A synergistic multi-specialist knowledge reasoning model for molecular science

Pengfei Liu
Pengcheng Laboratory
School of Computer Science and Engineering,
Sun Yat-sen University

Shuang Ge
Pengcheng Laboratory
Tsinghua Shenzhen International Graduate School,
Tsinghua University

Jun Tao
School of Computer Science and Engineering,
Sun Yat-sen University

Zhixiang Ren*
Pengcheng Laboratory
jason.zhixiang.ren@outlook.com

ABSTRACT

The rapid evolution of artificial intelligence in molecular science necessitates a shift from data-driven predictions to knowledge-guided reasoning. Existing molecular models are predominantly proprietary, lacking general molecular intelligence and generalizability. To address this, we propose a task-adaptive large reasoning model that integrates molecular scientific logic to emulate the thinking of molecular scientists, with capabilities for reasoning and reflection. Our approach incorporates multi-specialist modules to provide versatile molecular expertise and a chain-of-thought (CoT) framework enhanced by reinforcement learning infused with molecular knowledge, enabling structured and reflective reasoning. The model outperforms over 20 state-of-the-art multi-task large language models (LLMs) across 10 molecular tasks on 47 metrics, including property prediction, molecule generation, and reaction prediction. It achieves a 50.3% improvement over the base model while ensuring interpretability. It can bridge data-driven and knowledge-integrated approaches for intelligent molecular design.

***K*eywords** Reasoning Model · Molecular Science · Multi-task Learning

Artificial intelligence (AI) has rapidly become a driving force in molecular science, with significant applications in drug discovery [1], materials design [2], and chemical synthesis planning [3]. Traditional experimental methods and quantum chemical calculations, while foundational, are often time-consuming, expensive, and limited by their inability to handle large-scale molecular datasets efficiently. Conventional machine learning approaches have improved prediction accuracy for small datasets but struggle with scalability, feature engineering complexity, and capturing intricate molecular interactions [4]. In particular, deep learning techniques, such as graph neural networks (GNNs), have been widely applied to model molecular structures as graphs, enabling tasks like property prediction and reaction forecasting [5]. These methods excel in capturing local structural features but often suffer from high computational costs, limited scalability to large datasets, and challenges in modeling long-range dependencies within complex molecular sequences.

In contrast, the Transformer architecture [6] addresses these limitations through its self-attention mechanism, which efficiently processes sequential data like SMILES [7] strings, facilitating better generalization and interpretability in molecular representations. The advent of Transformer-based models [8][9][10] has strengthened the understanding and generation capabilities of sequential data. The emergence of ChatGPT [11] has sparked a surge in LLMs, inspiring their adaptation in molecular science. For instance, MolT5 [12] enables translation between molecules and natural language, which supports tasks such as molecule captioning and property description. BioT5+ [13] extends this paradigm to

*Corresponding author

biological contexts by integrating IUPAC [14] names for enhanced understanding in drug discovery. GIT-Mol [15] and MoleculeSTM [16] further incorporate multi-modal inputs for comprehensive molecular analysis. Those methods have demonstrated strong predictive power across diverse molecular tasks, enabling advances in property prediction [17], de novo molecular design, reaction prediction [18], and molecule captioning. Nevertheless, the majority of existing approaches are still primarily data-driven, focusing on statistical correlations [19] rather than capturing the underlying scientific principles [20]. They are typically **designed for single tasks**, limiting their generalizability to diverse multi-task scenarios. Moreover, they often **lack explicit reasoning mechanisms**, which necessitate logical inference and seamless integration of domain-specific knowledge.

Recent advances in multi-task molecular LLMs have begun to address these gaps by scaling foundation models for broader applicability. For example, LLaSMol [21], ChemLLM [22], and InstructBioMol [23] have achieved excellent performance in multi-task and multi-modal molecular learning, highlighting the potential of scaling molecular foundation models. However, these models primarily serve as high-performing predictors, lacking robust interpretability and explicit reasoning capabilities, which limits their application in scientific discovery scenarios that require transparent decision-making. While models such as o1 [24] and DeepSeek-R1 [25], among other advanced large-scale reasoning models, have introduced advanced reasoning frameworks, they also adapt to Mixture-of-Experts (MoE) [26] architectures, enabling flexibility across diverse tasks. However, these models encounter challenges with domain-specific accuracy and scalability when applied to molecular contexts. DeepMind's TxGemma model [27] combines conversational capabilities with molecular multi-task learning, though this integration results in decreased task performance. These limitations highlight a critical need for a multi-task molecular reasoning model that achieves high predictive accuracy and possesses robust reasoning capabilities.

Since molecular science relies on established chemical rules, structured knowledge, and logical reasoning, there is a growing need for models that integrate such knowledge to deliver accurate predictions and science-based reasoning. To tackle these challenges, we introduce a synergistic multi-specialist knowledge reasoning model that embeds molecular logic into a task-adaptive framework, as illustrated in Figure **??**. Unlike prediction-focused models, our approach incorporates chemical knowledge via CoT reasoning [28]. We construct an instruction dataset comprising 93K instances to train prediction specialists, focusing on molecular knowledge representation, and a dataset with chemical CoT reasoning, embedding molecular science constraints to align AI decisions with scientific principles, consisting of 3.5K high-quality instances for inference specialists. In contrast to MoE architectures, we leverage **data synergy** (joint training of data with similar knowledge while isolating disparate knowledge) and **specialist synergy** (collaboration between prediction specialists excelling in molecular representation and inference specialists proficient in structured reasoning) to enhance molecular knowledge representation and reasoning.

Enhanced by molecule-informed reinforcement learning, this approach aligns reasoning and answers with chemical accuracy through sparse CoT data, boosting data efficiency and performance. Our model outperforms over 20 multi-task LLMs across 10 molecular tasks, achieving improvements of up to 10% compared to the state-of-the-art baseline. This capability fosters reliable scientific exploration by bridging empirical data and science-driven insights, supporting innovations in drug discovery and materials design.

Results

This section presents the key outcomes of our model, which is designed to advance molecular science through enhanced reasoning and performance. We evaluate the model across a diverse set of tasks, conduct ablation studies to assess the impact of data and specialist synergy, and analyze the evolution of specialist adaptations and the distribution of reasoning chains. These analyses collectively demonstrate the model's superiority over existing approaches and its potential to bridge data-driven and knowledge-integrated paradigms.

Molecular multi-task reasoning framework

Molecular science tasks. We follow the experimental design of the representative molecular multi-task model LLaSMol, selecting ten representative molecular science tasks to assess our model's capabilities. These tasks encompass text-generation challenges, including molecule captioning and SMILES generation [12] for creating textual descriptions and molecular structures, respectively, as well as SMILES-to-IUPAC and IUPAC-to-SMILES translations for interconverting chemical notations. Additionally, the benchmark includes property prediction [29] with classification tasks, such as blood-brain barrier penetration (BBBP) and clinical toxicity (ClinTox), as well as regression tasks like water solubility (ESOL) and lipophilicity. It also contains forward reaction prediction and retrosynthesis, focusing on reaction outcome forecasting and the design of synthetic pathways. These tasks serve as key evaluations for LLMs in molecular science, as they are ideally suited for reasoning-oriented assessments due to their reliance on logical inference and domain-specific knowledge.

Knowledge-infused dataset design. We employed a stratified sampling method based on molecular feature distributions to construct a 93K instruction-following dataset from over 2 million data points, training prediction specialists to enhance molecular knowledge representation while ensuring efficient model performance. Additionally, we conducted deep sampling to build a 3.5K high-quality dataset, completed with CoT annotation, to train inference specialists focused on molecular knowledge inference. Further details are available in Supplementary Information A.

Synergistic multi-specialist architecture. The model integrates multiple specialist modules, coordinated by a router mechanism, to tackle diverse molecular tasks. It harnesses data synergy through the joint training of data containing similar knowledge and the isolation of disparate knowledge. This amplifies related knowledge, with specialist assignments tailored to task knowledge types. For instance, text-based SMILES generation and IUPAC-to-SMILES are handled by the same specialist. Forward reaction and retrosynthesis prediction are managed by a different specialist. For property prediction, BBBP and ClinTox are assigned to separate experts despite both being classification tasks. Their distinct feature distributions optimize performance. Moreover, it leverages specialist synergy through collaboration between prediction specialists excelling in molecular representation and inference specialists proficient in structured reasoning. Prediction specialists and inference specialists work together as pairs to jointly execute reasoning tasks. This makes reasoning more directed, avoiding errors caused by overly long reasoning chains.

The framework integrates multiple specialist modules, coordinated by a router mechanism, to tackle diverse molecular tasks. It harnesses data synergy to amplify related knowledge and specialist synergy to sharpen reasoning, setting it apart from conventional methods by embedding chemical logic directly. Prediction specialists and inference specialists operate as pairs to jointly execute reasoning tasks, with specialist assignments tailored to task knowledge types—e.g., text-based SMILES generation and IUPAC-to-SMILES are handled by the same expert, while forward and retrosynthesis are managed by another. For property prediction, BBBP and ClinTox, despite being classification tasks, are assigned to separate experts due to their distinct feature distributions, optimizing performance.

Knowledge-guided alignment strategy. Prediction specialists and inference specialists undergo initial independent training. As reasoning length increases, inference specialists may deviate from the initial answers provided by prediction specialists. To address this challenge, we implement reinforcement learning with task-specific molecular science reward functions. For instance, the SMILES generation reward integrates SMILES validity and similarity to the target molecule. This approach enhances the consistency of reasoning between prediction specialists and inference specialists.

Performance evaluation and ablation studies

Metrics. We evaluate the effectiveness of our model by assessing its performance across ten representative molecular science tasks. These assessments follow established evaluation protocols as outlined in [29, 21]. Each task is comprehensively evaluated using five or more assessment metrics, with representative metrics carefully selected to address task-specific challenges. METEOR scores [30] measure the quality of text-generation tasks by evaluating semantic similarity between generated and reference descriptions. MACCS molecular fingerprint [31] similarity gauges structural accuracy for SMILES generation and reaction prediction tasks by comparing molecular fingerprints. Accuracy is crucial for the success of classification tasks, such as BBBP and ClinTox, as it ensures the correct identification of molecular properties. For regression tasks like ESOL and Lipophilicity, we use 1/RMSE, which serves as an inverse measure of prediction error. The complete experimental setups, evaluation metrics, and results are comprehensively detailed in Supplementary Information B.

Baselines. To benchmark our model, we compare it against over 20 baselines, selecting representative models from three distinct series, as illustrated in Figure **??** (a), which highlights a subset of competitive models for clarity. From API-based LLMs, we include DeepSeek-V3 [32], Gemini-2.5-Flash [33], Claude-3.5 [34], and o4-mini [35], chosen for their advanced prompt-guided capabilities. Open-source general LLMs are represented by Qwen3-235B [36], DeepSeek-70B [25], MiniStral-8B [37], and LLaMA-4 [38], selected for their broad adaptability and community validation. For molecular multi-task models, we opt for TxGemma-9B-predict, TxGemma-9B-Chat, and LLaSMol [21], reflecting specialized designs in this domain. The TxGemma series [27] lacks training for text generation, molecular generation, and reaction prediction tasks. To ensure fairness, we exclude these tasks from comparison with TxGemma. API-based LLMs and open-source general LLMs depend on prompt-guided responses, while molecular multi-task models are locally deployed and reproduced for consistent testing.

Overall performance insights. Notably, our model outperforms all baselines in overall performance, achieving superior results in all tasks except Lipophilicity. As depicted in Figure **??** (a), the three radar charts clearly illustrate our model's significant advantage over other models. API-based LLMs tend to generate more hallucinations, with longer reasoning processes increasingly deviating from accurate answers or knowledge, leading to reduced accuracy. In contrast, LLaMA-4 exhibits moderate molecular knowledge and capability in sequence/text generation, but underperforms in quantitative tasks such as property prediction. Knowledge-enriched Molecular multi-task models, including ours, demonstrably

outperform API-based LLMs and Open-source general LLMs, reflecting the benefit of domain-specific expertise. However, our model falls slightly behind in the Lipophilicity regression task, potentially due to its specialized training on regression-focused datasets that our model has not fully optimized for. Compared to other methods, LLaSMol achieves a solid performance in multi-molecular task accuracy, as shown in Figure **??** (b). Relative to the state-of-the-art molecular multi-task model LLaSMol, our task metrics demonstrate an improvement of nearly 6%.

Ablation study findings. To quantify the benefits of our core strategies—**data synergy** and **specialist synergy**—we conduct ablation studies using DeepSeek-7B as the base model, with the mean performance across tasks serving as the aggregate evaluation metric. As shown in Figure **??** (c), instruction fine-tuning enhances molecular task capabilities, and subsequent specialist synergy further boosts reaction prediction and SMILES generation, elevating the aggregate metric from 0.705 to 0.801. In contrast, CoT-only fine-tuning on DeepSeek-7B yields modest gains (0.517 to 0.569) due to the smaller dataset size and increased inference length. Applying data and specialist synergy to this variant then improves performance to 0.704. Finally, the integration of reinforcement learning, incorporating task-specific reward functions to guide reasoning, raises the metric to 0.777. Starting from a baseline model of DeepSeek-7B fine-tuned on the instruction dataset with an initial metric of 0.705, this enhancement achieves an improvement of over 10%, realizing a multi-task molecular reasoning model that delivers high predictive accuracy and robust reasoning capabilities.

Specialist evolution analysis

To gain deeper insights into the internal dynamics of our model, we analyze the evolution of specialist modules before and after training, focusing on changes in molecular representation capabilities and internal weight distributions. This examination reveals how data synergy and specialist synergy foster distinct adaptations, enabling the model to balance knowledge representation (via predict specialists) and logical inference (via inference specialists) across molecular tasks.

As illustrated in Figure **??** (a), pre- and post-training Uniform Manifold Approximation and Projection (UMAP) [39] projections vividly demonstrate the evolution of task-specific embeddings developed by specialists. This demonstrates unique molecular knowledge interpretations in the latent vector space for both prediction specialists and inference specialists. Prediction specialists cluster representations more tightly for structural fidelity. Inference specialists expand clusters to accommodate reasoning pathways. Figure **??** (b) further quantifies these shifts through L2 norms, which measure the magnitude of weight changes in each specialist module.

The L2 norm of a weight matrix W is defined as

$$\|W\|_2 = \sqrt{\sum_{i,j} w_{i,j}^2}, \tag{1}$$

where $w_{i,j}$ denotes the elements of W. The difference in L2 norms is then computed as

$$\Delta\|W\|_2 = \max\left(\|W_{\text{specialist}}\|_2 - \|W_{\text{init}}\|_2, 0\right), \tag{2}$$

where $\Delta\|W\|_2$ is the L2 norm difference, $\|W_{\text{specialist}}\|_2$ is the L2 norm of the trained specialist weights, $\|W_{\text{init}}\|_2$ is the L2 norm of the initial weights, and $\max(\cdot, 0)$ ensures the difference is non-negative for visualization purposes. Prediction specialists undergo more aggressive fine-tuning to capture domain-specific patterns, which reflect substantial updates for enhanced knowledge representation. Inference specialists prioritize stability for consistent CoT application. They show moderate changes, indicating refined integration of reasoning logic.

The weight difference heatmap in Figure **??** (c) underscores the granularity of specialization. The molecule captioning specialist serves as the reference due to its foundational role, as all task reasoning processes can be fundamentally transformed into captioning tasks. In contrast, the compared text-generation tasks further embody greater scientific rigor. As illustrated in Figure **??** (c), the weight difference heatmap reveals distinct activation patterns between prediction specialists and inference specialists across these text-generation tasks. This highlights our framework's adaptability to multi-task scenarios by leveraging specialized routing. Such differentiation validates the efficacy of the design and supports enhanced performance in diverse molecular reasoning applications.

Reasoning chain distribution analysis

We assess the reasoning capabilities of our model by analyzing the distribution of reasoning chains across its outputs. This study focuses on three representative tasks: text-based molecule generation, molecular property prediction (BBBP), and retrosynthesis. We examine the diversity of reasoning elements and their proportional contributions. Figure **??** provides an overview of these distributions, highlighting the model's adaptability to varied molecular reasoning demands, with further details in Supplementary Information C.

The analysis reveals distinct reasoning patterns tailored to each task. Text-based molecule generation emphasizes structural construction, while BBBP prediction focuses on comparative property assessment, and retrosynthesis prioritizes structural disassembly. These variations highlight the model's capacity to adjust its reasoning process, driven by the synergy between prediction specialists and inference specialists. This adaptability enhances overall performance across diverse molecular tasks, as validated by consistent output accuracy. The case studies demonstrate the practical impact of these reasoning chains. Each task employs a distinct sequence of logical steps, yielding high-fidelity outputs, including validated SMILES strings, accurate permeability predictions, and feasible reactant pairs.

Discussion

Molecular multi-task models. Our model offers distinct advantages over existing molecular multi-task models by integrating domain-specific reasoning and achieving superior performance across a wide range of molecular tasks. Unlike traditional approaches that rely heavily on data-driven predictions, our framework leverages data and specialist synergy to enhance adaptability, distinguishing itself from conventional MoE models through task-tailored specialist collaboration, as evidenced by its outperformance of baseline LLMs. However, a limitation arises from the computational overhead of managing multiple specialists, which may hinder scalability for extremely large datasets. Future research could focus on optimizing resource allocation and exploring continuous learning strategies to adapt dynamically to evolving datasets.

Knowledge-integrated reasoning models. Building on this multi-task foundation, our approach excels in embedding chemical knowledge into reasoning processes, setting it apart from knowledge-integrated models that often lack dynamic adaptability. By employing CoT reasoning and reinforcement learning, the model generates interpretable outputs, bridging the gap between data-driven and science-grounded paradigms. A notable limitation is the dependency on high-quality CoT datasets, which are currently limited in size and scope, potentially restricting generalization. Looking ahead, we aim to enrich these datasets through automated knowledge extraction and iterative refinement, broadening the model's versatility across diverse molecular scenarios.

Interpretability. The strength of our model lies in its ability to deliver transparent reasoning chains through CoT processes, thereby enhancing trustworthiness in molecular science with decision-making insights grounded in scientific principles. This clarity distinguishes it from opaque black-box models, boosting reliability for applications such as drug discovery. Yet, the complexity of certain molecular tasks can lead to overly detailed reasoning chains, potentially confusing users or masking key insights. To overcome this, future research may explore interactive visualization interfaces or condensed reasoning formats, thereby strengthening the model's practical dependability in scientific settings.

LLM-based molecular science agents. Leveraging the multi-task and reasoning capabilities of our model, it serves as a solid framework for advancing LLM-based molecular science agents. Its capacity to blend specialist expertise and adapt to varied demands establishes it as a flexible core for intelligent agents. For instance, Google DeepMind's AI Co-Scientist, a multi-agent system powered by Gemini 2.0, acts as a virtual collaborator for generating hypotheses and research proposals [40]. Similarly, Sakana AI's "The AI Scientist" exemplifies how such agents can automate complex research processes through an automated pipeline for end-to-end scientific paper generation and discovery [41]. These agents are adept at navigating complex workflows, employing reasoning akin to that of human experts. To overcome the limitation of static knowledge integration identified earlier, future developments could incorporate automated tool invocation and data querying. Additionally, real-time feedback mechanisms and interactive training could be integrated. This paves the way for fully autonomous molecular science assistants. Moreover, it drives innovation in fields like drug discovery and materials design.

Conclusion

In summary, this study introduces a reasoning model that redefines molecular science through enhanced reasoning and superior multi-task performance. By integrating chemical knowledge via CoT reasoning, leveraging data and specialist synergy, and employing reinforcement learning, the model outperforms baseline LLMs across 10 molecular tasks, demonstrating significant gains in accuracy and interpretability. These advancements pave the way for a new paradigm in molecular research, fostering intelligent design solutions beyond traditional data-driven approaches. Future directions include optimizing computational efficiency, expanding high-quality CoT datasets, and developing real-time interactive agents, promising further breakthroughs in the field.

Methods

Dataset construction

The dataset is sourced from a variety of tasks aligned with the LLaSMol framework [21], encompassing 10 molecular science challenges. These include property prediction tasks such as BBBP, ESOL (Solubility), Lipophilicity, and ClinTox, drawn from the MoleculeNet benchmark [29], alongside molecule captioning and generation derived from ChEBI-20 [12]. Additionally, IUPAC-to-SMILES and SMILES-to-IUPAC translations, as well as forward and retrosynthesis reaction predictions, are sampled from LLaSMol. This diverse task set ensures comprehensive coverage of molecular reasoning and prediction capabilities, forming the foundation for our model's multi-task performance. The construction of the 93K instruction dataset and 3.6K curated dataset involves a structured process, as depicted in Figure **??** (a).

Instruction dataset. The 93K instruction dataset is meticulously crafted from an extensive initial pool, diverging from LLaSMol's approach by refining its voluminous IUPAC-to-SMILES, SMILES-to-IUPAC translations, and reaction prediction data. While these data offer depth, their scale can lead to diminishing training efficiency beyond a certain threshold. To counter this, we employ a strategic sampling compression technique, which reduces the dataset size while preserving its quality. This process, guided by Figure **??** (a), ensures optimal performance. The dataset is further shaped through a stratified sampling method based on molecular feature distributions, as depicted in Figure **??** (b). Starting with over 2 million source data points, we curate the 93K instruction dataset, for which the data volume for each remains at or below the ten-thousand level to optimize training efficiency. This process leverages molecular representations derived from the DeepSeek-7B model, followed by dimensionality reduction and selection based on feature distributions to obtain subsets. The sampling can be approximated as:

$$S_i = P_i(\text{UMAP}(f_i(R_i))), \tag{3}$$

where S_i represents the sampled subset for the i-th task, $f_i(R_i)$ is the feature extraction function from the DeepSeek-7B model followed by UMAP for dimensionality reduction of the representations R_i, and P_i denotes a sampling probability function based on the resulting feature distribution. For the remaining six tasks, the datasets are directly incorporated into the instruction dataset without undergoing this sampling process, ensuring a comprehensive integration of all task-specific data into the final 93K dataset. The complete 93K instruction dataset I is then aggregated from these subsets, formulated as:

$$I = \sum_{i=1}^{N} S_i, \tag{4}$$

where I is the integrated instruction dataset, and N is the number of task-specific subsets. Subsequently, we partition the dataset I into training, validation, and test sets at an 8:1:1 ratio.

Curated dataset. The 3.6K curated dataset is derived as a high-quality subset from the instruction dataset I, specifically designed to support the subsequent annotation of CoT data. This dataset undergoes a rigorous selection process, aligning with the stratified sampling approach to maintain feature diversity and task relevance. Its construction complements I by providing a focused resource for refining inference capabilities, with proportional adjustments mirroring the task distributions. The data extraction method for the curated dataset is consistent with the approach used for the instruction dataset subsets (see equation 3).

Knowledge annotation and denoising. The annotation and denoising of the Molecular CoT dataset follow a rigorous workflow, illustrated in Figure **??** (c). Initially, we utilize DeepSeek-R1 in a dialogue-driven format to prompt reasoning on our questions, validated against correct answers. Successful query-think-answer triplets are extracted into the CoT dataset. For incorrect responses, DeepSeek-R1 is prompted to self-correct, though these answers often contain anchor segments (e.g., "I already know the correct answer is ..."), necessitating denoising. Using DeepSeek-V3, we cleanse contradictory content, with 5% of the data manually validated to confirm the quality of the resulting 3.5K refined CoT dataset.

Model

Our reasoning model for molecular science is built upon the DeepSeek-7B pre-trained LLM, leveraging a multi-specialist layer integrated within its decoder architecture to handle diverse molecular tasks. The overall framework combines data synergy across tasks with specialist synergy for enhanced reasoning, as illustrated in Figure **??**. The training strategy follows a three-step process: representation learning via instruction fine-tuning on 74.5K data points,

knowledge infusion through CoT fine-tuning on 3.6K samples, and knowledge alignment using reinforcement learning with knowledge-guided rewards, as depicted in Figure **??**. This approach embeds chemical logic into the model's inference, enabling task-adaptive reasoning that mimics molecular scientists.

Model architecture. The model architecture extends the standard Transformer decoder with a multi-specialist layer, allowing for efficient handling of multi-task molecular scenarios. At its core, the model employs the Qwen architecture, a GPT variant, which processes tokenized inputs through self-attention mechanisms. The attention computation for a given input sequence is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{5}$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the keys. This mechanism captures dependencies in molecular representations such as SMILES strings. To enable efficient adaptation without full parameter tuning, we incorporate Low-Rank Adaptation (LoRA) [42] into the Transformer layers. LoRA approximates weight updates with low-rank matrices, formulated as:

$$W' = W + \Delta W = W + BA, \tag{6}$$

where W is the original weight matrix, $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$ are low-rank matrices with rank $r \ll \min(d, k)$, and ΔW represents the update.

Multi-specialist layer. This layer introduces specialist groups and a router for task-specific processing, promoting data synergy and specialist synergy among specialists. We define eight specialist groups, grouped by output formats and data distributions to realize data synergy (joint training of data containing similar knowledge, such as text-based tasks like molecule captioning and SMILES-to-IUPAC, while isolating disparate knowledge, such as separate groups for classification tasks like BBBP and ClinTox due to distinct feature distributions): (1) Molecule captioning (text output); (2) SMILES to IUPAC (text output, distinct data); (3) Molecular generation and IUPAC to SMILES (SMILES output); (4) Forward reaction prediction and retrosynthesis (reaction/SMILES output); (5) BBBP (Yes/No classification); (6) Clinical Toxicity (Yes/No classification, separate for data variance); (7) Solubility (float regression); (8) Lipophilicity (float regression).

The synergistic output is computed as:

$$O = \sum_{i=1}^{N} r_i \cdot sg_i(q), \tag{7}$$

where q is the input query, sg_i is the function of specialist group i, r_i is the routing weight, and $N = 8$ is the number of groups. This weighted aggregation enables specialist synergy (collaboration between prediction specialists excelling in molecular representation and inference specialists proficient in structured reasoning), resolving knowledge conflicts and enhancing multi-task inference.

Training strategy. The training strategy begins with representation learning through instruction fine-tuning on 74.5K multi-task instruction data, unifying diverse molecular tasks under a common prompt format. This step optimizes the cross-entropy loss:

$$\mathcal{L}_{\text{inst}} = -\sum_{t=1}^{T} \log p(y_t | y_{<t}, x), \tag{8}$$

where x is the input prompt, y is the target sequence, and T is the sequence length. Next, knowledge infusion via CoT fine-tuning on 3.6K curated samples embeds chemical reasoning into the model. At its core, CoT prompting enhances LLMs by decomposing complex problems into intermediate reasoning steps, mimicking human-like logical progression to improve accuracy and interpretability in tasks requiring multi-step inference. This is particularly suited to molecular science, where it bridges structural analysis to predictive outcomes. By augmenting inputs with explicit reasoning traces and structuring outputs as ⟨think⟩ steps followed by ⟨answer⟩, our approach fosters deliberate, scientist-like cognition. The loss remains cross-entropy but targets CoT-augmented sequences to align generation with chemical logic.

Finally, knowledge alignment employs the REINFORCE algorithm, a policy gradient method that optimizes the model's generation policy by maximizing expected rewards over reasoning-augmented sequences. Rewards are first standardized to reduce variance:

$$\hat{r} = \frac{r - \mathbb{E}[r]}{\sigma_r + \epsilon}, \tag{9}$$

where $\mathbb{E}[r]$ is the mean reward, σ_r is the standard deviation, and $\epsilon = 10^{-6}$ is a small constant for numerical stability. The REINFORCE loss is then computed as

$$\mathcal{L}_{\text{RL}} = -\mathbb{E}\left[\sum_t \log \pi(a_t|s_t) \cdot \hat{r}\right], \tag{10}$$

where $\pi(a_t|s_t)$ denotes the log-probability of action a_t (token) given state s_t (context). The overall reward r integrates task performance and reasoning quality:

$$r = \alpha \cdot r_{\text{answer}} + \beta \cdot r_{\text{think}}, \tag{11}$$

with $\alpha = 0.8$, $\beta = 0.2$, r_{answer} task-specific (e.g., accuracy for classification like BBBP/Clintox, inverse RMSE for regression like ESOL/Lipo, BLEU or SMILES similarity for generation/reaction tasks), and r_{think} assessing reasoning via length (Gaussian centered at 1569 characters) and diversity (unique word ratio). This guides the model to resolve conflicts and align with molecular scientific logic.

Hardware and software conditions. The hardware configuration includes a single H20-NVLink GPU with 96GB memory, paired with 16 vCPUs from an AMD EPYC 9K84 96-Core Processor, operating on Ubuntu 22.04. The software environment is built on Python 3.12, with PyTorch 2.5.1 as the core deep learning framework, utilizing CUDA 12.4 for GPU acceleration. Key libraries include transformers 4.51.3 for the pre-trained LLM, pandas 2.2.3 for data handling, peft 0.15.2 for LoRA implementation, and rdkit 2025.3.2 for molecular structure processing. Optimization is performed using the Adam optimizer with a learning rate of 1×10^{-4}, with batch sizes ranging from 4 to 16 for prediction specialist training and 1 to 2 for inference specialist training.

Acknowledgment

This work is supported by the Guangdong Science and Technology Programme (Grant No. 2024B0101010003). This work is also supported by grants from the National Natural Science Foundation of China (62172456 and 62372484) and Pengcheng Cloud-Brain.

References

[1] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10):573–584, 2020.

[2] JS Huang, JX Liew, AS Ademiloye, and Kim Meow Liew. Artificial intelligence in materials modeling and design. *Archives of Computational Methods in Engineering*, 28(5):3399–3413, 2021.

[3] Yu Zhang, Yang Han, Shuai Chen, Ruijie Yu, Xin Zhao, Xianbin Liu, Kaipeng Zeng, Mengdi Yu, Jidong Tian, Feng Zhu, et al. Large language models to accelerate organic chemistry synthesis. *Nature Machine Intelligence*, pages 1–13, 2025.

[4] Daniel S Wigh, Jonathan M Goodman, and Alexei A Lapkin. A review of molecular representation in the age of machine learning. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1603, 2022.

[5] Longlong Li, Yipeng Zhang, Guanghui Wang, and Kelin Xia. Kolmogorov–arnold graph neural networks for molecular property prediction. *Nature Machine Intelligence*, pages 1–9, 2025.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[7] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[9] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[10] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020.

[11] John Schulman, Barret Zoph, Christina Kim, Jacob Hilton, Jacob Menick, Jiayi Weng, Juan Felipe Ceron Uribe, Liam Fedus, Luke Metz, Michael Pokorny, et al. Chatgpt: Optimizing language models for dialogue. *OpenAI blog*, 2(4), 2022.

[12] Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. Translation between molecules and natural language. In *2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, pages 375–413. Association for Computational Linguistics (ACL), 2022.

[13] Qizhi Pei, Lijun Wu, Kaiyuan Gao, Xiaozhuan Liang, Yin Fang, Jinhua Zhu, Shufang Xie, Tao Qin, and Rui Yan. Biot5+: Towards generalized biological understanding with iupac integration and multi-task tuning. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 1216–1240, 2024.

[14] Gary L Long and James D Winefordner. Limit of detection. a closer look at the iupac definition. *Analytical chemistry*, 55(7):712A–724A, 1983.

[15] Pengfei Liu, Yiming Ren, Jun Tao, and Zhixiang Ren. Git-mol: A multi-modal large language model for molecular science with graph, image, and text. *Computers in biology and medicine*, 171:108073, 2024.

[16] Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Chaowei Xiao, and Animashree Anandkumar. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12):1447–1457, 2023.

[17] Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, Anh TN Nguyen, Lauren T May, Geoffrey I Webb, and Shirui Pan. Large language models for scientific discovery in molecular property prediction. *Nature Machine Intelligence*, pages 1–11, 2025.

[18] Pengfei Liu, Jun Tao, and Zhixiang Ren. A self-feedback knowledge elicitation approach for chemical reaction predictions. *Engineering Applications of Artificial Intelligence*, 156:111112, 2025.

[19] Taojie Kuang, Pengfei Liu, and Zhixiang Ren. Impact of domain knowledge and multi-modality on intelligent molecular property prediction: A systematic survey. *Big Data Mining and Analytics*, 7(3):858–888, 2024.

[20] Pengfei Liu, Jun Tao, and Zhixiang Ren. A quantitative analysis of knowledge-learning preferences in large language models in molecular science. *Nature Machine Intelligence*, 7(2):315–327, 2025.

[21] Botao Yu, Frazier N Baker, Ziqi Chen, Xia Ning, and Huan Sun. Llasmol: Advancing large language models for chemistry with a large-scale, comprehensive, high-quality instruction tuning dataset. In *First Conference on Language Modeling*.

[22] Di Zhang, Wei Liu, Qian Tan, Jingdan Chen, Hang Yan, Yuliang Yan, Jiatong Li, Weiran Huang, Xiangyu Yue, Dongzhan Zhou, Shufei Zhang, Mao Su, Hansen Zhong, Yuqiang Li, and Wanli Ouyang. Chemllm: A chemical large language model, 2024.

[23] Xiang Zhuang, Keyan Ding, Tianwen Lyu, Yinuo Jiang, Xiaotong Li, Zhuoyi Xiang, Zeyuan Wang, Ming Qin, Kehua Feng, Jike Wang, et al. Advancing biomolecular understanding and design following human instructions. *Nature Machine Intelligence*, pages 1–14, 2025.

[24] Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card, 2024.

[25] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[26] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

[27] Eric Wang, Samuel Schmidgall, Paul F. Jaeger, Fan Zhang, Rory Pilgrim, Yossi Matias, Joelle Barral, David Fleet, and Shekoofeh Azizi. Txgemma: Efficient and agentic llms for therapeutics, 2025.

[28] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[29] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

[30] Satanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72, 2005.

[31] Joseph L Durant, Burton A Leland, Douglas R Henry, and James G Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of chemical information and computer sciences*, 42(6):1273–1280, 2002.

[32] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2025.

[33] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, 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.

[34] Anthropic. Claude 3.5 sonnet model card and technical overview, 2025. Model Card Addendum, Anthropic Research.

[35] OpenAI. Openai o3 and o4-mini system card, 2025. Technical Report, OpenAI Safety.

[36] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

[37] Mistral AI Team. Un ministral, des ministraux, 2024. Technical Report, Mistral AI Blog.

[38] Meta AI Team. The llama 4 herd: The beginning of a new era of natively multimodal ai innovation, 2025. Technical Report, Meta AI Blog.

[39] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29), 2018.

[40] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Anil Palepu, Petar Sirkovic, Artiom Myaskovsky, Felix Weissenberger, Keran Rong, Ryutaro Tanno, Khaled Saab, Dan Popovici, Jacob Blum, Fan Zhang, Katherine Chou, Avinatan Hassidim, Burak Gokturk, Amin Vahdat, Pushmeet Kohli, Yossi Matias, Andrew Carroll, Kavita Kulkarni, Nenad Tomasev, Yuan Guan, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R D Costa, José R Penadés, Gary Peltz, Yunhan Xu, Annalisa Pawlosky, Alan Karthikesalingam, and Vivek Natarajan. Towards an ai co-scientist, 2025.

[41] Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery, 2024.

[42] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.