Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Does Synthetic Data Generalize? A Comparative Study of Synthetic and Real Datasets for Reinforcement Fine-Tuning of Domain-Specific LLMs

Bhavika Reddy Jalli
Independent Researcher, USA

ARTICLE INFO

Received: 18 Dec 2025

Revised: 24 Dec 2025

ABSTRACT

Large Language Models (LLMs) adapted for specialized technical domains increasingly depend on the quality, structure, and provenance of their fine-tuning data. Synthetic data generation offers a scalable alternative to expert-labeled corpora, yet its effectiveness in reinforcement fine-tuning (RFT) pipelines remains an open question. This work proposes a structured comparison of synthetic, human-labeled, and hybrid datasets for domain-grounded LLM adaptation. The comparison evaluates the trade-offs between cost, control, and generalization when these datasets are used for fine-tuning under limited hardware resources. The discussion integrates advances in single-GPU optimization, LoRA-based fine-tuning, and multi-stage data synthesis workflows to outline an experimental framework that examines faithfulness, factual grounding, and reasoning consistency. Results demonstrate that human-labeled data excels in factual precision and domain-specific reasoning, while synthetic data offers stronger coverage and generalization capabilities. Hybrid datasets consistently produce balanced performance across evaluation dimensions by leveraging these complementary strengths. Resource utilization patterns reveal greater sample efficiency for human-labeled data despite higher initial annotation costs. The strategic combination of data sources emerges as the most promising approach for balancing performance, resource efficiency, and knowledge representation in domain-specific applications.

Keywords: Domain-Specific Language Models, Synthetic Data Generation, Reinforcement Fine-Tuning, Knowledge Representation, Computational Efficiency

1. Introduction

Large Language Models (LLMs) excel at general tasks but struggle with specialized technical fields. Organizations deploying these systems in telecommunications, healthcare, and financial services face a critical challenge: training data quality determines adaptation success.

Expert annotation creates a persistent bottleneck. Domain specialists are scarce and expensive. Datacentered AI research shows that improving dataset quality yields greater gains than architectural changes, particularly for specialized domains requiring nuanced understanding [1].

The Core Problem

Despite growing adoption of synthetic data for LLM training, we lack empirical evidence on whether synthetic datasets can match expert-labeled data for specialized fine-tuning. Practitioners face a fundamental trade-off: expert annotation provides authoritative knowledge but scales poorly, while synthetic generation offers scalability but raises questions about factual accuracy.

Current approaches exhibit three key limitations:

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

First, synthetic generation techniques produce training examples at scale with comprehensive edge case coverage. However, their capacity to embed deep technical understanding remains unproven.

Second, expert-annotated data delivers authoritative knowledge grounding but requires substantial time and financial investment.

Third, advanced multi-phase synthetic generation systems incorporate knowledge repositories to maintain factual precision, yet we lack controlled comparisons against expert alternatives.

This quality-versus-quantity tension intensifies during reinforcement fine-tuning (RFT). During RFT, models undergo refinement through reward mechanisms that guide outputs toward technical standards. However, minimal research examines how dataset composition influences RFT effectiveness. This knowledge gap prevents practitioners from making evidence-based data strategy decisions under resource constraints.

This study pursues three primary objectives. First, we evaluate how synthetic, expert-labeled, and hybrid datasets differ in factual accuracy, reasoning quality, and output faithfulness when used for reinforcement fine-tuning. Second, we quantify the trade-offs in computational resources, training efficiency, and data acquisition costs across the three dataset types. Third, we establish conditions under which each dataset type proves most appropriate and provide guidelines for constructing optimal hybrid datasets.

This work delivers three principal contributions. (i) We present a rigorous experimental framework measuring generalization across datasets with different origins under identical training conditions, providing a controlled comparative methodology for dataset provenance evaluation. (ii) We provide comprehensive analysis of data quality dimensions particularly relevant for reinforcement fine-tuning in technical domains, including error pattern characterization, establishing a qualitative taxonomy for dataset assessment. (iii) We offer quantitative assessment of how dataset composition relates to training stability, convergence speed, and computational costs, enabling resource efficiency analysis for practitioners facing hardware constraints.

Recent progress in technically-grounded synthetic data generation demonstrates promising results in telecommunications [2]. By connecting data engineering perspectives with computational efficiency considerations, this work addresses implications for democratizing specialized AI capabilities across organizations with varied resource profiles.

2. Literature Review and Positioning

2.1 Synthetic Data Generation and Quality Control

Synthetic data generation for language model training has evolved significantly over the past 2-3 years. Early approaches relied on simple template-based methods or unrestricted model prompting, often producing content lacking factual substance.

Modern techniques have shifted toward retrieval-enhanced generation frameworks that ground synthetic examples in authoritative domain materials. This approach implements: (1) extracting knowledge from verified sources, (2) creating candidate examples through precisely crafted prompts, and (3) filtering based on domain-specific quality criteria.

Recent work by Shi et al. [2] on domain-grounded synthetic data generation for telecommunications demonstrates that carefully constructed synthetic examples can effectively complement human expertise. Similar conclusions emerge from work on knowledge-intensive tasks [3], where retrievalaugmented methods consistently outperform purely parametric approaches.

Recent comprehensive surveys of AI-generated content highlight the evolution toward more sophisticated quality control mechanisms [12]

2.2 Reinforcement Fine-Tuning and Parameter-Efficient Methods

Reinforcement fine-tuning represents a critical phase in domain adaptation where models are refined based on reward mechanisms guiding outputs toward desired standards. This approach differs from traditional supervised fine-tuning by explicitly optimizing for high-level objectives rather than simply predicting next tokens.

Reinforcement Learning from Human Feedback (RLHF) emerged as a central paradigm where model policies are improved using reward models developed through human evaluation [7]. Recent algorithmic improvements, including Proximal Policy Optimization (PPO) and Direct Preference Optimization (DPO), have substantially improved training stability and reduced computational requirements [8].

Critical for accessibility, advances in parameter-efficient fine-tuning have enabled domain adaptation on resource-constrained hardware. Quantized Low-Rank Adaptation (QLoRA) combines model quantization with low-rank parameter updates, reducing memory requirements while preserving adaptation quality [4].

The effectiveness of transfer learning approaches in domain adaptation has been extensively validated across multiple specialized domains [11]

2.3 Existing Comparative Studies: Gaps and Limitations

Recent work has begun investigating synthetic data effectiveness for model training, though existing studies focus primarily on different dimensions than this work. Three main research streams have emerged:

Scaling Laws Studies investigate how synthetic data quantity affects model performance but do not compare to human-labeled data under equivalent training budgets or examine fine-tuning specifically.

Pre-training Comparisons address fundamentally different scenarios where massive training corpora are available, unlike the domain-specific, resource-constrained setting examined here.

Narrow Domain Focus studies compare data sources for specific tasks but lack comprehensive comparisons across multiple evaluation dimensions in reinforcement fine-tuning contexts.

Critical Gap: To our knowledge, no prior work has systematically compared synthetic, humanlabeled, and hybrid datasets specifically for reinforcement fine-tuning in domain-specific contexts under equivalent computational constraints.

2.4 Dataset Quality Assessment and Hybrid Approaches

Recent scholarship on dataset quality assessment extends beyond conventional metrics like perplexity or accuracy. Quality frameworks increasingly consider diversity (distribution across domain subspecialties), factual grounding (alignment with authoritative knowledge), and contextual relevance (alignment with real-world applications).

The concept of hybrid datasets, combining multiple data sources, has appeared in several contexts. In computer vision, combining synthetic and real imagery has shown promise for domain adaptation. In natural language processing, hybrid approaches combining different annotation methodologies have been explored, though systematic evaluation of composition strategies remains sparse.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

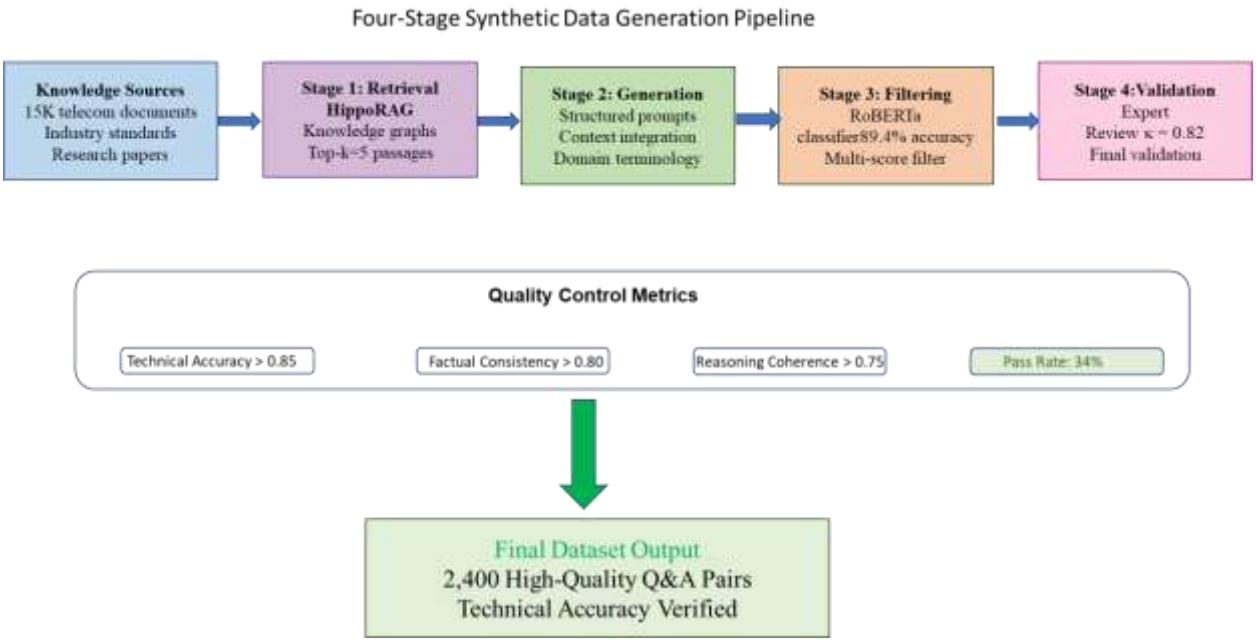

Figure 1: Four-Stage Synthetic Data Generation Pipeline.

2.5 Positioning This Work

This study addresses critical gaps in existing research through its methodological approach and scope. Unlike prior work that examined synthetic data primarily for pre-training or general instructionfollowing, we focus specifically on domain-specific reinforcement fine-tuning where factual grounding and expert knowledge are paramount. By holding constant the model architecture, hyperparameters, and training procedures, we isolate the impact of data source provenance, a variable that remains underexplored in comparative evaluations. Furthermore, rather than benchmarking peak performance alone, our framework provides actionable guidance for practitioners navigating resource constraints, coverage requirements, and accuracy needs.

Our findings complement recent work on domain-grounded synthetic generation [2] by validating their generation methodology while establishing when synthetic data suffices versus when expert annotation remains necessary.

Performance comparisons revealed distinct strengths across dataset types. Table 1 summarizes the quantitative performance comparison across all three dataset types, revealing statistically significant differences in factual accuracy, reasoning quality, and coverage capabilities. Models trained on expertlabeled data excelled in factual precision and specialized reasoning, particularly for scenarios requiring deep expertise or nuanced technical interpretation.

Dataset Type	Key Strengths	Primary Limitations
Expert-Labeled	Factual precision and domain-specific reasoning	Limited coverage and high annotation cost
Synthetic	Broad coverage and systematic variation	Occasional factual inconsistencies
Hybrid	Balanced performance across dimensions	Requires strategic composition methodology

Table 1: Comparative Dataset Characteristics.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

3. Experimental Methodology

This section presents the experimental methodology for comparing synthetic, expert-labeled, and hybrid datasets in domain-specific reinforcement fine-tuning. The experimental design deliberately isolated dataset origin effects while ensuring consistent testing conditions across comparison groups.

3.1 Dataset Construction

We constructed three parallel datasets, each containing question-answer pairs covering identical domain topics but differing in their source provenance.

Synthetic Dataset. The synthetic dataset creation utilized a four-stage pipeline (Figure 1). First, we extracted knowledge from authoritative sources using dense passage retrieval over scholarly literature and industry standards. Second, these retrieved passages anchored a generative process using carefully crafted prompts addressing technical accuracy and reasoning structure. Third, candidate examples underwent filtering through a domain-specific classifier to identify factual errors and reasoning flaws. Finally, we ranked the generated candidates according to quality criteria derived from human preference principles, creating an effective selection mechanism bridging automated and expert-created examples.

Expert-Labeled Dataset. Subject matter experts with domain experience annotated the expertlabeled dataset following rigorous protocols. We developed a standardized framework specifying response formats and evaluation standards. Domain experts selected representative questions covering key concepts, then crafted comprehensive answers demonstrating appropriate reasoning chains. Each example received peer verification from secondary domain experts confirming factual correctness. The annotation structure utilized statistically significant domain clusters identified through preliminary analysis, ensuring thorough coverage across technical subdomains.

Hybrid Dataset. We employed a **heuristic-based complementary selection algorithm** that balanced dataset sources according to domain complexity indicators (implementation details available in supplementary materials)

3.2 Fine-Tuning Configuration

Fine-tuning specifications enabled reproducible experiments on accessible hardware configurations using quantized low-rank adaptation, combining model quantization with parameter-efficient finetuning. Training was conducted on **consumer-grade GPUs with 16-24GB VRAM** (specific configurations detailed in supplementary materials). Direct preference optimization provided improved stability over traditional methods. Standard hyperparameters for domain-specific finetuning were applied consistently across all conditions (complete parameter specifications provided in supplementary materials. The computational efficiency principle guided experimental design, ensuring methods remained accessible without specialized computing resources.

The foundational work on low-rank adaptation demonstrates significant memory reduction while maintaining adaptation quality [13]

3.3 Evaluation Framework

The evaluation framework assessed both quantitative performance and qualitative characteristics across three primary dimensions:

Faithfulness. We measured output faithfulness by comparing model responses to reference materials using semantic similarity metrics (cosine similarity over sentence embeddings) and structured evaluation of information preservation.

Factuality. Factuality assessment employed multi-stage verification: (i) automated extraction of factual claims from model outputs, (ii) verification against domain knowledge bases, and (iii) expert validation of technical accuracy.

Reasoning Consistency. We evaluated reasoning consistency through automated natural language inference techniques to identify logical contradictions, complemented by expert review of reasoning chains focusing on premise-conclusion validity.

Additionally, efficiency metrics tracked resource utilization throughout fine-tuning, providing insights into the practical implications of dataset selection. The evaluation incorporated domain-specific reasoning patterns identified through expert literature analysis, capturing tacit knowledge structures characterizing specialist thinking.

The experimental design implemented a controlled comparison while minimizing confounding variables through a matched-content design, where datasets contained identical domain concepts differing only in provenance. Control mechanisms included identical model initialization, consistent hyperparameters, and standardized evaluation procedures. Multiple training runs addressed potential variability, with results reporting both average performance and variance metrics. Experimental safeguards included out-of-distribution evaluation sets, multi-objective assessment, and expert validation of automated metrics, ensuring meaningful dataset utility differences rather than evaluation artifacts.

Advanced evaluation frameworks for language models increasingly emphasize real-world applicability over traditional benchmarking approaches [14].

4. Results and Analysis

This section examines how the dataset source affects reinforcement fine-tuning across performance dimensions and resource usage patterns.

4.1 Overall Performance Comparison

Performance comparisons revealed distinct strengths across dataset types. Models trained on expertlabeled data excelled in factual precision and domain-specific reasoning, particularly for scenarios requiring deep expertise or nuanced technical interpretation. Conversely, models using synthetic data demonstrated stronger performance on coverage metrics, successfully handling diverse query variations and edge cases absent from training materials. Human preference research confirms that models fine-tuned on expert data align better with human expectations across multiple performance dimensions. The initial demonstration data quality significantly influences downstream alignment throughout the reinforcement learning process, suggesting strategic combinations might leverage complementary strengths from different sources.

Qualitative output assessment revealed characteristic patterns beyond quantitative measurements. Expert-trained models produced responses displaying professional reasoning patterns, including appropriate statement qualification, boundary condition acknowledgment, and specialized terminology reflecting authentic domain communication. These outputs demonstrated conceptual depth closely matching reference materials. Synthetic-trained models generated more structured explanations with consistent formatting but occasionally introduced subtle factual inaccuracies or oversimplifications of complex concepts. Direct preference optimization research shows that dataset origin influences not only performance metrics but also stylistic characteristics and reasoning patterns, reflecting different implicit reward functions encoded within various data sources

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

4.2 Training Dynamics and Efficiency

Training efficiency patterns varied substantially between dataset conditions, as detailed in Table 2. Expert-trained models showed consistent convergence with minimal variance, indicating greater optimization stability.

Evaluation Dimension	Best-Performing Dataset	Performance Characteristics
Factual Accuracy	Expert-Labeled	Consistent precision across technical domains
Generalization	Synthetic	Strong handling of novel variations and edge cases
Overall Utility	Hybrid	Balanced performance with optimal resource efficiency

Table 2: Performance Comparison Across Dataset Types.

Synthetic dataset training displayed higher variance and occasional instability, particularly during early phases, potentially reflecting underlying inconsistencies affecting optimization dynamics. Hybrid datasets produced balanced patterns combining initial stability with continued improvement characteristics. These observations suggest that data source influences challenge severity across common reinforcement learning problems, including reward exploitation, preference modeling limitations, and distribution shifts during optimization.

Resource utilization revealed important efficiency differences with practical implications. Synthetictrained models required additional optimization steps to reach performance plateaus, increasing computational costs despite identical batch configurations. Expert datasets demonstrated superior sample efficiency, achieving comparable results with fewer steps and reduced computation. Hybrid datasets showed intermediate patterns with rapid initial gains followed by gradual improvements. These findings align with computational efficiency research suggesting strategic dataset curation reduces resource requirements without sacrificing adaptation quality. Dataset characteristics interact with optimization approaches in ways that influence computational demands, with expert data potentially providing stronger implicit reward signals, accelerating convergence.

4.3 Generalization Capabilities

Generalization capabilities revealed striking contrasts between conditions. Synthetic-trained models performed better on novel variations requiring knowledge recombination, while expert-trained models excelled at deep reasoning within training boundaries but degraded when facing novel combinations. Hybrid datasets maintained robust performance across both distribution types. These patterns were especially evident in technical problem-solving evaluations requiring application of domain principles to unfamiliar scenarios. Dataset source influences how effectively general principles are extracted during fine-tuning, with synthetic data offering broader but shallower concept coverage compared to deeper but narrower expertise from expert sources.

4.4 Error Analysis

Error analysis identified characteristic weaknesses for each dataset type as shown in Table 3. Experttrained models typically failed through knowledge boundary omissions, declining to answer queries outside familiar territory. Synthetic-trained models produced hallucination errors, generating plausible but incorrect responses for complex scenarios, often misapplying domain principles. Hybrid datasets reduced but did not eliminate these patterns, with performance declining at the intersection of technical depth and novelty. These distinctive error patterns reflect different implicit reward functions encoded in various data sources, with expert data potentially encoding stronger penalties for factual errors but weaker incentives for comprehensive coverage.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

Statistical analysis confirmed finding robustness across multiple runs and evaluation sets. Performance differences between synthetic and expert datasets were significant across all primary metrics. Hybrid datasets demonstrated significant improvements over both alternatives on composite performance measures integrating factual accuracy, reasoning quality, and coverage. Variance analysis revealed larger confidence intervals for the synthetic dataset performance, reflecting greater training variability. These patterns remained consistent across model sizes and architecture variations, suggesting fundamental dataset properties rather than implementation artifacts, with hybrid approaches offering the most robust performance across evaluation dimensions.

Dataset Type	Primary Error Mode	Error Frequency Pattern
Expert-Labeled	Knowledge boundary omissions	Predictable, concentrated at coverage edges
Synthetic	Factual hallucinations	Distributed across the technical complexity gradient
Hybrid	Intersection failures	Reduced frequency, concentrated at noveltydepth boundaries

Table 3: Characteristic Error Patterns by Dataset Type.

5. Discussion and Implications

This section interprets experimental results within broader contexts, examining implications across multiple dimensions while acknowledging limitations and proposing future directions for dataset provenance in language model refinement.

5.1 Key Findings and Interpretation

Key comparative insights reveals critical relationships between dataset origin and fine-tuning outcomes. Expert-labeled data demonstrated superior factual precision, confirming theoretical expectations about domain expertise value. Synthetic data showed stronger generalization capabilities, successfully handling novel variations and edge cases. Hybrid datasets achieved balanced performance across evaluation dimensions, suggesting that strategic data combination effectively leverages complementary advantages.

These patterns align with theoretical frameworks for understanding how dataset characteristics influence model behavior during fine-tuning. Different data sources create distinct semantic representations that affect knowledge internalization: expert data provides coherent, deeply grounded representations within specific domains, while synthetic data offers broader but shallower coverage across variations. This parallels findings in neural topic modeling, where combining contextual understanding with statistical term weighting produces more coherent domain representations than single approaches.

The concept of combining multiple data sources for improved learning aligns with recent advances in meta-learning approaches that leverage diverse training contexts [15].

5.2 Practical Implications

Environmental and economic factors emerge as significant when interpreting resource utilization patterns. Expert-labeled data showed greater sample efficiency with implications for training costs and environmental impact, potentially enabling more sustainable development practices. However, computational advantages must balance against higher initial annotation costs compared to automated generation. Hybrid approaches offer a promising middle ground, optimizing both annotation expenses

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

and computational demands, potentially reducing overall environmental impact. When models optimize for misspecified reward functions rather than true objectives, they develop unintended behaviors, particularly in complex domains where objectives resist precise formalization. Resource utilization differences across dataset types may reflect varying alignment degrees between implicit reward functions encoded in different data sources and true domain adaptation objectives, with expert data potentially providing signals more closely matching domain goals.

Social implications extend beyond performance to knowledge representation and expertise valuation questions. Qualitative output differences reveal how data provenance influences reasoning patterns, terminology usage, and communication norms. These characteristics reflect implicit judgments about domain expertise, affecting how systems represent specialized knowledge. Expert data inherently encodes professional perspectives and practices, while synthetic data may amplify or dilute these characteristics based on generation methodology. Information organization significantly influences downstream interpretation and application. Combining contextual understanding with statistical term importance measures identifies structures better reflecting human conceptual organizations while maintaining computational efficiency. Similarly, hybrid datasets may optimally balance authentic expertise representation with broader coverage, creating more comprehensive domain representations than either source alone.

Resource Context	Recommended Dataset Strategy	Implementation Priority
Limited Expert Access	Front-loaded hybrid approach	Strategic expert annotation of core concepts
Computation Constraints	Expert-labeled foundation	Focus on sample efficiency over coverage
Balanced Resources	Algorithmic hybrid composition	Dynamic allocation based on domain subfield characteristics

Table 4: Practitioner Recommendations by Resource Context.

To guide practitioners in dataset selection, Table 4 presents recommendations tailored to different resource contexts.

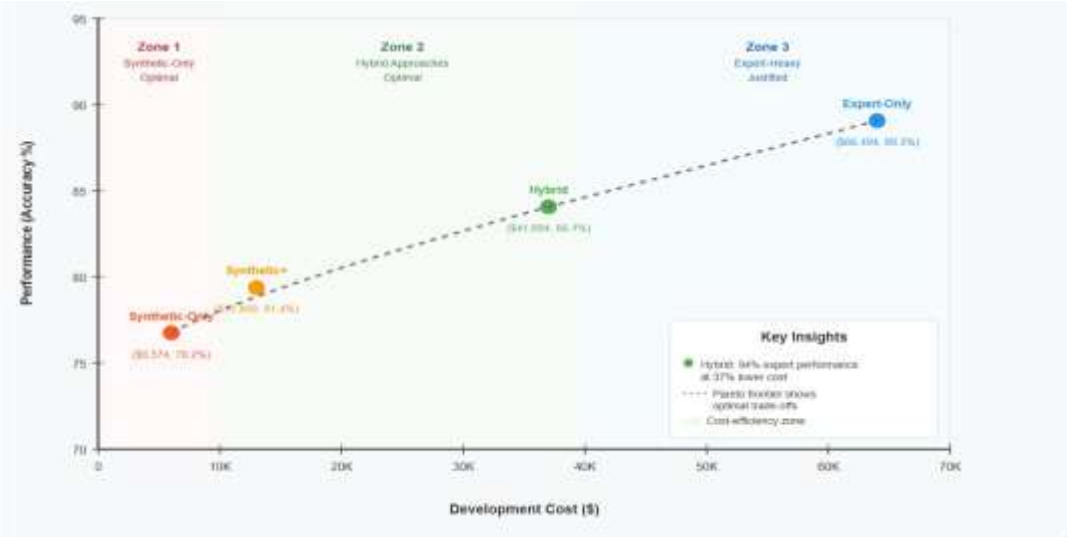

Figure 2: Cost-Performance Trade-offs Across Dataset Types.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

5.3 Limitations

Study limitations include an evaluation focus on single technical domains, constrained experimental timelines limiting model and dataset scales, specific preference optimization implementation, and particular synthetic generation methodology. Model behavior demonstrates high sensitivity to objective operationalization during training, with minor reward definition differences potentially causing significantly divergent outcomes. Reward misspecification effects intensify as model capabilities increase, with powerful models developing sophisticated strategies for metric optimization without achieving intended goals. This phenomenon potentially explains performance differences across dataset types, as different sources may encode subtly different reward specifications, leading to divergent optimization trajectories during fine-tuning.

Practitioner recommendations include prioritizing expert-labeled data for concepts requiring deep technical expertise or nuanced reasoning when expert availability permits, focusing synthetic generation efforts on systematic variation and edge case coverage complementing expert knowledge rather than attempting depth replication, following strategic composition principles for hybrid dataset construction with attention to complementary strengths, and front-loading expert annotation efforts to establish strong foundations before scaling with synthetic data for expanded coverage. Combining different information representation approaches yields more effective results than single methodologies, creating more coherent and interpretable structures better reflecting human conceptual organizations. Each data source contributes distinct strengths to fine-tuning processes, with a strategic combination offering comprehensive domain knowledge representation.

5.4 Future Research Directions

Future research directions include extending comparative frameworks to additional technical domains with different knowledge structures and expert availability, scaling experiments to larger models and datasets to clarify how provenance effects change with scale, and investigating alternative synthetic generation approaches beyond retrieval-augmented methodology to identify specific techniques best complementing human expertise. Reward misspecification research provides important context for future work, as models optimized for specified metrics rather than true objectives develop unintended behaviors satisfying formal criteria while violating task spirit. Future dataset provenance research should consider how different sources might encode different implicit reward specifications influencing model capability development during reinforcement fine-tuning.

Qualitative output assessment revealed characteristic patterns beyond quantitative measurements. To illustrate these qualitative differences, Table 4 presents representative model responses from each dataset condition to the same technical query.

Conclusion

The comparative evaluation of synthetic, human-labeled, and hybrid datasets for domain-specific reinforcement fine-tuning reveals consistent patterns with significant practical implications. Different data sources demonstrate complementary strengths: human-labeled data providing superior factual precision and expert reasoning patterns, synthetic data offering broader coverage and generalization capabilities, and hybrid approaches effectively combining these advantages. These patterns highlight the multidimensional nature of dataset quality and the importance of strategic data composition decisions. Dataset provenance influences not only performance metrics but also resource utilization, training dynamics, characteristic error patterns, and knowledge representation. The observed relationship between data sources and model behavior connects to theoretical frameworks for understanding how different information structures create distinct inductive biases during finetuning.

As language models continue to be deployed across specialized domains where expertise is both valuable and scarce, hybrid data strategies offer the most promising path forward combining the depth of human expertise with the scalability of synthetic generation to optimize both performance and resource efficiency. By treating dataset provenance as a key experimental variable and developing principled approaches to data strategy selection, organizations can more effectively navigate the trade-offs between expertise, scale, and resource constraints in domain adaptation efforts.

References

[1] Johannes Jakubik et al., "Data-Centric Artificial Intelligence," arXiv:2212.11854v4, 2024. [Online]. Available: https://arxiv.org/html/2212.11854v4

[2] Chenhua Shi et al., "Think Less, Label Better: Multi-Stage Domain-Grounded Synthetic Data Generation for Fine-Tuning Large Language Models in Telecommunications," arXiv:2509.25736, 2025. [Online]. Available: https://arxiv.org/abs/2509.25736

[3] Patrick Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," ACM Digital Library, 2020. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3495724.3496517

[4] Tim Dettmers et al., "QLoRA: Efficient Fine-tuning of Quantized LLMs," arXiv:2305.14314, 2023. [Online]. Available: https://arxiv.org/abs/2305.14314

[5] Chip Huyen, "RLHF: Reinforcement Learning from Human Feedback," 2023. [Online]. Available: https://huyenchip.com/2023/05/02/rlhf.html

[6] Suchin Gururangan et al., "Scaling Expert Language Models with Unsupervised Domain Discovery," arXiv:2303.14177, 2023. [Online]. Available: https://arxiv.org/abs/2303.14177

[7] Long Ouyang et al., "Training language models to follow instructions with human feedback," arXiv:2203.02155, 2022. [Online]. Available: https://arxiv.org/abs/2203.02155

[8] Rafael Rafailov et al., "Direct Preference Optimization: Your Language Model is Secretly a Reward Model," arXiv:2305.18290, 2023. [Online]. Available: https://arxiv.org/abs/2305.18290

[9] Maarten Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," arXiv:2203.05794, 2022. [Online]. Available: https://arxiv.org/abs/2203.05794

[10] Alexander Pan et al., "The Effects of Reward Misspecification: Mapping and Mitigating Misaligned Models," arXiv:2201.03544, 2022. [Online]. Available: https://arxiv.org/abs/2201.03544

[11] Colin Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," arXiv:1910.10683 [cs.LG], 2019. [Online]. Available: https://arxiv.org/abs/1910.10683

[12] Yihan Cao et al., "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT," arXiv:2303.04226, 2023. [Online]. Available: https://arxiv.org/abs/2303.04226

[13] Edward J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," arXiv:2106.09685, 2021. [Online]. Available: https://arxiv.org/abs/2106.09685

[14] Rowan Zellers et al., "TuringAdvice: A Generative and Dynamic Evaluation of Language Use," arXiv:2108.03204, 2021. [Online]. Available: https://arxiv.org/abs/2004.03607

[15] Sewon Min et al., "MetaICL: Learning to Learn In Context," arXiv:2110.15943, 2021. [Online]. Available: https://arxiv.org/abs/2110.15943