Agentic-NWDAF: Enabling Intent-driven Agentic Intelligence for Autonomous 6G Network Analytics

Mazene Ameur*, *Member, IEEE*, Bouziane Brik‡, *Senior Member, IEEE,* and Adlen Ksentini*, *Senior Member, IEEE*
* Communication Systems Department, EURECOM, Sophia Antipolis, France.
‡ Computer Science Department, University of Sharjah, Sharjah, UAE.
Emails: firstname.lastname@eurecom.fr, bbrik@sharjah.ac.ae

***Abstract*—The transition from 5G to 6G envisions Operations Support Systems (OSS) that embody Zero-touch Service Management (ZSM) and Intent-Based Networking (IBN) as core enablers of autonomous network operation. Ongoing standardization efforts by the 3rd Generation Partnership Project (3GPP) and the European Telecommunications Standards Institute (ETSI) have defined how OSS interfaces with Network Data Analytics Functions (NWDAF) through structured, service-based APIs that support performance assurance, fault detection, and closed-loop automation. However, current approaches remain largely procedural, limited to static analytics subscriptions and rule-based triggers and thus fall short of enabling intent-aware, adaptive, and democratized OSS–NWDAF interactions. To overcome these limitations, this paper introduces Agentic-NWDAF, an intent-aware orchestration layer that translates natural language OSS intents into executable NWDAF service calls. Agentic-NWDAF leverages Agentic Artificial Intelligence (AI) and the Model Context Protocol (MCP) to enable natural language understanding, adaptive analytics orchestration, and autonomous decision-making through collaborative AI agents. We validated Agentic-NWDAF through an extensive benchmark of 100 NWDAF intents derived from 3GPP specifications, demonstrating high accuracy, strong generalization, and reliable performance under LLM-as-a-Judge evaluation.**

***Index Terms*—Intent-Based Networking, Zero-touch Service Management, Agentic AI, LLM, NWDAF.**

I. Introduction

The evolution from fifth-generation (5G) to sixth-generation (6G) networks will unlock transformative applications, including immersive metaverse experiences, fully autonomous driving, and ubiquitous Internet-of-Everything (IoE) connectivity [1]. These applications require telecommunication infrastructures with unprecedented scalability, intelligence, and automation, demanding a fundamental redesign of network architecture and management paradigms [2]. To address these needs, Operations Support Systems (OSS) must evolve from rule-based management to intelligent, autonomous orchestration frameworks guided by two key paradigms: Zero-touch Service Management (ZSM) and Intent-Based Networking (IBN) [3]. ZSM, defined by the European Telecommunications Standards Institute (ETSI), promotes end-to-end automation across service and network domains to minimize human intervention, while maintaining assurance and efficiency [4]. Complementarily, IBN, as specified by the TM Forum, enables high-level intent expression that allows networks to translate operator goals into automated actions [6]. In this context, the 3rd Generation Partnership Project (3GPP) has progressively embedded these concepts through management data analytics and closed-loop automation in its specifications [7]. The convergence of ZSM and IBN within next-generation OSS architectures underpins autonomous, self-optimizing 6G networks that align operational behavior with business intent.

Meanwhile, recent and rapid progress in Large Language Models (LLMs) has given rise to a new computational paradigm known as Agentic AI. This paradigm encompasses systems composed of multiple autonomous agents, each capable of reasoning through complex tasks, breaking down problems, evaluating intermediate results, and collaborating to achieve broader objectives [8]. Compared to traditional pipeline-based automation, LLM-driven agents provide enhanced flexibility through dynamic decision-making and contextual understanding. Their ability to interpret unstructured technical specifications, integrate domain-specific knowledge, and perform nuanced reasoning makes them a powerful enabler for automating natural-language-driven workflows aligned with evolving telecom standards. Recently, this ecosystem has been further strengthened by the introduction of the Model Context Protocol (MCP) by Anthropic in 2024 [9], which defines a standardized interface for connecting LLMs with external tools, data sources, and memory systems. MCP enables structured interaction between models and their environments, supporting seamless retrieval, execution, and coordination across diverse resources

Within the evolving 3GPP Service-Based Architecture (SBA), the Network Data Analytics Function (NWDAF) has become a cornerstone for enabling network intelligence [10]. Introduced in 3GPP Release 15 to provide analytics to 5G Core Network Functions, NWDAF has matured through subsequent releases to support distributed analytics, real-time feedback loops, and machine learning model management. Release 17 further advanced this framework by separating the Model Training Logical Function (MTLF) and Analytics Logical Function (AnLF), enabling modular training and inference workflows and supporting external consumers such as OSS and Self-Organizing Network (SON) entities [11]. Ongoing standardization efforts by 3GPP Technical Specification (TS) 29.520 [11] and ETSI Experiential Networked Intelligence (ETSI ENI) group [12] have defined mechanisms for OSS to interface with the NWDAF through structured, service-based Application Programming Interfaces (APIs) supporting performance assurance, anomaly detection, and closed-loop automation.

Despite these advancements, a persistent challenge in the standards remains in translating high-level operator intents into the structured analytics interfaces of NWDAF. Recent studies leveraging Natural Language Processing (NLP) and LLMs for intent-based network management [13]–[15] often address limited scenarios with predefined intents and lack OSS integration. To bridge this gap, this paper proposes Agentic-NWDAF, a multi-agent, intent-aware framework that translates natural-language OSS intents into executable analytical service calls.

To the best of our knowledge, this work presents the first framework that employs LLM-based collaborative agents utilizing the MCP for seamless integration of 3GPP API documentation knowledge and NWDAF service execution. This architecture represents a key step toward intent-driven, agentic network intelligence in 6G, bridging the gap between declarative intent formulation and actionable analytics operations.

The key contributions of this paper are outlined as follows:

- We propose Agentic-NWDAF, a novel framework that automatically translates high-level, natural-language OSS intents into executable NWDAF service calls, enabling seamless realization of IBN and ZSM principles within 6G networks.
- Our approach harnesses the Agentic AI paradigm by orchestrating a swarm of collaborative, reasoning agents that collectively perform intent parsing, analytics service discovery, parameter extraction, workflow composition, and result synthesis.
- Agentic-NWDAF advances beyond current research and standardization efforts by introducing a MCP-based architecture for intent-driven NWDAF service management. To the best of our knowledge, this is the first work to propose MCP integration specifically designed for NWDAF, enabling a unified framework that bridges intent interpretation and standards-based network analytics execution.
- We construct a comprehensive benchmark dataset comprising 100 "golden intents" that serve as ground-truth labels for systematic evaluation. These intents encompass all NWDAF analytics categories as defined in 3GPP technical specifications and are validated by subject-matter experts to ensure domain fidelity and practical relevance.
- Through extensive evaluations, including expert assessments and an LLM-as-a-Judge approach, we demonstrate that Agentic-NWDAF achieves high task success rates, precise tool selection, and robust generalization across diverse analytics scenarios.

The rest of this paper is organized as follows. Section II provides a review of the related literature and highlights the key research gaps. Section III introduces the design and workflow of the proposed Agentic-NWDAF framework. Section IV describes the experimental setup, including the selected benchmark and evaluation metrics, followed by an in-depth analysis of the results. Finally, Section V offers concluding remarks and directions for future work.

II. Related Work

Recent research studies [13]–[15] have increasingly positioned IBN as a cornerstone for autonomous network management. The work in [16] frames IBN as a closed-loop automation paradigm encompassing intent profiling, translation, resolution, activation, and assurance, emphasizing the role of NLP in translating human intents into enforceable policies. Building on this, Velasco et al. [13] introduce an ML-driven orchestration framework that enables end-to-end IBN by continuously monitoring network conditions and dynamically executing management actions through a programmable data plane. Complementary to this, the authors in [14] propose an NLP-based intent translation approach that converts unstructured user intents into structured representations defined by parameters such as actor, objective, network action, target resource, and temporal scope, demonstrating measurable gains on industry-standard benchmarks. More recent research by Wang et al. [17] explores LLMs, such as OpenAI's GPT, for autonomous network management, and [15] presents an early LLM-based framework for intent extraction in 5G Core operations.

Despite recent advancements, existing approaches exhibit several key limitations: (i) they do not explicitly address NWDAF analytics orchestration, focusing instead on broader network management tasks such as slicing or configuration management; (ii) they rely mainly on NLP-based or single fine-tuned LLM solutions, limiting scalability and adaptability to evolving 3GPP standards; (iii) they depend on manually predefined structured intent representations (e.g., JSON), which restrict flexibility and require expert intervention for new use cases; (iv) they are not designed for telecom-grade OSS environments, lacking the integration depth required for operational deployment; and (v) most importantly, existing OSS–NWDAF interactions in current standards (i.e. 3GPP/ETSI) remain largely procedural, relying on static analytics subscriptions and rule-based triggers that limit their ability to dynamically interpret and operationalize high-level service intents, ultimately creating a bottleneck on the path toward fully autonomous networks. To address these challenges, this paper introduces Agentic-NWDAF, a comprehensive intent translation framework that transforms natural-language OSS analytics requests into executable NWDAF operations, enabling adaptive and autonomous analytics orchestration in next-generation networks.

III. Agentic-NWDAF Framework

In this section, we present the architectural design of Agentic-NWDAF and outline the translation workflow.

A. System Architecture

The proposed Agentic-NWDAF framework introduces a multi-layered architecture that connects human operators with autonomous analytics in next-generation telecom networks. It employs LLM-based reasoning agents to translate natural-language intents into executable NWDAF service calls, enabling intent-aware and zero-touch analytics orchestration. As shown in Fig. 1, the system consists of three layers: the User Plane Layer, where operators express high-level analytic intents (e.g., detecting abnormal network behavior or forecasting service degradation); The Agentic OSS–NWDAF Layer represents the cognitive core of the framework. It hosts a coordinated ensemble of autonomous reasoning agents responsible for understanding, decomposing, validating, and executing user intents. This layer connects the OSS domain to the analytical intelligence of NWDAF through MCP-enabled interfaces; and the Infrastructure Layer, which delivers 3GPP-compliant analytics through the MTLF and AnLF. The MTLF manages model training, provisioning, and lifecycle control, exposing services such as MLModelTraining and MLModel-Provisioning, while the AnLF handles analytics generation, subscription management, and inference execution through services like AnalyticsInfo and SubscriptionAnalytics [11].

A core innovation of Agentic-NWDAF lies in its MCP-driven service exposure, where the MCP server provides real-time access to 3GPP NWDAF API documentation. This allows agents to dynamically discover, interpret, and invoke services using structured schemas without manual configuration or model retraining. By adopting retrieval-based knowledge integration through MCP instead of fine-tuned models, the framework ensures continuous alignment with evolving 3GPP standards while remaining flexible, maintainable, and scalable.

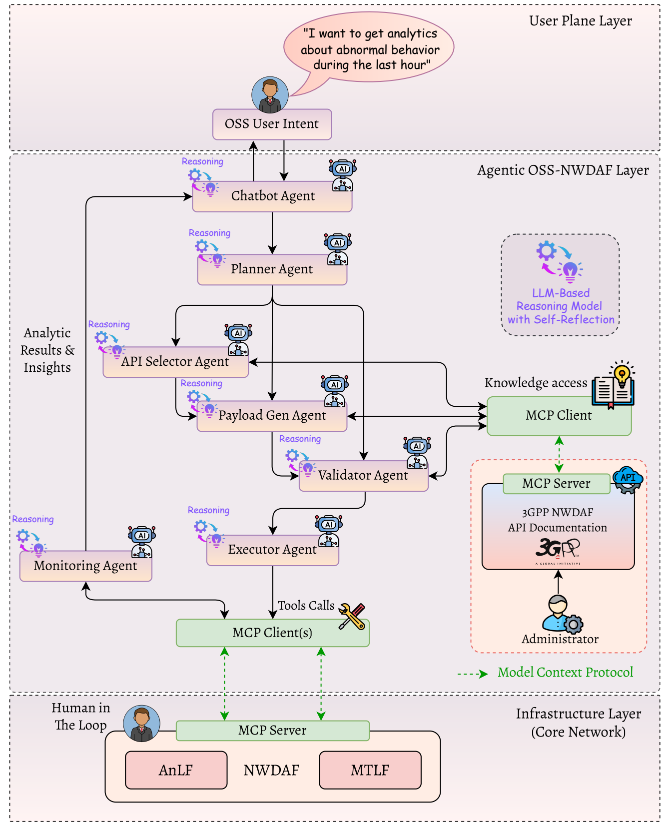

Fig. 1: Agentic-NWDAF System Design for OSS-NWDAF Intent Translation Using LLM-Based Reasoning.

B. System Workflow & Functional Roles

The proposed Agentic-NWDAF workflow operates through a coordinated system of self-reasoning agents that collaboratively translate natural-language intents into executable NWDAF operations, as illustrated in Fig. 1. The process begins when an operator submits an analytics request through the OSS interface. The Chatbot Agent first validates the intent to ensure it aligns with supported NWDAF use cases and adheres to semantic and ethical constraints (e.g., avoiding privacy-sensitive data). It then produces a structured representation, which the Planner Agent transforms into an actionable plan outlining analytical objectives and corresponding NWDAF services. The API Selector Agent consults the MCP client to retrieve 3GPP-compliant API specifications and determine the appropriate endpoints.

Once the target services are identified, the Payload Generation Agent constructs valid and schema-aligned payloads for the selected NWDAF services using knowledge retrieved via the MCP server. The Validator Agent verifies parameter integrity and adherence to NWDAF API specifications, after which the Executor Agent invokes the corresponding services through the MCP server interfacing with the AnLF and/or MTLF. The Monitoring Agent then supervises execution results, maintains feedback loops, and reports analytic outcomes back to the OSS. Through this distributed reasoning and MCP-based integration, Agentic-NWDAF ensures that each intent is contextually grounded, standards-compliant, and executed autonomously, enabling seamless and adaptive analytics orchestration for next-generation networks.

IV. Performance Evaluation

This section describes the benchmarking methodology used to evaluate the performance of the proposed Agentic-NWDAF framework. It outlines the evaluation setup and performance metrics, followed by a discussion of the obtained experimental results.

A. 3GPP NWDAF Intents Benchmark

To ensure a rigorous evaluation of the Agentic-NWDAF framework, a dedicated benchmark was constructed comprising 100 NWDAF intents derived from 3GPP technical specifications [11] and spanning a wide range of analytics scenarios. Each intent was carefully annotated by human experts with combined expertise in telecommunications and AI. To enhance reliability and objectivity, all annotations were independently reviewed by external domain specialists from EURECOM and partners from EU projects, ensuring consistency, transparency, and high overall evaluation quality.

B. Evaluation Metrics

The performance of the proposed Agentic-NWDAF framework was evaluated using state-of-the-art metrics to assess both individual agent performance and overall system effectiveness. The following metrics are employed [18]:

- **Task Success Rate (TSR):** Measures the proportion of user intents that are successfully translated and executed by the framework. It is defined as:

$$\text{TSR} = \frac{N_{\text{successful tasks}}}{N_{\text{total tasks}}} \times 100\%, \tag{1}$$

where $N_{\text{successful tasks}}$ is the number of tasks executed correctly and $N_{\text{total tasks}}$ is the total number of intents tested.

- **Tool Selection Accuracy (TSA):** Evaluates how accurately the reasoning agents select the appropriate NWDAF API/tool for a given task:

$$\text{TSA} = \frac{N_{\text{correct tool selections}}}{N_{\text{total tool selections}}} \times 100\%. \tag{2}$$

- **Execution Time:** Measures the average time taken by the system to translate a user intent into a completed NWDAF service call, including agent reasoning, payload generation, and service execution.
- **Token Usage:** Quantifies the number of tokens consumed by the LLM-based agents during the whole intent translation and reasoning processes.

C. Evaluation Setup

The proposed Agentic-NWDAF framework was implemented using CrewAI[1], an open-source platform for orchestrating autonomous AI agents with defined roles, objectives, and toolsets. All models were locally hosted via Ollama[2] to ensure deployment control and data privacy, and experiments were executed on an NVIDIA Jetson AGX Orin featuring a 2,048-core Ampere GPU, 64 Tensor Cores, and 64 GB of memory.

For the evaluation, we selected a set of high-performing, open-source LLMs that represented the state of the art at the time of development and were compatible with the available computational resources. Specifically, we employed non-reasoning models, including Llama3.1:8B, Qwen3:8B, Mistral:7B, and Falcon3:7B, together with reasoning-oriented models such as Deepseek-R1:14B, Phi4-Reasoning:14B, and GPT-OSS:20B. This combination enabled a comprehensive and diverse assessment across both general-purpose and reasoning-intensive settings. For LLM-as-a-Judge evaluation. we used the Mixtral 8×7B Mixture-of-Experts (MoE) model [19], which is a highly popular choice for LLM-as-a-Judge assessments. The model routes tokens through two expert groups per layer, activating about 12.9 billion of its 46.7 billion parameters.

For knowledge management, ChromaDB[3] was deployed as a vector database to support scalable embedding storage and efficient retrieval. The core network infrastructure was built by extending the open-source OAI NWDAF[4] implementation within the OAI 5G Core Network. The resulting architecture integrates the MTLF and AnLF components, facilitating advanced analytics across both training and inference workflows. Two MCP servers were implemented using FastMCP[5] to (i) expose NWDAF analytics services and (ii) provide access to 3GPP API documentation. Correspondingly, the agents were equipped with MCP clients, enabling them to access these servers as both tools and knowledge sources.

It is worth noting that our evaluation relied exclusively on open-source LLMs, guided by considerations of cost efficiency, model capability, and research transparency. Proprietary models were excluded due to their high operational costs in agentic systems, where frequent orchestration and generation cycles incur substantial usage-based expenses. Moreover, recent open-source models demonstrate competitive or superior performance and adaptability compared to proprietary counterparts. This choice also reinforces the principles of openness, reproducibility, and democratization in LLM research. All experimental configurations employed a single LLM instance across agents to ensure consistency. Multi-model or heterogeneous setups were not explored and are left for future work.

D. Evaluation Results

In this section, we comprehensively examine the experimental results across various metrics. To ensure a thorough evaluation of the proposed framework, we tested Agentic-NWDAF with two subsets of LLMs, distinguishing between reasoning and non-reasoning models. Reasoning LLMs perform multi-step inference and contextual analysis, whereas non-reasoning LLMs rely primarily on pattern-based text generation.

1) Overall Framework Performance: To evaluate the end-to-end effectiveness of Agentic-NWDAF, a series of experiments were conducted to examine how reasoning-enabled LLMs improve intent translation and decision accuracy in executing NWDAF service calls. Fig. 2 illustrates the Task Success Rate across different reasoning and non-reasoning LLMs. The results indicate a clear performance distinction between the two categories. Among the non-reasoning models, Mistral-7B and Qwen2.5-8B achieve moderate success rates around 75–78%, while Falcon-7B performs the lowest, highlighting limited reasoning capability in complex intent understanding. Conversely, reasoning-oriented models, particularly gpt-oss-20B and DeepSeek-R1-14B, demonstrate substantially higher success rates exceeding 85–90%. This improvement signifies that reasoning-oriented architectures are more effective in interpreting and operationalizing high-level OSS intents into precise NWDAF functions.

Fig. 3 presents the MCP Tools Selection Accuracy, which measures how accurately each LLM identifies the correct NWDAF analytical tools required for a given intent. The performance pattern mirrors that of the previous experiment.

[1]https://www.crewai.com/
[2]https://ollama.com/
[3]https://www.trychroma.com/
[4]https://gitlab.eurecom.fr/oai/cn5g/oai-cn5g-nwdaf
[5]https://gofastmcp.com/getting-started/welcome

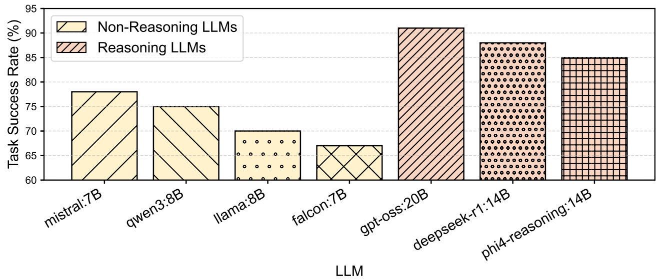

Fig. 2: Task Success rate across different reasoning vs non-reasoning LLMs.

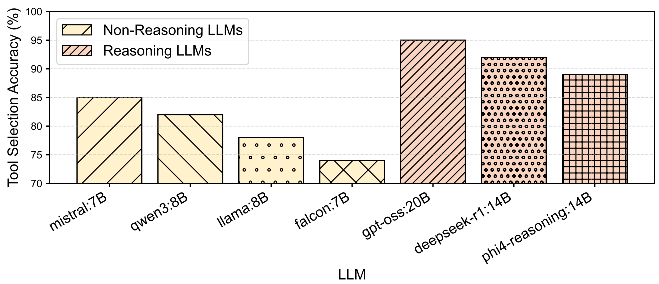

Fig. 3: MCP Tools Selection Accuracy across different reasoning vs non-reasoning LLMs.

Reasoning LLMs consistently outperform non-reasoning counterparts, with gpt-oss-20B reaching near-perfect accuracy 95%, followed closely by DeepSeek-R1-14B and Phi-4-Reasoning-14B. In contrast, smaller non-reasoning models such as Falcon-7B and LLaMA-8B exhibit noticeable drops, with accuracies between 70–78%. These results confirm that the integration of reasoning-enhanced LLMs within the Agentic-NWDAF framework significantly improves both task execution reliability and tool selection precision.

Building on the preceding analysis of task success and tool selection accuracy, these experiments evaluate the computational efficiency of the Agentic-NWDAF framework across different LLMs by analyzing execution time and token usage, providing insights into the trade-offs between reasoning capability and operational cost in intent translation and service invocation.

Fig. 4 depicts the Execution Time of different reasoning and non-reasoning LLMs across fifteen OSS intent categories. The data clearly differentiate reasoning-enabled models (gpt-oss-20B, DeepSeek-14B, Phi4-14B) from smaller non-reasoning models (Mistral-7B, Llama-8B, Falcon-7B, Qwen-8B). Reasoning models exhibit substantially higher execution times, ranging from approximately 90s to over 170s as intent complexity increases. This overhead stems from their deeper multi-step reasoning and context retention processes during NWDAF service selection. In contrast, lightweight non-reasoning models complete executions in under 50s, with only gradual increases for more complex intents. Although faster, these models sacrifice interpretative precision, reflecting a trade-off between latency and reasoning depth.

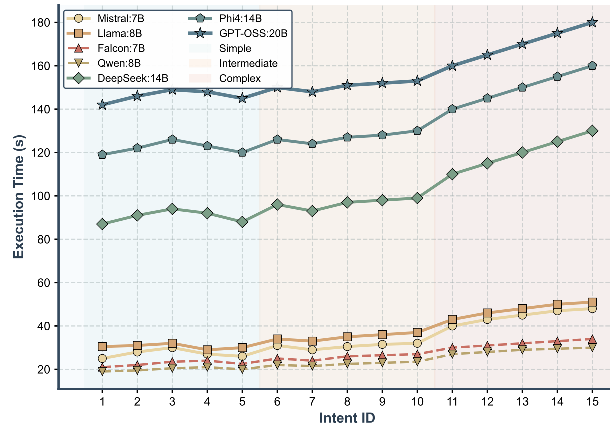

Fig. 4: Execution Time across different LLMs.

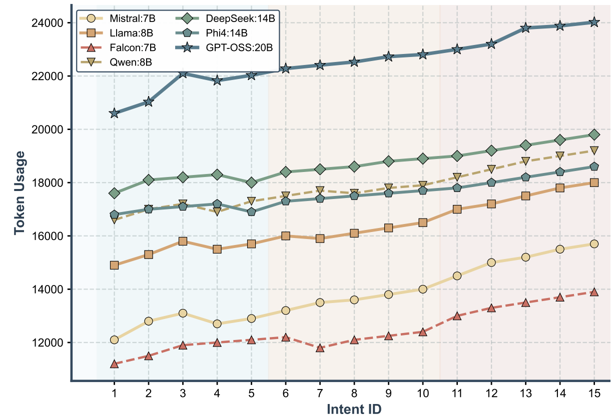

Fig. 5: Token Usage across different LLMs.

Fig. 5 illustrates Token Usage for the same set of models and intents. A consistent pattern emerges: reasoning-oriented LLMs consume considerably more tokens, reflecting longer intermediate reasoning chains and richer contextual elaboration. gpt-oss-20B exhibits the highest token count, surpassing 22,000 tokens for complex intents, followed by DeepSeek-14B and Phi4-14B, both exceeding 18,000 tokens. Conversely, smaller non-reasoning models maintain lower token consumption, typically between 12,000 and 16,000 tokens, indicating more direct but less context-aware processing.

In summary, reasoning LLMs offer more accurate intent-to-service translations but at higher computational and token costs, underscoring a trade-off in the Agentic-NWDAF framework between reasoning depth and execution efficiency.

2) LLM-as-a-Judge Evaluation: To conclude the performance assessment, Fig. 6 presents the LLM-as-a-Judge evaluation, assessing the quality and consistency of OSS intent translation and NWDAF service mapping within the Agentic-NWDAF framework. To ensure objectivity and reproducibility, the Mixtral 8×7B model was used as a unified evaluator across all models (rating out of 5). Results reveal clear performance variation: Falcon-7B scores lowest (2.4), Qwen-8B, LLaMA-

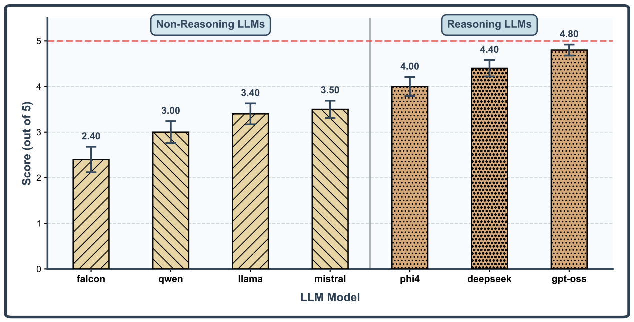

Fig. 6: LLM-as-a-Judge Evaluation.

8B, and Mistral-7B achieve moderate results (3.0–3.5), while larger models such as Phi4-14B, DeepSeek-14B, and gpt-oss-20B perform best (4.0–4.8), producing outputs most consistent with expert expectations.

E. Discussion & Learned Lessons

The extensive analysis and evaluation of the proposed Agentic-NWDAF framework underscore several important insights regarding its role in advancing intelligent automation within next-generation telecom operations. Results show that the proposed framework effectively connects high-level, natural-language OSS intents with executable NWDAF analytical services. By using reasoning-capable LLMs and agentic decision flows, it achieves robust intent understanding, precise tool invocation, and adaptive execution planning. These capabilities collectively support key industry goals such as ZSM and IBN, which are central to the 6G vision defined by 3GPP and ETSI. The framework demonstrates how agentic-driven orchestration can translate operational intents into measurable analytics actions, enabling more autonomous, context-aware, and resilient network management. However, several limitations remain. The main challenge involves the computational demands and infrastructure requirements of deploying large-scale LLMs in production. Ensuring performance consistency, cost efficiency, and effective lifecycle management of these models will be critical for practical adoption in future telecom environments.

V. Conclusion

In this paper, we propose Agentic-NWDAF, a novel intent-aware orchestration framework that bridges human intent and autonomous network analytics in 6G OSS. By leveraging LLM-based Agents and MCP, Agentic-NWDAF translates natural-language OSS intents into executable NWDAF service calls, enabling adaptive, context-aware, and fully automated analytics orchestration. Through a comprehensive benchmark of 100 validated NWDAF intents and rigorous LLM-as-a-Judge evaluation, we demonstrate that Agentic-NWDAF achieves high accuracy, robust generalization, and effective realization of ZSM and IBN principles. This work marks a significant step toward self-evolving, intent-driven network intelligence for the 6G era.

Future research will aim to minimize the computational footprint of the agents by investigating and implementing optimization techniques.

Acknowledgments

This work is supported by the European Union's Program under the 6G-DALI project (Grant No. 101192750).

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan and M. Zorzi, ”Toward 6G Networks: Use Cases and Technologies,” in IEEE Communications Magazine, vol. 58, no. 3, pp. 55-61, March 2020.

[2] M. Ameur, B. Brik and A. Ksentini, ”Dual Self-Attention is What You Need for Model Drift Detection in 6G Networks,” in IEEE Transactions on Machine Learning in Communications and Networking, vol. 3, pp. 690-709, 2025.

[3] A. Mekrache, A. Ksentini, and C. Verikoukis, “Next-Generation 6G Network Management with OSS-GPT,” in Proceedings of the ACM SIGCOMM 2025 Posters and Demos, 2025, pp. 158–160.

[4] GROUP SPECIFICATION, “ETSI GS ZSM 016 V1.1.1 (2024-10),” Etsi.org. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/016/01.01.01_60/gs_ZSM016v010101p.pdf.[Accessed: 14-Oct-2025].

[5] T. Forum, “Tmf921a intent management api profile,” tech report, 2022.

[6] J. Fonseca et al., “Realizing intent-driven network management with TM forum standards,” in 2024 IEEE 29th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2024, pp. 1–5.

[7] 3GPP, Management and orchestration; Intent driven management services for mobile networks, 3rd Generation Partnership Project (3GPP), Technical Specification 28.312, V18.0.0, Rel. 18, June 2024.

[8] D. Acharya, K. Bhaskar, and B. Kuppan, “Agentic AI: Autonomous Intelligence for Complex Goals-A Comprehensive Survey,” IEEE Access, 2025.

[9] “Introducing the Model Context Protocol,” Anthropic.com. [Online]. Available: https://www.anthropic.com/news/model-context-protocol. [Accessed: 26-Oct-2025].

[10] A. Tiwari, S. Das, A. Kumar and S. Srivastava, ”NWDAF in 5G: Architecture, Use Cases, and Evolution Across 3GPP Releases,” 2025 National Conference on Communications (NCC), New Delhi, India, 2025, pp. 1-6.

[11] 3GPP, 5G System; Network Data Analytics Services; Stage 3, 3rd Generation Partnership Project (3GPP), Technical Specification 29.520, V18.6.0, Rel. 18, July 2024.

[12] GROUP REPORT, “ETSI GR ENI 051 V4.1.1 (2025-02),” Etsi.org. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/ENI/001_099/051/04.01.01_60/gr_ENI051v040101p.pdf.[Accessed:14-Oct-2025].

[13] L. Velasco, M. Signorelli, O. G. De Dios, C. Papagianni, R. Bifulco, J. J. V. Olmos, S. Pryor, G. Carrozzo, J. Schulz-Zander, M. Bennis et al., “End-to-end intent-based networking,” IEEE communications Magazine, vol. 59, no. 10, pp. 106–112, 2021.

[14] Y. Njah, A. Leivadeas, J. Violos, and M. Falkner, “Toward intent-based network automation for smart environments: A healthcare 4.0 use case,” IEEE Access, vol. 11, pp. 136565–136576, 2023.

[15] D. M. Manias, A. Chouman and A. Shami, ”Towards Intent-Based Network Management: Large Language Models for Intent Extraction in 5G Core Networks,” 2024 20th International Conference on the Design of Reliable Communication Networks (DRCN), Montreal, QC, Canada, 2024, pp. 1-6.

[16] A. Leivadeas and M. Falkner, ”A Survey on Intent-Based Networking,” in IEEE Communications Surveys & Tutorials, vol. 25, no. 1, pp. 625-655, Firstquarter 2023.

[17] J. Wang et al., ”Network Meets ChatGPT: Intent Autonomous Management, Control and Operation,” in Journal of Communications and Information Networks, vol. 8, no. 3, pp. 239-255, Sept. 2023.

[18] M. Mohammadi, Y. Li, J. Lo, and W. Yip, “Evaluation and benchmarking of LLM agents: A survey,” in Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, 2025, pp. 6129–6139.

[19] A. Q. Jiang, et al., “Mixtral of Experts,” arXiv preprint arXiv:2401.04088, Jan. 2024. [Online]. Available: doi.org/10.48550/arXiv.2401.04088