document is provided «as is» with no warranties whatsoever including any warranty of merchantability, non-infringement, or fitness for any particular purpose. All liability (including liability for infringement of any property rights) relating to the use of information in this document is disclaimed. No license, express or implied, to any intellectual property rights are granted herein. This document is distributed for informational purposes only and is subject to change without notice. Readers should not design products based on this document.

ABSTRACT: SHORT INTRODUCTION AND PURPOSE OF DOCUMENT

This document describes guidelines and requirements for enabling technologies, as an extension to the Autonomous system framework – Phase 2 . This document focuses on general MLOps, and will explores MLOps applicable to an autonomous network, with advanced levels of autonomy for automation, in terms of examining the MLOps process, requirements, architecture, and end-to-end deployment and standardization recommendations.

DOCUMENT HISTORY

Date	Version	Author	Changes
08/10/2023	V 0.0	Lingli Deng, CMCC, Yuhan Zhang, CMCC	Initial working draft
14/03/2024	V0.1	Lingli Deng, CMCC, Yuhan Zhang, CMCC	Chapter4,5
15/04/2024	V0.2	Lingli Deng, CMCC, Yuhan Zhang, CMCC	Chapter 6
15/04/2024	V0.3	Lingli Deng, CMCC, Yuhan Zhang, CMCC	Chapter7
5/08/2024	V0.4	Lingli Deng, CMCC, Yuhan Zhang, CMCC, Sheng Gao, CMCC	Chapter1,8
17/08/2024	V0.5	Lingli Deng, CMCC	Editing
19/08/2024	V0.6	Yuhan Zhang, CMCC, Sheng Gao, CMCC	
26/08/2024	V0.7	Lingli Deng, CMCC	Chapter 8
29/08/2024	V0.8	Yuhan Zhang, CMCC, Sheng Gao, CMCC，Cheng Cheng Feng, CMCC	Chapter 9
09/10/2024	V0.9	Sebastian Thalanany, UScellular	Review and editing - All chapters
11/10/2024	V1.0	Yuhan Zhang, CMCC, Jean Paul Pallois Huawei	Merge the revisions
12/12/2024	V1.1	Yuhan Zhang, CMCC	Merge the revisions from BT
07/01/2025	V1.2	Yuhan Zhang, CMCC	Merge the revisions from UScellular

CONTENTS

01 INTRODUCTION

This document, titled «MLOps for Highly Autonomous Network,» aims at providing comprehensive guidelines and requirements, specific to enabling technologies for highly autonomous networks. It focuses on a generalized ML model development and operation, elaborating the MLOps process, requirements, architecture, and end-to-end deployment for Level 4+ autonomous network AI applications. In this context it examines the existing and relevant management standards applicable to AI applications and provides standardization recommendations.

02 DEFINITIONS

DEVOPS

A set of processes, methods and systems established to facilitate joint development, technology operations, and quality assurance, across different teams for software products.

ML MODEL

Output of a Machine Learning (ML) algorithm trained with a training dataset that generates predictions using patterns in the input data.[1]

MLOPS

A coined term from Machine Learning (ML) and DevOps, which stands for a set of processes, methods and systems established to facilitate the requirements management, data engineering, model development, model delivery and model operations, across different teams for ML models.

NOTE:
MLOps integrates machine learning (ML), DevOps, and data engineering to bring ML systems into production, enabling the development of machine learning products. It is based on principles of continuous integration and delivery (CI/CD), collaboration, orchestration, reproducibility including data, model, and code versioning, and continuous monitoring.[2]

03 MOTIVATION

Machine learning (ML) technologies, an integral subfield of artificial intelligence (AI), empowers a machine or systems to automatically acquire knowledge and improve from human experiences. This aspect is significant for operators to address both existing and emerging challenges within communication networks. As the implementation of intelligent network applications expands on a large scale, the hurdles impeding progress towards advanced levels of autonomy are also emerging in terms of rising complexity. As autonomous networks deploy various ML models across multiple management layers or domains ranging from resource operation, service operation, to business operation the challenges associated with the large-scale deployment and maintenance of ML models are becoming increasingly apparent.

MLOps is a suite of management processes, aiming to link up the development, deployment, and operation of ML models, connecting algorithms, together with delivery and operational teams, so as to improve the efficiency of life cycle management of ML models, and to promote their large-scale application. It is a specialized version of DevOps to manage ML models as special type of software products.[2].

Applying MLOps would effectively solve the practical problems of adopting ML technologies in a scalable manner, for highly autonomous network, through a systematic and automated life cycle management of ML models.

04 PROCESS

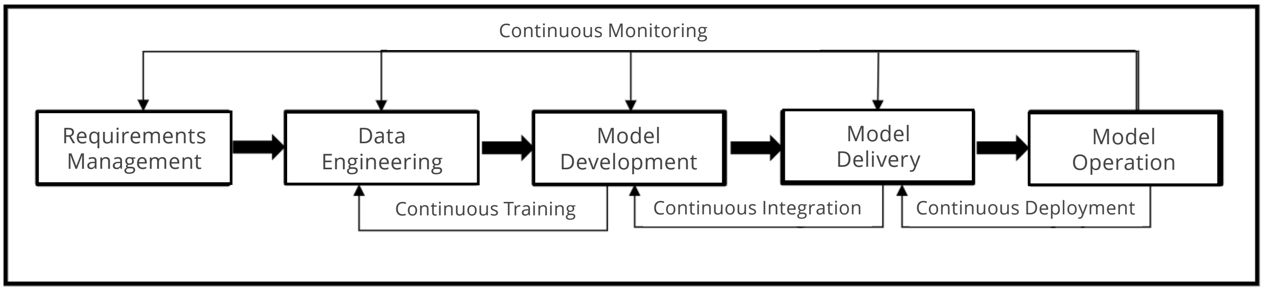

Figure 1 - ML model life cycle driven by MLOps workflow

As shown in Figure 1, MLOps workflow usually includes the following five phases:

- **Requirement Management:**

  This consists of feasibility analysis and prepares technical solutions based on business goals and requirements.

- **Data Engineering:**

  This transforms source data into a format optimized for ML model training.

- **Model Development:**

  This consists of ML model training (including ML model validation) and testing with selected base models, ensuring that the output models are optimized and ready for delivery.

- **Model Delivery:**

  This consists of a packaging of the developed ML model with configuration, code, and scripts, generates deliverables, and deploys them to the production environment.

- **Model Operation:**

  This consists of a monitoring (such as model inference performance) and operational maintenance (such as model activation/ deactivation) of deployed ML models in the production environment.

MLOps workflow automates ML models life cycle management through the following four closed-loop feedback pipelines:

» **Continuous Training (CT):**

This is triggered by a detection of preset conditions, for continuously training an ML model, ensuring the validation of the ML model, and the associated testing results, such that the performance requirements for accuracy and general requirements are met.

» **Continuous Integration (CI):**

This is triggered by updates from trained ML models and source codes, to continuously integrate ML models and source codes from various branches into a shared main branch, and uses automated testing, to ensure that the newly integrated ML models and source codes conform to the release criteria for subsequent delivery.

» **Continuous Deployment (CD):**

This is triggered by newly released ML models, to continuously deliver released ML models to an emulation environment to emulate the ML model operating conditions, which resemble real-world resources, and facilitates continuous deployment of the ML models to a relevant production environment automatically.

» **Continuous Monitoring (CM):**

This monitoring process is continuous, throughout the end-to-end life cycle of the ML model, where the monitoring process automatically identifies and monitors any risk and anomalous events, while sampling and collecting relevant data from events, managing and controlling based on a planned program or procedure, to ensure that the ML model is iteratively refined and prepared for a new life cycle.

05 REQUIREMENTS

The MLOps management process is primarily delivered through a workflow management function and coupled with a curated suite of 15 specialized functions. These functions are designed to support the full life cycle management of ML models, ensuring systematic oversight from development to deployment and operation phases. The functional requirements are delineated as follows:

- **Workflow Management Function:**
  This function spans t individual phases, prioritizing the oversight and governance of all functionalities. This function receives delivery or update notifications of ML models from model providers and complements other management functions to ensure whole life cycle management of ML models.

- **Data Processing Function:**
  This function transforms raw data into clean data usable by ML models, through a series of operations, and ultimately provides high-quality data for ML model development. These operations should include data cleaning, data conversion, data enhancement and other processing to reduce problems, such as data anomalies, data missing, and data duplication.

- **Model Training Function:**
  This function trains and validates the ML model based on data sets and machine learning algorithms based on specific business scenarios and operator management requirements, while adjusting and optimizing the ML model as needed. This is also triggered on demand (e.g., When an ML model performance degradation is detected) to retrain the ML model, ensuring that the ML model maintains performance and accuracy in a changing environment. It automatically executes training tasks triggered by preset conditions (e.g., the number of training data sets) and automatically updates the ML model online, based on target indicators (e.g., accuracy). Furthermore, the model training function validates ML models to evaluate their performance, including but not limited to margin of error validation, cross-validation, etc.

- **Model Testing Function:**
  This function tests the performance differences of the ML model on the test dataset, are identified through pre-set evaluation indicators, after the ML model is trained, to determine the generalization ability of the ML model, to ensure that the ML model meets the expected requirements.

- **Model Build and Integration Function:**
  This function builds and packages the code, ML models, dependencies, and other elements to produce deliverables, where the form of deliverables includes deployment packages, mirrors, etc., which can then be flexibly deployed to a production environment.

- **Model Emulation Function:**
  This function emulates the performance and effectiveness of ML models on independent emulation datasets, in order to discover defects of ML models, before deploying to a production environment.

- **Deployment Function:**
  This function deploys the ML model to a production environment and announces a new release.

- **Model Parsing Function:**
  This function extracts and analyses the ML model release profiles to understand what needs to be deployed, including environment configuration, interface configuration, model algorithm, etc., for configuring and launching ML model services.

- **Resource Checking Function:**
  This function checks whether the production environment meets the deployment requirements.

- **Resource Orchestration Function:**
  This function orchestrates resources (i.e., containers, pods, networks, etc.) based the on

environmental resource information to support the operating of ML models.

- **Model Orchestration Function:**
  This function executes ML model deployment, ML model configuration and network configuration. This function also has built in ML model registry and metadata store for tracking and logging the metadata of each ML model related workflow task, e.g., training date and time, duration, performance metrics, model lineage, etc.[3][4],

- **Model Onboarding Function:**
  This function initiates ML model inference service.

- **Monitoring Function:**
  This function continuously monitors the full life cycle of the ML model, including the input data monitoring, the ML model monitoring, and the business monitoring. The input data monitoring refers to a monitoring of the quality and distribution of inference input data. The ML model monitoring aspect refers to a monitoring of the performance of the ML model, during the model operation phase. The business monitoring aspect refers to a monitoring of the performance and effectiveness of ML model services in the business dimension, through pre-set business indicators.

- **Measurement Function:**
  This function analyses the ML model improvement direction, based on monitoring results and business requirements, to generate feedback reports on demand.

- **Data Collection Function:**
  This function collects the operational monitoring data, inference input data, and model training data of the ML model.

- **Model Inference Function:**
  This function executes inference analysis and provides the associated result feedback.

06 ARCHITECTURE

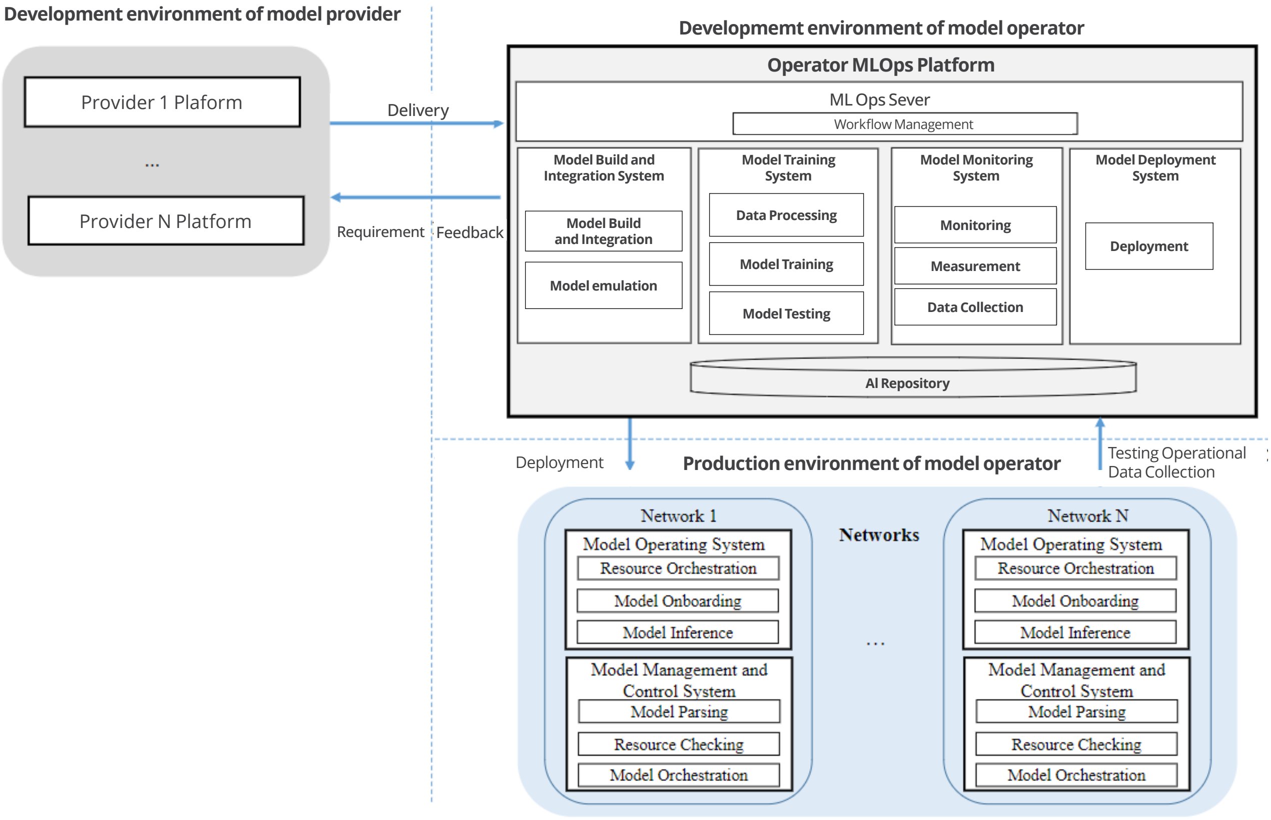

Figure 2 - Architecture of MLOps

The general architecture of MLOps is shown in Figure 2, which includes the development environment of a model provider, the development environment of a model operator, and the production environment of a model operator.

- The development environment of model providers refers to the development environment formed by the development workflow of different model providers. The ML model provider completes the entire process from requirements management, design, development, build, testing, and generates an ML model for delivery, based on its own development workflow. Afterwards, the model provider sends a release notification of the specific ML model to the operator subscribing to the model provider. After the ML model is deployed, when the model provider receives feedback from an operator on abnormal behaviours observed about the current ML model, the model provider analyses the issues, troubleshoots the ML model's faults, and updates the ML model as accordingly.

- The development environment of model operator, it refers to the development environment of the operator's MLOps workflow. After the operator MLOps platform receives the ML models delivered by the model provider, the operator MLOps platform implements the operator-side MLOps workflow and enables a testing of issues and provides abnormal behaviours to the model provider. The operator MLOps platform is mainly composed of the following type of server and system:

  » The MLOps server provides the workflow management function.

  » The model build and integration system provides the model build and integration function and the model emulation function.

- » The model training system provides the data processing function, the model training function, and the model testing function.

- » The model monitoring system provides the monitoring function, the measurement function, and the data collection function.

- » The model deployment system provides the deployment function.

- » The AI repository is used to store the ML models, and its related assets (including data, metadata, features, codes, parameters, images, etc.),records and tracks the life cycle status of the ML model.

- The production environment of model operator includes the emulation and production environments. The emulation environment is used for the emulation of an ML model before delivery. The production environment refers to the environment in which an ML model provides services to consumer. The production environment of the model operator should be credible and controllable. The production environment of model operator is mainly composed of the following systems:

  - » The model management and control system provide the model parsing function, the resource checking function, and the model orchestration function.

  - » The model operating system provides the resource orchestration function, the model onboarding function, and the model inference function.

07 DEPLOYMENT OPTIONS

For the network elements with AI/ML capabilities the operator MLOps platform should provide end-to-end model management services. The operator MLOps platform and the various systems in the production environment are be deployed as part of network elements, in cross-domain network management layers, single-domain network management layers, or within the network element management layers.

7.1 EMBEDDED NETWORK ELEMENT DEPLOYMENT

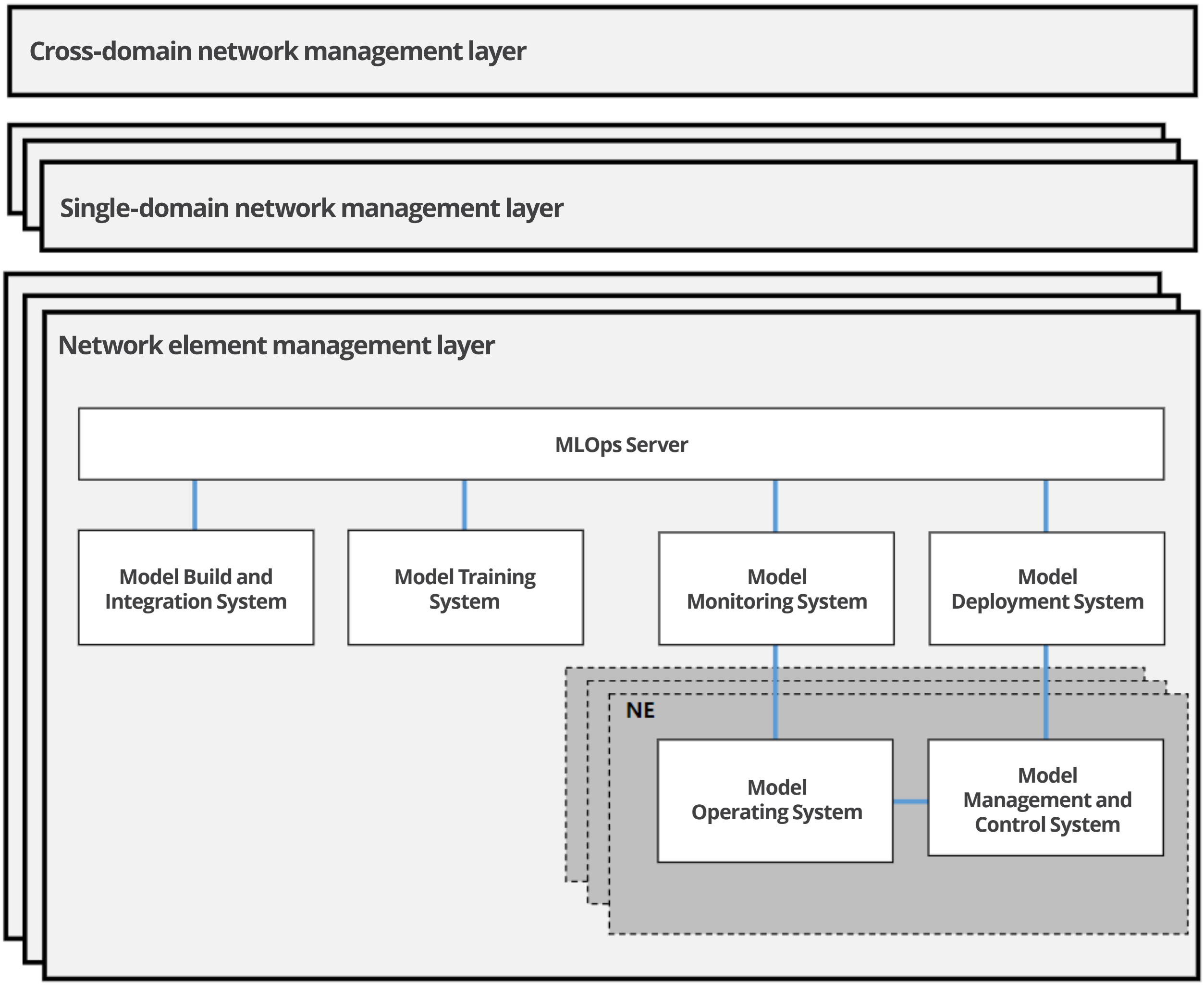
Cross-domain network management layer
Single-domain network management layer
Network element management layer
MLOps Server
Model Build and Integration System
Model Training System
Model Monitoring System
Model Deployment System
NE
Model Operating System
Model Management and Control System

Figure 3 - Embedded Network Element Deployment

In this deployment scenario, the MLOps server, the model build and integration system, the model training system, the model monitoring system, and the model deployment system are deployed in the network element management layer, to provide model training and other services for the network elements. The model management and control system and the model operating system are deployed in the network element of the network element management layer to provide model inference, and other services to the network element. This scenario is applicable to the MLOps workflow of a specific network element. For example, if the data collected by a network element being part of the training dataset is limited to the network, the model training cannot be performed in the network management layer.

7.2 INTEGRATED SINGLE-DOMAIN DEPLOYMENT

In this deployment scenario, the MLOps server, the model build and integration system, the model training system, the model monitoring system, and the model deployment system are deployed in a single-domain network management layer. The model management and control system and the model operating system are deployed in the single-domain network management layer to provide model inference and other services.

This scenario applies to single-domain MLOps workflow. For example, for the RAN or core network domain, the related functions may be integrated with existing management functions, such as MDAF [5].

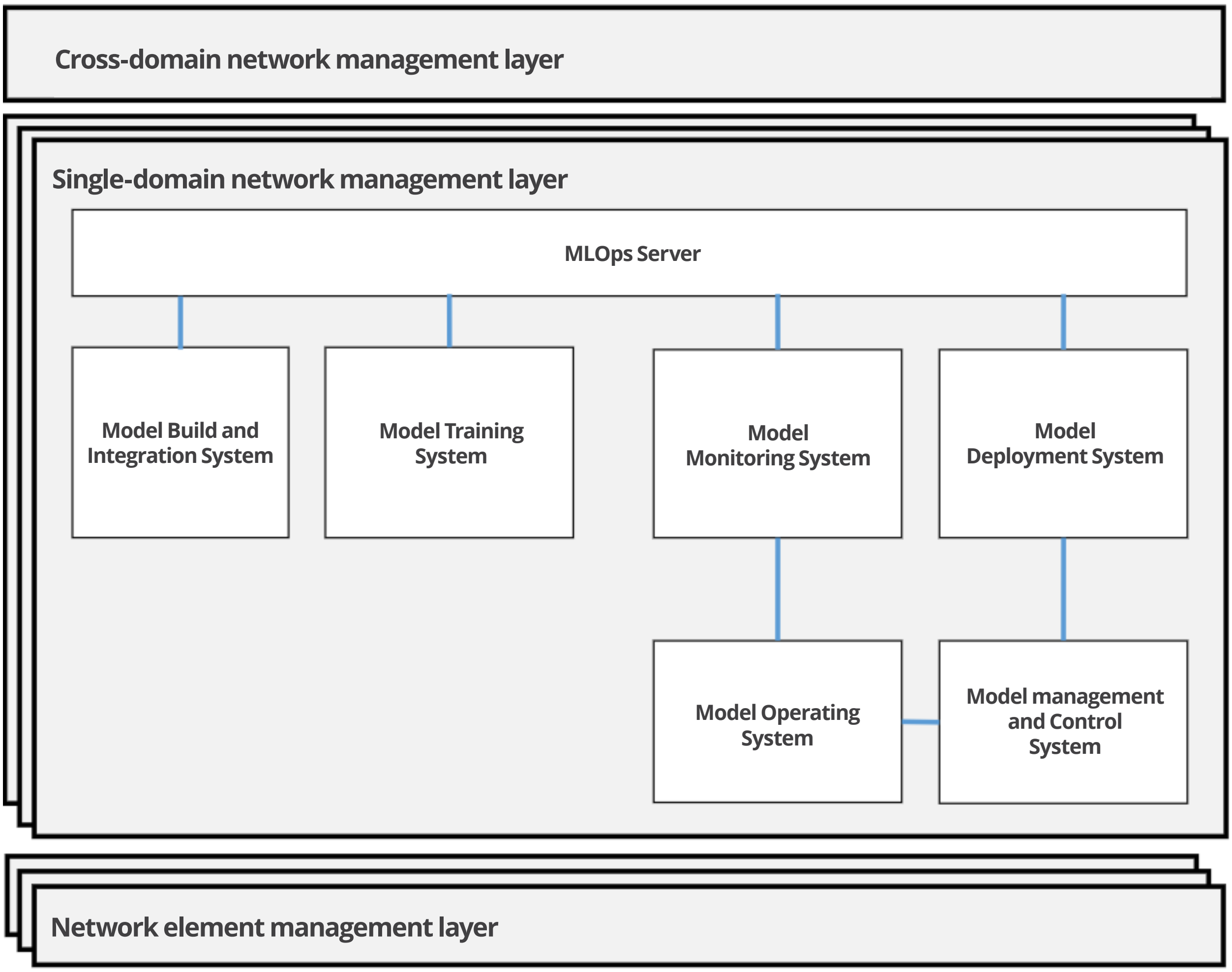
Cross-domain network management layer
Single-domain network management layer
MLOps Server
Model Build and Integration System
Model Training System
Model Monitoring System
Model Deployment System
Model Operating System
Model management and Control System
Network element management layer

Figure 4 - Integrated Single-Domain Deployment

7.3 REAL-TIME SINGLE-DOMAIN DEPLOYMENT

In this deployment scenario, the model training system is divided into two parts: the online and the offline model training systems. The online model training system is responsible for the real-time data collection and the online model training. In other words, the online model training system allows for the model to be continuously updated, to maintain its accuracy and effectiveness, as new data arrives continuously. The offline model training system is responsible for the initial training of the model, or periodic retraining.

In the real-time single-domain deployment, the online model training system, the model operating system, and the model management and control system are deployed in the network element, while the MLOps server, the model build and integration system, the offline model training system, the model monitoring system, and the model deployment system are deployed in the single-domain network management layer.

This scenario is applicable for single-domain MLOps workflows (e.g., MLOps workflows in the RAN domain network management system, or in the core network domain network management system), where real-time data can be collected within the network element, for example, in the gNB or in the NWDAF) for online model training, and where the ML model can be adjusted in time to adapt to a changing environment.

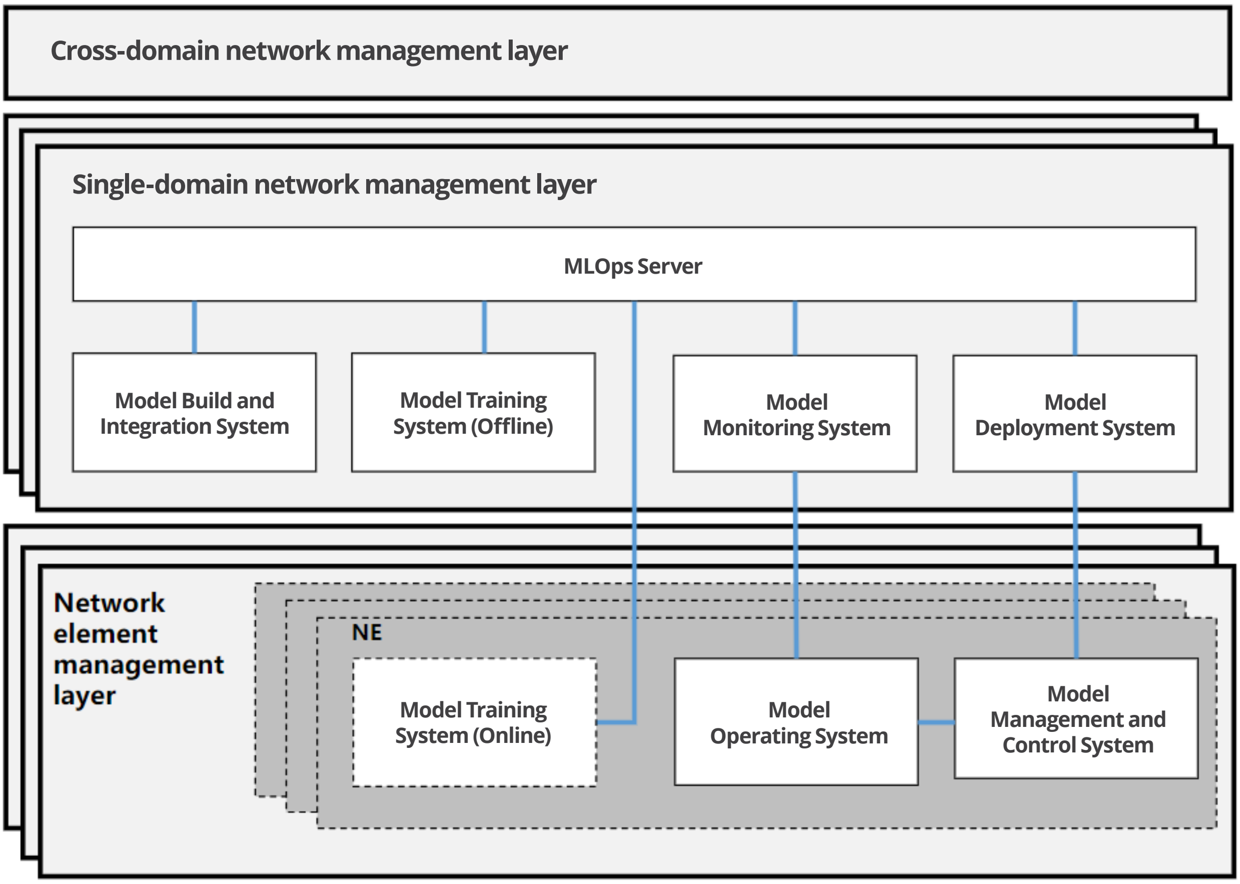
Cross-domain network management layer
Single-domain network management layer
MLOps Server
Model Build and Integration System
Model Training System (Offline)
Model Monitoring System
Model Deployment System
Network element management layer
NE
Model Training System (Online)
Model Operating System
Model Management and Control System

Figure 5 -Real-Time Single-Domain Deployment

7.4 CROSS-DOMAIN DEPLOYMENT

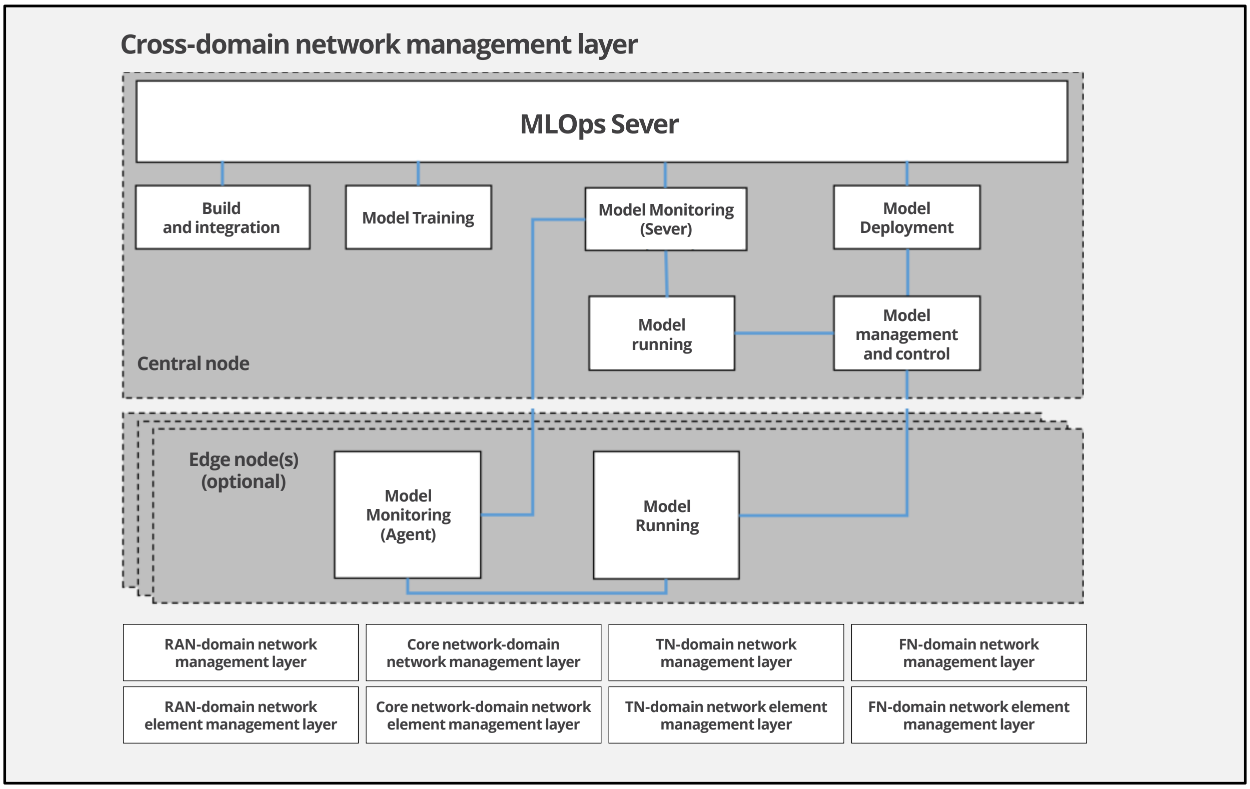
Cross-domain network management layer
MLOps Sever
Build and integration
Model Training
Model Monitoring (Sever)
Model Deployment
Model running
Model management and control
Central node
Edge node(s) (optional)
Model Monitoring (Agent)
Model Running
RAN-domain network management layer
Core network-domain network management layer
TN-domain network management layer
FN-domain network management layer
RAN-domain network element management layer
Core network-domain network element management layer
TN-domain network element management layer
FN-domain network element management layer

Figure 6 - Cross-Domain Deployment

In this deployment scenario, the central node is the core of the cross-domain network management layer, and it is responsible for centralized processing, coordination, and monitoring. The cross-domain network management layer is utilized for executing the overall MLOps workflow management and related functions. The edge nodes, which are located in various different geographical locations such as provincial nodes or nodes in different regions, perform model inference and monitoring, providing feedback to the central node.

The MLOps server, the model build and integration system, the model training system, the model deployment system and the model management and control system are deployed at the central node of the cross-domain network management layer.

The model monitoring system is divided into two parts, namely, a model monitoring system (Server), and a model monitoring system (Agent). The model monitoring system (Server) monitors ML models on central nodes, it collects and analyses data generated by the central node model operations and feedback data from the relevant agents and comprehensively evaluates the aggregated feedback data. The model monitoring system (Agent) monitors ML models deployed on the local edge node, performs monitoring and analysis of the collected data, and further provides preliminary measurement feedback. When the model monitoring system (Agent) deployed on the edge node detects a degradation in the model performance, it may report the degradation to the model monitoring system (Server) at the central node. The model monitoring system (Server) at the central node may report the incident to the MLOps Server to decide whether and how to deal with the situation (e.g., to retrain the model or select a new ML model and deploy it to the edge node). The model monitoring system (Agent) at different edge nodes track their own unique historical and inferential data, which may be aggregated by the model monitoring system (Server) and used by the model training system for (online or offline) retraining, to enhance the performance of the models.

The model operating systems may be deployed at both the central node and at the edge nodes. The model operating system at the central node is capable of handling complex data analysis and resource-demanding inference tasks. The model operating systems at the edge nodes are responsible for preliminary data analysis, providing decision-making for rapid local response for simple issues. This scenario is applicable for cross-domain MLOps workflows.

08 STANDARDIZATION RECOMMENDATIONS

8.1 CURRENT STATUS

In summary, the following observations are concluded regarding MLOps for autonomous networks standardization:

To start with, TM Forum highlights the significance of MLOps in advancing the capabilities of autonomous networks by providing a structured and efficient framework for managing ML models, necessitating industry-wide efforts towards the standardization of MLOps practices.

TMF:

The document on «Autonomous Networks: Empowering digital transformation - from strategy to implementation (IG1305)» [6], released by the TMF in 2023 identifies MLOps as a critical technology for the future development of autonomous networks.

Furthermore, despite the fact that multiple SDOs, such as those listed below, are currently engaged in the development of MLOps-related standards.

» From a generalized framework for end-to-end, and cross domain management perspective, the following considerations are relevant:

ISO/IEC:

The ISO/IEC JTC1/SC42 Artificial Intelligence Subcommittee has initiated the project: «Information Technology - Artificial Intelligence - AI System Lifecycle Processes.» [7]. By establishing robust life cycle standards, this initiative aims to provide guidance for application development, thereby supporting the standardization of applications. However, it does not address automatic closed loops from an MLOps perspective.

ITU-T:

The ITU-T has already established requirements related to autonomous networks: (ITU-T M.3016 [8]), AI-enhanced telecommunications operation and management framework (ITU-T M.3080 [9]), and requirement of joint development and operation for IMT-2020 and beyond (ITU-T Y.3164 [10]), but there is still a lack of management standards for AI applications, within these autonomous networks and with respect to the life-cycle management of ML models.

» From the domain-specific management perspective, 3GPP and O-RAN have completed preliminary research on AI/ML management for mobile communication networks, with varying levels of maturity and consistency.

3GPP:

in Rel-18 and Rel-19, a majority of working groups within TSG SA and TSG RAN have undertaken Study Items (SIs) and/or Work Items (WIs) focused on an integration of AI/ML technologies. These initiatives encompass a diverse array of usage scenarios and specific use cases, where AI/ML is applied to optimize various facets of the 3GPP System. For instance, 3GPP RAN1 is establishing a general framework for enhancing the air interface, leveraging the AIML life cycle management framework in TR 38.843 [11].

3GPP RAN3 is utilizing AIML technology to enhance the network performance and user experience for energy saving, load balance and mobility optimization in TR 37.817 [12]. 3GPP SA5 has addressed many aspects of an AI/ML mediated management for Network OAM aspects related to 5GS (5G System), including the management and orchestration subsystem that spans the5G Core (5GC) and Next Generation -RAN (NG-RAN) domains, in TR 28.908 [13] and TS 28.105 [14].

The current phase of a study spearheaded by 3GPP TSG SA is dedicated to examining the ongoing AI/ML initiatives within the TSG RAN, and TSG SA working groups. It seeks to pinpoint any potential misalignments or inconsistencies. However, the study and the standardization purview of the existing 3GPP framework is yet to encompass the autonomous principles of DevOps. The realization of a consistent and consensus-oriented approach for the establishment of an end-to-end AI/ML framework and lifecycle management is a work in progress. Such an approach is crucial for the development of AI-native networks, which are poised to play a pivotal role in the forthcoming 6G architecture.

In Release 19, the 3GPP SA5 has initiated a study on Generative AI, encompassing the exploration of novel AI/ML technologies and concepts associated with Language Models, which delves into pre-training, fine-tuning, distributed learning, reinforcement learning, and other pertinent learning methodologies.

O-RAN:

This initiative conducts research on «AI/ML workflow description» [15] and requirements, within the O-RAN architecture, addressing AI/ML workflows, typical use cases, and functional requirements. O-RAN has already commenced informative work on MLOps. However, the architecture and interfaces of MLOps within the O-RAN system require formal standardization to ensure seamless integration and operation.

» From a network agnostic management perspective, TMF has initiated research on AI/ML management for network management layers in several aspects, with varying levels of maturity and consistency.

TMF:

The ODA (Open Digital Architecture) project and AI Operations project are continuously exploring how AI can enable digital transformation. The ODA project released GB1022 [16] In 2021, which outlines the functional framework for ODA. By breaking down the key capabilities of a digital business or enterprise, this functional architecture fosters business and operational agility. The framework is divided into five functional blocks, namely, engagement management, party management, core commerce management, production, and intelligence management. Intelligence management is defined as being responsible for knowledge-defined automation. It leverages big and fast data to enable cross-functional intelligence functions and cognitive workflows. While this guidebook addresses the functional requirements of intelligence management, further elaboration of the architecture or functionality of the intelligent platform is required. The integration of AI/ML capabilities with ODA is still under active discussion. Meanwhile, the AIOps project has released a series of guidebooks focused on applying AIOps to network management processes. The IG1190 [17] series explores how to redesign operations and service management processes to support and manage the large-scale deployment of AI. GB1065 [18] provides guidance for an end-to-end, tool-agnostic, and use-case-agnostic software lifecycle to manage AI components from their initial conception, through design, development, deployment, production (including runtime and daily operations), maintenance (including the ML retraining process), until their eventual decommissioning. However, it does not detail how to build a system that effectively manages and ultimately meets specific requirements.

Overall, there is currently a lack of comprehensive and universally applicable standards for MLOps, which is urgently needed to establish common MLOps specifications, across network domains and organizations.

8.2 RECOMMENDATIONS

Overall, given the lack of generalized standards directions across domains, it is recommended that ITU-T initiate overall enabling considerations for an end-to-end and cross domain generalized MLOps framework by extending existing DevOps framework [10]. This would serve as guidance for related domain-specific standardization, enabling harmonized MLOps standards, while also promoting a collaborative development in the industry.

In the domain of mobile communication networks, the recommendations are two-fold:

» It is recommended that 3GPP considers a convergence of the life cycle management of the MLOps and DevOps processes into the end-to-end ML model life cycle, which would remedy and expedite the automation aspects of AI/ML management, within the entire 3GPP management system.

- It is recommended that O-RAN further promotes the application of MLOps in O-RAN by following a generalized framework to be set by 3GPP for mobile communication networks and focus its own subsequent work on standardizing interactions among decoupled RIC controllers and X-applications to satisfy MLOps requirements within O-RAN.

Additionally, a framework of Large models (e.g., Large Language Models (LLMs)), small models (e.g., small Language Models (LSMs)) and their application in the context of Generative AI (Gen AI) represents a pivotal direction in the ongoing evolution, and a transformative progression towards the B5G and 6G networks. These considerations are significant for standardization endeavours, which benefits all relevant SDOs in terms of incorporating Large Language Models Operations (LLMOps), thereby facilitating the seamless integration of an inherently intelligent architecture, within the network's foundational framework.

09 LIST OF ABBREVIATIONS

5GC	5th Generation Core
5GS	5th Generation System
6G	6th Generation
AI	Artificial Intelligence
AN	Autonomous Network
B5G	Beyond 5G
CI	Continuous Integration
CD	Continuous Deployment
CT	Continuous Training
CM	Continuous Monitoring
IMT	International Mobile Telecommunications
LLMOps	Large Language Models for Operations
MDAF	Model Data Analytics Function
ML	Machine Learning
NWDAF	Network Data Analytics Function
OAM	Operation Administration and Maintenance
RAN	Radio Access Network
SA	Service and System Aspects
SDO	Standards Development Organization
SI	Study Item
TN	Transport Network
TSG	Technical Specification Group
WI	Work Item

10 REFERENCES

[1] ISO/IEC 22989:2022(en) Information technology — Artificial intelligence — Artificial intelligence concepts and terminology, 2022

[2] GTI, «GTI AUTONOMOUS NETWORK V3.0», 2022

[3] D. Kreuzberger, N. Kühl and S. Hirschl, «Machine Learning Operations (MLOps): Overview, Definition, and Architecture,» in IEEE Access, vol. 11, pp. 31866-31879, 2023, doi: 10.1109/ACCESS.2023.3262138.

[4] Diaz-De-Arcaya, J., Torre-Bastida, A. I., Zárate, G., Miñón, R., & Almeida, A. (2023). A joint study of the challenges, opportunities, and roadmap of mlops and aiops: A systematic survey. ACM Computing Surveys, 56(4), 1-30 doi: 10.1145/3625289

[5] 3GPP, « Management Data Analytics (MDA)», TS 28.104, 2024

[6] TM Forum, «IG1305 Autonomous Networks Empowering digital transformation», 2023

[7] ISO/IEC 5338:2023, «Information technology — Artificial intelligence — AI system life cycle processes», 2023

[8] ITU-T, «Autonomous networks - Architecture framework», Y.3016, 2023

[9] ITU-T, «Framework of artificial intelligence enhanced telecom operation and management (AITOM)»,M.3080, 2021

[10] ITU-T, «Requirement of joint development and operation for IMT-2020 and beyond» Y.3164, 2024

[11] 3GPP,« Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface», TR 38.843, 2023

[12] 3GPP, «Study on enhancement for Data Collection for NR and EN-DC», TR 37.817, 2022

[13] 3GPP, «Study on Artificial Intelligence/Machine Learning (AI/ ML) management», TR 28.908, 2023

[14] 3GPP, «Management and orchestration; Artificial Intelligence/ Machine Learning (AI/ML) management», TS 28.105, 2024

[15] O-RAN, «AI/ML workflow description», 2022

[16] TM Forum, «GB1022ODA Functional Architecture Guidebook v1.1.0», 2021

[17] TM Forum, «IG1190 AIOps Service Management v4.0», 2021

[18] TM Forum, «GB1065 E2E AIOps Lifecycle», 2024

11 FIGURES

12 ACKNOWLEDGEMENTS

China Mobile, Lingli Deng

China Mobile, Sheng Gao

China Mobile, Yuhan Zhang

Deutsche Telekom, Sebastian Zechlin

Huawei, Jean Paul Pallois

Huawei, Luigi Licciardi

UScellular, Sebastian Thalanany

ZTE, Manchang Ju

NEXT GENERATION MOBILE NETWORKS ALLIANCE

NGMN is a global, operator-driven leadership network founded in 2006 by leading international mobile network operators (MNOs). As a global alliance of nearly 70 companies and organisations - including operators, vendors, and academia - NGMN provides industry guidance to enable innovative, sustainable and affordable next-generation mobile network infrastructure.

NGMN drives global alignment of technology standards, fosters collaboration with industry organisations and ensures efficient, project-driven processes to address the evolving demands of the telecommunications ecosystem.

VISION

The vision of NGMN is to provide impactful industry guidance to achieve innovative, sustainable and affordable mobile telecommunication services to meet the requirements of operators and address the demands and expectations of end users. Key focus areas include Mastering the Route to Disaggregation, Green Future Networks and 6G, while supporting the full implementation of 5G.

MISSION

The mission of NGMN is:

- To evaluate and drive technology evolution towards the three **Strategic Focus Topics:**
  - **Mastering to the Route to Disaggregation:**

    Leading in the development of open, disaggregated, virtualised and cloud native solutions with a focus on the E2E Operating Model
  - **Green Future Networks:**

    Developing sustainable and environmentally conscious solutions
  - **6G:**

    Anticipating the emergence of 6G by highlighting key technological trends and societal requirements, as well as outlining use cases, requirements, and design considerations to address them.
- To define precise functional and non-functional requirements for the next generation of mobile networks
- To provide guidance to equipment developers, standardisation bodies, and collaborative partners, leading to the implementation of a cost-effective network evolution
- To serve as a platform for information exchange within the industry, addressing urgent concerns, sharing experiences, and learning from technological challenges
- To identify and eliminate obstacles hindering the successful implementation of appealing mobile services.