Hybrid Architectural Paradigms for Neural Combinatorial Optimization: Integrating Multi-Head Self-Attention Structure Learning with Intermittent Message Passing

The landscape of combinatorial optimization (CO) is currently undergoing a foundational shift as researchers transition from rigid, handcrafted heuristic algorithms toward flexible, data-driven deep learning architectures. At the center of this evolution is the challenge of the Capacitated Vehicle Routing Problem (CVRP) and its various extensions, which require solvers to navigate non-Euclidean search spaces and high-dimensional constraints. Traditional methods, such as Genetic Algorithms or Simulated Annealing, while robust, often suffer from scalability limits and a lack of sensitivity to the underlying geometric and structural properties of modern urban logistics. The emergence of Graph Neural Networks (GNNs) and Transformers has provided a new vocabulary for modeling these problems, representing optimization instances as graphs where entities are nodes and relationships are edges.

A particularly potent architectural configuration involves a deep neural network that processes node features, including spatial coordinates (x, y) and a dynamic waste value (w), alongside edge features represented by a comprehensive distance matrix. This architecture utilizes multi-head self-attention (MHSA) layers not merely as feature extractors, but as structure-learning modules that generate a latent adjacency matrix. This learned adjacency matrix is then passed intermittently to message passing graph neural network (MPNN) layers for localized structural refinement and neighborhood aggregation. This hybrid approach effectively addresses the historical limitations of standalone GNNs, such as the over-squashing of long-range dependencies, while maintaining the inductive bias necessary for localized spatial reasoning.

Theoretical Framework of Graph-Based Combinatorial Optimization

Combinatorial optimization tasks, particularly those involving routing and scheduling, are inherently relational. The problem state is typically defined by a set of locations, each with specific attributes, and a cost associated with the transitions between them. Modeling these problems as graphs allows for the preservation of permutation equivariance, a critical inductive bias which ensures that the solver's output is invariant to the arbitrary indexing of nodes in the input data.

In the message passing paradigm, node representations are refined through iterative local exchanges. For a node v with a hidden state h_v, the message passing update at layer k can be generalized as:

Here, \psi represents a message function and \phi an update function, while \mathcal{N}(v)

denotes the neighborhood of node v as defined by the graph's adjacency matrix. However, the efficiency of this process is heavily dependent on the quality of the adjacency matrix provided. In many real-world combinatorial problems, the physical adjacency (e.g., direct road connections) does not fully capture the strategic importance of relationships between nodes, especially when capacity constraints and dynamic priority values are involved.

Node and Edge Feature Integration

The efficacy of a neural solver begins with its ability to encode the multifaceted nature of the problem instance. For urban waste logistics, the node features must encapsulate both spatial and functional data. The inclusion of x and y coordinates provides the spatial grounding, allowing the model to compute relative distances and orientation. Parallel to this, the "waste value" w serves as a critical stochastic parameter, representing the fill level or overflow risk of a collection point.

Feature Category	Data Points	Representation
Node Spatial	Cartesian coordinates	(x_i, y_i)
Node Functional	Waste fill level (w), bin type	$f_i = [w_i, \tau_i]$
Edge Metric	Cost/Distance matrix	$D_{i,j}$
Global Context	Vehicle capacity, depot location	$G = [C_{max}, (x_d, y_d)]$

Edge features, primarily the distance matrix, define the cost landscape of the optimization problem. In advanced architectures, these are not treated as static weights but are often transformed using sinusoidal Gaussian mappings to ensure numerical stability and to align their dimensionality with the latent space of the node embeddings. This alignment is vital because it allows the model to learn complex interdependencies between the cost of a path and the priority of its endpoints.

Multi-Head Self-Attention as a Dynamic Adjacency Operator

Traditional Transformers use self-attention to relate elements of a sequence, regardless of their distance. When generalized to graphs, the multi-head self-attention mechanism can be re-conceptualized as a "graph-forming operator". Instead of relying on a fixed, pre-defined topology, the MHSA layers learn to construct a dynamic, task-specific adjacency matrix \hat{A[span_16](start_span)[span_16](end_span)} at each layer.

The Mechanism of Attention-Derived Topology

The interaction between node embeddings $H \in \mathbb{R}^{N \times d}$ in an attention layer is governed by the learned projections for Query (Q), Key (K), and Value (V). The computation of the attention score for each head h results in a pairwise affinity matrix:
This matrix $\alpha^{(h)}$ represents a learned topology where the edge weight between node i and node j is determined by their contextual relevance to the optimization task. In the context of CVRP, this allows a node with a high waste value w to generate a strong attention signal across the graph, effectively "notifying" all potential collection vehicles of its high priority.
By interpreting attention maps as weighted adjacency matrices, the model gains the ability to discover "soft meta-paths"—higher-order interaction patterns that are not explicitly represented

in the physical road network but are crucial for efficient tour construction. This capability is particularly useful in environments where road distances alone are misleading due to traffic congestion or varying collection times.

Sparsity and Isomorphism in Learned Structures

A significant challenge with dense self-attention is its quadratic complexity $O(N^2)$, which can lead to memory exhaustion when processing large-scale urban graphs. To mitigate this, researchers have introduced sparsification techniques that retain only the most significant relationships in the learned adjacency matrix.

One influential framework, the Sparse-GIN-Attention model, implements a thresholding mechanism (e.g., retaining only values above 0.2) to convert the dense attention map into a sparse causal graph. This sparse structure is then passed to Graph Isomorphism Network (GIN) layers. GINs are uniquely suited for CO problems because they are theoretically guaranteed to be as expressive as the Weisfeiler-Lehman (WL) graph isomorphism test, enabling the model to distinguish between different structural configurations and generalize to unseen problem instances. The combination of global reasoning (via MHSA) and isomorphic consistency (via GIN) creates a representation space that is both robust and context-aware.

Intermittent Information Exchange and Architectural Hybridization

The core innovation of the architecture described in the query is the intermittent passing of the learned adjacency matrix to message passing layers. This design choice is motivated by the need to balance global context with local structural refinement, avoiding the common pitfalls of both pure GNNs and pure Transformers.

Addressing Over-Smoothing and Over-Squashing

Standard MPNNs suffer from over-smoothing, where node features converge to similar values after multiple layers, and over-squashing, where the aggregation of information from an exponentially growing neighborhood into a fixed-length vector creates a bottleneck. By intermittently using MHSA layers to "reset" the graph structure and learn a new latent adjacency, the model can effectively bypass these bottlenecks.

The MHSA layer acts as a global reasoning module that identifies long-range dependencies—for instance, connecting two geographically distant clusters of high-priority waste bins that should be served in the same route. Once this global relationship is encoded in the adjacency matrix, the subsequent MPNN layers can perform local neighborhood aggregation based on this refined topology. This ensures that the local message passing is informed by global strategic insights, leading to more coherent and efficient routing solutions.

Cascade vs. Interleaved Architectures

Research on the optimal way to combine Graph Transformers and relational deep learning has highlighted two primary configurations: cascade and interleaved.

Architecture Type	Information Flow	Structural Evolution
Cascade	GNN layers followed by Transformer layers	Sequential; global reasoning after local aggregation
Interleaved	Alternating GNN and Transformer layers	Iterative; structure and features refine each other
Proposed Hybrid	MHSA learned structure passed to MPNN	Dynamic; topology is optimized for local passing

Interleaved architectures are generally superior for complex tasks where the structure of the problem itself is dynamic. In the case of CVRP with stochastic waste values, the "optimal" graph structure changes as bins are filled or emptied. Intermittently passing a learned adjacency matrix allows the model to continuously re-evaluate the most efficient neighborhood for each node to communicate with, aligning the computational graph with the evolving problem state.

Application to Smart Municipal Waste Management

The integration of node features like (x, y, w) is most salient in the domain of smart city logistics, specifically municipal solid waste (MSW) collection. Static collection schedules, which represent the traditional approach, often result in vehicle underutilization or bin overflows. A deep neural solver utilizing the hybrid MHSA-MPNN architecture offers a dynamic alternative that responds to real-time IoT telemetry.

IoT-Enabled Telemetry and Priority Scoring

In a modern SWMS (Smart Waste Management System), bins are equipped with ultrasonic sensors and LPWAN modules (NB-IoT or LoRaWAN) to transmit their fill levels (w). These raw values are often passed through predictive models, such as XGBoost or LSTM, to generate overflow risk scores.
The hybrid architecture then receives these risk scores as the node attribute w. The learned adjacency matrix $\hat{A}$ effectively creates high-weight edges between bins with high overflow risk, regardless of their physical distance, if a strategically efficient route can connect them. This transforms the CVRP from a purely distance-minimization problem into a multi-objective optimization problem that balances travel cost, service priority, and vehicle capacity.

Operational Efficiency and Sustainability Gains

The impact of integrating AI, IoT, and graph-theoretic methods into waste logistics is significant. Comparative studies have shown that dynamic routing using these technologies can drastically improve key performance indicators (KPIs).

Performance Metric	Gain with Hybrid AI/IoT System	Reference
Overflow Events	50% Reduction	
Missed Pickups	72.7% Reduction	
Fuel Consumption	15.5% Reduction	
Bin Utilization	35.5% Improvement	
Total Collection Distance	50.31% Reduction	

In large urban areas like Surat, India, the application of such models has demonstrated the ability to cut overall collecting distances by half while reducing the vehicle fleet size by 24%.

This not only lowers operational costs but also contributes to sustainability goals by reducing greenhouse gas emissions and environmental degradation.

Optimization Landscapes and Policy Learning

Because combinatorial optimization involves discrete decision-making, the training of these deep neural solvers often moves beyond standard supervised learning into reinforcement learning (RL) and non-differentiable optimization.

Policy-Based Reinforcement Learning

In an RL framework, the hybrid MHSA-MPNN encoder generates high-dimensional node embeddings that capture the current state of the bin network. A decoder, typically a Pointer Network or a temporal pointing decoder, then acts as a policy to sequentially select nodes for the collection tour. The policy $\pi(a|s)$ is trained to maximize a reward function R, which is usually the negative of the total tour length or a weighted sum of fuel efficiency and collection coverage.
Entropy-regularized policy learning is often employed to maintain a balance between exploration and exploitation. This ensures that the model does not become overly focused on local distance-minimizing heuristics and instead learns to discover global strategies that account for long-term capacity constraints.

Zeroth-Order Optimization for Non-Differentiable Operators

A fundamental challenge in training graph Transformers for CO is the presence of non-differentiable operators, such as shortest-path computations or hard-threshold sparsification. To address this, Zeroth-Order Optimization (ZOO) provides a means of gradient estimation without direct differentiation.
However, ZOO can lead to sharp loss landscapes that impair generalization. The integration of Sharpness-Aware Minimization (SAM) helps the optimizer seek out flatter regions of the parameter space. By calculating gradients from a neighborhood of points rather than a single point, SAM improves the model's resilience to small variations in problem instances, which is critical when dealing with noisy real-world data like fluctuating waste generation rates.

Scalability and Distributed Architecture

As neural solvers are deployed in megacities, the size of the graphs they must process grows exponentially. Full-scale training of GNNs on massive adjacency matrices is often blocked by memory limitations of individual GPUs.

Graph Partitioning and Subgraph Optimization

Distributed training frameworks address this by partitioning the large urban graph into several small subgraphs. Each individual subgraph is trained to local optimality, and reinforcement learning is then used to coordinate actions across subgraph boundaries. This hierarchical approach ensures that the model can learn both the detailed local routing patterns and the

broader cross-zone coordination requirements.

Quasi-Linear Attention Alternatives

For tasks where global context is indispensable but dense MHSA is too costly, architectures like the Graph-Enhanced Contextual Operator (GECO) offer a scalable alternative. GECO replaces dense pairwise attention with global convolutions that encompass all nodes, achieving a quasi-linear time complexity $O(N \log N)$. Experimental results indicate that GECO can provide a $169\times$ speedup on two-million-node graphs while preserving, and often improving, prediction quality.

Scalability Method	Computational Complexity	Primary Constraint Addressed
Dense MHSA	$O(N^2)$	Long-range dependencies
Sparse GIN-Attention	$O(E)$ (where $E < N^2$)	Memory efficiency and isomorphism
GECO (Convolutions)	$O(N \log N)$	Training speed and time complexity
Graph Partitioning	$O(N/P)$ (where P is partitions)	Memory limit of single machines

Categorical Insights and the Isomorphic Lens

A deep theoretical understanding of the hybrid MHSA-MPNN architecture can be found in the intersection of graph theory and category theory. In this view, the Transformer architecture is interpreted as a mapping that strengthens relationships within a category (e.g., node features), dynamically adjusting based on contextual importance.

Transformers as Categorical Functors

The attention mechanism, by dynamically constructing adjacency matrices, captures critical intra-level and inter-level relationships. Research from the LAMM laboratory at MIT suggest that by evolving Transformers into hierarchical GIN models, we can reveal an implicit capacity for graph-level relational reasoning. This allows the model to map macroscopic properties (like total tour cost) to microscopic features (like individual bin fill levels) while maintaining abstract structural consistency.

The use of a trainable scale parameter γ in these hybrid models often shows an increasing trend in deeper layers. This suggests that while lower layers focus on raw feature extraction, deeper layers rely more heavily on the structural adjustments provided by the learned adjacency matrix. This "layer-wise" specialization is precisely what makes the intermittent passing to MPNN so effective; it allows the model to refine its structural understanding as the features become more abstract.

Positional Encoding as a Topological Prior

To provide the model with a sense of the graph's global structure, positional encodings (PE) are often integrated into the MHSA layers. Unlike sequence-based Transformers that use absolute indices, graph-aware models use topological PEs such as Laplacian eigenvectors or shortest-path distances. These encodings act as a topological prior, informing the attention mechanism about the "shape" of the urban landscape. This helps the learned adjacency matrix

respect the physical constraints of the distance matrix while still allowing for context-driven global reasoning.

Comparative Analysis of Routing Performance

The effectiveness of hybrid neural solvers is confirmed through extensive benchmarking across multiple CO domains, from network dismantling to dynamic routing.

Network Dismantling and Centrality

The MIND (Message Iteration Network Dismantler) model demonstrates the power of utilizing node embeddings from all message-passing iterations alongside an expressive attention mechanism. MIND achieves competitive performance in identifying vital nodes for network dismantling—a task mathematically similar to finding the critical stop points in a CVRP—without relying on manually engineered features. This proves that pure geometric learning, when properly structured with MHSA and MPNN, can decipher complex structural roles with high generalizability.

Dynamic Graph Constraints

In dynamic TSP and CVRP scenarios where traffic congestion and travel times vary, architectures like GTA-RL (Graph Temporal Attention) outperform static baselines. By simultaneously embedding temporal and graph features, these models capture the "evolution" of the distance matrix, allowing the tour to be optimized for the specific time it will be executed.

Task Domain	Optimal Hybrid Strategy	Key Performance Improvement
Network Dismantling	MIND (Iteration Profiles + Attention)	Generalizes to million-node networks
Waste Management	LSTM/XGBoost + CVRP Hybrid	50% distance reduction in Surat
Sports Scheduling	MARL + GNN Coordination	13% Operational Efficiency gain
Bitcoin Trust Nets	DGTEN (Gaussian MP + Attention)	12.34% MCC improvement
Molecular Modeling	ABT-MPNN (Atom-Bond Trans.)	SOTA across 9 major datasets

Nuanced Conclusions on Neural Architecture for Combinatorial Optimization

The research landscape indicates that the most effective neural solvers for combinatorial optimization are those that bridge the gap between global reasoning and local structural aggregation. The specific architecture of receiving node features (x, y, w) and edge features (distance matrix), using MHSA to learn an adjacency matrix, and passing it intermittently to MPNN layers, represents a state-of-the-art solution to the inherent tensions in graph learning. By interpreting multi-head self-attention as a graph-generating operator, the model gains the flexibility to redefine its own topology based on the specific constraints of the problem instance. This allows priority parameters like the waste value w to fundamentally reshape the connectivity of the graph, ensuring that "urgent" nodes receive adequate attention during the tour

construction phase. The intermittent passing to MPNN layers then enforces localized constraints and structural consistency, mitigating the representational collapse often seen in deep Transformer models.

While computational complexity and the non-differentiability of shortest-path operators remain significant hurdles, the development of distributed training frameworks, quasi-linear attention mechanisms like GECO, and sharpness-aware optimization techniques are paving the way for real-time deployment in massive urban infrastructures. The resulting systems offer not only improved efficiency and reduced costs but also a scalable foundation for the next generation of sustainable smart cities, aligning technological advancement with circular economy principles and global development goals.

Ultimately, the synergy of IoT-enabled real-time sensing, AI-driven predictive analytics, and graph-theoretic optimization provides a resilient decision-support system that can adapt to the complex, stochastic dynamics of modern urban environments. Future research will likely focus on even tighter integration between hardware-level edge processing and these hybrid neural architectures to ensure privacy and efficiency at the source, further solidifying the role of neural CO in the Industry 4.0 paradigm.

Works cited

1. Machine Learning for Combinatorial Optimization: a Methodological Tour d'Horizon - arXiv, https://arxiv.org/abs/1811.06128 2. A survey on deep learning-based algorithms for the traveling salesman problem, https://journal.hep.com.cn/fcs/EN/1159673955055362507 3. Graph Learning - arXiv, https://arxiv.org/html/2507.05636v2 4. AI-IoT-graph synergy for smart waste management: a scalable framework for predictive, resilient, and sustainable urban systems - Frontiers, https://www.frontiersin.org/journals/sustainability/articles/10.3389/frsus.2025.1675021/full 5. A stochastic optimization framework for planning of waste collection and value recovery operations in smart and sustainable cities | Request PDF - ResearchGate, https://www.researchgate.net/publication/325763190_A_stochastic_optimization_framework_for_planning_of_waste_collection_and_value_recovery_operations_in_smart_and_sustainable_citi es 6. An IP Core Mapping Algorithm Based on Neural Networks | Request PDF - ResearchGate, https://www.researchgate.net/publication/346780979_An_IP_Core_Mapping_Algorithm_Based_on_Neural_Networks 7. Generation of Routes Using Genetic Algorithms to Optimize the Domestic Solid Waste Collection - ResearchGate, https://www.researchgate.net/publication/389885308_Generation_of_Routes_Using_Genetic_Algorithms_to_Optimize_the_Domestic_Solid_Waste_Collection 8. Graph Neural Networks for City Policy Recommendations as a Link Prediction Task - DSpace@MIT, https://dspace.mit.edu/bitstream/handle/1721.1/162992/rozario-crozario-meng-eecs-2025-thesis.pdf?sequence=1&isAllowed=y 9. Differential Encoding for Improved Representation Learning over Graphs - arXiv, https://arxiv.org/html/2407.02758 10. Advances in graph neural networks for alloy design and properties predictions: a review, https://www.oaepublish.com/articles/jmi.2025.42 11. A Deep Learning Approach using Graph Neural Networks for Anomaly Detection in Air Quality Data Considering Spatiotemporal Correl - IEEE Xplore, https://ieeexplore.ieee.org/iel7/6287639/6514899/09877800.pdf 12. Biased Backpressure Routing Using Link Features and Graph Neural Networks - IEEE Xplore, https://ieeexplore.ieee.org/iel8/9882533/10356147/10681132.pdf 13. Jamming Detection in Cell-Free MIMO with Dynamic Graphs - arXiv, https://arxiv.org/html/2601.06075v1 14. Graph-aware isomorphic attention for adaptive dynamics in transformers - AIP Publishing,

https://pubs.aip.org/aip/aml/article/3/2/026108/3344859/Graph-aware-isomorphic-attention-for-a daptive 15. ABT-MPNN: an atom-bond transformer-based message-passing neural network for molecular property prediction - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC9968697/ 16. Exploring the benefits of Graph Transformers in Relational Deep Learning, https://repository.tudelft.nl/file/File_dc85ab72-c58e-461c-a995-f3f1cba48f7f 17. AMAG: Additive, Multiplicative and Adaptive Graph Neural Network For Forecasting Neural Activity - NeurIPS, https://proceedings.neurips.cc/paper_files/paper/2023/file/1c70ba3591d0694a535089e1c25888 d7-Paper-Conference.pdf 18. A Generalization of ViT/MLP-Mixer to Graphs - OpenReview, https://openreview.net/pdf?id=l7yTbEWuOQ 19. HFGNN: Efficient Graph Neural Networks Using Hub-Fringe Structures, https://edisonchan-szu.github.io/publication/HFGNN.pdf 20. A Scalable and Effective Alternative to Graph Transformers - AAAI Publications, https://ojs.aaai.org/index.php/AAAI/article/view/34231/36386 21. DeepGANTT: A Scalable Deep Learning Scheduler for Backscatter Networks, https://www.researchgate.net/publication/357366509_DeepGANTT_A_Scalable_Deep_Learnin g_Scheduler_for_Backscatter_Networks 22. Adaptive Multi-Agent Reinforcement Learning with Graph Neural Networks for Dynamic Optimization in Sports Buildings - MDPI, https://www.mdpi.com/2075-5309/15/14/2554 23. Graph-Based Neural Networks' Framework Using Microcontrollers for Energy-Efficient Traffic Forecasting - MDPI, https://www.mdpi.com/2076-3417/14/1/412 24. Hybrid Value Function Approximation for Solving the Technician Routing Problem with Stochastic Repair Requests | Transportation Science - PubsOnLine, https://pubsonline.informs.org/doi/10.1287/trsc.2022.0434 25. Do LLMs Offer a Robust Defense Mechanism Against Membership Inference Attacks on Graph Neural Networks? - MDPI, https://www.mdpi.com/2073-431X/14/10/414 26. ! (2022 IJCAI) A Survey On Graph Structure Learning - Progress and Opportunities | PDF, https://www.scribd.com/document/779297346/2022-IJCAI-A-Survey-on-Graph-Structure-Learnin g-Progress-and-Opportunities 27. Graph-based Tabular Deep Learning Should Learn Feature Interactions, Not Just Make Predictions - arXiv, https://arxiv.org/html/2510.04543v1 28. (PDF) Dynamic routing for efficient waste collection in resource constrained societies, https://www.researchgate.net/publication/368388380_Dynamic_routing_for_efficient_waste_coll ection_in_resource_constrained_societies 29. Incremental Strategy-based Residual Regression Networks for Node Localization in Wireless Sensor Networks - KoreaScience, https://www.koreascience.kr/article/JAKO202226461376786.page 30. Finding MNEMON: reviving memories of node embeddings - OpenBU, https://open.bu.edu/bitstream/2144/45998/2/2204.06963.pdf 31. DGTEN: A Robust Deep Gaussian based Graph Neural Network for Dynamic Trust Evaluation with Uncertainty-Quantification Support - arXiv, https://arxiv.org/html/2510.07620v1 32. Bridging Graph Neural Networks and Large Language Models: A Survey and Unified Perspective - Infoscience, https://infoscience.epfl.ch/server/api/core/bitstreams/7e6f86c2-04cf-4866-8d2e-5ad08272ca0b/c ontent 33. MSGformer: A Hybrid Multi-Scale Graph–Transformer Architecture for Unified Short- and Long-Term Financial Time Series Forecasting - MDPI, https://www.mdpi.com/2079-9292/14/12/2457 34. Pairwise Affinity Matrices in Feature Selection and the Emergence of Self-Attention - arXiv, https://arxiv.org/html/2507.14560v2 35. Graph Transformer Architecture, https://www.emergentmind.com/topics/graph-transformer 36. lamm-mit/converted_dataset_example · Datasets at Hugging Face, https://huggingface.co/datasets/lamm-mit/converted_dataset_example 37. Sharpness-aware Zeroth-order Optimization for Graph Transformers - IJCAI, https://www.ijcai.org/proceedings/2025/0348.pdf 38. The Origin of Self-Attention: From Pairwise

Affinity Matrices to Transformers - arXiv, https://arxiv.org/html/2507.14560v1 39. Dynamic Graph Combinatorial Optimization with Multi-Attention Deep Reinforcement Learning - Renata Borovica-Gajic, https://renata.borovica-gajic.com/data/2022_sigspatial_rl.pdf 40. [2511.09261] A Distributed Training Architecture For Combinatorial Optimization - arXiv, https://arxiv.org/abs/2511.09261 41. Rethinking Graph Transformer Architecture Design for Node Classification - arXiv, https://arxiv.org/html/2410.11189v1 42. A Scalable and Effective Alternative to Graph Transformers - arXiv, https://arxiv.org/pdf/2406.12059 43. SA-GNN: Prediction of material properties using graph neural network based on multi-head self-attention optimization - AIP Publishing, https://pubs.aip.org/aip/adv/article/14/5/055033/3295158/SA-GNN-Prediction-of-material-propert ies-using 44. AI-DRIVEN DECISION SUPPORT SYSTEM FOR OPTIMIZED MUNICIPAL SOLID WASTE MANAGEMENT: A MACHINE LEARNING AND ROUTE OPTIMIZATION FRAMEWORK FOR SMART CITIES - ResearchGate, https://www.researchgate.net/publication/399039158_AI-DRIVEN_DECISION_SUPPORT_SYS TEM_FOR_OPTIMIZED_MUNICIPAL_SOLID_WASTE_MANAGEMENT_A_MACHINE_LEARN ING_AND_ROUTE_OPTIMIZATION_FRAMEWORK_FOR_SMART_CITIES 45. (PDF) AI-IoT-graph synergy for smart waste management: a scalable framework for predictive, resilient, and sustainable urban systems - ResearchGate, https://www.researchgate.net/publication/396589928_AI-IoT-graph_synergy_for_smart_waste_ management_a_scalable_framework_for_predictive_resilient_and_sustainable_urban_systems 46. Learning Network Dismantling Without Handcrafted Inputs - arXiv, https://arxiv.org/html/2508.00706v2 47. DLMinTC+: A Deep Learning Based Algorithm for Minimum Timeline Cover on Temporal Graphs - MDPI, https://www.mdpi.com/1999-4893/18/2/113 48. Enhanced vehicle routing for medical waste management via hybrid deep reinforcement learning and optimization algorithms - Frontiers, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1496653/full 49. BEYOND CIRCUIT CONNECTIONS:ANON-MESSAGE PASSING GRAPH TRANSFORMER APPROACH TOWARDS QUANTUM ERROR MITIGATION, https://proceedings.iclr.cc/paper_files/paper/2025/file/c276c3303c0723c83a43b95a44a1fcbf-Pa per-Conference.pdf 50. A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics - arXiv, https://arxiv.org/abs/1703.02952