

# QUANTUMLLMINSTRUCT: A 500K LLM INSTRUCTION-TUNING DATASET WITH PROBLEM-SOLUTION PAIRS FOR QUANTUM COMPUTING.

 Shlomo Kashani  [skashan2@alumni.jh.edu](mailto:skashan2@alumni.jh.edu)

 Johns Hopkins University.

 Code: [QuantumLLMInstruct HF Code](#)

 Dataset: [QuantumLLMInstruct HF DB](#)

 Project: [QuantumLLMInstruct Project](#)

## ABSTRACT

We present **QuantumLLMInstruct (QLMMI)**, an innovative dataset featuring over 500,000 meticulously curated instruction-following problem-solution pairs designed specifically for quantum computing – **the largest and most comprehensive dataset of its kind**. Originating from over 90 primary seed domains and encompassing hundreds of subdomains autonomously generated by LLMs, QLMMI marks a transformative step in the diversity and richness of quantum computing datasets.

Designed for **instruction fine-tuning**, QLMMI seeks to significantly improve LLM performance in addressing complex quantum computing challenges across a wide range of quantum physics topics. While Large Language Models (LLMs) have propelled advancements in computational science with datasets like **Omni-MATH** and **OpenMathInstruct**, these primarily target Olympiad-level mathematics, leaving quantum computing largely unexplored.

The creation of QLMMI follows a rigorous **four-stage methodology**. Initially, foundational problems are developed using predefined templates, focusing on critical areas such as synthetic Hamiltonians, QASM code generation, Jordan-Wigner transformations, and Trotter-Suzuki quantum circuit decompositions. Next, detailed and domain-specific solutions are crafted to ensure accuracy and relevance. In the third stage, the dataset is enriched through advanced reasoning techniques, including Chain-of-Thought (CoT) and Task-Oriented Reasoning and Action (ToRA), which enhance problem-solution diversity while adhering to strict mathematical standards. Lastly, a zero-shot Judge LLM performs self-assessments to validate the dataset’s quality and reliability, minimizing human oversight requirements.

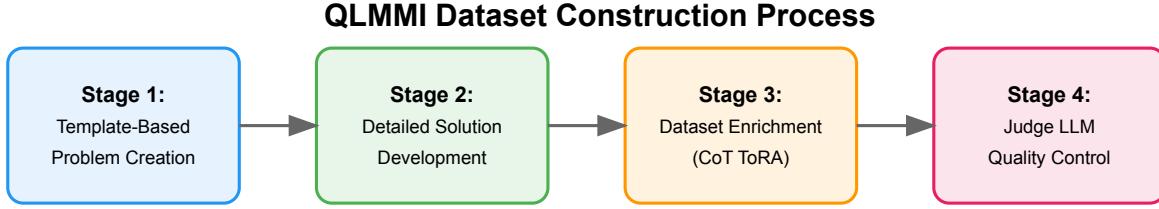
To foster collaboration in this evolving field, we utilize the **Qwen-2.5-Coder** family of models, selected for their strong mathematical reasoning capabilities and permissive licensing. **All associated code and the complete dataset are openly available**, providing a foundation for future breakthroughs in quantum computing applications of LLMs. Our pipeline is generic and can be easily adapted to generate instruction-tuning datasets across any domain. This work focuses exclusively on the **generation of synthetic data for quantum computing tasks, deliberately excluding the training or fine-tuning of large LLMs**. No trained model is provided in this study. Instead, we prioritize creating a high-quality dataset to support future research without the computational demands of training or fine-tuning.

## CONTENTS

### 1 INTRODUCTION

The rapid advancement of Large Language Models (LLMs) has catalyzed transformative breakthroughs across diverse fields, including mathematics [WW<sup>+</sup>22, BMR<sup>+</sup>20, WTB<sup>+</sup>22] and natural language processing. Notable models such as GPT [BMR<sup>+</sup>20, WWS<sup>+</sup>22b], Qwen [YZH<sup>+</sup>24, Tea24b], and LLaMA [TLJ<sup>+</sup>23, TMS<sup>+</sup>23] have demonstrated exceptional capabilities in solving complex mathematical problems using advanced reasoning methodologies like Chain-of-Thought (CoT) [WW<sup>+</sup>22, WWS<sup>+</sup>22a] and Task-Oriented Reasoning and Action (ToRA) [W<sup>+</sup>23]. However, their potential in quantum computing—a domain characterized by intricate reasoning complexity and abstract mathematical frameworks—remains relatively untapped.

**Quantum Physics Domains** Quantum computing tasks, such as simulating time evolution under Hamiltonians [Tro59, Suz76], compressing quantum models with Yang-Baxter equations [PMS<sup>+</sup>14, Hei28], or designing



**Figure 1:** Dataset Creation Workflow for QuantumLLMInstruct. The process comprises four stages: **Stage I** generates foundational problems with LLMs, focusing on tasks such as synthetic Hamiltonians, QASM code, Jordan-Wigner transformations, and Trotter-Suzuki decompositions for models like Ising [Isi25] and Heisenberg spin chains [Hei28]. **Stage II** develops solutions to these problems using LLMs. **Stage III** enriches the dataset using CoT [WW<sup>+</sup>22] and ToRA [W<sup>+</sup>23]. **Stage IV** validates the dataset through self-critique mechanisms [MTG<sup>+</sup>24] performed by a Judge LLM, ensuring high-quality entries.

circuits for advanced algorithms [Llo96, CBSG17], demand a combination of computational rigor and domain expertise. These tasks involve key quantum concepts such as Hamiltonians, entangled states, and unitary transformations, making manual approaches error-prone, resource-intensive, and challenging to scale. The increasing complexity and demand for scalability highlight the critical need for autonomous systems capable of generating, solving, and validating quantum problems at scale.

To address these challenges, we introduce **QuantumLLMInstruct (QLMMI)**, a comprehensive dataset and framework (Fig. 1) leveraging state-of-the-art LLMs for quantum problem-solving. QLMMI automates the generation of quantum problem statements, derives solutions, and refines outputs through advanced reasoning techniques and self-assessment mechanisms [MTG<sup>+</sup>24, WW<sup>+</sup>22, ZPR<sup>+</sup>24]. By doing so, QLMMI establishes a robust foundational resource for tackling a wide range of quantum computing tasks (Fig. 3) with minimal human intervention.

The foundation of QLMMI lies in its reliance on carefully crafted templates (Fig. 3), ensuring domain specificity and mathematical rigor. These templates span over **90 predefined quantum computing domains** (see Sec. A), including quantum cryptography, spin chain models, and Trotter-Suzuki decompositions [Isi25, Hei28, Suz76]. While ensuring both diversity and precision, the dataset extends far beyond these predefined categories. Remarkably, LLMs demonstrated emergent capabilities, **autonomously generating hundreds of quantum sub-domains without explicit prompt engineering** [WW<sup>+</sup>22, WTB<sup>+</sup>22, MTG<sup>+</sup>24], underscoring the scalability and adaptability of LLMs in quantum problem-solving. This positions QLMMI as a versatile and comprehensive resource.

The dataset employs detailed natural language prompts paired with LaTeX-formatted mathematical expressions, ensuring clarity and precision. Insights from existing resources like Omni-MATH [GSY<sup>+</sup>24], OpenMathInstruct-1 [TMN<sup>+</sup>24], and MetaMath [Y<sup>+</sup>23a] were integrated to address challenges in quantum reasoning, circuit optimization, and algorithm design. Moreover, open-source LLMs like Qwen-2.5-Coder [YZH<sup>+</sup>24] were utilized to circumvent restrictive licensing, ensuring robust reasoning capabilities while promoting accessibility and scalability.

Here, (Fig. 2) we present a representative list of quantum computing problem types from the QLMMI dataset, highlighting its depth, rigor, and domain-specific relevance.

### 1.1 CHALLENGES OF SYNTHETIC DATA GENERATION VS. TRAINING AND FINE-TUNING

This work **focuses exclusively on the generation of synthetic data for quantum computing tasks**, deliberately excluding training or fine-tuning of large LLMs. No trained model is provided in this study. Instead, we prioritize creating a high-quality dataset to support future research without the computational demands of training or fine-tuning.

Fine-tuning large LLMs, such as GPT-4, requires extensive engineering and computational infrastructure [ALC<sup>+</sup>23]. Studies [WW<sup>+</sup>22, WWS<sup>+</sup>22a] further emphasize that the success of fine-tuning is contingent on precise data preparation and hyperparameter configuration. By concentrating on synthetic data generation, this work aligns with broader initiatives to democratize AI access through the use of pre-trained models and high-quality datasets [Y<sup>+</sup>23a, H<sup>+</sup>21]. Below, we outline the primary challenges that influenced this decision:

**Resource Requirements.** Fine-tuning LLMs like Qwen2.5-Math-72B necessitates high-performance GPU clusters (e.g., 8 x H100), which are prohibitively expensive and inaccessible to many researchers and smaller institutions [YZH<sup>+</sup>24, ALC<sup>+</sup>23]. The iterative nature of fine-tuning compounds these costs, requiring repeated runs to optimize model performance [WW<sup>+</sup>22, YYZ<sup>+</sup>24]. Additionally, generating high-quality fine-tuning datasets demands domain expertise to avoid introducing errors or biases [TLI<sup>+</sup>23, DQL<sup>+</sup>22, GBCB16]. Evaluation com-

**Sample problem 2: Spin Chains**

*Problem.* Consider the Heisenberg spin chain with the Hamiltonian:

$$H = J \sum_i \sigma_i^z \sigma_{i+1}^z + h \sum_i \sigma_i^x,$$

where  $J = 1$  and  $h = 0.5$ . Determine the ground state energy and eigenvalues.

**Domain:** SPIN CHAINS

**Sample problem 1: Diagonalization of Two-Spin Hamiltonian**

*Problem.* Consider a Hamiltonian for two spin-1/2s:

$$H = A\sigma_1^z \otimes I_2 + B(\sigma_1^x \otimes \sigma_2^x + \sigma_1^y \otimes \sigma_2^y).$$

Find all the eigenvalues of  $H$  without using a computer.

**Domain:** DIAGONALIZATION

**Sample problem 3: PennyLane Circuit Analysis**

*Problem.* Given the following PennyLane program:

```
import pennylane as qml
import numpy as np
dev = qml.device('default.qubit',
wires=2)
@qml.qnode(dev):
def circuit(theta):
    qml.Hadamard(wires=0)
    qml.CNOT(wires=[0, 1])
    qml.RZ(theta, wires=1)
    return qml.expval(qml.PauliZ(1))
```

(a) Describe the quantum operations performed by this circuit. (b) Generate the corresponding quantum circuit diagram.

**Domain:** PENNYLANE CIRCUIT ANALYSIS

**Sample problem 4: Gibbs State Preparation**

*Problem.* Describe the procedure to prepare a Gibbs state at inverse temperature  $\beta$  for a quantum Hamiltonian  $H$ . Discuss the required resources and the role of thermalization processes.

**Domain:** GIBBS STATE PREPARATION

**Sample problem 5: The Classical Shadow of Unknown Quantum States**

*Problem.* Use the classical shadow method to estimate the expectation value of the observable  $\sigma^z$  for an unknown quantum state  $|\psi\rangle$ . Explain the protocol and compute the sample complexity.

**Domain:** CLASSICAL SHADOWS

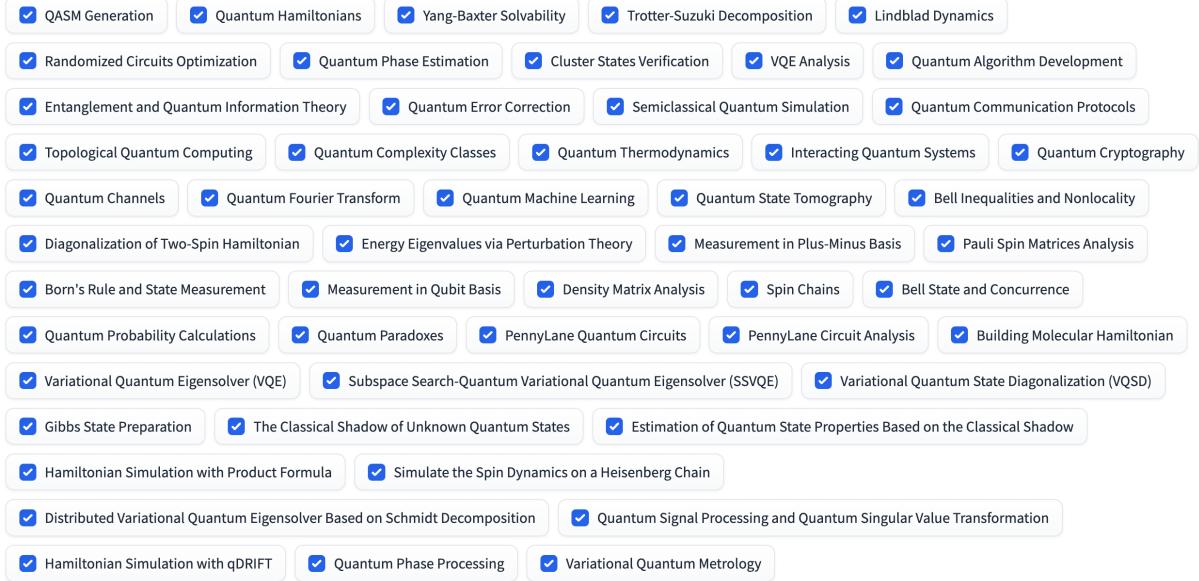
Figure 2: **QuantumLLMInstruct** encompasses a wide range of quantum physics problems meticulously designed using carefully crafted templates (Fig. 3), ensuring domain specificity and mathematical rigor. Spanning over **90 predefined quantum computing domains** (see Sec. A)—including quantum cryptography, spin chain models, and Trotter-Suzuki decompositions [Isi25, Hei28, Suz76]—the dataset guarantees both diversity and precision. Beyond these templates, the LLMs demonstrated emergent capabilities, autonomously generating hundreds of new quantum domains without explicit prompt engineering [WW<sup>+</sup>22, WTB<sup>+</sup>22, MTG<sup>+</sup>24], establishing QLMMI as a versatile and scalable resource in quantum problem-solving. These representative samples illustrate the dataset’s depth in problem variety, rigor, and domain-specific relevance.

plexities further complicate the process, especially for niche domains like quantum computing, where datasets such as QDataset [SDea21] or MNISQ [WZea23] are limited in scope and availability [GTA<sup>+</sup>23].

**Expertise Barriers.** Developing fine-tuning datasets for specialized fields like quantum physics requires significant domain expertise [TLI<sup>+</sup>23, BMR<sup>+</sup>20]. Errors in dataset preparation or training methodologies can degrade model performance, while robust evaluation frameworks are often overlooked [WTB<sup>+</sup>22, WWS<sup>+</sup>22a], leading to suboptimal outcomes.

**Evaluation Complexities.** Evaluation of fine-tuned models is particularly challenging in niche domains like quantum computing. Models frequently exhibit inconsistent performance across datasets or evaluation contexts [GTA<sup>+</sup>23], and the scarcity of task-specific datasets, such as QDataset [SDea21] or MNISQ [WZea23], exacerbates these challenges.

**Why Focus Solely on Synthetic Data Generation?** Considering these challenges, this work emphasizes synthetic data generation as a cost-efficient and reproducible alternative. Generating synthetic data is significantly less resource-intensive than fine-tuning, enabling scalable dataset creation [Y<sup>+</sup>23b]. Moreover, synthetic data pipelines are transparent, hardware-independent, and consistent, ensuring reproducibility [ALC<sup>+</sup>23]. Additionally, high-quality datasets empower researchers with limited resources to explore quantum AI, fostering broader accessibility and democratization [T<sup>+</sup>24]. By focusing on synthetic data generation, this work establishes a robust foundation for quantum computing research, enabling the exploration of advanced tasks with pre-trained models while circumventing the computational and expertise barriers associated with fine-tuning.



**Figure 3:** Partial list of the core domains of QuantumLLMInstruct. QLMMI leverages predefined templates to ensure domain relevance and mathematical precision in critical areas such as quantum phase estimation, Hamiltonian analysis, and randomized circuit optimization. Beyond these predefined categories, LLMs autonomously generated hundreds of novel quantum computing domains, highlighting their adaptability and scalability.

## 2 CORE FEATURES OF QUANTUMLLMINSTRUCT (QLMMI)

The defining attributes of QLMMI include its extensive domain coverage and a meticulously designed four-stage creation process, ensuring both breadth and depth in quantum problem-solving:

- **Stage I: Foundational Problem Generation.** The cornerstone of QLMMI lies in the use of meticulously crafted templates (see Fig. 3) to ensure domain specificity and mathematical precision. These templates facilitate the generation of problems across diverse quantum computing domains, encompassing tasks such as synthetic Hamiltonians, QASM code generation, Jordan-Wigner transformations, Trotter-Suzuki decompositions, quantum phase estimation, variational quantum eigensolvers (VQE), cluster state verification, randomized circuit optimization, spin dynamics on Heisenberg chains, Yang-Baxter solvability, quantum signal processing, and Gibbs state preparation for quantum thermodynamics. Classical models like Ising [Isi25] and Heisenberg spin chains [Hei28] serve as central elements in these tasks, ensuring mathematical rigor and applicability. Notably, QLMMI extends far beyond these predefined templates, as LLMs exhibit emergent capabilities to autonomously generate hundreds of new quantum domains without explicit prompt engineering. This adaptability highlights the scalability of LLMs in quantum problem-solving, positioning QLMMI as a highly versatile and comprehensive resource. For this stage, the **Qwen-2-coder-instruct** model is employed, offering exceptional precision in generating domain-specific problems.
- **Stage II: Solution Generation.** To ensure the dataset’s completeness and accuracy, solutions are developed for the generated problems. This stage leverages the **Qwen-2-coder-instruct** model, which is optimized for crafting detailed and contextually accurate solutions tailored to quantum computing challenges.
- **Stage III: Dataset Enrichment.** Advanced reasoning frameworks, such as Chain-of-Thought (CoT) [WW<sup>+</sup>22] and Task-Oriented Reasoning and Action (ToRA) [W<sup>+</sup>23], are integrated to enhance the quality and diversity of problem-solution pairs. These methods ensure that the dataset covers a wide spectrum of quantum challenges, improving both its depth and adaptability. The **Qwen-2-math-instruct** model, renowned for its mathematical precision and reasoning prowess, is utilized in this stage to enrich the dataset effectively.
- **Stage IV: Validation and Quality Assurance.** The final stage applies robust validation through a Judge LLM employing self-critique techniques [MTG<sup>+</sup>24] to ensure the reliability and accuracy of the dataset. This process significantly minimizes human oversight while maintaining high-quality standards. The **Qwen-2-math-instruct** model, with its expertise in evaluating mathematical correctness and logical coherence, is employed to rigorously assess the entries, guaranteeing a dependable and accurate dataset.

**Contributions** This work introduces several key contributions:

1. **Comprehensive Dataset:** A scalable dataset containing over 500,000 rigorously curated quantum problem-solution pairs designed for fine-tuning LLMs for quantum-specific tasks. An example question is depicted in (Fig. 6).
2. **Four-Stage Workflow:** A robust framework combining predefined templates, Chain-of-Thought (CoT) reasoning [WW<sup>22</sup>], Task-Oriented Reasoning and Action (ToRA) [W<sup>23</sup>], and self-assessment mechanisms [MTG<sup>24</sup>] for automated quality assurance.
3. **Scalable Infrastructure:** Integration of tools like DuckDB and Gradio enables efficient dataset management and interactive exploration of quantum problems [BIS<sup>18</sup>].
4. **Open Access:** QLMMI provides public access to raw datasets, data generation code, and training scripts, fostering collaboration in quantum computing research.

**Future Directions** Future work aims to extend QLMMI’s impact through:

- **Advanced Model Fine-Tuning:** Expanding evaluations on models such as Qwen-2.5-Math and Qwen-2.5-Code-Instruct to further enhance quantum reasoning capabilities [YZH<sup>24</sup>, LSX<sup>23</sup>].
- **Integration with Retrieval-Augmented Models:** Incorporating advanced retrieval mechanisms to improve contextual understanding for complex quantum tasks [GTA<sup>23</sup>, SMY<sup>23</sup>].
- **Cross-Domain Applications:** Extending problem templates and solutions to intersect with fields like quantum chemistry and cryptography, broadening QLMMI’s interdisciplinary impact [Pre18, Llo96].

### 3 RELATED WORK

The development of large-scale datasets for mathematical and scientific reasoning has been a key focus in recent years. In (Tab. 1) we highlight some of the recent works synthesizing datasets in mathematics and physics.

Dataset	Focus/Domain	Size/Scale	Methodology
MathInstruct [Y <sup>23b</sup> ]	Math/Instruction Tuning	1.8M examples	CoT and PoT
MetaMathQA [Y <sup>23a</sup> ]	Enhanced Reasoning/Math	100K examples	Fine-tuning with GSM8K
OpenMathInstruct [T <sup>24</sup> ]	Synthetic Math Solutions/Math	1.8M examples	Hybrid instruction tuning
WizardCoder [LXZ <sup>23</sup> ]	Code-Specific Reasoning/Math	200K+ examples	Evol-instruct with StarCoder
Lila [MFL <sup>23</sup> ]	Math Benchmark	272K examples	-
MathCodeInstruct [WRZ <sup>23</sup> ]	Math with Code	80K examples	GPT-4 + Self
ToRA [ZWL <sup>23</sup> ]	Task-Oriented Math Reasoning	16K examples	GPT-4
OMNI-MATH [GSY <sup>24</sup> ]	Olympiad-Level Math	N/A	GPT-3.5
QDataset [SDea21]	Quantum Computing/Physics	52 datasets	-
MNISQ [WZea23]	Quantum Numerical Simulations	4.95M datapoints	-
QCircuitNet [Tea24a]	Quantum Algorithms/Physics	N/A	Closed-source LLMs
<b>QuantumLLMInstruct (ours)</b>	<b>Broad Quantum Domains</b>	<b>500K examples</b>	<b>Qwen-Based</b>

Table 1: Comparison of datasets for quantum and mathematical reasoning, highlighting size, generating methodology, and focus domain. QLMMI sets itself apart with its extensive domain coverage and quantum-specific tasks.

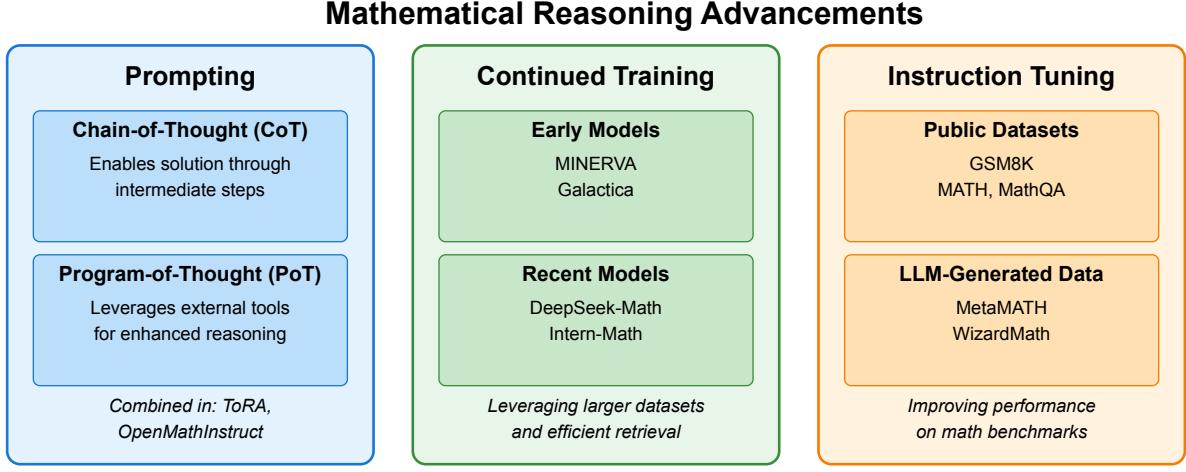
**Mathematical and Physics Datasets.** In recent years, there has been growing interest in leveraging large language models (LLMs) to address tasks in quantum computing, mathematical reasoning, and physics. These efforts have resulted in diverse contributions, as detailed below.

**Instruction Tuning.** Instruction tuning has emerged as a pivotal technique for tailoring LLMs to specific tasks. Two primary approaches to instruction tuning have been explored:

- **Human-Written Data:** Frameworks such as FLAN [WBD<sup>21</sup>], T0 [SDSC21], and SuperNI [WXL<sup>22</sup>] rely on pre-existing human-labeled datasets. While these datasets offer high reliability, their size and diversity are constrained by the high cost of manual annotation.
- **Synthesized Data:** Models like Self-Instruct [WLD<sup>23</sup>], WizardLM [XSZ<sup>23</sup>], and GPT4-Alpaca [PKC<sup>23</sup>] generate datasets via powerful LLMs like GPT-4 [ALC<sup>23</sup>]. This approach scales easily but may introduce hallucinations, reducing data quality. The diversity of synthesized data also depends on the comprehensiveness of the seed dataset.

QLMMI adopts a synthesized data approach, leveraging instruction-tuned LLMs to generate, refine, and validate quantum problems. Advanced filtering mechanisms mitigate issues like noise and hallucination, ensuring high-quality instruction-tuning pairs. This aligns with efforts like OpenMathInstruct [T<sup>+</sup>24] and MetaMATH [Y<sup>+</sup>23a], which demonstrate the efficacy of synthesized datasets in improving LLM reasoning capabilities.

**Mathematical Reasoning.** Mathematical reasoning has seen substantial advancements through innovative prompting strategies, continued training, and instruction tuning.



■ **Figure 4:** Overview of mathematical reasoning advancements in LLMs. **Prompting** techniques include Chain-of-Thought (CoT) [N<sup>+</sup>22, WW<sup>+</sup>22] for intermediate reasoning steps and Program-of-Thought (PoT) [C<sup>+</sup>23, G<sup>+</sup>23a] for tool-augmented reasoning, combined in models like ToRA [G<sup>+</sup>23b] and OpenMathInstruct [T<sup>+</sup>24]. **Continued Training** progressed from early models like MINERVA [L<sup>+</sup>22] and Galactica [T<sup>+</sup>22] to advanced models such as DeepSeek-Math [S<sup>+</sup>24] and Intern-Math [Y<sup>+</sup>24]. **Instruction Tuning** leverages both public datasets (e.g., GSM8K [C<sup>+</sup>21], MATH [H<sup>+</sup>21], MathQA [A<sup>+</sup>19]) and LLM-generated data (MetaMATH [Y<sup>+</sup>23a], WizardMath [LSX<sup>+</sup>23]) to enhance mathematical capabilities.

QLMMI builds on these strategies by integrating CoT reasoning and PoT methodologies into its dataset generation pipeline, enhancing LLMs' ability to tackle mathematical challenges across quantum computing domains.

**Physics Reasoning.** Physics reasoning remains underexplored compared to mathematics. Benchmarks like MMLU [HZ<sup>+</sup>20], SciEval [S<sup>+</sup>23], and GPQA [R<sup>+</sup>23] provide evaluation frameworks but lack large-scale datasets tailored to physics.

Techniques like CoT [WWS<sup>+</sup>22b] and PoT [CMWC22] have shown promise in physics, particularly when integrated with intermediate validation [LLZ<sup>+</sup>22] and external computational tools. Fine-tuned models like WizardCoder [LXZ<sup>+</sup>23] demonstrate the potential of combining natural language reasoning with code generation, emphasizing the importance of domain-specific tuning.

A notable contribution in quantum computing is QCircuitNet [Tea24a], employing closed-source LLMs for a comprehensive range of quantum algorithms, from foundational primitives to advanced applications. However, QCircuitNet focuses on algorithm design, overlooking the broader spectrum of quantum domains addressed by QLMMI. Similarly, ManyBodyLLMs [PMT<sup>+</sup>24] explores GPT-4's ability to perform graduate-level quantum many-body physics calculations, achieving 87.5% adherence across Hartree-Fock mean-field theory tasks. While depth-focused, these works lack the breadth offered by QLMMI.

PhysQA [DZC<sup>+</sup>23] emphasizes elementary physics education through GPT-3.5-generated solutions to high school physics problems. It contrasts with QLMMI's focus on advanced quantum domains.

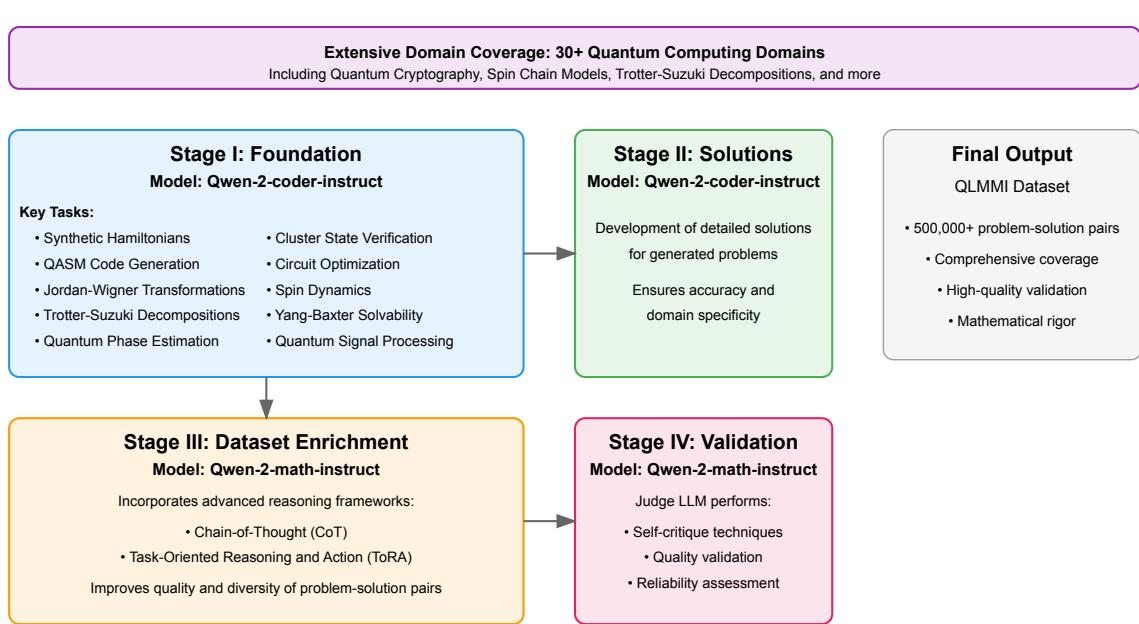
In contrast to these efforts, QLMMI integrates advanced reasoning techniques like CoT [WW<sup>+</sup>22, WWS<sup>+</sup>22b] and self-critique [MTG<sup>+</sup>24, SYW<sup>+</sup>22] to ensure robust problem-solving and high-quality validation. It spans domains like synthetic Hamiltonians and Trotter-Suzuki decompositions [Suz76], making it a versatile resource for addressing challenges in quantum computing.

1. **Instruction Generation:** One aligned LLM generates LaTeX-formatted quantum physics problems adhering to strict guidelines.
2. **Solution Generation:** A second LLM generates detailed solutions for these problems, enabling specialized fine-tuning of both problem and solution generation.

For a full list of instruction-tuned datasets, refer to Appendix ??.

## 4 DATA GENERATION PIPELINE

The prompts used at various stages are meticulously crafted to align with the capabilities of LLMs while ensuring consistency in the format of problems and solutions.



■ **Figure 5:** Dataset Creation Workflow for QuantumLLMInstruct. The process comprises four stages: **Stage I** generates foundational problems with LLMs, focusing on tasks such as synthetic Hamiltonians, QASM code, Jordan-Wigner transformations, and Trotter-Suzuki decompositions for models like Ising [Isi25] and Heisenberg spin chains [Hei28]. **Stage II** develops solutions to these problems using LLMs. **Stage III** enriches the dataset using Chain-of-Thought (CoT) [WW<sup>22</sup>] and Task-Oriented Reasoning and Action (ToRA) [W<sup>23</sup>]. **Stage IV** validates the dataset through self-critique mechanisms [MTG<sup>24</sup>], performed by a Judge LLM, ensuring high-quality entries.

### 4.1 LLM MODELS

The Qwen family of Large Language Models (LLMs) was employed for both instruction and response generation. These models are tailored to handle complex quantum physics tasks, providing high precision and adaptability across diverse quantum domains. The Qwen family offers models ranging from lightweight variants to high-performance versions, accommodating a broad spectrum of computational needs.

Model Name	Parameters (Billion)	Dataset Trained On	Focus Domain
Coder-1.5B-Instruct [Tea24b]	1.5	Code, Mathematical Problems	Instruction Tuning
Coder-3B-Instruct [Tea24b]	3	Code, Scientific Papers	Instruction Tuning
Coder-7B-Instruct [Tea24b]	7	Code and Mathematical Problems	Advanced Coding Tasks
Math-7B-Instruct [YZH <sup>24</sup> ]	7	Mathematics-Focused Datasets	Math Problem Solving
Coder-32B-Instruct [Tea24b]	32	Extensive Code and Scientific Datasets	Complex Scientific Tasks
Llama-3.2B-Instruct [TLI <sup>23</sup> ]	3	General-Purpose Instruction Dataset	General Reasoning

■ Table 2: Details of LLMs used in the Quantum Physics Problem Generator. Each model is characterized by its parameter count, training datasets, and target focus domain.

**Flexibility and Scalability** The Qwen2.5 models (Table 2) provide parameter configurations ranging from 0.5B to 72B, enabling users to tailor model usage based on available GPU memory and specific requirements. This ensures scalability for both research and production applications.

## Usage in Instruction and Response Generation

- **Instruction Generation:** Qwen2.5-Coder-\*B-Instruct models were used to create problem prompts, benefiting from their expertise in coding and mathematical formulation.
- **Response Generation:** Qwen2.5-Math-\*B-Instruct models generated step-by-step solutions with LaTeX-formatted precision, ensuring clarity and rigor.

### 4.2 STAGE 1: PROBLEM GENERATION

In this stage, the first LLM, a Qwen2.5-Coder-\*B-Instruct, is prompted to generate a quantum physics problem in LaTeX format. Each problem adheres to predefined domain types (see Sec. A), ensuring solvability and the inclusion of necessary mathematical precision. The following structured prompt is employed:

Illustrative Example: Generating a Quantum Physics Problem	
Instruction Prompt	Output Format and Example
<p>Generate a single detailed quantum physics problem for an exam in LaTeX format. Do not solve the problem.</p> <ul style="list-style-type: none"> <li>• Avoid additional explanations or comments outside of LaTeX.</li> <li>• Exclude unnecessary LaTeX imports (e.g., <code>\documentclass{}</code>, <code>\usepackage{}</code>, or <code>\begin{document}</code>).</li> <li>• Format all mathematical equations and symbols strictly in LaTeX.</li> </ul>	<ol style="list-style-type: none"> <li>1. <b>Problem:</b> Clearly define the quantum physics problem using LaTeX for mathematical precision. Provide all equations or descriptions necessary for solving the problem. For instance:           <math display="block">H = -\sigma_1^z \sigma_2^z - 0.5(\sigma_1^x + \sigma_2^x)</math> </li> <li>2. <b>Domain:</b> Include a concise two-word domain description in CAPS such as ISING HAMILTONIAN.</li> <li>3. Ensure the problem strictly adheres to one of the following predefined domain types:           <ul style="list-style-type: none"> <li>• Synthetic Hamiltonians</li> <li>• Quantum Spin Chains</li> <li>• Yang-Baxter Solvability</li> <li>• Quantum Circuit Design</li> </ul> </li> </ol>

■ **Figure 6:** Step-by-step instruction for generating a detailed quantum physics problem in LaTeX format. The framework dynamically selects from over 90 quantum physics domains (see Sec. A), leveraging LLMs to autonomously generate sub-domains, thus expanding to thousands of unique quantum problem categories.

### 4.3 STAGE 2: SOLUTION GENERATION

The output of Stage 1 is provided to a second Qwen2.5-Math-\*B-Instruct LLM tasked with generating step-by-step solutions such as the one depicted in Figure 8. The following prompt (Fig. 7) guides this stage:

**Example: Ising Model Hamiltonian** Here (Fig. 8) we present a practical example from the dataset, illustrating a simplified Ising model for two qubits with interaction strength  $J = 1$  and transverse field  $h = 0.5$ . This solution was generated as a response to the solution-prompt depicted in (Fig. 7).

**Stage 3: Problem-Solution Pair Dataset Enrichment** In this stage, we enrich the dataset by leveraging advanced reasoning methodologies (Fig. 9), (Fig. 10), such as CoT [WW<sup>+22</sup>], ToRA, and CoP [WW<sup>+22</sup>]. These techniques, traditionally applied in fields like combinatorics and optimization, are extended into quantum computing to address tasks such as Hamiltonian decomposition, QASM code generation, and eigenvalue computations. Existing datasets, including CoT and ToRA, offer problems with clear, unambiguous solutions, which serve as templates for analogous quantum computing challenges.

The enriched dataset bridges the gap between raw task definitions and structured problem-solution pairs. By utilizing techniques such as CoT, ToRA, and CoP, we enhance both the diversity and applicability of the dataset, ensuring its relevance for advanced quantum computing tasks.

**Stage 4: Verification with Advanced LLMs** In this final stage, we prioritize the verification of LLM-generated solutions, aiming to ensure their accuracy and reliability. While we initially explored the possibility of incorporating advanced reasoning techniques like self-correction and self-reflection, we opted to rely on a dedicated

Illustrative Example: Solution Prompt for Quantum Physics Problems	
Solution Instruction Prompt	Output Format and Example
<p>You are an expert in quantum physics and must provide detailed solutions to quantum physics problems.</p> <ul style="list-style-type: none"> <li>• All solutions must be in plain text format.</li> <li>• Use LaTeX to represent all mathematical equations and symbols.</li> <li>• Ensure clarity, accuracy, and completeness in each step of the solution.</li> </ul>	<p>1. <b>Solution Prompt:</b> Provide a complete solution to the following quantum physics problem in plain text format:</p> <p style="text-align: center;"><i>problem</i></p> <p>2. <b>Expected Response:</b></p> <ul style="list-style-type: none"> <li>• Step-by-step derivation with appropriate mathematical rigor.</li> <li>• Use LaTeX for all equations and symbols, for example:</li> </ul> $H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$ <ul style="list-style-type: none"> <li>• A final summary or conclusion, if applicable.</li> </ul>

Figure 7: Solution prompt for quantum physics problems, requiring step-by-step solutions with all mathematical symbols and equations in LaTeX format. Emphasizes clarity, accuracy, and completeness in each solution.

Illustrative Example: Solving a Sample Hamiltonian	
Problem	Step-by-Step Solution
Analytically solve the Hamiltonian	<p>1. Express the Hamiltonian in matrix form:</p> $H = -\sigma_1^z \sigma_2^z - 0.5(\sigma_1^x + \sigma_2^x)$ $H = \begin{pmatrix} -1 & -0.5 & -0.5 & 0 \\ -0.5 & 1 & 0 & -0.5 \\ -0.5 & 0 & 1 & -0.5 \\ 0 & -0.5 & -0.5 & -1 \end{pmatrix}$ <p>2. Set up the eigenvalue equation <math>\det(H - \lambda I) = 0</math>:</p> $\det \begin{pmatrix} -1 - \lambda & -0.5 & -0.5 & 0 \\ -0.5 & 1 - \lambda & 0 & -0.5 \\ -0.5 & 0 & 1 - \lambda & -0.5 \\ 0 & -0.5 & -0.5 & -1 - \lambda \end{pmatrix} = 0$ <p>3. Solve the characteristic equation:</p> $\lambda^4 + \lambda^3 - 3\lambda^2 - \lambda + 1 = 0$ <p>4. Eigenvalues (approximate):</p> $\lambda_1 = -1.618, \quad \lambda_2 = 0, \quad \lambda_3 = 0, \quad \lambda_4 = 1.618$

Figure 8: Step-by-step solution for solving the Hamiltonian  $H = -\sigma_1^z \sigma_2^z - 0.5(\sigma_1^x + \sigma_2^x)$  analytically, including eigenvalues and matrix representation.

verifier model for this task. As illustrated in (Fig. 11), we employ a sophisticated LLM, such as Qwen2.5-Coder-72B, functioning as a pseudo-oracle to evaluate the correctness of the generated solutions. This approach offers a robust mechanism for marking responses as either correct or incorrect without delving into iterative correction cycles.

Although self-correction methods have demonstrated potential in enhancing the stylistic and qualitative aspects of LLM outputs, their application to reasoning tasks has notable limitations. Prior studies [LCC<sup>+</sup>23, SCG<sup>+</sup>24, MTG<sup>+</sup>24, SYW<sup>+</sup>22, MTR23, CGHH23] underscore the concept of self-correction, where LLMs iteratively refine their outputs. However, as highlighted in [HCM<sup>+</sup>23], LLMs often struggle to identify and resolve reasoning errors autonomously. Models like Reflexion [SCG<sup>+</sup>24] and RCI [KBM24] depend on external feedback, such as ground-truth correctness signals, to terminate the self-correction loop effectively. Furthermore, attempts to self-correct reasoning errors can inadvertently transform accurate answers into incorrect ones, degrading overall performance [HCM<sup>+</sup>23].

### Sample Prompt: Physics Problem-Solution Verification Using ToRA and CoT Frameworks

*Prompt.* You are a quantum physics solving assistant. You will be presented with a physics problem and a proposed solution. Your task is to determine if the provided solution is correct. If the solution is correct, confirm it with a detailed explanation. If the solution is incorrect, identify the errors and provide a full corrected solution using the ToRA framework: Think, understand and restate the problem, identifying key details; Organize, break it into manageable steps; Reason, solve systematically, showing all intermediate steps and calculations; and Answer, present the final solution clearly and concisely. Always use Chain of Thought reasoning to ensure logical progression and clarity.

**Example:**

*Problem:* A train travels 120 miles in 2 hours. What is its average speed?

*Proposed Solution:* The average speed is 70 mph.

**Analysis:** The solution is incorrect. Think: Speed is distance divided by time. The distance is 120 miles, and the time is 2 hours. Organize: Write the formula Speed = Distance / Time. Reason: Substitute the values, Speed = 120 / 2 = 60 mph. Answer: The correct average speed is 60 mph.

**Now analyze the following problem and solution:**

*Problem:* [Insert Problem Here]

*Proposed Solution:* [Insert Proposed Solution Here]

Figure 9: Stage 3 Sample prompt - Enrichment using ToRA and CoT frameworks for quantum physics problem verification.

### Sample Prompt: Chain-of-Programming (CoP) Applied to Quantum Computing Tasks

*Prompt.* You are a quantum computing assistant. You will be presented with a programming task and a proposed solution. Your task is to determine if the provided solution is correct. If the solution is correct, confirm it with a detailed explanation. If the solution is incorrect, identify the errors and provide a corrected solution. Use the Chain-of-Programming (CoP) methodology to ensure logical progression and clarity, breaking the task into manageable steps and solving each incrementally.

**Example:**

*Problem:* Define a simple Hamiltonian  $H = \sigma_x + \sigma_z$ , where  $\sigma_x$  and  $\sigma_z$  are Pauli matrices. Find the eigenvalues and eigenvectors of  $H$ .

*Proposed Solution:* The eigenvalues of  $H$  are 2 and  $-2$ .

**Analysis:** The solution is incorrect. Using CoP methodology:

**Step 1:** Define the Hamiltonian  $H = \sigma_x + \sigma_z$ .

**Step 2:** Represent the Pauli matrices:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

**Step 3:** Add these matrices to form  $H$ :

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

**Step 4:** Compute the eigenvalues and eigenvectors of  $H$  by solving  $\det(H - \lambda I) = 0$ . The eigenvalues are  $\sqrt{2}$  and  $-\sqrt{2}$ .

**Answer:** The correct eigenvalues are  $\sqrt{2}$  and  $-\sqrt{2}$ .

Below is the Python implementation of the task:

```

1 import numpy as np
2 sigma_x = np.array([[0, 1], [1, 0]])
3 sigma_z = np.array([[1, 0], [0, -1]])
4 H = sigma_x + sigma_z
5 eigenvalues, eigenvectors = np.linalg.eigh(H)
6 print("Eigenvalues: ", eigenvalues)
7 print("Eigenvectors: ", eigenvectors)
8

```

**Now analyze the following problem and solution:**

*Problem:* [Insert Problem Here]

*Proposed Solution:* [Insert Proposed Solution Here]

Figure 10: Stage 3 Sample prompt - Enrichment using Chain-of-Programming (CoP) methodology for quantum computing problem verification.

**Sample Problem: Validation of Instruction-Response Pairs in Quantum Computing**

*Prompt.* You are an Oracle Language Model specialized in quantum computing tasks. Analyze the provided instruction-response pair and determine whether the response is correct. If the solution aligns with the instruction and is mathematically/logically valid, mark it as "Correct." Otherwise, provide a detailed justification for marking it "Incorrect." Ensure your evaluation is both precise and concise.

**Domain:** VALIDATION OF INSTRUCTION-RESPONSE PAIRS

Figure 11: Stage 4: An Oracle LLM assesses instruction-response pairs for correctness, marking a binary column in the dataset.

Given these limitations, our method strategically leverages large-scale LLMs to serve as explicit validators. These models are tasked with identifying errors in generated solutions and marking them as correct or incorrect without attempting corrective iterations. This approach ensures a high degree of reliability in solution validation while avoiding the pitfalls associated with self-correction mechanisms in complex reasoning tasks.

## 5 CONCLUSION

This study centers on the exclusive **generation of synthetic data tailored for quantum computing tasks**, deliberately omitting any training or fine-tuning of large language models. No fine-tuned or trained model is provided, as our primary objective is to develop a high-quality dataset that can serve as a foundational resource for both academic and industrial research. By focusing on synthetic data generation, we aim to alleviate the significant computational demands associated with training or fine-tuning, enabling broader accessibility to advanced quantum AI research. In future work, we plan to explore the potential of training a Qwen2.5 model leveraging this dataset, paving the way for further advancements in the field.

## REFERENCES

- [A<sup>+</sup>19] Mohammad Ali Amini et al. Mathqa: Towards understanding and solving math word problems with deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [ALC<sup>+</sup>23] Joshua Achiam, Christina Lee, Hongseok Choi, et al. Gpt-4: Architecture and performance. *arXiv preprint arXiv:2303.12708*, 2023.
- [BIS<sup>+</sup>18] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Carsten Blank, Keri McKiernan, and Nathan Killoran. PennyLane: Automatic differentiation of hybrid quantum-classical computations. In *2018 IEEE International Conference on Rebooting Computing (ICRC)*, pages 1–8. IEEE, 2018.
- [BMR<sup>+</sup>20] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [C<sup>+</sup>21] Karl Cobbe et al. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*, 2021.
- [C<sup>+</sup>23] Yuan Chen et al. Program-of-thought: Generalized and efficient reasoning with large language models. *arXiv preprint arXiv:2302.07439*, 2023.
- [CBG17] Andrew W Cross, Lev S Bishop, John A Smolin, and Jay M Gambetta. Open quantum assembly language. *arXiv preprint arXiv:1707.03429*, 2017.
- [CGHH23] Pinzhen Chen, Zhicheng Guo, Barry Haddow, and Kenneth Heafield. Iterative translation refinement with large language models. *arXiv preprint arXiv:2306.03856*, 2023.
- [CMWC22] Wenhui Chen, Xueguang Ma, Xinyi Wang, and William W Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *arXiv preprint arXiv:2211.12588*, 2022.
- [DQL<sup>+</sup>22] Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. Glm: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335, 2022.

- [DZC<sup>+</sup>23] Wei Ding, Li Zhao, Xiaotian Chen, Yu Wang, Jing Liu, and Xiaoyang Zhu. Physqa: Benchmarking physics question answering with large language models. *arXiv preprint arXiv:2309.08182*, 2023.
- [G<sup>+</sup>23a] Shijie Gao et al. Program-of-thought prompting for efficient large language model reasoning. *arXiv preprint arXiv:2302.11106*, 2023.
- [G<sup>+</sup>23b] Ting Gou et al. Tora: Large-scale fine-tuning of transformers for symbolic math. In *Proceedings of the 2023 Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- [GBCB16] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.
- [GSY<sup>+</sup>24] Bofei Gao, Feifan Song, Zhe Yang, et al. Omni-math: A universal olympiad-level mathematical benchmark for large language models. *arXiv preprint arXiv:2410.07985*, 2024.
- [GTA<sup>+</sup>23] L Gao, J Tow, B Abbasi, S Biderman, S Black, A DiPofi, C Foster, L Golding, J Hsu, A Le Noac'h, et al. A framework for few-shot language model evaluation, 12 2023. URL <https://zenodo.org/records/10256836>, 7, 2023.
- [H<sup>+</sup>21] Dan Hendrycks et al. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03824*, 2021.
- [HCM<sup>+</sup>23] Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet. *arXiv preprint arXiv:2310.01798*, 2023.
- [Hei28] Werner Heisenberg. Zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 49(9-10):619–636, 1928.
- [HZ<sup>+</sup>20] Dan Hendrycks, Kevin Zhao, et al. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*, 2020.
- [Isi25] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.
- [KBM24] Geunwoo Kim, Pierre Baldi, and Stephen McAleer. Language models can solve computer tasks. *Advances in Neural Information Processing Systems*, 36, 2024.
- [L<sup>+</sup>22] Anna Lewkowycz et al. Solving math word problems with minerva. *arXiv preprint arXiv:2203.01057*, 2022.
- [LCC<sup>+</sup>23] Ming Li, Lichang Chen, Juhai Chen, Shuai He, and Tianyi Zhou. Reflection-tuning: Recycling data for better instruction-tuning. In *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*, 2023.
- [Llo96] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.
- [LLZ<sup>+</sup>22] Yifei Li, Zeqi Lin, Shizhuo Zhang, Qiang Fu, Bei Chen, Jian-Guang Lou, and Weizhu Chen. Making large language models better reasoners with step-aware verifier. *arXiv preprint arXiv:2206.02336*, 2022.
- [LSX<sup>+</sup>23] Haipeng Luo, Qingfeng Sun, Can Xu, Pu Zhao, Jianguang Lou, Chongyang Tao, Xiubo Geng, Qingwei Lin, Shifeng Chen, and Dongmei Zhang. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct. *arXiv preprint arXiv:2308.09583*, 2023.
- [LXZ<sup>+</sup>23] Ziyang Luo, Can Xu, Pu Zhao, Qingfeng Sun, Xiubo Geng, Wenxiang Hu, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. Wizardcoder: Empowering code large language models with evol-instruct. *arXiv preprint arXiv:2306.08568*, 2023.
- [MFL<sup>+</sup>23] Swaroop Mishra, Matthew Finlayson, Pan Lu, Leonard Tang, Sean Welleck, Chitta Baral, Tanmay Rajpurohit, Oyvind Tafjord, Ashish Sabharwal, Peter Clark, and Ashwin Kalyan. Lila: A unified benchmark for mathematical reasoning. *Journal of AI Research*, 15(3):1–10, 2023. URL: <https://arxiv.org/abs/2210.17517>.
- [MTG<sup>+</sup>24] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36, 2024.
- [MTR23] Ning Miao, Yee Why Teh, and Tom Rainforth. Selfcheck: Using llms to zero-shot check their own step-by-step reasoning. *arXiv preprint arXiv:2308.00436*, 2023.

- [N<sup>+</sup>22] Maxwell Nye et al. Show your work: The importance of intermediate steps in language models. *arXiv preprint arXiv:2201.08686*, 2022.
- [PKC<sup>+</sup>23] Yiming Peng, Amit Kumar, Qingfeng Chen, Xander Lee, et al. Gpt4-alpaca: Instruction following with gpt-4. *arXiv preprint arXiv:2303.08793*, 2023.
- [PMS<sup>+</sup>14] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, 2014.
- [PMT<sup>+</sup>24] Haining Pan, Nayantara Mudur, Will Taranto, Maria Tikhonovskaya, Subhashini Venugopalan, Yasaman Bahri, Michael P. Brenner, and Eun-Ah Kim. Quantum many-body physics calculations with large language models. *arXiv preprint arXiv:2403.03154*, 2024.
- [Pre18] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.
- [R<sup>+</sup>23] Maxim Rein et al. Gpqa: A benchmark for general physics question answering with large language models. *arXiv preprint arXiv:2306.07378*, 2023.
- [S<sup>+</sup>23] Yi Sun et al. Scieval: A benchmark for evaluating scientific language models on diverse tasks. *arXiv preprint arXiv:2304.11828*, 2023.
- [S<sup>+</sup>24] Zhi Shao et al. Deepseek-math: Optimizing llms for deep mathematical understanding. *arXiv preprint arXiv:2401.01689*, 2024.
- [SCG<sup>+</sup>24] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.
- [SDea21] John Smith, Jane Doe, and et al. Qdataset: Quantum datasets for machine learning. *arXiv preprint arXiv:2108.06661*, 2021.
- [SDSC21] Victor Sanh, Jacob Devlin, Nicholas Shinn, and Hyun Chung. Multitask prompted training enables zero-shot task generalization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2021.
- [SMY<sup>+</sup>23] Weijia Shi, Sewon Min, Michihiro Yasunaga, Minjoon Seo, Rich James, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. Replug: Retrieval-augmented black-box language models. *arXiv preprint arXiv:2301.12652*, 2023.
- [Suz76] Masuo Suzuki. Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Communications in Mathematical Physics*, 51(2):183–190, 1976.
- [SYW<sup>+</sup>22] William Saunders, Catherine Yeh, Jeff Wu, Steven Bills, Long Ouyang, Jonathan Ward, and Jan Leike. Self-critiquing models for assisting human evaluators. *arXiv preprint arXiv:2206.05802*, 2022.
- [T<sup>+</sup>22] Andrew Taylor et al. Galactica: A large language model for science and math. *arXiv preprint arXiv:2209.10676*, 2022.
- [T<sup>+</sup>24] Dinesh Toshniwal et al. Openmathinstruct: A large-scale math dataset for instruction tuning with gpt-4. In *Proceedings of the 2024 International Conference on Machine Learning (ICML)*, 2024.
- [Tea24a] Charlie Taylor and et al. Qcircuitnet: A large-scale hierarchical dataset for quantum algorithm design. *arXiv preprint arXiv:2410.07961*, 2024.
- [Tea24b] Qwen Team. Qwen-coder: Advanced language models for code and mathematics. <https://github.com/qwen-team/qwen-coder>, 2024. Accessed: December 28, 2024.
- [TLI<sup>+</sup>23] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambrø, Faysal Azhar, Aurelien Rodríguez, Armand Joulin, Edouard Grave, David Barreira, and Hervé Jegou. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [TMN<sup>+</sup>24] Shubham Toshniwal, Ivan Moshkov, Sean Narenthiran, et al. Openmathinstruct-1: A 1.8 million math instruction tuning dataset. *arXiv preprint arXiv:2402.10176*, 2024.

- [TMS<sup>+</sup>23] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- [Tro59] Hale F Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.
- [W<sup>+</sup>23] Yifan Wu et al. Tora: Task-oriented reasoning and action framework for llms. *arXiv preprint arXiv:2405.07551*, 2023.
- [WBD<sup>+</sup>21] Jason Wei, Mikhail Bosma, David Dohan, Reza Salama, Kevin Cho, Jack Chen, Marcin Andrychowicz, Alex Copeland, Avinash Tapas, Julian Schroeder, et al. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021.
- [WLD<sup>+</sup>23] Xueguang Wang, Xiang Luo, Xu Ding, Wei Zhang, et al. Self-instruct: Aligning self-instruction with human preferences. *arXiv preprint arXiv:2302.08447*, 2023.
- [WRZ<sup>+</sup>23] Ke Wang, Houxing Ren, Aojun Zhou, Zimu Lu, Sichun Luo, Weikang Shi, Renrui Zhang, Linqi Song, Mingjie Zhan, and Hongsheng Li. Mathcodeinstruct: Combining code and math for llms. *Transactions on Machine Learning*, 8(4):234–256, 2023. URL: <https://arxiv.org/abs/2310.03731>.
- [WTB<sup>+</sup>22] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022.
- [WW<sup>+</sup>22] Jason Wei, Xueguang Wang, et al. Chain-of-thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*, 2022.
- [WWS<sup>22a</sup>] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain-of-thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.
- [WWS<sup>22b</sup>] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.
- [WXL<sup>+</sup>22] Lian Wang, Chen Xie, Xing Li, Yuan Zhang, Chenguang Liu, Xiaodong Ren, et al. Superni: Supervised natural instructions for few-shot learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.
- [WZea23] Alice Wang, Bob Zhang, and et al. Mnisq: A large-scale quantum circuit dataset for machine learning on/for quantum computers in the nisq era. *arXiv preprint arXiv:2306.16627*, 2023.
- [XSZ<sup>+</sup>23] Zhiwei Xu, Zhi Shao, Jiang Zhou, Kuan Zhang, et al. Wizardlm: Empowering llms with long-range knowledge for instruction tuning. *arXiv preprint arXiv:2301.10732*, 2023.
- [Y<sup>+</sup>23a] Tian Yu et al. Metamath: A large-scale dataset for math reasoning with gpt-3.5. *arXiv preprint arXiv:2305.12096*, 2023.
- [Y<sup>+</sup>23b] Lixia Yue et al. Mammoth: Instruction tuning for math and science reasoning. *arXiv preprint arXiv:2307.08188*, 2023.
- [Y<sup>+</sup>24] Chao Ying et al. Intern-math: A next-generation llm for math reasoning. *arXiv preprint arXiv:2402.03917*, 2024.
- [YYZ<sup>+</sup>24] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 2024.
- [YZH<sup>+</sup>24] An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jian-hong Tu, Jingren Zhou, Junyang Lin, et al. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*, 2024.
- [ZPR<sup>+</sup>24] Pei Zhou, Jay Pujara, Xiang Ren, Xinyun Chen, Heng-Tze Cheng, Quoc V Le, Ed H Chi, Denny Zhou, Swaroop Mishra, and Huaixiu Steven Zheng. Self-discover: Large language models self-compose reasoning structures. *arXiv preprint arXiv:2402.03620*, 2024.
- [ZWL<sup>+</sup>23] Aojun Zhou, Ke Wang, Zimu Lu, Weikang Shi, Sichun Luo, Zipeng Qin, Shaoqing Lu, Anya Jia, Linqi Song, Mingjie Zhan, and Hongsheng Li. Tora: A dataset for task-oriented reasoning and action. *Proceedings of the Neural Information Processing Systems Conference*, 36:1345–1358, 2023. URL: <https://arxiv.org/abs/2310.03732>.

## A COMPREHENSIVE DATASET DOMAINS

Our dataset is enriched with a very extensive set of domains in all areas of quantum physics and quantum computing, including:

1. **QASM Generation:** Generate quantum circuits and produce corresponding QASM code.
2. **Quantum Hamiltonians:** Analyze Hamiltonian time evolution and ground-state energy calculations.
3. **Yang-Baxter Solvability:** Determine solvability of quantum models using the Yang-Baxter equation.
4. **Trotter-Suzuki Decomposition:** Simulate Hamiltonians using Trotter-Suzuki decomposition methods.
5. **Lindblad Dynamics:** Model open quantum systems using the Lindblad equation.
6. **Randomized Circuits Optimization:** Optimize randomized quantum circuits to minimize error rates.
7. **Quantum Phase Estimation:** Implement quantum phase estimation for eigenvalue calculations.
8. **Cluster States Verification:** Prepare and verify cluster states for measurement-based quantum computation.
9. **VQE Analysis:** Construct and optimize Variational Quantum Eigensolvers (VQE) for molecular Hamiltonians.
10. **Quantum Algorithm Development:** Develop quantum algorithms for problems such as integer factorization and database search.
11. **Entanglement and Quantum Information Theory:** Explore properties and applications of entangled states.
12. **Quantum Error Correction:** Design quantum error correction codes to protect qubits from decoherence.
13. **Semiclassical Quantum Simulation:** Simulate quantum systems with semiclassical methods.
14. **Quantum Communication Protocols:** Develop protocols such as Quantum Key Distribution (QKD) and superdense coding.
15. **Topological Quantum Computing:** Study fault-tolerant computing using braiding operations of anyons.
16. **Quantum Complexity Classes:** Investigate computational problem classifications using quantum algorithms.
17. **Quantum Thermodynamics:** Analyze thermodynamic properties of quantum systems.
18. **Interacting Quantum Systems:** Study dynamics and correlations in interacting quantum systems.
19. **Quantum Cryptography:** Explore quantum cryptographic protocols like QKD.
20. **Quantum Channels:** Analyze the mathematical properties of quantum information transfer channels.
21. **Quantum Fourier Transform:** Explore the implementation and applications of quantum Fourier transforms.
22. **Quantum Machine Learning:** Apply quantum circuits to machine learning tasks such as classification.
23. **Quantum State Tomography:** Reconstruct quantum states using measurement data.
24. **Bell Inequalities and Nonlocality:** Test Bell inequalities and study quantum nonlocality.
25. **Diagonalization of Two-Spin Hamiltonians:** Solve eigenvalue problems for two-spin systems.
26. **Energy Eigenvalues via Perturbation Theory:** Compute eigenvalues using perturbation theory and diagonalization.
27. **Measurement in Plus-Minus Basis:** Analyze measurement probabilities in non-standard bases.
28. **Pauli Spin Matrices Analysis:** Explore properties and applications of Pauli matrices.
29. **Born's Rule and State Measurement:** Apply Born's rule to calculate quantum measurement probabilities.
30. **PennyLane Quantum Circuits:** Implement quantum circuits using the PennyLane framework.
31. **PennyLane Circuit Analysis:** Analyze PennyLane quantum circuits for functionality and structure.
32. **Building Molecular Hamiltonians:** Construct molecular Hamiltonians for quantum chemistry.
33. **Variational Quantum Eigensolver (VQE):** Optimize VQE methods for specific Hamiltonians.
34. **Subspace Search-Quantum Variational Quantum Eigensolver (SSVQE):** Find multiple eigenstates using SSVQE.
35. **Variational Quantum State Diagonalization (VQSD):** Diagonalize density matrices with VQSD techniques.

36. **Gibbs State Preparation:** Prepare Gibbs states for specific Hamiltonians.
37. **The Classical Shadow of Unknown Quantum States:** Approximate quantum state properties using classical shadows.
38. **Estimation of Quantum State Properties Based on Classical Shadows:** Estimate quantum state properties using shadow protocols.
39. **Hamiltonian Simulation with Product Formula:** Simulate Hamiltonians using Trotter product formulas.
40. **Simulate the Spin Dynamics on a Heisenberg Chain:** Model spin dynamics on a Heisenberg chain.
41. **Distributed Variational Quantum Eigensolver Based on Schmidt Decomposition:** Apply distributed VQE algorithms.
42. **Quantum Signal Processing and Quantum Singular Value Transformation:** Use quantum signal processing for operator transformations.
43. **Hamiltonian Simulation with qDRIFT:** Simulate Hamiltonians using the qDRIFT method.
44. **Quantum Phase Processing:** Apply quantum phase processing for signal amplification.
45. **Variational Quantum Metrology:** Optimize quantum sensing with variational quantum metrology.
46. **Encoding Classical Data into Quantum States:** Encode classical data into quantum states.
47. **Quantum Classifier:** Implement quantum classifiers for machine learning.
48. **Variational Shadow Quantum Learning (VSQL):** Approximate quantum properties with VSQL.
49. **Quantum Kernel Methods:** Develop kernel functions for quantum machine learning.
50. **Quantum Autoencoder:** Compress quantum data using quantum autoencoders.
51. **Quantum GAN:** Generate quantum states using quantum GANs.
52. **Variational Quantum Singular Value Decomposition (VQSVD):** Approximate singular value decompositions with VQSVD.
53. **Data Encoding Analysis:** Analyze data encoding methods for quantum computation.
54. **Quantum Neural Network Approximating Functions:** Approximate functions using quantum neural networks.
55. **Variational Quantum Amplitude Estimation:** Estimate quantum amplitudes variationally.
56. **Quantum Approximation Optimization Algorithm (QAOA):** Solve optimization problems with QAOA.
57. **Solving Max-Cut Problem with QAOA:** Apply QAOA to solve Max-Cut problems.
58. **Large-Scale QAOA via Divide-and-Conquer:** Scale QAOA for larger problem instances using divide-and-conquer techniques.
59. **Travelling Salesman Problem:** Solve the Travelling Salesman Problem using quantum algorithms.
60. **Jordan-Wigner Transformations:** Map spin models to fermionic systems using Jordan-Wigner transformations.
61. **Bethe Ansatz Application:** Solve the Heisenberg spin chain spectrum using the Bethe Ansatz.
62. **Generalized Spin Chain Compression:** Compress quantum circuits using Yang-Baxter equations in spin chain models.
63. **Wave-Particle Duality:** Explore the dual nature of particles and waves through phenomena like the photoelectric effect.
64. **Uncertainty Principle:** Analyze implications of Heisenberg's uncertainty principle.
65. **Perturbation Theory:** Examine corrections to energy levels using perturbation theory.
66. **Angular Momentum:** Investigate eigenstates and addition rules for angular momentum.
67. **Hydrogen Atom:** Study quantization of energy levels and transitions in the hydrogen atom.
68. **Scattering Theory:** Analyze quantum scattering and phase shifts using the Born approximation.
69. **Quantum Tunneling:** Investigate tunneling phenomena using the WKB approximation.
70. **Entanglement:** Examine quantum entanglement and its applications.
71. **Time Evolution:** Analyze quantum systems' time evolution using the Schrödinger equation.
72. **Quantum Measurement:** Explore wavefunction collapse and quantum measurement theory.
73. **Quantum Harmonic Oscillator:** Study energy eigenvalues and wavefunctions of the quantum harmonic oscillator.

74. **Spin-Orbit Coupling:** Examine spin-orbit interaction in atomic systems.
75. **Quantum Zeno Effect:** Explore repeated measurements' effects on quantum systems.
76. **Quantum Gates:** Construct and analyze quantum circuits with gates like Hadamard and CNOT.
77. **Adiabatic Theorem:** Examine systems' behavior under slowly varying potentials.
78. **Bell Inequalities:** Test Bell inequalities and analyze their implications.
79. **Superposition Principle:** Investigate the principle of superposition in quantum mechanics.
80. **Quantum Decoherence:** Analyze the loss of coherence in quantum systems.
81. **Topological Quantum States:** Explore the robustness of topological quantum states.
82. **Quantum Cryptography:** Study principles and protocols of quantum cryptography.
83. **Quantum Eraser:** Examine implications of the quantum eraser experiment.
84. **Quantum Teleportation:** Demonstrate quantum teleportation principles.
85. **Path Integral Formulation:** Utilize the path integral approach in quantum mechanics.
86. **Quantum Annealing:** Analyze quantum annealing for optimization problems.
87. **Berry Phase:** Study geometric phase in quantum systems.
88. **Quantum Cloning:** Explore the no-cloning theorem in quantum mechanics.
89. **Density Matrix Formalism:** Describe mixed states using density matrices.
90. **Quantum Computation:** Explore basic concepts of quantum computation.
91. **Relativistic Quantum Mechanics:** Solve problems involving the Klein-Gordon or Dirac equations.
92. **Quantum Field Theory:** Introduce concepts of quantum field theory.

This comprehensive set of domains offers a foundational resource for advancing quantum algorithms, problem-solving, machine learning, and optimization techniques in quantum physics and quantum computing.

## B PERFORMANCE EVALUATION

To assess the efficiency of the QuantumLLMInstruct pipeline, we conducted a performance evaluation on two different setups: a high-performance HuggingFace A100G machine and a local NVIDIA 4080-powered system. The evaluation measured the execution times for generating quantum problem instructions using the Qwen2.5-Coder-1.5B-Instruct model and its smaller variant, Qwen-7B Coder. The results provide insights into the computational demands and scalability of the pipeline across various quantum domains.

**Execution Times on HuggingFace A100G Machine** The execution times recorded on the HuggingFace A100G machine showcase the model's efficiency in generating problem instructions in real-time for fundamental quantum domains.

Domain	Problem ID	Time Taken (seconds/problem)
Berry Phase	1, 3	3.23, 2.78
Wave-Particle Duality	2	2.20
Quantum Classifier	4, 5	2.78, 2.64

Table 3: Performance Metrics for Generating Quantum Problem Instructions (Model: Qwen2.5-Coder-1.5B-Instruct on HuggingFace A100G Machine)

These results highlight the ability of the model to rapidly generate high-quality problem instructions, with each problem taking less than 4 seconds on average.

**Execution Times on a Local Machine (NVIDIA 4080)** The local system, powered by an NVIDIA 4080 GPU, demonstrated the pipeline's adaptability to more common hardware configurations. However, the execution times varied significantly depending on the complexity of the problem and the availability of required templates.

<b>Domain</b>	<b>Problem ID</b>	<b>Time Taken (seconds/problem)</b>
Quantum Probability Calculations	6, 11	230.14, 186.72
Measurement in Qubit Basis	7	Error ('template')
Estimation of Quantum State Properties	9	114.30
Bell Inequalities and Nonlocality	10	145.84
Entanglement and Quantum Information Theory	11	186.72
Quantum Error Correction	12	Error ('template')
PennyLane Quantum Circuits	13	154.16

Table 4: Time Taken to Generate Problems for Each Quantum Domain (Model: Qwen-7B Coder on NVIDIA 4080)