Journal of Electrical Engineering and Information Technologies, Vol. **9**, No. 2 pp. 108-115 (2024)
Article 225
Received: November 19, 2024
Accepted: November 25, 2024
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 004.8:621.39]:339.13.025
doi.org/10.51466/JEEIT2492225108t

Original Scientific Pape

SURVEY ON TECHNICAL, BUSINESS AND REGULATORY ASPECTS OF ARTIFICIAL INTELLIGENCE IN TELECOM NETWORKS AND SERVICES

Dimitar Tanevski [1)] Toni Janevski Venceslav Kafedziski

Ss. Cyril and Methodius University in Skopje, Faculty of Electrical Engineering and Information Technologies
1000 Skopje, Macedonia

[1)] dimitartanevski@yahoo.com

Abstract: The advancement of AI/ML models, along with the complexity of new-generation networks, including IMT-2020, has made AI functionality indispensable in telecom architecture and business strategies. Standardization bodies such as ETSI and ITU-T have established the foundation for architectural frameworks, with ITU-T Y.3172 and ETSI's ENI models, leading to various recommendations and supplementary documents covering internal communication, quality of service, fault and recovery, and security frameworks and procedures. Globally, regulatory efforts surrounding AI in telecom, such as the EU's Artificial Intelligence (AI) Act and initiatives from the U.S. and China, emphasize risk categorization, fines for non-compliance, and ethical principles to ensure responsible deployment. Industry collaborations, exemplified by Nokia and NVIDIA's partnership for tech efficiency, and the deployment of AI-driven chatbots by telecom giants like AT&T, Deutsche Telekom, and Vodafone, highlight tangible benefits such as reduced costs, improved performance, and enhanced customer experience. These advancements not only drive down CAPEX and OPEX but also fuel revenue growth and bolster customer satisfaction, underscoring the transformative impact of AI in telecommunications

Keywords: Artificial Intelligence (AI), AI governance, AI regulation, AI standardization, business aspects of AI

ТЕХНИЧКИ, БИЗНИС И РЕГУЛАТОРНИ АСПЕКТИ НА ВЕШТАЧКАТА ИНТЕЛИГЕНЦИЈА ВО ТЕЛЕКОМУНИКАЦИСКИТЕ МРЕЖИ И СЕРВИСИ

Апстракт: Напредокот на AI/ML моделите, паралелно со зголемената комплексност во мрежите од новата генерација, вклучително IMT-2020, ја направија AI функционалноста неразделив дел од телекомуникациската архитектура и бизнис стратегијата. Стандардизационите тела како ETSI и ITU-T ја воспоставија основата за архитектурните рамки, со ITU-T Y.3172 и ETSI ENI моделите, презентирајќи различни препораки и пропратни документи кои покриваат внатрешна комуникација, квалитет на услугата, управување со грешки и безбедносни рамки и процедури. Глобално, регулаторните напори околу AI во телекомуникациите, како што се AI актот на ЕУ и легислативните иницијативи на САД и Кина, нагласуваат категоризација на ризиците, казни за непочитување и етички принципи, обезбедувајќи одговорна имплементација. Индустриските соработки, како што е партнерството меѓу Nokia и NVIDIA за технолошка ефикасност, и имплементацијата на AI-управувани чатботови од телекомуникациски гиганти како AT&T, Deutsche Telekom и Vodafone, истакнуваат конкретни придобивки како намалување на трошоци, подобрени перформанси и подобрено корисничко искуство. Овие напредоци не само што ги намалуваат CAPEX и OPEX, туку го поттикнуваат растот на приходите и го зголемуваат корисничкото задоволство, нагласувајќи го трансформативното влијание на AI во телекомуникациите.

Клучни зборови: ретроспективна дозиметрија, термолуминисценција, оптички стимулирана луминисценција вештачка интелигенција (ВИ), деловни аспекти на ВИ, управување со ВИ, ргулатива за ВИ, стандардизација на ВИ.

I. INTRODUCTION

FUTURE networks face evolving requirements, including increased bandwidth, QoS, personalized services, and heterogeneous access technologies, adding complexity to network management. Efficiently handling these complexities through manual intervention is becoming infeasible. In recent years, AI/ML models have become integral to technological frameworks, with the path to 6G and next-gen networks emphasizing AI alongside traditional network requirements [2]-[4]. The growing complexity of networks and business demands underscore the need for AI/ML to optimize CAPEX and OPEX in deployment, operation, and maintenance. Projections predict a CAGR of 41.4% from 2022 to 2031. A recent *NVIDIA* survey found 56% of telecom companies see AI as crucial to their success, with 90% actively implementing AI. Notably, 48% are piloting AI, while 41% have integrated it into operations. AI adoption aims to improve customer experience (48%) and reduce costs (35%), with most companies seeing revenue increases and lower operational expenses, yielding ROI within five years.

In 2023, generative AI emerged as a leading model, particularly in chatbots and digital assistants. However, AI implementation comes with challenges, such as shortage of professionals and interoperability issues across telecom network segments. To address these, ITU-T has issued several recommendations. ITU-T Y.3115 provides guidelines for integrating AI into existing networks, while Y.3170 specifies ML-based QoS assurance for IMT-2020 networks [6] [7]. Y.3172 outlines an ML framework for IMT-2020, detailing interfaces and enablers for efficient deployment [8]. Y.3183 focuses on QoE translation into network parameters with AI-driven policies [9], and Y.3177 suggests a framework for AI-based network automation [10].

Parallel efforts by ETSI's ENI ISG define Cognitive Network Management using AI, proposing an architectural framework for personalized services [11]. Recommendations like ENI ETSI 002 outline the requirements for implementing ENI, while ENI ETSI 003 discusses context-aware policies [12] [13]. The ENI group has conducted 22 Proof of Concepts (PoCs) to demonstrate the feasibility of this framework [17].

Implementation of AI/ML models in telecom networks requires extensive data, often involving customer personal data. This data must be handled carefully to avoid privacy breaches and comply with GDPR regulations. Additionally, AI models must be constructed to avoid discrimination or manipulation of consumer behavior. Various countries have introduced legislation to regulate AI use, with the EU having the strictest regulations, followed by China, while the U.S. has more lenient policies.

The structure of this paper is as follows: Section II explores ITU and ETSI architectural frameworks, their components, interfaces, and a Fault and Recovery Model for AI validation. Section III discusses global legislative measures relevant to AI in telecom. Section IV highlights practical AI implementations and business motivations. Section V concludes the paper.

II. AI/ML Architectures

Machine Learning (ML) will be essential tool in every segment of telecommunication networks, giving a way of generalized learning, prediction, optimization and preventive actions such as fault detection and recovery. However, implementation of ML algorithms is a novelty in operations and maintenance procedures for telco ISPs. This poses the question for cost-effective and seamless ways for incorporating ML in networks of the new generation. These algorithms and models are data driven, making it essential to detect the sources of information, connections and resources needed for efficient implementation.

A. ITU-T AI/ML Framework

The new generation networks are introducing a whole new level of heterogeneity, i.e. diversity in the RAN segment (different RAT, devices), linked to diversity in the core network (CN), and possibility for network slicing. In general, challenges brought about with introduction of ML in telco networks are grouped in 4 categories according to ITU-T Y.3172 [8]:

- Coping with the heterogeneous nature of the networks;
- Unified development and roadmap alignment between ML functionalities and networks;
- Finding efficient way for cost-effective and seamless integration in telco network architectures;
- Estimating and minimizing the impact on operations and maintenance procedures of communication networks.

To mitigate disputed points of ML implementation and integration, the ITU-T Y.3172 Recommendation proposes a generalized architectural framework.

The framework defines 5 types of high-level architectural requirements, on the basis of which multiple recommendations for lower-level architectures and functionalities are developed – enablers for data correlation sourced from heterogeneous technologies, enablers for deployment, requirements related to interfaces between different architectural components, requirements for declarative specification of ML applications, and requirements related to management.

The enablers for correlation define a requirement for ML architecture that is able to support correlation between data sets from different sources in the network. This is due to multi-level dependencies of a given parameter linked with different network functions (NFs) deployed in separate segments in the network. A good example of this is the QoS parameter in end-to-end user flows, which is interdependent with performance in the access network (AN) and the core network (CN). Also, ML architecture is required to provide support of different technologies at the same time. This can be illustrated with an example where slices designed for different technologies should coexist and be optimized for providing a required end-to-end QoS. Since multilevel and heterogeneous data is introduced in this section, it is obvious that postprocessing and analysis will be carried in distributed manner. That means that ML architecture should support distributed instantiation of machine learning functionalities and data transfer. Often, data analysis required for CN (core network) is dependent on information gathered in AN (access network), thus preprocessed AN data should be transferred and correlated at CN level.

The enablers for deployment specify the required flexible placement of ML functions, based on defined resource (e.g., computational load, availability of data, dedicated HW) and parameters constraints (e.g., latency, QoS). For instance, ML algorithms perform better in GPU environment than in CPU, so deployment should be flexible enough to meet these requirements regardless of data sources. In addition, a well-known practice for real-time latency applications is to place ML functionalities closer to the edge segment of the network (Edge Cloud Computing).

On top of the aforementioned, ML architecture should provide dynamic pluggable interfaces for new data sources, based on the requirements for detection. This is especially popular in anomaly detection for mIoT (massive IoT) cases.

Another essential feature is the distinction between model training and model usage. This is introduced through variable interfaces linked to ML functionalities. Certain segments of the network may have the data needed for model training, but they could lack HW resources for such task, and vice versa. The framework suggests flexible training and deployment of trained models, based on the needs and capabilities of the specific network segment. Moreover, this represents a dependency for data exchange between different network segments which, as mentioned above, are the main enabler for data correlation.

Given the plethora of services and vendors building and using the telecommunication network, the framework defines a need for declarative specification for representing ML applications, which could later be translated into a specific set of ML functionalities and parameters enabled by the underlay network architecture. The framework specifies requirements for declarative specification, making the underlay network architecture agnostic for ML applications linked to a specific network function. ML architecture should be able to provide dynamic specification of data sources, ML models, output targets, and constraints.

Finally, all of ML functionalities have to be managed in efficient and seamless manner. Different applications and optimization problem spaces demand different ML models with specific parameters and optimization techniques. Moreover, as previously explained, the framework offers dynamical addition of various data sources. These data sources, in general, are of heterogeneous nature, requiring that the ML model selection for a given application is selected dynamically and adaptively. Consequently, ML functionalities in the network should provide performance monitoring and suitable orchestration based on the output of the models. In addition to optimization and orchestration of VNFs and the underlay network, the output could be compared to given thresholds and reused for model corrections. All of the activities performed in ML functions need to be transparent for the network, meaning that training, updating and scaling of the model must not cause any impact or disturbance to the live network. These requirements are sublimed in the high-level architectural model recommended in ITU-T Y.3172 specification, illustrated in Figure 1.

B. ETSI ENI Framework

Parallel to recommendations presented by ITU, ETSI derived its own recommendation regarding the AI framework, named ENI (Experiential Network Intelligence) [15]. The goal behind the architecture presented is to provide a unified distributed platform, capable of addressing problems of the next generation networks using AI. The basis is a policy-driven closed loop model, which can dynamically adapt the network behavior in accordance with change of inputs. The system utilizing services offered by the ENI system is denoted as “Assisted System” (AS). While ITU architecture seems more modular, in terms of ML functionality placement, the ETSI architecture describes the ENI system as isolated segment/box with API interface for communication with the Assisted System (Figure 2).

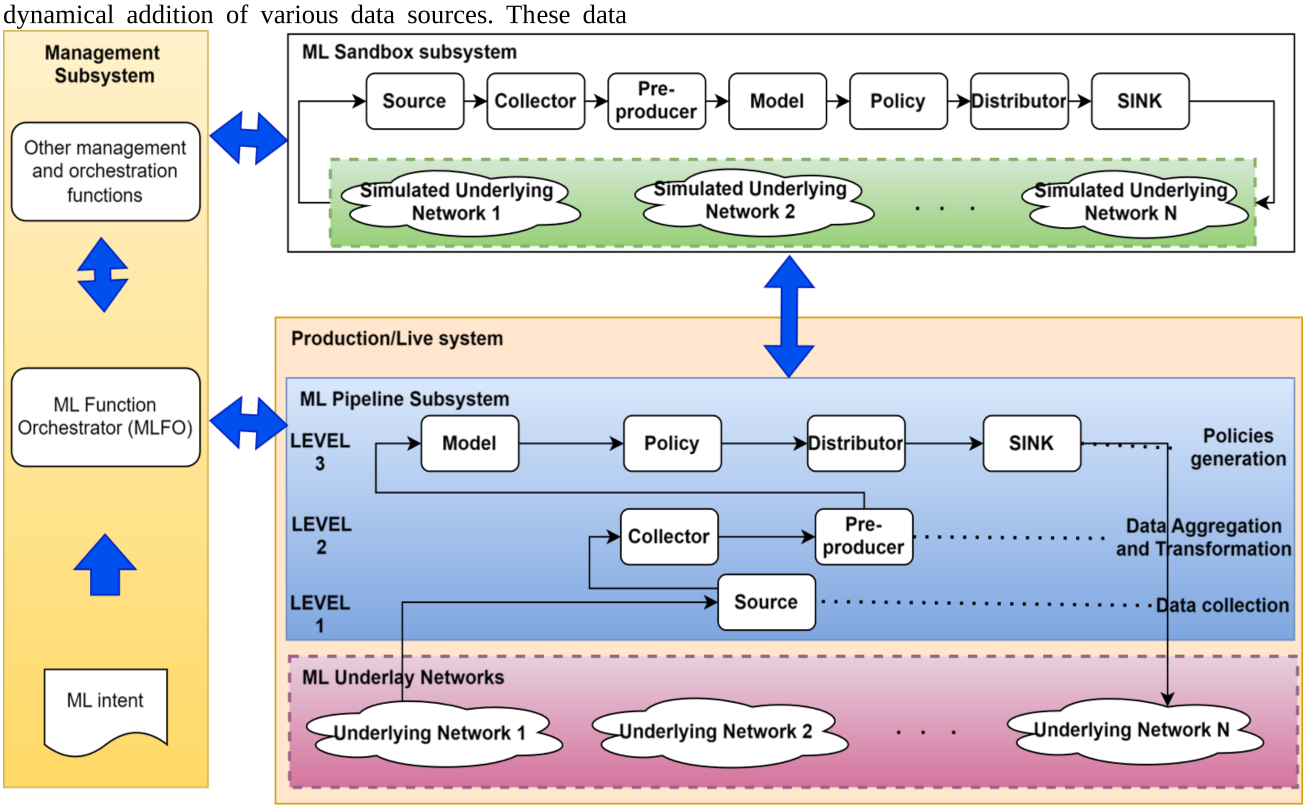

Fig. 1 ITU-T.3172 AI/ML Architecture Framework

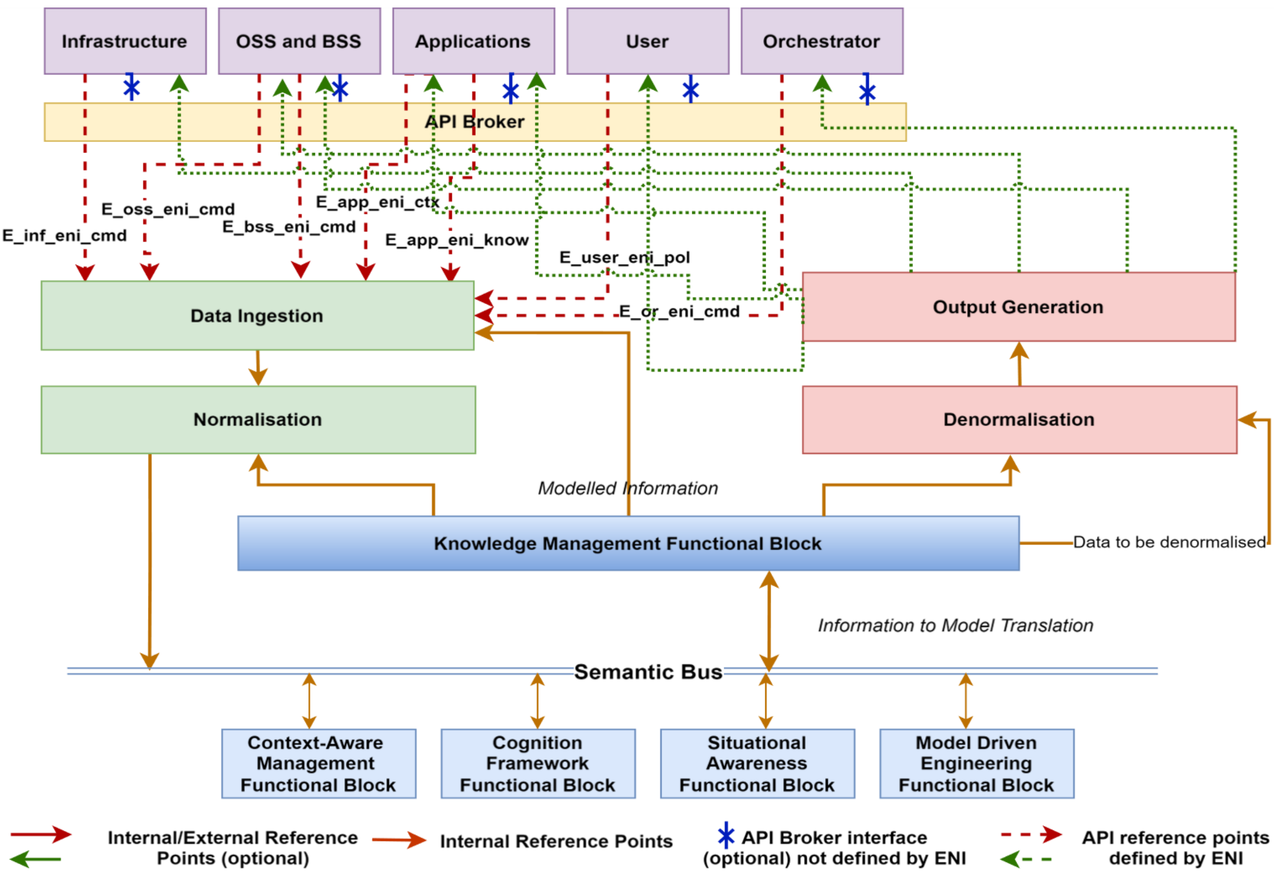

Fig. 2 ENI System architecture with defined interfaces and reference points

A major part is devoted to the API Broker, which is crucial for communication with assisted systems that does not support ENI External Reference Points (RESTful interfaces) and ENI APIs. Since AI presents a new level of functionality in the networks, it is logical to assume that NFs interfaces are not directly compatible with the aforementioned. The recommendation classifies the AS in three levels, based on AI capabilities – none, only for non-real-time decisions, and real and non-real time capable AI AS systems. The ENI system presents two modes of operation, i.e. “recommendation mode” and “management mode”. The first is used only for providing recommendations to the AS, which can later be used in the internal (AS) deduction of policies and rules. The second is used for providing decisions and commands to the Assisted System. Furthermore, a hybrid mode is presented where, after the AS validates that recommendations provided in “recommendation mode” are compliant with the internally derived results, it can switch to “management mode”. The ENI system is divided in three logical functional blocks: input processing, analysis, and output generation. The input generation is consisted of two sub-blocks: data ingestion and normalization functional blocks. As the name suggest, the purpose of data ingestion is to collect data from various sources and additionally apply common data processing techniques to enable further processing by other ENI functional blocks. As previously defined in ITU architecture, data streams can be ingested by different network domains (Access Network, Transport, and Core) or Assisted Systems (separate segments in the network domains). The communication is performed by using External Reference Points or API Broker. This stage can be mapped to data preprocessing in classical ML, with stages as data filtering, data correlation, data augmentation, and data labeling. In addition, if private or biometric data is used, ENI suggests a data anonymization and pseudonymization phase. The normalization functional block serves as mediator between the external systems and the ENI system, i.e. it unifies data parameters towards internal ENI functional blocks. The generalization enables vendor interoperability and neutrality. Next in line is the analysis phase. It consists of knowledge management, context awareness, cognition management, situation awareness, model-driven engineering, and policy management functional block. The knowledge management block serves for differentiation between known facts, axioms and inference. It uses different mathematical models for presenting information and entropy. This block is the heart of the ENI system and is used by every other functional block.

Context Awareness Management Function serves for extracting the state and environment from which rules can be derived for a given subset of entities in the Assisted System.

The brain of the ENI system is the cognition management functional block. It evaluates the meaning of processed data, considering both the context and predefined objectives of the system of interest.

The situational awareness block provides real-time monitoring and understanding of the surrounding environment, enabling informed decision-making. Meanwhile, the model-driven function block utilizes predefined models to guide its actions and responses, ensuring efficient and effective system operation.

The last layer is output generation. It consists of two blocks: denormalization and output generation functional blocks. Denormalization is used for translating ENI system information and data from various internal formats into standardized form. For example, customer data can be saved in different formats as Lightweight Directory Access Protocol (LDAP) and Relational Database Management System (RDBMS). The output generation functional block can be integrated or separated from the denormalization block. It converts data received from the denormalization step (recommendations or management configurations) into format suitable for the Assisted System of interest. The overall ENI system architecture is depicted in Figure 2.

C. Fault and Recovery Model

One of the most accentuated features in terms of operations, maintenance and regulatory legislation is the fault and recovery function, presenting an obligatory part of every AI system. ITU described such model in the recommendation ITU-T Y.3177 (Fig.3), which is compliant and follows the basic framework introduced in ITU-T Y.3172. The fundamental components outlined in the model largely adhere to the recommendations outlined in the ENI architecture. The main constituents of the model are resource management and fault management pipelines [10]. The sandbox is a crucial part of the framework, used for simulating scenarios and improving the ML model. This component is particularly vital for fault management as the anticipated number of faults at the production level is minimal, rendering it unsuitable for ML training and refinement. The resource management segment consists of two main components: resource prediction and resource adaptation. The first is used for data gathering (performance metrics, load, and resource utilization) and AI/ML prediction. The inference obtained as output of AI/ML models is further used as basis for resource prediction, subsequently guiding resource allocation and adaptation procedures. The latter is presented by multistage pipeline of decision functions, with tendency for optimal resource slicing. Resource adaptation can be carried in two ways: reactively and proactively. It is defined by three functions: resource arbitration, network function migration, and network slice reconfiguration. The main aim of the resource arbitration function is to optimally allocate available resources of the underlying network across different network slices, taking into consideration the plethora of parameters that describe the requirements and importance of services (QoS, delay, jitter, bandwidth, etc.). AI/ML models are crucial for executing optimization procedures of this caliber. Network function migration functionality is responsible for agile and effective migration of vNFs/cNFs, aiming at meeting the service requirements while optimizing the overall performance of the network. It delegates the process of vNFs/cNFs placement and resource utilization of different physical nodes in the underlay infrastructure.

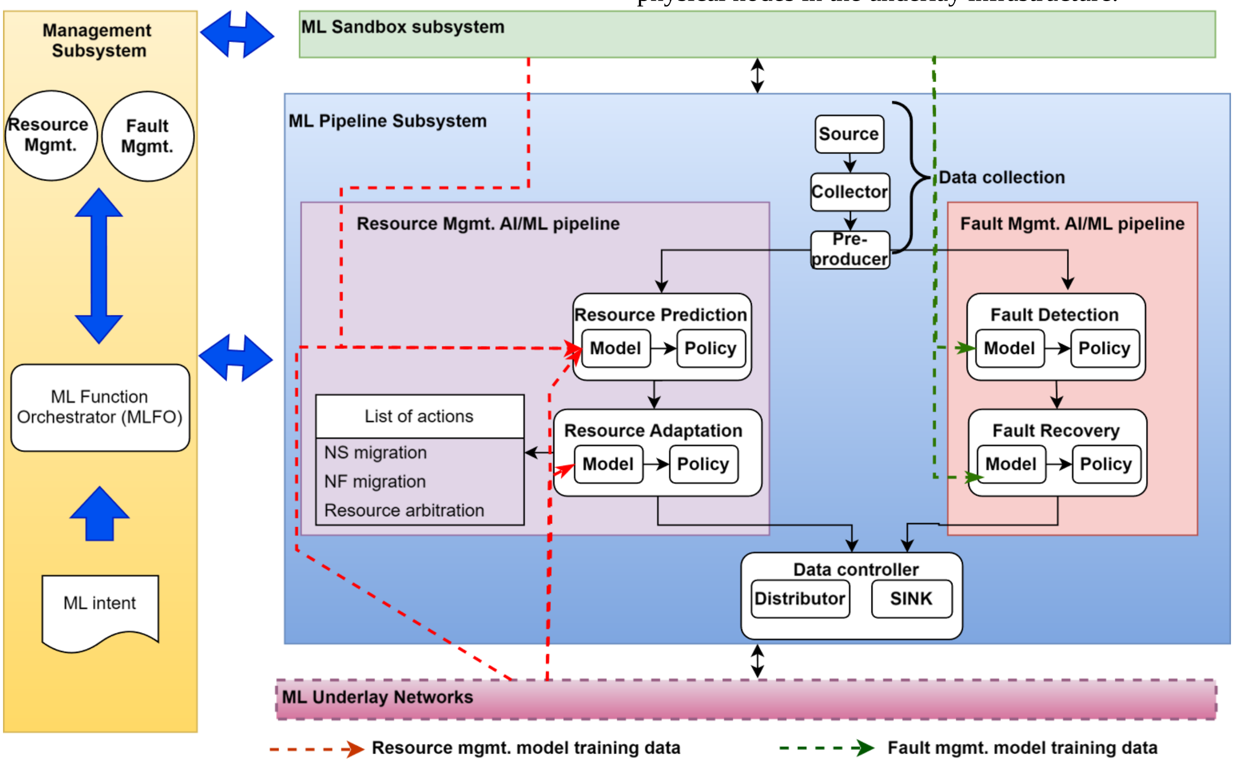

Fig. 3 ITU-T Y.3177 Framework Recommendation for Resource and Fault Management

However, this function cannot instantiate additional physical resources, i.e. can only manipulate with the ones linked to the network slices of interest. To make the model more flexible, network slice reconfiguration is introduced. This function operates slower than the previous two, because it requires interactions with underlay resource controllers. The network slice reconfiguration enables scale in/out of new physical nodes, supporting specific NFs.

Fault management is one of the most important segments of implementing different security mechanisms in the AI/ML network model. It consists of fault detection and fault recovery. Verification and improvement of different AI/ML models is generally performed by using the AI/ML sandbox.

I. III. Standardization and Regulation

With the growing popularity of AI, regulatory bodies worldwide are focusing on standardizing and regulating its use. In April 2021, the European Commission proposed the first European AI regulatory framework. Over the next two years, various groups worked on defining AI use, its weaknesses, and security flaws. In December 2023, an agreement was reached with the European Parliament, leading to the final adoption of the EU AI Act on March 13, 2024, with 523 votes in favor, 46 against, and 49 abstentions [18]. The Act classifies AI use into four categories: unacceptable (banned), high risk, limited risk, and minimal risk. It also covers general-purpose AI models used by major tech companies. Non-compliance with banned AI practices can result in fines up to EUR35 mil. or 7% of a company's global annual turnover, whichever is higher [19].

In the United States., AI regulation is largely state-level, with no comprehensive federal framework. Between 2016 and 2022, 14 states have passed AI-related legislation, with Maryland being the leader. However, only 9 out of 88 federal AI legislative acts were passed [20]. In October 2023, President Joe Biden issued an executive order (EO) on “Safe, Secure, and Trustworthy Development and Use of AI” [21]. This move suggests that the U.S. telecommunications sector may face similar regulations to the EU AI Act [22].

China's AI legislation is divided into two segments: government AI systems and the private sector [23]. The government sector faces looser regulations, with most categories permitted, including social scoring, already implemented in some areas. China has stricter laws for “algorithmic recommendation” due to the backlash against *ByteDance*'s media apps in 2017. In 2021, China introduced legislation requiring algorithm providers to uphold societal values and comply with standards similar to the GDPR. Recommendations must align with ethical principles and avoid promoting addictive content. All recommendation algorithms must be submitted for government approval.

On July 13, 2023, China introduced the “Interim Measures for Management of Generative Artificial Intelligence Services”, in effect from August 15, 2023 [24]. These regulations are vital for AI models like chatbots and AI agents in telecommunications and personalized content delivery.

Although AI regulations in telecommunications will likely be similar, this paper adheres to the EU AI Act. Given the definition of high-risk AI systems, many telecommunications use cases will fall under this category. High-risk AI includes non-banned biometrics, critical infrastructure (e.g., digital infrastructure, road traffic, utilities), education, law enforcement, and administration of justice and democratic processes. Telecommunications, as part of the critical infrastructure, are subject to these regulations. Key use cases involve data processing for network management, personalized content, targeting, trend analysis, and churn analysis. Profiling, a key business driver, is considered high-risk under Article III of the EU AI Act and aligned with GDPR definitions [19].

Telecommunications network equipment, categorized as critical infrastructure [Annex I, point (6)], relies on software packages on COTS hardware with APIs (HTTP/REST) between network functions. Annex IV mandates detailed documentation for software and APIs related to ML functionalities. To ensure compliance with the legislation, high-risk systems must adhere to requirements and provide appropriate documentation. AI system providers must ensure their products comply with the EU harmonization legislation [Annex I, Section A]. The Act mandates a continuous risk management system, supported by documentation as specified in Article 11, for testing high-risk systems against predefined thresholds and metrics.

While some network functions (RAN, CN, MEC) can operate without personal data, many business use cases involve processing biometric and personal data. Access to customer data (e.g., CRM, CDR/EDR, DPI systems, chatbots) must comply with GDPR Article 9(1) and national laws, unless prohibited by the EU Act. The risk management system must address risks of biased or discriminatory results based on biometric data such as social status, race, or ethnicity. Data management must align with GDPR, which requires a legitimate basis for data processing, such as consent or contractual agreement.

Chatbots, often general-purpose AI systems (GPAIS), must follow special design criteria laid out in Annex XI and Annex XIII. In addition to internal controls, ML providers are subject to external regulatory oversight. Article 12 mandates automatic log keeping for high-risk AI systems throughout their lifetime, with retention of at least six months (Article 19). High-risk AI systems also require certification, with a validity period of four years (Annex III).

In addition to formal legislation, numerous ethical playbooks and recommendations from organizations like GSMA provide methodologies and self-assessment tools for implementing lawful AI/ML practices in telecom networks. GSMA’s AI for Impact (AI4I) initiative outlines three levels of responsibility for ethical AI deployment. The first level involves AI product managers and teams who manage low-risk AI applications using self-assessment tools. The second level, ethics committee, evaluates higher-risk applications with expert input. The executive board, at the third level, makes final deployment decisions, often in consultation with the ethics committee. Many CSPs have established internal bodies for AI/ML risk and security assessments, both pre-deployment and during the system's lifecycle.

Orange formed its Data and AI Ethics Council in March

2021, consisting of 11 independent experts from diverse professional backgrounds. This independent body advises on ethical issues and appoints Data and AI Ethics Officer in each country for local assessment and reporting to the Council. Similarly, *Telstra* created a Risk Council for AI and Data (RCAID), which evaluates AI/ML software at all stages of deployment. For high-risk applications, the evaluation is escalated to *Telstra*'s Data and AI Council. *Telefonica* follows the Responsible AI by Design methodology, which incorporates a three-tiered structure to assess AI/ML risk levels.

These regulations aim to shape AI use cases and business models while ensuring fair market competition and interoperability between ML functions and network interfaces [26].

II. IV. Business: Aims, Opportunities and Realizations

The main drivers of commercially available technologies are business use cases and opportunities. Emergence of AI/ML models has created new business models, profit avenues, and savings. Many companies have started implementing these models across various segments of their networks and business portfolios, aligning with forecasted growth. Numerous papers and reports highlight potential use cases. The Body of European Regulators for Electronic Communications (BEREC) classifies AI use cases into six main areas: network and capacity planning, channel modeling, dynamic spectrum sharing (DSS), QoS optimization, security optimization, and fraud detection [27].

In the RAN segment, *Nokia* and *NVIDIA* announced a collaboration to combine Nokia Cloud RAN with NVIDIA's AI-powered processing [28]. This may enhance the benefits of Nokia Cloud RAN, including a 15% reduction in base station power consumption, 70% fewer alarms, and 80% efficiency gains with zero-touch optimization. Additionally, Open Radio Access Networks (O-RAN) promote open standards, enabling cost-efficiency and flexibility in telecom networks. AI/ML integration in O-RAN supports real-time network optimization, intelligent resource management, and automation. Companies like *Nokia* and *NVIDIA* are integrating AI-driven O-RAN solutions for improved service delivery and scalability [27] [28].

Intelligence and automation are crucial across all telecom networks, with the ultimate goal being end-to-end service optimization for guaranteed QoS. This is known as zero-touch operation. ETSI formed the Zero-Touch Network and Service Management (ZSM) group in 2017 to create a framework for interoperability among vendor-specific solutions [29] [30]. *Nokia* has reported an 80% operational efficiency increase in a proof-of-concept with a North American operator using their Self-Organizing Network (SON) software, which utilizes AI/ML for cognitive automation. *SK Telecom*'s TANGO (Telco Advanced Next-Generation OSS) is another successful implementation, offering an AI-assisted network operating system that integrates and analyzes data from various network segments [31]. TANGO is expected to recoup CAPEX and OPEX within five years and further reduce costs by 40% over the next five years. The zero-touch provisioning market was valued at USD2,772.13 mil.in 2022 and is forecasted to grow to USD7,544.15 mil.by 2032, with CAGR of 10.6% [32].

NTT, *Huawei* and other industry leaders are advancing development of versatile language models, targeting a wide range of applications across different sectors. These models, similar to ChatGPT, offer natural language understanding and generation across sectors [33]. The model comes in two versions: an ultra-lightweight version (600 million parameters) and a lightweight version (7 billion parameters), far smaller than OpenAI's GPT-3 (175 billion parameters). This reduced complexity allows efficient deployment on single GPU or even CPU, lowering training and maintenance costs and providing faster ROI. For context, GPT-3's training requires 1300 MWh, equivalent to one hour of nuclear power generation. NTT's "tsuzumi" model outperforms GPT-3.5, scoring 81.3% on the Rakuda benchmark. Its ability to adapt to different business models makes it suitable for customer care operations in telecoms.

Huawei has developed telecommunications-specific models, NET4AI and AI4Net, which cater to the new generation of networks. NET4AI supports AI-as-a-service (AIaaS), forming the backbone for AI-powered services, while AI4Net focuses on intelligent operations, enhancing network performance, resource management, security, and user experience [34].

In addition to backend applications, several use cases directly engage end consumers. For instance, *AT&T* enhanced its "Call Before You Dig" platform with AI/ML, improving geospatial recognition and report analysis. This led to a 7% reduction in miles traveled per dispatch and 5% increase in productivity [35]. *Deutsche Telekom* deployed three chatbots: Tinka, Vanda, and Hub:raum. Tinka, used in Austria, answers 80% of customer queries, forwarding the rest for human analysis. Vanda is a similar NLP-based chatbot, and Hub:raum assists with recruitment queries [36].

Vodafone introduced its chatbot, TOBi, which enhances customer communication through mobile apps, WhatsApp, and SMS, automating two-thirds of customer interactions in Italy, leading to a 68% improvement in customer experience [37]. *Colt Technology Services* launched the Sentio AI/ML assistant, automating service provisioning and providing flexible bandwidth in real time.

Aggressive adoption of AI/ML in telecom networks drives technological advancement, reduces CAPEX/OPEX, and increases revenues. The AI telecom market was valued at USD1.45 billion in 2022, with forecasted CAGR of 28.2% from 2023 to 2030. North America accounted for 34.8% of the market in 2022, but Asia Pacific is expected to have the fastest growth due to investments in China and India [38]. China leads AI patents with 61.1%, while the U.S. invest USD67.2 billion in the private AI sector, about 8.7 times more than China [39]. McKinsey reports that 42% of companies reduce costs with AI, and 59% see revenue increases, with tech, media, and telecom sectors leading AI adoption for product and service development [40]. AI-driven predictive maintenance can cut telecom maintenance costs by 30-40%, and behavior prediction models may reduce churn by 10-15%. Over 75% of telecom companies plan to invest in AI systems in the next three years [41].

III. V. CONCLUSION

This paper explores AI/ML advancements and challenges in the telecom sector. A robust technological and regulatory framework is essential for ensuring interoperability, privacy, safety, and business sustainability. Key recommendations from ITU-T (Recommendation Y.3172) and ETSI (ENI framework) serve as the foundation, with fault and recovery functionalities crucial for model sustainability. Regulatory developments, including the EU AI Act, the U.S. state-level legislation, and China's dual-focused framework, emphasize the need for ethical AI deployment. Industry-led initiatives like GSMA's Ethical Requirements playbook further support responsible AI use. Legislative frameworks will shape telecom network development, especially for high-risk applications, though diverse global regulations complicate deployments. AI/ML integration unlocks new business opportunities, particularly in network optimization and customer service. AI-powered chatbots by *AT&T*, *Deutsche Telekom*, and *Vodafone* have improved customer experience with faster query resolution and accessibility. Statistical trends indicate AI's transformative impact on telecom operations, maintenance, and architecture.

REFERENCES

[1] Toni Janevski. "QoS for Fixed and Mobile Ultra-Broadband". Wiley – IEEE Press, 2019
[2] Toni Janevski. "Future Fixed and Mobile Broadband Internet, Clouds, and IoT/AI". Wiley – IEEE Press, 2024.
[3] Huawei. "6G: The Next Horizon. From Connected People and Things to Connected Intelligence", 2020.
[4] Nokia. "Responsible AI for Telecom", 2021.
[5] NVIDIA. "State of AI in Telecommunications: 2024 Trends Survey Report", https://resources.nvidia.com/en-us-ai-in-telco/state-of-ai-in-telco-2024-report [last retrieved in October 2024].
[6] Recommendation ITU-T Y.3115, "AI-enabled cross-domain network architectural requirements and framework for future networks including IMT-2020". 2022.
[7] Recommendation ITU-T Y.3170, "Requirements for machine learning-based quality of service assurance for the IMT-2020 network". 2018.
[8] Recommendation ITU-T Y.3172, "Architectural framework for machine learning in future networks including IMT-2020". 2019.
[9] Recommendation ITU-T Y.3183, "Framework for network slicing management assisted by machine learning leveraging quality of experience feedback from verticals". 2023.
[10] Recommendation ITU-T Y.3177, "Architectural framework for artificial intelligence-based network automation for resource and fault management in future networks including IMT-2020". 2021.
[11] ETSI. ETSI GS ENI 001 V3.2.1, "Experiential Networked Intelligence (ENI); ENI Use Cases". May 2023.
[12] ETSI. ETSI GS ENI 002 V3.2.1, "Experiential Networked Intelligence (ENI); ENI Requirements". April 2023.
[13] ETSI. ETSI GR ENI 003 V1.1.1, "Experiential Networked Intelligence (ENI); Context-Aware Policy Management Gap Analysis". May 2018.
[14] ETSI. ETSI GR ENI 004 V3.1.1, "Experiential Networked Intelligence (ENI); Terminology". July 2023.
[15] ETSI. ETSI GS ENI 005 V3.1.1, "Experiential Networked Intelligence (ENI); System Architecture". June 2023
[16] ETSI. ETSI GS ENI 006 V2.1.1, "Experiential Networked Intelligence (ENI); Proof of Concepts Framework". May 2020.
[17] ETSI. Ongoing PoCs, https://eniwiki.etsi.org/index.php?title=Ongoing_PoCs, [last retrieved in October 2024] .
[18] European Parliament. "Artificial Intelligence Act: MEPs Adopt Landmark Law". 2024.
[19] European Parliament. "Artificial Intelligence Act". 2024.
[20] HAI. "AI Index Report 2023". 2023.
[21] White House. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, (2023-10), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/, [last retrieved in October 2024].
[22] European Parliament. "United States' Approach to Artificial Intelligence" 2024.
[23] Matt Sheehan/ "China's AI Regulations and How They Get Made". Carnegie Endowment for International Peace, Washington, DC, 2023.
[24] China Law Translate. "Interim Measures for Management of Generative Artificial Intelligence Services". July 13, 2023,
[25] Official Journal of the European Union, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, April 27, 2016.
[26] GSMA. "The AI Ethics Playbook: Implementing Ethical Principles into Everyday Business". 2022.
[27] BEREC. "Report on the impact of Artificial Intelligence (AI) solutions in the telecommunications sector on regulation". June 8, 2023.
[28] Nokia. "Nokia to Revolutionize Mobile Networks with Cloud RAN and AI Powered by NVIDIA". February 24, 2024.
[29] ETSI. "Zero-touch Network and Service Management". December 7, 2017.
[30] ETSI. "Zero-touch network Service Management group renewed for two years". October 5, 2023.
[31] GSMA. "Innovator Profile: SKT Tango". October 14, 2019.
[32] POLARIS. "AI in Telecommunication Market Share, Size, Trends, Industry Analysis Report, By Region, And Segment Forecasts, 2023 – 2032". January 2024.
[33] NTT R&D. NTT's Large Language Models 'tsuzumi'. April 25, 2024, https://www.rd.ntt/e/research/LLM_tsuzumi.html, [last retrieved in October 2024].
[34] HuaweiTech, NET4AI: Supporting AI as a Service in 6G, November, 2022, https://www.huawei.com/en/huaweitech/future-technologies/net4ai-supporting-ai-as-a-sevice-6g, [last retrieved in October 2024].
[35] Eric Zavesky. "Call Before You Dig: Using AI and Data Science to Protect Buried Cables and Keep Construction Projects on Track". December 1, 2023, https://about.att.com/blogs/2023/call-before-you-dig.html, [last retrieved in October 2024].
[36] AIT Staff Writer. "Top 10 AI-Powered Telecom Companies in World", AITHORITY. March 4, 2020, https://aithority.com/technology/analytics/top-10-ai-powered-telecom-companies-in-world, [last retrieved in October 2024].
[37] Neil Blafden. "Artificial brains and predictive care: Vodafone's digital journey". Vodafone. December 15, 2019, https://www.vodafone.co.uk/newscentre/viewpoint/artificial-brains-and-predictive-care-vodafones-digital-journey/, [last retrieved in October 2024].
[38] GVR. "AI in Telecommunication Market Size, Share & Trends Analysis Report By Application (Network Security, Network Optimization, Customer Analytics, Virtual Assistance, Self-Diagnostics), By Region, And Segment Forecasts, 2023 – 2030", https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-telecommunication-market#, [last retrieved in October 2024].
[39] HAI. "AI Index Report 2024", 2024. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024.pdf, [last retrieved in October 2024].
[40] McKinsey. "The State of AI in 2023: Generative AI's Breakout Year". August, 2023.
[41] GITNUX. "MARKETDATA REPORT 2024: AI in the Telecommunications Industry Statistics". April 2024.