ETSI White Paper No. 64

AI Technologies in Experiential Networked Intelligence to Increase Autonomous Operation

1st edition – November 2024

Authors:

Yu Zeng, John Strassner, Jingyu Wang, Fabrizio Granelli, Pietro Cassarà, Raymond Forbes, Luigi Licciardi, Aldo Artigiani

ETSI
06921 Sophia Antipolis CEDEX, France
Tel +33 4 92 94 42 00
info@etsi.org
www.etsi.org

Contents

1. Executive Summary

ISG ENI focuses on defining the functionalities and architecture to increase the Autonomous operation of a communication network, thereby enhancing the overall operator experience. Artificial Intelligence (AI) technologies in ENI can be used throughout the lifecycle of the various domains making up the infrastructure (e.g., Campus, Radio, Fixed Access, Backbone, Core, Data centre) as well as the lifecycle of the services provided to the end users (e.g., VPN, IoT, mobile, customized slices of application-specific functionality, and fixed access). The ENI architecture contains two main functionalities, (1) policy definition and management, and (2) cognition management. The former enables any user to input policies and requests to the ENI system in a standard form that ENI can process and respond to in a standard format. The latter provides enhanced and explainable learning and decision-making, and uses those policies to grant the expected end-user experience in a safe and scalable manner.

2. Autonomous Operation using Level Categorization

There is a common understanding in the autonomous network (AN) expert's community to identify Levels of autonomy from Level 0 (no use of autonomy) to Level 5 (full autonomy), as shown in Table 2.1. The concept expressed is related to a similar definition used in Autonomous Driving Cars, where Level 0 is fully manual and Level 5 means fully autonomous driving (no human intervention). This means that the level of Autonomy increases with the gradual introduction of autonomation in Network creation and management from Level 0 to Level 5.

In a similar way to Autonomous Driving Car, a progressive evolution is planned to move forward to a fully AN. AI models, in both their predictive and generative forms, can support the Autonomous Networks evolution. Increasing their use will increase the Level of Autonomy.

Telco and digital service providers are taking care of the evolution of Autonomy in their Networks with increasing interest. Some are evaluating the opportunity to assess the Level of Autonomy in their Networks. This implies common definitions of the Levels of autonomy, introducing clear rules and measures (e.g., Key Product and Quality Indicators, or KPIs and KQIs) to make this assessment.

Even if most Operators are now positioned in Level 1 or 2, there is a clear wish to move forward to Levels 3 and 4 in the next 5 years. This goal seems reasonable to reach using AI. The assessment could be in terms internal and external to Network creation and management in terms of Autonomy, and might be proposed to Customers as a KPIs and/or KQIs. Nevertheless, vendors can also be impacted by the AN-level assessment, when operators require vendors to provide them with systems and equipment compliant with a specific level of autonomy.

Table 2.1: Categories of network intelligence from a technical point of view [i.3]

Level	Name	Definition	Man-Machine Interface	Decision Making Participation	Decision Making and Analysis	Degree of Intelligence	Environment Adaptability	Supported Scenario
0	Manual network	OAM Personnel manually manage the network and obtain alarms and logs	How (command)	All-manual	Shallow awareness (from events and alarms)	Manual understanding	Fixed	Single scenario
1	Partially automated network with some automated diagnostics	Automated scripts are used in service provisioning, network deployment, and maintenance. Shallow perception of network status and decision-making suggestions of machine	How (command)	Provide suggestions for machines or humans and help decision-making	Local awareness (SNMP/YANG events, alarms, KPIs, and logs)	A small amount of analysis	Little change	Few scenarios
2	Automated network	Automation of most service provisioning, network deployment, and maintenance Comprehensive perception of network status and local machine decision-making	HOW (declarative)	The machine provides multiple opinions, and then chooses one	Increased awareness (Telemetry-provided basic data)	Powerful analysis	Little change	Few scenarios
3	Self-optimization network	Deep awareness of network status and automatic network control, meeting users' network intentions	HOW (declarative)	Most of the machines make decisions	Comprehensive and adaptive sensing (such as data compression and optimization technologies)	Comprehensive knowledge with enhanced prediction	Changeable	Multiple scenarios and combinations
4	Partial autonomous network	In a limited environment, people do not need to participate in decision-making and network can self-adapt	WHAT (intent)	Optional decision-making response (decision typically needs human approval)	Adaptive posture awareness (edge collection and judgment)	Comprehensive knowledge and forward forecast	Changeable	Multiple scenarios and combinations
5	Autonomous network	In different network environments and network conditions, the network can automatically adapt to and adjust to meet people's intentions	WHAT (intent)	Machine self-decision	Adaptive optimization (E2E closed-loop, including collection, judgment, and decision-making)	Self-evolution and knowledge-based reasoning	Any change	Any scenario & combination

When upgrades move toward Level 4 it is a clear challenge for Network Operators in terms of a significant and tangible evolution, in the 3-dimensional domain: network, service, and operation. The Level assessment seems to be of significant interest to monitoring and measuring progress to reach the targeted Level of Autonomy. It should be significantly easier to reach for vendors, since they are focussed on supplying a component of the system or network.

ENI proposed adding Autonomous levels at different levels in the Telco Network, from single equipment in a single Domain to the overall combination of network resources and services. Detailed information and documentation can be found in ETSI GR ENI 007 [27] and ETSI GR ENI 010 [28].

3. Use cases and requirements

3.1. Introduction

As one of the key technologies that can help network operations move towards L4 autonomous operation, mainstream operators are considering adopting the latest AI technologies (e.g., Large Language Model (LLM), Generative AI (GenAI), transformers, etc.) to help improve operational efficiency.

3.2. Use Case 1: Network Awareness

Network awareness refers to the operator understanding the state and status of the network at any given time. Scenarios include knowledge Question and Answering (Q&A) and intelligent Management, and Administration Operation, Operation and Maintenance (OAM) assistants in a single- and multi-agent form. It will be fully integrated into other tools and systems to support network management operations and the whole production process of AN. The European Commissioner for Internal Market Thierry Breton's MWC Barcelona keynote speech mentioned on Feb 2024, "The EU should … set up end-to-end integrated infrastructures and platforms for **telco cloud** and **edge**, which could be used to orchestrate the development of innovative technologies and **AI applications** for various **use cases**" [6]. Table 3.2.1 summarizes this use case.

Table 3.2.1: Use Case 1 – Network Awareness

Use Case name		Problem Description	Solution Highlights	Business Value
Comprehensive monitoring and network status awareness.		Cloud networks and services are becoming increasingly complex, making detecting and preventing faults in advance difficult. The root cause of the problem is complex to locate and takes a long time to evaluate.	Predictive maintenance enables problems to be inferred before they occur.	Improving network awareness is mandatory for improving network OAM operations. This also includes more efficient scheduling and troubleshooting.

Network status data are scattered in multiple systems in different locations. This complicates cross-system coordination, which in turn makes analysis and remediation difficult and slow. The process of analyzing network status requires significant labour effort and specialized skills. In addition, traditional rule-based and pattern-matching methods cannot realize intelligent prediction and analysis of unknown and new cyber threats and lack risk assessment optimization schemes.

New network status awareness applications based on a LLM enable the ingestion and analysis of massive network telemetry and other related information. Operational domain knowledge is infused into a LLM to enable its analysis of network status data to be customized to a particular application and environment. A LLM can provide a comprehensive analysis of different data and customized inference of what those data mean. This can then be used for a variety of other tasks, such as operation optimization and enhanced operation decision-making.

3.3. Use Case 2: Intelligent IP Network Simulation

IP networks carry a large number of voice and data services across different regions, and minor network configuration errors can cause huge economic losses and social impacts. According to statistics, more than 70% of major IP network failures are caused by manual configuration errors. Online simulation and verification tools can be used to assess proposed network configuration risks in advance, thereby reducing network failures. Table 3.3.1 summarizes this use case.

Table 3.3.1: Use Case 2 – Online Intelligent Simulation for IP Network

Use Case name	Problem Description	Solution Highlights	Business Value
IP network intelligent online simulation	Reduce major failures caused by manual configuration errors and improve configuration efficiency and security	A digital map is built based on digital twin technology, with embedded high-precision simulation capabilities, and network verification algorithms	Increased risk prevention and associated annual cost prevention will be significant

Incorrect Quality of Service (QoS) configurations may adversely affect millions of users. Incorrect QoS configuration can lead to several problems, including service outages, network congestion, poor service due to increased latency, jitter, packet loss, and the introduction of security vulnerabilities. This employed digital twin technology with an embedded high-precision simulation system that generated data to assess the risk of proposed network changes. The network risk assessment is carried out by using CPV/DPV (Control/Data Plane Verification). Highlights of the solution include:

- A network digital twin provides a realistic digital verification environment, records the status and behaviour of the digital twin in real-time, supports the traceability and playback of historical data, and greatly reduces the cost of trial and error.
- High-precision network protocol simulation supports multi-vendor devices that use more than 20 mainstream routing protocols to generate realistic traffic simulation. The impact of changes on routes, traffic paths, link loads, and other pertinent factors affecting performance can be identified in advance.
- Network verification algorithm formalizes network verification intents and rules for network-wide connectivity verification, network-wide loop verification, network-wide problems and anomalies, and output verification reports.

It was shown to effectively take care of IP network security, increase the accuracy rate of preventing network change risks before most risk happen, intercept high-risk operations with very good results, and avoid economic losses and social impacts caused by potential problems and risks. It is expected to prevent economic losses of more than 130 million Euro in value per year when applied nationwide.

3.4. Use Case 3: Network Maintenance

Operators face the following challenges in network maintenance scenarios:

- How to accurately identify and output results that meet user requirements based on user and business policies conveyed in the form of intents?
- How to improve OAM, which will also improve the user experience?
- How to accurately locate cross-layer network faults and shorten the locating time?
- How to overcome the lack of in-depth cross-layer fault analysis tools and applications?
- How to increase knowledge dissemination for on-site maintenance operations?
- How to achieve dynamic orchestration and end-to-end task execution and make network operations more autonomous?

Table 3.4.1: Use Case 3 – Using a Network LLM for Network Maintenance

Use Case name	Problem Description	Solution Highlights	Business Value
Innovative practice of network LLM in network maintenance scenario.	The operator faces network maintenance challenges including the inability to understand and disambiguate intent policies, conduct in-depth cross-layer fault analysis, perform efficient OAM decision-making, and execute end-to-end management tasks.	The solution has been applied to 8 scenarios such as on-site maintenance and emergency support, with an average fault location accuracy of more than 80%.	10-second quick Q&A, minute-level cross-layer fault locating, 20% increase in fault handling efficiency, and 50% reduction in on-site installation and maintenance.

The high-level technology evolution requirements for its network LLM include intent creation, management, and processing; improved analysis of ingested and inferred data; and enhanced decision-making and execution in AN closed-loop network maintenance scenarios. By learning professional knowledge and business rules in the telco field, the LLM is enhanced with application-specific knowledge and understanding.

Business rules teach it intelligent scheduling and help improve its decision-making. Since network maintenance is made up of multiple specialized applications, this overall solution can scale through the use of a mixture-of-experts model [17] and finetuning mechanisms [17] to effectively empower AI applications and improve the value of network operations.

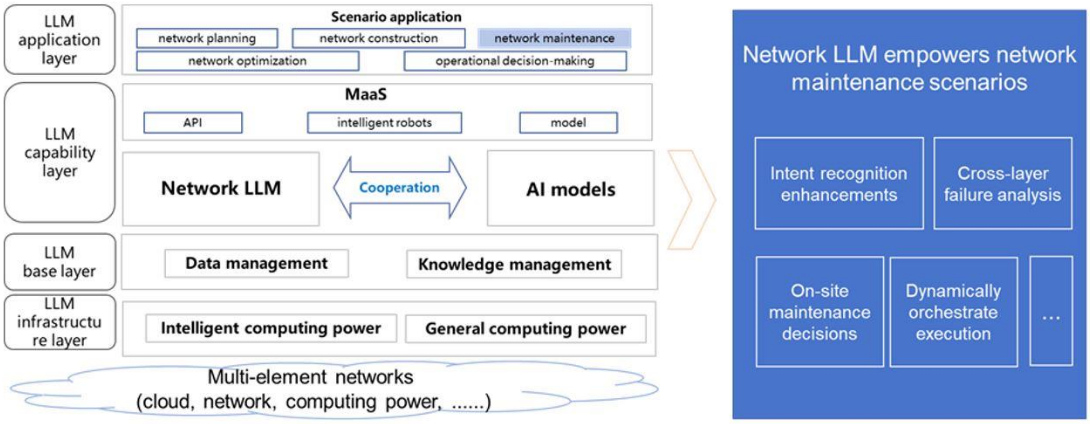

Figure 3.4.1: Architectural Diagram of the Network Maintenance Scenario

Accurate intent recognition: ENI [15] contains an innovative policy model [16] that can represent imperative, declarative, and intent policies. This enables each to call the other, and simplifies their parsing and application through the use of common abstractions. The next phase of this use case will connect the network LLM to knowledge management in ENI (e.g., knowledge repositories as well as a knowledge graph for enhanced reasoning with explanations). This combination will also reduce the possibility of hallucinations from the LLM.

Cross-layer fault analysis: cross-layer fault location can be realized by using a combination of knowledge graphs [15] [17] to represent cross-layer operation, knowledge retrieval enhancements using retrieval augmented generation [17] technology, and additional software to perform root cause analysis such as a Cognitive Assistant [22].

Intelligent interactive decision-making: Natural language makes OAM more accessible to business users. A LLM can improve data query efficiency by enabling users to understand data and relationships between data. The LLM, when paired with ENI cognitive capabilities, can assist in finetuning the context for a query, thereby producing more accurate results. It can also seed the knowledge of the LLM and knowledge graph by incorporating expert knowledge from maintenance personnel.

This architecture supports the use of mobile phones to query multi-system equipment and professional network management indicators anytime and anywhere, and the results can be returned within 10 seconds, the cross-layer fault locating is improved to the minute level, the fault neutralization efficiency is increased by 20%, the on-site installation and maintenance hour is reduced by 50%, and the automatic completion rate of complaint work orders is increased by 10%.

3.5. Getting Ready for Partial Autonomous Operation

The evolution of network LLM capabilities can be aligned with the evolution of ANs, encouraging the addition of new features to move upwards in the autonomous level categorization. The following sections explore this relationship.

Challenges in Achieving Partial Autonomous Operation

In order to be applied to communication networks on a large scale, network LLMs face challenges in three aspects: security and reliability, knowledge refinement for specific network scenarios and incorporating additional technologies to augment the power of a LLM.

1. Secure and reliable. The communication network is a complex production network serving the national economy and people's livelihood, and any failure may affect tens of millions of users. Hence, it is necessary to ensure that AI systems can operate securely and reliably. Indeed, this is a fundamental principle of the AI Act [24]. In addition, the AI Act requires the decision-making process to be explainable and traceable [25].
2. Knowledge Refinement. The communication network is a very complex network, involving different domains (e.g., wireless, core, and transmission) that support diverse applications and serve hundreds of millions of users. The LLM can be used in scenarios such as fault location, service improvement, network monitoring and troubleshooting, intelligent professional Q&A, and personalized intelligent customer service. However, a LLM cannot *prove* that a problem was fixed, it can only provide a probability. Hence, cognitive technologies [19] must be used in conjunction with a LLM (see clause 4 and particularly clause 4.3 of the present document.)
3. Technology Augmentation. The network LLM needs to be augmented with other appropriate technologies as described in [15] [16] [17] to perform the following: intent recognition, parsing, and management to serve more constituencies; policy management for providing recommendations and commands in a standard format; knowledge graphs for proving and explaining decisions; natural language processing and understanding for conversing with and answering questions from the user.

Ready for Partial Autonomous Operation

The network LLM will reshape how AI is used in the development of autonomous network operations. Key points include:

1. Inject knowledge into management and orchestration systems to achieve more intelligent operations and improve the level of network security. This includes, for certain specific tasks and services, the ability to:
   A. Optimize data flow and resource allocation to provide the best “Quality of Experience” (QoE) and “Quality of Service” (QoS) for preferred sets of customers.

   NOTE: “Quality of Experience” (QoE) and “Quality of Service” (QoS) are terms used widely to measure experience and service useability by end customers with Mean opinion scores.

   B. Dynamically adjust the allocation of cloud-network resources to maximize cost-effective service operation.

C. Predict and identify network vulnerabilities and automatically repair or isolate affected components from the rest of the network.

D. Automate network and system management operations to reduce manual intervention and lower costs.

2. Integrate the network LLM into network operation planning, construction, operation, and maintenance, including:

   A. Real-time analysis and optimization of network behaviour, dynamic adjustment of resources, response to emergencies and changes, and improvement of system stability.

   B. Identify and defend against security threats to prevent attacks in advance.

   C. Optimize the energy allocation of 5G base stations and data centres to achieve energy savings and cost reduction.

The integration of AI technologies into a telecommunications network environment presents a number of different challenges. This has resulted in a more cautious phased integration approach, as follows:

- Phase 1 is mostly chatbots for question-answering and similar applications. For example, a network optimization chatbot could help with problem diagnosis to decision execution.
- Phase 2 adds role-oriented digital assistants. Exemplary applications include on-site installation and maintenance, customer service, and fault diagnosis and remediation. Integrating LLMs with knowledge graphs will usher in explainability and transparency.
- Phase 3 adds agents and multi-agent systems. One example is [26], which leverages the collective strengths of multiple LLMs through a layered Mixture-of-Agents (MoA) approach. It enhances response quality through iterative refinement. LLMs generate better responses when they have access to outputs from other models, even if those outputs are of lower quality. Similar to mixture-of-experts, this is a layered architecture where each layer is made up of multiple LLM agents. There are typically 3 agents per layer and 4 layers. Each agent takes all the outputs from agents in the previous layer as auxiliary information in generating its response. MoA achieves state-of-the-art performance on benchmarks like AlpacaEval 2.0, MT-Bench, and FLASK, surpassing GPT-4 Omni. This exemplary approach, after appropriate finetuning, could be used to further automate network OAM operations.

4. Management Technologies to Increase Autonomy of OAM Operations

4.1. Deep learning

4.1.1. Introduction

Deep learning is a subset of artificial intelligence that investigates how to use neural networks with many layers (i.e., "deep") to model complex patterns in data. It is particularly effective for tasks involving large amounts of unstructured data, such as images, audio, and text. Deep learning has become a crucial component in enhancing autonomous systems through the development of increasingly computationally efficient models. This section will provide several examples of its use in ICT technology.

4.1.2. Generic AI for OAM Use Cases

Deep learning has shown significant potential in enhancing the autonomy of network management systems by enabling automatic detection and classification of network anomalies, predicting network congestion, and optimizing resource allocation based on traffic patterns and user behaviour. The following are examples of deep learning for network management autonomous operation in different domains for different applications:

1. **AI Networking and AIOps[1]:** Autonomous networks leverage AI and machine learning to continuously monitor the networks, analysing data in real-time. This capability allows the network management systems to identify patterns that reveal anomalies, enabling proactive problem-solving and self-healing mechanisms. This helps ensure that network issues are detected and resolved before they impact users [11]. The needed data to feed the AI models can be gathered using telemetry procedures. ETSI ISG ENI is specifying a *cognitive network* (see clause 4.3 and [22]) that uses a sophisticated cognition model to understand normalised ingested data (e.g., what caused the data to occur and its significance to the operation of the system) and information (e.g., what is the urgency of fixing this problem, and how likely is the fix to cause other problems), as well as the context that defines how those data were produced[2]. Once that understanding is achieved, ENI then evaluates the meaning of the data, and determines if any actions need to be taken to ensure that the goals and objectives of the system are met. This includes improving or optimising performance, reliability, and/or availability. In short, it uses AIOps and cognitive networking to improve the operator experience.

[1] AIOps is the use of AI technologies to automate and optimise IT service management and operations.

[2] Many types of data, and especially information and knowledge, are *context-dependent*. This means that the significance and relevance of a problem may change in different contexts.

2. **Machine Learning for Network Management:** The complexity of modern networks, especially with the move towards cloud-native and disaggregated architectures, necessitates the use of AI and machine learning for effective management. These technologies enable the network to learn from historical data and predict future network states, such as congestion, allowing for proactive adjustments in routing to prevent bottlenecks [12]. Some additional examples include:

    - Automation of Configuration, Policy Management, and Network Monitoring, where simple and repetitive, but manually intensive, tasks, can be performed using ML.
    - Anomaly Detection, where ML identifies trends and outliers that indicate potential outages, failures, or bottlenecks.
    - Security Incident Alerts and Remediation, by recognising trends and anomalies that indicate security issues.
    - Predictive Maintenance, which uses ML to proactively replace hardware before it fails and address nascent issues while they are still relatively minor and can be resolved more easily and quickly.

3. **Deep Reinforcement Learning (DRL) for Resource Allocation in 5G Networks:** DRL algorithms learn from the interaction with the environment, making decisions based on the state of the system and receiving feedback in the form of rewards. This allows the model to learn the optimal policy over time. DRL can handle complex, uncertain, and dynamic environments by learning to adapt to different environmental changes and make decisions that maximize the long-term reward. One approach uses deep Q-learning to dynamically allocate network resources based on traffic demands and user behaviour. The model learns to optimize spectrum and computing resource allocation to maximize network performance and user quality of experience [13]. Some additional examples include:

    - DRL enables networks to self-optimise by continually fine-tuning parameters and paths to minimise some metrics such as congestion and latency while maximizing other metrics, such as throughput and reliability;
    - DRL can be used to model resource allocation as a dynamic programming problem for optimising one or more goals, such as energy efficiency and cost;
    - DRL can be used to improve the reliability of data transmission in 5G networks. For example, it can be used to determine the number of repeated transmissions of emergency data to reach the target outage probability; and
    - DRL can make real-time decisions based on the current state of the system.

4. **Transfer Learning for Cross-Domain Resource Optimization:** Transfer learning can be applied to adapt optimization models trained on one network domain to new domains with limited data. This improves the generalization capability for network optimization models [14]. Some examples include:

    - transfer learning applied to resource management on network cross-domains;
    - transfer learning applied to data traffic analysis on multimodal traffic;
    - transfer learning applied to computing task allocation on multimodal tasks.

4.2. Policy Management

ISG ENI uses a novel policy model [16] to manage the behaviour of the system. Management involves monitoring the activity of a system, making decisions about how the system is acting, and performing control actions to modify the behaviour of the system [15]. Policy Management ensures that consistent and scalable decisions are made governing the behaviour of a system. Policy controls the behaviour of an Entity, not the actual end result. For example, an access control list may be created and managed using policy, but is not a policy instance or type of policy.

The actions of a policy should always be verified. Past architectures have not done this (i.e., they usually have a policy decision entity and an entity to enforce the decision, but no entity to verify that the policy was executed correctly). This is an important feature of the ENI Policy Management system, as shown below. Also, a goal of ENI is to continually evaluate and optimize policy, so that it becomes more effective with experience.

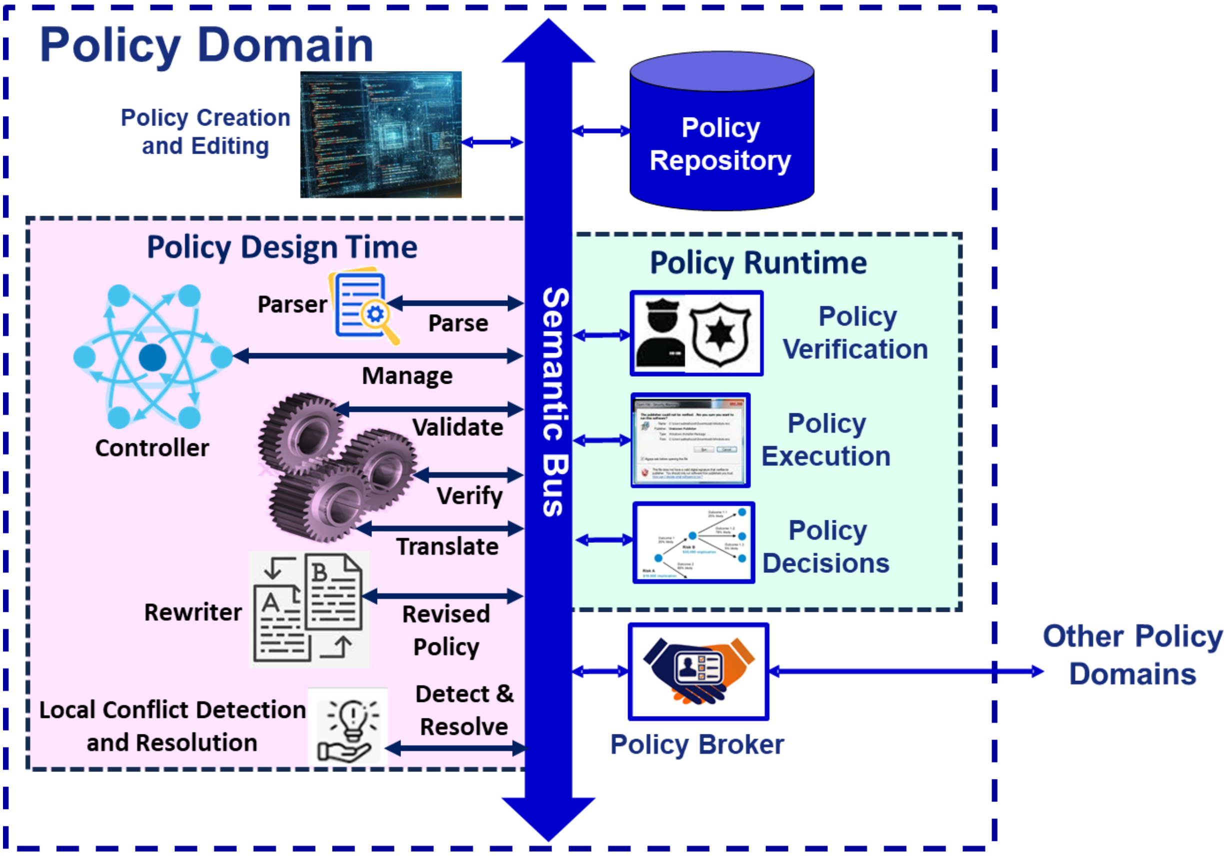

Figure 4.2.1: Simplified Functional Block Diagram of the ENI Policy Model

4.2.1. Defining Input and Output Policies

The Policy Continuum [15] differentiates between the needs of different constituencies in defining and expressing policy. Each constituency is made up of a set of users that have similar business needs, and more importantly, use similar concepts and terminology. The Policy Continuum formally differentiates between the needs of different constituencies in defining and expressing policy. The number of continua in the Policy Continuum is determined by the applications using it. This enables ENI to formally capture policies as expressed in each continuum and translate them into its own internal format. This also provides traceability and explanations (crucial for compliance with the EU AI Act) in policy management.

A set of five External Reference Points [15] are used to send policies to and from the ENI System. There is also an External Reference Point to ingest information and knowledge that applies to policies from a particular source (e.g., a LLM using RAG).

4.2.2. Types of Policies

ENI defines a novel information model that can represent different types of policies, each with its own content [16]. For the current release, three types of policies are defined:

1. **Imperative policy:** a type of policy that uses statements to explicitly change the state of a set of targeted objects. Hence, the order of statements that make up the policy is explicitly defined. It is typically made up of Event, Condition, and Action clauses. An example of an imperative policy, using informal English, is:
    - *WHEN an Alarm is received, IF the severity of the Alarm is Critical THEN execute the CriticalAlarm Policy.*
2. **Declarative policy:** a type of policy that uses statements from a formal logic to describe a set of computations that need to be done without defining how to execute those computations. Hence, the state is not explicitly manipulated, and the order of statements that make up the policy is irrelevant. An example of a declarative policy, using First Order Logic, is:
    - $\exists x \exists y\ (Customer(x) \land SLA(y) \land have(x, y))$

    The English equivalent is: “Some Customers have a SLA”.
3. **Intent policy:** a type of policy that uses statements from a restricted natural language (e.g. an external Domain Specific Language, or DSL) to express the goals of the policy, but does not specify how to accomplish those goals. In particular, formal logic syntax is not used. Therefore, each statement in an Intent Policy may require the translation of one or more of its terms to a form that another managed functional entity can understand. An example of an intent policy is:
    - *No processor shall run at more than 75% utilization.*

    The above example indicates different types of ambiguity that may exist in an intent statement. For example, does the term "processor" include both CPUs and GPUs? What about ASICs that have processing capabilities? As another example, the term "utilization" could refer to memory, I/O operations, or processor utilization.

4.2.3. The ENI Policy Model

Figure 4.2.3.1 shows a simplified functional block diagram of the ENI Policy Model. At the top, a PolicyObject may aggregate 0..n metadata objects. This is inherited by all subclasses of the PolicyObject.

There are 4 subclasses. PolicySource defines the author and other contact info for a policy, and PolicyTarget defines the set of managed objects that this policy may affect. PolicyStructure and PolicyComponentStructure define the types of policies and components of a policy, respectively. Conceptually, the "left side" represents the type of policy, and the "right side" represents the contents of the policy. When a given policy is defined on the left side, the set of components that can be used to populate its content are then defined on the right side. Once a particular subclass of PolicyStructure is chosen, this restricts the types of policy components that can be used to define its content.

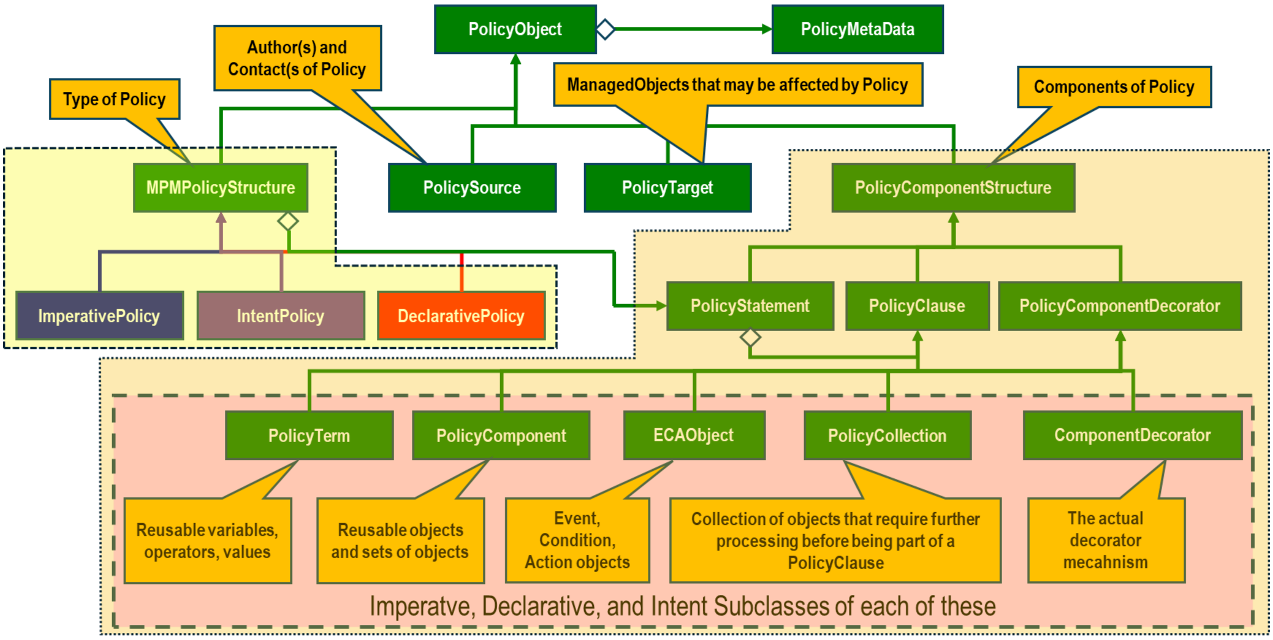

Figure 4.2.3.1: Simplified Functional Block Diagram of the ENI Policy Information Model

4.2.4. ENI Policy Execution

There are multiple ways to execute ENI policies [15]. An exemplary functional block diagram is shown in Figure 4.2.3.1 using a model-driven engineering approach, where code is generated from UML information and data models. An exemplary recipe for executing policies is:

1. Annotate the information model with tag-value pairs to add information for tools to manipulate the model.
2. Generate one or more DSLs that are used to describe the object model(s) created from the information model. This enables these models to be shared across different tools and programming languages to enable a consistent and coherent view of the information to be accessible. Each DSL corresponds to a particular Policy Continuum level.
3. Generate a final DSL to model the behaviour of the system being managed. This ensures that behaviours defined by policies are supported in and consistent with the information model.
4. Analyse the produced policies. This includes syntactic and semantic checking and conflict resolution. Reasoning is used to help resolve any conflicts found. This can also be used to transform a policy of one type (e.g. intent) into a policy of another type.
5. Generate monitoring criteria, then Implement and deploy the policy.

4.3. Cognitive Management

Cognition is the process of acquiring and understanding data and information to produce new data, new information, and new knowledge. Cognitive networks use situation-awareness[3] to evaluate the performance of a system against its goals. Cognitive systems understand the underlying meaning of why telemetry-based information occurred as a function of context. Cognitive systems reason by using formal logic or other mechanisms to form and prove hypotheses as to why a problem occurred and how to fix it. Cognitive systems learn experientially and become more efficient and accurate over time as their knowledge increases and is refined. An example of Knowledge produced is the understanding of how telemetry affects KPIs, and the trends that lead to faults and violations of Service Levels (SL) and Service Level agreements (SLAs).

A Cognitive Network is aware of its goals, and can actively protect them from being violated even in the presence of change. Similarly, if its goals change, then it takes measures to change the services it provides to meet those changed goals. This is one of the primary use cases for Cognitive Networks.

The ENI Cognitive Management Functional Block uses cognitive processes to understand how past behaviour, coupled with currently ingested contextual data and information, affect the goals that the ENI System is trying to achieve. The ENI Cognitive Management system draws from human decision-making processes to better comprehend the relevance and meaning of ingested data. Cognitive management enables the ENI System to experientially learn to improve its operation and performance, thereby providing autonomic behaviour.

4.3.1. Cognition Model

The ENI Cognitive Management Functional Block is based on an innovative cognition model [15] [19]. A cognition model defines how cognitive processes, such as comprehension, action, and prediction, are performed and influence decisions. The ENI cognition model draws heavily on how human cognition is performed. More specifically, cognitive psychology defines three interacting layers, called reactive (or subconscious), deliberative, and reflective. Reactive processes take immediate responses based upon the reception of an appropriate external stimulus. Deliberative processes receive data from and can send recommendations and/or commands to the reactive processes.

In ENI, these processes accumulate and generalise knowledge from experience, and combine that with what is learned from other people and systems. They can achieve more complex goals by applying short- and long-term memory in order to create and carry out more elaborate plans. Reflective processes consider what predictions turned out wrong, along with what obstacles and constraints were encountered, in order to prevent sub-optimal performance from occurring again. This may require the reformulation of the problem in a way that leads to more effective solutions.

ENI Cognitive Management learns from experience to improve its performance. This includes acquiring new knowledge from instruction or experience, revising and correcting existing knowledge, and combining existing data and information to infer and deduce new knowledge.

[3] Situation awareness is perceiving data and behavior that pertain to the relevant circumstances and/or conditions of a system or process, understanding the meaning and significance of these data and behaviors, and how processes, actions, and new situations inferred from these data and processes are likely to evolve in the near future.

All recommendations and commands from ENI are embedded in Policies. The key advantage of this combined approach is the transformer's ability to process and generate natural language, while leveraging the knowledge graph's capability to integrate and reason over disparate data sources and domain knowledge within the telco ecosystem. This could drive more intelligent, automated and context-aware decision-making across various telco operations and services.

Figure 4.3.1.1 shows a simplified Functional block diagram of the ENI Cognition closed control loops.

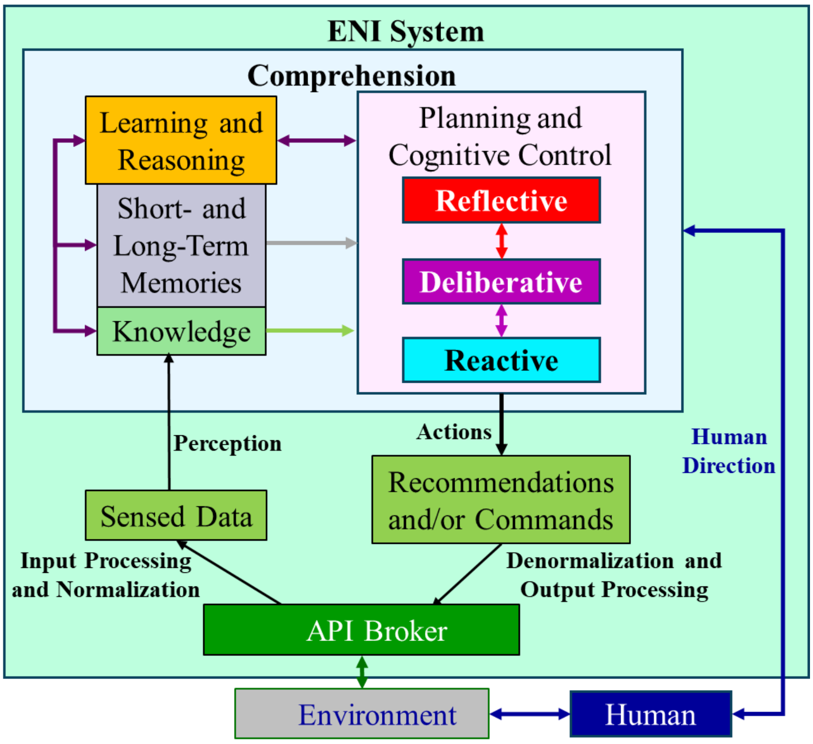

Figure 4.3.1.1: Simplified Functional Block Diagram of the ENI Cognition Model

4.3.2. Cognitive Planning and Execution

Managing networks is complex, since business rules, environmental conditions, and user needs are all constantly changing. Therefore, ENI uses a set of agents, each optimised to perform a particular function, track changes, assess their meaning and relevance, and plan any changes desired. This is pictured below.

Figure 4.3.2.1 shows some of our extensions to the Observe-Orient-Decide-Act (OODA) control loop [19]. We control the loop using a Finite State Machine (FSM); we insert an AI-based planning module between the orient and decide functions; we add a reinforcement learning module that observes each stage of the FSM and produces recommendations to optimize one or more of the control loops. The system can take “shortcuts” if certain criteria are met (also not part of the original OODA loop), shown as “Normal” (no shortcuts), High Priority (bypasses planning), and “Urgent” (bypasses planning and decision-making).

We use a hybrid approach to implement the above cognition model by using a DRL agent framework for implementing a state machine. This combination makes the state machine more robust to change. DRL agents can be data-driven, learning to adapt their behaviour online by continuously updating their policy based on new observations and rewards from the environment. DRL can optimize sophisticated reward functions that encode complex objectives like safety constraints, fairness criteria, or multi-objective trade-offs, enabling finer-grained behaviour management.

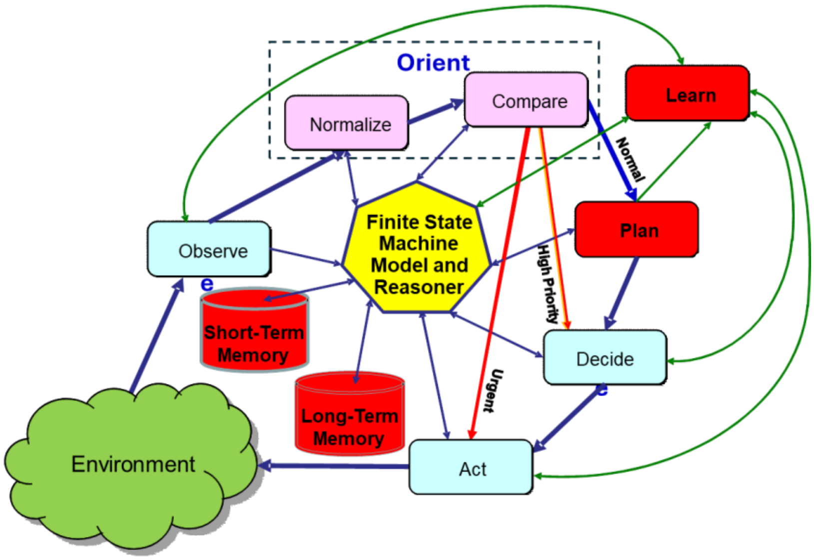

Figure 4.3.2.1: ENI Cognition Model as an Extension of the OODA Control Loop [18]

4.4. How Generative AI Applications Enhance Autonomicity

Generative models are being increasingly used in network management applications. Indeed, Generative AI models can be used to autonomously detect and resolve network issues, optimizing performance and minimizing downtime. These models also enhance network security by identifying and mitigating threats in real-time. Additionally, AI-driven automation streamlines configuration management, reducing human error and ensuring consistent policies across the network. These are collectively examples of intelligent management, which leads to more efficient resource utilization, improved user experiences, and significant cost savings.

This section provides examples of how transformers and knowledge graphs enhance the autonomy of network operations.

4.4.1. Overview of Generative AI

With the advent of 5G and the upcoming 5.5G, one of the goals of 3GPP is integrating cellular and non-cellular communication technologies to provide significant improvement of the network connectivity, accessibility, and data rates to support future services such as tactile internet, augmented reality, metaverse, cloud gaming, telepresence, autonomous remote driving, and navigation.

This integration of heterogeneous communication systems is leading to an increase in the complexity of networks, which makes their control and management ever more complex.

In this context, Generative AI can support the design and management of complex systems according to gathered observations. For example, the output of generative models can provide:

- enhanced predictability based on historical information and context, which can be used to estimate when and where problems might occur in a network infrastructure;
- diagnosis of network problems in real-time, reducing downtime and improving reliability;
- optimised allocation of network resources, ensuring efficient utilisation;
- data analytics and trends to make reliable decisions about infrastructure management by exploiting data-driven strategies;
- personalised services to customers, enhancing their experience and reducing customer churn.

Some applications of a Generative AI model to networking can be found in the literature. In [1], authors generate a semantic model for the received information, starting from the original complex content, to make the transmission to the channel corruptions more robust. In [2], the authors define a method for the virtual representation of physical objects of a 5G and beyond network. In [3], the authors describe a method for extracting a model of a city's entire mobility network, a weighted directed graph in which nodes are geographic locations and weighted edges represent people's movements between those locations, thus describing the entire mobility set flows within a city.

In [4], authors investigate a generative pre-trained model NetGPT for both traffic understanding and generation tasks. The authors use multi-pattern network traffic modelling to construct unified text inputs and support traffic understanding and generation tasks. Finally, in [5], the authors define a model to generate a synthetic CPS topology with realistic network feature distribution. This model can learn different complex network parameters and capture the distribution of different network features of the input networks.

4.4.2. Transformers and Knowledge Graphs as Used in ENI

In [17], there is the description of the use of Transformers and Knowledge Graphs in ENI.

Transformers are a specific neural network architecture [20] that is designed to process sequential data, such as text, by using self-attention mechanisms to capture relationships between different parts of the input sequence. Transformers excel at understanding context and long-range dependencies in text, making them highly effective for various language tasks, including machine translation, summarization, and question-answering. LLMs are typically based on the transformer architecture and are characterized by their massive scale, often containing billions of parameters. LLMs can perform various natural language processing tasks, including text generation, conversational AI, and content creation.

The Transformer Management Functional Block is located within the Policy Management Functional Block for two reasons: (1) the most common function of the Transformer Management Functional Block is to be used to create and edit ENI Policies, and (2) External Policy Users (i.e. the End-User, an Application, the OSS, the BSS, and the Orchestrator) do not need direct access to the functionality provided by the Transformer Management Functional Block. This also reduces the attack surface of ENI.

The Transformer Management Functional Block may require significant computing, memory, and/or other resources for its operation. This is one reason why it is designed as a Functional Block. Hence, it may use a set of Internal Reference Points. A simplified Functional Block Diagram is shown in Figure 4.4.2.1. It contains four Functional Blocks: Retrieval Augmented Generation (RAG), Prompting Framework, Transformer Processing, and Output Generation.

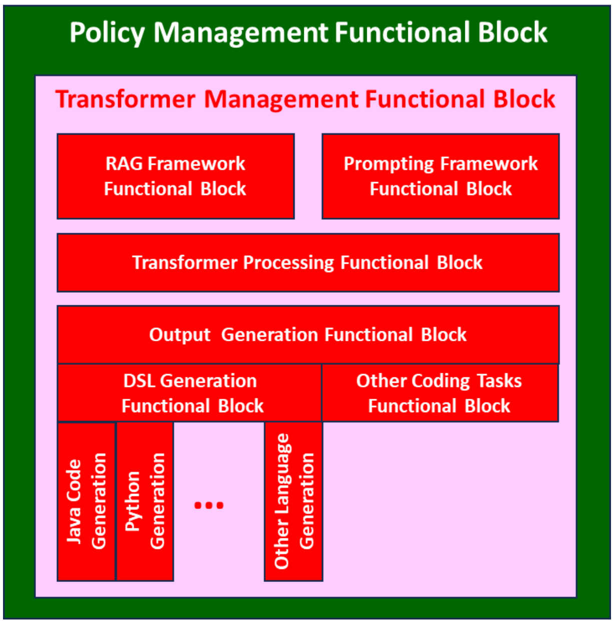

Figure 4.4.2.1: Simplified Functional Block Diagram of the Transformer Management Functional Block

RAG enhances the output of transformers by combining the strengths of retrieval-based and generation-based approaches, leading to more accurate, relevant, and contextually informed responses. It enables ENI to use open-source models and finetune them with telco-specific documentation and business rules.

The prompting framework enables different prompting techniques to be used. It can significantly enhance transformer performance in generative AI by improving logical reasoning, handling complex problems, and mimicking non-linear human thought processes.

The Transformer Processing Functional Block enables different Transformers to be used. Its primary focus is to provide additional information for parsing input policies to the parser components in the Policy Management Functional Block.

The Output Generation Functional Block is used to generate code corresponding to the processed policy. It is designed as a modular set of hierarchical Functional Blocks. Two examples are shown, one for processing DSLs, and the other for generating code, such as Java® and Python™.

A Knowledge Graph[4] uses a formal logic model to represent the semantics of nodes, edges, and their relationships as a rich set of structures. This formal representation allows logical inference for retrieving implicit knowledge rather than only allowing queries requesting explicit knowledge. [21] discusses how Knowledge Graphs may be used to significantly improve complex database analysis and reasoning by LLMs.

This includes (1) providing enhanced context and background knowledge, (2) enhancing question answering by forcing the LLM to reason in steps, (3) enhancing recommendations by training the LLM to follow step-by-step reasoning through the graph, (4) improving data integration and retrieval by ensuring the retrieved information (e.g., from a RAG system) is more relevant, and (5) helping knowledge transfer and generation across different domains and tasks, enabling better performance on unseen or out-of-distribution data.

The inclusion of a Knowledge Graph also provides two important benefits for LLMs:

1. Mitigating Hallucinations by grounding LLMs with factual knowledge from Knowledge Graphs, hallucinations may be reduced, improving the reliability of the system's outputs.

2. Enabling Explainable and Traceable Reasoning by providing a structured and interpretable representation of knowledge, allowing LLMs to generate more explainable and traceable reasoning paths for their outputs. The use of formal logic in Knowledge Graphs enables hypotheses and theorems to be proven. This is critical for compliance with the EU's AI Act.

Likewise, a Transformer can provide several benefits for Knowledge Graphs, including: (1) automated graph construction and completion, (2) linking entities in text to entities in a Knowledge Graph to form a semantic network, (3) process dynamically changing graph data to enable reasoning over evolving knowledge, and provide richer context definition, contextual knowledge, and dependencies.

Possible applications of this combination include:

- **Knowledge Assistant.** The transformer model processes network data streams (logs, events, metrics) to extract insights and generate natural language reports/alerts. The Knowledge Graph integrates static network topology data, device configurations, service models etc, to provide context for reasoning. The combination enables advanced network monitoring, root cause analysis, predictive maintenance, and automated remediation workflows.

- **Service/Resource Discovery.** The Knowledge Graph models complex interdependencies & hierarchies of network services & apps. The Transformer processes user requests/intents and queries the Knowledge Graph to discover and recommend relevant services or resources based on user entitlements, device capabilities etc.

- **Intent-based Service Automation.** The Knowledge Graph models intricate network policies, constraints and best practices. Transformer queries the Knowledge Graph to generate low-level device configurations or orchestration workflows while adhering to policies.

[4] A knowledge graph is not a type of Generative AI; rather, it is a type of symbolic logic. However, when coupled with a transformer, a neuro-symbolic Ai is realized, which does include generative AI functionality.

4.5. Network Digital Twins

The advent of the 5G Service-Based Architecture, and in general the softwarization of the network, has made network status and configuration vital for run-time adaptation and optimization of network performance and resource utilization. For this reason, future networks will need to embed such dynamic data in the Network Digital Twins, granting real-time alignment of the real network status.

The concept of Digital Twins is well-known, especially in manufacturing and IoT, as a tool to enable predictive maintenance and accurate control of objects and complex systems. The concept is to build a digital replica of a real system, that is supported by a continuous flow of data from the physical twin (the original system) in such a way as to maintain the two entities "synchronized".

The same concept can be applied to networks and networked services, leading to Network Digital Twins: an accurate replica of a real network, complete with modelled equipment and traffic. The Network Digital Twin represents an important asset for future networks, and in particular future mobile networks where resources are scarcer and coverage status of every single user can change overtime.

Network Digital Twins are a key component of future mobile networks beyond 5G, as it can be used to orchestrate and manage the emulation environment following NFV/SDN principles [23]. Data generated by the Digital Twin, based on network flows and device behaviours, are made accessible to other components and functions of the system, so they can perform intelligent analysis and predictions.

In particular, the Network Digital Twin directly supports three main AI-driven autonomous functionalities:

- **Generate datasets for training AI/ML algorithms:** representing a faithful replica of the actual network infrastructure, the Network Digital Twin can generate diverse datasets for training AI/ML algorithms without affecting or impacting the actual physical component.
- **Perform prediction and prevention:** the Network Digital Twin can predict different future scenarios and forecast different problems and vulnerabilities.
- **Analyze "what-if" scenarios:** Network Digital Twins can provide different scenarios to gain in-depth knowledge of the network behaviour and analyze the different management strategies without requiring direct actions on the physical counterpart. The Network Digital Twin can provide a playground for AI to perform tests and learn the potential impact of its actions without generating potential performance issues on the real system.

However, several challenges are still to be addressed. Arguably the most relevant is the two-way continuous data flow between the physical and the digital twin. However, the amount of information to perform an accurate replica of the state of the physical twin and the related data flows and services might be massive and difficult to handle. This leads to the requirement of applying AI/ML to reduce the amount of actual data to transfer by modelling or predicting some aspects of the data flows and network state information.

In addition, the time scale in networks can be as short as milli-, micro- or even nano-seconds, requiring the introduction of prediction and modelling to minimize the distance between the time on the physical twin and that in its digital version. The amount of time required to transfer/synchronize state information will impact the overall performance and, in some cases, even the feasibility of some of the scenarios above.

5. Network Technologies for AI Service Delivery (Net4AI)

The fast evolution towards investigating how AI can be used in telecom networks raises the expectations for service agility in both business and consumer markets. While technological development is growing fast, the critical question is whether the service provider's existing underlying data communications networks can fulfil the expectations, delivering high-demanding services intelligently and with low latency and ensuring a deterministic experience-centric network architecture.

To accomplish the above expectations, the network has to provide deterministic QoS and QoE and autonomous operation of both infrastructure and services.

As reported in Figure 5.1, the following concept could help in focalizing the relationship between AI and Network.

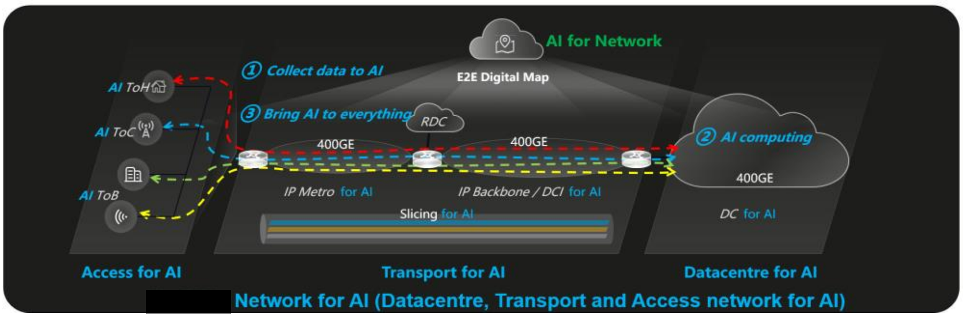

Figure 5.1: AI for Network and Network for AI

- "Network for AI" enable the Network to properly deliver the services based on AI to the customer, enabling dynamic service creation and granting SLA proactively. Conversely, Network has to adopt technologies enabling simple and effective monitoring and control from AI technologies.
- "AI for network" increase the Autonomous lifecycle of infrastructure and service.

Clause 5 is concentrating on defining the aspects related to "Network for AI". Convergence to a fewer network protocols and simpler routing infrastructure will simplify its representation in digital twins and help enable AI-based automation. The Network will be capable of supporting multiple scenarios by 2030, including:

- **Transport Network:** The base station access is upgraded from 10GE to 50GE, driving the transmission speed of the Metro Area Network (MAN) aggregation network and backbone network to 400/800GE and increased usage of Segment Routing IPv6 (SRv6), which is the latest evolution of source routing technology. With the development of the industry ecosystem and related standards, SRv6 and its compression technologies are increasingly deployed on global IP networks, helping telecom carriers, industry customers, and enterprise customers deploy more cost-effective and intelligent networks and provide convenient and high-quality service experiences based on network slicing functionality.

- **Data centre network:** The server network scale increases from 100,000 nodes to millions of nodes. The maximum interface rate of data centre switches is upgraded to 400/800GE, reducing the latency from microseconds to nanoseconds. Additionally, AI training scenarios mainly involve a small number of high-density flows, which are small in number and high in flow density. Traditional Equal Cost Multi-Path (ECMP) is based on hash mechanisms not considering the throughput of each flow. This causes traffic conflict and backpressure, reducing traffic throughput and affecting training efficiency. Traffic paths between AI training cards can be planned to avoid hash conflicts between traffic. Uplink conflicts on leaf switches and downlink conflicts on spine switches must be considered during path planning. Load balancing functionalities considering the status of the whole data centre network are recommended to realize efficient network scale load balancing.

- **Campus network:** Wi-Fi is upgraded from Wi-Fi 6 to Wi-Fi 7, giving users a peak access capability of up to 30 Gbit/s. With the development of WLAN technologies, homes, and enterprises rely increasingly on Wi-Fi to access their Network. The Wi-Fi 7 improves the data transmission rate and ensures low latency and high reliability. Therefore, Wi-Fi 7 better matches the requirements for robustness and delay performance for data transmission in scenarios such as voice conference, real-time operation, industrial Internet of Things (IIoT), interactive telemedicine, and similar (for example, Industrial Automated Guided Vehicles (AGVs) require 100 ms @99.999%) services.

- **Intelligent OAM:** The network is upgraded from L3 (conditionally autonomous) to L4 (highly autonomous) or later to L4.5 to almost achieve fully intelligent AN L5. Intelligent OAM will benefit from the listed technologies. They simplify the creation of a Digital Map enabling network modeling, dynamic data collection, and actuation of the needed network changes. This ensures a high-level service customer experience based on uniform management for all the network domains.

- **NaaS interface:** Customers expect one-step solutions to network and cloud service management, selection, and subscription.

- **Deterministic SLA:** Customers require deterministic service-level SLA assurance, obtained with a flexible allocation of cloud and network resources and intelligent traffic control based on different production scenarios.

- **Customer-level assurance:** Customers pay more attention to service-related network quality and there is the need to prevent network faults in advance and provide the possibility to perform service monitoring and maintenance by themselves.

6. Conclusions

As one of the key technologies that can help network operation move towards L4, mainstream operators are considering adopting the latest AI and ML technologies (e.g., Generative AI LLM and transformers) to help improve network operation management efficiency. Automation involves several kinds of use cases, as highlighted in Clause 3. They involve the Network infrastructure lifecycle management (creation, expansion, troubleshooting), service delivery, and end-user interactions.

The first class of use cases involves the management of a specific network infrastructure domain, like the one involving the IP network and digital twin simulation as described in this document, or the entire network status to calculate the KPI indicator to be aware of the network status.

For the second class, services can be requested and delivered by interpreting the user intent or by an API interface, connecting the end user directly to the data centre/cloud where the applications are implemented. SLA breach avoidance and root cause analyses are examples of service management automation.

In Clause 4, several technologies have been analyzed to address those use cases, and the combination of them is suggested to fulfill each use case at best. Significantly, no single AI technology satisfies all requirements of each use case. Hence, an analysis and recommendations of which technologies mix to be used to fulfill each scenario are described. In particular, the functions of policy management to standardize recommendations and commands generated for the system being managed and the Cognition Model to understand the meaning and implications of ingested data, and define which actions are needed to fulfill the policies are described:

- ETSI ISG ENI uses a novel policy model [16] to manage the behaviour of the system. Management involves monitoring the activity of a system, making decisions about how the system is acting, and performing control actions to modify the behaviour of the system [15]. Policy management ensures that consistent and scalable decisions are made to govern the behaviour of a system. Policy controls the behaviour of an Entity, not the actual end result. For example, an access control list may be created and managed using policy but is not a policy instance or type of policy.
- A cognition model defines how cognitive processes, such as comprehension, action, and prediction, are performed and influence decisions. The ENI cognition model draws heavily on how human cognition is performed.

Generative models are increasingly used in network management applications and, in particular, in the above Policy and Cognition Model functionalities. Indeed, Generative AI models can be used to autonomously detect and resolve network issues, optimizing performance and minimizing downtime. By analyzing vast amounts of network data, generative AI can predict potential failures and recommend proactive maintenance. It also enhances network security by identifying and mitigating threats in real-time. Additionally, AI-driven automation streamlines configuration management, reducing human error and ensuring consistent policies across the network. This intelligent management leads to more efficient resource utilization, improved user experiences, and significant cost savings, making generative AI an invaluable tool for modern network operations.

The network infrastructure itself has to evolve to interact efficiently with highlighted AI technologies, providing e2e real-time interaction. The network infrastructure has to be easily represented in a digital map, with a simple way to dynamically collect the status of the Network and actuate the necessary adaptations. The Network has to deliver state-of-the-art services (like AR, unmanned industry, and 8K video) to every user and device, providing ubiquitous 10Gbps access on campus, at home, on mobile, and in industry. This implies using a 400/800GE connection in the Network and the data centre. Digital Map utilization enables to properly optimize the connection in the data centre implementing network scale load balancing mechanisms. E2e SRv6-based network slicing from Campus to data centre and public cloud represents one way to diminish the overall number of protocols and stitching points in the network, enabling simple modelling, monitoring, provisioning and optimization.

Technologies are available to increase the overall AN level of the telecommunication networks, and ENI wants to highlight the possible way ahead continuing to drive technology evolution and suggesting best practices to implement them in live environments. Fervent activities in Work Items and PoCs will need a joint effort by the whole industry to synchronize the effort in a common direction and speed the increase of the Autonomous Network level.

7. References

[1] Eleonora Grassucci and Yuki Mitsufuji and Ping Zhang and Danilo Comminiello; Enhancing Semantic Communication with Deep Generative Models; arXiv 2023

[2] Mozo, A.; Karamchandani, A.; Gómez-Canaval, S.; Sanz, M.; Moreno, J.I.; Pastor, A. B5GEMINI: AI-Driven Network Digital Twin. *Sensors* 2022, 22, 4106

[3] Mauro, Giovanni and Luca, Massimiliano and Longa, Antonio and Lepri, Bruno and Pappalardo, Luca; Generating mobility networks with generative adversarial networks; Springer Science and Business Media LLC, EPJ Data Science vol. 11 issue 1 2022

[4] Xuying Meng and Chungang Lin and Yequan Wang and Yujun Zhang; NetGPT: Generative Pretrained Transformer for Network Traffic; arXiv 2023

[5] Y. Liu, H. Xie, A. Presekal, A. Stefanov and P. Palensky, "A GNN-Based Generative Model for Generating Synthetic Cyber-Physical Power System Topology" in *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4968-4971, Nov. 2023

[6] European Commission Representation in Cyprus, “Commission presents new initiatives for digital infrastructures of tomorrow" in EU Europa News, read June 18 2024, published February 2024 https://cyprus.representation.ec.europa.eu/news/commission-presents-new-initiatives-digital-infrastructures-tomorrow-2024-02-21_en

[7] DENG Letian, ZHAO Yanru, "Deep Learning-Based Semantic Feature Extraction: A Literature Review and Future Directions" in Northwest Agriculture and Forestry University, Xianyang 712100, China, June 2023 https://www.zte.com.cn/content/dam/zte-site/res-www-zte-com-cn/mediares/magazine/publication/com_en/article/202302/202302003.pdf

[8] Han, Z., Zheng, X., Ren, Y., Li, X., Wang, Q. "Personalized Adaptive Cruise Control with Deep Reinforcement Learning", In: Francisco Rebelo and Zihao Wang (eds) Ergonomics In Design. AHFE (2023) International Conference. AHFE Open Access, vol 77. AHFE International, USA. https://doi.org/10.54941/ahfe1003421

[9] Ali Irshayyid,Jun Chen and Enrico Meli, "Comparative Study of Cooperative Platoon Merging Control Based on Reinforcement Learning" in U.S. National Institutes of Health's National Library of Medicine, January 2023 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9865798/

[10] Mrinal R. Bachute (Ph.D [Electronics], Industry Liaison Officer and Associate Professor), with Javed M. Subhedar (M.S.[Automotive Electronics]), "Autonomous Driving Architectures: Insights of Machine Learning and Deep Learning Algorithms" in ScienceDirect: Machine Learning with applications Dec. 2021 https://www.sciencedirect.com/science/article/pii/S2666827021000827

[11] Anon "What Is an Autonomous Network? Definition & How It works" in Nile Publications Dec. 2021 https://nilesecure.com/ai-networking/what-is-an-autonomous-network-definition-how-it-works

[12] Lingli Deng (China Mobile), Yuhan Zhang (China Mobile), Amit Dass (Cisco), Tony Verspecht (Cisco), Sebastian Zechlin (Deutsche Telekom), Andreas Volk (HPE), Jean Paul Pallois (Huawei), Luigi Licciardi (Huawei), Gary Li Intel, Jermin Girgis TELUS, Sebastian Thalanany UScellular, Manchang Ju ZTE; "Automation and Autonomous system Architecture Framework – Phase 2" NGNM publications October 2024 https://www.ngmn.org/publications/automation-and-autonomous-system-architecture-framework.html

[13] Priyadarshan Patil, (The University of Texas at Austin) "Applications of Deep Learning in Traffic Management: A Review" In tensorgate publications https://research.tensorgate.org/index.php/IJBIBDA/article/view/26

[14] Xing Wang, Zhendong Wang, Kexin Yang, Zhiyan Song, Chong Bian, Junlan Feng, and Chao Deng (China momile reseach intitute) "A Survey on Deep Learning for Cellular Traffic Prediction" https://spj.science.org/doi/10.34133/icomputing.0054

[15] John Strassner ETSI rapporteur "ENI System Architecture" ETSI open area version 4.0.x draft https://docbox.etsi.org/ISG/ENI/Open/ENI005

[16] John Strassner ETSI rapporteur "ENI Models inference and APIs" ETSI open area version 4.0.x draft https://docbox.etsi.org/ISG/ENI/Open/ENI019

[17] John Strassner ETSI rapporteur "ENI Transformer Architure" ETSI open area version 4.0.x draft https://docbox.etsi.org/ISG/ENI/Open/ENI030%20(Release%204)

[18] Alastair Luft US Navy "The OODA Loop and the Half-Beat" The Strategy Bridge March 2020 https://thestrategybridge.org/the-bridge/2020/3/17/the-ooda-loop-and-the-half-beat

[19] John Strassner (Futurewei) Raymond Forbes (Huawei) Haining Wang (Intel), Ziting Zhang, (China Telecommunications) and Antonio Gamelas (Portugal Telecom) “ENI Vision: Understanding the Operator Experience Using Cognitive Management” ETSI Whitepaper January 2023 https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_WP51_Understanding_the_Operator_Experience_Using_Cognitive_Manage.pdf

[20] A. Vaswani, et al., “Attention is All You Need”, 2017

[21] J. Sequeda, et al., “A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model’s Accuracy for Question Answering on Enterprise SQL Databases”, November 2023

[22] J. Strassner, “A Cognitive Assistant using a Semantic Knowledge Graph-Enabled Transformer for Reasoning and Decision-Making”, August 2024

[23] E. Rodriguez *et al*., "A Security Services Management Architecture Toward Resilient 6G Wireless and Computing Ecosystems," in *IEEE Access*, vol. 12, pp. 98046-98058, 2024, doi:10.1109/ACCESS.2024.3427661

[24] EU Parliament “The European Union: AI Act” Official Journal proceedings of the EU, July 2024 https://artificialintelligenceact.eu/the-act/

[25] F. Walke, L. Bennek, and J.T. Winkler, "Artificial Intelligence Explainability Requirements of the AI Act and Metrics for Measuring Compliance" (2023). Wirtschaftsinformatik 2023 Proceedings. 77

[26] J. Wang, et al., “Mixture-of-Agents Enhances Large Language Model Capabilities”, June 2024

[27] Luca Pesando ETSI rapporteur “ENI Categorization” ETSI catalogue version 1.1.1 https://docbox.etsi.org/ISG/ENI/Open/ENI007

[28] Yu Zeng ETSI rapporteur “ENI Evaluation of Categorization” ETSI catalogue version 1.2.1 https://docbox.etsi.org/ISG/ENI/Open/ENI010

Annex A. Autonomous Driving

Deep learning provides significant improvements in various aspects of vehicle perception, decision-making, and control. Key functionalities provided include:

1. **Semantic Feature Extraction**: Deep learning models can extract semantic features from various inputs, improving the vehicle's ability to understand its environment. This is particularly useful in tasks such as object detection, scene understanding, and natural language processing for interpreting road signs and navigation instructions [7].
2. **Personalized Adaptive Cruise Control**: Deep reinforcement learning algorithms, such as Dueling Double Deep Q-Network, can be used to develop personalized adaptive cruise control systems. These systems can learn and adapt to individual driving styles, categorizing them as aggressive, general, or conservative, and adjust the vehicle's behaviour accordingly [8].
3. **Cooperative Platoon Merging**: Deep reinforcement learning approaches can optimize the merging behavior of connected and automated vehicles in platoons. This can lead to significant reductions in energy consumption (up to 76.7%) and improvements in passenger comfort [9].
4. [10] provides an overview of the **architecture and algorithms used for common autonomous driving tasks**, including motion planning, platooning, pedestrian detection lane recognition, and others.

The Standards People

ETSI
06921 Sophia Antipolis CEDEX, France
Tel +33 4 92 94 42 00
info@etsi.org
www.etsi.org