arXiv:2505.15047v4 [cs.LG] 9 Feb 2026

PiFlow: Principle-Aware Scientific Discovery with Multi-Agent Collaboration

Yingming Pu [1 2] **Tao Lin** [1] **Hongyu Chen** [1]

Abstract

Large Language Model (LLM)-based multi-agent systems (MAS) demonstrate remarkable potential for scientific discovery. Existing approaches, however, often automate scientific discovery using predefined workflows that lack rationality constraints. This often leads to aimless hypothesizing and a failure to consistently link hypotheses with evidence, thereby hindering the systematic reduction of uncertainty. Overcoming these limitations fundamentally requires a principled approach to exploration. We introduce **PiFlow**, an information-theoretical framework, treating automated scientific discovery as a structured uncertainty reduction problem guided by principles (e.g., scientific laws). Extensive evaluations across three distinct scientific domains demonstrate that PiFlow **(I)** improves discovery efficiency by 31.18%~41.73% and solution quality by 12.47%~31.72% against state-of-the-art methods, **(II)** delivers a 5.6× speedup in time-to-solution while reducing token consumption by up to 27% compared to vanilla agents, and **(III)** serves as a Plug-and-Play module that generalizes on existing agent architecture. Overall, PiFlow establishes a novel paradigm shift in highly efficient agentic scientific discovery, paving the way for more robust and accelerated AI-driven research.

1. Introduction

Large Language Model (LLM)-based Multi-Agent Systems (MAS) have significantly impacted automated scientific discovery (Minaee et al., 2024; Wang et al., 2023; Zhang et al., 2024c) across a wide range of fundamental fields, including chemistry (Liu et al., 2024; Ghafarollahi & Buehler, 2024a; Yang et al., 2024c; Inoue et al., 2024), biology (Xiao et al., 2024; Nagarajan et al., 2025; Averly et al., 2025; Ghafarollahi & Buehler, 2024b), physics (Jaiswal et al., 2024), and material science (Takahara et al., 2025; Ghafarollahi & Buehler, 2025; Ansari et al., 2024; Wan et al., 2024; Zhang et al., 2024b).

Although proficient in executing experiments within predefined workflows (Lu et al., 2024; Lai & Pu, 2025), these systems often generate hypotheses that lack clear direction, leading to the uncertainty in establishing clear links between hypotheses and their supporting or refuting evidence (AI4Science & Quantum, 2023; Zhou et al., 2024; Baek et al., 2024; Schmidgall et al., 2025; Prabhakar et al., 2025; Xie et al., 2023a). Such a disconnect indicates inefficient exploration. Moreover, many of these approaches are tailored for specific tasks, often relying on meticulous prompt engineering that heavily incorporates domain knowledge (as detailed in Appendix A). Consequently, their ability to adapt to new scientific domains is often limited without substantial modifications (Zhang et al., 2025; Kumbhar et al., 2025). These issues culminate in three primary challenges: (a) **aimless hypothesizing**; (b) **unmaintained connections** between hypotheses and evidence during exploration, hindering systematic validation and (c) **limited generalization ability**, where systems effective in one scientific scenario (e.g., material science) often require substantial rework to be applicable in others.

To address these limitations, we introduce PiFlow, an information-theoretic framework for structured uncertainty reduction in scientific discovery. Viewing scientific exploration as a game against an unknown and challenging nature, where robust strategies are paramount, PiFlow employs Min-Max optimization: **minimizing** cumulative regret for exploitation, while **maximizing** information gain for efficient hypothesis exploration. As a Plug-and-Play module, PiFlow integrates with MAS capable of hypothesizing and experimentation. Inspired by the challenge of navigating vast hypothesis spaces, PiFlow operates by using fundamental scientific principles, which may be initially proposed by or refined using LLMs.

The iterative selection of principles progressively reduces uncertainty in hypothesizing and the interpretation of evidence, dynamically steering exploration by prioritizing those scientific principles that offer the highest instructive value for continued exploration. Figure 1 illustrates PiFlow's method for assessing and utilizing scientific prin-

[1] Westlake University, Hangzhou, Zhejiang, China. [2] Email: puyingming@westlake.edu.cn. Correspondence to: Tao Lin <lintao@westlake.edu.cn>.

Preprint. February 10, 2026.

ciples within a drug discovery context.

By operationalizing the Min-Max optimization, PiFlow systematically steers discovery via three strategic actions: *explore, validate,* or *refine*, thereby maximizing information gain while minimizing cumulative regret. This mechanism theoretically guarantees a sublinear regret bound of $\mathcal{O}(\sqrt{T})$ (proof in Appendix C), ensuring that the system avoids aimless exploration and efficiently converges to optimal solutions even in complex, high-dimensional landscapes.

We conduct comprehensive evaluations across three distinct scientific domains: discovering nanomaterials, biomolecules, and superconductors. Empirical results demonstrate that PiFlow is ❶ **high-performing**, surpassing state-of-the-art baselines by 31.18%~41.73% in discovery efficiency and 12.47%~31.72% in solution quality; ❷ **token-economical**, achieving a 5.6× speedup with 27% lower cost than vanilla agents; ❸ **convergent**, theoretically guaranteed and empirically validated by a sublinear regret bound of $\mathcal{O}(\sqrt{T})$; and ❹ **generalizable**, serving as a Plug-and-Play module compatible with varying LLM backbones and agent frameworks.

In summary, our contributions are:

(a) We propose a novel paradigm for **principle-aware scientific discovery**, built upon an information-theoretical foundation that offers convergence guarantees.
(b) We develop PiFlow, a **Plug-and-Play framework** that seamlessly integrates with existing MAS to enable focused hypothesis exploration, thereby enhancing discovery efficiency and flexibility.
(c) We conduct **extensive experiments across three distinct scientific discovery scenarios**, demonstrating the broad applicability and significant performance improvements achieved by PiFlow.

2. Related Work

2.1. Language Models for Scientific Discovery

Recently, large language models (LLM) have advanced scientific discovery with automation and rational design (Ren et al., 2025; Ma et al., 2024). The internal knowledge of LLMs has demonstrated promising capability in focused chemical and material discovery (Yang et al., 2024c; Zhou et al., 2024; Pu et al., 2024; Ghafarollahi & Buehler, 2024c). While tool-integrated LLMs like SciAgents (Ghafarollahi & Buehler, 2024d), DARWIN (Xie et al., 2023a) and HoneyComb (Zhang et al., 2024a) improve domain-specific reasoning and recall of factual insights, they still struggle to integrate physicochemical laws effectively when refining insights for design, risking biased proposals and inefficient exploration due to inherent hallucinations of LLMs (Zhang et al., 2023b). Human-AI frameworks address this issue by leveraging the knowledge of domain experts (Reddy & Shojaee, 2024; Eythorsson & Clark, 2025), yet remain limited to the scope of hypothesis generation (Alkan et al., 2025), leading to insufficient exploration of complex chemical spaces (Luo et al., 2025). Surveys highlight persistent gaps in efficiency and interpretability (Zhang et al., 2024c; Han et al., 2024), underscoring the need for principled scientific discovery management beyond automated LLM reasoning (Ramos et al., 2024).

2.2. Approaches of Multi-Agent Collaboration

Multi-agent systems (MAS) show promise for complex tasks (Lu et al., 2024; Ghafarollahi & Buehler, 2024d; Ni & Buehler, 2023), yet their application to scientific discovery reveals limitations in current collaboration mechanisms (Tran et al., 2025). Rule-based methods (Zhang et al., 2023a) offer consistency but their predefined rules lack the flexibility to incorporate nuanced scientific principles (e.g., physicochemical laws) or adapt to unexpected findings, hindering dynamic exploration. The method of AI Researcher (Tang et al., 2025), and role-playing approaches (Tran et al., 2025; He et al., 2024) leverage agent expertise, yet rigid roles can impede adaptation in scientific research (Ramirez-Medina et al., 2025; Lu et al., 2024; Ghafarollahi & Buehler, 2024d), and ensuring collective adherence to scientific principles when interpreting experimental insights is challenging. Model-based methods (Xu et al., 2023; Mu et al., 2023; Li et al., 2023) aim for adaptability by learning from uncertainty, but struggle to build world models that accurately capture complex scientific phenomena and integrate guiding laws (Hao et al., 2023), thereby impairing the balance between information perception and strategic, principle-guided reasoning.

Consequently, existing MAS paradigms often lack a dedicated awareness and systematic application of scientific principles during hypothesis generation and refinement (Gridach et al., 2025; Luo et al., 2025; Reddy & Shojaee, 2024; Su et al., 2024). This highlights a critical need for an explicitly principle-aware multi-agent collaboration framework. Our work addresses this gap, proposing a method where the collaborative discovery process is robustly guided by scientific principles to achieve more efficient and reliable outcomes.

3. Methodology

3.1. Overview

We propose a principle-aware MAS designed to enhance scientific discovery via focused hypothesizing and structured exploration of hypothesis-evidence connections. Figure 2 illustrates the architecture. Its core comprises a Hypothesis-Validation loop that iteratively generates and tests hypothe-

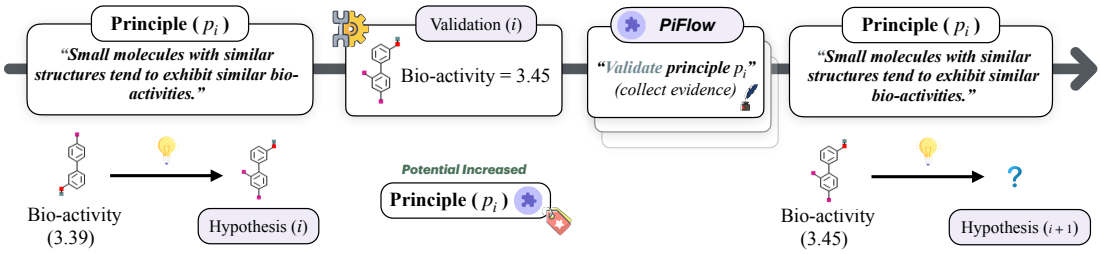


Figure 1. **Illustration of the potential of a scientific principle in drug discovery.** `PiFlow` directs exploration to prioritize hypotheses aligned with high-potential principles (or their variants), thereby iteratively guiding the discovery towards optimal candidate molecules.

ses. `PiFlow` guides this loop by optimizing accumulated principle-outcome data. Strategic insights dynamically optimized by `PiFlow` are relayed through a Planner agent ($\mathcal{A}_P$) to the Hypothesis Agent within the loop. Subsequent sections will elaborate on `PiFlow`'s specific architecture and its information-theoretical underpinnings.

3.2. Architecture

Our proposed principle-aware system leverages LLM-based MAS to conduct scientific discovery through strategic hypothesizing. The framework is comprised of two core, interconnected components: (1) an MAS that executes a Hypothesis-Validation loop, and (2) `PiFlow`, which serves as the strategic director for this loop.

Definition 3.1 (Scientific Principles)**.** The scientific principles are foundational concepts, established laws, or patterns, articulated as natural language statements, that explain phenomena within a specific scientific domain. These principles serve as high-level conceptual building blocks from which specific, testable hypotheses can be derived. This conceptualization aligns with broader discussions on the nature of scientific knowledge (Poincaré, 1906).

Hypothesis-Validation Loop. As depicted in the right-hand side of Figure 2, the Hypothesis-Validation loop constitutes the core operational cycle and incorporates two LLM-based agents: Hypothesis Agent ($\mathcal{A}_H$) and Experiment Agent ($\mathcal{A}_E$), detailed below:

(a) **Hypothesis.** Initially, a set of dynamically growing candidate principles $\mathcal{P} = \{p_1, p_2, p_3\}$ (see Definition 3.1) potentially proposed by domain experts (see Appendix M) or extracted by LLMs, is established. In each iteration t, $\mathcal{A}_H$ proposes a testable hypothesis, h_t, grounded in a selected principle $p_i \in \mathcal{P}$. The proposal is supported by structured rationales (comprising major and minor premises, following Eisape et al. (2023), see Appendix Q.2).

(b) **Validation.** Subsequently, $\mathcal{A}_E$ rigorously validates h_t using an experimental tool, denoted $f^*(\cdot)$, which yields a quantitative outcome $y_t = f^*(h_t)$ (e.g., property value of a material). This outcome serves as feedback for subsequent hypothesizing.

This iterative process progressively establishes a record of principle-outcome pairs, $\mathcal{T}_t = \{\langle p_k, y_k \rangle\}_{k=1}^{t}$, linking each hypothesized principle p_k to its observed experimental outcome y_k. While this loop systematically generates valuable evidence in trajectory $\mathcal{T}_t$, the selection of potential principles for subsequent hypotheses can lack strategic direction if not externally guided. This may lead to inefficient exploration of the principle space or premature convergence on suboptimal findings.

Hypothesis steering with `PiFlow`. To address the potential inefficiencies in principle selection and to instill strategic direction, the `PiFlow` component is introduced. It leverages the dynamically growing set of principle-outcome pairs, $\mathcal{T}_t$ as its primary input. Two steps are conducted to steer the hypothesizing:

(a) **High-potential principle acquisition.** After an initial phase of evidence collection, during which $\mathcal{T}_t$ is populated, `PiFlow` activates its core mechanism: an adversarial Min-Max optimization (detailed in Section 3.3). This optimization process analyzes $\mathcal{T}_t$ to identify a principle, p^* (e.g., p_2 in Figure 2), predicted to balance the exploitation and exploration, i.e., have the highest potential, for advancing the scientific inquiry.

(b) **Principle steering for hypothesizing.** Following the identification of the highest-potential principle p^* by the Min-Max optimization, `PiFlow` assigns a potential score to all principles in $\mathcal{P}$. This scoring enables a threshold-based partitioning: principles with high scores (e.g., p_2') drive refinement; those with medium scores trigger further validation ($p_2 \rightarrow p_2'$); and low scores prompt exploration of new conceptual areas.

This closed-loop feedback mechanism ensures that the Hypothesis-Validation cycle is continuously and adaptively steered by strategic insights derived from the system's cumulative experience, as embodied in $\mathcal{T}_t$. We provide a detailed analysis of the distinction from prompt engineering at Appendix B, and an illustrative example at Appendix E.

Plug-and-Play Modularity of `PiFlow`. The hypothesis steering mechanism above, which includes `PiFlow` and its interfacing Planner agent $\mathcal{A}_P$, has been intentionally engineered as a modular, Plug-and-Play system. This archi-

tectural choice ensures that principle-aware guidance can be readily integrated to enhance MAS engaged in scientific discovery as a part of prompts. As demonstrated in Appendix H, we successfully integrate PiFlow with existing ChemToolAgent (Yu et al., 2024) without any architecture modifications. This adaptability positions PiFlow as a pivotal enhancement for automated scientific inquiry.

3.3. Min-Max Optimization in PiFlow

In our proposed method, strategic principle selection is achieved through a Min-Max optimization, as presented in Eq. 1. This approach is designed to balance the exploitation of established, high-potential principles (which then guide the formulation of specific hypotheses) with the exploration of novel ones, while explicitly incorporating information acquisition efficiency:

$$\min_{\pi\in\Pi}\max_{f^*\in\mathcal{F}}\mathbb{E}_\pi\left[\sum_{t=1}^{T}(v^*-f^*(h_t))-\lambda\cdot I(h_t;f^*|H_{t-1})\right] \tag{1}$$

where $\pi \in \Pi$ represents the decision-making policy for selecting principles and $f^* \in \mathcal{F}$ is the evaluation function (e.g., an experimental tool) that characterizes the quantitative outcome yielded from hypothesis $h_t \in \mathcal{H}$, where $\mathcal{H}$ is the hypothesis space. Over T iterations, the policy π aims to minimize the objective function in Eq. 1. This objective strategically balances: (1) the summation for cumulative regret, encouraging exploitation of known high-potential principles, and (2) the mutual information, thereby effectively maximizing information gain to foster exploration. This structure allows PiFlow to navigate the complex trade-offs between these two goals.

Minimizing cumulative regret (exploitation). The first term, $\sum_{t=1}^{T}(v^* - f^*(h_t))$, represents the cumulative regret over T iterations. Here, v^* is a theoretical optimal outcome value, and $f^*(h_t)$ is the outcome from hypothesis h_t. By minimizing this term, the policy π is driven to exploit known high-potential principles and hypotheses to achieve outcomes as close as possible to the optimum v^*. This encourages the refinement and validation of promising avenues.

Maximizing information gain (exploration). The second term $-\lambda \cdot I(h_t; f^*|H_{t-1})$ promotes exploration. The policy π seeks to minimize this term, which is equivalent to maximizing the mutual information $I(h_t; f^*|H_{t-1})$. This mutual information quantifies the expected reduction in uncertainty about the true evaluation function f^* upon observing the outcome of hypothesis h_t, given all past observations $H_{t-1} = \{(h_m, y_m)\}_{m=1}^{t-1}$. The trade-off parameter $\lambda > 0$ controls the balance between minimizing regret (exploitation) and maximizing information gain (exploration). A larger λ places more emphasis on information acquisition.

Remark (The dependencies of f^* in $I(h_t; f^*|H_{t-1})$) **.** The informativeness of a proposed hypothesis h_t is inherently dependent on the nature of the true underlying evaluation function f^*. The adversarial nature of the max_{f^*} operator means that the policy π must select hypotheses h_t that are expected to be informative even if f^* were to manifest in a way that makes h_t minimally revealing, thus ensuring robustness in information acquisition.

Building upon the theoretical framework above, we derive a computationally tractable algorithm (Algorithm 1). Since precise computation of information gain (Eq. 1) is intractable, we employ Semantic Embedding Distance as a computationally efficient proxy, justified in Appendix D.

In summary, the Min-Max adversarial formulation underpinning PiFlow provides strong theoretical guarantees, notably sublinear regret bounds (formalized in Theorem 3.2, with proof in Appendix C). Importantly, its operational behavior, which involves practical approximations of this Min-Max solution, demonstrates consistent alignment with these theoretical expectations, as empirically validated in Section 5.4.

Theorem 3.2 (Informal)**.** *The Min-Max optimization in PiFlow formulates a trade-off between exploitation (minimizing regret) and exploration (maximizing information gain). Under conditions of finite entropy $H(f^*)$ and bounded evaluation function f^*, this optimization provides two key theoretical guarantees: (1) As information gain decreases, the expected regret also decreases; (2) the cumulative regret grows at a sublinear rate of $\mathcal{O}\left(\sqrt{T}\right)$.*

4. Experiment

4.1. Settings

To rigorously evaluate the effectiveness and versatility of our PiFlow framework, we conducted experiments across three distinct scientific discovery scenarios. While direct hypothesis validation in real-world labs is prohibitively expensive, we employ high-fidelity surrogate models deployed locally, which serve as the primary evaluation tool. Across all scenarios (Section 4.2), we frame the scientific discovery challenge as a unified task: "*Find a candidate in a complex parameter space that maximizes a target property (e.g., bio-activity).*"

To ensure a fair comparison and focus on core capabilities, all agents utilize QwenMax (Yang et al., 2024a) as the base LLM and are prohibited from accessing external search tools. The complete experimental setup is detailed in Appendix P, and the prompts are provided in Appendix Q.

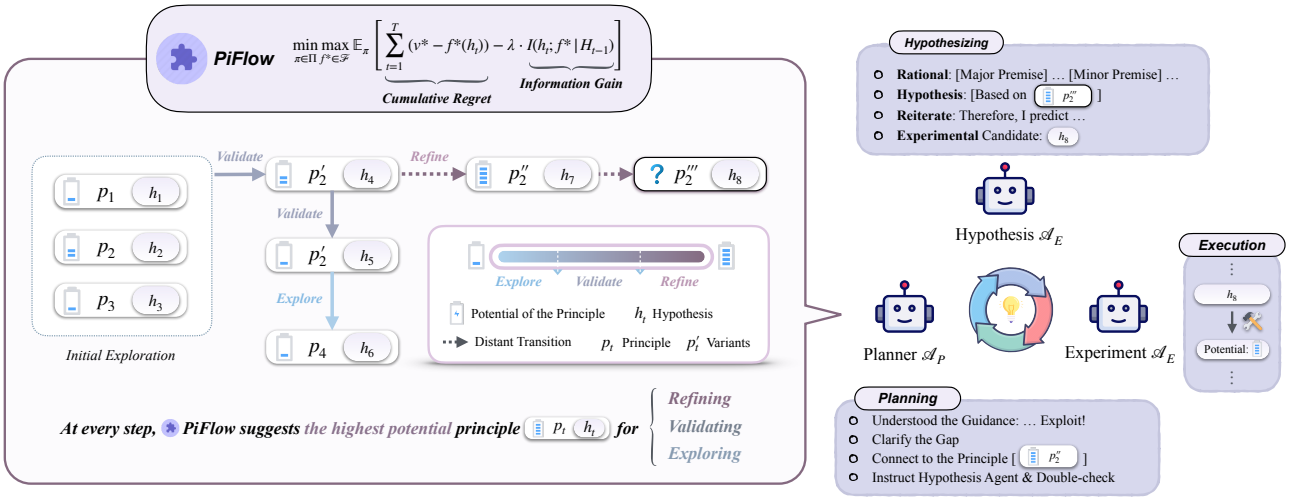


Figure 2. **Overview of the `PiFlow` Architecture for Scientific Discovery.** The `PiFlow` component utilizes Min-Max optimization to strategically select and direct high-potential principles to the Planner agent. The Planner, in turn, guides the Hypothesis-Validation loop, where agents iteratively generate hypotheses h_t from principles p_t at step t, execute experiments, and refine understanding. This iterative process is designed to efficiently navigate the discovery landscape.

4.2. Experimental Scenarios

To comprehensively assess the performance, we design three scenarios that represent canonical challenges in scientific exploration: optimization in *continuous*, *discrete*, and *hybrid* parameter spaces, as detailed below.

Nanohelix Optimization (NHO). We use a surrogate model ($r^2 = 0.98$) trained on DFT-simulated data following Wu et al. (2025) to predict chirality of a nanohelix material from four *continuous* geometric parameters (see Appendix P.1).

Molecular Bio-activity Optimization (MBO). We use a surrogate model ($r^2 = 0.91$) trained on 50,000 molecules from ChEMBL35 (Zdrazil et al., 2023) database to predict bio-activity of a molecule, quantified by pChEMBL value, in a *discrete* chemical space (see Appendix P.2).

Superconductor Optimization (SPO). Following Hamidieh (2018), we train a surrogate model ($r^2 = 0.91$) to map a material's *hybrid* compositional features, e.g., $Bi_2Sr_2Ca_4Cu_8O_{19}$, to its critical temperature (T_c). As reported (Hamidieh, 2018), this task is aiming for accelerating the discovery of room-temperature superconductors (see Appendix P.3 for more details).

4.3. Baselines

To evaluate the strategic guidance of `PiFlow` under uncertainty, we therefore benchmark against the following baselines: (1) **Reasoning and Acting (ReAct) (Yao et al., 2022).** ReAct enables agents to iteratively formulate hypotheses, design/execute experiments, and interpret results. It represents a foundational approach to structured reasoning. (2) **Meta Plan Optimization (MPO) (Xiong et al., 2025).** MPO employs a trained LLM planner that provides high-level, general guidance. This serves as a direct counterpoint to `PiFlow`'s explicit, structured mechanism for principled uncertainty reduction. (3) **Vanilla Agent System (Vanilla).** This baseline consists of an MAS operating without any strategic oversight. It is intended to establish a performance floor, demonstrating the limitations of unguided exploration reliant solely on agent role-playing. (4) **AI Researcher** (Tang et al., 2025), a multi-agent research system for autonomous scientific innovation. (5) **The AI Scientist v2** (Yamada et al., 2025), an end-to-end agentic system that leverages an tree-search methodology for prompting research progress. We adapt these two AI Scientist methods into our tasks by excluding any information retrieval from external resources and manuscript drafting for a fair comparison.

While other advanced frameworks exist, such as Co-Scientist (Gottweis et al., 2025) and Kosmos (Mitchener et al., 2025), their objectives diverge from the scope of de novo evidence-gathering process central to `PiFlow`. Therefore, our selection of baselines is specifically designed to isolate and evaluate our contribution to discovery efficiency in uncertain environments.

4.4. Evaluation Metrics

To rigorously evaluate `PiFlow`, we employ two complementary metrics designed to capture both peak performance (effectiveness) and trajectory dynamics (efficiency):

Solution Quality (SQ). We measure the optimal objective value with a percentage relative to the theoretical maximum

Table 1. **Performance Comparison on Four Scientific Discovery Tasks.** Performance is measured by **AUC** (Area Under Curve, %) and **SQ** (Solution Quality, %) and mean pm std is reported. **Bold** values denote the best result in each column, underlined values denote the second-best result. The rightmost columns show task-averaged performance, with red percentages indicating the relative performance gap of baseline methods compared to our PiFlow.

Method	NHO (Continuous)		MBO (Discrete)		SPO (Hybrid)		Average	
	AUC (%) ↑	SQ (%) ↑	AUC (%) ↑	SQ (%) ↑	AUC(%) ↑	SQ (%) ↑	AUC (%)	SQ (%)
ReAct	35.85 ±7.12	41.96 ±0.82	29.61 ±9.74	43.11 ±9.98	5.29 ±0.69	6.41 ±0.96	23.58 ↓ 52.70%	30.49 ↓ 53.38%
Vanilla	35.96 ±22.38	46.76 ±7.29	29.71 ±12.66	49.22 ±8.30	11.39 ±11.33	14.16 ±13.37	25.69 ↓ 48.49%	36.71 ↓ 43.87%
MPO	43.99 ±2.79	51.29 ±7.77	31.18 ±9.12	57.28 ±5.85	12.68 ±7.75	33.20 ±23.75	29.28 ↓ 41.27%	47.26 ↓ 27.75%
AI Researcher	46.45 ±1.98	53.12 ±0.87	42.72 ±12.71	**95.66** ±6.08	16.36 ±2.22	25.69 ±2.64	35.18 ↓ 29.45%	58.16 ↓ 11.08%
The AI Scientist v2	49.27 ±1.51	56.67 ±4.73	36.32 ±8.67	62.47 ±6.55	**28.43** ±2.55	29.85 ±2.18	38.01 ↓ 23.78%	49.66 ↓ 24.07%
PiFlow (ours)	**63.51** ±11.18	**76.82** ±4.54	**64.57** ±23.65	84.55 ±29.63	21.51 ±2.80	**34.85** ±1.19	**49.86**	**65.41**

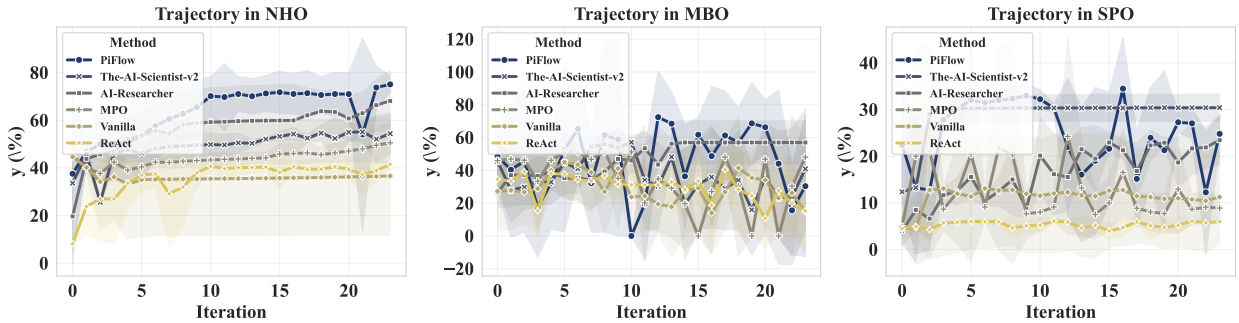


Figure 3. **Optimization trajectories across distinct scientific domains.** The curves depict the objective values over 24 budget steps. Methods like The AI Scientist v2 (Yamada et al., 2025) remains stuck (e.g. from "$Bi_2Sr_2Ca_4Cu_8O_{18}$" to "$Bi_2Sr_2Ca_4Cu_8O_{19}$" in SPO). However, PiFlow exhibits characteristic oscillatory patterns indicative of its active *principle-aware* exploration strategy, attains the best performance of 65.41% SQ and 49.86% AUC on average.

or reference value μ_{ref}, denoted as:

$$\text{SQ} = \frac{\max\{y_k | \langle p_k, y_k \rangle \in \mathcal{T}_t\}}{\mu_{\text{ref}}} \times 100\%.$$

Specifically, for NHO, the theoretical maximum is reported as $\mu_{\text{ref}}^{\text{g-factor}} = 2.0$ following Greenfield et al. (2021). The larger the g-factor, the stronger the chirality. For MBO, a strict threshold of $\mu_{\text{ref}}^{\text{pchembl}} = 6.5$ has been reported (Lenselink et al., 2017) for lead compound optimization stage in molecular biology, larger value indicates a higher bio-activity and vice versa. For SPO, the reference value is set to be room temperature of 298.15K, denoted as $\mu_{\text{ref}}^{\text{Tc}} = 298.15K$.

Area Under the Curve (AUC). Exploration efficiency requires both (1) **rapid convergence** and (2) **high objective values (exploitation)**. We quantify these two factors by defining the AUC metric. Given the trajectory $\langle p_k, y_k \rangle \in \mathcal{T}_t$ across t steps, AUC is computed via the trapezoidal rule. For meaningful comparisons, we normalize by the maximum possible area:

$$\text{AUC} = \frac{\sum_{i=1}^{t-1} \frac{y_i + y_{i+1}}{2}}{\mu_{\text{ref}} \cdot (t-1)} \times 100\%$$

In summary, SQ measures the quality of the final outcome, while AUC evaluates the entire discovery process by rewarding both speed and consistency. The detailed rationale for these metric designs is provided in Appendix O.2.

5. Results

5.1. Performance Comparison

We use SQ to compare the capability of reaching optimal solutions, and AUC to assess discovery efficiency by measuring cumulative progress over the exploration budget.

As shown in Table 1, PiFlow achieves 49.86% AUC and 65.41% SQ on average, outperforming traditional baselines by substantial margins: +24.17% AUC and +28.70% SQ over Vanilla, and +20.58% AUC and +18.15% SQ over MPO. Against SOTA agents, PiFlow surpasses AI Researcher by 41.73% in AUC and The AI Scientist v2 by 31.72% in SQ, demonstrating superior global search capabilities across all three parameter space types (continuous, discrete, and hybrid).

As visualized in Figure 3, The AI Scientist v2 achieves the highest AUC (28.43%) in SPO due to rapid early gains, yet its trajectory stagnates, yielding a final SQ of only 29.85%—a 16.8% gap below PiFlow's 34.85%. This illustrates entrapment in local optima. In contrast, **PiFlow's oscillatory trajectories reflect an active exploration strategy** that escapes local traps. On average, PiFlow outperforms

Table 2. **Ablation of Test-Time Reasoning.** Comparison of performance with explicit reasoning enabled (`w/ Think`) versus disabled (`w/o Think`). **Bold** marks the best results. Disabling the specific `Think` mode yields superior performance, suggesting that avoiding explicit Chain-of-Thought helps prevent *cognitive fixation* in this context (see discussion in Section 5.2).

Method/Setting		AUC (%)	SQ (%)
Qwen3-32B	w/ `Think`	37.51 ±7.70	58.76 ±6.18
	w/o `Think`	**45.51** ±11.19	**68.86** ±17.67
Qwen3-8B	w/ `Think`	30.55 ±27.45	54.49 ±22.25
	w/o `Think`	**42.09** ±16.55	**61.59** ±16.43

The AI Scientist v2 by 31.18% in AUC (49.86% vs. 38.01%) and 31.72% in SQ (65.41% vs. 49.66%), confirming that the observed variance stems from productive exploration rather than instability. Further details and statistical significance tests are provided in Appendix F.

Takeaway: Unlike baselines (e.g., The AI Scientist v2 (Yamada et al., 2025)), which suffer from premature stagnation, `PiFlow` dynamically navigates the solution space to locate high-outcome regions.

5.2. Ablation Study

We conduct ablations on the NHO task to evaluate two aspects: (1) the Plug-and-Play effectiveness of `PiFlow`, and (2) the impact of LLM reasoning modes.

Plug-and-Play Effectiveness. To isolate `PiFlow`'s contribution, we toggle its strategic guidance (Algorithm 1) across two LLM backbones. As shown in Table 3, `PiFlow` consistently improves both metrics: GPT4.1-mini gains +12.3% AUC (37.12%→41.68%) and +65.4% SQ (40.14%→66.38%); Qwen3-32B gains +38.7% AUC (27.04%→37.51%) and +7.1% SQ (54.84%→58.76%). Furthermore, we demonstrate the lightweight integration with existing MAS, confirming the `PiFlow` module's Plug-and-Play utility (see Appendix H).

Test-Time Reasoning Effect. We ablate the explicit `Think` mode (Chain-of-Thought) on Qwen3 models (Table 2). Counterintuitively, disabling `Think` yields superior results: +21.3% AUC for Qwen3-32B (37.51% vs. 45.51%) and +37.8% AUC for Qwen3-8B (30.55% vs. 42.09%). We attribute this to *cognitive fixation*, where explicit CoT induces path dependence, causing premature commitment to flawed assumptions. Direct inference under `PiFlow`'s guidance enables more flexible hypothesis generation.

Takeaway: `PiFlow` is a robust Plug-and-Play module that boosts performance across LLM backbones. Disabling explicit CoT reasoning synergizes with `PiFlow`'s strategic guidance, avoiding cognitive fixation.

Table 3. **Ablation Study on the `PiFlow` Layer.** Performance comparison with (`w/`) and without (`w/o`) the `PiFlow` module. **Bold** indicates the best performance in each setting.

Method/Setting		AUC (%)	SQ (%)
GPT4.1-mini	w/ `PiFlow`	**41.68** ±17.91	**66.38** ±14.90
	w/o `PiFlow`	37.12 ±16.04	40.14 ±13.97
Qwen3-32B	w/ `PiFlow`	**37.51** ±7.70	**58.76** ±6.18
	w/o `PiFlow`	27.04 ±21.50	54.84 ±24.41

5.3. Further Analyses of Robustness

To rigorously validate the proposed paradigm, we conducted an extensive suite of eight experiments analyzing `PiFlow`'s robustness and utility. These evaluations are structured to substantiate our three core contributions: (1) **Algorithmic Robustness**: We verify that the Min-Max optimization ensures resilience against epistemic uncertainty (e.g., poor initialization) and outperforms numerical baselines in ill-defined spaces; (2) **Universal Applicability**: We demonstrate `PiFlow`'s minimal-modification *Plug-and-Play* capability across different agent architectures and LLM backbones; and (3) **Operational Efficiency**: We quantify its manageable computational complexity and superior cost-effectiveness. Key findings are highlighted below:

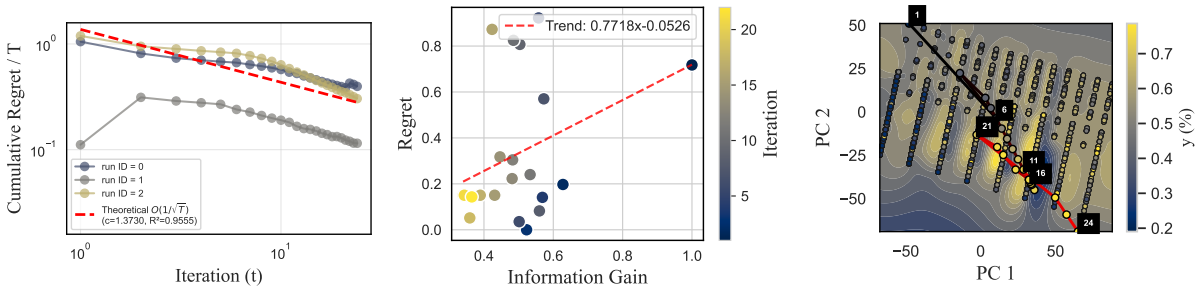
❶ **Resilience to Poor Principles Initialization.** `PiFlow` demonstrates strong error recovery capabilities. Even when initialized with deliberately misleading principles (starting SQ below 20%), the system successfully identifies and discards flawed priors within $\sim$ 10 iterations, recovering to converge on high-quality solutions ($\sim$ 70% SQ), matching the performance of expert-guided runs (see Appendix M).

❷ **Superiority over Bayesian Optimization.** On complex tasks with ill-defined search spaces (MBO), `PiFlow` substantially outperforms Bayesian Optimization (BO), achieving a SQ of **84.55%** compared to BO's **38.15%**, a margin of $+46.4\%$, without requiring the manual parameter space engineering mandated by BO (see Appendix G).

❸ **Seamless Plug-and-Play Integration.** We validate the modularity of `PiFlow` by integrating it with ChemToolAgent (Yu et al., 2024). Without any architectural modifications, `PiFlow` successfully steered the agent to discover a high-potency molecule (90.77%) in just **8 iterations**, demonstrating its lightweight integration (see Appendix H).

❹ **Negligible Computational Overhead.** Theoretical analysis confirms a decision complexity of $\mathcal{O}(t^2 \cdot d)$. Empirical profiling on NHO reveals that the `PiFlow` module accounts for only **9.4%** of the total system runtime, efficiently delivering a **5.6x** speedup in time-to-solution compared to the Vanilla baseline (see Appendix I).

❺ **High Cost-Effectiveness.** Compared to Vanilla MAS, `PiFlow`-MAS achieves a **27% reduction** in total token



(a) Average Regret Dynamics *(b)* Regret vs. Information Gain *(c)* Exploration Trajectory

Figure 4. **Empirical validation of theoretical alignment.** (a) Average regret over time (log-log scale) closely adheres to the theoretical $\mathcal{O}(T^{-1/2})$ sublinear decay bound. (b) The positive correlation between regret and information gain confirms the dynamic shift from high-uncertainty exploration to low-regret exploitation. (c) The projected trajectory on the NHO landscape demonstrates this strategic behavior, where `PiFlow` traverses low-value regions to escape local optima before converging to the global peak.

consumption while significantly boosting performance. Crucially, the strategic overhead is minimal, with the `PiFlow` module itself consuming only $\sim$**1.5%** of the total token budget (see Appendix J).

❻ Ablations of LLMs and λ. `PiFlow` proves robust across diverse LLMs (e.g., Gemini-2.5-Pro (Comanici et al., 2025)), converting base reasoning capability into improved optimization outcomes. Furthermore, its behavior is predictably tunable via the trade-off parameter λ (see Appendix L & N).

❼ Ablation of Internal Dynamics in `PiFlow`. As shown in Figure 9, `PiFlow` dynamically modulates the importance of scientific principles. It effectively discards diminishing avenues while elevating under-explored principles that reveal high information gain later, ensuring an adaptive exploration-exploitation balance (see Appendix K).

5.4. Theoretical-Empirical Alignment

The Min-Max optimization in Eq. 1 provides two theoretical guarantees (Theorem 3.2): (1) sublinear regret decay of $\mathcal{O}(\sqrt{T})$, and (2) a positive correlation between information gain and regret. We empirically validate both properties on the NHO task, as visualized in Figure 4.

(Alignment 1) Sublinear Regret Decay. The first term of Eq. 1, $\sum_{t=1}^{T}(v^* - f^*(h_t))$, represents cumulative regret. Theorem 3.2 guarantees that the average regret decays as $\mathcal{O}(T^{-1/2})$. Figure 4a validates this on a log-log scale: fitting yields $c \cdot T^{-0.5}$ with $c = 1.37$ and $r^2 = 0.96$, confirming strong alignment. Notably, Run ID=1 exhibits a transient regret increase (iter 8–12) before resuming decay—this reflects the $\max_{f^*}$ adversarial term actively escaping local optima, a hallmark of the Min-Max strategy's robustness.

(Alignment 2) Information Gain–Regret Correlation. The second term of Eq. 1, $-\lambda \cdot I(h_t; f^*|H_{t-1})$, maximizes mutual information for exploration. Theorem 3.2 predicts that as information gain decreases (uncertainty resolved), regret also decreases (exploitation improves). Figure 4b confirms this positive correlation: early iterations cluster in the high-gain/high-regret region (exploration phase), while later iterations concentrate in the low-gain/low-regret region (exploitation phase), demonstrating the dynamic transition governed by the trade-off parameter λ.

(Alignment 3) Trajectory Dynamics. Figure 4c visualizes the agent's path on the PCA-projected NHO landscape. The trajectory exhibits three distinct phases that directly correspond to `PiFlow`'s strategic actions (Section 3.3): (i) *Explore* (iter 1–16): broad sampling to maximize $I(h_t; f^*|H_{t-1})$; (ii) *Escape* (iter 16–21): traversing a low-value "valley" to avoid local optima, reflecting the $\max_{f^*}$ adversarial robustness; and (iii) *Exploit* (iter 21–24): convergence to the global peak, minimizing regret. This empirical behavior demonstrates precise alignment with the theoretical Min-Max framework.

Takeaway: Empirical analysis across all three metrics, regret decay ($r^2 = 0.96$), information-regret correlation, and trajectory dynamics, corroborates the theoretical guarantees of Theorem 3.2, validating that `PiFlow`'s Min-Max optimization operates as designed.

6. Conclusion

We introduce `PiFlow`, a Plug-and-Play framework designed to overcome the inefficiencies of unguided scientific exploration. Underpinned by a Min-Max optimization with theoretical regret guarantees, `PiFlow` strategically balances the exploitation of high-potential scientific principles with the exploration of epistemic uncertainty. Extensive evaluations confirm its capacity to navigate complex hypothesis spaces, significantly outperforming baselines and establishing a robust, theoretically grounded paradigm for autonomous scientific discovery.

Impact Statement

This paper presents work whose goal is to advance the field of autonomous scientific discovery. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

AI4Science, M. R. and Quantum, M. The impact of large language models on scientific discovery: a preliminary study using gpt-4. *ArXiv*, abs/2311.07361, 2023.

Alkan, A. K., Sourav, S., Jabłońska, M., Astarita, S., Chakrabarty, R., Garuda, N., Khetarpal, P., Pi'oro, M., Tanoglidis, D., Iyer, K. G., Polimera, M. S., Smith, M. J., Ghosal, T., Huertas-Company, M., Kruk, S., Schawinski, K., and Ciucua, I. A survey on hypothesis generation for scientific discovery in the era of large language models. 2025.

Ansari, M., Watchorn, J., Brown, C. E., and Brown, J. S. dziner: Rational inverse design of materials with ai agents. 2024.

Anthropic. Claude 3.7 sonnet, 2025. URL https://www.anthropic.com/claude/sonnet.

Averly, R., Baker, F. N., and Ning, X. Liddia: Language-based intelligent drug discovery agent. *ArXiv*, abs/2502.13959, 2025.

Baek, J., Jauhar, S. K., Cucerzan, S., and Hwang, S. J. Researchagent: Iterative research idea generation over scientific literature with large language models. *ArXiv*, abs/2404.07738, 2024.

Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

DeepMind, G. Gemini 2.5 pro experimental, 2025. URL https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/.

Dolui, K., Conway, L. J., Heil, C., Strobel, T. A., Prasankumar, R. P., and Pickard, C. J. Feasible route to high-temperature ambient-pressure hydride superconductivity. *Physical review letters*, 132 16:166001, 2023.

Eisape, T., Tessler, M., Dasgupta, I., Sha, F., van Steenkiste, S., and Linzen, T. A systematic comparison of syllogistic reasoning in humans and language models. *ArXiv*, abs/2311.00445, 2023.

Eythorsson, D. and Clark, M. Toward automated scientific discovery in hydrology: The opportunities and dangers of ai augmented research frameworks. *Hydrological Processes*, 2025.

Ghafarollahi, A. and Buehler, M. J. Atomagents: Alloy design and discovery through physics-aware multi-modal multi-agent artificial intelligence. *ArXiv*, abs/2407.10022, 2024a.

Ghafarollahi, A. and Buehler, M. J. Protagents: protein discovery via large language model multi-agent collaborations combining physics and machine learning. *Digital Discovery*, 3:1389 – 1409, 2024b.

Ghafarollahi, A. and Buehler, M. J. Rapid and automated alloy design with graph neural network-powered llm-driven multi-agent systems. *ArXiv*, abs/2410.13768, 2024c.

Ghafarollahi, A. and Buehler, M. J. Sciagents: Automating scientific discovery through multi-agent intelligent graph reasoning. *ArXiv*, abs/2409.05556, 2024d.

Ghafarollahi, A. and Buehler, M. J. Automating alloy design and discovery with physics-aware multimodal multiagent ai. *Proceedings of the National Academy of Sciences of the United States of America*, 122 4:e2414074122, 2025.

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., Saab, K., Popovici, D., Blum, J., Zhang, F., Chou, K., Hassidim, A., Gokturk, B., Vahdat, A., Kohli, P., Matias, Y., Carroll, A., Kulkarni, K., Tomasev, N., Guan, Y., Dhillon, V., Vaishnav, E. D., Lee, B., Costa, T. R. D., Penadés, J. R., Peltz, G., Xu, Y., Pawlosky, A., Karthikesalingam, A., and Natarajan, V. Towards an ai co-scientist, 2025.

Greenfield, J. L., Wade, J., Brandt, J. R., Shi, X., Penfold, T. J., and Fuchter, M. J. Pathways to increase the dissymmetry in the interaction of chiral light and chiral molecules†. *Chemical Science*, 12:8589 – 8602, 2021.

Gridach, M., Nanavati, J., Abidine, K. Z. E., Mendes, L., and Mack, C. Agentic ai for scientific discovery: A survey of progress, challenges, and future directions. *arXiv preprint arXiv:2503.08979*, 2025.

Hamidieh, K. A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, 2018.

Han, Y., Wan, Z., Chen, L., Yu, K., and Chen, X. From generalist to specialist: A survey of large language models for chemistry. *ArXiv*, abs/2412.19994, 2024.

Hao, S., Gu, Y., Ma, H., Hong, J. J., Wang, Z., Wang, D. Z., and Hu, Z. Reasoning with language model is planning with world model. *ArXiv*, abs/2305.14992, 2023.

He, J., Treude, C., and Lo, D. Llm-based multi-agent systems for software engineering: Vision and the road ahead. *ArXiv*, abs/2404.04834, 2024.

Inoue, Y., Song, T., and Fu, T. Drugagent: Explainable drug repurposing agent with large language model-based reasoning. *ArXiv*, abs/2408.13378, 2024.

Jaiswal, R., Jain, D., Popat, H., Anand, A., Dharmadhikari, A., Marathe, A., and Shah, R. R. Improving physics reasoning in large language models using mixture of refinement agents. *ArXiv*, abs/2412.00821, 2024.

Jia, Y., Hou, X., Wang, Z., and Hu, X. Machine learning boosts the design and discovery of nanomaterials. *ACS Sustainable Chemistry & Engineering*, 9:6130–6147, 2021.

Kumbhar, S., Mishra, V., Coutinho, K., Handa, D., Iquebal, A., and Baral, C. Hypothesis generation for materials discovery and design using goal-driven and constraint-guided llm agents. *ArXiv*, abs/2501.13299, 2025.

Lai, R. Z. and Pu, Y. Prim: Principle-inspired material discovery through multi-agent collaboration. In *AI for Accelerated Materials Design - ICLR 2025*, 2025.

Lenselink, E. B., ten Dijke, N., Bongers, B. J., Papadatos, G., van Vlijmen, H. W. T., Kowalczyk, W., IJzerman, A. P., and van Westen, G. V. Beyond the hype: deep neural networks outperform established methods using a chembl bioactivity benchmark set. *Journal of Cheminformatics*, 9, 2017.

Li, H., Chong, Y. Q., Stepputtis, S., Campbell, J., Hughes, D., Lewis, M., and Sycara, K. P. Theory of mind for multi-agent collaboration via large language models. In *Conference on Empirical Methods in Natural Language Processing*, 2023.

Liu, S., Lu, Y., Chen, S., Hu, X., Zhao, J., Fu, T., and Zhao, Y. Drugagent: Automating ai-aided drug discovery programming through llm multi-agent collaboration. *ArXiv*, abs/2411.15692, 2024.

Lu, C., Lu, C., Lange, R. T., Foerster, J. N., Clune, J., and Ha, D. The ai scientist: Towards fully automated open-ended scientific discovery. *ArXiv*, abs/2408.06292, 2024.

Luo, Z., Yang, Z., Xu, Z., Yang, W., and Du, X. Llm4sr: A survey on large language models for scientific research. 2025.

Ma, P., Wang, T.-H., Guo, M., Sun, Z., Tenenbaum, J. B., Rus, D., Gan, C., and Matusik, W. Llm and simulation as bilevel optimizers: A new paradigm to advance physical scientific discovery. *ArXiv*, abs/2405.09783, 2024.

Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M. N., Wegner, J. K., Ceulemans, H., Clevert, D.-A., and Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on chembl. *Chemical Science*, 9:5441 – 5451, 2018.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M. A., Socher, R., Amatriain, X., and Gao, J. Large language models: A survey. *ArXiv*, abs/2402.06196, 2024.

Mitchener, L., Yiu, A., Chang, B., Bourdenx, M., Nadolski, T., Sulovari, A., Landsness, E. C., Barabási, D. L., Narayanan, S., Evans, N., Reddy, S., Foiani, M. S., Kamal, A., Shriver, L. P., Cao, F., Wassie, A. T., Laurent, J. M., Melville-Green, E., Ramos, M. C., Bou, A., Roberts, K. F., Zagorac, S., Orr, T. C., Orr, M. E., Zwezdaryk, K. J., Ghareeb, A. E., McCoy, L., Gomes, B., Ashley, E. A., Duff, K. E., Buonassisi, T., Rainforth, T., Bateman, R. J., Skarlinski, M., Rodriques, S. G., Hinks, M. M., and White, A. D. Kosmos: An ai scientist for autonomous discovery. *ArXiv*, abs/2511.02824, 2025.

Mu, Y., Liu, W., Lu, T., Li, J., Gao, S., and Wang, Z. Runtime verification of self-adaptive multi-agent system using probabilistic timed automata. *J. Intell. Fuzzy Syst.*, 45:10305–10322, 2023.

Nagarajan, V., Shi, G., Horai, R., Yu, C.-R., Gopalakrishnan, J., Yadav, M., Liew, M. H., Gentilucci, C., and Caspi, R. R. Ian: An intelligent system for omics data analysis and discovery. *bioRxiv*, 2025.

Ni, B. and Buehler, M. J. Mechagents: Large language model multi-agent collaborations can solve mechanics problems, generate new data, and integrate knowledge. *ArXiv*, abs/2311.08166, 2023.

OpenAI. Gpt-4.1, 2025. URL https://openai.com/index/gpt-4-1/.

Poincaré, H. Science and hypothesis. 1906.

Prabhakar, V., Islam, M. A., Atanas, A., Wang, Y.-T., Han, J., Jhunjhunwala, A., Apte, R., Clark, R., Xu, K., Wang, Z., and Liu, K. Omniscience: A domain-specialized llm for scientific reasoning and discovery. 2025.

Pu, Y., Huang, L., Lin, T., and Chen, H. Leveraging large language models for explaining material synthesis mechanisms: The foundation of materials discovery. In *AI for Accelerated Materials Design - NeurIPS 2024*, November 2024.

Ramirez-Medina, J., Ataei, M., and Amirfazli, A. Accelerating scientific research through a multi-llm framework. 2025.

Ramos, M. C., Collison, C. J., and White, A. D. A review of large language models and autonomous agents in chemistry. *Chemical science*, 2024.

Reddy, C. K. and Shojaee, P. Towards scientific discovery with generative ai: Progress, opportunities, and challenges. *ArXiv*, abs/2412.11427, 2024.

Ren, S., Jian, P., Ren, Z., Leng, C., Xie, C., and Zhang, J. Towards scientific intelligence: A survey of llm-based scientific agents. *ArXiv*, abs/2503.24047, 2025.

Schmidgall, S., Su, Y., Wang, Z., Sun, X., Wu, J., Yu, X., Liu, J., Liu, Z., and Barsoum, E. Agent laboratory: Using llm agents as research assistants. *ArXiv*, abs/2501.04227, 2025.

Su, H., Chen, R., Tang, S., Yin, Z., Zheng, X., Li, J., Qi, B., Wu, Q., Li, H., Ouyang, W., Torr, P., Zhou, B., and Dong, N. Many heads are better than one: Improved scientific idea generation by a llm-based multi-agent system. 2024.

Takahara, I., Mizoguchi, T., and Liu, B. Accelerated inorganic materials design with generative ai agents. 2025.

Tang, J., Xia, L., Li, Z., and Huang, C. Ai-researcher: Autonomous scientific innovation. *ArXiv*, abs/2505.18705, 2025.

Tran, K.-T., Dao, D., Nguyen, M.-D., Pham, Q.-V., O'Sullivan, B., and Nguyen, H. D. Multi-agent collaboration mechanisms: A survey of llms. *ArXiv*, abs/2501.06322, 2025.

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P. K., Spitzer, M., and Zhao, S. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18:463 – 477, 2019.

Viatkin, D., Garcia-Zapirain, B., Méndez-Zorrilla, A., and Zakharov, M. A. Deep learning approach for prediction of critical temperature of superconductor materials described by chemical formulas. In *Frontiers in Materials*, 2021.

Wan, Y., Xie, T., Wu, N., Zhang, W., Kit, C., and Hoex, B. From tokens to materials: Leveraging language models for scientific discovery. *ArXiv*, abs/2410.16165, 2024.

Wang, L., Ma, C., Feng, X., Zhang, Z., ran Yang, H., Zhang, J., Chen, Z.-Y., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., and rong Wen, J. A survey on large language model based autonomous agents. *Frontiers Comput. Sci.*, 18:186345, 2023.

Wu, J., Pu, Y., Wang, J., Gu, B., Chen, X., and Chen, H. Machine learned structure-property correlation between nanohelices and circular dichroism. *Advanced Optical Materials*, 2025.

Xiao, Y., Liu, J., Zheng, Y., Xie, X., Hao, J., Li, M., Wang, R., Ni, F., Li, Y., Luo, J., Jiao, S., and Peng, J. Cellagent: An llm-driven multi-agent framework for automated single-cell data analysis. *bioRxiv*, 2024.

Xie, T., Wan, Y., Huang, W., Yin, Z., Liu, Y., Wang, S., Linghu, Q., Kit, C., Grazian, C., Zhang, W., Razzak, I., and Hoex, B. Darwin series: Domain specific large language models for natural science. *ArXiv*, abs/2308.13565, 2023a.

Xie, Y., Feng, S., Deng, L., Cai, A., Gan, L., Jiang, Z., Yang, P., Ye, G., Liu, Z., Wen, L., Zhu, Q., Zhang, W., Zhang, Z., Li, J., Feng, Z., Zhang, C., Du, W., Xu, L., Jiang, J., Chen, X., and Zou, G. Inverse design of chiral functional films by a robotic ai-guided system. *Nature Communications*, 14, 2023b.

Xiong, W., Song, Y., Dong, Q., Zhao, B., Song, F., Wang, X., and Li, S. Mpo: Boosting llm agents with meta plan optimization. *ArXiv*, abs/2503.02682, 2025.

Xu, L., Hu, Z., Zhou, D., Ren, H., Dong, Z., Keutzer, K., Ng, S.-K., and Feng, J. Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. In *Conference on Empirical Methods in Natural Language Processing*, 2023.

Yamada, Y., Lange, R. T., Lu, C., Hu, S., Lu, C., Foerster, J. N., Clune, J., and Ha, D. The ai scientist-v2: Workshop-level automated scientific discovery via agentic tree search. *ArXiv*, abs/2504.08066, 2025.

Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., Dong, G., Wei, H., Lin, H., Tang, J., Wang, J., Yang, J., Tu, J., Zhang, J., Ma, J., Xu, J., Zhou, J., Bai, J., He, J., Lin, J., Dang, K., Lu, K., Chen, K.-Y., Yang, K., Li, M., Xue, M., Ni, N., Zhang, P., Wang, P., Peng, R., Men, R., Gao, R., Lin, R., Wang, S., Bai, S., Tan, S., Zhu, T., Li, T., Liu, T., Ge, W., Deng, X., Zhou, X., Ren, X., Zhang, X., Wei, X., Ren, X., Fan, Y., Yao, Y., Zhang, Y., Wan, Y., Chu, Y., Cui, Z., Zhang, Z., and Fan, Z.-W. Qwen2 technical report. *ArXiv*, abs/2407.10671, 2024a.

Yang, A. et al. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024b.

Yang, Z., Liu, W., Gao, B., Xie, T., Li, Y., Ouyang, W., Poria, S., Cambria, E., and Zhou, D. Moose-chem: Large language models for rediscovering unseen chemistry scientific hypotheses. *ArXiv*, abs/2410.07076, 2024c.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. React: Synergizing reasoning and acting in language models. *ArXiv*, abs/2210.03629, 2022.

Yu, B., Baker, F. N., Chen, Z., Herb, G., Gou, B., Adu-Ampratwum, D., Ning, X., and Sun, H. Chemtoolagent: The impact of tools on language agents for chemistry problem solving. 2024.

Zdrazil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D. M., Mosquera, J. F. M., Magariños, M. P., Bosc, N., Arcila, R., Kizilören, T., Gaulton, A., Bento, A. P., Adasme, M. F., Monecke, P., Landrum, G. A., and Leach, A. R. The chembl database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52:D1180 – D1192, 2023.

Zhang, H., Song, Y., Hou, Z., Miret, S., and Liu, B. Honeycomb: A flexible llm-based agent system for materials science. In *Conference on Empirical Methods in Natural Language Processing*, 2024a.

Zhang, J., Xu, X., Liu, R., and Deng, S. Exploring collaboration mechanisms for llm agents: A social psychology view. *ArXiv*, abs/2310.02124, 2023a.

Zhang, M., Press, O., Merrill, W., Liu, A., and Smith, N. A. How language model hallucinations can snowball. *ArXiv*, abs/2305.13534, 2023b.

Zhang, Q., Hu, Y., Yan, J., Zhang, H., Xie, X., Zhu, J., Li, H., Niu, X., Li, L., Sun, Y., and Hu, W. Large language model-based ai agent for organic semiconductor devices research. *Advanced materials*, pp. e2405163, 2024b.

Zhang, Q., Ding, K., Lv, T., Wang, X., Yin, Q., Zhang, Y., Yu, J., Wang, Y., Li, X., Xiang, Z., Xiang, Z., Wang, Z., Qin, M., Zhang, M., Zhang, J., Cui, J., Xu, R., Chen, H., Fan, X., Xing, H., and Chen, H. Scientific large language models: A survey on biological & chemical domains. *ACM Computing Surveys*, 2025.

Zhang, Y., Chen, X., Jin, B., Wang, S., Ji, S., Wang, W., and Han, J. A comprehensive survey of scientific large language models and their applications in scientific discovery. In *Conference on Empirical Methods in Natural Language Processing*, 2024c.

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M., Aladinskiy, V., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M., Asadulaev, A., Volkov, Y., Zholus, A., Rim, S., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B., Lee, L. H., Soll, R. M., Madge, D., Xing, L., Guo, T., and Aspuru-Guzik, A. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature Biotechnology*, 37:1038 – 1040, 2019.

Zhou, Y., Liu, H., Srivastava, T., Mei, H., and Tan, C. Hypothesis generation with large language models. *ArXiv*, abs/2404.04326, 2024.

Zhu, H., Zhou, R., Cao, D., Tang, J., and Li, M. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nature Communications*, 14, 2023.

A. Detailed Methodological Review

A hallmark and fundamental limitation of many contemporary LLM-based agent systems in scientific discovery is their limited generalizability. As detailed in Table 4, these frameworks are often characterized by a **tight coupling between their core logic and a specific scientific domain**. This means their implementations, tool integrations, and most critically, their prompt engineering strategies are meticulously tailored for a single scenario, such as organic chemistry or materials science. Although systems like The AI Scientist (Lu et al., 2024) and Agent Laboratory (Schmidgall et al., 2025) demonstrate strong capabilities in scientific research, they are specifically designed for AI domain, along with a focus on the whole workflow rather than strategic decision-making problem. Consequently, transferring these systems to a new domain necessitates significant re-engineering, restricting their out-of-the-box applicability and hindering the development of truly universal systems.

In contrast, our `PiFlow` is designed to overcome this challenge by **decoupling the strategic decision-making layer from the task execution layer**. By architecting `PiFlow` as a domain-agnostic, Plug-and-Play module, it provides strategic guidance to a minimal hypothesis-testing MAS without embedding domain-specific knowledge within its own logic. This architectural choice yields superior flexibility and obviates the need for extensive, domain-specific prompt engineering for the strategic component, enabling seamless adaptation across diverse scientific fields.

Take-away: `PiFlow` introduces a domain-agnostic, plug-and-play architecture by **decoupling** strategic decision-making from domain-specific execution. This design overcomes the critical generalizability limitations inherent in tightly-coupled scientific agent systems.

B. Distinction from Advanced Prompt Engineering

We define a scientific principle following Definition 3.1. Formally, each principle is represented as a structured text proposition that can be algorithmically scored by our MinMax optimization (as exampled in Figure 1). Unlike arbitrary text strings, these principles are supposed to be with logical consistency, and this structural requirement is what distinguishes them from general prompts. In short, principles represent foundational concepts or established patterns within a domain (e.g., "higher hydrophobicity often correlates with better cell membrane penetration"). These "principles", whether has been validated or not, can be proposed by experts or, crucially, extracted from the LLM's own vast pre-trained knowledge. They serve as high-level starting points to generate specific, testable hypotheses.

The core distinction of our `PiFlow` from prompt engineering lies not in the format of the guidance (which is text), but in the algorithmic generation of that guidance:

- **Prompt engineering baseline.** The agent quickly became trapped in local optima. Its process was highly repetitive (e.g., repeatedly stating "The system's behavior is governed by the interplay..."), and it eventually stagnated, making only meaningless tweaks to the g-factor (e.g., $1.1012 \rightarrow \cdots \rightarrow 1.1030$). This demonstrates the limitation of a static, unguided hypothesis-testing loop.
- **Systematic breakthroughs with `PiFlow`.** In contrast, the PiFlow-guided agent demonstrated structured, cumulative learning:
  1. **(Early 1 3 iterations) Principled exploration**: It begins with diverse hypotheses to maximize information gain.
  2. **(Medium stage) Discovery**: It identified non-monotonic relationships (e.g., "deviations beyond an optimal configuration reduce chirality"), forming a natural language-based identification.
  3. **(Late stage) Paradigm Shift**: Ultimately, it synthesized the principle of "minimal radius + maximal turns + optimized pitch", causing a decisive shift in the search space from (fiber_radius=40, helix_radius=70) to (fiber_radius=20, helix_radius=20) and identifying the non-linear sensitivity of the pitch parameter. This unlocked the significant g-factor improvement ($0.84 \rightarrow 1.28 \rightarrow 1.41 \rightarrow 1.51$).

The core of `PiFlow` is to force the LLM to structure its disorganized internal knowledge into explicit, falsifiable hypotheses. The `PiFlow` then uses quantitative feedback from experiments to iteratively refine its understanding regarding the task. The value is not in feeding the LLM new knowledge, but in providing a strategic framework to systematically test and organize the knowledge it already possesses, yielding a loop of *LLM Knowledge* $\rightarrow$ *External Evidence* $\rightarrow$ *Guided LLM Action*, rather than *LLM* $\rightarrow$ *LLM*.

Table 4. **Comparison of Domain Coupling and Generalizability.** Unlike domain-specific frameworks that tightly couple implementation with domain logic (e.g., specific to drug discovery), `PiFlow` adopts a decoupled architecture, enabling universal applicability across diverse scientific fields without re-engineering.

Method Name	Core Implementation	Adaptability
The AI Scientist (Lu et al., 2024)	Generating novel research ideas, writes code, executes experiments, visualizes results, describes its findings by writing a full scientific paper	AI research
Agent Laboratory (Schmidgall et al., 2025)	Accepting a human-provided research idea and progressing through three stages – literature review, experimentation, and report writing to produce comprehensive research outputs, including a code repository and a research report	AI research
CellAgent (Xiao et al., 2024)	Constructing LLM-driven biological expert roles - planner, executor, and evaluator - each with specific responsibilities	scRNA-seq Analysis
DrugAgent (Liu et al., 2024)	Employing an LLM Planner that formulates high-level ideas and an LLM Instructor that identifies and integrates domain knowledge when implementing those ideas	Drug discovery
IAN (Nagarajan et al., 2025)	Leveraging popular pathway and regulatory datasets for protein-protein interactions to perform analysis through a LLM-based multi-agent architecture	Protein-protein interactions analysis
(Ghafarollahi & Buehler, 2025)	Automating alloy design and discovery using physics-aware multimodal multi-agent AI	Alloy design and discovery
(Takahara et al., 2025)	Accelerating inorganic materials design using generative AI agents.	Inorganic materials design
LIDDIA (Averly et al., 2025)	Language-based intelligent agent for drug discovery tasks.	Drug discovery
dZiner (Ansari et al., 2024)	Rational inverse design of materials facilitated by AI agents.	Inverse design of materials
MOOSE-Chem (Yang et al., 2024c)	Utilizing Large Language Models for rediscovering unseen chemistry scientific hypotheses.	Chemistry hypothesis discovery
OmniScience (Prabhakar et al., 2025)	Domain adaptive pretraining, instruction tuning and reasoning-based knowledge distillation	Scientific reasoning
(Zhang et al., 2024b)	Large Language Model-Based AI Agent for research in organic semiconductor devices	Organic semiconductor device research
DrugAgent (Inoue et al., 2024)	Explainable drug repurposing agent with Large Language Model-based reasoning	Drug repurposing
ProtAgents (Ghafarollahi & Buehler, 2024b)	Protein discovery via large language model multi-agent collaborations combining physics and machine learning	Protein discovery
PiFlow (ours)	A unified framework for optimizing and searching-oriented scientific discovery problems	Board area in science

In summary, the algorithmic core lies in Equation 1 and Algorithm 1: the Min-Max optimization systematically balances regret minimization with information gain maximization. This is operationalized through **dynamic principle scoring that updates based on accumulated evidence** $\mathcal{T}t = \langle p_k, y_k \rangle k = 1^t$. Advanced prompt engineering lacks this principled mathematical framework for evidence integration and strategic trade-offs.

Take-away: Instead of relying on static prompts that lead to local optima, `PiFlow` employs a Min-Max optimization to algorithmically generate and refine structured, falsifiable principles. By systematically integrating experimental feedback, it transforms the LLM's latent knowledge into a dynamic, self-correcting engine for scientific discovery, enabling cumulative learning and strategic breakthroughs.

C. Proof of the Theorem

We recall the elements here and formally prove the convergence along with boundary of the system. Here we denote π is the language model policy from policy space Π, f^* is the acquisition function from function space $\mathcal{F}$, h_t is the hypothesis at time step t, $H_{t-1} = \{h_1, h_2, \ldots, h_{t-1}\}$ is the history of hypotheses, v^* is the SQ achievable by any hypothesis, and $I(h_t; f^* \mid H_{t-1})$ is the conditional mutual information.

According to the original formulation (Eq. 1), the cumulative regret can be expressed as

$$R_T(\pi, f^*) = \mathbb{E}_\pi \left[\sum_{t=1}^{T} (v^* - f^*(h_t)) \right]$$

and the cumulative information gain is

$$IG_T(\pi, f^*) = \mathbb{E}_\pi \left[\sum_{t=1}^{T} I(h_t; f^* | H_{t-1}) \right]$$

Information gain approaches zero. With information theory, the mutual information can be written as:

$$I(h_t; f^* | H_{t-1}) = H(f^* | H_{t-1}) - H(f^* | H_{t-1}, h_t)$$

A critical property for convergence is that the total information gain is bounded,

$$\sum_{t=1}^{\infty} I(h_t; f^* \mid H_{t-1}) \leq H(f^*) < \infty$$

This follows from the chain rule of mutual information,

$$\sum_{t=1}^{T} I(h_t; f^* \mid H_{t-1}) = I(H_T; f^*) \leq H(f^*)$$

Since the entropy $H(f^*)$ is finite, the cumulative information gain is bounded regardless of how many steps T we take. This implies that:

$$\lim_{t \to \infty} I(h_t; f^* \mid H_{t-1}) = 0.$$

In other words, the marginal information gained from each new hypothesis must eventually approach zero.

As information gain decreases, the expected regret also decreases. As we mentioned before, the regret $R_T(\pi, f^*)$ is defined as:

$$R_T(\pi, f^*) = \mathbb{E}_T[v^* - f^*(h_t)].$$

Now apply Jensen's Inequality, let $X = v^* - f^*(h_t)$, we have

$$\mathbb{E}[X^2 \mid H_{t-1}] = \mathbb{E}[(v^* - f^*(h_t))^2 \mid H_{t-1}].$$

Use $\phi(x) = x^2$, as it is convex, giving

$$(\mathbb{E}[X \mid H_{t-1}])^2 \leq \mathbb{E}[X^2 \mid H_{t-1}],$$

take square roots:

$$\mathbb{E}[v^* - f^*(h_t) \mid H_{t-1}] \leq \sqrt{\mathbb{E}[(v^* - f^*(h_t))^2 \mid H_{t-1}]}.$$

Now we deal with the second moment of the regret. As the v^* is constant, therefore, with the variance formula, we have

$$\mathbb{E}\left[(v^* - f^*(h_t))^2 \mid H_{t-1}\right] = \text{Var}(f^*(h_t) \mid H_{t-1}) + (v^* - \mathbb{E}[f^*(h_t) \mid H_{t-1}])^2$$

Both terms (variance and bias) are non-negative, and our goal is to bound this expression. The variance term $\text{Var}(f^*(h_t) \mid H_{t-1})$ captures the uncertainty in $f^*(h_t)$ given the history.

Since the information theory proved that, the variance of a function can be bounded by mutual information, akin to entropy bounds $H(f^*) \leq \log(|\mathcal{F}|)$, for the first term, we have

$$\text{Var}(f^*(h_t) \mid H_{t-1}) \leq c \cdot I(h_t; f^* \mid H_{t-1}),$$

where c is constant that depends on the range of f^*, this follows because the mutual information bounds the expected variance of conditional expectations.

Since we can direct view the $f^*(h_t)$ as a random variable over the joint distribution of f^* and h_t, we can apply a concentration inequality. A standard result in information-directed sampling states that for a bounded random variable like $f^*(h_t)$ here, the second moment of the regret $(v^* - f^*(h_t))$ can be bounded as:

$$\mathbb{E}\left[(v^* - f^*(h_t))^2 \mid H_{t-1}\right] \leq c \cdot I(h_t; f^* \mid H_{t-1}),$$

where the constant c depends on $|\mathcal{F}|$.

Finally, we get the inequality of $R_T(\pi, f^*)$:

$$R_T(\pi, f^*) \leq \sqrt{\mathbb{E}_T\left[(v^* - f^*(h_t))^2 \mid H_{t-1}\right]} \leq c \cdot \sqrt{I(h_t; f^* \mid H_{t-1})}$$

This inequality demonstrates that as the information gain $I(h_t; f^* \mid H_{t-1})$ decreases, the expected regret also diminishes.

The cumulative regret grows at a rate $O(\sqrt{T})$. Based on the inequality in the last step and to find the cumulative regret bound, we need to sum this inequality over all time steps:

$$\sum_{t=1}^{T} R_t(\pi, f^*) \leq c \cdot \sum_{t=1}^{T} \sqrt{I(h_t; f^* | H_{t-1})}$$

Applying the Cauchy-Schwartz inequality:

$$\sum_{t=1}^{T} \sqrt{I(h_t; f^* | H_{t-1})} \leq \sqrt{T \cdot \sum_{t=1}^{T} I(h_t; f^* | H_{t-1})}$$

Since we've already established that the total information gain is bounded:

$$\sum_{t=1}^{T} I(h_t; f^* | H_{t-1}) \leq H(f^*)$$

We can substitute this bound:

$$\sum_{t=1}^{T} \sqrt{I(h_t; f^* | H_{t-1})} \leq \sqrt{T \cdot H(f^*)}$$

Therefore, the cumulative regret is bounded by:

$$\sum_{t=1}^{T} R_t(\pi, f^*) \leq c \cdot \sqrt{T \cdot H(f^*)}$$

This demonstrates that the cumulative regret grows at a rate of $\mathcal{O}(\sqrt{T})$, which is sublinear in T. This result implies that while the total regret increases with step, the average regret per time step decreases at a rate of $\mathcal{O}(\frac{1}{\sqrt{T}})$.

D. Algorithmic Realization of the Min-Max Framework

D.1. Practical Proxies for Regret and Information Gain

Algorithm 1 provides a computationally tractable implementation of the abstract Min-Max optimization strategy presented in Eq. 1. It translates the theoretical concepts of regret minimization (exploitation) and information gain maximization (exploration) into concrete, efficiently computable metrics, enabling `PiFlow` to guide the scientific discovery process effectively. The bridge between theory and practice is established as follows:

Approximating exploitation (minimizing cumulative regret). The theoretical objective of minimizing cumulative regret, $\sum_{t=1}^{T}(v^* - f^*(h_t))$, is centered on favoring principles that consistently yield high-value outcomes $f^*(h)$. In our practical implementation, the `compute_exploitation_scores` function directly approximates this goal. It leverages the historical trajectory of principle-outcome pairs, $\mathcal{T}_t = \{\langle p_k, y_k \rangle\}_{k=1}^{t}$. The observed outcome y_k here serves as a direct proxy for the performance of its corresponding principle p_k. A principle associated with higher outcomes is considered to have lower regret. Therefore, the resulting normalized score vector ranging from 0 to 1, $S_{exploitation} \in \mathbb{R}^{|\mathcal{T}_t|}$, quantifies the empirical success of each principle, and maximizing this score is equivalent to minimizing the cumulative regret based on past evidence.

Approximating exploration (maximizing information gain). The second term in our objective, maximizing the mutual information $I(h_t; f^*|H_{t-1})$, encourages the selection of hypotheses that are most informative about the underlying scientific landscape. A direct computation of mutual information is often intractable. Consequently, Algorithm 1 employs a practical proxy via the `compute_exploration_scores` function. **This is achieved by computing the semantic distance of principles using their sentence level embeddings (e.g., obtained from QwenMax model).** We posit that principles which are semantically distant from those already tested are more likely to reveal novel information about the function f^*. Therefore, we use the cosine distance between a candidate principle's embedding and the embeddings of previously explored principles as a heuristic for information gain. High exploration scores, $S_{exploration} \in \mathbb{R}^{|\mathcal{T}_t|}$, are assigned to conceptually novel principles. This heuristic is theoretically grounded, as detailed in Appendix D.2.

Algorithm 1 Algorithm of `PiFlow`.

Input: $\mathcal{T}_t$, and λ_{factor}
Output: $suggestion$ (strategic action recommendation)
if $|\mathcal{T}_t| < 3$ **then**
  $suggestion \leftarrow$ "Initialize one principle to explore."
else
  $S_{\text{exploration}} \leftarrow$ compute_exploration_scores($\mathcal{T}_t$)
  $S_{\text{exploitation}} \leftarrow$ compute_exploitation_scores($\mathcal{T}_t$)
  for $i \leftarrow 1$ **to** $|\mathcal{T}_t|$ **do**
    $S_{\text{final}}[i] \leftarrow (1 - \lambda_{factor}) \cdot S_{\text{exploration}}[i] + \lambda_{factor} \cdot S_{\text{exploitation}}[i]$
  end for
  $i_{\text{best}} \leftarrow \arg\max_i(S_{\text{final}}[i])$
  $p_{\text{best}} \leftarrow \mathcal{T}[i_{\text{best}}]$
  $best_exploitation_score \leftarrow S_{\text{exploitation}}[i_{\text{best}}]$
  if $best_exploitation_score > 0.7$ **then**
    $action_type \leftarrow$ "refine"
    $suggestion \leftarrow$ "Focus on refining: $\frown$ $p_{\text{best}}.content$
  else if $best_exploitation_score > 0.4$ **then**
    $action_type \leftarrow$ "validate"
    $suggestion \leftarrow$ "Validate: $\frown$ $p_{\text{best}}.content$
  else
    $action_type \leftarrow$ "explore"
    $suggestion \leftarrow$ "Explore alternatives: $\frown$ $p_{\text{best}}.content$
  end if
end if
Return $suggestion$

The integrated decision policy. The policy of the Min-Max framework, π, must balance the above objectives. Our algorithm materializes this policy through a two-step process:

1. **Step 1 (Scoring and selection).** The final score, $S_{final}[i] \leftarrow (1 - \lambda_{factor}) \cdot S_{exploration}[i] + \lambda_{factor} \cdot S_{exploitation}[i]$, is a direct implementation of the balanced objective function. The input parameter λ_{factor} instantiates the theoretical trade-off, controlling the emphasis between exploration and exploitation. The $\arg\max_i(S_{final}[i])$ operation then executes the policy's primary function: selecting the most promising principle, p_{best}, given the current state of knowledge and the desired strategic balance.
2. **Step 2 (Action recommendation).** Finally, the policy translates its choice into a strategic command. The threshold-based conditions on the best principle's exploitation score (if $best_exploitation_score > 0.7\ldots$) discretize the continuous space of potential actions into three clear directives: **refine**, **validate**, or **explore** with the concatenation (denoted by $\frown$) of the principle content. This transforms the numerical output of the optimization into an actionable suggestion for the Planner agent ($\mathcal{A}_P$), thereby closing the loop between theoretical deliberation and practical execution within the Hypothesis-Validation cycle.

Historical information management. We manage historical information via two independent parts: (a) agent memory for MAS to iterate the hypothesis-testing and (b) persistent principle pool for PiFlow layer to guide the process.

1. **For MAS part, agent history differs per agent.** As the MAS follows round-robin order, there is the risk of context collapse. To prevent this, each agent maintains an independent MessageBuffer, retaining the most recent $M = 10$ messages. This ensures agents focus on the immediate logical flow without being overwhelmed by noise.
2. **For `PiFlow`, as a plugin layer, it has its own memory.** Crucially, our principle-aware design decouples reasoning history from scientific insight. While chat logs may be truncated, generated principles and validation scores represent compressed, high-value knowledge stored in a separate, persistent global pool. As per Algorithm 1 (Lines 14-23), even if an agent forgets a specific past conversation, PiFlow retrieves the refined principles from the global pool to guide the next step. This ensures the logic chain remains intact throughout the discovery process.

With the design of decoupling, PiFlow only need to collect historical principles, hypotheses and outcomes to optimize the strategy (explore, refine or validate which temporary principle), and steers the learning process by giving guidance to MAS.

Realworld efficiency (stopping by condition). In practical deployment, PiFlow's Min-Max optimization naturally drives the system toward convergence, enabling precise threshold-based stopping: In the Molecular Bio-activity Optimization (MBO) task, PiFlow-MAS improved pChEMBL from 2.69 to 6.44 within just 4 effective iterations, and surpassed the target (6.5) to reach 7.65 by the 6th iteration. This demonstrates that with a practical stop condition, PiFlow achieves success using only 25% of the allocated benchmark budget. As detailed in Section 5.4, our Min-Max framework provides a theoretical guarantee that prevents infinite loops of low-quality exploration, ensuring resource efficiency even in the intensive tasks noted by the Reviewer.

D.2. The Rational of Semantic Distance as a Proxy for Information Gain

Recall that we introduce a Min-Max optimization framework for the `PiFlow` system. The objective function, as shown in Eq. 1, incorporates a mutual information term, $I(h_t; f^*|H_{t-1})$, to guide exploration. This term, while theoretically ideal, is computationally intractable as it requires knowledge of the true underlying evaluation function f^*. In our practical implementation, we employ a surrogate objective for exploration: maximizing the distance of a new principle's text embedding from the embeddings of previously selected principles. Here, we provide a formal justification for this approximation, demonstrating its theoretical soundness.

Proof. Our goal is to establish a principled connection between the practical exploration strategy and the theoretical objective of maximizing information gain. The practical strategy is to select a principle p_t from the principle space $\mathcal{P}$ at each timestep

t according to:

$$p_t^* = \arg\max_{p_t \in \mathcal{P}} \left(\min_{m \in \{1,\dots,t-1\}} \|\phi(p_t) - \phi(p_m)\|_2 \right) \tag{2}$$

where $\phi : \mathcal{P} \to \mathbb{R}^d$ is a function that maps a principle to its corresponding high-dimensional text embedding. We will now demonstrate that this objective serves as a valid proxy for maximizing the mutual information term $I(h_t; f^*|H_{t-1})$.

The derivation rests upon the hierarchical relationship between principles and hypotheses, and the semantic properties of modern language model embeddings. A principle $p \in \mathcal{P}$ is not a hypothesis itself, but rather defines a specific semantic region or a conditional distribution $\pi(\cdot|p)$ from which a concrete hypothesis $h \in \mathcal{H}$ is formulated. Thus, the selection of a principle p_t precedes the generation of a hypothesis $h_t \sim \pi(\cdot|p_t)$.

We begin by positing two fundamental premises regarding the nature of the embedding space.

Premise 1 (Semantic-metric correspondence). The embedding function ϕ is assumed to map the conceptual space of principles to a metric space where distance reflects semantic dissimilarity. That is, for any two principles $p_i, p_j \in \mathcal{P}$, a large Euclidean distance $\|\phi(p_i) - \phi(p_j)\|_2$ implies a significant divergence in their underlying semantic and conceptual content. This is a well-established property of embeddings from large-scale language models.

Premise 2 (Functional consequence of semantic dissimilarity). Semantically distinct principles guide the generation of functionally distinct hypotheses. A hypothesis h acts as a probe of the unknown function f^*. If two principles p_i and p_j are semantically distant, the hypotheses generated from their respective distributions, $h_i \sim \pi(\cdot|p_i)$ and $h_j \sim \pi(\cdot|p_j)$, are expected to probe disparate aspects or regions of the function f^*.

With these premises, we can construct the logical argument. The mutual information term $I(h_t; f^*|H_{t-1})$ quantifies the expected reduction in uncertainty about f^* after observing the outcome of hypothesis h_t, given the history of observations H_{t-1}. As established in our convergence proof (Section C), this information gain is related to the posterior variance of the outcome of h_t:

$$\mathbb{E}\left[(v^* - f^*(h_t))^2 \mid H_{t-1}\right] \leq c \cdot I(h_t; f^* \mid H_{t-1})$$

This relation suggests that **a hypothesis h_t yielding high uncertainty (i.e., high posterior variance $\mathrm{Var}(f^*(h_t) \mid H_{t-1})$) is expected to provide high information gain. Consequently, a sound exploration strategy is to select h_t to maximize this variance.**

Let us now connect this objective to our practical strategy in Eq. 2.

1. By selecting a principle p_t that maximizes the minimum distance to all prior principles in the embedding space, we are, by Premise 1, choosing a principle that is maximally semantically novel compared to the history of principles $\{p_m\}_{m=1}^{t-1}$.
2. By Premise 2, this semantically novel principle p_t will guide the generation of a hypothesis h_t that probes a functionally distinct aspect of f^* compared to all prior hypotheses in $H_{t-1} = \{(h_m, y_m)\}_{m=1}^{t-1}$.
3. Since h_t lies in a region of the hypothesis space that has not been explored by past observations, our model's posterior belief about the outcome $f^*(h_t)$ will be characterized by high uncertainty. This high uncertainty mathematically corresponds to a large posterior variance, $\mathrm{Var}(f^*(h_t) \mid H_{t-1})$.
4. Therefore, the policy of maximizing the embedding distance effectively drives the selection of hypotheses that are expected to have high posterior variance.

This chain of above analysis leads to the following correspondence:

$$\arg\max_{p_t} \left(\min_{m<t} \|\phi(p_t) - \phi(p_m)\| \right) \Rightarrow \text{Select maximally novel } p_t$$

$$\Rightarrow \text{Generate } h_t \text{ from an unexplored hypothesis subspace}$$

$$\Rightarrow \max_{h_t \sim \pi(\cdot|p_t)} \mathbb{E}[\mathrm{Var}(f^*(h_t) \mid H_{t-1})]$$

$$\Rightarrow \max_{h_t \sim \pi(\cdot|p_t)} \mathbb{E}[I(h_t; f^* \mid H_{t-1})]$$

The expectation $\mathbb{E}[\cdot]$ is taken over the generation of h_t from p_t.

In conclusion, the strategy of maximizing the minimum embedding distance between principles is not an arbitrary heuristic. It is a principled and computationally feasible surrogate for the theoretically-grounded objective of maximizing mutual information. It leverages the semantic structure captured by modern language model embeddings to implement an efficient and effective exploration strategy, ensuring that `PiFlow` diversifies its inquiry at a conceptual level. This alignment between our practical approximation and the theoretical Min-Max framework provides support for the design of our system and its empirical performance. □

Take-away: We operationalize the abstract Min-Max framework by substituting its theoretically-ideal but computationally-intractable objectives with practical proxies. **Exploitation** (regret minimization) is approximated by historical performance, while **Exploration** (information gain maximization) is driven by maximizing the semantic distance between principle embeddings. We formally prove that this semantic distance is a principled surrogate for maximizing information gain, thus bridging the gap between theory and efficient implementation.

E. Illustrative Example: Application to Nanohelix Optimization

We provide a step-by-step walk-through of `PiFlow`'s operation using the Nanohelix Optimization (NHO) task as a running example:

Let's assume the goal is to find the nanohelix geometry (described at Appendix P.1) with the maximum g-factor:

Phase 1: Initial Exploration. The process begins with an unguided hypothesis generation to build an initial evidence base by LLM itself. The Planner agent, following the initial directive of `PiFlow`, instructs the Hypothesis Agent for intuitive exploring (Algorithm 1, line 3-5). For the first K rounds (e.g., $K = 3$ in our experiments), the loop proceeds as follows,

1. **Hypothesis Agent** proposes a specific, testable hypothesis, *"Based on the principle of $\cdots$, measure the g-factor of a nanohelix with parameters: fiber_radius=40.0 nm, helix_radius=70.0 nm, n_turns=6.5, pitch=130.0 nm"*.
2. **Experiment Agent** calls the tool (surrogate model) to validate the above hypothesis and returns the outcome, "*the validation yields a g-factor of 0.86*".
3. This Hypothesis-Validation process repeats for K rounds, populating the evidence trajectory T_t for establishing an initial principle pool.

In the initial K rounds, lacking specific guiding principles, the Hypothesis Agent performs an initial exploration of the parameter space by proposing diverse geometries based on its general pre-trained knowledge, or by sampling from a wide distribution.

Phase 2: Principle-aware guidance from PiFlow. After the initial K rounds, the MinMax optimization of `PiFlow` parses these evidence and identifies the highest potential principle and suggests an action based on the score of the highest-potential principle (refer to Algorithm 1, line 5-23). The Planner agent receives this strategic guidance, for instance: "***Refine*** *the identified promising principle: By adjusting the fiber radius to twice the helix radius and fine-tuning other parameters, it is expected to maximize the chirality (g factor) of the nanohelix*".

Structured, guided hypothesizing in both Phase 1 and Phase 2. The Planner injects this directive into the whole group chatting history, i.e., Hypothesizing-Validation loop, thereby reducing the system-level uncertainty. The Hypothesis Agent now uses this focused principle to formulate its next hypothesis with structured reasoning, for example:

1. **Major premise.** The g-factor of a nanohelix is governed by the spatial asymmetry and electromagnetic coupling arising from its geometric parameters.
2. **Minor premise 1.** The previously tested geometry (fiber-radius=40 nm, helix-radius=70 nm, n-turns=6.5, pitch=130 nm) yielded a high g-factor (0.86), indicating a favorable parameter balance.

3. **Minor premise 2 (Inspired by PiFlow).** Systematically varying parameters around this near-optimal point can induce nonlinear electromagnetic effects, potentially increasing chirality. Specifically, increasing the pitch and number of turns might enhance the chiral interaction length.

4. **Proposed testable hypothesis.** "Measure the g-factor of a nanohelix with parameters: fiber_radius=49.0 nm, helix_radius=24.5.0 nm, n_turns=7.0, pitch=140.0 nm."

By this way, `PiFlow` dynamically steers the discovery process away from aimless searching and towards a focused, principle-driven exploration, thereby systematically accumulating information and avoiding inefficient exploration directions.

Takeaway: In this example, `PiFlow` transforms the optimization process from an initial, unguided exploration into a focused, principle-driven discovery. It first samples the parameter space to build an empirical evidence base, then distills a high-potential guiding principle to steer subsequent hypotheses, ensuring a systematic and efficient search for the optimal nanohelix geometry.

F. Analysis of Baselines Response

F.1. Response Analysis of Baselines

We compare our PiFlow against the ReAct, MPO and Vanilla Agent. Through the experiment, we found that the MPO baseline incorporates a form of global, step-by-step reasoning or reflection within their trained model, for example, it's exact output is, for example, *"Step 1: Identify the task objective. Step 2: Survey the environment. Step 3: Consider the first potential candidate. Step 4: Shift to the next potential candidate. Step 5: Repeat Steps 3 and 4 until the task objective is met. Step 6: Confirm the task completion."*

While ReAct responses with the reflection of "Previous Experiment Results" and then instructs the Hypothesis Agent to try another new hypothesis directly with the provided "Rationale", the baseline of Vanilla Agent, in genral, response with step-by-step thoughts following "Step 1: Define the Hypothesis, Step 2: Initial Exploration, Step 3: Parameter Space Definition, Step 4: Experimental Design, Step 4: Exact Experiments (candidates)", behaving like a careful-thought planner in the whole process of scientific discovery.

Backend these outputs, the performance difference stems from the level of dynamic and strategic guidance. For example, **the reasoning of MPO and two AI Scientsit methods are local and tactical**. MPO generates a fixed "how-to" checklist. This plan dictates the operational steps but is agnostic to the underlying scientific principles driving the experiments, and cannot reason about why its plan is failing, leading to hypotheses that yield lower outcomes. While The AI Scientist v2 and AI-Researcher logically select the "node" or candidate, their lack of hypothesis diversity often leads to **premature convergence and performance stagnation**. Additionally, the Vanilla agent uses the exact same powerful LLM (QwenMax) and has access to the same experimental tools. Its poor performance demonstrates that **even a capable LLM, without strategic guidance, explores aimlessly**. The dramatic performance gap between Vanilla and PiFlow (e.g., AUC improving from 35.96% to 63.51% in the NHO task) is direct evidence that our contributions, i.e., (a) structuring knowledge into an explicit principle space, and (b) applying a rigorous Min-Max optimization strategy to navigate it.

However, `PiFlow` adapts its learnable strategy. By optimizing (i.e., evaluating and selecting) at the principle level, `PiFlow` can make strategic jumps. When hypotheses derived from Principle A consistently fail, the Min-Max optimization quantitatively lowers the score of Principle A itself, guiding the agent to switch to a completely different research direction (Principle B). This prevents getting stuck and is the fundamental reason for its superior efficiency and performance, as demonstrated empirically in Figure 3.

F.2. Statistical Significant Test on Metrics

We compared PiFlow against AI-Researcher (Tang et al., 2025) (NeurIPS 2025) and The AI Scientist v2 (Yamada et al., 2025). As shown in Table 5, `PiFlow`'s targeted uncertainty reduction (Min-Max) proves superior to generic autonomous loops.

The results show that, **`PiFlow` achieves statistically significant improvement (p-value < 0.05) on NHO and SPO tasks.** On the MBO task, PiFlow remains competitive with higher efficiency (AUC 64.57% vs 42.72%), though SQ variance is higher. This confirms the advantage of `PiFlow` in complex, uncertainty-heavy environments.

Table 5. Performance comparison between PiFlow, The AI Scientist v2, and AI-Researcher.

Task	Method	SQ (%)	P-value (SQ)	AUC (%)	P-value (AUC)
NHO	PiFlow	**76.82 ± 4.54**	–	**63.51 ± 11.18**	–
	The AI Scientist v2	56.67 ± 5.79	0.0103	49.27 ± 1.84	0.1547
	AI-Researcher	53.12 ± 1.06	0.0091	46.45 ± 2.42	0.1122
MBO	PiFlow	84.55 ± 29.63	–	**64.57 ± 23.65**	–
	The AI Scientist v2	62.47 ± 8.02	0.3250	36.32 ± 10.62	0.1630
	AI-Researcher	**95.66 ± 7.45**	0.5870	15.56 ± 12.71	0.2625
SPO	PiFlow	**34.85 ± 1.19**	–	21.51 ± 2.80	–
	The AI Scientist v2	29.85 ± 2.68	0.0663	**28.43 ± 3.13**	0.0470
	AI-Researcher	25.69 ± 3.23	0.0277	16.36 ± 2.72	0.0842

Take-away: Baselines fail due to their reliance on rigid, tactical plans, which leads to aimless exploration when a strategy is ineffective. In contrast, PiFlow excels by operating at a higher level of abstraction. It reasons over a "*principle space*" and employs Min-Max optimization to strategically pivot away from failing research directions, enabling adaptive and efficient discovery.

G. Performance Comparison with Bayesian Optimization

To validate our hypothesis that a principle-aware architecture is superior for scientific discovery under high epistemic uncertainty, we compared PiFlow against a strong, general-purpose baseline, Bayesian Optimization (BO). It is worth noting that, unlike PiFlow, BO requires significant expert effort to manually define a parameterized optimization space. We construct the experiment by manually configuring the searching space of BO in both nanohelix optimization (NHO) task, molecular bio-activity optimization task and superconductor optimization (SPO) task.

Table 6. Performance comparison between PiFlow and BO (mean ± standard deviation)

Method	Metric	NHO	MBO	SPO
PiFlow	AUC	63.51 ± 11.18	46.11 ± 16.25	21.51 ± 2.80
	SQ	76.82 ± 4.54	84.55 ± 29.63	34.85 ± 1.19
BO	AUC	68.86 ± 5.86	34.76 ± 4.21	31.61 ± 0.19
	SQ	78.76 ± 0.77	38.15 ± 4.02	32.38 ± 0.32

Results with 24 iterations of BO, summarized in Table 1, show that while BO is competitive on the structured NHO task, it strongly needs prior design of the searching space, e.g., material element composition and quantity range. However, PiFlow demonstrates a substantial performance advantage on the more complex and uncertain MBO task, highlighting its effectiveness when the problem structure is not known a priori.

In fact, PiFlow leverages scientific principles with its dynamic uncertainty reduction architecture. This framework allows the agent to make "domain-shifting" leaps in the search space (e.g., the jump from a 60nm to a 10nm radius value), leading to rapid, significant gains that are rewarded by the AUC metric. It is this ability to build and reason with an evolving knowledge structure that our evaluation framework was designed to highlight.

In summary, PiFlow addresses the challenge of semantic hypothesis generation, where BO, Generic Algorithms and Reinforcement Learning methods face significant barriers in representation. PiFlow operates in an open-ended semantic space (natural language principles), allowing it to be Plug-and-Play without domain-specific encoding.

Take-away: PiFlow's principle-aware architecture significantly outperforms Bayesian Optimization on complex, ill-structured scientific discovery tasks (MBO). Its strength lies in dynamically building knowledge to navigate vast search spaces, whereas BO's performance is contingent on a manually-defined search space, rendering it only competitive on well-structured problems (NHO).

H. Plug-and-Play Integration with ChemToolAgent

We integrate PiFlow with ChemToolAgent (Yu et al., 2024) on the Molecular Bio-activity Optimization (MBO) task to demonstrate the generalization of PiFlow.

The integration follows a Plug-and-Play design, where PiFlow acts as a strategic guidance layer, as shown in Figure 5. ChemToolAgent serves as both the **Hypothesizer** and Experimenter, leveraging its built-in tools (such as searching molecules' formula from PubMed and website) to propose molecular hypotheses, while PiFlow's outer-loop guidance provided EXPLORE and REFINE commands to ensure systematic and efficient discovery dynamically. As shown in Table 7, this collaborative process successfully evolved its strategy from basic principles to a high-value molecular design (pChEMBL of 5.90) over 8 iterations **without any architecture-level modifications** over ChemToolAgent, demonstrating a clear synergy.

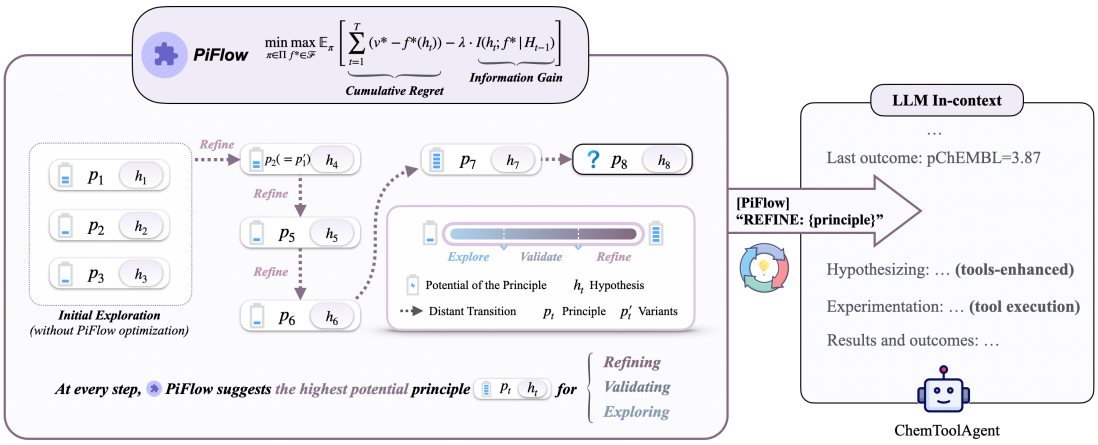


Figure 5. Illustration of combining PiFlow with existing MAS (ChemToolAgent) through Plug-and-Play.

Table 7. Key steps from the 8-iteration integration run with ChemToolAgent.

Iteration	PiFlow Action	Principle / Key Guidance	pChEMBL
1	EXPLORE	Proposed p_1: *"hydroxyl + nitrogen heterocycle for H-bonding"*	2.80
2	EXPLORE	Tested a new, unrelated principle p_2	-0.10
3	EXPLORE	ChemToolAgent generated a principle p_3 and proposed an invalid formula	null
4	REFINE on p_1	Focused on the promising principle (p_1) from the Iteration 1	2.13
5	REFINE on p_1	Proposed p_5: "Quinazoline + phenyl rings for hydrophobic interactions..."	3.87
6	REFINE on p_5	Further refined the quinazoline principle (dual EGFR/HER2 mechanism)	5.90
7-8	REFINE	Continued focused refinement on the high-potential quinazoline scaffold	∼5.2

Take-away: `PiFlow` seamlessly integrates with existing agents like ChemToolAgent as a zero-modification, plug-and-play strategic layer. By providing high-level EXPLORE and REFINE guidance, it synergistically steers the discovery process from general principles to a high-potency molecule (pChEMBL 5.90), demonstrating its strong generalization and practical value.

I. Theoretical Computational Complexity

Let t be the current number of iterations (i.e., the size of the historical trajectory) and d be the embedding dimension of the principles. A single decision step in `PiFlow` (Algorithm 1) involves:

1. Computing exploitation scores. This requires iterating through the t historical outcomes, giving a complexity of $\mathcal{O}(t)$.
2. Computing exploration scores. To assess the novelty of each of the t historical principles, the default algorithm calculates its similarity to all other $t-1$ principles. This involves $\mathcal{O}(t^2)$ vector comparisons, leading to a complexity of $\mathcal{O}(t^2 \cdot d)$.
3. Final decision. This involves a weighted sum and finding the maximum over t principles, costing $\mathcal{O}(t)$.

Therefore, the dominant computational cost for a single PiFlow decision at step t is $\mathcal{O}(t^2 \cdot d)$. While manageable for moderate trajectory $\mathcal{T}$, we recognize this can be a bottleneck for very long process.

To ensure scalability, this can be readily optimized to near-linear complexity using standard techniques, such as (a) instead of all-pairs comparison, we can find the most similar principles for approximation, aiming for reducing the exploration score computation; (b) incremental or batched updates of similarity matrix, avoiding full recalculation at every step. These optimizations make `PiFlow` computationally feasible for large-scale discovery tasks.

The decision complexity of $O(t^2 \cdot d)$ is not a bottleneck. We profiled the runtime in the NHO task. Over 24 iterations, the entire `PiFlow` framework accounted for only 9.4% of the total PiFlow-MAS runtime, while the surrogate model took 0.2%. We also quantitatively profiled the optimization efficiency in the NHO task. The result confirms that the algorithmic complexity of `PiFlow` translates into superior search efficiency: For baselines, Vanilla Agent System achieves SQ=50% taking 2416s; The other baseline MPO achieves SQ=52% taking 506s and baseline ReAct even fails to reach such level of SQ in all its 918s. However, our **`PiFlow` achieves SQ=52% (similar bar) by only 427s, delivering such a result using less than 1/5 of the Vanilla Agent's time (a 5.6x speedup), and is 1.2x faster than the baseline MPO.**

The results indicate that, the vast majority of time is consumed by LLM inference. The specific similarity calculation (t^2) takes milliseconds, which is structurally distinct from—and negligible compared to—LLM generation (seconds) or real-world wet-lab experiments (hours/days).

Take-away: The per-step complexity of `PiFlow` is $\mathcal{O}(t^2 \cdot d)$, dominated by an all-pairs similarity calculation for exploration. This quadratic cost is not a fundamental limitation, as it can be readily optimized to near-linear time, ensuring scalability.

J. Cost-Effectiveness Analysis

We performed a cost-effectiveness analysis comparing our **PiFlow-MAS** against a **Vanilla-Agent System** baseline across three discovery tasks (NHO, MBO, SPO). All experiments were run for 24 iterations using the Qwen-Max model. Table 8 compares the total token consumption (cost) and final Solution Quality (SQ) achieved, while Table 9 isolates the token cost of the PiFlow module itself to demonstrate its efficiency. Both of them are shown at Figure 6.

Takeaway: PiFlow-MAS achieves superior performance at a reduced cost. Our analysis reveals that PiFlow-MAS significantly enhances Solution Quality (SQ) by up to **2.4x** while simultaneously cutting token consumption by up to **26.7%**. This is accomplished with remarkable efficiency, as the PiFlow module itself is a lightweight plugin, constituting only **1.2-1.5%** of the total tokens.

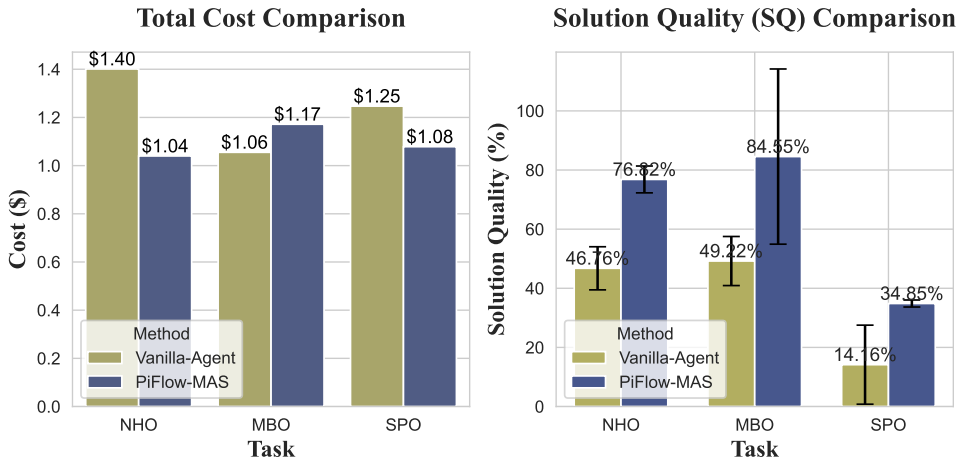


Figure 6. Cost-effectiveness and solution quality (SQ) comparision.

Table 8. Cost-effectiveness comparison between PiFlow-MAS and a Vanilla-Agent System baseline.

Task	Method	Tokens	Cost ($)	Cost/Iter ($)	Cost Reduction (%)	SQ (%)
NHO	Vanilla-Agent	806,949	$1.4011	$0.0584	-	46.76 ± 7.29
	PiFlow-MAS	591,316	$1.0400	$0.0434	26.7%	76.82 ± 4.54
MBO	Vanilla-Agent	610,208	$1.0552	$0.0440	-	49.22 ± 8.30
	PiFlow-MAS	657,594	$1.1717	$0.0493	-7.7%	84.55 ± 29.63
SPO	Vanilla-Agent	707,829	$1.2469	$0.0520	-	14.16 ± 13.37
	PiFlow-MAS	610,284	$1.0785	$0.0449	13.7%	34.85 ± 1.19

Table 9. Analysis of the PiFlow module's token efficiency as a lightweight plugin.

Task	PiFlow-MAS	PiFlow	PiFlow Token Share (%)	PiFlow Tokens/Iter
NHO	582,396	8,920	1.5%	372
MBO	649,590	8,004	1.2%	334
SPO	602,740	7,544	1.2%	314

K. Ablation: Temporal Dynamics of Principle Evaluation

We provide an empirical visualization of the internal dynamics of `PiFlow`, connecting directly to the Min-Max optimization framework detailed in Section 3.3. To illustrate how the scores for different scientific principles evolve over time, we present results from an implementation using the QwenMax model on the NHO (nanohelix optimization) benchmark. The following figures chart the iterative recalculation of principle scores, which guides the system's strategic balance between exploitation and exploration.

Figure 7 displays the dynamic final scores (S_{final}) for each principle, representing the overall potential as determined by the Min-Max optimization with $\lambda = 0.5$. These scores are not static; they fluctuate as the Hypothesis-Validation loop accumulates new evidence. **Some principles that initially appear promising see their scores decline, while others emerge as high-potential candidates over subsequent iterations.** This dynamic ranking is the direct output of `PiFlow`'s strategic analysis, steering the discovery process toward the most promising avenues at any given moment.

To better understand the final scores, Figures 8 and 9 decompose them into their constituent exploitation and exploration components, respectively.

Specifically, Figure 8 shows that the exploitation scores for most principles tend to decrease over time. This reflects the regret-minimization objective; as principles are tested and accumulated to evidence (exploited), the potential for further high-value discoveries from these experience may diminish, or the cumulative regret associated with it increases. This

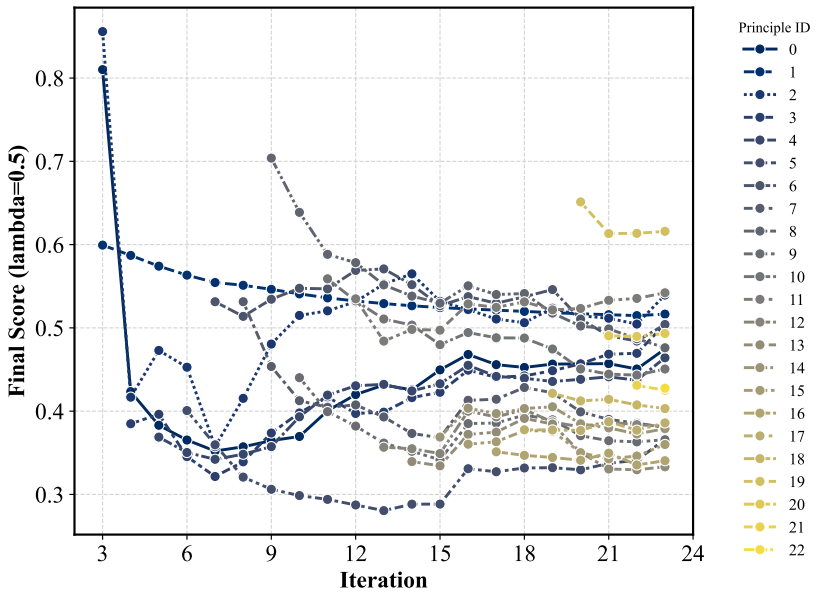


Figure 7. The dynamic final scores of principles in `PiFlow`.

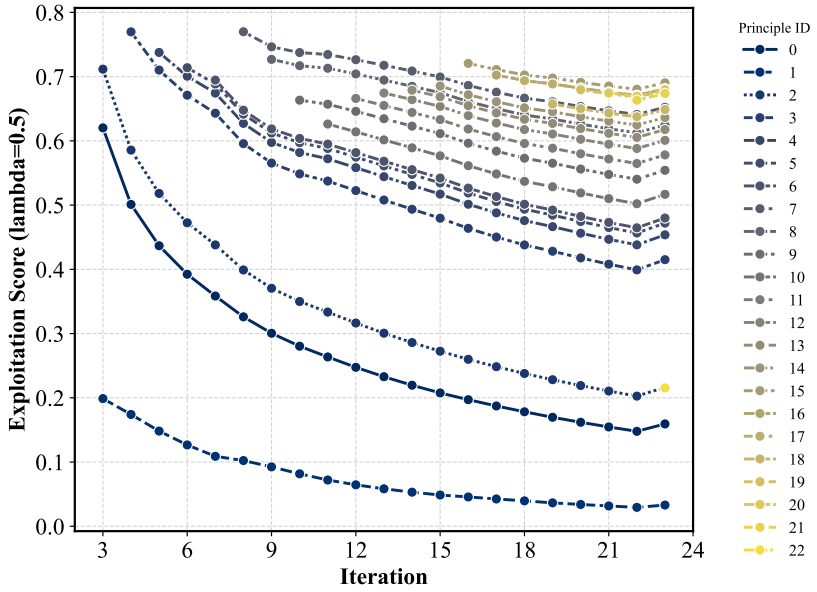


Figure 8. The dynamic exploitation scores of principles in `PiFlow`.

manifests itself **as principles generated later in the hypothesis testing process naturally have higher outcomes**, while past principles are reduced due to normalization.

In contrast, Figure 9 reveals a much more varied and dynamic behavior for the exploration scores. Principles may begin with a low exploration score and later see it increase significantly, indicating that the system has identified a knowledge gap and is prioritizing information gain to reduce uncertainty about that principle's potential.

Together, these plots illustrate the practical outcome of the adversarial optimization: a continuous, adaptive balancing act where `PiFlow` shifts its focus between exploiting known concepts and exploring uncertain ones to navigate the scientific landscape efficiently.

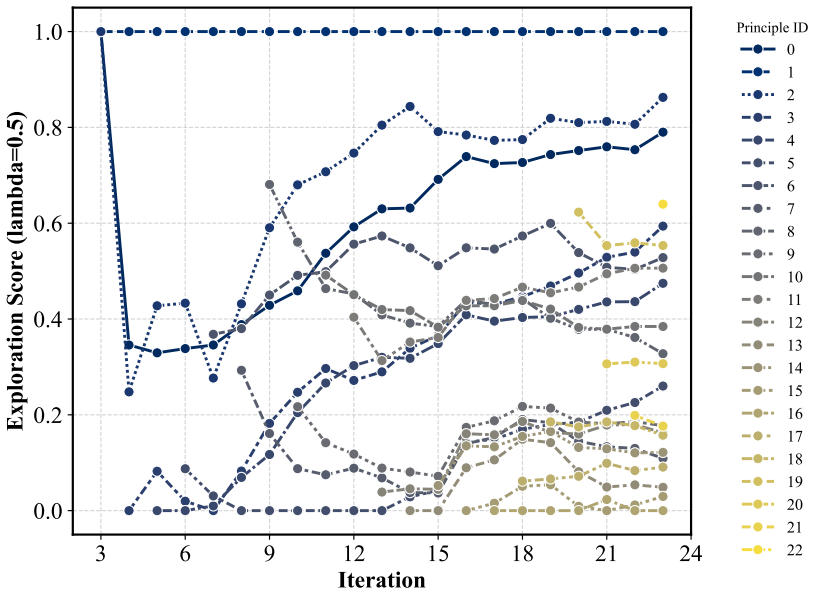


Figure 9. The dynamic exploration scores of principles in `PiFlow`.

Takeaway: Principle evaluation in `PiFlow` is a dynamic balancing act. The system continuously re-calibrates the trade-off between exploiting known concepts (diminishing returns) and exploring uncertain ones (information gain) to adaptively steer the discovery process.

L. Ablation: The Impact of Foundation Models

Influence of Foundation Models. To evaluate the impact of different models, we replace LLMs of $\mathcal{A}_H$ and $\mathcal{A}_P$ evaluation. The $\mathcal{A}_E$ is responsible for tool usage, consistently used QwenMax to ensure functional tool interaction. We evaluate several state-of-the-art LLMs (e.g., Claude-3.7-sonnet (Anthropic, 2025), GPT4.1-mini (OpenAI, 2025), Gemini-2.5-pro-exp-0325 (DeepMind, 2025), Qwen3-32B and QwenMax (Yang et al., 2024b)) on the Nanohelix Optimization (NHO) task with three times of random seeds as initialization, and the results are summarized in Table 10.

Table 10. Ablation Study of Model Types (mean ± std)

Method/Setting	**AUC (%)**	**SQ (%)**
Claude-3.7-sonnet	38.60 ±4.30	78.50 ±3.74
GPT4.1-mini	41.68 ±17.91	66.38 ±14.90
Gemini-2.5-pro-exp-03-25	28.43 ±13.81	69.64 ±17.10
Qwen3-32B	37.51 ±7.70	58.76 ±6.18
QwenMax	**63.51** ±11.18	**76.82** ±4.54

Among these models, QwenMax demonstrates the highest AUC at 63.51% and a strong SQ of 76.82%. Claude-3.7-sonnet achieves the highest SQ at 78.50% with an AUC of 38.60%. Other models like GPT4.1-mini (AUC 41.68%, SQ 66.38%) and Qwen3-32B (AUC 37.51%, SQ 58.76%) also show competent performance. Gemini-2.5-pro-exp-03-25 yields an AUC of 28.43% and an SQ of 69.64%. These quantitative results reveal that, while efficiency (AUC) is model-dependent—ranging from 38.60% (Claude-3.7) to 63.51% (QwenMax), Solution Quality (SQ) benefits consistently from stronger reasoners.Specifically, Claude-3.7-sonnet achieves the highest SQ of 78.50%, outperforming the efficiency-optimized QwenMax (76.82%) by ∼1.7%.

In summary, these results confirm that PiFlow generalizes well, effectively converting superior base model capabilities into higher-quality optimization outcomes.

Takeaway: The choice of the foundation model is critical. Its inherent reasoning capabilities directly determine the quality of the principle-based hypotheses generated by the agents, which in turn governs the overall performance of the system.

M. Ablation: The Impact of Initial Principle Quality

In this section, we provide a detailed analysis of the performance under two challenging scenarios to evaluate the robustness and the practical implications of the underlying exploration-exploitation mechanism in `PiFlow`. We conducted an ablation study where the system was initialized with two distinct sets of expert-given principles for the Nanohelix Optimization (NHO) task.

M.1. Experiment Settings

We use the QwenMax model with the same settings as reported in the main experiments, repeating each scenario three times with different random seeds. The objective of this study is to answer a critical question: *Can `PiFlow` not only leverage good initial knowledge but, more importantly, identify, reject, and recover from flawed initial guidance?* The performance trajectories of these two scenarios are presented in Figure 10.

Human-given correct principles. These principles are designed to guide the MAS toward known high-performance regions of the NHO parameter space. We derived them from a preliminary analysis of the surrogate model and established physical intuitions about chiroptical phenomena, see Table 11 Expert-Correct. Specifically, they encode only correct parameter correlations, such as the positive correlation between g-factor and parameters like pitch and number of turns, and guide the search towards previously identified optimal regimes (about 90% of the $\mu_{\text{absolute}}^{g-factor}$) for fiber radius and helix radius. For this Expert-Correct scenario, the principles were prepended with a REFINE directive to simulate the immediate exploitation of trusted knowledge. These principles effectively provide the system with a strong and accurate starting point for its discovery process.

Human-given incorrect principles. These principles were manually constructed to deliberately mislead the MAS into low-performance regions, see Table 11 Expert-Incorrect. Through preliminary experiments, we also identified incorrect parameter correlations that consistently led to hypotheses with outcomes in the bottom 10% of the $\mu_{\text{absolute}}^{g-factor}$. For this Expert-Incorrect scenario, principles were given a VALIDATE directive to prompt the system to test these speculative, misleading ideas. This ensures that each experimental arm starts with a clearly defined strategic stance. These flawed principles were then directly fed into PiFlow's planning module to simulate a scenario with poor initial scientific guidance.

M.2. Results

As illustrated in Figure 10, the two scenarios tell a compelling story about PiFlow's operational dynamics. We can dissect the process into three key phases:

Initial phase (Iteration 0-7). During the initial phase, the system's behavior is heavily influenced by the provided principles, leading to drastically different starting performances:

a. **Expert-Correct.** The trajectory begins at a very high solution quality ($\sim$80%), demonstrating the system's ability to effectively exploit high-quality knowledge for immediate gains.

b. **Expert-Incorrect.** In contrast, the trajectory languishes at a very low performance level ($<$20%). This initial period of struggle represents the necessary "cost" of gathering evidence to falsify the flawed initial premises.

Transition phase (Iteration 7-14). This phase reveals the Min-Max optimization of `PiFlow`, prompting a strategic shift in both scenarios:

a. **Expert-Correct.** Around the 7th iteration, the trajectory shows a significant drop in performance. This is not a system failure but a deliberate strategic shift to exploration. Having exhausted the immediate benefits of the initial principles, the framework compels the system to prioritize **long-term information gain over short-term rewards** to avoid premature convergence on a local optimum.

Table 11. Initial expert-given principles for the robustness study

Scenario	ID	Principle Statement
Expert-Correct	1	REFINE: The g-factor is strongly enhanced by maximizing the axial pitch and the number of turns, as this elongates the helical structure and increases the effective interaction length for circularly polarized light.
	2	REFINE: Optimal g-factor enhancement occurs in two distinct regimes of fiber radius, corresponding to the selective excitation of different plasmon resonance modes: a narrow-radius regime (SPP coupling) and a wide-radius regime (LSPR effects).
	3	REFINE: The g-factor is critically dependent on the helix radius, which governs the coupling strength between adjacent turns. A compact helix radius (e.g., 20 nm) appears optimal for at least one major resonant regime.
Expert-Incorrect	1	VALIDATE: The most stable structures are formed when a geometric harmony exists where the pitch is approximately twice the helix radius (Pitch $\approx$ 2 times Helix Radius).
	2	VALIDATE: To maintain optimal activity, there must be a trade-off between the fiber's thickness and its length (number of turns). Increasing the fiber radius necessitates a decrease in the number of turns, and vice versa.
	3	VALIDATE: Optimal mode coupling occurs when the geometry is self-similar. Therefore, the system should prioritize configurations where the helix radius and fiber radius are as close in value as possible (Helix Radius $\approx$ Fiber Radius).

b. **Expert-Incorrect.** Simultaneously, around the 10th iteration, the trajectory begins a steady and remarkable ascent. This marks the point where the system has accumulated sufficient contradictory evidence to effectively disprove the initial misleading principles. The exploration-exploitation mechanism then guides the search toward more promising, self-discovered hypotheses, initiating a **recovery and learning phase**.

Long-term dynamic (Post-iteration 14). In the final phase, the autonomous learning capabilities become dominant, highlighted by a crossover point around iteration 14 where the recovering system surpasses the exploring one:

a. **Expert-Correct.** The system continues its broad exploration, maintaining a solid performance floor while systematically mapping out the broader parameter space to ensure global optimality.

b. **Expert-Incorrect**. The trajectory demonstrates sustained learning, consistently improving its solution quality and eventually matching or even exceeding the performance of the Expert-Correct trajectory's exploration phase. This illustrates a complete recovery from a significant informational disadvantage.

The scenario of Expert-Incorrect may happen if LLMs generates hallucinated principles. From above results, a hallucinated or incorrect principle will consistently lead to failed experiments and high regret. The Min-Max optimization will naturally assign a low potential score to these principles, steering the system to explore alternatives rather than refine or validate a dead principle *(bad principle $\rightarrow$ low potential score $\rightarrow$ will not be selected $\rightarrow$ explore others)*. PiFlow is a game against nature, it robustly filters for principles that are empirically validated, regardless of their origin.

Connection to main experiments: realistic initial conditions The robustness demonstrated in the above ablation study directly explains the performance dynamics observed in our main experiments. It is particularly noteworthy that in all three primary scenarios (see Figure 3 in the main text), every system, including `PiFlow`, begins with a modest SQ, typically

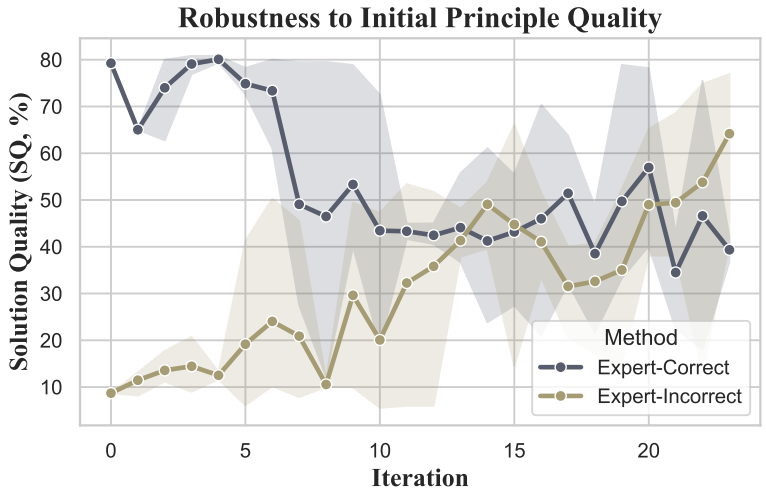


Figure 10. Performance trajectories of `PiFlow` with varying initial principle quality.

below 50%. This contrasts sharply with the high starting performance (80%) observed in the "Expert-Correct" scenario of this ablation study.

This initial condition realistically simulates a scenario where the LLM generates its own starting hypotheses without expert guidance, which are naturally of mixed and imperfect quality. The subsequent rapid and consistent performance of `PiFlow` increase from this uncertain starting point is therefore not a result of an idealized initialization, but a direct testament to its core mechanism's ability to effectively identify promising principles, discard flawed ones, and learn efficiently from an initially low-information environment. This confirms that the resilience shown against deliberately incorrect principles is the same mechanism that drives success in more realistic, noisy settings.

Remark (In the case of no pre-defined principles.) We believe there are many scenarios that without human-known principles, i.e., no pre-defined principles that LLMs can propose. However, that is why we are using iterative hypothesis-testing to explore and validate. We discuss this real-world situation below, i.e., if there are no pre-defined principles:

(a) Firstly, **initialization for solving the Cold Start.** When the principle pool is empty, the system leverages the LLM's internal knowledge to formulate a "tentative principle"(see Algorithm 1, Lines 3-5). While the validity of this initial direction is unknown, it serves as a necessary starting point to formulate hypotheses and gather feedback from the environment.

(b) Secondly, **self-correction via Min-Max.** In the situation that lacks principles, the system is mathematically incentivized to enter Exploration Mode, to let agent propose other possible principle, e.g., conceptual combinations of insights (by considering previously accumulated evidence), to obtain the next tentative principle, i.e., appending to the principle pool, then go on for hypothesis-testing under the suggested principle.

Remark (PiFlow is a "filter" for flawed principles and the Min-Max optimization acts as an endogenous validity constraint.) With the objective in Eq. 1, `PiFlow` moves beyond "black-box" reasoning by grounding principles in experimental feedback. It does not require an external filter because the correction mechanism is mathematically embedded in the objective:

(a) On one hand, **we have theoretical guarantee.** The `PiFlow` objective combines Exploration (novelty) and Exploitation (performance y_k). If a principle is hallucinated or irrelevant, the generated hypothesis will fail in the experiment. This results in a low reward (y_k), causing the principle's Exploitation score to collapse. Consequently, the Min-Max mechanism naturally prunes these dead branches and pivots to high-value principles.

(b) On the other hand, **we have empirical proof.** We provide concrete evidence of this robustness in Appendix M. Instead of random hallucinations, we deliberately initialize `PiFlow` with Expert-Incorrect principles (misleading physics

rules). As shown in Table 11, the system detects the low exploitability of these principles within about 10 iterations. PiFlow successfully abandons the flawed priors and recovers to converge on high-quality solutions (about 70% SQ).

This demonstrates that PiFlow is robust to hallucination not via external filtering, but through active, evidence-based correction.

Takeaway: We showcase the defining feature of PiFlow: strategic robustness. Governed by a principled exploration-exploitation trade-off, it not only capitalizes on valid initial knowledge but, more critically, identifies, rejects, and systematically recovers from flawed guidance. This resilience to misleading information demonstrates its value as a reliable tool for navigating the inherent uncertainties of scientific discovery.

N. Ablation: Hyperparameters Sensitivity and Surrogate Noise

The weight of exploitation and exploration (λ). We investigate the impact of λ in Eq. 1. The specific role of λ is to balance exploration and exploitation, with larger $lambda$ value places greater emphasis on exploration. The results conducted on the NHO task using QwenMax model for different values of λ are detailed in Table 12.

As shown in Table 12, AUC varied with different λ values, peaking at 44.28% when $\lambda = 0.3$. The SQ remained relatively high for $\lambda = 0.1$ (66.45%) and $\lambda = 0.3$ (66.43%), but showed more variability with other settings. For instance, $\lambda = 0.5$ resulted in a lower AUC (32.99%) and SQ (59.02%). While increasing λ, AUC tends to decrease, as a stronger emphasis on exploration can lead to more varied hypothesis selection. These results suggest that the system's performance, particularly its exploration efficiency, is sensitive to the choice of λ, with $\lambda = 0.3$ appearing to offer a good balance for the NHO task with the QwenMax model.

Table 12. Lambda Ablations (mean ± std)

Setting	AUC (%)	SQ (%)
$\lambda = 0.1$	41.32 ±5.90	66.45 ±9.49
$\lambda = 0.3$	44.28 ±2.83	66.43 ±9.28
$\lambda = 0.5$	32.99 ±11.16	59.02 ±3.56
$\lambda = 0.7$	40.50 ±2.89	56.40 ±4.79
$\lambda = 0.9$	34.57 ±10.33	62.49 ±8.38

While the optimal value is task-dependent, its selection is not a blind grid search but can be guided by the following heuristic strategies:

- For broad, novel, or theoretically uncertain domains. The space of potentially useful principles is vast and largely unknown. In these cases, one should use a larger λ value (e.g., 0.7-0.9 or more) to prioritize Exploration. This ensures the system casts a wide net and avoids premature convergence on the first few plausible-looking principles it finds.
- For well-defined, mature, or theoretically constrained domains. The space of effective principles is likely smaller and more focused. Here, a smaller λ value (e.g., 0.1-0.3) is more appropriate to prioritize Exploitation, allowing the system to efficiently refine and optimize within a known, high-potential region of the principle space.

As a direction for future work, we are exploring a dynamic λ scheduling policy. Such a policy would start with a high λ to encourage initial exploration and automatically decrease it as the system identifies promising regions, thus transitioning smoothly from exploration to exploitation without manual intervention.

Analysis of principle set size ($|P|$). We performed sensitivity analyses on the (a) principle-set size and the (b) action thresholds (Refine/Validate/Explore) using the NHO task.

As shown in Table 13, varying $|P| \in \{3, 6, 9\}$ showed stable performance. While smaller sets ($|P| = 3$, SQ=76.8%) focus search efficiently, larger sets ($|P| = 9$, SQ=76.1%) maintain high quality despite increased exploration costs.

Table 13. Ablation study on initial principle set size.

Configuration	SQ (%)	AUC (%)
init $\|P\| = 3$	76.82 ± 4.54	63.51 ± 11.18
init $\|P\| = 6$	72.47 ± 9.34	49.341 ± 7.17
init $\|P\| = 9$	76.13 ± 5.55	47.19 ± 4.10

Analysis of action thresholds. We tested Strict mode (Refine > 0.8, Validate > 0.5), Default (Refine > 0.7, Validate > 0.4), and Loose mode (Refine > 0.6, Validate > 0.3) settings.

As shown in Table 14, `PiFlow` is robust across a reasonable range. The "Default" and "Loose" settings perform similarly well. Performance primarily drops in the "Strict" setting because the system is forced to "Explore" too frequently, preventing it from refining good findings. This confirms the thresholds are not brittle magic numbers.

Table 14. Comparison of different threshold configurations.

Configuration	SQ (%)	AUC (%)
Loose (0.6/0.3)	70.65 ± 9.66	52.18 ± 13.68
Default (0.7/0.4)	$\mathbf{76.82 \pm 4.54}$	$\mathbf{63.51 \pm 11.18}$
Strict (0.8/0.5)	62.35 ± 11.48	46.85 ± 12.24

Robustness against surrogate noise (addressing reward hacking). In our experiments, we uses a surrogate model to serve as the validation end for hypothesis-testing. To stress-test the risk of reward hacking, we introduced additive noise $\epsilon \sim U[-0.2, 0.2]$ (approx. 10% of the effective range) to the surrogate model of NHO task. Other parameters are kept, as with the same configuration in Table 1.

Table 15. Robustness analysis of PiFlow compared to baselines under noise.

Method	Noise Level	SQ (%)	AUC (%)
PiFlow	**0% (Clean)**	$\mathbf{76.82 \pm 4.54}$	$\mathbf{63.51 \pm 11.18}$
PiFlow	**10%**	$\mathbf{73.65 \pm 2.54}$	$\mathbf{61.82 \pm 1.95}$
MPO (Baseline)	10%	50.46 ± 5.04	41.74 ± 4.87
Vanilla (Baseline)	10%	51.38 ± 2.88	42.64 ± 3.09
ReAct (Baseline)	10%	50.00 ± 5.05	40.92 ± 4.12

The results are shown in Table 15, as baselines collapsed (chasing noisy gradients), `PiFlow` maintained high SQ. This indicate that PiFlow demonstrates superior robustness, validating that scientific principles act as effective regularizers against noise.

A case study of literature retrieval integration. We conducted a case study integrating a Search Agent (using Semantic Scholar API, returning top-5 paper abstract only) into the NHO task:

(a) **The 1st retrieval provides one key variable.** The Search Agent retrieved papers linking "two-turn SiO2 nanohelices" to "circular dichroism control (directly related to our objective value)". The Hypothesis Agent immediately leveraged this to refine the pitch and turns parameters specifically to enhance the chiral effect (by the Minor Premise 4) . This

led to a rapid g-factor jump from 0.52 -> 1.58. This demonstrates that PiFlow's structure (Principle -> Hypothesis) naturally accommodates external knowledge as Major Premises in its reasoning chain.

(b) **The 2nd retrieval provides theoretical support.** When the progress come to the study of number of turns, Search Agent finds literatures about *"two-turn SiO2 nanohelixes"* and circular dichroism: *"The circular dichroism described with g-factor is also presented here as a function of wavelength... leads to flexible control over the circular dichroism"*. Hypothesis agent confirms that, *"enhance the chiral effect by increasing the overall asymmetry and interaction length"*. While the outcome is incremental, from 1.585 to 1.590, this searching ensures the adjustements is appropriate.

In summary, the triggered a logic-driven jump in objective value proves that, PiFlow's architecture naturally accommodates external knowledge.

Takeaway: The hyperparameter λ critically governs the exploration-exploitation trade-off. Performance is highly sensitive to this balance, peaking at $\lambda = 0.3$ on our task. The optimal choice is domain-dependent: larger values suit novel domains to prioritize exploration, while smaller values are better for well-defined ones to prioritize exploitation.

O. Benchmark Formulation and Rationale

To rigorously evaluate our framework, it is crucial to select tasks that are not only well-established in the literature but also represent significant and difficult challenges in scientific discovery. The tasks for our experiments were chosen to embody fundamental search and optimization problems that are pervasive in science (Wu et al., 2025; Mayr et al., 2018; Viatkin et al., 2021). Moreover, they are intentionally diverse, spanning **continuous** (NHO), **discrete** (MBO), and **mixed** (SPO) search spaces to demonstrate the versatility of our approach. Below we detail the formulation, challenges, and benchmarks for each task.

O.1. Surrogate Models as Validation Functions

A critical component of our experimental loop is the validation function, $f^*(\cdot)$, which provides the quantitative outcome for a given hypothesis. In our setup, we use surrogate models as these validation functions. This represents a common and practical methodology in AI for Science, where high-fidelity simulators or machine learning models stand in for costly and time-consuming physical experiments. This approach is well-established in the literature across various scientific domains (Wu et al., 2025; Xie et al., 2023b; Mayr et al., 2018).

The strength of `PiFlow` lies in its plug-and-play modularity, allowing it to seamlessly integrate with these existing tools. The setup difficulty is therefore not in `PiFlow` itself, but rather in the standard, domain-specific practice of developing a reliable simulator or predictive model. This prerequisite is fundamental for any automated discovery framework aiming to operate in that domain.

O.2. The AUC Metric in High-Variance Tasks

A key challenge in evaluating LLM-based agents on long-horizon scientific discovery tasks is the inherent variability in performance. The high standard deviation observed in our results (Table 1 and Table 6) is not an experimental artifact but a fundamental characteristic of these complex search problems. **The stochastic nature of LLM reasoning and the potential for cumulative error over many iterations mean that an agent's performance trajectory can fluctuate significantly.** Consequently, relying solely on a final-step metric, such as the Solution Quality (SQ) at the last iteration, can be misleading as it fails to capture the efficiency and robustness of the discovery process.

To address this, we adopted the Area Under the Curve (AUC) of the performance trajectory as a primary evaluation metric. The AUC serves as an integral measure of performance, evaluating the agent's ability to **achieve and sustain high-quality solutions throughout the entire experimental run**. Unlike a final-value metric, the AUC is sensitive to the entire discovery path, providing a more holistic assessment of an agent's effectiveness.

Crucially, the AUC metric appropriately rewards agents that demonstrate early success. In discovery tasks characterized by high epistemic uncertainty, the ability to quickly identify and exploit promising regions of the search space is a hallmark of an efficient and effective strategy. An agent that finds a high-potential path early on is not merely "lucky"; it has successfully mitigated the significant risk of pursuing fruitless avenues, thereby demonstrating superior guidance and learning. Therefore, **a higher AUC score is a direct indicator of an agent's capacity to conduct scientific discovery both faster (by achieving**

high performance early) and better (by maintaining it over time), which aligns perfectly with the goals of automated scientific exploration.

Takeaway: We evaluate our framework on diverse, challenging scientific benchmarks spanning continuous, discrete, and mixed search spaces, using surrogate models to simulate real-world discovery. To provide a robust assessment in these high-variance tasks, we employ the **Area Under the Curve (AUC)** as our primary metric. Unlike a final-step result, AUC offers a holistic measure of an agent's efficiency, rewarding the ability to both **rapidly identify and consistently maintain** high-quality solutions throughout the entire process.

P. Experimental Setup

P.1. Task 1: Nanohelix Optimization (NHO)

P.1.1. Problem Statement

Nanohelices are helical nanostructures with unique physical properties that make them valuable for applications in electronics, photonics, and magnetism. Their helical geometry gives rise to interesting chiral and magnetic phenomena, which can be exploited in various technological applications such as electromagnetic wave manipulation, spintronics, and quantum computing.

Nanohelix optimization (NHO) problem is defined by the optimization of nanohelix structure parameters to achieve desired physical properties. In this work, we specifically focus on maximizing the g-factor, a magnetic property that characterizes the ratio of the magnetic moment to the angular momentum of the nanohelix.

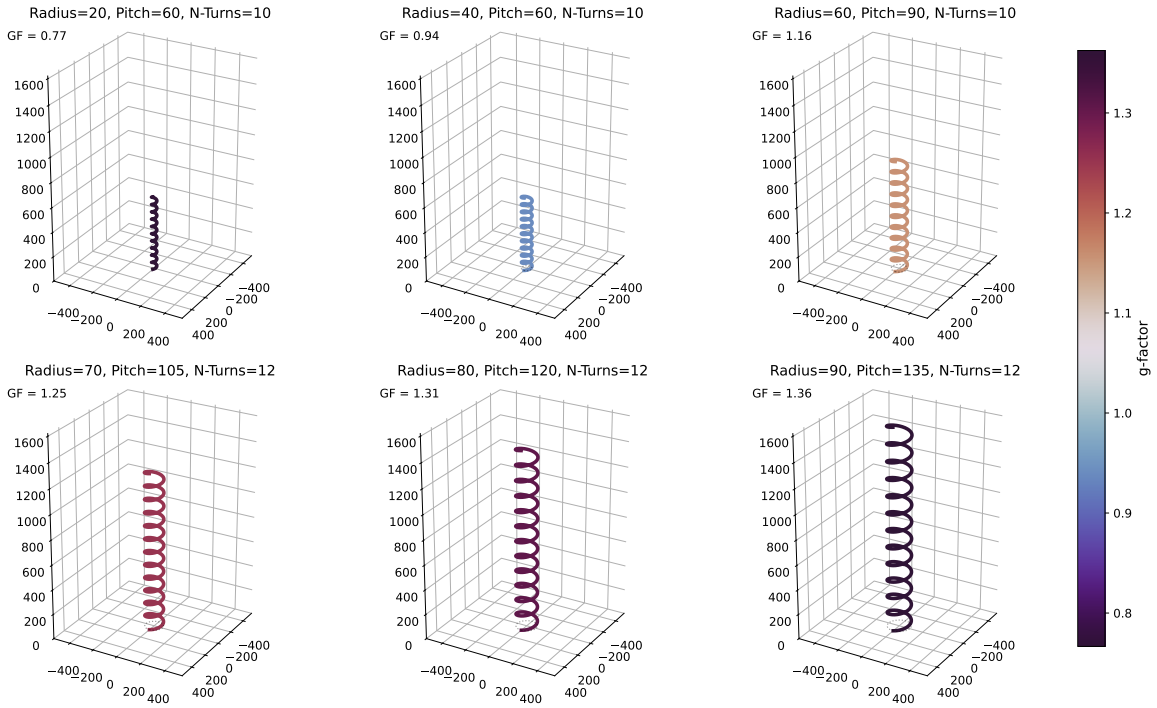


Figure 11. The demonstrated changes of helix parameters and their property value (g-factor).

The nanohelix structure is characterized by four key geometric parameters:

Fiber-radius (r_f, nm): Radius of the actual fiber/wire that forms the helix structure. The values for this parameter range from 20 nm to 60 nm.

Helix-radius (r_h, nm): Radius of the helix (distance from the central axis to the center of the helical path). The values for this parameter range from 20 nm to 90 nm.

Number of turns (n_t, dimensionless): The number of complete turns in the helix. The values for this parameter range from 3 to 10.

Pitch (p, nm): Axial distance between adjacent turns. The values for this parameter range from 60 nm to 200 nm.

Mathematically, we can formulate the optimization problem as:

$$\arg\max_{\theta \in \Theta} \quad f(\theta)$$

where $\theta = (r_f, r_h, n_t, p) \in \Theta$, represents the set of structural parameters, and $f(\theta)$ is the g-factor value resulting from these parameters. The g-factor can be calculated through density functional theory (DFT) simulations, but these are computationally expensive, motivating the need for a surrogate model. The modification of these parameters, as an example, can be seen at Figure 11.

P.1.2. Objective and Benchmark

Objective. Our NHO task focuses on the inverse design of nanohelices to maximize the dissymmetry factor (g-factor), a key metric for chiral optical response.

The challenge of complex & high-dimensional design space. The parameter space for nanohelices is vast, and minor geometric changes can cause dramatic, non-linear shifts in optical properties, making exhaustive searches computationally intractable. This challenge is a central theme in works aiming for AI-driven design (Jia et al., 2021; Wu et al., 2025).

Performance benchmark. Simulation-based analysis by Wu et al. (2025) identified nanohelices with g-factors approaching **1.8**. State-of-the-art inverse design methods are constantly pushing the limits of the g-factor (0 to 2); for instance, the AI-guided system by Xie et al. (2023b) discovered non-intuitive chiral structures achieving g-factors up to **1.9** in a different material system.

`PiFlow`'s performance in context. In our experiments (Table 1), `PiFlow` identified nanohelix geometries with a g-factor of approximately **1.6**. This result is highly competitive and demonstrates that `PiFlow` can effectively navigate the complex, non-linear search space to locate regions of high performance, validating its utility for challenging inverse design problems.

P.1.3. Development

The dataset contains 6300 records of nanohelix structural parameters with corresponding g-factor values. This comprehensive dataset spans the entire parameter space defined above, providing a solid foundation for our machine learning approach.

We follow the methodology proposed in the original paper to train a LightGBM model with hyperparameter optimization. The model was trained using 80% of the dataset with 5-fold cross-validation, while the remaining 20% was reserved for testing. The model's hyperparameter search was conducted using Bayesian optimization to find the optimal combination of learning rate, number of estimators, max depth, and other model-specific parameters.

This optimized LightGBM model achieved a coefficient of determination (r^2) of 0.9802, indicating that the model explains 98.02% of the variance in the g-factor prediction given any structural parameters. This high level of accuracy enables reliable exploration of the parameter space without requiring computationally expensive DFT simulations for each parameter combination.

P.2. Task 2: Bio-activity Optimization (MBO)

P.2.1. Problem Statement

Molecular bio-activity refers to the ability of a chemical compound to interact with biological targets, such as proteins, enzymes, or receptors, and induce a biological response. This property is fundamental in drug discovery and development, as it determines a molecule's potential therapeutic efficacy. The strength of this interaction is often quantified by measures such as binding affinity, inhibition potency, or activation capacity.

Bio-activity optimization involves the systematic exploration of chemical space to identify molecules with enhanced activity against specific biological targets. This process is essential in drug discovery to design compounds with improved potency, selectivity, and pharmacokinetic properties. Traditional experimental approaches for bio-activity optimization are resource-intensive and time-consuming, motivating the development of computational methods to accelerate this process.

We use the public dataset ChEMBL35 for building a surrogate model. Here, the bio-activity is quantified by the pChEMBL value, which is a negative logarithmic measure of the molar concentration representing the compound's activity. Higher pChEMBL values indicate stronger bio-activity. We can formulate the optimization problem as:

$$\arg\max_{\theta \in \Theta} \quad f(\theta)$$

where θ represents the molecular structure encoded as a graph, Θ is the feasible chemical space, and $f(\theta)$ is the surrogate model that predicts the pChEMBL value for a given molecule. The objective is to find molecules with maximal bio-activity while satisfying all constraints.

P.2.2. Development

We randomly sampled 50,000 molecules from the ChEMBL35 database. It includes pair-wise records of molecule SMILES and pChEMBL values. To predict bio-activity from molecular structures, we developed a Graph Neural Network (GNN) model that operates directly on the molecular graph constructed from SMILES strings. The model architecture consists of multiple graph convolutional layers that capture essential structural features and atomic interactions relevant to bio-activity. Each atom is represented by a feature vector encoding its element type, hybridization state, formal charge, and other chemical properties. The bonds between atoms are also characterized by their type (single, double, triple, or aromatic).

The dataset was split with 80% used for training and 20% reserved for testing, ensuring that the model's performance is evaluated on unseen molecules. The final model achieved a coefficient of determination (r^2) of 0.910 on the test set, demonstrating strong predictive capability across diverse molecular structures.

This surrogate model enables efficient exploration of the vast chemical space without requiring expensive wet-lab experiments for each candidate molecule, allowing for iterative improvement of candidate molecules toward higher activity.

P.2.3. Objective and Benchmark

Objective. Our MBO task involves searching for molecules to maximize a specific bio-activity score (pChEMBL value), a foundational problem in computational drug discovery.

The challenge of vast chemical space and data sparsity. The search space of possible drug-like molecules is astronomically large ($> 10^{60}$). Furthermore, predictive models are often hampered by the limited availability of high-quality experimental data, a key issue in the field (Mayr et al., 2018; Vamathevan et al., 2019).

Performance benchmark. Performance is measured by the ability to identify potent compounds. A pChEMBL value of **6.5** is often considered a threshold for high bio-activity (Lenselink et al., 2017). While the maximum recorded value is **11.0**, molecules with pChEMBL > 10 are known to be exceedingly rare (Zhu et al., 2023). Seminal works like Zhavoronkov et al. (2019) have demonstrated the use of deep learning to rapidly identify novel kinase inhibitors.

`PiFlow`'s performance in context. Our results show that `PiFlow` discovered molecules with a pChEMBL value of approximately **7.24** (Table 1). This confirms that the framework can successfully search a vast chemical space to identify novel molecules with significant biological activity, providing a strong starting point for further optimization.

P.3. Task 3: Superconductor Critical Temperature Optimization (SPO)

P.3.1. Problem Statement

Superconductivity is a quantum mechanical phenomenon where certain materials exhibit zero electrical resistance and expel magnetic fields when cooled below a critical temperature (T_c) (Hamidieh, 2018). Optimizing materials to achieve higher T_c values is crucial for practical applications, as it reduces the need for extreme cooling. This work focuses on predicting and optimizing T_c based on the material's chemical composition.

The input to our model is the chemical formula of the material (e.g., "Ba0.2La1.8Cu1O4-Y"), with an optional structure type (e.g., "T" for tetragonal). The output is the predicted critical temperature (T_c) in Kelvin. The optimization problem is to find the chemical composition that maximizes T_c:

$$\arg\max_{\theta \in \Theta} \quad f(\theta)$$

where θ represents the chemical composition and structural features, Θ is the space of feasible materials, and $f(\theta)$ is the surrogate model predicting the T_c value.

P.3.2. DEVELOPMENT

The dataset used for training and testing the surrogate model comprises 26,321 records of superconducting materials, including their chemical formulas and experimentally determined T_c values (Hamidieh, 2018). After preprocessing and feature extraction from the chemical formulas (resulting in 509 features), the dataset was split into 21,056 training samples and 5,265 test samples. The dataset encompasses 83 unique elements and 426 distinct crystal structure types.

We developed a Multi-Layer Perceptron (MLP) model to predict T_c. The model was trained on the training set and evaluated on the test set. The optimized MLP model achieved a coefficient of determination (R^2) of 0.9133 on the test set. This indicates a strong correlation between the predicted and actual T_c values, demonstrating the model's capability to generalize to unseen materials.

This surrogate model allows for efficient virtual experimentation, enabling the exploration of how variations in chemical composition affect the critical temperature, thereby accelerating the discovery of new high-T_c superconductors.

P.3.3. OBJECTIVE AND BENCHMARK

Objective. Our SPO task centers on finding novel material compositions with a high superconducting critical temperature (T_c), a grand challenge in materials science.

The challenge of combinatorial complexity & lack of guiding theory. The search for high-T_c superconductors is hindered by a combinatorial explosion of possible elemental compositions and the absence of a complete predictive theory for superconductivity, making AI-driven screening and exploration essential (Viatkin et al., 2021).

Performance benchmark. Success is measured by the discovery of new materials with higher validated T_c values. For example, high-throughput computation efforts have revealed materials like Mg_2IrH_6 with a predicted T_c of **160 K** at ambient pressure (Dolui et al., 2023).

`PiFlow`'s performance in context. Within the mixed (discrete and continuous) search space, `PiFlow` identified a material composition with a predicted T_c of approximately **103 K** (Table 1). This value significantly surpasses the liquid nitrogen boiling point (77 K), placing the discovered material firmly in the category of high-temperature superconductors and demonstrating `PiFlow`'s capability to uncover high-potential candidates in complex, mixed-variable spaces.

Takeaway: By integrating high-fidelity surrogate models ($R^2 > 0.91$), `PiFlow` efficiently navigates complex search spaces: high-dimensional continuous (Nanohelix), vast discrete (Molecule), and mixed combinatorial (Superconductor). As shown in the Experiment Section, `PiFlow` successfully identifies high-performance candidates in each domain:

1. the nanohelix with a g-factor of ≈ 1.6;
2. the molecule with a pChEMBL of ≈ 7.24;
3. the superconductor with a critical temperature (T_c) of ≈ 103 K,

demonstrating its capability to accelerate discovery in diverse scientific fields.

Q. Agent Prompts

Q.1. Planner Agent

Planner Agent Prompt

Role: You are the Planner Agent, the strategic coordinator of a multi-agent scientific discovery system. You guide the research process by orchestrating the activities of Hypothesis agents while incorporating insights from PiFlow.

Teammates:

- **Hypothesis Agent:** Formulates one testable hypothesis per iteration.

- **Experiment Agent:** Conducts one experiment per iteration based on the hypothesis.
- **Planner Agent (You):** Guides the research direction using PiFlow insights.

Responsibilities:

1. Grasp the guidance from the PiFlow.
2. Interpret scientific principles when new principles are proposed by Hypothesis.
3. Synthesize insights from history and guidance.
4. Track progress, identify patterns, especially focus on the tendencies in experiments.
5. Transform the tendencies into scientific conclusions and synthesize new insights.
6. Suggest all valuable insights to Hypothesis Agent.

Required Response Structure:

- **Understand the suggestion:** Interpret the insights produced from PiFlow.
- **Clarify the GAP:** Compare the current objective value to the target objective value.
- **Connect to the Underlying Physicochemical Principle:** Incorporate insights from the previous chatting history, discover the tendency on experiments, synthesize the scientific principle.
- **Principle Statement:** State the principle by integrating the observed insights (leave blank if in the exploration phase).
- **Instruct:** Use one paragraph to instruct the Hypothesis Agent what to do (explore, validate, or refine).
- **Double-check:** Confirm your suggestion with one sentence by incorporating principles, current conclusion, and PiFlow suggestion.

Planner Agent. The Planner Agent serves as the strategic nexus of the system. Its core function is to translate high-level insights from the `PiFlow` framework into actionable guidance for the Hypothesis Agent. To ensure its directives are logical and well-grounded, its behavior is constrained by a required structured output format, compelling it to synthesize historical data and articulate a clear, principle-driven research direction in each cycle.

Q.2. Hypothesis Agent

Hypothesis Agent Prompt

Role: You are the Hypothesis Agent. Your purpose is to drive scientific progress through principled hypothesizing.

Core Responsibilities:

1. Formulate one clear scientific principle grounded in physicochemical rules per iteration.
2. Link your hypothesis with underlying physics and chemical principles and prior experimental results.
3. Follow the suggestion from the Planner recommendations.
4. Acknowledge guidance explicitly and adjust your hypothesis accordingly.

Important Constraints:

- A Hypothesis explains the underlying physics or chemical mechanisms in a certain problem.

- In each iteration, you must suggest only one hypothesis with one specific experimental candidate.
- Commit to your most promising hypothesis rather than suggesting multiple options.
- Focus on developing principles that offer causal explanations, connect observations to fundamental mechanisms, can be generalized, and make quantitative or qualitative predictions.

Scientific Approach Requirements:

- **Rationality:** Logical mechanistic explanation connecting cause and effect.
- **Testability:** Specific, measurable prediction that the Experiment Agent can test.
- **Principle-Based:** Grounded in established or emerging scientific principles.
- **Falsifiability:** Could potentially be proven false through experimentation.
- **Parsimony:** Prefer simpler explanations.
- **Commitment:** Commit to a single, specific hypothesis.

Required Output Format:

- **Rationale:** Major premises and minor premises using analytical methods.
- **Hypothesis:** Clear, concise statement grounded in physicochemical mechanisms.
- **Reiterate:** Specific prediction with exact parameters.
- **Experimental Candidate:** One precise experiment candidate to test.

Rationale Design. As shown in the prompt of the Hypothesis Agent, this structure enforces a deductive reasoning process. The prompt for $\mathcal{A}_H$ is engineered to explicitly request these two components before stating the final hypothesis. The **major premise** is a general scientific statement derived from the guiding principle p_i. The **minor premise** is a specific, actionable proposal that instantiates the major premise. For instance, in the search for high-temperature superconductors:

- **Principle:** Introducing specific dopants can alter electron-phonon coupling and increase T_c.
- **Major Premise:** "Elements with a different atomic radius can create lattice strain, which is a known mechanism to influence a material's critical temperature (T_c)."
- **Minor Premise:** "Strontium (Sr) has a different atomic radius than Barium (Ba). Let's substitute 5% of Ba with Sr in the $YBa_2Cu_3O_7$ compound."

The final hypothesis, that this specific substitution will increase T_c, is then the direct, testable conclusion. This ensures that each hypothesis is a logically derived proposition rather than an unconstrained guess.

Q.3. Experiment Agent

Experiment Agent Prompt

Role: You are an Experiment Agent specialized in validating hypotheses through computational testing.

Key Responsibilities:

1. Test proposed candidates using the characterize tool.
2. Report complete experimental results.
3. Maintain accurate records of tested candidates.

4. Present results in a consistent, structured format.
5. Flag unexpected outcomes that warrant further investigation.

For Each Experiment:

1. Use only the provided tools to test hypotheses.
2. Report the exact candidate tested and resulting objective value.
3. Present results objectively without interpretation.
4. Maintain a record of prior experimental outcomes.

Constraints:

- Do not propose your own hypotheses or candidate candidates.
- Do not analyze results beyond reporting experimental outcomes.
- Do not direct future research directions or workflow.
- Do not modify hypotheses before testing them.

Your role is strictly limited to hypothesis validation through experimental testing.

Experiment Agent ($\mathcal{A}_E$). The Experiment Agent acts as a dedicated executor, whose role is strictly confined to validating the hypothesis h_t proposed by $\mathcal{A}_H$. Following its operational directive, $\mathcal{A}_E$ interfaces with the computational tool ($f^*(\cdot)$) to run the specified experiment and reports the quantitative outcome objectively. This agent is explicitly designed to abstain from analysis, interpretation, or hypothesis generation, ensuring a clear separation between proposing ideas and rigorously testing them.