Large Language Models for Base Station Siting: Intelligent Deployment based on Prompt or Agent

Yanhu Wang, Muhammad Muzammil Afzal, Zhengyang Li, Jie Zhou, Chenyuan Feng, *Member, IEEE*, Shuaishuai Guo, *Senior Member, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

***Abstract*—Traditional base station siting (BSS) methods rely heavily on drive testing and user feedback, which are laborious and require extensive expertise in communication, networking, and optimization. As large language models (LLMs) and their associated technologies advance, particularly in the realms of prompt engineering and agent engineering, network optimization will witness a revolutionary approach. This approach entails the strategic use of well-crafted prompts to infuse human experience and knowledge into these sophisticated LLMs, and the deployment of autonomous agents as a communication bridge to seamlessly connect the machine language based LLMs with human users using natural language. This integration represents the future paradigm of artificial intelligence (AI) as a service and AI for more ease. As a preliminary exploration, this research first develops a novel LLM-empowered BSS optimization framework, and heuristically proposes four different potential implementations: the strategies based on Prompt-optimized LLM (PoL), human-in-the-Loop LLM (HiLL), LLM-empowered autonomous BSS agent (LaBa), and Cooperative multiple LLM-based autonomous BSS agents (CLaBa). Through evaluation on real-world data, the experiments demonstrate that prompt-assisted LLMs and LLM-based agents can generate more efficient, cost-effective, and reliable network deployments, noticeably enhancing the efficiency of BSS optimization and reducing trivial manual participation.**

***Index Terms*—Base station siting, large language model (LLM), Generative Pretrained Transformers (GPT), prompt engineering, agent engineering, AI as a Service**

I. Introduction

As the backbone of mobile communication networks, base stations not only provide seamless connectivity for mobile users, ensuring the continuity and reliability of communication, but also support the ever-increasing demand for high data throughput [1], [2]. This capability allows users to enjoy high-speed network services even while on the move. With the proliferation of smartphones and mobile devices, the number of mobile users has surged and the demand for data speed and quality has increased accordingly [3]–[5]. On-demand base station siting (BSS) has therefore become particularly critical and challenging, as it directly affects the breadth and depth of network coverage and the quality of service experienced by users [6]–[8].

Yanhu Wang, Muhammad Muzammil Afzal, Zhengyang Li, Jie Zhou, and Shuaishuai Guo are with School of Control Science and Engineering, Laboratory of Wireless Communication Technologies, Shandong University, Jinan 250062, China (e-mail: yh-wang,zhengyang_li,jiezhou@mail.sdu.edu.cn; muzammil_afzaal@yahoo.com; shuaishuai_guo @sdu.edu.cn)

C. Feng is with the Department of Communication Systems, EURECOM, Biot 06410, France (e-mail: Chenyuan.Feng@eurecom.fr).

T. Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

Conventional BSS techniques assess network performance and pinpoint areas for improvement predominantly through road testing and user feedback. Specifically, road testing involves the collection of signal quality data by deploying measurement equipment across predetermined geographic areas. Although this method can yield insightful insights, it is time-consuming and often impractical to implement, particularly in densely populated urban environments [9]. Moreover, road testing collects data at a specific time and location, which might not accurately represent the dynamic changes in the entire network performance over time [10]. User complaints and feedback constitute another essential method for identifying coverage issues and service quality deficiencies in traditional approaches [11]. However, activating this reactive strategy means that network improvements are typically not initiated until problems become severe enough to provoke user complaints. Additionally, the collected feedback may not be representative of the entire user population, potentially leading to biased or incomplete data [12], [13]. Engineers must engage in a continuous and iterative process that includes analyzing feedback, modeling BSS problems, developing solutions, deploying new base stations, and reevaluating network performance [14].

Given these limitations, traditional BSS approaches [15] demand engineers to possess extensive expertise in communications, networking, optimization, and programming, coupled with strong problem-formative, analytical, and problem-solving skills. The continuous advancements in telecommunications technology [16], [17] and the evolving patterns of user behavior [18] necessitate that engineers continually learn and adapt, further increasing the complexity of the task. Additionally, the dynamic character of urban environments, which is characterized by fluctuating traffic patterns [19], dynamic user mobility [20] and also time-varying user requirements [21], further complicates the BSS optimization.

A promising solution to these challenges is the integration of AI into BSS processes, especially large language models (LLMs) such as Generative Pretrained Transformers (GPT)-3.5, GPT-4 and GPT-4o. These LLMs can not only generate human-like text [22], [23], but also solve complex problems in a wide range of domains, including mathematics [24], programming [25], visual arts [26], medicine [27], and psychology [28]. Users can now describe requirements in natural language, enabling modeling and coding processes from semi-automatic to fully automatic, freeing up professionals to concentrate more on intricate problem solving and innovative design. For instance, LLMs are capable of writing numerical algorithms in mathematical and physical [29], generating computer code related to chemical equations and structures in chemistry [30], and assisting power engineers in problem decomposition and code generation in electronic engineering [31].

Regarding BSS problem, LLMs offer the following benefits: i) LLMs can process enormous amounts of real-time data from various sources, providing a comprehensive analysis of network performance, which allows for more efficient and accurate identification of weak coverage regions and service deficiencies; ii) LLMs can mitigate the lag associated with passive feedback mechanisms by proactively suggesting improvements based on continuously learning from network data and user feedback; iii) The ability of LLMs to adjust dynamically to alterations in traffic patterns and user behavior can be utilized to guarantee that the generated BSS solutions remain relevant and effective in rapidly evolving urban environments.

In light of this, this research investigates how AI, in particular LLMs, may revolutionize BSS by improving the effectiveness of the siting procedure as well as the general caliber of mobile network services. Precisely, we explore an innovative LLM-empowered BSS method, which can be described as four main strategies according to the degree of human intervention and interaction between agents as follows: i) **Prompt-optimized LLM-based (PoL-) strategy** aims to using crafting-optimized interaction prompts to guide the LLMs to autonomously accomplish BSS tasks with minimal human intervention. This strategy is centered on employing well-designed prompts to guide the LLM effectively comprehending and managing intricate BSS requirements, generating accurate and reliable siting solutions. ii) **Human-in-the-Loop LLM-based (HiLL-) strategy** aims to simplify user involvement in BSS decision making by allowing even users with limited expertise to express their needs in plain natural language descriptions. HiLL strategy comprises intelligent prompts, policy recommendations, problem decomposition, task planning, and comprehensive evaluation to improve efficiency and intuitiveness of the user's interaction with the LLM. iii) **LLM-empowered autonomous BSS agent-based (LaBa-) strategy** aims to develop an autonomous agent that is capable of managing the complete BSS process completely independently. To mitigate the hallucination problem, which is a typical issue in natural language processing (NLP), the system leverages external tools and databases to verify and validate the LLM's outputs. With the help of external validations, this agent is intended to assume complete control over BSS tasks, including data analysis and decision-making, and to increase the dependability of the generated siting strategies. iv) **Cooperative multiple LLM-based autonomous BSS agents-based (CLaBa-) strategy** allows multiple agents work collaboratively to solve the BSS problem. This strategy first translates the problem into mathematical formulas and then automatically generates solver code based on those formulas. The generated code is executed to produce the solution and saved to a file for further analysis. The effectiveness of the solution is verified by a series of unit tests generated by the LLMs and altered by the user. If an execution error or test failure is encountered, the system automatically feeds back the error message, triggering the LLM to revise the code in an iterative process until the proper solution is found.

To the best of our knowledge, we are the first to explore the use of LLM to solve the BSS problem. Our main contributions in this work can be summarized as follows:

- We formulate a novel LLM-empowered BSS framework and heuristically propose four different potential implementations. Specifically, the proposed PoL strategy enables LLMs to autonomously perform BSS tasks with minimal human intervention; the HiLL strategy achieves a more intuitive and user-friendly BSS process by incorporating human insight and preferences into decision-making procedure; the LaBa strategy realizes the independent completion of the entire BSS process; CLaBa strategy is designed to ulteriorly mitigate hallucinations.
- We conduct an empirical comparative analysis by using a dataset from the Mathorcup Undergraduate Mathematical Modeling Challenge 2022 [32]. This analysis not only assesses the performance of LLM-based tactics compared to traditional techniques in terms of traffic and cost-effectiveness, but also offers compelling data supporting the efficacy of LLM strategies in real-world applications.
- Besides propelling technological advancements in BSS, we also introduce fresh perspectives and tools to the telecom network design domain. Through these innovative approaches, we are able to achieve more efficient, cost-effective and reliable network deployment to fulfill the expanding demand for communications, while optimizing resource allocation and reducing operational costs. In addition, our study provides several frameworks for future researchers to further explore and improve methods for applying LLMs to solve complex engineering problems.

The rest of this paper is structured as follows. Section II provides a description of the BSS problem. The prompt-optimized LLM and the LLM-based autonomous BSS agent are demonstrated in Sections III and IV, respectively. In Section V, a thorough comparison study is performed. The entire paper is concluded in Section VI.

II. Problem Description

This section offers a thorough formulation of the BSS problem and defines the key performance metrics used to evaluate network performance. We outline the optimization problem for determining the optimal locations for new base stations and introduce the metrics of traffic coverage and cost, which are critical for effective network planning and deployment.

A. Problem Formulation

The primary task of BSS is to identify areas with poor coverage areas in the current network and strategically deploy new base stations to enhance the coverage in these regions [33]. As shown in Fig. 1, the bottom layer shows the communication coverage of the existing network in a given urban area, where the red zones indicate the areas with poor communication

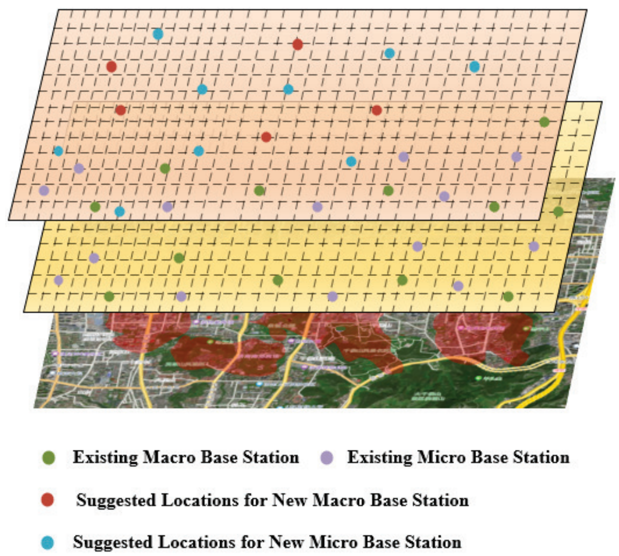


Fig. 1. **The coverage and planning of base stations within a given region.** The real-world map is shown on the bottom layer; existing macro and micro base stations are displayed on the middle layer; both planned and existing macro and micro base stations are marked in the top layer, along with proposed upgrades to address areas with poor coverage.

coverage. Such areas can be identified by expensive and time-consuming road tests or can be located by user complaints about signal quality. In practical network planning, it is infeasible to solve all weak coverage problems simultaneously due to the high cost of base station construction. Therefore, it is necessary to prioritize weak coverage areas with high traffic volume.

To simplify calculations and facilitate understanding, we divide the given area into small grids, as shown in the middle layer of Fig. 1, and focus on the central point of each grid. When dividing the grid, we assume that the coverage radius of either macro base station or micro base station is an integer multiple of the grid radius, so as to ensure that the base station deployed in the center of the grid can provide communication coverage for the whole grid. Such a partition ensures that no matter the size of the region, the location candidate set of new base stations can be treated as a finite number of points. The BSS decision-making is based on specific attributes of each point, including coordinates, quality of communication coverage, and also volume of traffic. The main notations are listed in Table I.

When deploying a new base station, the primary goal is to achieve seamless coverage service as much as possible in terms of the traffic flow, in addition, the distance between any two stations must be larger than a certain threshold in order to account for interference mitigation and deployment expenses. Telecom operators pursue lower costs while meeting signal coverage requirements via intelligent deployment of macro base stations and micro base stations, the former is characterized by a large coverage radius and higher construction cost, while the latter is more suitable for supplementary coverage and network capacity enhancement in particular places, such as hotspots with high traffic flow. Comprehensively considering

TABLE I: Main Notations and their Descriptions

Notation	System Parameter
N	Total number of grid points
$\mathcal{N}$	Coordinate set of all candidate locations for new base stations
$\mathcal{T}$	Coordinate set of all existing base stations;
d_h	Coverage radius of macro base station
d_d	Coverage radius of micro base station
C_h	Deployment cost of macro base station
C_d	Deployment cost of micro base station
T	Set of existing base station locations
$D_{\min}$	Minimum distance between any two base stations
p_i	Boolean variable indicating whether a macro base station is deployed at the central point of grid i or not
q_i	Boolean variable indicating whether a micro base station is deployed at the central point of grid i or not
w_t	Traffic volume associated with the point t
(x_i, y_i)	Coordinates of the central point of grid i
(x_j^e, y_j^e)	Coordinates of the jth existing base station

the characteristics of base stations, network coverage and deployment costs, the BSS optimization problem could be formulated as follows:

$$
\begin{aligned}
&(\mathrm{P1})\ \arg\min_{\{(p_i,q_i)\}_{i=1}^{N}} \sum_{i=1}^{N} (p_i C_h + q_i C_d)\,,\\
&s.t.(\mathrm{C1})\ \sum_{t\in G_i} w_t (P_{i,h,t} + P_{i,d,t}) \ge 0.9 \sum_{t\in G_i} w_t, \forall i \in \mathcal{N},\\
&(\mathrm{C2})\ p_i \in \{0,1\},\quad q_i \in \{0,1\},\quad \forall i \in \mathcal{N},\\
&(\mathrm{C3})\ p_i + q_i \le 1,\quad \forall i \in \mathcal{N},\\
&(\mathrm{C4})\ \sqrt{(x_i - x_j^e)^2 + (y_i - y_j^e)^2} \ge D_{\min},\\
&\qquad \text{if } p_i + q_i = 1\ ,\ \forall i \in \mathcal{N}, \forall j \in \mathcal{T}\\
&(\mathrm{C5})\ \sqrt{(x_i - x_n)^2 + (y_i - y_n)^2} \ge D_{\min},\\
&\qquad \text{if } p_i + q_i = 1 \text{ and } p_n + q_n = 1,\ \forall i, n \in \mathcal{N}
\end{aligned}
\tag{1}
$$

with

$$
\begin{aligned}
P_{i,h,t} &= P(p_i \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \le d_h),\\
P_{i,d,t} &= P(q_i \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \le d_d)
\end{aligned}
\tag{2}
$$

where $P_{i,h,t}$ and $P_{i,d,t}$ denote the probabilities that devices at location $t \in G_i$ (with the coordinate (x_t, y_t)) is covered by a macro base station or micro base station located at (x_i, y_i), respectively; G_i denotes the entire area in grid i; w_t denotes the traffic volume associated with location (x_t, y_t); $\mathcal{N}$ denotes the coordinate set of all candidate locations for new base stations, and N represents the total number of grid points; $\mathcal{T}$ is the coordinate set of all existing base stations; d_h and d_d represent the coverage radii of macro base station and micro base station, respectively; C_h and C_d are the costs of macro base station and micro base station, respectively; D_{min} is the minimum distance between arbitrary two base stations; p_i and q_i denote the Boolean variables indicating whether a macro base station or a micro base station is deployed at the central point of grid i, respectively; (x_i, y_i) is the coordinates of grid i's central point; (x_j^e, y_j^e) is the coordinates of existing base station j.

In the optimization problem P1, C1 indicates that the

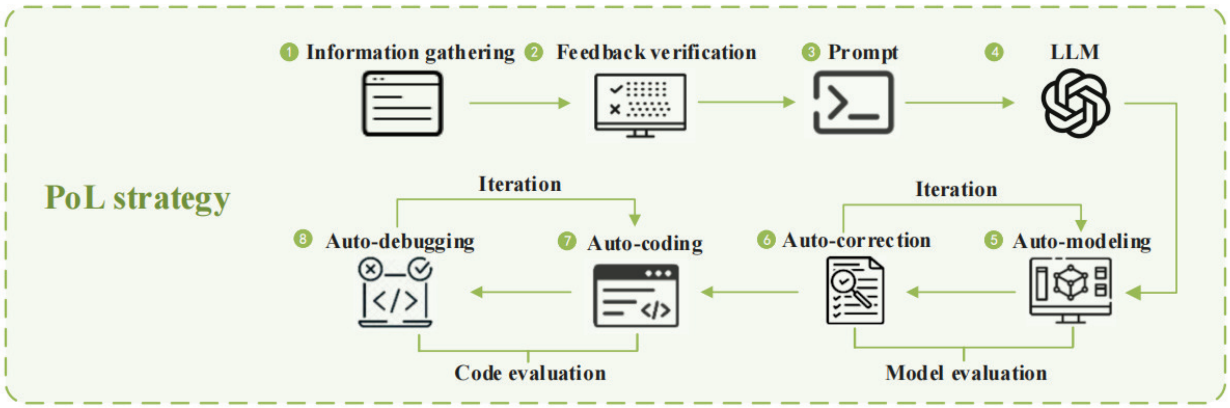


Fig. 2. **PoL strategy**: Guide LLM to effectively comprehend and manage intricate BSS requirements and produce solutions by employing optimized prompts.

communication coverage probability for data traffic should be greater than 90%; C2 shows the value of the deployment decision; C3 indicates that Only one new base station, either macro or micro base station, can be built in one single location; C4 and C5 indicate that the distance between any two base stations should be at least greater than $D_{\min}$.

B. Performance Metrics

To evaluate the effectiveness of BSS solution, we adopt traffic coverage and deployment cost as performance metrics, which are essential for BSS design and optimization since they could comprehensively reflect network performance and resource utilization efficiency. Specifically, traffic coverage is a critical factor in BSS optimization as it has a direct impact on network performance and user satisfaction. It refers to the percentage of total traffic in a particular area that is covered and served by base stations. Ensuring a high level of traffic coverage is essential for maintaining service quality and preventing congestion. Telecom networks can function more effectively and handle high traffic volumes without encountering congestion by covering a sizable portion of traffic, which is set at 90% in this work. This paper aims to develop a LLM-empowered framework to provide an efficient method for optimizing base station layout under the premise of maintaining satisfactory traffic coverage. Deployment cost refers to the total expenditure required to deploy and maintain the base stations needed to achieve the desired traffic coverage.

C. Traditional Method for BSS

In urban areas, traditional BSS method relies on road testing of communication signals and user feedback regarding call quality. This process requires communications engineers to perform multiple steps to ensure that the deployment of new base stations effectively improves network coverage and user experience. The details are as follows.

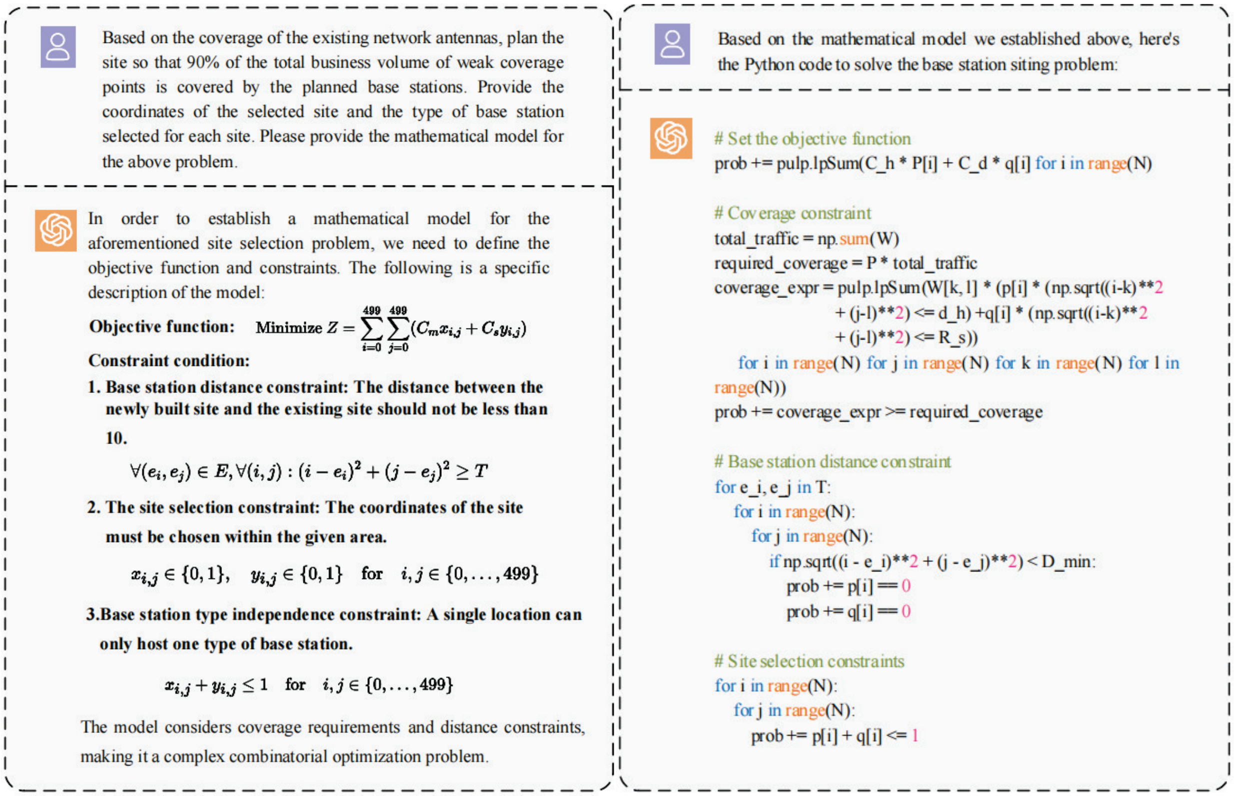
First, communications engineers conduct road tests to measure and record signal strength, coverage, and data transmission rates at different locations by driving test vehicles on urban roads. These test data provide engineers with a dispassionate assessment of current network performance and help identify areas of weak coverage and blind spots. The next step for engineers is to collect user feedback. Users typically report call quality issues, such as dropped calls, dropped calls, weak signals, or unstable data connections, by calling customer service or using mobile apps. The engineer then compiles all of these user feedback into a thorough problem report. This feedback reveals potential issues with the real user experience in addition to issues discovered during the road test. The engineers started modeling the base station location issue after gathering enough information and feedback. During this phase, several aspects must be taken into account, including topography, building barriers, subscriber density and the layout of existing base stations.

Developing the solution is the next critical step. Based on the results of the model analysis, engineers determine the optimal location for the new base station. They can use optimization algorithms, such as simulated annealing [34], genetic algorithms [35], or particle swarm optimization (PSO) [36], to balance coverage effectiveness, construction costs, and operational efficiency to find the optimal solution. Once the solution is determined, engineers will implement the construction of the new base station. Once the new base station is operational, engineers need to collect feedback again. They monitor the performance of the new base station and evaluate its improvement in network coverage and call quality through road tests and user feedback. Based on feedback, engineers may need to make further tweaks and optimizations to ensure that the new base station is getting the most out of it.

Despite being efficient, this conventional method has numerous drawbacks. Time and human resources are needed for the entire process, particularly during the data collection and processing stage. Furthermore, there is frequently a delay in the data gathering and processing of client feedback, which might lead to problems not being resolved in a timely manner.

III. LLM-EMPOWERED BSS OPTIMIZATION STRATEGY BASED ON PROMPT ENGINEERING

By introducing the advanced technology of LLMs, the efficiency of the BSS process can be significantly improved, reducing the need for manual intervention and making the optimization of BSS more intelligent and automatic. Thus, we



(a) Prompt for modeling

(b) Prompt for coding

Fig. 3. Auto-modeling and auto-coding in **PoL** strategy for the BSS problem. This figure demonstrates the basic modeling and coding process of LLM under the prompt of researchers. (a) In response to the prompt for modeling, ChatGPT 4o gives the basic mathematical model of BSS, including the objective function to minimize the total cost of the new base station and the constraints such as coverage constraints and base station distance constraints. (b) In response to the prompt for coding, ChatGPT 4o uses the pulp optimizer as the solver for the coding hint, and writes Python code to solve the BSS problem based on the mathematical model established above.

propose the LLM-enabled framework for BSS optimization and develop two LLM prompt engineering based strategies, namely, PoL- strategy, and HiLL-strategy. Next, we describe their workflow in detail.

A. *Prompt-optimized LLM-based (PoL-) Strategy*

The PoL strategy is presented in this subsection as a potential solution to the BSS issue in wireless networks. PoL strategy harnesses the sophisticated capabilities of LLMs to streamline the workflow for engineers by automating key steps, including prompt formulation, model validation, error identification, and iterative optimization. This automation not only expedites the BSS process but also reduce human intervention.

As depicted in Fig. 2, the PoL-based framework encompasses the entire workflow from preliminary site selection to final configuration determination. Through carefully crafted prompts, LLM can efficiently comprehend and handle the intricate requirements of BSS, consequently producing solutions. The purpose of optimizing these prompts is to guide LLM become more proficient in issue comprehension, modeling, and coding. Furthermore, Fig.3 illustrates the application of the framework in the processes of automatic modeling and coding. In response to researcher prompts, LLM is capable of providing the basic mathematical model for the BSS problem, including an objective function for minimizing the total cost of new base stations, as well as constraints such as coverage and base station distance. Additionally, the LLM utilizes the pulp optimizer as a solver and writes Python code to solve the BSS problem based on the aforementioned mathematical model.

In addition to showcasing the potential of artificial intelligence to automate the resolution of intricate combinatorial optimization problems, the LLM framework presented in this study also highlights the potent synergy between AI and human expert knowledge, providing a novel and reliable approach for the field of wireless network planning.

B. *Human-in-the-loop LLM-based (HiLL-) Strategy*

As illustrated in Fig. 2, the concept of the PoL framework is straightforward. However, our primary goal is to evaluate various LLMs' capabilities in addressing routine BSS tasks. In the site selection modeling phase, the precision of the LLM's suggested solution needs to be confirmed by engineering expertise. If there are discrepancies, feedback is provided through fresh prompts that specify the error type without explicitly stating the correction method. In the coding phase for base station configuration, the LLM-generated code is executed to identify any functional issues. Errors are communicated back to the LLM with the associated system error messages, prompting necessary adjustments. Upon successful

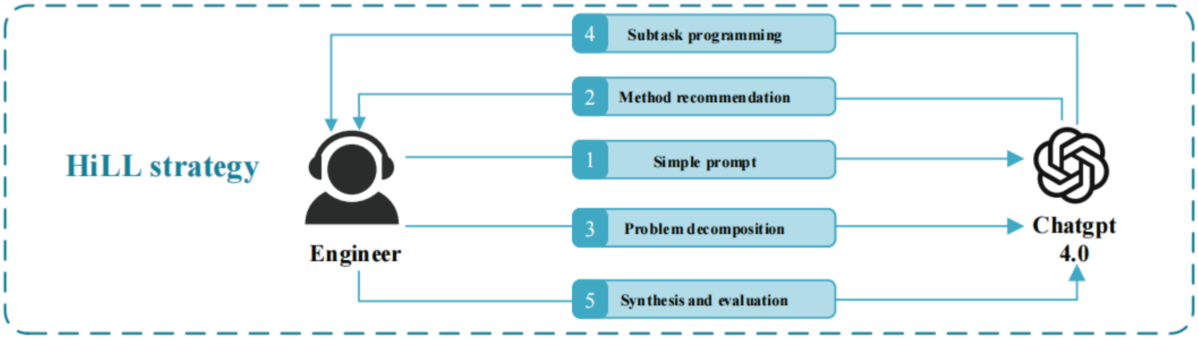
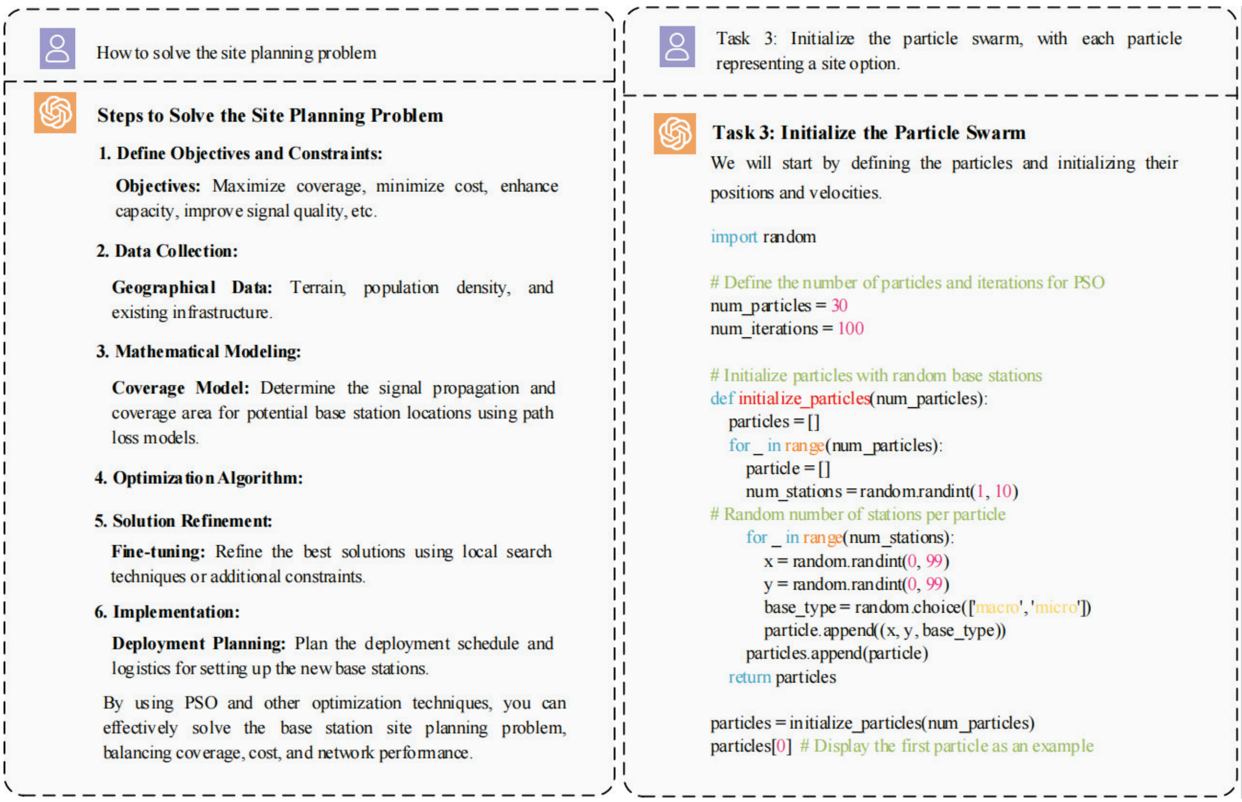


Fig. 4. In the **HiLL strategy**, the engineer and the LLM collaborate to solve issues and combine their respective advantages through human-machine interaction.



(a) Problem-solving strategies

(b) Demonstration of subtasks question and answer

Fig. 5. Problem-solving strategies and demonstration of subtasks question and answer for the BSS problem using **HiLL** strategy. This figure shows how users guide the LLM through heuristic questioning and feedback to generate a comprehensive solution. (a) In response to user questions, the LLM describes the BSS problem in depth, breaks it down into six phases for solution, and suggests attempting algorithm. (b) In response to detailed user questions, the LLM provides answers and solves the corresponding code based on the previous steps.

execution, the proposed base station configurations are validated against benchmarks set by established network planning tools. Any deviations are fed back into the LLM to refine the solution further. Through this iterative collaboration between communication engineers and LLMs, it is feasible to automate the programming and optimization of BSS problems.

A simple prompt, such as "How to optimize the layout of 5G base stations in an urban environment," may not provide sufficient details to allow an LLM to generate a comprehensive solution on its own. To address this challenge, we propose a HiLL framework that combines the computational advantages of LLMs with the expertise of telecommunications engineers. As shown in Fig. 4, the framework intended to serve as an algorithmic consultant for the project, using sophisticated models such as ChatGPT 4.0 to provide more targeted analysis, as illustrated in Fig. 4. For instance, ChatGPT can recommend literature on "heuristic methods for optimizing network infrastructure layout" and suggest the use of algorithms such as genetic and PSO algorithms [36] for dynamic BSS.

Following the framework's guidance, ChatGPT can break down the BSS problem into the following actionable steps, as shown in Fig. 5 (a). First, define objectives and constraints. Second, collect data. Third, perform mathematical modeling. Fourth, develop optimization algorithms. Fifth, refine solu-

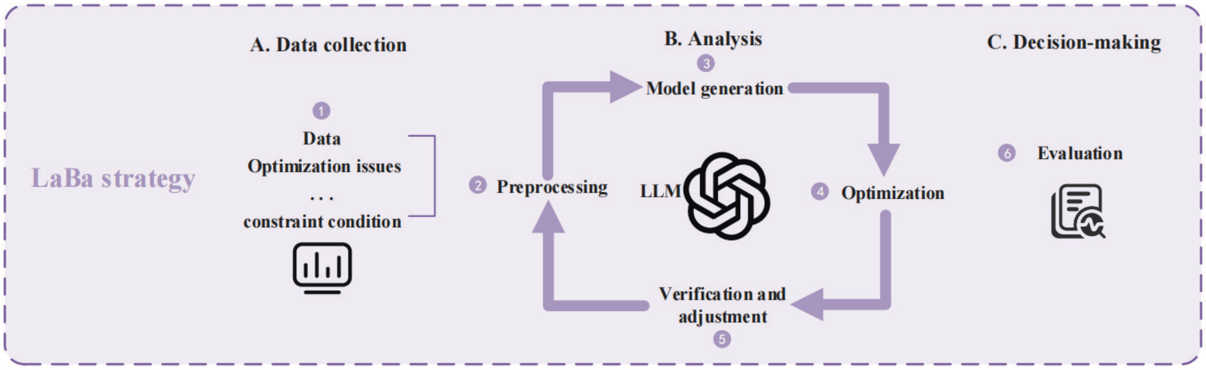


Fig. 6. **LaBa strategy**: LLM engages in the entire process from data collection and analysis to decision-making, introducing external tools and databases to verify and validate the output of LLM, achieving comprehensive automation and intelligence.

tions. And sixth, plan and implement deployment. This step-by-step approach requires a combination of domain-specific knowledge and programming skills, where human input is invaluable throughout the process.

Following an examination of the recommended literature and ChatGPT's deconstruction of the BSS issue, telecom engineers can further refine these tasks into more specific coding subtasks that LLMs can help with. Subtasks 1 through 5 include, for instance, reading and processing data, designing a fitness function to assess each site option, initializing the particle swarm and designating each particle as a site option, updating the particle position and velocity in accordance with the fitness function to find the best solution, and deciding on the site selection scheme based on the final group optimal location. ChatGPT's coding implementation for subtask 3 is displayed in Fig. 5 (b).

In summary, the HiLL strategy enables LLMs to recommend relevant literature and methods to initiate BSS projects. Once engineers grasp the technical roadmap, they need to decompose the problem into subtasks that LLMs can help with. The ability of LLMs to generate code snippets and provide programming guidance significantly reduces development time and complexity, allowing engineers to focus on synthesizing key components and optimizing the system. Through this collaboration, engineers and LLMs can precisely and effectively complete BSS tasks.

IV. LLM-empowered BSS Optimization Strategy based on Agent Engineering

To address the challenge of BSS in complex urban environments, we propose two LLM-empowered solutions based on human-computer interaction in Section III. To perform BSS tasks autonomously with minimal human intervention, we propose two fully intelligent LLM-based schemes, that is, strategies based on LaBa and CLaBa in this section. Next, we'll cover the details and advantages of both schemes.

A. LaBa Strategy

As depicted in Fig. 6, the LaBa is a highly automated system designed to independently handle the entire BSS process. This system leverages the natural language processing and understanding capabilities of LLMs to manage tasks from data collection and analysis to decision-making, with minimal human intervention. To mitigate the risk of "hallucinations" [37]—where the LLM might generate inaccurate or erroneous information—the system incorporates external tools and databases to verify and validate the outputs of the LLM, ensuring the accuracy and reliability of its decisions.

The agent begins by collecting a large amount of network performance data and user feedback. These data sources are diverse, including road test data and user feedback data. Road test data is collected by communications engineers who drive test vehicles through urban roads, measuring and recording signal strength, coverage, and data transmission rates. The agent uses these data to objectively assess current network performance, identifying areas with weak coverage and blind spots. User feedback is collected through customer service calls or mobile apps, where users report issues with call quality, such as dropped calls, weak signals, or unstable data connections. The agent aggregates this user feedback into detailed problem reports, supplementing the findings from the road tests.

After gathering sufficient data, the agent processes and deeply analyzes these data. Data processing steps include data cleaning and pre-processing to remove noise and invalid data, ensuring data quality. Subsequently, the agent extracts key features from the data, such as signal strength distribution and user density distribution, which serve as inputs for subsequent modeling. Using advanced modeling techniques and optimization algorithms, the agent generates BSS strategies. These algorithms, including simulated annealing, genetic and PSO algorithms, help the agent balance coverage effectiveness, construction costs, and operational efficiency to find the optimal solution.

To ensure the reliability of the generated siting strategies, the agent uses external tools and databases for validation. This step includes data accuracy checks and strategy effectiveness assessments. To be more precise, data accuracy checks verify the accuracy of the input data, ensuring a reliable data foundation. Strategy effectiveness assessments use simulation tools to evaluate the generated siting strategies in practical applications, considering metrics such as coverage, signal quality, and user experience. Through these external validation steps, the

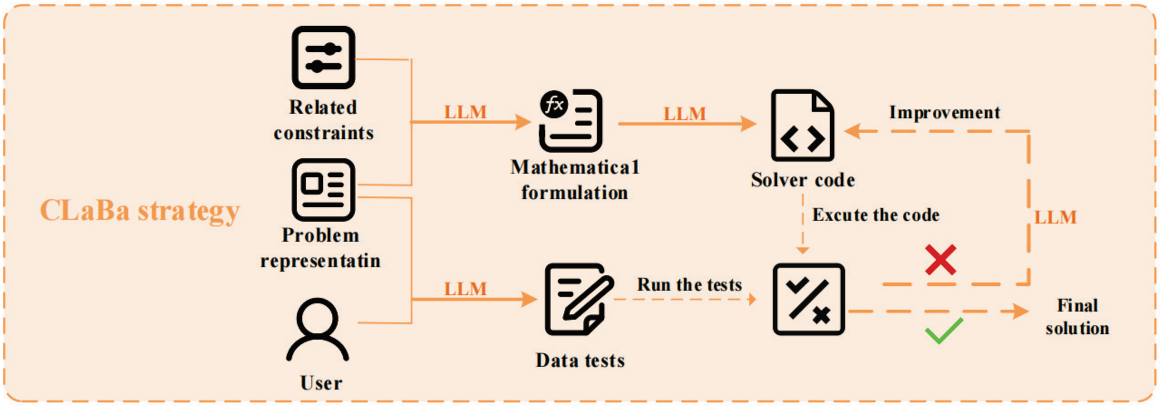


Fig. 7. **CLaBa strategy**: Multiple LLMs work together, where each is responsible for different aspects of BSS task. The LLMs convert optimization problems into mathematical formulas and generates solving codes. The obtained LLM data is verified based on users and issues through testing, and the code is improved and corrected.

agent significantly enhances the reliability and practicality of the generated strategies.

The validated siting strategies are then used to guide the construction and deployment of new base stations. Once the new base stations are operational, the agent continues to monitor their performance. Through regular road tests and user feedback, the agent assesses the improvements in network coverage and call quality brought by the new base stations. Based on the feedback received, the agent may need to make further adjustments and optimizations to ensure the continuous optimization of the new base stations' performance.

This LaBa strategy significantly improves the efficiency and accuracy of BSS, reducing the need for human intervention. However, the system also faces challenges, including ensuring data quality and consistency, the reliability of external validation, and the capability for continuous optimization. By integrating advanced natural language processing technology and external validation mechanisms, the LaBa strategy achieves comprehensive automation and intelligence in the BSS process, providing an efficient and reliable solution for the planning and optimization of wireless communication networks.

B. *CLaBa Strategy*

The CLaBa strategy represents an innovative solution designed to tackle the BSS problem through collaborative efforts of multiple LLMs. As shown in Fig. 7, this approach first translates the problem into mathematical formulas and then automatically generates solver code based on those formulas. The generated code is executed to produce the solution, which is then saved to a file for further analysis. To verify the solution's effectiveness, a series of unit tests, generated by the LLMs and modified by the user, are conducted. If an execution error or test failure occurs, the system automatically feeds back the error message, prompting the LLM to revise the code iteratively until the correct solution is found.

In the implementation process, multiple LLMs work together, each contributing to different aspects of the BSS task. Initially, these LLMs use natural language understanding and processing technologies to parse the input BSS requirements and related constraints, translating them into mathematical models. This process involves identifying key variables and constraints and expressing them as mathematical formulas for subsequent processing. Next, based on these mathematical formulas, the LLMs automatically generate solver code. This code, typically written in advanced programming languages, is designed to efficiently solve the BSS problem. Once the code is generated, the system executes it immediately to obtain a preliminary solution. To ensure the reliability of this solution, the system saves it to a file for further analysis and validation. After generating the solution, the system validates its effectiveness through a series of unit tests. These unit tests are automatically generated by the LLMs and adjusted by the user as needed. The purpose of these tests is to comprehensively examine the solution's performance under various conditions, ensuring its feasibility and effectiveness in practical applications. If any errors or test failures are detected during execution, the system logs the error messages and feeds them back to the LLMs. The LLMs then automatically correct the code, iteratively optimizing and refining the solution until the correct and effective solution is found.

Through this cooperative approach, multiple LLMs can fully leverage their own proficiencies to jointly tackle the intricate BSS issue. This method not only enhances problem-solving efficiency but also significantly reduces the requirement for human intervention. With the collaborative efforts of multiple LLMs, the system can quickly adapt to changing requirements and environmental conditions, providing flexible and efficient BSS solutions.

V. Experimental Result and Discussions

In this section, extensive experimental results are provided to verify the effectiveness of the proposed schemes.

A. *Experimental Setup*

The dataset for the 12th MathorCup College Mathematical Modeling Challenge in 2022, Problem D, involves a 2500×2500 grid area, where the coordinates of each grid point range from 0 to 2499[1]. The dataset consists of two

[1] http://www.mathorcup.org

parts: the first part includes the coordinates and traffic volume information of all weak coverage points within the region, and the second part contains the coordinates of existing network base stations. In this work, the coverage radii of macro and micro base station are set as 30 grids and 10 grids, respectively. The deployment costs of macro and micro base station are set as 10 and 1, respectively. The minimum distance between any two base stations is set as 10 grids. In order to ensure a realistic architecture and network optimization, the objective is to rationally plan the positions and types of base stations to cover at least 90% of the traffic volume of weak coverage points, taking traffic volume and cost into consideration.

In this paper, we proposed four LLM-based strategies, the first two of which emphasize the interactive engagement between humans and LLMs. Considering the practical limitations of time and efficiency, we have imposed a cap on the number of permissible interactions. Specifically, if a solution is not successfully formulated in ten interactions, or if the solution generated post this threshold fails to align with the stipulated criteria, the endeavor is deemed unsuccessful. This approach starkly contrasts with fully automated methods, enabling a stringent evaluation of the efficacy and efficiency of HiLL strategy.

B. Experiment Results and Analysis

1) Performance of PoL Strategy: The performance of PoL, based on ChatGPT 4o, was evaluated to determine the optimal placement of macro and micro base stations for effective coverage and cost minimization. The results, visualized in Fig. 8, show that PoL achieved a traffic coverage of 90.01% with a total cost of 26. The figure includes a clear legend, with red diamonds representing existing base stations, blue squares representing new macro base stations, and black hexagrams representing new micro base stations. The new macro base stations were placed to maximize coverage over larger areas, while the new micro base stations were deployed to fill coverage gaps and handle high-traffic areas. This approach ensures a balance between coverage efficiency and cost. The PoL's ability to achieve high traffic coverage with low cost can be attributed to its efficient use of prompts to guide the LLM in generating optimal base station placements. This approach leverages the LLM's ability to process complex requirements and generate reliable solutions.

2) Performance of HiLL Strategy: The HiLL, also based on ChatGPT 4o, was evaluated using the same metrics of traffic coverage and cost. This approach leverages both human expertise and the computational power of the LLM. The iterative process involves human experts providing feedback on the LLM-generated solutions, which the LLM then uses to refine its outputs continuously. This synergy allows the model to address complex problems that may be challenging for an LLM to solve autonomously. The results, illustrated in Fig. 9, indicate that the HiLL strategy achieved a traffic coverage of 90.01% with a total cost of 26. The HiLL's ability to achieve a higher traffic coverage with a lower cost can be attributed to the integration of human insights and preferences into the decision-making process. This approach allows for more nuanced and context-aware solutions.

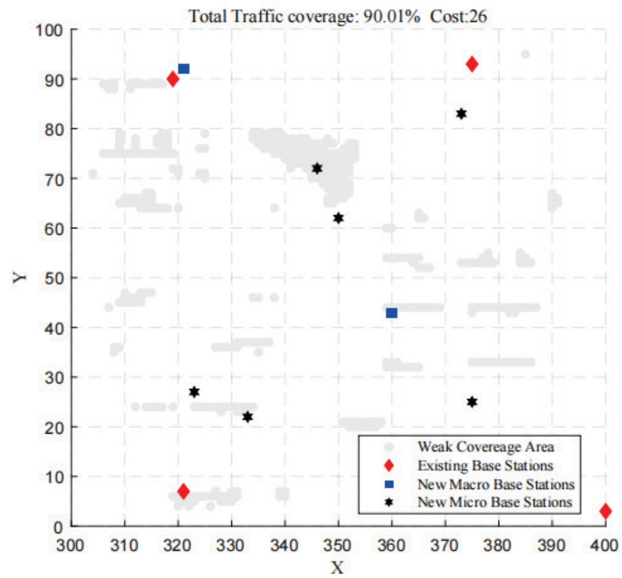


Fig. 8. PoL strategy employs ChatGPT 4o to solve the BSS problem and obtains the results of base station planning, including the distribution of existing base stations, new macro and new micro base stations.

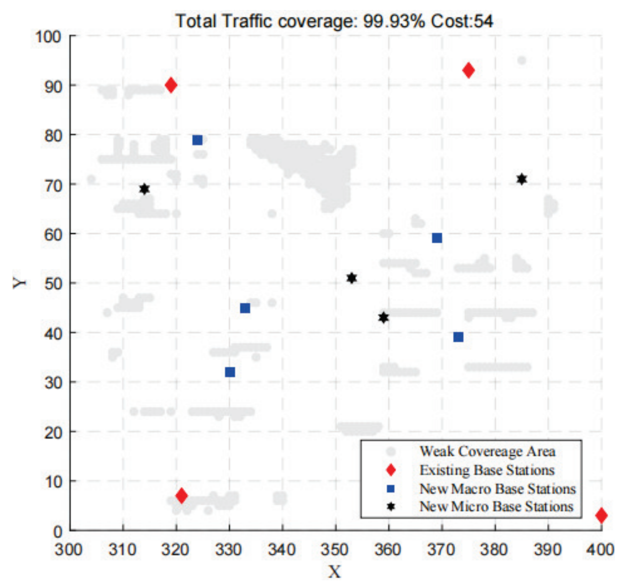


Fig. 9. HiLL strategy uses ChatGPT 4o to solve the BSS problem and obtains the solution to base station planning, including the distribution of existing base stations, new macro and new micro base stations.

3) Performance of LaBa Strategy: In this subsection, we evaluate the performance of the LaBa strategy in addressing the BSS selection problem. Unlike PoL and HiLL strategies, which rely on continuous human interaction, the LaBa strategy operates with minimal human intervention, autonomously managing the entire BSS process from data collection to decision-making. Similarly, the primary metrics used for evaluation were traffic coverage and cost.

Fig. 10 illustrates the distribution of weak coverage areas, existing base stations, new macro base stations, and new micro base stations. The LLM-based agent achieves 100% traffic coverage with a cost metric of 91, effectively balancing

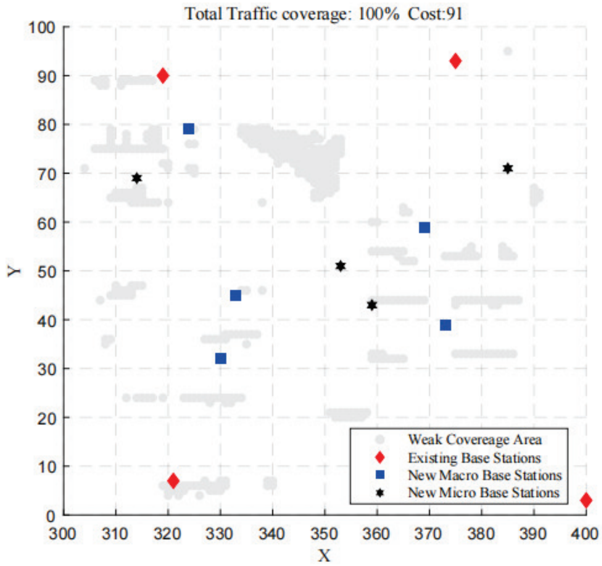


Fig. 10. LaBa strategy based on ChatGPT 4o is used to solve the BSS problem, and the results of base station planning are obtained, including the distribution of existing base stations, new macro and new micro base stations.

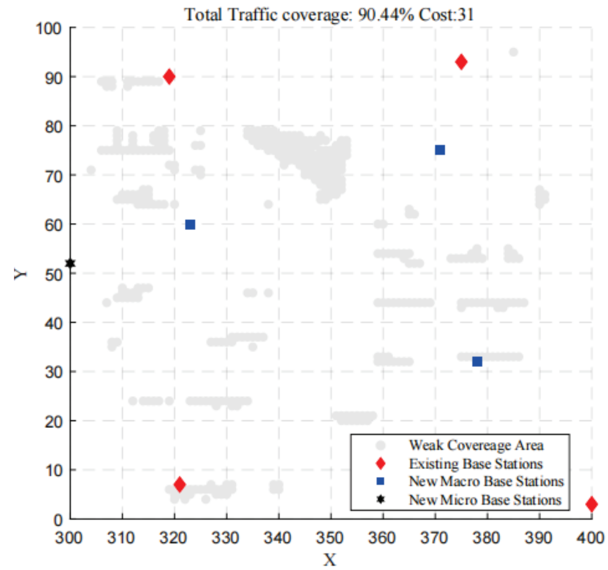


Fig. 11. CLaBa strategy is used to solve the BSS problem, and the results of base station planning are obtained, including the distribution of existing base stations, new macro and new micro base stations.

coverage needs with cost constraints. This comprehensive automation highlights the potential of LLM-based solutions in achieving efficient and effective base station planning, significantly reducing the need for extensive human involvement.

4) Performance of CLaBa Strategy: In this subsection, we evaluate the performance of CLaBa strategy in solving the BSS problem. Fig. 11 illustrates that the cooperative multiple LLM-based agents achieved a total traffic coverage of 90.44% with a cost metric of 31. This method effectively balances coverage needs with cost constraints, demonstrating the ability of multiple LLMs working together to provide a robust and cost-effective solution for base station planning.

Unlike the LaBa strategy, the CLaBa strategy emphasizes collaboration among multiple LLMs. Each LLM specializes in different aspects of the BSS task, such as problem representation, mathematical formulation, code generation, and solution verification. This cooperative approach enhances problem-solving efficiency and adaptability by allowing each LLM to focus on its strengths, significantly reducing the need for extensive human intervention. Furthermore, this collaborative effort facilitates continuous improvement and iterative refinement of solutions through feedback mechanisms, ensuring that the final outcomes are reliable and effective. The choice of traffic coverage and cost as performance metrics is justified because they directly impact the efficiency and effectiveness of base station placement. High traffic coverage ensures that the majority of weak coverage points are served, while low cost ensures that the solution is economically viable.

5) Performance Comparison and Analysis: In addition, we tested the success rate of four proposed LLM-based methods for solving the BSS problem, each of which was tested 10 times. From Fig. 12, it is evident that the CLaBa strategy has the highest success rate among the four approaches. The CLaBa approach's multi-agent cooperation mechanism

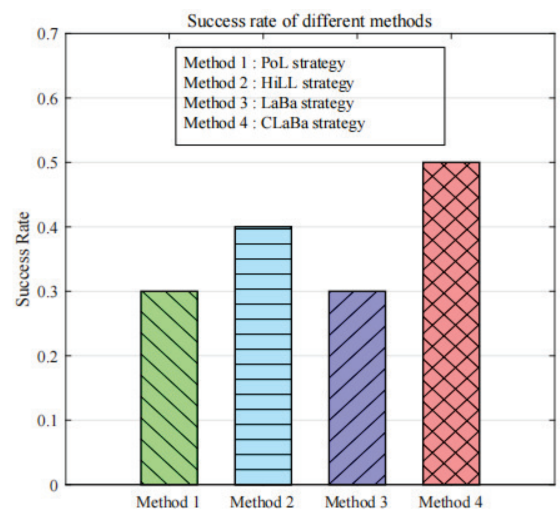


Fig. 12. Success rate of the proposed four LLM-based strategies for the BSS optimization problem.

primarily accounts for its superior performance in resolving the BSS issue. This approach divides duties among agents with different areas of expertise, resulting in a notable boost in efficiency and flexibility. Each agent focuses on its expertise, such as mathematical modeling, code generation, and solution validation. In addition, the CLaBa strategy enables continuous improvement and iterative optimization of the solution through a feedback mechanism, ensuring the reliability and effectiveness of the final result.

VI. Open Issues & Future Directions

In this section, we explore several open issues and also promising directions for future research and development in the integration of LLMs with next-generation networks and communications.

A. LLM-empowered AI Native Next-Generation Networks

The native intelligence of the next generation communication network can be rapidly established and boosted by fully utilizing LLM's potent natural language processing capability, the native intelligence of the upcoming generation of communication networks. For example, in future communication and networks, resource management is a core task to ensure efficient network operation. LLMs and other AI technologies play a significant role in resource management by improving the utilization efficiency of network resources through intelligent scheduling and optimization. Specifically, LLMs can analyze historical data and current network status, predict future network needs, optimize resource allocation in advance, and reduce network congestion and latency. The integration of LLMs does, however, come with certain difficulties, including designing flexible interfaces to adapt to different network environments, developing efficient algorithms to meet real-time requirements, and optimizing models to accommodate the resource constraints of network devices. By adopting a modular design, different components of the LLMs can be integrated into the network system as needed. Algorithm optimization can reduce computing resource consumption to ensure fast responses. Additionally, flexible interface design ensures that LLMs can operate efficiently in various network environments.

B. Task-oriented Selection in Human-LLM Interaction or Pure LLMs

For the future generation of networking and communication systems, it is essential to make the task-oriented decision between a human-LLM interaction framework or a fully automated LLM framework. On the one hand, the purely autonomous LLM-based framework can significantly improves efficiency by reducing human involvement. However, LLMs are known to suffer from the hallucination problem, where models can produce inaccurate or misleading information. This issue is particularly severe in automated network management and communication systems, where it can result in hazards and faults in the system. On the other hand, human-LLM interaction can mitigate the impact of hallucinations, improving system reliability. Human involvement can serve as a verification and correction mechanism to detect and correct erroneous information generated by LLMs promptly. For example, in an automated customer service system, customer service personnel can review and adjust the model's responses to ensure users receive accurate and reliable service. Although this approach may reduce overall efficiency, it enhances the accuracy and reliability of information, increasing user trust and reducing potential risks.

C. Lightweight Design Adaptive to Wireless Communication Networks

The emergence of mobile applications and edge computing in the realm of next-generation networks and communications has put further demands on the computing power and storage of models. Resource-constrained devices need to effectively reduce model complexity and size without compromising their functionality. Therefore, it is crucial to develop device-adaptive model lightweight and fast inference technologies.

Lightweight deployment aims to lower the volume and computing demands of models through a variety of techniques, including quantization, pruning, knowledge distillation, compression, hardware acceleration, etc. For instance, quantization is a technique that can drastically reduce the size of the model and speed up inference while maintaining acceptable accuracy levels, whose key ides is decreasing the precision of the numbers used in the model's weights and activations from floating-point numbers to integers. Pruning strategies simplify the model structure by removing superfluous connections from the neural network, which further reduces complexity. Rapid inference techniques focus on improving the execution speed of models on edge devices, especially when real-time or near-real-time performance is required. Adaptive inference technology can dynamically adjust the operating parameters of the model according to the current network environment and device performance, achieving optimal inference speed and accuracy.

All in all, lightweight deployment and fast inference are indispensable for applying LLMs to next-generation networks and communications. These technologies enable the efficient deployment and operation of LLMs on resource-constrained devices, meeting the need for low latency and high performance in areas such as real-time communications, smart devices, and the Internet of Things.

VII. Conclusion

This work demonstrates the significant potential of LLMs in addressing the BSS problem through innovative strategies. Specifically, the PoL strategy autonomously performs BSS tasks with minimal human intervention, efficiently generating reliable siting solutions through well-designed prompts. The HiLL strategy is particularly effective in scenarios where validation or feedback from user expertise is required to improve the quality of the decision-making process. The LaBa strategy independently manages the entire BSS process, utilizing external tools and databases to validate its outputs, ensuring the reliability of the generated strategies. The CLaBa strategy further enhances system adaptability and reduces the need for human intervention through collaborative problem-solving. These strategies can be extended to other domains such as urban planning, logistics, and resource management, where autonomous decision-making is crucial.

This innovative framework is expected to be swiftly extended to other task-oriented network optimization process as an first but significant baby step. This will allow humans to breezily inject experiences or specific requirements into various computation-intensive tasks, relieving them from a great deal of laborious and tedious work and facilitating more effective decision-making through human-LLM interaction. This cornerstone will serve as the foundation for more future work on AI as a Service and AI for insightful Science/Engineering.

References

[1] W. Cui, Y. Zhang, L. Wei, and M. Lu, "Optimizing base stations deployment in wireless access networks for smart grids," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, 2015, pp. 1–5.

[2] Z. Zhu, Y. Li, Y. Guo, Z. Zhou, and Z. Liu, "K-mean clustering algorithm and particle swarm algorithm based on base station siting optimization problem," in *Proc. IEEE Conf. Telecommun. Opt. Comput. Sci. (TOCS)*, Online, 2022, pp. 1042–1046.

[3] R. Chen and S. Guo, "Look-ahead task offloading for multi-user mobile augmented reality in edge-cloud computing," *IEEE Netw.*, vol. 37, no. 4, pp. 40–46, Aug. 2023.

[4] Y. Wang, S. Guo, Y. Deng, H. Zhang, and Y. Fang, "Privacy-preserving task-oriented semantic communications against model inversion attacks," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 8, pp. 10 150–10 165, 2024.

[5] X. Ge, L. Pan, Q. Li, G. Mao, and S. Tu, "Multipath cooperative communications networks for augmented and virtual reality transmission," *IEEE Trans. Multimedia*, vol. 19, no. 10, pp. 2345–2358, Oct. 2017.

[6] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.

[7] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 2, pp. 803–826, Feb. 2015.

[8] W. Mei and R. Zhang, "Joint base station and irs deployment for enhancing network coverage: A graph-based modeling and optimization approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8200–8213, Nov. 2023.

[9] S. Tayal, P. Garg, and S. Vijay, "Optimization models for selecting base station sites for cellular network planning," *Appl. Geomatics Civ. Eng.*, pp. 637–647, June 2019.

[10] M. Malekzadeh, "Performance prediction and enhancement of 5G networks based on linear regression machine learning," *EURASIP J. Wirel. Commun. Netw.*, vol. 2023, no. 1, pp. 1–34, Aug. 2023.

[11] M. Kulin, T. Kazaz, E. De Poorter, and I. Moerman, "A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer," *Electronics*, vol. 10, pp. 2079–9292, Oct. 2021.

[12] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Commun. Surv. Tutor.*, vol. 23, no. 2, pp. 668–695, Feb. 2021.

[13] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, Apr. 2014.

[14] H. Ganame, L. Yingzhuang, H. Ghazzai, and D. Kamissoko, "5G base station deployment perspectives in millimeter wave frequencies using meta-heuristic algorithms," *Electronics*, vol. 8, no. 11, pp. 2079–9292, Aug. 2019.

[15] C. V. Murudkar and R. D. Gitlin, "User-centric approaches for next-generation self-organizing wireless communication networks using machine learning," in *2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)*, Nov. 2019, pp. 1–6.

[16] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 4, pp. 2595–2621, Jun. 2018.

[17] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, Jul. 2020.

[18] J. Ryu, D. Im, and H.-J. Yoo, "AI SoCs for AR/VR user-interaction," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, San Francisco, CA, USA, 2021, pp. 1–4.

[19] J. Liu, Z. Wang, and L. Zhang, "Integrated vehicle-following control for four-wheel-independent-drive electric vehicles against non-ideal V2X communication," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 3648–3659, Apr. 2022.

[20] C. Feng, H. H. Yang, D. Hu, Z. Zhao, T. Q. S. Quek, and G. Min, "Mobility-aware cluster federated learning in hierarchical wireless networks," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 10, pp. 8441–8458, Oct. 2022.

[21] R. Lal, R. Singh, T. Raj, and Aditay, "ANN-based AR/VR for serving disabled people: A novel & comprehensive approach," in *Proc. Int. Conf. Artif. Intell. Innov. Healthc. Ind. (ICAIIHI)*, vol. 1, Raipur, India, 2023, pp. 1–6.

[22] S. Guo, Y. Wang, S. Li, and N. Saeed, "Semantic importance-aware communications using pre-trained language models," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2328–2332, Sep. 2023.

[23] J. Chang, K. Brantley, R. Ramamurthy, D. Misra, and W. Sun, "Learning to generate better than your LLM," in *Proc. NeurIPS Workshop Instr. Tuning Instr. Follow.*, New Orleans, LA, USA, 2023.

[24] A. R. Didolkar, A. Goyal, N. R. Ke, S. Guo, M. Valko, T. P. Lillicrap, D. J. Rezende, Y. Bengio, M. C. Mozer, and S. Arora, "Metacognitive capabilities of LLMs: An exploration in mathematical problem solving," in *Proc. Int. Conf. Mach. Learn. Workshop (ICML Workshop)*, 2024.

[25] E. Nijkamp, B. Pang, H. Hayashi, L. Tu, H. Wang, Y. Zhou, S. Savarese, and C. Xiong, "Codegen: An open large language model for code with multi-turn program synthesis," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Kigali, Rwanda, 2023.

[26] W. Hu, Y. Xu, Y. Li, W. Li, Z. Chen, and Z. Tu, "Bliva: A simple multimodal LLM for better handling of text-rich visual questions," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 3, Vancouver, BC, Canada, 2024, pp. 2256–2264.

[27] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature Med.*, vol. 29, no. 8, pp. 1930–1940, Jul. 2023.

[28] M. Sadeghi, B. Egger, R. Agahi, R. Richer, K. Capito, L. H. Rupp, L. Schindler-Gmelch, M. Berking, and B. M. Eskofier, "Exploring the capabilities of a language model-only approach for depression detection in text data," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Pittsburgh, PA, USA, 2023, pp. 1–5.

[29] A. AhmadiTeshnizi, W. Gao, and M. Udell, "OptiMUS: Optimization modeling using mip solvers and large language models," *arXiv preprint arXiv:2310.06116*, 2023.

[30] A. D. White, G. M. Hocky, H. A. Gandhi, M. Ansari, S. Cox, G. P. Wellawatte, S. Sasmal, Z. Yang, K. Liu, Y. Singh *et al.*, "Assessment of chemistry knowledge in large language models that generate code," *Digit. Discov.*, vol. 2, no. 2, pp. 368–376, Jan. 2023.

[31] R. Li, C. Pu, J. Tao, C. Li, F. Fan, Y. Xiang, and S. Chen, "LLM-based frameworks for power engineering from routine to novel tasks," *arXiv preprint arXiv:2305.11202*, 2023.

[32] "Mathorcup undergraduate mathematical modeling challenge 2022," http://www.mathorcup.org.

[33] J. Chen, Y. Shi, J. Sun, J. Li, and J. Xu, "Base station planning based on region division and mean shift clustering," *Math.*, vol. 11, no. 8, p. 1971, Apr. 2023.

[34] L. Du, Y. Xiang, and N. Xu, "Base station siting optimization based on greedy and simulated annealing algorithms," in *Proc. Int. Signal Process. Commun. Eng. Manag. Conf. (ISPCEM)*, Montreal, Quebec, Canada, 2022, pp. 229–232.

[35] K. Li, W. Wang, and H.-L. Liu, "6G shared base station planning using an evolutionary bi-level multi-objective optimization algorithm," *Inf. Sci.*, vol. 642, p. 119224, Sept. 2023.

[36] J. Li and X. Luo, "Particle swarm algorithm-based analysis of signal base station siting and signal coverage problems," in *Proc. IEEE Conf. Telecommun. Opt. Comput. Sci. (TOCS)*, Online, 2022, pp. 1024–1030.

[37] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin *et al.*, "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," *arXiv preprint arXiv:2311.05232*, 2023.