Large Language Models in Intent-Based Networking: A Comprehensive Survey Across the Intent Lifecycle

Marwa Tageldien, Bassant Selim, and Lokman Sboui
Systems Engineering Department, École de technologie supérieure, Montréal, QC, Canada
Emails: marwa.tageldien@ieee.org, bassant.selim@etsmtl.ca, lokman.sboui@etsmtl.ca

2025 International Telecommunications Conference (ITC-Egypt) | 978-1-6654-5800-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ITC-EGYPT66095.2025.1186656

***Abstract*—The rising complexity of modern networks, particularly in 6G environments, demands scalable and autonomous management frameworks. Intent-Based Networking (IBN) addresses this challenge by enabling users to specify high-level operational goals rather than low-level configurations. However, traditional IBN approaches remain limited by their reliance on strict intent interpretation mechanisms. Large Language Models (LLMs), with their advanced semantic understanding and contextual reasoning capabilities, can offer a promising enhancement to the IBN lifecycle. Hence, in this survey, we present the first dedicated and structured analysis of how LLMs are being integrated into the IBN paradigm. We examine the most recent literature to trace the application of LLMs across all five phases of the IBN lifecycle: intent profiling, translation, conflict resolution, policy activation, and assurance. Unlike prior works that treat LLMs and network management in isolation, this survey emphasizes their convergence, detailing how LLMs support context-aware interpretation, flexible policy generation, and dynamic adaptation in response to network variability. Additionally, we present a comprehensive taxonomy that maps current research efforts of LLM to IBN phases and the specific LLM models used in each phase. Furthermore, the survey offers an analysis of the limitations and open challenges associated with deploying LLMs into IBN systems.**

***Index Terms*—Intent-Based Networking (IBN), Large Language Models (LLMs)**

I. INTRODUCTION

The evolution of modern communication networks, particularly with the appearance of fifth generation (5G) and the ongoing development of sixth generation (6G) architectures, has brought an unprecedented scale of complexity, heterogeneity, and dynamic behavior [1]. Managing such environments through traditional manual or semi-automated approaches, including rule-based automation and reactive configuration mechanisms, has become increasingly impractical, exposing critical limitations in scalability, adaptability, and operational efficiency [2] [3]. Moreover, the resolution of network issues often relies on skilled operators, resulting in long recovery times and potential misconfigurations, particularly in critical scenarios. In addition, the exclusive dependency on technical users or network administrators for routine and strategic operations may hinder system responsiveness and accessibility. To address these limitations, Intent-Based Networking (IBN) has emerged as a paradigm shift, aiming to abstract user goals, often expressed in natural language, into formalized intent representations that can be automatically interpreted, validated, and enforced by the underlying network infrastructure. However, conventional IBN frameworks often suffer from rigid intent parsing mechanisms that struggle with natural language variability, lack robustness in handling ambiguous or conflicting user goals, and fail to adapt to unforeseen network conditions, particularly in highly dynamic and heterogeneous environments [4]. These limitations hinder automated intent interpretation, result in misaligned configurations, and ultimately reduce system responsiveness. The integration of Large Language Models (LLMs) into IBN systems presents a promising direction to overcome these obstacles.

LLMs, especially those based on transformer architectures, such as T5, GPT-4, and LLaMA2, have demonstrated remarkable performance in natural language understanding, semantic abstraction, and few-shot learning [5] [6], making them suitable for dynamic, intent-driven systems. Their capability to generalize across diverse linguistic structures and infer contextual relationships allows them to bridge semantic gaps between human expressions and machine-executable policies [7]. Recently, LLMs have been increasingly leveraged to support various stages of the IBN lifecycle: from parsing user intents, translating them into enforceable configurations, resolving policy conflicts, and enabling autonomous assurance and self-healing [8]. However, this convergence introduces new challenges, such as the interpretability of generated policies, hallucination in translation stages, and scalability under constrained token windows.

In this survey, we provide the first structured and comprehensive review of the integration of LLMs into IBN systems. We present a detailed taxonomy of contributions across the five phases of the IBN lifecycle, categorize the role of various LLM architectures within each phase, and compare the scope of this survey with other IBN surveys. Additionally, we discuss the open research questions and challenges facing the field. Hence, this survey is intended to serve as a foundational reference for both researchers and practitioners seeking to leverage LLMs for more autonomous and adaptive network management.

The structure of this paper is organized as follows: Section II provides background on both IBN and LLMs. Section III

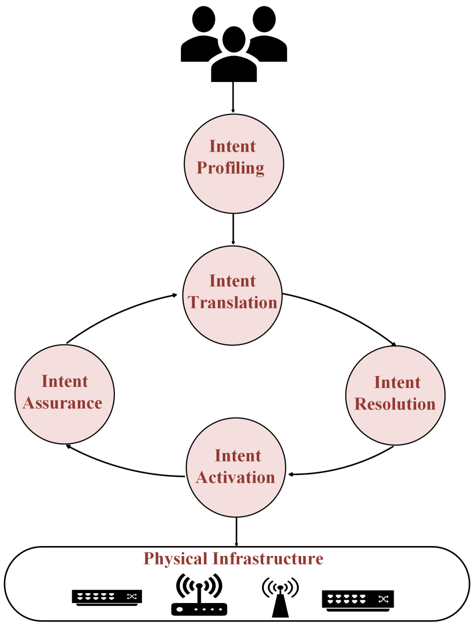

Fig. 1. Intent based Networking (IBN) Components [9].

presents a detailed discussion of how LLMs are utilized across the five phases of the IBN workflow, followed by a comparison between this survey and prior surveys in the field. Section IV outlines key challenges associated with integrating LLMs into IBN and highlights potential future research directions. Finally, section V concludes this work.

II. Essentials of IBN and LLMs

IBN and LLMs represent two converging paradigms at the forefront of autonomous and human-centric network management. This section provides foundational context for both technologies, highlighting the principles of IBN, the capabilities of LLMs, and setting the stage for understanding their integration in modern networking architectures.

A. Intent-Based Networking Overview

IBN is a type of network management that has an abstraction layer to enable Users (e.g., non-technical users, developers, network administrators, etc.) to specify high-level operational goals, known as intents, rather than detailed configurations [10]. Based on the 3rd Generation Partnership Project (3GPP) standards [11], intent is ”the expectations including requirements, goals, and constraints given to a system, without specifying how to achieve them”. For instance, within a private 5G network in a factory, a non-technical user might express an intent such as: “*Ensure reliable, low-latency video communication between office staff and engineers on the production floor during work hours*”. Once the intent is expressed, the IBN system is responsible for interpreting, translating, enforcing, and assuring it autonomously within the network infrastructure.

To achieve the business user’s intent, it should go through the IBN Life-cycle. As shown in Fig. 1, the conventional IBN system consists of the following components:

- **Intent Profiling**: This component serves as the initial interface between the user and the IBN system. It is responsible for capturing the user’s intent expressed in natural language and constructing a structured, intermediate representation of that intent (e.g., a semantic JSON object with fields like objective, QoS requirements, etc.). Then, this structured intent passed to the translation component for further processing.
- **Intent Translation**: This component takes the structured intent generated during profiling and converts it into network policy. It maps high-level goals and constraints into formalized network policies, which are then rendered into low-level configurations.
- **Intent Resolution**: In multi-user environment, the users submit the intents independently and this may lead to intent conflict. This component is responsible for detecting and solving the conflict between the new user’s intent and the deployed one. If it cannot solve the issue, it alert the user and/or the network administrator to solve it.
- **Intent Activation**: This component contains the network devices (e.g., routes, switches, access points (AP), etc), and it is responsible for deploying the network configuration into the appropriate network devices.
- **Intent Assurance**: This component is responsible for monitoring the network after deploying the network configuration, ensuring user’s intent is satisfied after deployment. It works as a reactive, detecting network errors, and proactive, predicting network errors, to achieve self-configuration and self-healing of the network.

Despite the potential of IBN, several practical challenges hinder its full realization. These include difficulties in accurately interpreting user intents expressed in natural language, managing conflicting goals across multiple stakeholders, and dynamically adapting to changing network conditions. Another key challenge lies in ensuring that the network continuously meets the users’ goals. Existing IBN systems, often built on predefined rules or lightweight machine learning models, face limitations in handling these issues, particularly in scenarios requiring contextual awareness and real-time adaptability [9].

Recent advancements in LLMs offer new opportunities to enhance IBN capabilities. Their strength in contextual understanding, semantic abstraction, and adaptive reasoning positions them as valuable tools to improve the accuracy of intent interpretation and the resolution of complex or ambiguous user goals [12]. Rather than replacing existing mechanisms, LLMs can complement and extend current IBN systems to support more intelligent, flexible, and user-centric network management.

B. Large Language Models Overview

LLMs are advanced deep learning architectures designed to process and generate human language. They are primarily based on transformer architectures [13], which leverage self-attention mechanisms to capture contextual relationships between words across long sequences. LLMs are trained on massive textual datasets to perform a wide range of natural language processing (NLP) tasks including text classification, summarization, question answering, dialogue generation, and language translation [14]. Recent advances in LLMs, PaLM, LLaMA, BERT, and GPT-4, have introduced powerful capabilities in contextual understanding, semantic generalization, and adaptive reasoning, making them suitable for applications requiring nuanced language comprehension [15]. These models exhibit emergent behaviors such as multi-step reasoning, commonsense inference, and domain adaptation, which make them highly suitable for applications that require accurate language understanding in complex and dynamic environments [16].

III. Integration of LLMs in IBN systems

The integration of LLMs into IBN frameworks offers a promising avenue for enhancing network management and automation. For example, as discussed in section II-A, IBN aims to simplify network operations by allowing administrators to specify high-level intents, which are then translated into network policies and configurations. Traditionally, translating these intents into executable policies has relied on rigid parsers or rule-based systems with limited semantic flexibility. LLMs, with their advanced language understanding capabilities, can facilitate this translation process by accurately interpreting user intents expressed in natural language and converting them into actionable network commands. This capability is valuable in dynamic network environments where rapid adaptation to changing conditions is essential. For instance, an LLM can map a user's goal, e.g., *"ensure minimum latency for video conferencing"*, into a series of intent objects and policy rules that configure traffic prioritization and QoS parameters.

In addition, LLMs have the ability to operate in bidirectional conversation to enhance IBN systems with interactive capabilities, allowing operators to refine intents through dialogue [17]. Furthermore, LLMs can support conflict resolution by reasoning over multiple conflict intents and suggesting optimal solutions based on network context and historical data [12].

Thus, the integration of LLMs in IBN architectures not only automates intent understanding and policy generation, but also aligns with the broader vision of cognitive, intent-driven, and human-centric networks envisioned for next-generation network management. The following subsections explore the contributions of LLMs across the key phases of the IBN lifecycle.

A. LLM for Intent Profiling

Traditional intent profiling methods in IBN systems often rely on fixed, template-based approaches such as Template/GUI-Based and Grammar/Keyword-Based [15] [18]. These methods often perform well in constrained scenarios, while they exhibit limited capacity to interpret the diverse and evolving expressions of user intents. They require extensive

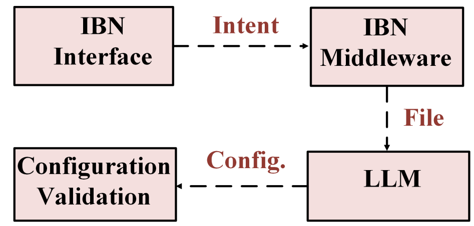

Fig. 2. LLM in Intent Translation [21].

manual rule creation and struggle when facing the syntactic and semantic variability inherent in natural language. As a result, these traditional methods are poorly suited to accommodate emerging application domains, new user goals, or personalized network configurations. To overcome these challenges, recent research has leveraged LLMs in the intent profiling phase to enhance semantic understanding and adaptability, enabling systems to extract user intents expressed in natural language without relying on manually crafted rules.

For example, LLMs are utilized in the intent profiling phase by enabling a conversational interface through a chatbot that captures user intents in natural language, as proposed in [19]. This system adapts its understanding and responses by considering each user's past interactions and contextual information, allowing it to personalize intent interpretation for different user profiles. In addition, the authors in [20] introduced an LLM-based Generative Intent Abstraction (GIA) framework to construct high-level intent representations. Rather than mapping user inputs directly to configurations, GIA transforms the intent into semantically enriched, context-aware structures that capture not only explicit user goals but also embedded constraints such as QoS requirements, service preferences, and environmental context. This system leverages a dynamic embedding mechanism to enrich intent profiles with real-time data, enabling the system to adapt to current network states and operational policies. Furthermore, the intent profiling was facilitated by a lightweight, fine-tuned LLM to process natural language inputs from network operators and extract actionable intent representations as proposed in [22]. The LLM employed a few-shot prompt engineering approach to classify each intent into predefined optimization categories such as throughput improvement, delay reduction, or energy efficiency, and to identify key attributes including the optimization target and its magnitude. These structured outputs served as semantic abstractions of the user's intent and were passed to downstream modules for validation and orchestration.

Moreover, LLMs and NLP models have demonstrated strong performance in intent profiling, particularly within private network environments. In private 5G networks, the proposed system in [23] employed an NLP-based Intent Engine, built using the APEX framework, to classify natural language intents and extract key attributes such as bandwidth requirements, latency sensitivity, and targeted service types. A semantic similarity algorithm was used to match user inputs with predefined system functions, and parameter values were normalized into a structured format compatible with downstream orchestration modules. While in [24], a GPT-3.5-based LLM was utilized to extract and classify multiple intents from user-generated natural language requests in the 5G Core network. The model was prompted to identify standardized intent types, such as deployment, modification, and performance assurance, based on 3GPP specifications. Additionally, the LLM was designed, by prompts, to provide an explanation for each classification, enhancing transparency and supporting future fine-tuning.

B. LLM for Intent Translation

Traditional translation approaches often rely on rule-based systems or static templates (e.g., blueprint mapping, declarative language mapping, etc.). These methods often lack generalization capabilities and struggle to handle the linguistic diversity and dynamic nature of user intents. To address these limitations, LLMs have emerged as promising tools for enabling flexible, context-aware, and prompt-driven translation of high-level intents into structured policy which aligns with programmable network infrastructure.

Several recent works have proposed leveraging LLMs to automate this translation process. For example, in [19], the LLM was employed for intent acquisition and translation within an IBN framework to improve user interaction and automation. For the translation phase, a fine-tuned LLM was utilized to convert the intents into structured configurations using Retrieval-Augmented Generation (RAG), function calling, and domain knowledge. In addition, a unified LLM management system was proposed in [21], where a fine-tuned LLM was adapted to generate router configuration in the network domain. As shown in Fig. 2, the LLM receives a structured file from the IBN Middleware component. This structured file contains the prompt role, intent, network information in a JSON file, and instructions about how to use the JSON file to generate the router configuration. After that, the LLM generates specific router configurations using a one-shot prompting strategy.

Another system proposed in [25] introduced an LLM-driven AI agent integrated into a Software-Defined Networking (SDN) to support zero-touch configuration in optical networks. In this system, the LLM receives structured requirements, extracted during the profiling phase as key-value pairs derived from natural language inputs, and translates them into machine-readable commands. These commands comply with YANG models and are formatted as XML-based configurations executable by the SDN controller using the NETCONF protocol. In addition, a proposed framework in [26] decomposed the user-defined intents into structured intermediate forms, which were subsequently processed by LLMs to generate network configurations. For the translation part in this framework, open source LLMs (Mistral 7B and Llama 13B) were fine-tuned using few-shot learning to generate Network Service Descriptor (NSD) JSON structures. This NSD aligns user goals with infrastructure requirements to enable seamless interpretation by orchestration engines. Similarly, another

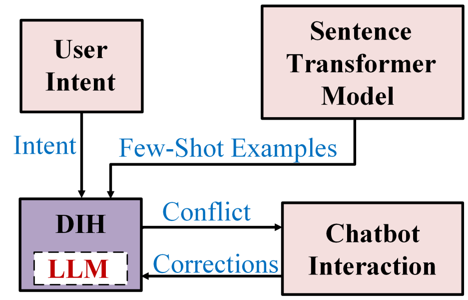

Fig. 3. LLM in Intent Conflict Resolution [26].

proposed framework in [27] utilized a few-shot prompting approach, where the LLM, Code Llama, was guided by previous validated examples stored in a knowledge base to generate NSDs based on the European Telecommunications Standards Institute-Network Functions Virtualization (ETSI NFV) standards. This system incorporated a human feedback loop to iteratively refine the knowledge base without modifying the LLM's parameters. Furthermore, another system in [28] leveraged LLMs, ChatGPT API (GPT-3.5 and GPT-4) for intent translation through few-shot prompting. This system works as follows: examples of valid policies in JSON format, key-value pairs, are embedded in the prompt to guide the LLM's output. These valid examples follow the monitor-analyze-plan-execute and knowledge (MAPE-K) structure. This enables the model to produce well-formed configurations that translate user objectives into enforceable configurations.

Additionally, a two-stage translation framework, called LIT, was introduced in [29]. This framework employs LLMs, fine-tuned on device manuals, to generate correct policy sequences and parameter values aligned with heterogeneous hardware requirements and real-time network conditions. The translation accuracy was enhanced through the use of Retrieval-Augmented Generation (RAG) and Mixture of Experts (MoE) models, where the LLM leverages manual knowledge and orchestrates expert policy models to produce optimized outputs. Furthermore, another system in [30] incorporated an Intent Rendering and Aggregation Generator (IRAG) to segment user intents into sub-intents. These sub-intents are processed by GPT-4o to generate partial configuration blocks, which are subsequently validated through an Intent Rendering and Validation (IRAV) module using syntactic and semantic consistency checks. This modular structure improves correctness and reduces translation latency compared to end-to-end prompting approaches.

Moreover, in [23], an LLM-assisted NLP engine was deployed to translate user-defined intents into structured API calls for the orchestration layer. The engine relies on semantic similarity, leveraging the LLM, to interpret user parameters like throughput and latency, and applies rule-based mappings on top of that semantic interpretation to convert them into precise, machine-readable configurations.

TABLE I
TAXONOMY OF LLMS CONTRIBUTIONS ACROSS IBN LIFECYCLE PHASES

Paper	Publication Year	Profiling	Translation	Resolution	Activation	Assurance
Kou et al. [20]	2025	✓	✓	✓		
Kou et al. [31]	2025	✓	✓		✓	✓
Wei et al. [30]	2025		✓			
Mekrache et al. [26]	2024		✓	✓		
Mekrache et al. [27]	2024		✓			
Dubey et al. [32]	2024					
Lira et al. [33]	2024			✓		
Araujo et al. [34]	2024					
Fuad et al. [21]	2024		✓			
Fontana et al. [19]	2024	✓	✓			
Guo et al. [29]	2024		✓			
Dzeparoska et al. [35]	2024					✓
Zhou et al. [25]	2024		✓			
Van Tu et al. [36]	2024					
Habib et al. [22]	2024	✓	✓			
Shah et al. [37]	2024	✓				
Manias et al. [24]	2024	✓				
Manias et al. [38]	2024	✓				
Mcnamara et al. [23]	2023	✓	✓			
Dzeparoska et al. [28]	2023		✓		✓	✓
Chatzistefanidis et al. [12]	2021	✓	✓	✓		✓

C. LLM for Intent Resolution

Traditional systems often rely on static conflict resolution rules or administrative overrides, which lack adaptability in dynamic environments and are difficult to scale. In contrast, recent research explores how LLMs, sometimes in combination with optimization techniques or human-in-the-loop guidance, can enhance automated resolution by reasoning over conflicting intent semantics and operational constraints.

As instance, the work in [26] proposed an LLM-centric IBN architecture in which conflict detection is performed before intent translation. As shown in Fig. 3, the architecture integrated LLMs within the Domain Intent Handler (DIH) to analyze user intents for semantic inconsistencies. This system targets intra-intent conflicts, where multiple objectives within a single intent may be logically incompatible; for example, requesting high throughput alongside minimal interference in radio configurations. Therefore, prior to translating the intent into network configurations, the LLM employs few-shot learning techniques to detect such contradictions and interacts with the user through a chatbot interface for clarification. This dialog-driven resolution process ensures that only valid and conflict-free intents are passed to the translation phase.

In addition, a framework was proposed in [12] to support pre-deployment multi-agent intent negotiation. It was designed to handle conflicting service requirements submitted by different business operators (BOs). The framework introduces a central entity, called MAESTRO, which identifies conflicts and potential collaboration opportunities among service intents. LLM-based agents, embedded within each BO and within MAESTRO, are activated to resolve these conflicts through structured negotiation. These agents are equipped with contextual modules, regulatory constraints, and optimization engines, enabling them to semantically interpret competing service-level objectives, such as divergent throughput requirements, and iteratively negotiate using prompt-driven reasoning.

D. LLM for Intent Activation

Activating the network configuration often depends on traditional methods such as Flow Activation and Service Function Chain (SFC) activation. These methods rely on rigid deployment templates or administrator intervention, lacking the flexibility to adapt to dynamic environments and heterogeneous network states. In contrast, LLM-enhanced activation frameworks aim to automate the translation-to-deployment bridge by generating executable outputs such as device configurations, API calls, or orchestration scripts in real time.

For example, the framework proposed in [28] integrates the activation phase within a closed-loop system, where LLMs generate policies from high-level user intents during the translation phase. These policies are mapped to domain-specific API calls and gradually executed within a MAPE-K loop that supports continuous monitoring and adaptive correction. Although the actual activation is carried out by downstream components, the LLM facilitates the process by producing deployment-aware translations aligned with the system's feedback and control mechanisms.

E. LLM for Intent Assurance

Intent assurance mechanisms incorporate closed-loop control, AI-driven anomaly detection, and self-healing capabilities as part of zero-touch networking. However, many of these solutions rely on predefined rules or specialized models that may lack flexibility across domains. LLM-based frameworks offer a complementary approach by introducing semantic understanding and context-aware reasoning to the assurance process. They can enhance proactive assurance by predicting violations, recommending corrective actions, and facilitating adaptive reconfiguration in dynamic environments.

TABLE II
LLM MODELS USED ACROSS SURVEYED PAPERS

LLM Model	Parameters	Open Source	Papers
GPT-3.5	175B	✗	[19] [20] [21] [23] [24] [25] [28] [36]
GPT-4	∼500B est.	✗	[20] [21] [28] [29] [31] [32] [35] [37]
GPT-4o	Unknown	✗	[12] [20] [30]
Mistral	7B	✓	[21] [26] [27] [33] [36] [37]
LLaMA	7B / 13B	✓	[26] [27]
LLaMA2	7B / 13B	✓	[21] [36]
CodeLLaMA	34B	✓	[26]
ChatGLM	6B	✓	[29]
Baichuan-LoRA	13B	✓	[29]
OpenChat	7B	✓	[36]
NeuralChat	7B	✓	[36]
LLaVA	7B	✓	[36]
BERT	110M / 340M	✓	[22]
Hermes	Unknown	✓	[37]
Unspecified LLM	Unknown	✗	[34] [38]

As evidence, an LLM-driven assurance framework was proposed in [35] to continuously monitor network alignment with user-defined intents by detecting "intent drift." In this framework, the LLM is employed to extract key performance indicators (KPIs) from the original intents and to interpret real-time monitoring data in order to identify potential deviations. When drift is detected, the LLM generates updated policies aimed at restoring alignment between network behavior and the user's initial intent. Additionally, the Emergence system introduced in [28] incorporates LLMs into the assurance phase by embedding them within a MAPE-K control loop. This loop enables the system to monitor policy execution in real time, assess compliance with the original intent, and, if necessary, prompt the LLM to revise or generate new policies based on observed deviations.

Furthermore, in [12], the LLM contributed indirectly to the assurance phase by participating in a closed-loop process guided by continuous monitoring. Although the LLM does not perform the monitoring itself, it leverages inputs from the monitoring unit to detect deviations from agreed service-level objectives (SLOs). When violations or resource shortages are observed, the system initiates a new round of negotiation, during which LLM-based agents revise the previously established policies. In addition, as proposed in [31], LLMs play a central role in intent assurance by generating executable reconfiguration actions in response to detected violations. Upon identifying an intent violation, such as increased response time, the system invokes an LLM (e.g., GPT-4o) to analyze real-time monitoring data and recommend system-level adjustments. These recommendations are then translated into configurations and automatically applied through orchestration platforms such as K3s and SDN controllers.

To contextualize our contribution within the broader landscape of IBN research, a comprehensive taxonomy of the surveyed contributions is provided in Table I, highlighting the integration of LLMs within each functional component. While Table II provides a detailed overview of the specific LLM models employed in the surveyed papers. Furthermore, Table III compares our survey with the prior IBN-surveys, highlighting the scope and contribution of each one.

IV. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

The integration of LLMs within IBN architectures presents notable advancements, but several fundamental limitations remain. These problems should be addressed to ensure LLM-enabled IBN systems are robust, reliable, and ready for large-scale deployment. This section outlines these key limitations and proposes future research directions.

A. Lack of Public Datasets and Benchmarks

A major barrier to progress in IBN systems is the absence of standardized, publicly available datasets and evaluation benchmarks. Most current systems are trained or tested on synthetic datasets or internal traces, which limits reproducibility, transparency, and meaningful comparison across approaches.

Suggested Research Direction: Future work should prioritize the creation of open datasets and community-shared benchmarks, ideally structured to cover the full IBN lifecycle across various network types. To ensure broad utility, the datasets should include: user intent expressions (natural language, CLI, GUI), corresponding ground truth configurations and policy outcomes, varying network states, traffic loads, topologies (e.g., 5G, SD-WAN, IoT), and annotated conflict scenarios and assurance violations. In parallel, shared evaluation benchmarks should be developed to assess translation accuracy, conflict detection, activation reliability, and semantic consistency. These resources would provide a critical foundation for training and fine-tuning, helping to accelerate progress and foster a reproducible research ecosystem.

B. Ambiguity in Multi-Intent and Vague User Expressions

In intent profiling, LLMs exhibit difficulty in disambiguating vague or multi-intent user expressions, particularly when statements combine several goals or omit explicit contextual anchors. This leads to partial or imprecise intent extraction.

Suggested Research Direction: To improve intent understanding, LLM-IBN systems should incorporate domain-aware prompts enriched with user metadata and telecom-specific ontologies. In addition, RAG techniques can contextualize new

TABLE III
COMPARISON BETWEEN THE IBN SURVEY PAPERS

Survey Paper	Year	Scope	Contribution
Zeydan et al. [39]	2020	Overview of IBN with focus on standardization, architectural layers, and early platforms.	Highlighted emerging IBN tools, identified early-stage challenges, and recommended future directions including intent languages and AI integration.
Pang et al. [40]	2020	Comprehensive survey on intent-driven networks and their enabling technologies.	Explained intent lifecycle model and layered architecture, introduced technologies like NLP, ML, and telemetry.
Leivadeas et al. [9]	2022	Extensive survey of IBN from both technical and operational perspectives, with use-case analysis.	Mapped IBN architecture and enabling technologies, categorized expression/translation approaches, and discussed open challenges in deployment, scalability, and semantic interpretation.
Ouyang et al. [41]	2022	Intent refinement in IBN including taxonomy and system design.	Defined refinement phases, categorized schemes by user type and method, proposed NLP+DFA implementation.
Gharbaoui et al. [10]	2023	General overview of IBN fundamentals, layered architecture, and benefits across industrial and SDN contexts.	Detailed integration challenges, highlighted the need for standardized frameworks and discussed compatibility with SDN/NFV.
Hurtado et al. [42]	2023	Focused on intent parsing in IBN using NLP and ML integration.	Reviewed parsing techniques, proposed methodology based on ML/NLP integration to enhance intent recognition and translation.
Mehmood et al. [43]	2023	Structured review of autonomous IBN for future cellular networks.	Discussed architectural framework, classified intents by stakeholder, analyzed integration with SDN/NFV and automation trends.
Minhas et al. [44]	2024	General overview of IBN concepts, architectural layers, and its application benefits in modern networks.	Summarized IBN's operational model, identified scalability and security challenges, and proposed standardization and AI inclusion.
Mihaeljans et al. [45]	2024	Overview of IBN core functions, layered architecture, and operational workflow, with insights into AI/SDN/NFV integration.	Defined IBN lifecycle and system components, identified assurance and scalability challenges, and proposed domain-specific models for autonomous and supervised IBN.
Our Survey	**2025**	**First survey to analyze the integration of LLMs into all phases of IBN and discuss the challenges and limitations for this integration.**	**Pioneers the analysis of LLMs in IBN through the 5-IBN phase, explains a taxonomy aligned with LLM capabilities, and highlights the challenges of integrating LLM with IBN.**

intents using past profiles, while structured entity recognition and intent extraction can improve parameter coverage.

C. Context Window Limitations in Transformer-Based LLMs

Transformer-based LLMs (e.g., GPT-3.5, GPT-4) are limited by fixed context windows that restrict the number of tokens they can process per query. This hampers the model's ability to maintain coherence over long, multi-turn dialogues or track historical intent evolution.

Suggested Research Direction: Future research should explore mechanisms for expanding LLM context capacity in IBN scenarios. Possible approaches include hierarchical prompt structuring, external memory modules, and the use of LLMs with extended context capabilities (e.g., GPT-4 Turbo, Claude 3). Combining these techniques with RAG pipelines can help bridge the short context limitations and preserve intent continuity across sessions.

D. Semantic Drift in Intent Translation

LLMs can produce syntactically valid but semantically incorrect or unsafe configurations during translation, a phenomenon known as semantic drift. Such misconfigurations pose serious operational risks, including policy violations.

Suggested Research Direction: A promising direction is the development of multi-stage validation pipelines that verify LLM outputs before deployment. These could include schema conformance checks (e.g., JSON or YANG formats), formal policy validation, and human-in-the-loop (HITL) checkpoints. Moreover, integrating Digital Twin (DT) can support pre-deployment simulation of generated configurations, allowing for testing and correction before execution.

E. Lack of LLM-Driven Conflict Detection Mechanisms

Despite the growing use of LLMs in IBN systems, the conflict detection phase remains largely underexplored. Few works have integrated mechanisms that can identify and resolve semantic or operational contradictions between co-existing or overlapping intents. Most current systems either omit this step entirely or rely on external validation logic and human oversight.

Suggested Research Direction: To address this gap, future research should investigate LLM-driven conflict reasoning frameworks capable of detecting both intra-intent (within a single intent) and inter-intent (across multiple intents) inconsistencies. This could include the development of domain-specific prompting strategies, rule-augmented LLM architectures, or hybrid systems that combine LLMs with formal conflict-checking engines.

F. Underexplored Role of LLMs in Intent Activation

While significant progress has been made in using LLMs for intent profiling, translation, and assurance, there is limited work focused on the intent activation phase.

Suggested Research Direction: Future research should include the design of LLM-compatible deployment pipelines that output structured representations (e.g., JSON) suitable for programmable APIs such as NETCONF. Additionally, integrating these pipelines with DT could enable safe testing and pre-deployment validation of activation strategies.

V. CONCLUSION

This paper presented the first comprehensive survey focused on the integration of LLMs within the IBN lifecycle. By reviewing and analyzing recent research contributions, this work categorized the roles of LLMs across all five IBN components: profiling, translation, resolution, activation, and assurance. It provided a detailed taxonomy of the LLM models adopted in current IBN systems and offered a comparison with existing IBN survey papers. In addition, it outlined the limitations associated with integrating LLMs into IBN frameworks and proposed future research directions. As LLMs continued to shape the evolution of intelligent network management, this survey aimed to serve as a foundational reference to guide future development and help bridge the gap between natural language-driven intent and autonomous network orchestration.

References

[1] W. Saad *et al.*, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[2] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE access*, vol. 8, pp. 133 995–134 030, 2020.

[3] C. De Alwis *et al.*, *6G Frontiers: Towards Future Wireless Systems*. John Wiley & Sons, 2022.

[4] A. S. Jacobs *et al.*, "Hey, lumi! using natural language for {intent-based} network management," in *2021 usenix annual technical conference (usenix atc 21)*, 2021, pp. 625–639.

[5] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[6] R. Taylor *et al.*, "Galactica: A large language model for science," *arXiv preprint arXiv:2211.09085*, 2022.

[7] J. Devlin and other, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[8] K. Dzeparoska, "Intent-based network management," Ph.D. dissertation, University of Toronto (Canada), 2024.

[9] A. Leivadeas and M. Falkner, "A survey on intent-based networking," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 625–655, 2022.

[10] M. Gharbaoui *et al.*, "Intent-based networking: Current advances, open challenges, and future directions," in *23rd International Conference on ICTON*. IEEE, 2023, pp. 1–5.

[11] 3GPP, "3GPP SA5 intent-driven management for mobile network," Rapporteur Presentation, Dec. 2024.

[12] L. Chatzistefanidis *et al.*, "Maestro: LLM-Driven Collaborative Automation of Intent-Based 6G Networks," *IEEE Networking Letters*, 2024.

[13] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[14] S. Minaee *et al.*, "Large language models: A survey," *arXiv preprint arXiv:2402.06196*.

[15] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang *et al.*, "A survey on evaluation of large language models," *ACM transactions on intelligent systems and technology*, vol. 15, no. 3, pp. 1–45, 2024.

[16] L. Berti, F. Giorgi *et al.*, "Emergent abilities in large language models: A survey," *arXiv preprint arXiv:2503.05788*, 2025.

[17] J. Liu *et al.*, "Balancing Accuracy and Efficiency in Multi-Turn Intent Classification for LLM-Powered Dialog Systems in Production," *arXiv preprint arXiv:2411.12307*, 2024.

[18] A. Leivadeas and M. Falkner, "VNF placement problem: A multi-tenant intent-based networking approach," in *24th Conference on ICIN*. IEEE, 2021, pp. 143–150.

[19] M. Fontana *et al.*, "Exploring large language models in intent acquisition and translation," in *10th International Conference on NetSoft*. IEEE, 2024, pp. 231–234.

[20] S. Kou *et al.*, "GIA: LLM-enabled Generative Intent Abstraction to Enhance Adaptability for Intent-Driven Networks," *Transactions on Cognitive Communications and Networking*, 2025.

[21] A. Fuad *et al.*, "An intent-based networks framework based on large language models," in *10th International Conference on NetSoft*. IEEE, 2024, pp. 7–12.

[22] M. A. Habib *et al.*, "LLM-based intent processing and network optimization using attention-based hierarchical reinforcement learning," *arXiv preprint arXiv:2406.06059*, 2024.

[23] J. a. Mcnamara, "NLP powered intent based network management for private 5G networks," *IEEE Access*, vol. 11, pp. 36 642–36 657, 2023.

[24] D. M. Manias *et al.*, "Towards intent-based network management: Large language models for intent extraction in 5G core networks," in *20th International Conference on the DRCN*. IEEE, 2024, pp. 1–6.

[25] A. Zhou *et al.*, "Large language model-driven AI agent in SDN controller towards intent-based management of optical networks," in *50th European Conference on Optical Communication*. VDE, 2024, pp. 1595–1598.

[26] A. Mekrache *et al.*, "Intent-based management of next-generation networks: An LLM-centric approach," *IEEE Network*, 2024.

[27] A. Mekrache and A. Ksentini, "LLM-enabled intent-driven service configuration for next generation networks," in *10th International Conference on NetSoft*. IEEE, 2024, pp. 253–257.

[28] K. Dzeparoska *et al.*, "Llm-based policy generation for intent-based management of applications," in *19th International Conference on CNSM*. IEEE, 2023, pp. 1–7.

[29] L. Guo *et al.*, "Following the Compass: LLM-Empowered Intent Translation with Manual Guidance," in *32nd International Conference on ICNP*. IEEE, 2024, pp. 1–12.

[30] Y. Wei *et al.*, "Leveraging LLM Agents for Translating Network Configurations," *arXiv preprint arXiv:2501.08760*, 2025.

[31] N. Akbari, J. Grundy *et al.*, "IntentContinuum: Using LLMs to Support Intent-Based Computing Across the Compute Continuum," *arXiv preprint arXiv:2504.04429*, 2025.

[32] A. Dubey *et al.*, "Leveraging large language models for intent-based generation of cloud-native configurations," in *International Conference on Advanced ANTS*. IEEE, 2024, pp. 1–6.

[33] O. G. Lira *et al.*, "Large language models for zero touch network configuration management," *Communications Magazine*, 2024.

[34] A. S. Araujo *et al.*, "An agentic approach for dynamic software-defined network management using large language models," in *Conference on NFV-SDN*. IEEE, 2024, pp. 221–226.

[35] K. Dzeparoska *et al.*, "Intent assurance using LLMs guided by intent drift," in *Network Operations and Management Symposium*. IEEE, 2024, pp. 1–7.

[36] N. Van Tu *et al.*, "Towards intent-based configuration for network function virtualization using in-context learning in large language models," in *Network Operations and Management Symposium*. IEEE, 2024, pp. 1–8.

[37] C. Shah *et al.*, "Using large language models to generate, validate, and apply user intent taxonomies," *arXiv preprint arXiv:2309.13063*, 2023.

[38] D. Manias *et al.*, "Semantic routing for enhanced performance of LLM-assisted intent-based 5G core network management and orchestration," *arXiv preprint arXiv:2404.15869*, 2024.

[39] E. Zeydan and Y. Turk, "Recent advances in intent-based networking: A survey," in *91st Vehicular Technology Conference (VTC)*. IEEE, 2020, pp. 1–5.

[40] L. Pang, C. Yang *et al.*, "A survey on intent-driven networks," *IEEE Access*, vol. 8, pp. 22 862–22 873, 2020.

[41] Y. Ouyang, C. Yang *et al.*, "A brief survey and implementation on refinement for intent-driven networking," *IEEE Network*, vol. 35, no. 6, pp. 75–83, 2022.

[42] R. Hurtado, C. Picón *et al.*, "Survey of intent-based networks and a methodology based on machine learning and natural language processing," in *International Congress on Information and Communication Technology*. Springer, 2023, pp. 363–382.

[43] K. Mehmood, K. Kralevska *et al.*, "Intent-driven autonomous network and service management in future cellular networks: A structured literature review," *Computer Networks*, vol. 220, p. 109477, 2023.

[44] S. Minhas, R. Jaswal *et al.*, "Revolutionizing networking: A comprehensive overview of intent-based networking," in *International Conference on INNOCOMP*. IEEE, 2024, pp. 463–468.

[45] M. Mihaeljans, A. Skrastins *et al.*, "Paramounts of intent-based networking: Overview," *Elektronika ir Elektrotechnika*, vol. 30, no. 6, pp. 53–59, 2024.