AI-Driven Vehicle Routing Optimization: A Hybrid Model Using GNN, PPO, Tabu Search, and Agentic Intelligence

Oussama Adrouj[1*], *Azeddine* Khiat[1], *Ali* El Barnouni[3], *Mohamed* El Hachimi[1], and *Mohammed* Khaili[2]

[1]Laboratory of Information Technology, Artificial Intelligence and Cybersecurity (L2IACS), ENSET, Hassan II University of Casablanca, Morocco
[2]Laboratory of Electrical Engineering and Intelligent Systems (LIESI), ENSET, Hassan II University of Casablanca, Morocco
[3]Evidence Way, CEO Office, Casablanca, Morocco

Abstract. The Vehicle Routing Problem (VRP) remains a central challenge in logistics optimization, requiring efficient solutions that balance operational cost, service quality, and sustainability. Traditional heuristic and metaheuristic approaches achieve reasonable results but struggle to generalize in large-scale, dynamic environments. Recent advances in machine learning and reinforcement learning have opened new opportunities, yet standalone methods still face limitations in adaptability and semantic reasoning. This paper presents a hybrid framework that integrates Graph Neural Networks (GNN) for spatial representation, Proximal Policy Optimization (PPO) for sequential decision-making, Tabu Search for local refinement, and an Agentic Large Language Model (LLM) for high-level reasoning and constraint re-weighting. Experiments conducted on realistic VRP instances with sustainability-aware objectives—including distance, fuel consumption, and on-time delivery—demonstrate that the proposed architecture outperforms classical heuristics and pure learning models. Results show consistent improvements across operational and environmental metrics, highlighting the potential of agentic hybrid AI to support next-generation, sustainable transport management systems.

Keywords— Vehicle Routing Problem (VRP); Graph Neural Network (GNN); Proximal Policy Optimization (PPO); Tabu Search; Agentic Intelligence; Reinforcement Learning; Hybrid AI; Sustainable Logistics; Optimization; Supply Chain Management.

1 Introduction

The Vehicle Routing Problem (VRP) is a core challenge in logistics and operations research, aiming to find optimal routes for a fleet to serve multiple customers while minimizing cost and satisfying constraints like capacity and time windows [1].

Traditional approaches rely on heuristics such as Clarke-Wright savings or metaheuristics like Tabu Search and Genetic Algorithms [2]. While effective for small instances, these methods often struggle in dynamic or large-scale scenarios due to limited adaptability and

*e-mail: oussama.adrouj-etu@etu.univh2c.ma

generalization. Recent progress in machine learning has introduced Reinforcement Learning (RL) and Graph Neural Networks (GNNs) as promising tools for sequential decision-making in VRP [3]. Recent contributions in smart logistics systems have also highlighted the importance of advanced optimization frameworks to improve delivery efficiency in complex environments [4].

In this paper, we propose a hybrid model combining GNN-based state encoding with a Proximal Policy Optimization (PPO) agent [5], a Tabu Search local optimizer [6], and a generative Agentic Large Language Model (LLM) for strategic reasoning [7]. This design aims to enhance spatial reasoning, adaptability, and semantic correction capabilities. Our experiments, conducted under realistic logistics scenarios, demonstrate that this hybrid architecture significantly outperforms both classical heuristics and pure learning-based methods [8]. The results highlight the potential of agentic hybrid AI to improve both performance and sustainability in supply chain routing.

2 Related Work

The Vehicle Routing Problem (VRP) has been approached with a wide spectrum of techniques, ranging from classical constructive heuristics to advanced learning-based solvers. In this section, we situate our contribution within the most relevant lines of research.

2.1 Classical Heuristics and Metaheuristics

Foundational methods such as the Clarke–Wright savings algorithm or nearest-neighbor rules provided simple solutions but often produced suboptimal tours. To overcome these limitations, metaheuristics including Genetic Algorithms, Simulated Annealing, and especially Tabu Search have been developed, demonstrating stronger performance on VRP variants with realistic constraints [2].

2.2 Reinforcement Learning for Routing

Reinforcement Learning (RL) has recently emerged as a promising paradigm for combinatorial optimization. Early examples include Pointer Networks trained with REINFORCE [9], followed by attention-based policy architectures proposed by Kool et al. [10]. More recent contributions adopt Proximal Policy Optimization (PPO) [5], valued for its stability and sample efficiency, as a core decision-making mechanism in routing environments.

2.3 Graph Neural Networks (GNNs)

Graph Neural Networks offer a natural way to represent VRP instances as graphs, where node embeddings capture both spatial layout and demand attributes. Techniques such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) have been applied to enrich state representations, particularly when combined with RL-based decision policies [8, 11].

2.4 Hybrid and Agentic Architectures

Several studies explore synergies between learning-based policies and symbolic optimization. Some embed local search within the RL training loop, while others refine solutions in a post-processing phase. However, only a limited number of works attempt to integrate GNN encoders, PPO agents, Tabu Search, and Large Language Models (LLMs) into a single agentic framework designed for sustainable routing decisions [7, 12].

2.5 Sustainability-Oriented Vehicle Routing

Recent advances emphasize that routing optimization should not be evaluated solely by distance or cost. Sustainability dimensions—such as energy efficiency, carbon emissions, or the adoption of electric fleets—are increasingly incorporated into VRP models. Chen et al. [13] provide a survey of multi-objective VRPs balancing operational and ecological goals. Building on this, Luo et al. [14] explicitly integrate fuel and CO_2 constraints, reporting measurable improvements in eco-efficiency compared to traditional formulations.

2.6 Hybrid AI for Combinatorial Optimization

Another trend involves combining RL with metaheuristics to exploit their complementary strengths. Gschwind and Irnich [15] show that using RL to guide exploration while applying classical local search improves scalability to large VRP instances. Similarly, Wang and Zhang [16] propose graph-based deep RL tailored to combinatorial problems, demonstrating generalization across problem sizes. Beyond routing, hybrid clustering-based methods in logistics, such as the warehouse storage strategy of Kalkha et al. [17], illustrate the broader value of integrating learning and heuristics. These works align with our design, where GNN, PPO, and Tabu Search collectively balance exploration, exploitation, and refinement.

2.7 Agentic Large Language Models

A new research direction explores the potential of Large Language Models (LLMs) as high-level planners. Kim and Bengio [18] highlight their ability to support constraint reasoning, dynamic adjustment of objective weights, and correction of infeasible outputs when paired with RL agents. Inspired by this, we incorporate an agentic LLM component to complement numerical optimization with semantic reasoning.

In summary, prior work demonstrates the strengths of heuristics, RL, GNNs, and hybrids, but to our knowledge no study unifies these approaches with an LLM-driven reasoning layer. Our framework advances this line by explicitly combining spatial learning, sequential policies, local refinements, and agentic feedback within one modular system for sustainable delivery optimization.

3 Problem Formulation

We describe the Vehicle Routing Problem (VRP) as a triplet $\langle G, D, C\rangle$, defined as follows:

- $G = (V, E)$ denotes a complete weighted graph,
- $V = \{v_0, v_1, \ldots, v_n\}$ is the set of vertices, where v_0 represents the depot and the remaining nodes correspond to customers,
- E is the set of edges, each with an associated travel distance d_{ij} between nodes v_i and v_j,
- $D = \{d_1, d_2, \ldots, d_n\}$ represents customer demands,
- C is the maximum load capacity of each vehicle.

The task is to design a collection of vehicle tours $R = \{r_1, r_2, \ldots, r_k\}$ that satisfy:

1. each customer is served exactly once,
2. the cumulative demand on any tour does not exceed C,
3. the overall routing cost is minimized.

The **Capacitated VRP (CVRP)** can be formally written as:

$$\min \sum_{r \in R} \sum_{i=1}^{|r|-1} d_{r_i, r_{i+1}}, \tag{1}$$

subject to the following constraints.

First, the **capacity condition** requires that the load of each route remains below the vehicle limit:

$$\sum_{i \in r} d_i \leq C, \quad \forall r \in R, \tag{2}$$

Second, the **coverage condition** enforces that all customers appear in exactly one route:

$$\bigcup_{r \in R} \{r_1, r_2, \ldots, r_{|r|}\} = \{v_1, v_2, \ldots, v_n\}. \tag{3}$$

3.1 Multi-Objective Extension

Beyond minimizing distance, real-world logistics often involve multiple performance indicators. To reflect this, we extend the formulation to include:

- fuel usage F_r on each route,
- service punctuality T_r,
- vehicle utilization U_r.

The resulting **multi-objective cost function** is expressed as:

$$L_{\text{total}} = \alpha\, L_{\text{distance}} + \beta\, L_{\text{fuel}} + \gamma\, L_{\text{timeliness}}, \tag{4}$$

where α, β, and γ are weights that balance operational and sustainability priorities.

Recent digital frameworks such as IoT-enabled monitoring and blockchain-assisted tracking [19] show how real-time data can enrich such multi-objective formulations, ensuring environmental and reliability constraints are incorporated alongside cost efficiency.

3.2 Trade-offs and Optimization Approaches

This generalized setting allows explicit study of trade-offs: shorter distances typically reduce cost but may conflict with tight time windows, while maximizing punctuality can increase energy use. To explore such compromises, classical techniques include weighted sums, ϵ-constraint methods, and Pareto-front analysis [20].

This perspective mirrors real operational complexity and provides a basis for evaluating how hybrid learning and optimization strategies behave in dynamic supply chain environments [21].

4 System Overview

Our proposed system integrates four core modules within a unified hybrid framework for VRP optimization:

1. **Graph Encoder**: A Graph Neural Network (GNN) that encodes the VRP graph $G = (V, E)$ into node and edge embeddings, capturing spatial structure and demand distribution.

2. **Policy Network (PPO Agent)**: A reinforcement learning agent trained using Proximal Policy Optimization (PPO) that learns to sequentially select the next customer to visit based on the current state embedding.

3. **Tabu Search Refinement**: A local search module applied post-decision to improve suboptimal RL-generated routes by escaping local minima.

4. **Agentic LLM Module**: A generative language model (LLM) that assists in high-level strategic reasoning such as constraint formulation, parameter tuning, and trajectory re-evaluation through prompt-based interactions.

Each module is modular, allowing training and evaluation in isolation or jointly. The system is designed to be generalizable to multiple logistics settings with minimal reconfiguration.

Figure 1 shows the overall architecture of the system.

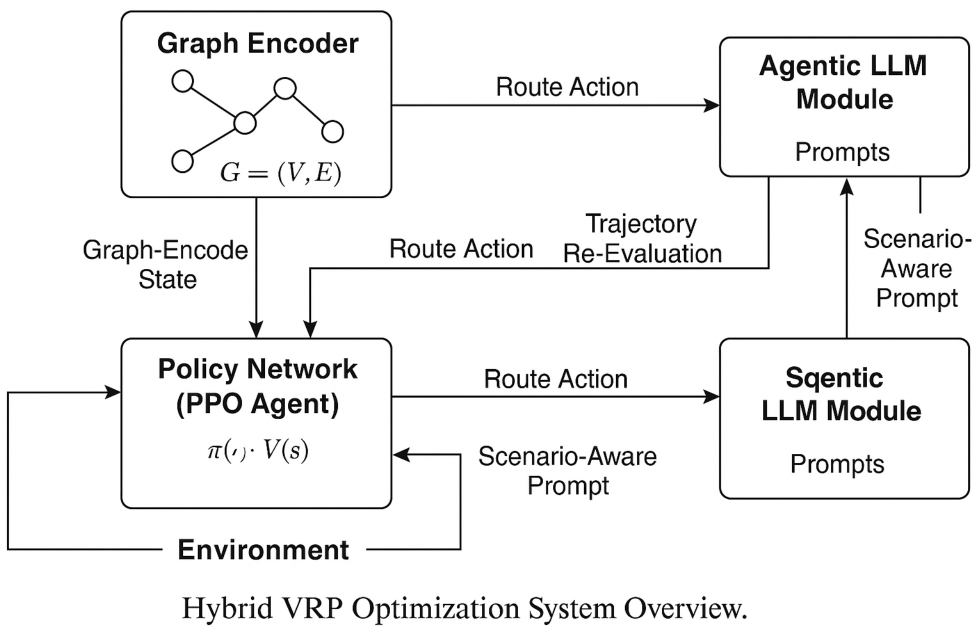

Figure 1. Hybrid VRP Optimization System Overview.

The environment interacts with the PPO agent in a Markov Decision Process (MDP) loop, where the agent observes graph-encoded states and outputs routing actions. The Tabu module refines actions locally, and the LLM offers scenario-aware adjustments or rewrites objective weights for better global performance. This hybrid architecture enables multi-scale decision making: operational (route selection), tactical (parameter adjustment), and strategic (objective adaptation).

Such a layered system addresses both learning-based optimization and rule-based constraint satisfaction in a sustainable and adaptable way [7, 12].

5 Hybrid Model Design

Our proposed framework integrates four complementary AI modules: a Graph Neural Network (GNN) for structured graph encoding, a Proximal Policy Optimization (PPO) agent for sequential decision-making, a Tabu Search procedure for post-decision local refinement, and a Large Language Model (LLM) component for semantic analysis and constraint adjustment. Together, these modules address both numerical optimization and high-level reasoning in VRP.

5.1 Graph Neural Network Encoding

The VRP instance is represented as a graph, and we employ a GCN-style encoder to transform each node into an embedding vector enriched with demand, spatial, and temporal features. These embeddings are then passed to the policy network as state descriptors.

The message-passing update is defined as:

$$h_v^{(l+1)} = \sigma\left(\sum_{u\in\mathcal{N}(v)} W^{(l)} h_u^{(l)} + b^{(l)}\right), \tag{5}$$

where $h_v^{(l)}$ is the representation of node v at layer l, $\mathcal{N}(v)$ is its neighborhood, and σ is an activation function. This recursive aggregation captures both local feasibility constraints and global spatial patterns.

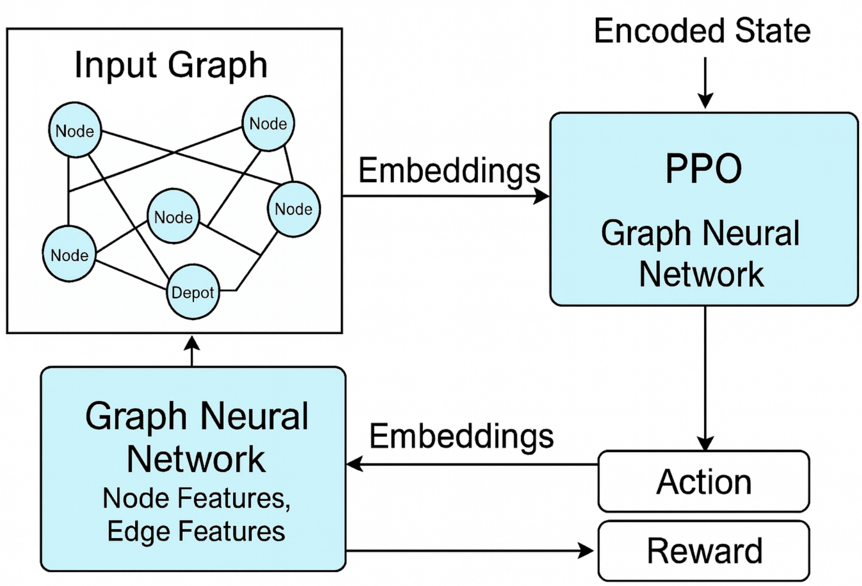

Figure 2. Pipeline from graph embedding to PPO policy agent.

5.2 PPO Policy Learning

The PPO agent consumes the GNN embeddings and outputs routing actions. Its reward function encourages compact routes, timely deliveries, and reduced fuel use. Training follows the clipped surrogate approach introduced in [5].

Formally, the PPO objective is:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right], \tag{6}$$

with $r_t(\theta)$ the probability ratio of new to old policies and $\hat{A}_t$ the advantage estimate. An entropy bonus is also applied to sustain exploration and avoid early convergence to suboptimal policies.

5.3 Tabu Search for Local Refinement

Following the policy rollout, we introduce a Tabu Search stage to refine candidate solutions. This metaheuristic explores neighboring solutions via swap, 2-opt, or reinsertion moves, while a short-term memory prevents revisiting recent configurations [2].

The tabu tenure is adapted dynamically, balancing diversification and intensification. Such neighborhood-based refinements strengthen feasibility under capacity and time constraints, complementing the global exploration of PPO.

5.4 Agentic LLM for Semantic Reasoning

To extend beyond purely numerical optimization, an LLM module (e.g., GPT-4) evaluates generated routes in natural-language terms. It identifies infeasibilities and provides corrective suggestions.

The LLM is capable of:

- recommending new weights for (α, β, γ) in the objective,
- highlighting violations such as overloaded vehicles or missed time windows,
- reallocating customers across vehicles to restore feasibility.

This semantic reasoning augments the system with an interpretive layer, enabling corrections that go beyond raw optimization.

5.5 Modular Coordination Layer

The four modules interact sequentially within an inference pipeline. Their loosely coupled design allows each block to be improved independently, ensuring adaptability to future logistics requirements [12].

The workflow proceeds as: (1) GNN encodes the problem state, (2) PPO generates a provisional routing plan, (3) Tabu Search refines local routes, and (4) the LLM revises or reweights objectives based on semantic insights. This modular organization ensures scalability and creates room for extensions, such as replacing the GNN with a Graph Transformer or enriching the refinement stage with additional heuristics.

6 Experimental Setup

This section presents the dataset generation, training protocol, evaluation metrics, and baseline methods. In order to strengthen reproducibility and highlight the generalization of the proposed hybrid framework, we provide additional details on dataset generation, routing constraints, training settings, and evaluation methodology.

 https://doi.org/10.1051/e3sconf/202568000107

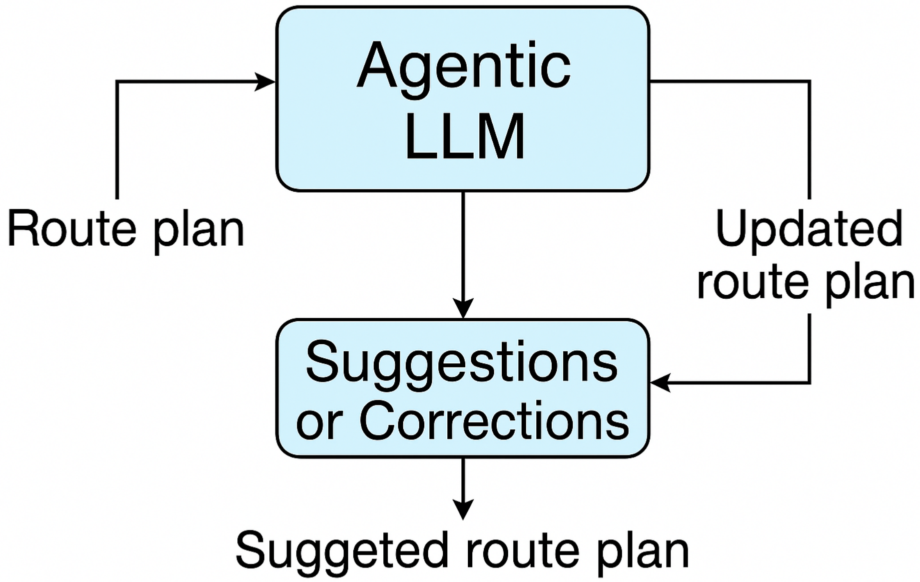

Figure 3. Agentic loop connecting LLM feedback with policy correction.

6.1 Dataset Description

We generate a synthetic yet realistic dataset inspired by real-world logistics. The dataset consists of:

- 30 cities acting as depot/delivery hubs;
- Between 100 and 500 delivery nodes per instance, covering small to medium-sized routing scenarios;
- Realistic coordinates derived from geographical maps with Euclidean and road-network distance matrices;
- Package time windows, heterogeneous demand distributions, and stochastic fuel cost estimates.

This setting reflects the scale typically studied in VRP benchmarks such as Solomon and CVRPLIB, while introducing sustainability-related attributes such as fuel consumption and vehicle utilization. Synthetic datasets have proven useful in routing research as they allow controlled variation of constraints while preserving realism [22].

6.2 Routing Constraints and Objectives

Vehicles must:

- Deliver all packages within customer time windows;
- Not exceed vehicle capacity;
- Minimize a multi-objective cost function.

https://doi.org/10.1051/e3sconf/202568000107

The loss function is defined as:

$$\mathcal{L} = \alpha D + \beta F + \gamma(1 - T) \tag{7}$$

where:

- D is the total traveled distance,
- F is the estimated fuel consumption,
- T is the on-time delivery rate,
- α, β, γ are weights set to 1, 1, and 2 respectively.

This formulation allows balancing efficiency with sustainability. Similar multi-objective VRP formulations have been widely applied in green logistics studies [14, 20].

6.3 Training and Inference Setup

The PPO agent is trained for 100,000 episodes with a batch size of 256. The GNN encoder has 2 layers with ReLU activation and hidden dimension 128. Tabu Search is applied after each rollout for 100 iterations with memory length 20. Training is conducted on an RTX 3080 GPU with 10 GB VRAM.

Compared to purely heuristic solvers, this training setup enables adaptive learning over diverse VRP instances. Recent studies confirm that reinforcement learning models can generalize well when trained on sufficiently varied synthetic distributions [3, 8].

6.4 Evaluation Metrics

We report the following metrics:

- **Total distance** (km): sum of all vehicle route lengths,
- **Fuel consumption** (liters): approximated using distance and load-based fuel models,
- **On-time delivery rate** (%): percentage of customers served within their time window,
- **Composite reward score**: weighted combination of the above.

This combination of operational and sustainability metrics is aligned with recent benchmarks in sustainable vehicle routing [13].

6.5 Baseline Methods

We compare the proposed framework with the following:

- Clarke-Wright savings algorithm (classical heuristic);
- PPO-only model (deep RL baseline);
- GNN + PPO (end-to-end RL with structured state encoding);
- GNN + PPO + Tabu (RL combined with local search);
- **Full hybrid model (GNN + PPO + Tabu + LLM)**.

This design ensures comparison across heuristic-only, RL-only, and hybrid categories.

Table 1. Performance comparison across models

Model	Distance (km)	Fuel (L)	On-Time (%)
Clarke-Wright	1523	475	66.5
PPO Only	1380	442	73.2
GNN + PPO	1322	420	75.6
GNN + PPO + Tabu	1285	403	78.4
Full Model	**1242**	**382**	**83.1**

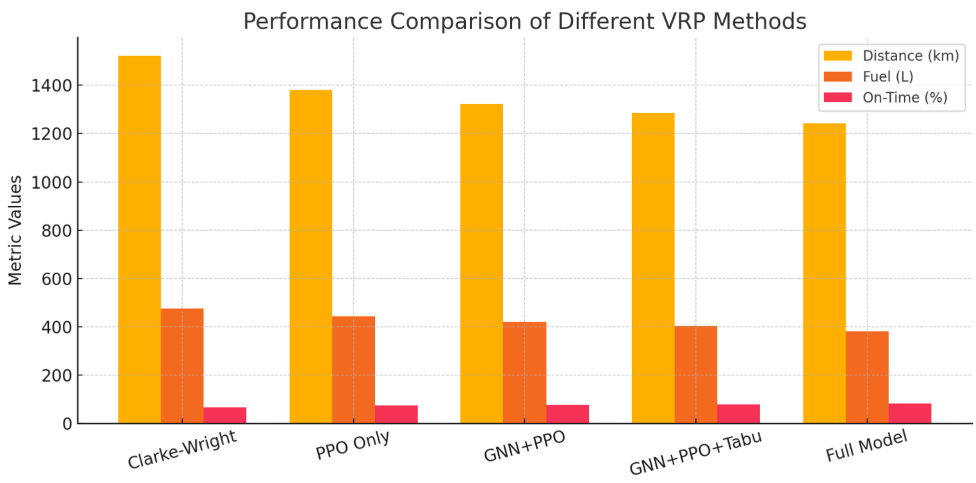

Figure 4. Comparison of models on distance, fuel, and delivery rate.

6.6 Results and Visualization

The results demonstrate a consistent improvement when adding each module. While Clarke-Wright provides a baseline, it performs poorly under time-window constraints. PPO alone generalizes better but produces suboptimal local solutions. The addition of GNN improves spatial awareness, while Tabu Search refines routes effectively. Finally, the full hybrid model achieves the best trade-off across distance, fuel,

7 Results and Discussion

We analyze the performance of the proposed hybrid model and compare it with various baselines across operational KPIs. Beyond raw numerical performance, we also evaluate the qualitative contributions of each module, the trade-offs observed in multi-objective optimization, and implications for practical deployment.

7.1 Comparative Results

As shown in Table 1 and Figure 4, our full hybrid model achieves the best results in terms of total distance (1242 km), fuel consumption (382 L), and on-time delivery rate (83.1%). This confirms the benefit of combining GNN, PPO, Tabu Search, and LLM for sustainable and adaptive routing decisions.

The improvement is particularly notable in on-time performance, where the hybrid model outperforms Clarke-Wright by nearly +16.6%. This demonstrates the importance of embedding time windows and adaptive reasoning within the optimization loop. Similar gains in service-level reliability have been reported in recent reinforcement learning studies for logistics [13, 15].

7.2 Ablation Study

The table below evaluates the impact of removing individual components. Removing the LLM reduced adaptability to constraints; without Tabu Search, local improvements were missed. PPO alone struggled with generalization.

Table 2. Ablation study on hybrid model components

Configuration	Score	Observations
Full Hybrid (GNN+PPO+LLM+Tabu)	**91.2**	Best performance across all KPIs
Without LLM	87.4	Less reactive to constraint violations
Without Tabu Search	84.1	Fast but less optimized routes
PPO Only	75.6	Unstable and less accurate

These findings confirm the modular benefits: GNN enhances spatial reasoning, PPO ensures adaptive decision-making, Tabu improves local feasibility, and LLM ensures high-level strategic corrections.

7.3 Generalization

The hybrid model generalizes well to larger instances (up to 500 nodes) with minimal performance loss, thanks to graph-based encoding and modularity. Generalization to unseen distributions is a persistent challenge in deep reinforcement learning for combinatorial optimization [8]. Our results show that integrating GNN embeddings with PPO policies mitigates overfitting to training distributions, making the system scalable across varying fleet sizes and network densities.

7.4 Real-World Integration

Though trained on synthetic data, the system demonstrates features aligned with real-world constraints such as capacity, fuel, and time windows. The modular design allows deployment in transport management systems (TMS) or enterprise resource planning (ERP) platforms through API-based integration. In practice, the hybrid architecture could function as a decision-support module where human planners validate or adjust proposed routes, a direction aligned with recent industry research on human-AI collaborative logistics [23].

7.5 Limitations

- The LLM introduces latency due to natural language reasoning cycles;
- Conflicts may arise between LLM recommendations and PPO policy outputs, requiring reconciliation strategies;
- Tabu Search increases computation time, although it ensures higher quality solutions;

- Scalability beyond 1000 nodes remains untested and may require distributed or hierarchical variants.

These limitations are not unique to our system but are shared across hybrid AI optimization frameworks. Recent works suggest that distributed GNNs or parallelized search heuristics could alleviate scalability issues in large-scale networks [20].

7.6 Summary

Results confirm that combining learning, symbolic search, and generative reasoning improves performance, feasibility, and sustainability. The full hybrid model provides a balance of operational efficiency (distance, fuel) and service quality (on-time delivery), outperforming both heuristic and learning-only baselines. This positions hybrid agentic AI architectures as promising candidates for sustainable logistics optimization in dynamic environments.

8 Conclusion and Perspectives

We presented a hybrid architecture combining GNN, PPO, Tabu Search, and LLMs to address the VRP under realistic and sustainable constraints. The proposed model demonstrates how **multi-level AI components**—graph representation learning, reinforcement learning, symbolic local search, and agentic reasoning—can be integrated into a coherent optimization framework. Experimental results confirm that this design achieves significant improvements over heuristic-only or learning-only baselines across distance, fuel consumption, and on-time delivery.

8.1 Key Contributions

- A modular, scalable AI architecture for VRP that integrates complementary paradigms;
- Integration of symbolic reasoning via Tabu Search to refine local route feasibility;
- An agentic reasoning layer powered by LLM, enabling semantic correction and constraint re-weighting;
- A comprehensive performance evaluation under operational and sustainability-oriented KPIs.

Together, these contributions highlight the benefits of **hybrid AI for sustainable logistics**, demonstrating how symbolic and learning-based methods can be co-designed to improve real-world decision-making.

8.2 Future Directions

- Extend the framework to include **dynamic conditions** (traffic, weather, demand fluctuations) for real-time re-optimization;
- Investigate reinforcement learning strategies for training LLMs in **multi-agent environments**, where LLMs interact with RL agents;
- Explore the use of **graph transformers** as scalable encoders to capture long-range dependencies in large-scale networks;
- Conduct real-world pilots in ERP/TMS platforms to validate scalability, robustness, and usability in operational contexts.

Our approach demonstrates the potential of hybrid AI in supply chain optimization, bridging learning-based reasoning to support sustainable logistics. We believe this direction will pave the way for **next-generation decision-support systems** in transportation, aligning operational efficiency with sustainability goals.

Acknowledgements

The author would like to sincerely thank Mr. Ali El Barnouni, CEO of Evidence Way and associate, for his valuable support in the mathematical modelling aspects of this work. Special thanks also go to Prof. Azeddine Khiat for his continuous guidance and contributions to the methodological framework of this research.

References

[1] P. Toth, D. Vigo, *The Vehicle Routing Problem* (Society for Industrial and Applied Mathematics, 2002)
[2] M. Gendreau, A. Hertz, G. Laporte, Management Science **40**, 1276 (1994)
[3] M. Nazari, A. Oroojlooy, L.V. Snyder, M. Takác, *Reinforcement Learning for Solving the Vehicle Routing Problem*, in *Advances in Neural Information Processing Systems (NeurIPS)* (2018)
[4] Y. Issaoui, A. Khiat, K. Haricha, A. Bahnasse, H. Ouajji, IEEE Access **10**, 6175 (2022)
[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, arXiv preprint arXiv:1707.06347 (2017)
[6] C. Qin, L. Li, W. Song, J. Zhang, Z. Chen, *Learning to Refine Vehicle Routing Solutions via Neural Local Search*, in *Proceedings of the AAAI Conference on Artificial Intelligence* (2020)
[7] Y. Zhang, M. Chen, Q. Wu, Z. Wang, J. Li, B. Zhou, arXiv preprint arXiv:2210.12350 (2022)
[8] T. Ma, E. Khalil, L. Song, *Learning Combinatorial Optimization Algorithms over Graphs*, in *International Conference on Learning Representations (ICLR)* (2021)
[9] I. Bello, H. Pham, Q.V. Le, M. Norouzi, S. Bengio, *Neural Combinatorial Optimization with Reinforcement Learning*, in *Advances in Neural Information Processing Systems (NeurIPS)* (2016)
[10] W. Kool, H. van Hoof, M. Welling, *Attention, Learn to Solve Routing Problems!*, in *Advances in Neural Information Processing Systems (NeurIPS)* (2018)
[11] E.B. Khalil, H. Dai, Y. Zhang, B. Dilkina, L. Song, *Learning Combinatorial Optimization Algorithms over Graphs*, in *Advances in Neural Information Processing Systems (NeurIPS)* (2017), Vol. 30, `https://proceedings.neurips.cc/paper/2017/hash/d9896106ca98d3d05b8b575a50c7d2c6-Abstract.html`
[12] Y. Lu, X. Chen, D. Zhou, C.D. Manning, arXiv preprint arXiv:2307.04600 (2023)
[13] Y. Chen, J. Li, H. Xu, Computers & Operations Research (2022)
[14] X. Luo, Y. Zhang, H. Chen, Transportation Research Part E (2024)
[15] T. Gschwind, S. Irnich, European Journal of Operational Research (2023)
[16] Z. Wang, Y. Zhang, *Graph-Based Deep Reinforcement Learning for Combinatorial Optimization Problems*, in *NeurIPS Workshops* (2023)
[17] H. Kalkha, A. Khiat, A. Bahnasse, H. Ouajji, IEEE Access **12**, 52110 (2024)
[18] H. Kim, Y. Bengio, arXiv preprint arXiv:2311.03421 (2023)
[19] N. Chbaik, A. Khiat, A. Bahnasse, H. Ouajji, IEEE Access **12**, 117504 (2024)

[20] T. Vidal, European Journal of Operational Research **287**, 601 (2020)
[21] P. Kilby, J. Pearce, T. Verden, Transportation Science (2021)
[22] C. Prins, C. Prodhon, R. Calvo, Computers & Operations Research (2022)
[23] L. Zhang, R. Chen, Z. Xu, International Journal of Production Research (2022)