This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TMC.2026.3665626, IEEE Transactions on Mobile Computing

Large Language Model-Based Gray Wolf Optimization for Near-Field ISAC Networks

Zhen Chen, *Senior Member, IEEE,* Kezhi Wang, *Senior Member, IEEE,* Jianqing Li, *Senior Member, IEEE,* Xiu Yin Zhang, *Fellow, IEEE,* and Kai-Kit Wong, *Fellow, IEEE*

***Abstract*—The advent of extremely large antenna arrays and high-frequency signaling is expected to enable next-generation integrated sensing and communication (ISAC) networks to predominantly operate in the near-field region. Due to the dual influence of distance and angle on wave propagation characteristics in the near-field region, accurately modeling these characteristics remains a critical challenge. Motivated by the potential of large language models (LLMs) in angle prediction and distance estimation, an LLM-enhanced multi-objective optimization problem (MOOP) is developed to accurately capture the dependence of the channel on both the angular position and distance. The formulated LLM-enhanced MOOP framework is decomposed into a series of sub-problems, which can balance spectral efficiency for communication and localization accuracy for sensing. To overcome the computational and energy challenges associated with LLMs, a gray wolf optimization (GWO)-based algorithm is integrated as black-box search operator with LLM-specific prompt engineering to solve these sub-problems. Numerical results demonstrate that the proposed LLM-GWO scheme achieves an trade-off between communication and sensing performance, outperforming baseline approaches in terms of both Pareto front quality and convergence.**

***Index Terms*—Large language model, Grey wolf optimization, Multi-objective optimization, Near-Field ISAC, Pareto front.**

I. Introduction

THE forthcoming sixth-generation (6G) and beyond wireless networks are expected to support advanced applications, such as Integrated Sensing and Communications (ISAC), by leveraging much wider bandwidths at higher frequencies and significantly larger antenna arrays at base stations (BS) [1]–[3]. To address a massive number of access devices, the extremely large antenna arrays will likely be adopted to future communications. However, increasing the number of antennas also expands the antenna aperture, which may grow to several dozens or even hundreds of meters. This large size can lead to wireless communication occurring in the near field [4]. In conventional Fraunhofer (far-field) regions, electromagnetic (EM) propagation is typically approximated as a planar wave, where phase responses exhibit a linear relationship with angles [5]. However, as communications transition into the near-field region, the nature of signal propagation changes fundamentally. Unlike the far-field communication, the near-field communication requires the adoption of an accurate spherical-wave model, which captures nonlinear phase variations that depend not only on angles but also on corresponding distances [6]. This shift in modeling marks a critical distinction between the far-field and near-field regions. Importantly, it provides an exciting opportunities for near-field communications, as it enables more precise signal propagation characteristics.

Due to the unique characteristics of near-field propagation, research on ISAC systems from a near-field perspective has grown significantly [7]–[10]. For example., the authors of [7] provided a comprehensive overview of how near-field communication influences ISAC and discussed its potential applications. Meanwhile, the works in [11] and [12] focused on designing waveforms and beamforming methods for downlink near-field ISAC system. Specifically, [8] analyzed the uniform spherical wave, whereas [13] investigated the non-uniform spherical wave. Furthermore, [14] examined RIS-enabled UAV-to-vehicle channels in both near-field and far-field regions, offering a gain-loss mathematical model that emphasizes the importance of near-field systems. Despite these advances, the performance of near-field ISAC in both downlink and uplink settings were yet to be thoroughly explored. A major challenge lies in managing interference from target echo signals, which can vary depending on the shape, size, and material of the targets. Moreover, achieving a flexible balance between sensing and communication functions in practical near-field environments presents a significant challenge.

To address these issues, multi-objective optimization (MOO) techniques are crucial in near-field ISAC networks, as they allow sensing and communication performance goals to be optimized simultaneously [15], [16]. As a result, there has been increasing interest in incorporating metaheuristic algorithms into near-field ISAC network optimization, which are effective in handling complex environments. However, these objective functions often face conflicting objectives that include maximizing data throughput, minimizing Cramr-Rao lower bounds (CRLBs) to achieve accurate target localization, ensuring sufficient coverage and reducing communication en-

This work has been supported in part by the National Natural Science Foundation of China under Grant 62371197, in part by the Natural Science Foundation of Guangdong Province under Grant 2024A1515011172, in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (No. 2019D06). *(Corresponding author: Kezhi Wang.)*

Z. Chen is with the College of Cyber Security, Jinan University, Guangzhou, China. (e-mail:chenz.scut@gmail.com).

K. Wang is with the Department of Computer Science, Brunel University London, UB8 3PH Uxbridge, U.K. (e-mail: kezhi.wang@brunel.ac.uk).

J. Li is with the School of Computer Science and Engineering, Macau University of Science and Technology, Macau 999078, China (e-mail: jqli@must.edu.mo).

X. Zhang are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China. (e-mail: xiuyinzhang@foxmail.com).

K.-K Wong is with the Department of Electronic and Electrical Engineering, University College London, London, United Kingdom. (e-mail: kaikit.wong@ucl.ac.uk).

ergy consumption. Therefore, how to design MOO techniques are vital in enhancing the overall efficiency of near-field ISAC systems.

Recent developments have highlighted the integration of Large Language Models (LLMs) with MOO techniques to improve performance and adaptability. LLMs play two key roles in this setting [17], [18]. First, when provided with well-designed prompts, LLMs can generate solutions without requiring detailed problem encoding. These solutions can account for multiple objectives and constraints, that simplifying optimization by allowing users to describe goals in natural language, thus reducing the need for specialized technical knowledge. Second, LLMs can use the broad knowledge to design or choose effective operators for multi-objective evolutionary algorithms (MOEAs). By interpreting natural language descriptions of optimization problems, LLMs help devise robust strategies that improve search efficiency and convergence. This integration facilitates AI-driven control and interaction, advancing beyond traditional manual systems. One notable effort is presented in [19], where LLMs were used experimentally to generate novel optimization methods. Similarly, a few studies [20], [21] have demonstrated the feasibility of using LLMs as black-box solvers for single-objective optimization tasks, relying exclusively on prompting without additional training.

Nevertheless, how to integrate LLMs with the MOEA framework in a manner that ensures broad applicability across various MOEA implementations remains an open challenge. Since users/targets often operate in hybrid far-field and near-field environments with fluctuating parameters, existing studies primarily examine the effectiveness of these metaheuristic algorithms in static or predefined environments. However, real-world hybrid scenarios are inherently dynamic, and existing studies are limited by fixed or rule-based parameter tuning methods. This intrinsic coupling significantly increases the complexity of channel modeling and system optimization, especially when communication performance and sensing accuracy must be balanced simultaneously. This complexity requires accurate modeling of both angle-of-arrival and distance between the BS and the users or targets [22]. This intrinsic coupling significantly increases the complexity of channel modeling and system optimization, especially when communication performance and sensing accuracy must be balanced simultaneously. This complexity requires accurate modeling of both angle-of-arrival and distance between the base station and the users or targets. The motivation for employing LLMs for angle prediction and distance estimation arises from their natural language processing (NLP) capabilities, we propose a novel LLM-assisted MOEA framework for near-field ISAC networks that attempts to employ LLMs in MOPs. On the other hand, drawing inspiration from the social hierarchy and hunting behavior of grey wolves [23], a Grey Wolf Optimization (GWO)-based metaheuristic is integrated into the LLM as a context-aware layer. This enables the proposed model to dynamically adjust to environmental changes, thereby enhancing its responsiveness. Such context-aware adaptability directly addresses the dynamic requirements of near-field ISAC networks, which can enhance situational awareness and enable more informed decision-making throughout the optimization process. Our main contributions are summarized as follows

- A new LLMs model is developed that integrates GWO to balance the conflicting objectives of maximizing communication performance and minimizing sensing error in near-field ISAC systems. Unlike conventional GWO or other metaheuristics that rely on fixed strategies, our approach leverages natural language understanding for more adaptive decision-making. The proposed LLM-GWO model acts as a quasi-intelligent optimizer, potentially setting a new standard for adaptive metaheuristic algorithms.
- To tackle the proposed optimization challenging, the LLM-enhanced GWO is exploited in the MOO problem by decomposing the original problem into smaller subproblems. Unlike static heuristic operators, the LLM act as intelligent operators that can adjust based on the evolving optimization environments, which is integrated into the MOEA framework through tailored prompt engineering. This setup ensures that both objective functions and constraints are considered, aiming to maximize the SINR for communication while minimizing sensing error using the CRB criteria.
- Considering the complex nature of signal propagation in near-field ISAC systems, which is influenced by both angle and distance, the proposed LLM-GWO approach leverages natural language descriptions to define the relationship between antenna positions, target distance, and angle of arrival with respect to the physical parameters of the system. The LLM-GWO then generates mathematical expressions that accurately capture the nonlinear dependencies between these system parameters for angle prediction and distance estimation. Simulation results confirm the effectiveness of the proposed LLM-GWO operator, demonstrating its superior performance on an established benchmark.

The remainder of this paper is organized as follows. Section II introduces the system model and defines the optimization problem, including the objective functions and constraints. Section III presents an MOP optimization method using LLMs to solve the formulated problem. Section IV shows simulation results that demonstrate the effectiveness of our approach. Finally, the conclusion summarizes our findings.

Notation: $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, $\|\cdot\|$ and $\|\cdot\|^F$ denote the transpose, the conjugate transpose, pseudo-inverse, ℓ_2 norm, and Frobenius norm, respectively. $\otimes$ represents the Kronecker product operation. $a[i]$ is the i-th element of a vector $\boldsymbol{a}$.

II. Related Works

A. GWO-based Multiobjective evolutionary algorithm

The GWO algorithm draws its inspiration from the hierarchical leadership and hunting strategies exhibited by grey wolves in their natural environment [24]. In GWO algorithm, three top solutions are selected as leaders: best x_α, second-best x_β, and third-best x_δ. These leaders guide the remaining wolves, denoted as x_ω, to search for the global optimum.

The algorithm works through three main phases: encircling, hunting and attacking the prey.

1) Encircling: The encircling phase mimics how grey wolves surround the prey during the hunting process. This behavior can be mathematically represented as follows

$$x(t+1) = x_p(t) - A(t) \times |C(t) \times x_p(t) - x(t)|, \quad (1)$$

where $x(t)$ denotes the position vector of a grey wolf at the t-th iteration, and $x_p(t)$ is the position vector of the prey at the same iteration. The coefficients $C(t)$ and $A(t)$ influence the behavior of wolves, which are defined as:

$$\begin{aligned} A(t) &= 2 \times a(t) \times r_1 - a(t), \\ C(t) &= 2 \times a(t) \times r_2, \end{aligned} \quad (2)$$

where r_1 and r_2 are random variables within ranging between 0 and 1. These random factors introduce exploration capabilities into the algorithm. Meanwhile, the convergence factor $a(t)$ reduces linearly from 2 to 0 as the number of iterations increases, which is calculated as:

$$a(t) = 2 - \frac{(2 \times t)}{t_{\max}}. \quad (3)$$

This gradual reduction enables the wolves to shift from exploratory behavior in the early stages of the process to a more exploitative focus as the iterations advance.

2) Hunting: By considering the location of the three best hunters (x_α, x_β, and x_δ), the other wolves x_ω follow them to hunt according to

$$\begin{aligned} D_\alpha(t) &= |C_\alpha(t) \times x_\alpha(t) - x(t)| \\ D_\beta(t) &= |C_\beta(t) \times x_\beta(t) - x(t)| \\ D_\delta(t) &= |C_\delta(t) \times x_\delta(t) - x(t)| \end{aligned} \quad (4)$$

$$\begin{aligned} x_\alpha(t+1) &= x_\alpha(t) - A_\alpha(t) \times D_\alpha(t) \\ x_\beta(t+1) &= x_\beta(t) - A_\beta(t) \times D_\beta(t) \\ x_\delta(t+1) &= x_\delta(t) - A_\delta(t) \times D_\delta(t) \end{aligned} \quad (5)$$

$$x(t+1) = \frac{x_\alpha(t+1) + x_\beta(t+1) + x_\delta(t+1)}{3} \quad (6)$$

where $C_\alpha(t)$, $C_\beta(t)$, and $C_\delta(t)$ are calculated using Eq. (2); $x_\alpha(t+1)$, $x_\beta(t+1)$ and $x_\delta(t+1)$ are the first three optimal solutions at $(t+1)$-th iteration; and $A_\alpha(t)$, $A_\beta(t)$ and $A_\delta(t)$ can be obtained from Eq. (2).

3) Attacking the prey: When the prey stops moving, the wolves stop hunting. Otherwise, the wolf begins to attack. This stage is directly related to the convergence factor $a(t)$ in Eq. (3), which decreases linearly from 2 to 0 as the number of iterations increasing. As $a(t)$ decreases, each grey wolf updates its position by moving between its current location and the prey. The algorithm starts by randomly placing the wolves in the search space and evaluating their fitness using a given function. In each iteration, the top three wolves (x_α, x_β, and x_δ), are selected to guide the rest in the main steps: encircling, hunting, and attacking the prey. This iterative process continues until the maximum number of iterations, $t_{\max}$, is reached. The average of the three best solution is considered to be the final solution.

The conventional GWO algorithm has a small number of parameters in the search process and is easy to implement. But the global search ability of the algorithm is weak, resulting in falling into a local optimum in some cases easily. Recent studies have also examined their potential in various research areas [25], [26], including notable improvements in metaheuristic optimization algorithms, showing that LLMs can imitate mutation and crossover operations from evolutionary algorithms through careful tuning and prompting [27]. Despite these promising developments, the application of LLMs for designing metaheuristic optimization algorithms remains in its infancy. The main reason that there often depend on frequent interactions with LLMs during optimization, which can be expensive and limit scalability. While single-objective optimization has been studied, the use of LLMs in multi-objective evolutionary optimization is still largely unexplored. This gap motivates our work on a more efficient LLM-based GWO framework for near-field ISAC systems. Our goal is to harness the strengths of LLMs to overcome current challenges and expand the possibilities for multi-objective optimization in complex environments.

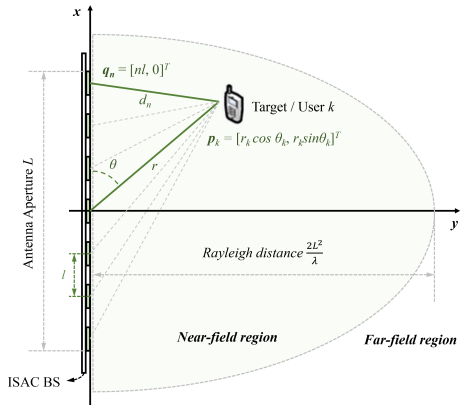


Fig. 1. Illustration of the near-field ISAC system.

III. System Model and Problem Formulation

The near-field ISAC system consists of a BS equipped with a uniform linear array (ULA) containing N antennas. The system communicates with K single-antenna entities, which could either be communication users or sensing targets. These entities are collectively represented by $\mathcal{K} = \{1, 2, ..., K\}$. As shown in Fig. 1, we consider a communication user or sensing target located at a distance of r and an angle of θ from the center of the ULA. Thus, the coordinate of user/target k is given by

$$\boldsymbol{p}_k = [r_k \cos\theta_k, r_k \sin\theta_k]^T. \quad (7)$$

However, due to the lack of spatial aperture in the vertical dimension, a conventional ULA cannot resolve elevation angles, which is a well-known limitation in array signal processing. Accurate elevation angle estimation would require a uniform planar array (UPA) or a more general three-dimensional array structure. Extending the system model to such geometries would increase the dimensionality and complexity of the channel and sensing models and is beyond the scope of the present work. The primary objective of this paper is

to investigate the integration of LLM-assisted multi-objective optimization with near-field ISAC system design, rather than to exhaustively study all possible array geometries. Nevertheless, we note that the proposed LLM-GWO framework is not inherently restricted to ULAs and can be readily extended to UPA or 3D array configurations by redefining the channel model and optimization variables. To this end, assuming the ULA is positioned parallel to the *x*-axis, with the *n*-th antenna element's coordinate specified as:

$$\boldsymbol{q}_n = (0, (n-1)l), \quad n = 1, ..., N \tag{8}$$

where l is the spacing between adjacent antennas. To determine the spatial relationship between an antenna element and a user/target, the distance between the *n*-th antenna element and the user/target *k* is given by:

$$\begin{aligned} d_n(r_k, \theta_k) &= \|\boldsymbol{p}_k - \boldsymbol{q}_n\| \\ &= \sqrt{r_k^2 + ((n-1)l)^2 - 2r_k(n-1)l\cos\theta_k}. \end{aligned} \tag{9}$$

According to [28], a critical aspect of the system is the classification of spatial regions into near-field and far-field zones, which depends on the distance between the BS and the user/target. The boundary between these zones is defined as the Rayleigh distance:

$$d = \frac{2L^2}{\lambda} = \frac{2(N-1)^2 l^2}{\lambda}, \tag{10}$$

where $L = (N-1)l$ is the aperture of the antenna array. N is the number of antennas, $l = \frac{\lambda}{2}$ is the antenna spacing with λ denoting for the carrier wavelength.

In general, when the transmission distance surpasses the Rayleigh distance, the system operates in the far-field region. In this case, electromagnetic waves exhibit a planar wavefront, and the beam directionality is focused on a single angle. However, when the transmission distance is shorter than the Rayleigh distance, the system transitions into the near-field region. This is characterized by a spherical wavefront, where the beam focuses on a localized area, providing enhanced precision in sensing and communication. For example., in a 28 GHz communication system, the Fraunhofer distance approximates the Rayleigh distance, varies with the number of antenna elements. Users located within a transmission range of a few dozen meters may be classified in the near-field region, particularly for systems with larger antenna arrays. It follows from the above that the electromagnetic wave characteristics differ significantly across the near-field and far-field regions. In the far-field, the beam directs energy uniformly in one direction, making it effective for long-distance communication. Conversely, the spherical wavefront enables focused energy in a small area that enhancing sensing accuracy in the near-field. This ability to adapt beamforming between planar and spherical wavefronts is pivotal in designing efficient ISAC systems.

A. Channel Model

Inspired by [22], the distance between ISAC BS and users depend on the Rayleigh distance characteristics, and there are two types of channels for near-field and far-field communications [28], [29]. We assume that different users transmit mutually orthogonal pilot sequences to the ISAC BS, enabling independent channel estimation for each user [30]. In the Fresnel region of the near-field, specifically within the range $1.2L \leq r \leq \frac{2L^2}{\lambda}$, it is worth noting that the channel gain of each link between the antenna elements and the user or target remains approximately identical [31]. The dominant near-field effect is phase curvature, whereas amplitude variation remains minimal for practical array sizes. The constant-gain approximation preserves the essential near-field characteristics while maintaining model tractability [32]–[34]. Building on this property, the channel between user *k* and the *n*-th antenna element can be expressed as:

$$h_n(r_k, \theta_k) = \hat{\rho}_k e^{-\frac{2j\pi d_n(r_k,\theta_k)}{\lambda}} = \rho_k e^{-\frac{2j\pi(d_n(r_k,\theta_k)-r_k)}{\lambda}}, \tag{11}$$

where $\rho_k = \hat{\rho}_k e^{-\frac{2j\pi r_k}{\lambda}}$ and $\hat{\rho}_k$ represents the free-space path loss of user/target *k*.

1) Far-field communications: A classical approach assumes that the distance between the user and the ULA is significantly larger than the length of the ULA. Under this assumption, the curvature difference of the spherical wavefronts across different antennas becomes negligible [35], [36]. Therefore, the distance difference from the user/target *k* to the reference antenna and the *n*-th antenna can be approximated as:

$$\Delta d = d_n(r_k, \theta_k) - r_k = (n-1)l\cos\theta_k. \tag{12}$$

Then, the far-field channel vector $\boldsymbol{h}(r_k, \theta_k) \in \mathbb{C}^{N\times 1}$ between the ISAC BS and the user or target *k* can be formulated as follow:

$$\begin{aligned} \boldsymbol{h}(r_k, \theta_k) &= [h_1(r_k, \theta_k), ..., h_N(r_k, \theta_k)]^T \\ &= \hat{\rho}_k e^{-\frac{2j\pi d_0}{\lambda}} \boldsymbol{a}(r_k, \theta_k), \end{aligned} \tag{13}$$

where $\boldsymbol{a}(r_k, \theta_k) \in \mathbb{C}^{N\times 1}$ is the far-field array response vector of user/target *k*, which can be written as

$$\boldsymbol{a}(r_k, \theta_k) = \frac{1}{\sqrt{N}}\left[1, e^{\frac{2j\pi l\cos\theta_k}{\lambda}}, ..., e^{\frac{2j\pi(N-1)l\cos\theta_k}{\lambda}}\right]. \tag{14}$$

2) Near-field communications: When the distance between user/target k and ISAC BS is smaller than the Rayleigh distance ($d_n(r_k, \theta_k) < L$), the wavefront is described with the spherical model. Thus, the distance (9) can be rewritten as

$$d_n(r_k, \theta_k) = \sqrt{r_k^2\left(1 + \frac{((n-1)l)^2}{r_k^2} - \frac{2(n-1)l\cos\theta_k}{r_k}\right)} \tag{15}$$

By applying the Fresnel approximation $\sqrt{1+x} \approx 1 + \frac{1}{2}x$, the distance between the *n*-th antenna and the user/target *k* is reduced to

$$\begin{aligned} d_n(r_k, \theta_k) &\approx r_k + \frac{1}{2}\left(\frac{((n-1)l)^2}{r_k} - 2(n-1)d\cos\theta_k\right) \\ &= r_k + \frac{\mu}{r_k} - \psi\cos\theta_k, \end{aligned} \tag{16}$$

where $\mu = \frac{((n-1)l)^2}{2}$ and $\psi = (n-1)d$. Since $l \ll r_k$ and the quadratic term is several orders of magnitude smaller than r_k, the resulting path-loss variation $\frac{1}{d_n^\alpha}$ is negligible across antenna

elements. This directly supports the approximation that the per-element channel magnitude remains nearly identical. Then, the near-field array response vector is

$$\boldsymbol{a}(r_k,\theta_k)=\frac{1}{\sqrt{N}}\left[...,e^{\frac{2\pi j}{\lambda}\left(\frac{\mu}{r_k}-\psi\cos\theta_k\right)},...\right]. \tag{17}$$

Unlike communication channel modeling, monostatic target sensing relies on the echo signal received at the BS. This process is characterized by the round-trip channel matrix $\boldsymbol{Q}\in\mathbb{C}^{N\times N}$, which plays a key role in sensing channel modeling. Let r_s, θ_s, and ρ_s represent the distance, angle, and complex channel gain of the sensing target, respectively. Based on these parameters, the near-field round-trip channel matrix for target s is given by:

$$\boldsymbol{Q}(r_s,\theta_s)=\rho_s\boldsymbol{a}(r_s,\theta_s)\boldsymbol{a}^T(r_s,\theta_s). \tag{18}$$

In view of the above-mentioned analysis, near-field modeling depends on antenna positions, while far-field modeling focuses on angles of arrival. The decoding vector for receiving the signal from user *k* or target *s* is formulated as $f_k=\boldsymbol{h}_k/\|\boldsymbol{h}_k\|$ with $\boldsymbol{h}_k$ being $\boldsymbol{h}(r_k,\theta_k)$ or $\boldsymbol{Q}(r_s,\theta_s)$ depending on the distance between user *k* or target *s* and ISAC BS.

B. Problem Formulation

In this subsection, we focus on the near-field network configuration, a coherent time block of length T is employed at the start of each coherent time block. Then, the communication channels are estimated using exsiting channel estimation methods. Following this initialization, the ISAC BS utilizes the remaining portion of the coherent time block to perform joint communication and sensing. To achieve this, the ISAC BS transmits a carefully designed joint communication and sensing signal:

$$\boldsymbol{x}=\sum_{k\in\mathcal{K}}\boldsymbol{f}_k z_k+\boldsymbol{s}, \tag{19}$$

where $\boldsymbol{f}_k\in\mathbb{C}^{N\times 1}$ denotes the beamforming vector used to transmit the information symbol $z_k\in\mathbb{C}$ to the user *k*. Additionally, a dedicated sensing signal $\boldsymbol{s}\in\mathbb{C}^{N\times 1}$ is integrated into the system to maximize the sensing degrees of freedom.

To ensure effective signal transmission and sensing, specific assumptions are made regarding the statistical properties of the transmitted symbols. The information symbols z_k, $k\in\mathcal{K}$ are independently distributed and normalized to unit power, satisfying $\mathbb{E}[z_k z_i]=1$, if $k=i$; and $\mathbb{E}[z_k z_i]=0$, otherwise. For the sensing signal $\boldsymbol{s}$, the statistical behavior is described by the covariance matrix $\boldsymbol{R}_s=\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H]$. The covariance matrix of the sensing signal provides a foundation for optimizing the joint communication and sensing performance. Then, the covariance matrix of the sensing signal is given by

$$\boldsymbol{R}_x=\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^H]=\sum_{k\in\mathcal{K}}\boldsymbol{f}_k\boldsymbol{f}_k^H+\boldsymbol{R}_s. \tag{20}$$

1) Communication Model: The communication signal received by user *k* is modeled as

$$\begin{aligned}\mathrm{y}_k=&\underbrace{\boldsymbol{h}_k^T(r_k,\theta_k)\boldsymbol{f}_k z_k}_{\textit{desired signal}}+\sum_{i\neq k}\boldsymbol{h}_k^T(r_k,\theta_k)\boldsymbol{f}_i z_i\\&+\boldsymbol{h}^T(r_k,\theta_k)\boldsymbol{s}+n_k,\end{aligned} \tag{21}$$

where $n_k\sim\mathcal{CN}(0,\sigma_k^2)$ denotes additive white Gaussian noise (AWGN). The first term represents the desired signal, while the remaining terms include interference from other users, the sensing signal and noise, respectively.

The achievable communication rate for all users is quantified as:

$$\begin{aligned}&\mathcal{R}^{com}(\boldsymbol{h}(r_k,\theta_k),\boldsymbol{f}_k,\boldsymbol{R}_x)\\&=\sum_{k\in\mathcal{K}}\log_2\left(1+\mathrm{SINR}(r_k,\theta_k)\right).\end{aligned} \tag{22}$$

where $\mathrm{SINR}(r_k,\theta_k)=\frac{|\boldsymbol{h}^T(r_k,\theta_k)\boldsymbol{f}_k|^2}{\boldsymbol{h}^T(r_k,\theta_k)\boldsymbol{R}_x\boldsymbol{h}^*(r_k,\theta_k)+|\boldsymbol{h}^T(r_k,\theta_k)\boldsymbol{f}_k|^2+\sigma_k^2}$.

2) Sensing Model: The sensing process involves receiving echo signals at the ISAC BS, which can be expressed as

$$\mathbf{y}_s=\boldsymbol{Q}(r_s,\theta_s)\boldsymbol{x}+\boldsymbol{w}_s \tag{23}$$

where $\boldsymbol{w}_s\in\mathcal{CN}(\mathbf{0}_N,\sigma_s^2\mathbf{I}_N)$ is the AWGN.

According to [31], the joint estimation of distance and angle relies on the near-field sensing channel $\boldsymbol{Q}(r_s,\theta_s)$. The sensing performance can be estimated with the mean squared error (MSE) metrics: $\zeta_{r_s}^2=\mathbb{E}[|r_s-\hat{r}_s|]$ and $\zeta_{\theta_s}^2=\mathbb{E}[|\theta_s-\hat{\theta}_s|^2]$. However, due to the difficulty of deriving closed-form expressions for these MSEs, the Cramr-Rao Bound (CRB) is used as an alternative metric. The inverse of the Fisher Information Matrix (FIM) is employed to CRB, which are calculated as

$$CRB(\boldsymbol{Q}(r_s,\theta_s))=\mathbf{J}_\zeta^{-1} \tag{24}$$

where $\mathbf{J}_\zeta$ is the FIM that can be calculated based on the derivatives of the received signal $\mathbf{y}_s$ with respect to the parameters $\zeta_i\in\{r_s,\theta_s\}$. Using the general FIM definition, $\mathbf{J}_\zeta$ can be partitioned into submatrices based on the parameters being estimated:

$$\mathbf{J}_\zeta=\begin{bmatrix}\mathbf{J}_{11}&\mathbf{J}_{12}\\\mathbf{J}_{12}^T&\mathbf{J}_{22}\end{bmatrix}, \tag{25}$$

where $\mathbf{J}_{11}$ captures the Fisher information for the distance and angle parameters $\{r_s,\theta_s\}$:

$$\mathbf{J}_{11}=\frac{2|\rho_s|^2N}{\sigma_s^2}Re\left\{\mathrm{tr}\left(\frac{\partial\boldsymbol{Q}}{\partial r_s}\boldsymbol{R}_x\frac{\partial\boldsymbol{Q}^H}{\partial r_s}\right)\right\} \tag{26}$$

and $\mathbf{J}_{12}$ represents the cross-correlation between $\{r_s,\theta_s\}$ and ρ_s, that is

$$\mathbf{J}_{12}=\frac{2N}{\sigma_s^2}Re\left\{\rho_s\mathrm{tr}\left(\frac{\partial\boldsymbol{Q}}{\partial r_s}\boldsymbol{R}_x\boldsymbol{Q}^H\right)\right\} \tag{27}$$

and $\mathbf{J}_{22}$ corresponds to the Fisher information for the complex channel gain

$$\mathbf{J}_{22}=\frac{2N}{\sigma_s^2}Re\left\{\mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{R}_x\boldsymbol{Q}^H\right)\right\} \tag{28}$$

and $\boldsymbol{R}_x=\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^H]$ is the signal covariance matrix.

By taking the expressions for $\mathbf{J}_{11}$, $\mathbf{J}_{12}$ and $\mathbf{J}_{22}$, the

$CRB(\boldsymbol{Q}(r_s,\theta_s))$ is provided as

$$CRB(\boldsymbol{Q}(r_s,\theta_s)) = \left(\mathbf{J}_{11} - \mathbf{J}_{12}\mathbf{J}_{22}^{-1}\mathbf{J}_{12}^{T}\right)^{-1}. \tag{29}$$

Based on the settings mentioned above, we have $\zeta_{r_s}^2 \geq [CRB]_{1,1}$ and $\zeta_{\theta_s}^2 \geq [CRB]_{2,2}$, where the submatrices $\mathbf{J}_{11}$, $\mathbf{J}_{12}$, and $\mathbf{J}_{22}$ are determined through detailed matrix operations. To evaluate near-field sensing capacity, the log sum of the CRBs is employed as the objective function, which provide a scalar measure of the overall estimation performance across parameters r_s and θ_s. Since CRB values can vary widely, taking the logarithm reduces the dynamic range, preserving the relative highlighting reductions in high-error components, especially when using LLM-based sampling techniques. Thus, the sensing object function is expressed as

$$\mathcal{R}^{rad}\big(\boldsymbol{Q}(r_s,\theta_s),\boldsymbol{R}_x\big) = \log_2\left(\mathrm{tr}\big(CRB(\boldsymbol{Q}(r_s,\theta_s),\boldsymbol{R}_x)\big)\right). \tag{30}$$

It follows that the utility of the network is maximized while minimizing the log sum of CRBs by jointly optimizing parameters r_s and θ_s, $\forall s \in \mathcal{K}$. Therefore, the optimization problem can be formulated as

$$\min_{\substack{\boldsymbol{h}(r_k,\theta_k),\\ \boldsymbol{f}_k,\boldsymbol{R}_x,\\ \boldsymbol{Q}(r_s,\theta_s)}} \left[-\mathcal{R}^{com}(\boldsymbol{h}(r_k,\theta_k),\boldsymbol{f}_k,\boldsymbol{R}_x), \mathcal{R}^{rad}(\boldsymbol{Q}(r_s,\theta_s),\boldsymbol{R}_x)\right] \tag{31a}$$

$$s.t.\ \mathcal{R}^{com}(\boldsymbol{h}(r_k,\theta_k),\boldsymbol{f}_k,\boldsymbol{R}_x) \geq R_{\min,k} \tag{31b}$$

$$0 \leq \mathrm{tr}(\boldsymbol{R}_x) \leq P_{\max} \tag{31c}$$

$$\theta_k,\theta_s \in \mathcal{P}, \quad \forall k,\ s \in \mathcal{K} \tag{31d}$$

$$\boldsymbol{R}_x \succ \sum_{k\in\mathcal{K}} \boldsymbol{f}_k\boldsymbol{f}_k^H \tag{31e}$$

where $R_{\min,k}$ denotes the minimum rate requirement of user k and $\mathrm{tr}(\boldsymbol{R}_x)$ denotes the transmit power of ISAC BS, which should be no less than 0; $P_{\max}$ denotes the maximum transmit power. (31d) denotes the constraints of user/target angle, which are limited in the range of $\mathcal{P}$. The last constraint (31e) is from that the covariance matrix of the sensing signal, $\boldsymbol{R}_s$, which is derived from the relationship:

$$\boldsymbol{R}_s = \boldsymbol{R}_x - \sum_{k\in\mathcal{K}} \boldsymbol{f}_k\boldsymbol{f}_k^H \tag{32}$$

where $\boldsymbol{f}_k$ denotes the beamforming vector for user k. The sensing signals are appropriately designed without compromising the performance of communication signals. The condition $\boldsymbol{R}_s \succ 0$ guarantees a positive semidefinite covariance matrix, which is necessary for the stability and feasibility of the sensing process.

In practical near-field ISAC systems, communication objective $\mathcal{R}^{com}(\boldsymbol{h}(r_k,\theta_k),\boldsymbol{f}_k,\boldsymbol{R}_x)$ and sensing objective $\mathcal{R}^{rad}(\boldsymbol{Q}(r_s,\theta_s),\boldsymbol{R}_x)$ often exhibit a trade-off due to shared system resources; however, the existence and strength of this trade-off depend on the characteristics of the underlying Pareto boundary. In formulated problem (31), it is difficult to optimize the physical channel realizations due to the nature of the channel stochastic and environment-dependent. Instead, we optimize channel-related parameters by configuring the positions of communication users and sensing targets, which determine the angle of arrival and distance. This abstraction enables learning spatial configurations to jointly enhance communication and sensing, which is also a core of our LLM-based framework.

IV. Proposed LLM-GWO-Based Optimization Framework

The multi-objective optimization problem formulated in (31) is challenging to solve using conventional convex optimization techniques. This difficulty does *not* arise merely from the coupling of decision variables, but rather from several intrinsic sources of non-convexity embedded in the problem structure. Specifically, the objective functions for communication and sensing are intertwined through parameters such as the distance and angle of users, as well as the precoding vectors for communication and sensing signals. These coupled variables create a non-linear dependency that results in a non-convex feasible region, particularly when these variables are optimized jointly in the context of the near-field ISAC system. Moreover, the optimization problem is formulated as a multi-objective problem, which inherently involves conflicting objectives that further contribute to the non-convexity. The non-convexity is particularly apparent in the interaction between the communication rate and sensing accuracy, where optimizing one often compromises the other. This complexity is addressed using the LLM-GWO approach, which decomposes the original non-convex problem into smaller sub-problems that can be solved more effectively. In this section, we first clarify the origin of non-convexity and then present the proposed LLM-assisted GWO framework for efficiently solving the problem.

A. Problem Reformulation via Decomposition

To enable tractable optimization, we reformulate problem (31) in a compact vector form as

$$\begin{aligned} \mathrm{P1}: \quad & \min_{\boldsymbol{x}} \mathcal{G}(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left\{-\mathcal{R}^{com}(\boldsymbol{x}), \mathcal{R}^{rad}(\boldsymbol{x})\right\} \\ & s.t. \quad \boldsymbol{x} \in \mathcal{C} \end{aligned} \tag{33}$$

where $\boldsymbol{x} = [r_1, r_2, ..., r_K, \theta_1, \theta_2, ..., \theta_K, f_1, f_2, ..., f_K]$ represents all optimization variables, $\mathcal{C}$ denotes the decision space, and $\mathcal{F}_a(\boldsymbol{x}) = \mathcal{R}^{com}(\boldsymbol{h}(r_k,\theta_k),\boldsymbol{f}_k,\boldsymbol{R}_x)$, $\mathcal{F}_b(\boldsymbol{x}) = \mathcal{R}^{rad}(\boldsymbol{Q}(r_s,\theta_s),\boldsymbol{R}_x)$. The proposed LLM-GWO problem involves decomposing (P1) into a series of single-objective sub-problems based on the weighted Tchebycheff method, in which the sub-problem is expressed as

$$\begin{aligned} & \min_{\boldsymbol{x}} \mathcal{U}(\boldsymbol{x}|\boldsymbol{\omega},\boldsymbol{\vartheta}) = \max_{i\in\{a,b\}} \{\omega_i(\mathcal{F}_i(\boldsymbol{x}) - \vartheta_i^*)\} \\ & s.t. \quad \boldsymbol{x} \in \mathcal{C} \end{aligned} \tag{34}$$

where $\boldsymbol{\omega} = [\omega_a, \omega_b]$ denotes the weight vector satisfying $0 < \omega_a, \omega_b < 1$, $\omega_a + \omega_b = 1$. The current minimum objective function values is defined as $\vartheta_i^* = \min\{\mathcal{F}_i(\boldsymbol{x})\}$ for $i = a, b$. This distributed weight vectors is provided by Das and Dennis [37], which is particularly suitable for controlling the trade-off between objectives in the proposed optimization setting. The weighted Tchebycheff method guarantees a fundamental relationship between Pareto-optimal solutions and specific weights [38]. For each Pareto-optimal point $\boldsymbol{x}^*$, there exists a corresponding weight vector $\boldsymbol{\omega}^*$ to ensure $\boldsymbol{x}^*$ as the optimal

solution of the weighted Tchebycheff formulation in (34). This dual relationship highlights the equivalence between solving the Tchebycheff formulation with appropriate weights and the Pareto-optimal solution.

For conventional GWO algorithm, the gray wolf population gradually converges around the optimal individual as the number of iterations increases, resulting in a decline in population diversity. This makes it difficult to avoid local optimal solution when the optimal individual becomes trapped, which result in premature termination and reduced accuracy in finding the optimal solution. Inspired by LLM [21], the LLM-driven search operator is integrated into the framework of MOO problem (34), whose main function is to effectively expand the search space and improve the global search ability of the algorithm. the LLM does not provide the objective function to GWO; rather, it assists by guiding the search parameters for each sub-problem. It is worth noting that this comparativeness between the LLM and GWO enhances the adaptability and efficiency of the optimization process, allowing the framework to solve complex multi-objective problems effectively. The steps are as follows.

1) Initialization: The initialized weight variables with a population size of M is selected, i.e., $\boldsymbol{\omega}_j, j = 1, ..., M$. LLMs adapt to tasks through in-context learning (ICL) by leveraging input examples; however, the performance heavily depends on the selection of diverse and high-quality inputs. Inspired by the principle of lens imaging, we introduce the reverse learning strategy that including both the solution and its opposite. By combining the reverse learning strategy with LLM search, the bidirectional search is considered to expand the search scope and improve the search diversity of ICL.

As shown in Fig. 2, the reverse learning strategy with the principle of lens imaging is illustrated. The reverse point of the individual ϑ can be calculated with the principle of lens imaging:

$$\frac{(a+b)/2 - \vartheta}{\vartheta^* - (a+b)/2} = \frac{h}{h^*}. \tag{35}$$

where ϑ^* denotes the reverse point of the individual ϑ.

If $\kappa = h/h^*$, then Eq. (35) is extend to the i-dimensional optimization problem, which can be rewritten as:

$$\vartheta_i^* = \frac{\vartheta_{\min} + \vartheta_{\max}}{2} + \frac{\vartheta_{\min} + \vartheta_{\max}}{2\kappa} - \frac{\vartheta_i}{\kappa} \tag{36}$$

where $\vartheta_{\min}$ and $\vartheta_{\max}$ denote the minimum and maximum values of the variables across all initial solutions. ϑ_i^* is the lens inverse solution of ϑ_i. The scaling factor κ of the lens is important for the learning performance of lens imaging. A smaller κ produces a wider range of inverse solutions, while a larger κ results in a narrower range, supporting more focused local searches.

However, this approach becomes ineffective when the lead wolves are trapped in a local minimum or are far from the optimal solution. To address this issue, a probabilistic strategy is exploited by selecting m inputs for the j-th subproblem: with probability P from a local neighborhood u^j, and with probability $1-P$ from the global population $\mathbf{\Phi}$. This balances local precision and global diversity, capturing specific patterns while avoiding overfitting. This expands the search scope

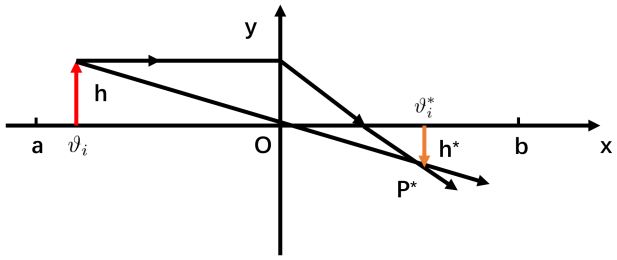


Fig. 2. Schematic diagram of lens imagingreverse learning.

and enables a bidirectional search for the optimal solution. Combining the reverse learning strategy with LLM improves the GWO algorithm's search performance, which can solve complex tasks effectively by leveraging both localized and global information. Thus, the original problem (33) is decomposed into M single-objective sub-problem based on GWO strategy, in which the j-th subproblem associated with $\boldsymbol{\omega}^j$ is given by

$$\mathcal{U}(\boldsymbol{x}^j|\boldsymbol{\omega}^j, \boldsymbol{\vartheta}^j) = \max_{i\in\{a,b\}} \left\{\omega_i^j(\mathcal{F}_i(\boldsymbol{x}) - \vartheta_i)\right\}. \tag{37}$$

Accordingly, an initial population of size M is uniformly and randomly generated from set $\mathcal{C}$, i.e., $\boldsymbol{x}^j \in \mathcal{C}$, $j = 1, ..., N$, where $\boldsymbol{x}^j$ is the initial solution for the j-th subproblem. The function $\mathcal{U}(\boldsymbol{x}^j|\boldsymbol{\omega}^j, \boldsymbol{\vartheta}^j)$ is continuous with respect to $\boldsymbol{\omega}^j$, which which implies that the optimal solutions of two subproblems j and j' will be close if their weight vectors $\boldsymbol{\omega}^j$ and $\boldsymbol{\omega}^{j'}$ are close in terms of Euclidean distance. Therefore, the neighbor set of $\boldsymbol{\omega}^j$ is defined as the set of indices of the S weight vectors nearest to $\boldsymbol{\omega}^j$, denoted as $\mathcal{N}_j = \{j_1, ..., j_S\}$, which can be expressed as

$$\boldsymbol{\omega}^j = 2 \times a(t) \times r_2, \tag{38}$$

where $a(t)$ can be given by (3). The difference is that the size of the selected individuals can scale flexibly due to the adaptable interaction with LLMs. This scalability enables the method to accommodate varying problem complexities or optimization requirements dynamically. After individuals are selected, there are ranked in descending order based on the evaluation metrics. Then, the top three candidate solutions (i.e. $\boldsymbol{x}_\alpha$, $\boldsymbol{x}_\beta$ and $\boldsymbol{x}_\delta$) are chosen. These candidates are subsequently assessed using the aggregation function corresponding to the j-th subproblem, ensuring an efficient evaluation process.

2) Evolution: The problem (33) is reformulated as a series of scalarized objective functions. Within the transformed framework, a LLM is employed as a black-box operator, performing two key tasks: *encircling* and *hunting*. These operations are executed through prompt engineering to generate new candidate solutions.

To design effective prompts, we incorporate the following three types of information into prompt engineering: problem description, in-context examples and task instructions. More specifically, for the in-context examples of the j-th subproblem at each generation, the LLM receives d selected parent solutions along with their objective values. These examples are partly drawn from $\mathcal{N}_j$ and partly from the entire population.

Algorithm 1 Proposed LLM-MOEA algorithm

Input The population size: N. The neighbor size: S. Parents number: d; Iteration number: $N_{\max}$; The number of points generated by LLM: n_o. .
Initialize: set $\mathbf{\Phi} = \varnothing$. $\boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^M$ based on Das and Dennis method.
Calculate the d the M closest weight variables to each weight variable $\omega_j, j = 1, ..., M$.
Update the neighbor set $\mathcal{N}_j$.
Generate an initial population $x_j, j = 1, ..., M$ and compute $\mathcal{U}(\boldsymbol{x}|\boldsymbol{\omega}, \boldsymbol{\vartheta})$.
Reference point generated by $\boldsymbol{\vartheta}^*$.
while $N_{\max}$ not reached **do**
    $i = i + 1$;
  Evolution
  for$j = 1, 2, ..., M$ **do**;
    Selection:
      Parent solution points d selected by partly from $\mathcal{N}_j$ with a probability of P and partly from the population with a probability of $1 - P$.
    Reproduction via the LLM:
      a) Generate textual *Prompt* for LLM given the subproblem $\mathcal{U}^j(\boldsymbol{x}|\boldsymbol{\omega}, \boldsymbol{\vartheta})$.
      b) Using LLM, a number of offspring points $\{\boldsymbol{x}'_1, ..., \boldsymbol{x}'_{n_o}\}$ is generated with given the instruction prompt.
    Update:
      a) update the best candidate solution $\boldsymbol{x}'$ from (40) and based on $\mathbf{\Phi}$.
      b) update variables $\boldsymbol{\vartheta}^*$, $\{\boldsymbol{x}^{j_1}, ..., \boldsymbol{x}^{j_M}\}$ and $\mathbf{\Phi}$
Output: $\mathbf{\Phi}$

Regarding constraint (31c), the LLM is instructed to strictly avoid generating any infeasible points. For constraint (31c), the LLM is asked to strictly output unexpected point. Then, we introduce the transmit powers as follows

$$\mathrm{tr}(\boldsymbol{R}_x) = \mathrm{tr}\left(\sum_{k\in\mathcal{K}} \boldsymbol{f}_k \boldsymbol{f}_k^H\right) + \mathrm{tr}(\boldsymbol{R}_s) \tag{39}$$

where $\mathrm{tr}(\sum_{k\in\mathcal{K}} \boldsymbol{f}_k \boldsymbol{f}_k^H)$ and $\mathrm{tr}(\boldsymbol{R}_s) = \mathrm{tr}(\boldsymbol{s}\boldsymbol{s}^H)$ denote the communication power and sensing power of ISAC BS, respectively. The communication vector $\boldsymbol{f}_k$ and sensing vector $\boldsymbol{s}$ can be obtained from the results of LLM. Moreover, each variable is normalized in the prompt to ensure the generated points within the viable ranges. A detailed example of the prompt and the generated solution are provided in Algorithm 1.

Fig. 2 shows the overall design of the proposed LLM-GWO algorithm for the ISAC system in the near-field region. Specifically, for the number of iterations t, the convergence coefficient $a(t)$ is obtained according to (3) and the weight vector $\boldsymbol{\omega}^j$ is computed according to (38). Subsequently, with the addition of the initialized reference point ϑ, the neighbor set $\mathcal{N}_j$, and the population $\mathbf{\Phi}$, the subproblem $\mathcal{U}^j(x|\omega, \vartheta)$ and the parent points $x_i, \cdots, x_d$ are determined to generate the sample set for the textual prompt entered into LLM with task goal, definition and rules. The LLM generates a number of individuals based on the textual prompt, and the best 3 individuals $x_\alpha, x_\beta, x_\delta$ are selected to calculate the best candidate solutions x' according to (6). After that, ϑ, $\mathcal{N}_j$ and $\mathbf{\Phi}$ are updated based on x' and the process will continue to repeat according to the next iteration number $t+1$.

An example of a prompt is illustrated as follow:

Example Prompt:
-"As a research expert in near-field ISAC and metaheuristic algorithm."
-"You are given a multi-objective evolutionary problem to maximize the spectral efficiency of users and the localization Cramer-Rao of the sensing target. "
-"Now you will help me minimize (34)"
-"Object 1 represents the ISAC BS communication performance and object 2 represents the ISAC BS sensing performance."
-"Variable 1 and Variable 2 represent the distance and angle of near-field communication users, respectively; Variable 3 represents the transmitted precoding of BS; Variable 4 and Variable 5 represent the distance and angle of near-field sensing user, respectively. "
-"Please generate two points that are different from all points above, and not dominated by any of the above."
-"The points with their objective values as shown below."
point: <start>7.304 -2.12 6.058 8.926 6.256<end>
value: 9.515
point: <start>8.468 4.460 7.970 1.244 6.217<end>
value: 9.360
...
point: <start>3.022 -5.726 2.688 6.290 4.906<end>
value: 9.519
point: <start>3.098 -5.724 4.940 8.694 1.682<end>
value: 9.876
Please generate two points that are different from all points above, and new points requires targeting unexplored areas of the search space, ensuring they remain lower than previously identified points. Each output new point must start with <start> and end with <end>.

1) Description of task: The objective of this task is to identify more optimal solutions by minimizing the aggregation function value for the j-th subproblem. The goal is to enhance solution quality while maintaining computational efficiency.

2) In-context samples: To assist the LLM in generating new candidate solutions, a set of demonstration samples is provided. These samples include solutions from the current population and the corresponding fitness values. Each sample follows a specific format: the variables are marked with a starting symbol <start> and an ending symbol <end>. This structured presentation ensures clarity and consistency in guiding the LLM.

3) Expected outputs: The desired result is to generate $s = 2$ new individuals (points) that are distinct from the provided input points. These new points should adhere to a recognizable format, such as beginning with <start> and ending with <end>, to allow the LLE-GWO algorithm to parse them

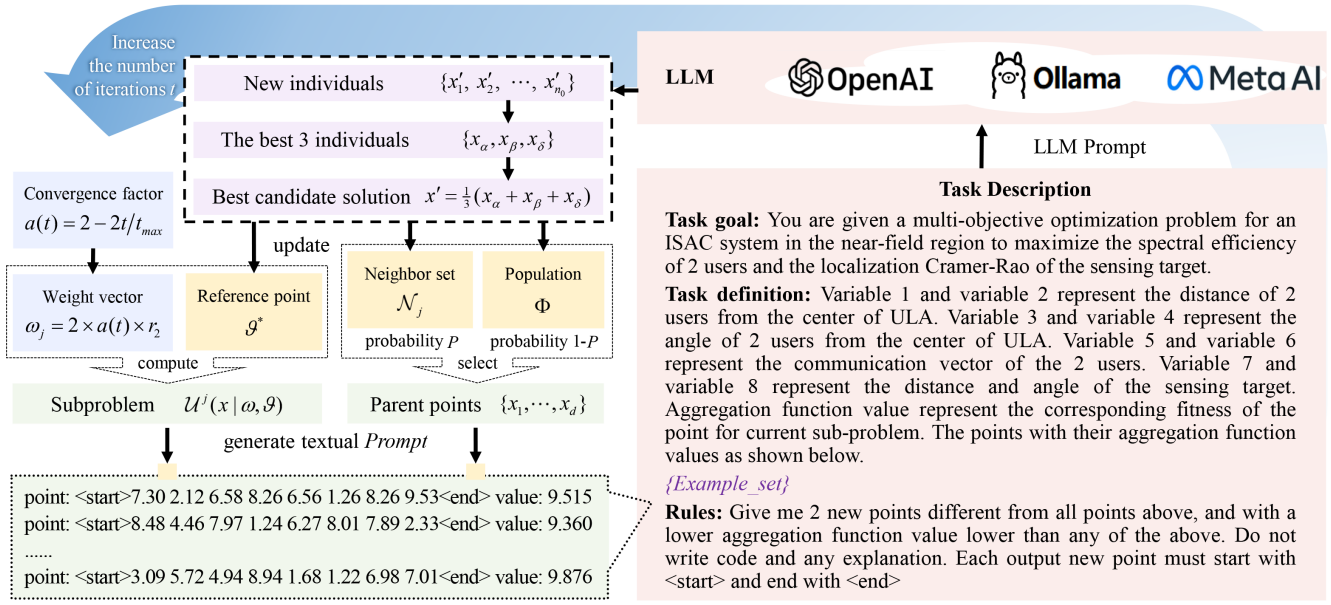


Fig. 3. A LLM-based GWO framework for near-field ISAC system.

effectively from the LLM's responses. To maintain stable interaction, restrictions are imposed to avoid redundancy, such as preventing the inclusion of code or unnecessary explanations.

4) Update: Within the search space, the best candidate solution, $\boldsymbol{x}_\alpha$, is considered the closest to the optimal point, followed sequentially by $\boldsymbol{x}_\beta$ and $\boldsymbol{x}_\delta$. To mathematically formalize the hunting behavior, all new output points $\boldsymbol{x}'$ generated by the LLM are sorted in descending order based on their performance or fitness values. This step ensures a focused search toward optimal solutions and aligns the algorithm's progression with the most promising candidates, and then, the best candidate solution $\boldsymbol{x}'$ is updated by adding up the mean values of the three best positions (i.e., $\boldsymbol{x}_\alpha$, $\boldsymbol{x}_\beta$ and $\boldsymbol{x}_\delta$) as (6), that is

$$\boldsymbol{x}' = \frac{\boldsymbol{x}_\alpha + \boldsymbol{x}_\beta + \boldsymbol{x}_\delta}{3} \tag{40}$$

For each output point $\boldsymbol{x}'$ generated from (40), the reference point $\boldsymbol{\vartheta}^*$ is updated as follows: for $i = 1, 2$, if $\vartheta_i^* > \mathcal{F}_i(\boldsymbol{x})$, then set $\vartheta_i^* = \mathcal{F}_i(\boldsymbol{x})$. Next, the neighboring population is updated. For each $j_i \in \mathcal{N}_j$, if $\mathcal{U}(\boldsymbol{x}'|\boldsymbol{\omega}^{j_i}, \boldsymbol{\vartheta}^*) \leq \mathcal{U}(\boldsymbol{x}^{j_i}|\boldsymbol{\omega}^{j_i}, \boldsymbol{\vartheta}^*)$, then update $\boldsymbol{x}^{j_i} = \boldsymbol{x}'$ and $\mathcal{G}(\boldsymbol{x}^{j_i}) = \mathcal{G}(\boldsymbol{x}')$. Finally, the evolutionary archive $\boldsymbol{\Phi}$ is updated by removing all solutions dominated by $\boldsymbol{x}'$. If no existing solution in $\boldsymbol{\Phi}$ dominates $\boldsymbol{x}'$, then add $\boldsymbol{x}'$ to $\boldsymbol{\Phi}$.

B. Structured Prompt Design for LLM-Assisted Operators

In Algorithm 1, the LLM is invoked as a context-aware reproduction operator through a structured prompt template. The prompt is designed to provide sufficient optimization context while maintaining a compact and consistent format, and consists of four key components:

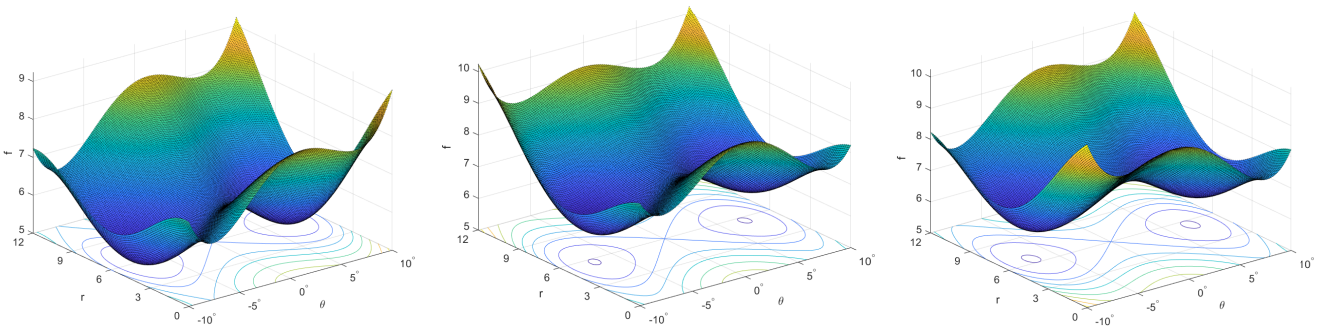
- **Subproblem specification:** A concise description of the current decomposed subproblem, including the associated weight vector $\boldsymbol{\omega}_j$ and the weighted Tchebycheff aggregation function.
- **Parent solution context:** A small set of representative parent solutions selected from the local neighborhood and the global population, together with their corresponding objective values. This enables the LLM to infer promising search directions from historical optimization trajectories.
- **Feasibility and constraint guidance:** Explicit constraints and feasibility requirements are included to discourage the generation of infeasible or dominated solutions.
- **Task-level instructions:** High-level instructions encouraging exploration of under-sampled regions and avoidance of premature convergence.

This structured prompt template is fixed across iterations, ensuring stable and reproducible behavior. The LLM does not perform online training or parameter updates; instead, it leverages in-context learning by conditioning its outputs on the supplied optimization history. This design allows the LLM to adapt its generation behavior across iterations while preserving the mathematical structure and interpretability of the underlying evolutionary process.

C. Convergence Behavior and Evaluation Guidelines

Since Algorithm 1 is a metaheuristic, LLM-assisted multi-objective evolutionary algorithm, deriving a closed-form theoretical convergence rate is generally intractable. Nevertheless, its convergence behavior can be systematically evaluated using well-established empirical criteria.

From a scalar optimization perspective, convergence can be examined by tracking the evolution of the weighted Tchebycheff aggregation function $U(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\vartheta})$ for each subproblem. In particular, convergence speed is characterized by the number of iterations required for the best or average aggregation value to stabilize within a prescribed tolerance, as well as by the decay trend of the aggregation value during early



(a) GWO (b) LEDMA (c) LLM-GWO

Fig. 4. 2-D version forms of the used real-valued benchmarking functions (r, θ, f).

iterations. From a multi-objective perspective, convergence is assessed through the stabilization of the external Pareto archive Φ. Typical indicators include the rate of newly generated non-dominated solutions and the diminishing variation of the Pareto-front shape over successive generations. When Φ exhibits negligible updates across multiple iterations, the algorithm is deemed to have converged.

The convergence dynamics of Algorithm 1 are further shaped by the GWO mechanism. The convergence factor $a(t)$ decreases linearly with the iteration index, enabling a gradual transition from global exploration to local exploitation, which promotes stable convergence in highly non-convex search spaces. Moreover, the LLM-assisted reproduction operator does not compromise convergence stability, as all generated candidates are strictly evaluated using the same aggregation function and feasibility constraints. Instead, the LLM primarily accelerates convergence in early and intermediate stages by producing higher-quality and more diverse candidate solutions, while the GWO update rules ensure reliable refinement in later stages. Therefore, these criteria provide a consistent and reproducible framework for analyzing and comparing the convergence behavior of Algorithm 1 with baseline approaches.

D. Complexity Analysis

The generation of weight vectors and neighborhood construction incurs a one-time cost of $\mathcal{O}(M^2)$, which can be computed offline. At each iteration, fitness evaluation for all subproblems requires $\mathcal{O}(MC_f)$ operations, where C_f denotes the cost of computing the communication and sensing objective functions. The GWO update steps involve simple vector operations and incur a complexity of $\mathcal{O}(MD)$ per iteration. The LLM-assisted reproduction introduces an additional cost of $\mathcal{O}(Mn_o)$ per iteration, where the LLM is treated as a black-box oracle with bounded invocation frequency. Updating the external Pareto archive Φ requires dominance checks with complexity $\mathcal{O}(M|\Phi|)$ per iteration, where $|\Phi|$ denotes the archive size. Therefore, the overall computational complexity of Algorithm 1 over T iterations can be summarized as

$$\mathcal{O}\big(M^2 + T\big(MC_f + MD + Mn_o + M|\Phi|\big)\big)\,.$$

This complexity is polynomial in the population size and problem dimension, and is comparable to conventional MOEA-based methods, with a controlled additional overhead introduced by the LLM-assisted operator.

V. Numerical Results

In this section, the prediction performance of the proposed is evaluated with the near-field communication setting of ISAC network. As similarly in [39], we consider the difference in the probability of near-field and far-field users maximize the SINR of the near-field communication objective function, while minimization near-field sensing performance. The magnitude of the RCS is assumed to be uniformly distributed in range in [0.8, 1]. The channel power at a reference distance of 1 meter is set to $\varrho_0 = -60\text{dB}$. Additionally, the ISAC BS operates with a bandwidth of B = 51.2 MHz, while the noise power is fixed at $\sigma^2 = -110\text{dBm}$. The transmission power limits for the ISAC BS are defined as $p_{\max} = 20\text{dBm}$ and $p_{\min} = 0\text{dBm}$, ensuring the system adheres to realistic operational constraints.

To verify the auxiliary role of the LLM in accelerating the convergence rate and enhancing the diversity of the population within the MOEA framework, we conducted a comparative analysis of existing LLM-enabled metaheuristic algorithm, named as LEDMA [40], GWO, MOEA/D [41] and our proposed LLM-GWO framework. Among them, LEDMA and GWO share the same algorithmic foundation as our framework, with the primary distinction lying in their reproductive strategies. These methods have proven to be effective for solving MOPs due to their ability to efficiently handle objectives, maintain solution diversity, and adapt to complex problems. However, there also face notable challenges, such as sensitivity to parameter settings, high computational intensity, and limitations in tackling high-dimensional problems. The effectiveness of these methods often depends on the specific problem, requiring a careful balance of trade-offs during implementation. For the proposed LLM-GWO algorithm, we integrate the GPT-4.0 model as the LLM search operator, the main reason is that it demonstrated capacity to perform reasoning, constraint-aware synthesis and multi-objective trade-off analysisłall of which are core to our formulation of LLM-based optimization. Additionally, GPT-4's reinforcement learning fine-tuning and instruction-following behavior make it well-suited to high-level control and policy generation in structured environments.

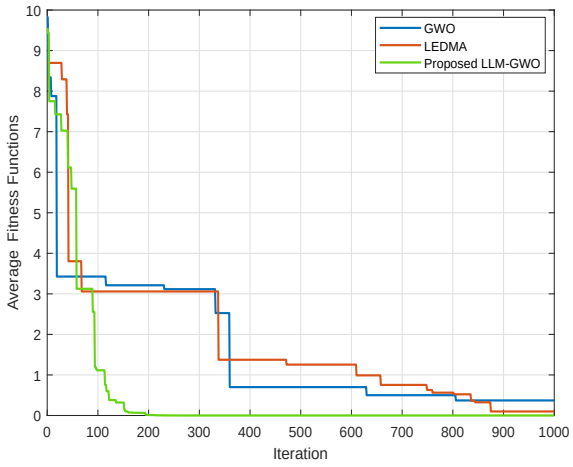


Fig. 5. The convergence curves for the average fitness functions.

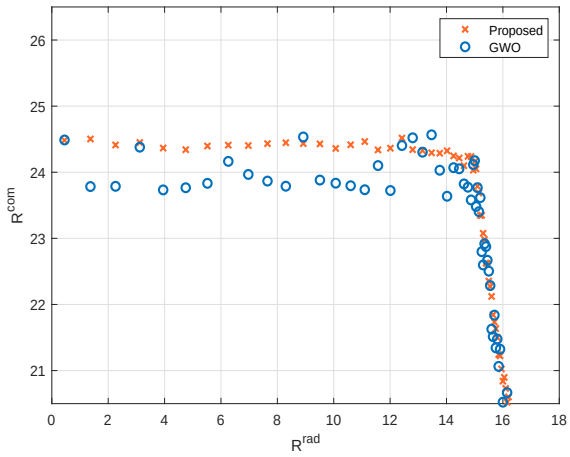


Fig. 6. Pareto-optimal points distribution achieved by existing GWO and LLM-GWO algorithms.

Fig. 4 compares the embeddings of samples with the highest and lowest Pareto front using three visualization methods: GWO, LEDMA and the proposed LLM-GWO. In the Fig. 4, the proposed LLM-GWO decision boundary clearly separates the two groups, showing that they occupy different areas in the embedding space. These findings indicate that Pareto front not only reflects the quality of jailbreak prompts but also strongly influences their representation in the embedding space.

Following the convergence-study guidelines outlined in Section IV, we evaluate the convergence behavior of the proposed Algorithm 1 from both scalar and multi-objective perspectives. Specifically, convergence speed is assessed by tracking the evolution of the average aggregation (fitness) value across iterations, while convergence quality is evaluated through the stabilization and distribution of the obtained Pareto fronts. Fig. 5 depicts the convergence curves of of our proposed LLM-GWO approach as well as two baseline algorithms with respect to the number of iterations. With the number of iterations increasing, the convergence performance of LEDMA scheme reduces more faster than that of GWO scheme. An interesting phenomenon is that with a small number of iterations, deploying the proposed LLM-GWO framework can converges rapidly to a stable, desired state, and it also has better fault tolerance and high stability. Besides, Fig. 5 demonstrates that the proposed LLM-GWO approach converges faster than the LEDMA and GWO scheme, and achieves the lowest average fitness function value of 0.007, while the LEDMA and GWO achieve average fitness function value of 0.267 and 0.029, respectively. This further demonstrates the superiority of our proposed LLM-GWO approach in terms of the average fitness function. As observed in Fig. 5, the proposed LLM-GWO algorithm converges faster than the baseline methods in terms of aggregation-value stabilization, which is consistent with the convergence guidelines discussed in Section IV.

Fig. 6 illustrates the relationship between the sensing rate $\mathcal{R}_{\min}^{\mathrm{rad}}$ for distance and angle estimation and the minimum communication rate $\mathcal{R}_{\min}^{\mathrm{com}}$. We observe that the Pareto boundary achieved by the proposed LLM-GWO outlined in (34). The curve connecting these two points signifies the Pareto boundary of the ISAC's rate region, thereby revealing a fundamental trade-off between sensing accuracy and communication performance. Despite this trend, it is worth emphasizing that the RCRBs remain at a relatively low level even when $\mathcal{R}_{\min}^{\mathrm{com}}$ is considerably high. This observation validates the effectiveness of the proposed LLM-GWO approach. Furthermore, although the proposed LLM-GWO scheme experiences a slight degradation in sensing performance at the same communication rate, it significantly reduces power consumption. This trade-off makes LLM-GWO a compelling solution in power-constrained scenarios.

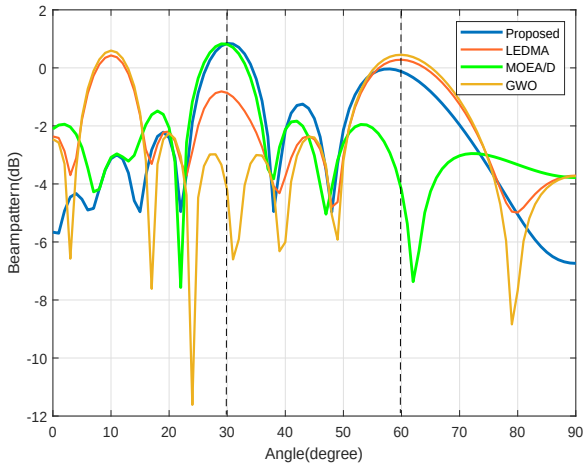


Fig. 7. Beampattern of radar functionality of LLM-GWO scheme.

Fig. 7 shows how Algorithm 1 enables radar functionality by steering transmit and receive beams toward specific directions, 30° and 60°. The algorithm focuses signals from multiple antennas to form a directional "beam" or "lobe" [42]. This beam is based on LLM-driven beam training, which enhances target detection and reduces interference. The transmit signal helps by directing energy toward the targets, while the receive beamforming is optimized to clearly capture the reflected signals. This demonstrates the beamforming gain achieved by Algorithm 1 in the far or near-field ISAC systems. As shown in the Fig. 7, the proposed LLM-GWO approach provides higher power at the target locations than the LEDMA and

GWO-based methods. These algorithms have been selected based on their relevance in multi-objective optimization and their common use in related research. Particular, MOEA/D is closely related to our decomposition-based framework, but the performance is weaker than that of the proposed LLM-GWO scheme. This result confirms that the proposed LLM-GWO approach improves target estimation performance.

VI. Conclusion

This paper presented a novel LLM-enabled near-field ISAC framework that jointly optimizes communication and sensing by maximizing network utility and minimizing localization CRBs. By integrating LLM with the Grey Wolf Optimizer through prompt-guided decomposition of a multi-objective problem, we introduced a scalable and effective solution strategy. Numerical evaluations confirmed the superior performance of the proposed LLM-GWO method over traditional GWO, achieving improved Pareto fronts and faster convergence. These results not only validate the feasibility of LLM-assisted GWO optimization in ISAC systems but also establish a promising foundation for future research into intelligent, adaptive optimization tools in ISAC applocation. However, we recognize that real-world ISAC systems are likely to involve multiple BSs in distributed environments, where coordination between BSs and efficient resource management across multiple nodes are crucial. Future work could focus on the development of a distributed version of the LLM-GWO framework to better handle multi-BS ISAC systems.

References

[1] N. Babu, A. Kosasih, C. Masouros, and E. Bjrnson, "Symbol-Level Precoding for Near-Field ISAC," *IEEE Communications Letters*, vol. 28, no. 9, pp. 2041–2045, 2024.

[2] Z. Chen, L. Huang, H. C. So, H. Jiang, X. Y. Zhang, and J. Wang, "Deep Reinforcement Learning Over RIS-Assisted Integrated Sensing and Communication: Challenges and Opportunities," *IEEE Vehicular Technology Magazine*, vol. 20, no. 3, pp. 97–105, 2025.

[3] Z. Chen, G. Chen, J. Tang, S. Zhang, D. K. So, O. A. Dobre, K.-K. Wong, and J. Chambers, "Reconfigurable-Intelligent-Surface-Assisted B5G/6G Wireless Communications: Challenges, Solution, and Future Opportunities," *IEEE Communications Magazine*, vol. 61, no. 1, pp. 16–22, 2023.

[4] Z. Zhang, Y. Liu, Z. Wang, J. Chen, and T. Q. S. Quek, "Dynamic MIMO Architecture Design for Near-Field Communications," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14 669–14 684, 2024.

[5] H. Jiang, W. Shi, X. Chen, Q. Zhu, and Z. Chen, "High-Efficient Near-Field Channel Characteristics Analysis for Large-Scale MIMO Communication Systems," *IEEE Internet of Things Journal*, pp. 1–1, 2024.

[6] S. Yang, C. Xie, W. Lyu, B. Ning, Z. Zhang, and C. Yuen, "Near-Field Channel Estimation for Extremely Large-Scale Reconfigurable Intelligent Surface (XL-RIS)-Aided Wideband mmWave Systems," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 6, pp. 1567–1582, 2024.

[7] J. Cong, C. You, J. Li, L. Chen, B. Zheng, Y. Liu, W. Wu, Y. Gong, S. Jin, and R. Zhang, "Near-Field Integrated Sensing and Communication: Opportunities and Challenges," *IEEE Wireless Communications*, pp. 1–8, 2024.

[8] K. Qu, S. Guo, N. Saeed, and J. Ye, "Near-Field Integrated Sensing and Communication: Performance Analysis and Beamforming Design," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6353–6366, 2024.

[9] Y. Xu, Y. Li, and T. Q. Quek, "Robust transmission design for ris-assisted integrated sensing and communication systems," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 11, pp. 17 151–17 164, 2024.

[10] X. Yang, R. Zhang, D. Zhai, F. Liu, R. Du, and T. X. Han, "Constellation design for integrated sensing and communication with random waveforms," *IEEE Transactions on Wireless Communications*, vol. 23, no. 11, pp. 17 415–17 428, 2024.

[11] D. Singh, A. Kumar, H. D. Joshi, A. K. Singh, W. Anwar, T. Myllyl, M. Magarini, L. Nkenyereye, and K. Dev, "Generalized Adaptive Spreading Modulation: A Novel Waveform for Integrated Sensing and Communication Oriented Vehicular Applications," *IEEE Internet of Things Journal*, vol. 11, no. 18, pp. 29 486–29 493, 2024.

[12] N. H. Chu, D. N. Nguyen, D. T. Hoang, Q.-V. Pham, K. T. Phan, W.-J. Hwang, and E. Dutkiewicz, "AI-Enabled mm-Waveform Configuration for Autonomous Vehicles With Integrated Communication and Sensing," *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 16 727–16 743, 2023.

[13] A. M. Elbir, K. V. Mishra, S. Chatzinotas, and M. Bennis, "Terahertz-Band Integrated Sensing and Communications: Challenges and Opportunities," *IEEE Aerospace and Electronic Systems Magazine*, pp. 1–10, 2024.

[14] H. Jiang, W. Shi, Z. Zhang, C. Pan, Q. Wu, F. Shu, R. Liu, Z. Chen, and J. Wang, "Large-Scale RIS Enabled Air-Ground Channels: Near-Field Modeling and Analysis," *IEEE Transactions on Wireless Communications*, pp. 1–11, 2024.

[15] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective Signal Processing Optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, 2014.

[16] Z. Chen, J. Tang, X. Y. Zhang, D. K. C. So, S. Jin, and K.-K. Wong, "Hybrid Evolutionary-Based Sparse Channel Estimation for IRS-Assisted mmWave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1586–1601, 2022.

[17] B. Liu, X. Liu, S. Gao, X. Cheng, and L. Yang, "LLM4CP: Adapting Large Language Models for Channel Prediction," *Journal of Communications and Information Networks*, vol. 9, no. 2, pp. 113–125, 2024.

[18] M. Wong, T. Rios, S. Menzel, and Y. S. Ong, "Prompt Evolutionary Design Optimization with Generative Shape and Vision-Language models," in *2024 IEEE Congress on Evolutionary Computation (CEC)*, 2024, pp. 1–8.

[19] M. Pluhacek, A. Kazikova, T. Kadavy, A. Viktorin, and R. Senkerik, "Leveraging large language models for the generation of novel metaheuristic optimization algorithms," in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, 2023, pp. 1812–1820.

[20] C. Yang, X. Wang, Y. Lu, H. Liu, Q. V. Le, D. Zhou, and X. Chen, "Large language models as optimizers," vol. arXiv preprint arXiv:2309.03409, 2024.

[21] F. Liu, X. Lin, Z. Wang, S. Yao, X. Tong, M. Yuan, and Q. Zhang, "Large Language Model for Multi-objective Evolutionary Optimization," *arXiv preprint arXiv:2310.12541*, 2024.

[22] C. Zhen, L. Jianqing, Z. Haijun, and Z. Wei, "Low-complexity channel estimation for RIS-assisted ISAC system," *China Communications*, vol. 22, no. 4, pp. 296–308, 2025.

[23] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[24] E. Emary, H. M. Zawbaa, and C. Grosan, "Experienced gray wolf optimization through reinforcement learning and neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 3, pp. 681–694, 2018.

[25] K. Sanderson, "GPT-4 is here: what scientists think," *Nature*, vol. 615, p. 773, 2023.

[26] J. Chen, Z. Liu, X. Huang, C. Wu, Q. Liu, G. Jiang, Y. Pu, Y. Lei, X. Chen, and X. Wang, "When large language models meet personalization: perspectives of challenges and opportunities," *World Wide Web*, vol. 27, no. 4, pp. 1–45, 2024.

[27] Y. Chen, X. Song, C. Lee, Z. Wang, R. Zhang, D. Dohan, K. Kawakami, G. Kochanski, A. Doucet, M. A. Ranzato, S. Perel, and N. de Freitas, "Towards learning universal hyperparameter optimizers with transformers," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 32 053–32 068.

[28] Z. Ding, R. Schober, and H. V. Poor, "NOMA-Based Coexistence of Near-Field and Far-Field Massive MIMO Communications," *IEEE Wireless Communications Letters*, vol. 12, no. 8, pp. 1429–1433, 2023.

[29] K. Wang, Z. Ding, and G. K. Karagiannidis, "User Clustering for Coexistence Between Near-Field and Far-Field Communications," *IEEE Communications Letters*, vol. 28, no. 10, pp. 2442–2446, 2024.

[30] X. Zhang, H. Zhang, and Y. C. Eldar, "Near-Field Sparse Channel Representation and Estimation in 6G Wireless Communications," *IEEE Transactions on Communications*, vol. 72, no. 1, pp. 450–464, 2024.

[31] Z. Wang, X. Mu, and Y. Liu, "Near-Field Integrated Sensing and Communications," *IEEE Communications Letters*, vol. 27, no. 8, pp. 2048–2052, 2023.

[32] A. Hussain, A. Abdallah, and A. M. Eltawil, "Redefining Polar Boundaries for Near-Field Channel Estimation for Ultra-Massive MIMO Antenna Array," *IEEE Transactions on Wireless Communications*, vol. 24, no. 10, pp. 8193–8207, 2025.

[33] C. Yang, G. Zhang, M. Cui, Q. Wu, and Y. Zeng, "Near-Field Codebook-Based 3-D Spherical Channel Estimation for UCA XL-MIMO Systems," *IEEE Wireless Communications Letters*, vol. 14, no. 10, pp. 3154–3158, 2025.

[34] H. Miao and M. Peng, "Near-field high-speed user sensing in wideband mmwave communications: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 73, pp. 919–935, 2025.

[35] K. Fang, Y. Ouyang, B. Zheng, L. Huang, G. Wang, and Z. Chen, "Security Enhancement for RIS-Aided MEC Systems With Deep Reinforcement Learning," *IEEE Transactions on Communications*, vol. 73, no. 4, pp. 2466–2479, 2025.

[36] Z. Chen, Y. Guo, P. Zhang, H. Jiang, Y. Xiao, and L. Huang, "Physical Layer Security Improvement for Hybrid RIS-Assisted MIMO Communications," *IEEE Communications Letters*, vol. 28, no. 11, pp. 2493–2497, 2024.

[37] I. Das and J. E. Dennis, "Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems," vol. 8, no. 3, 1998.

[38] Y. Hao, Q. Ni, H. Li, and S. Hou, "Robust Multi-Objective Optimization for EE-SE Tradeoff in D2D Communications Underlaying Heterogeneous Networks," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4936–4949, 2018.

[39] Q. Wu, Y. Zeng, and R. Zhang, "Joint Trajectory and Communication Design for Multi-UAV Enabled Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2109–2121, 2018.

[40] H. Li, M. Xiao, K. Wang, D. I. Kim, and M. Debbah, "Large Language Model Based Multi-Objective Optimization for Integrated Sensing and Communications in UAV Networks," *IEEE Wireless Communications Letters*, vol. 14, no. 4, pp. 979–983, 2025.

[41] L. Ke, Q. Zhang, and R. Battiti, "MOEA/D-ACO: A Multiobjective Evolutionary Algorithm Using Decomposition and AntColony," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1845–1859, 2013.

[42] H. Li, Z. Wang, X. Mu, P. Zhiwen, and Y. Liu, "Near-Field Integrated Sensing, Positioning, and Communication: A Downlink and Uplink Framework," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 9, pp. 2196–2212, 2024.