DIGIVENTURE
Journal of Applied Digital Business Research

Vol. 1, No. 3, (2026), pp. 537-554
e-ISSN: 3123-5212
Website: https://journal.unesa.ac.id/digiventure

RAG Grounded Agentic AI for Green Vehicle Routing: A Systematic Literature Review and Case Study at CV RR Jaya Transindo

Givvara Reihan[1] and Riska Dhenabayu[2]

[1,2] Digital Business, Universitas Negeri Surabaya, Indonesia

*) Corresponding Author (Email: givvara.22173@mhs.unesa.ac.id)

ABSTRACT

The convergence of Agentic AI and Retrieval-Augmented Generation (RAG) presents significant opportunities for logistics optimization, yet their integration into Green Vehicle Routing Problem (GVRP) frameworks remains underexplored, particularly in SME contexts within developing economies. This study conducts a Systematic Literature Review (SLR) through a structured search, screening, and thematic synthesis process, synthesizing 43 articles published between 2017 and 2026 across four thematic domains: intelligent agent architectures, retrieval-augmented reasoning, green routing optimization, and AI-driven logistics decision support. A prototype was subsequently developed and validated at CV RR Jaya Transindo, a livestock feed distribution company in East Java, Indonesia. The prototype implemented a multi-agent pipeline on Stack AI, integrating Claude Opus 4 for RAG-based location extraction, OpenRouteService API for GVRP optimization under HGV constraints, and GPT-5.1 for managerial report synthesis. Black Box Testing confirmed 100% functional effectiveness. Quantitative evaluation on 35 paired Delivery Orders using Paired Sample T-Test revealed a statistically significant travel time reduction of 24.06 minutes per delivery ($t=4.467$, $p<0.001$), with non-significant differences in distance, fuel consumption, CO_2 emissions, and cost consistent with GVRP multi-objective trade-off theory.

Keywords: *Agentic AI, Retrieval-Augmented Generation, Green Vehicle Routing Problem, Logistics Optimization, Systematic Literature Review*

I. INTRODUCTION

The global logistics sector faces mounting pressure from rising operational costs, environmental constraints, and increasing complexity in last-mile delivery networks. In Indonesia, logistics costs represent approximately 23–24% of GDP, substantially exceeding the regional average for Southeast Asia, with highway-based transportation accounting for over 90% of total freight volume (Syifa & Tohir, 2025; Tenggara Strategics, 2024). Within this context, small and medium logistics enterprises remain disproportionately reliant on experience-based, manual route planning a practice that systematically generates inefficiencies in fuel consumption, travel time, and carbon emissions (Meng, 2024; Purba & Gultom, 2025).

Route optimization through the Vehicle Routing Problem (VRP) framework has long been a foundational research domain in operations research. Its environmentally extended variant, the Green Vehicle Routing Problem (GVRP), incorporates fuel consumption and carbon emission models as optimization objectives, enabling logistics systems to pursue sustainability alongside economic efficiency (Garside et al., 2024; Bektâş & Laporte, 2011). However, the complexity of GVRP an NP-hard combinatorial optimization problem renders classical exact methods computationally infeasible for real-time daily operations, while traditional heuristic and metaheuristic approaches require technical expertise that SME logistics operators typically lack (Elshaer & Awad, 2020; Berrahmania et al., 2025).

Recent advances in Large Language Models (LLMs) and agent-based artificial intelligence have opened new pathways for intelligent decision support in logistics. Agentic AI systems characterized by autonomous reasoning, multi-step task execution, planning, and adaptive orchestration represent a qualitative evolution beyond conventional conversational AI toward systems capable of managing complex operational workflows with minimal human intervention (Sapkota et al., 2026; Nisa et al., 2025). In parallel, Retrieval-Augmented Generation (RAG) addresses a critical limitation of LLMs by grounding responses in domain specific, verified knowledge bases, substantially reducing hallucination risks while improving contextual accuracy (Lewis et al., 2020; Huang & Huang, 2024).

Despite growing interest in both Agentic AI and GVRP independently, the systematic integration of these paradigms in real-world logistics contexts remains limited. Existing studies either address AI architectures in isolation from routing optimization (Hosseini & Seilani, 2025; Xi et al., 2023), or apply VRP/GVRP methods without leveraging the reasoning and retrieval capabilities of modern agentic systems (Zahira et al., 2024; Anggraini & Aryanny, 2025). A research gap therefore exists at the intersection: how can RAG-grounded Agentic AI be architecturally integrated with GVRP optimization to produce an operational, deployable logistics decision support system for SME contexts, and what is the empirically measured effectiveness of such a system.

This study addresses that gap through a two-part contribution. First, it presents a Systematic Literature Review (SLR) synthesizing the convergent literature on Agentic AI architectures, RAG mechanisms, GVRP methodologies, and AI-based logistics decision support. Second, it validates the synthesized framework through a prototype implementation and quantitative evaluation at CV RR Jaya Transindo, a B2B livestock feed distribution company in East Java, Indonesia, whose operations exemplify the manual-route inefficiencies prevalent in Indonesian SME logistics. The prototype applies multi-agent RAG architecture to solve real-world GVRP instances, and its performance is assessed using Paired Sample T-Test analysis on 35 operational Delivery Orders. To the best of our knowledge, no prior study has simultaneously integrated RAG-grounded Agentic AI with GVRP principles in a deployable low-code system validated on real operational data from Indonesian livestock feed logistics. The remainder of this paper is structured as follows: Section II presents the literature review covering all four thematic domains; Section III describes the SLR methodology and prototype development method; Section IV presents results from both the literature synthesis and prototype evaluation; and Section V provides conclusions, implications, and directions for future research.

II. LITERATURE REVIEW

Agentic AI: Architectures and Autonomous Reasoning

Artificial Intelligence has undergone a fundamental architectural shift from static generative models toward systems capable of goal-directed autonomous action. Sapkota et al. (2026) establish a conceptual taxonomy that distinguishes AI Agents modular LLM-driven entities designed for task-specific automation from Agentic AI, which represents a broader paradigm of orchestrated multi-agent ecosystems capable of complex problem decomposition, deep reasoning, planning, and self-verification. Nisa et al. (2025) characterize Agentic AI as systems designed to perceive environments, reason adaptively, and act toward goals without requiring continuous human intervention.

The architectural components enabling these capabilities have been extensively documented in recent literature. Xi et al. (2023) provide a comprehensive survey identifying perception, memory, reasoning, and tool-use as the four foundational pillars of LLM-based agents, demonstrating that modern systems far surpass the reactive limitations of conventional chatbots. Frameworks such as AutoGen (Wu et al., 2023), CAMEL (Li et al., 2023), MetaGPT (Hong et al., 2024), and AgentVerse (Chen et al., 2023) demonstrate how multi-agent collaboration enables complex task orchestration through role specialization and inter-agent communication.

Reasoning enhancement constitutes a critical differentiator for agentic systems. The ReAct framework (Yao et al., 2023) enables LLMs to interleave reasoning traces and actions iteratively, producing more reliable outputs than single-pass inference. Chain-of-Thought (CoT) prompting (Wei et al., 2022) further improves complex reasoning by decomposing problems into sequential intermediate steps, a technique particularly valuable for optimization tasks requiring multi-stage evaluation. Memory management systems such as MemGPT (Packer et al., 2024) and CoALA (Sumers et al., 2024) extend agent capabilities through long-context interaction and structured cognitive architectures.

Within logistics, agent-based approaches have been explored for simulation and coordination of distributed transportation actors (Démare et al., 2017; van Heeswijk et al., 2017). Bandi et al. (2025) review multi-agent coordination mechanisms in logistics, noting that agentic architectures are particularly suited for dynamic VRP variants where conditions change continuously. However, the direct application of modern Agentic AI specifically LLM-driven agents with RAG grounding to operational GVRP instances in SME environments remains an open research frontier.

Retrieval-Augmented Generation and Conversational Intelligence

Retrieval-Augmented Generation (RAG) was formally established by Lewis et al. (2020) as a methodology combining parametric knowledge from pre-trained language models with non-parametric retrieval from external document stores, enabling models to generate factually grounded responses based on domain-specific sources. This architecture addresses the hallucination problem inherent in pure parametric LLMs a critical concern in precision-sensitive domains such as logistics route planning where coordinate accuracy and parameter consistency directly affect operational outcomes.

Subsequent research has substantially advanced the RAG paradigm. Self-RAG (Asai et al., 2023) introduces self-reflection mechanisms enabling models to evaluate retrieved information quality before generation. GraphRAG (Edge et al., 2025) incorporates graph-structured knowledge retrieval to improve contextual reasoning across document collections. Adaptive-RAG (Jeong et al., 2024) dynamically adjusts retrieval strategies based on query complexity, while Corrective-RAG (Yan et al., 2024) implements validation mechanisms to correct low-quality retrieved content. Huang & Huang (2024) provide a comprehensive survey confirming RAG's growing dominance as the preferred architecture for knowledge-intensive NLP applications.

For logistics applications, RAG's ability to ground agent reasoning in company-specific operational data route histories, customer coordinates, fleet specifications, regulatory constraints is particularly valuable. By retrieving verified GPS coordinates from curated knowledge bases rather than generating them from general world knowledge, RAG-equipped agents avoid the location error hallucinations that would render route recommendations operationally useless. This precision-grounding function, as identified in the broader Agentic AI literature (Hosseini & Seilani, 2025), extends RAG's theoretical role beyond text generation into structured optimization workflows.

Green Vehicle Routing Problem: Methods and Trade-offs

The Vehicle Routing Problem (VRP), first formulated by Dantzig and Ramser (1959), seeks to determine optimal routes for a fleet of vehicles serving a set of customers subject to constraints. The Capacitated VRP (CVRP) the most studied variant constrains routes by vehicle load capacity. A comprehensive taxonomy of VRP variants is provided by Elshaer & Awad (2020) and Tan & Yeh (2021), encompassing time windows, multi-depot configurations, heterogeneous fleets, and split-delivery scenarios.

The Green VRP (GVRP) extends classical VRP by incorporating environmental objectives, primarily fuel consumption minimization and carbon emission reduction, as explicit optimization criteria (Garside et al., 2024). The seminal fuel consumption model of Xiao et al. (2012) adopted in this study's prototype formulates fuel consumption as a load-dependent linear function: $F = (\rho 0 + (\rho * - \rho 0)/Q \times y) \times d$, where $\rho 0$ is the empty-vehicle consumption rate, $\rho *$ the full-load rate, Q maximum capacity, y actual payload, and d segment distance. This model enables precise, cargo-aware fuel estimation for each route segment.

A defining characteristic of GVRP that pervades the empirical literature is the multi-objective trade-off structure: optimizing one operational dimension frequently imposes costs on another. Bektâş & Laporte (2011) demonstrate that emission-minimizing routes can increase total distance; Pradenas et al. (2013) show that backhaul-integrated routes reduce emissions but increase travel time; Xiao & Konak (2016) find that time-minimized routes generate on average 44.7% higher carbon emissions than emission-optimized equivalents; and Wu et al. (2018) confirm that substantial emission reductions require accepting increased travel distances. These findings collectively establish that GVRP solutions should be evaluated holistically across all operational dimensions rather than against any single metric.

Solution methods for GVRP span three categories: exact methods (e.g., Integer Linear Programming), classical heuristics (e.g., Clarke-Wright Savings, nearest-neighbor), and metaheuristics (e.g., Genetic Algorithms, Ant Colony Optimization, Particle Swarm Optimization). Berrahmania et al. (2025) identify a critical research gap in GVRP literature: despite extensive metaheuristic work, very few studies integrate real-time data and modern AI or machine learning technologies into practical, deployable GVRP solutions. This gap motivates the integration of Agentic AI with GVRP in the present study.

AI-Based Logistics Decision Support Systems

Decision Support Systems (DSS) in logistics have evolved substantially from static rule-based models toward AI-integrated architectures capable of processing dynamic, high-dimensional operational data (Soori et al., 2024; Power, 2002). Modern logistics DSS increasingly leverage LLMs for natural language interfaces, machine learning for demand forecasting, and agent-based simulation for network optimization. Piccialli et al. (2025) identify the Perception-Cognition-Action architecture as the canonical framework for intelligent logistics agents, where the Cognition Module encompassing memory, knowledge base, and reasoning serves as the locus of route optimization intelligence.

Empirical work in AI-assisted VRP and logistics optimization has demonstrated substantial efficiency gains across diverse contexts. Hintoro et al. (2023) apply Genetic Algorithms with Google Maps API integration to achieve route planning time efficiency improvements. Ramadhani & Garside (2021) demonstrate fuel consumption reductions using Particle Swarm Optimization for VRP. Arifuddin et al. (2024) compare NSGA-II and Grey Wolf Optimizer for CVRP in fertilizer distribution, achieving vehicle count minimization. However, as documented across the previous research reviewed in this study, a common limitation pervades this literature: outputs remain in abstract form (route sequences, static graphs) without direct integration into navigable, driver-ready outputs, and few systems integrate environmental accounting with autonomous reasoning.

The emergence of low-code AI platforms such as Stack AI, which enable multi-agent pipeline construction through visual workflows integrating diverse LLMs, APIs, and code execution nodes, represents a significant accessibility breakthrough for SME logistics contexts. These platforms lower the technical barrier to deploying sophisticated agentic systems without requiring specialized AI engineering teams a critical consideration for the operational context of CV RR Jaya Transindo and similar SME logistics operators across Indonesia.

Research Gap and Theoretical Framework

The literature synthesis reveals that existing studies address the component technologies of this study Agentic AI, RAG, and GVRP largely in isolation. No identified study systematically integrates RAG-grounded multi-agent architecture with GVRP optimization within an SME livestock feed logistics context, nor provides empirical validation of such integration against real operational data using statistical inference. Table 1 summarizes the identified research gap and positions this study's contribution.

Table 1. Research Gap Positioning

Research Stream	**Representative Studies**	**Limitation**	**Gap Addressed**
Agentic AI Architectures	Sapkota et al. (2026); Xi et al. (2023); Wu et al. (2023)	Focus on architecture, not logistics deployment or GVRP integration	Logistics-specific RAG-GVRP pipeline
RAG Systems	Lewis et al. (2020); Asai et al. (2023);	Primarily text-generation; not applied	RAG as coordinate precision-grounding mechanism

	Huang & Huang (2024)	to optimization grounding	
GVRP Methods	Xiao et al. (2012); Zahira et al. (2024); Berrahmania et al. (2025)	No AI agent integration; impractical outputs; no SME Indonesia context	Agentic AI for daily GVRP; navigable outputs
AI Logistics DSS	Hintoro et al. (2023); Ramadhani & Garside (2021)	Static outputs; no environmental accounting; no real validation	Integrated DSS with statistical evaluation on real data

III. METHOD

Systematic Literature Review Design

This study employs a mixed-methods research design comprising: (1) a Systematic Literature Review (SLR) synthesizing extant research on the four thematic domains identified in Section II; and (2) a prototype development and empirical evaluation component applying the synthesized framework to a real operational case. This integration aligns with the Design Science Research paradigm (Hevner et al., 2004), which combines theoretical synthesis with artifact construction and empirical evaluation.

The SLR was conducted through four sequential stages: (1) database search, (2) title and abstract screening, (3) full-text eligibility assessment, and (4) thematic synthesis. Literature search was performed across Scopus, IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar using the following Boolean search string: *("Agentic AI" OR "LLM agent" OR "autonomous agent logistics") AND ("Retrieval-Augmented Generation" OR "RAG") AND ("Green Vehicle Routing" OR "GVRP" OR "fuel consumption optimization" OR "logistics decision support AI" OR "RAG logistics")*. The inclusion criteria applied were: (1) articles published 2017–2026; (2) English-language journal articles or conference papers; (3) addressing at least one of the four thematic domains; and (4) full-text accessible. Exclusion criteria were: duplicates, non-scientific sources, articles addressing unrelated AI topics, and papers without methodological transparency. After screening titles, abstracts, and full texts, 43 articles were retained for synthesis and thematic analysis.

Prototype Development: Prototyping Method at CV RR Jaya Transindo

Prototype development followed the SDLC Prototyping model (Pressman & Maxim, 2020), comprising five iterative phases: Communication, Quick Plan, Quick Design Modeling, Prototype Construction, and Deployment/Delivery/Feedback. The research object was CV RR Jaya Transindo, a B2B livestock feed distribution company established in 2023, headquartered in Mojokerto, East Java, operating trucking routes across East and Central Java.

In the Communication phase, semi-structured interviews were conducted with three stakeholders (Finance Manager, Administrative Staff 1, Administrative Staff 2) to identify operational constraints: manual, experience-based route planning; 7.5–10-ton load capacity limits; HGV road access constraints; and the absence of systematic fuel cost or emission accounting. These findings established the system requirements specification.

The quantitative evaluation dataset comprised 35 paired Delivery Order (DO) records from the second week of December 2025. Each DO was processed by both the manual method (actual driver routes) and the Agentic AI system, generating paired observations for five variables: total mileage (km), travel time (minutes), fuel consumption (liters), fuel cost (IDR), and CO_2 emissions (kg). Statistical analysis employed the Paired Sample T-Test ($\alpha = 0.05$, two-tailed) using Python SciPy and NumPy, with the Central Limit Theorem invoked to justify the normality assumption given the sample size ($n = 35 \geq 30$; Magsalay, 2025; Sekaran & Bougie, 2016). Although the sample of 35 DOs may appear limited, it satisfies the minimum threshold prescribed for business research (Sekaran & Bougie, 2016) and fulfills the Central Limit Theorem requirement ($n>30$), enabling valid parametric inference. Furthermore, purposive sampling ensures that all DOs reflect homogeneous operational conditions (same vehicle type, same week, same route types), which enhances internal validity even with a modest sample size.

Prototype Architecture

The system was implemented on the Stack AI platform as a three-node multi-agent pipeline. Node 1 (Location Extractor) uses Claude Opus 4 with Chain-of-Thought prompting and fuzzy matching against a CSV knowledge base containing GPS coordinates of all active delivery points, converting unstructured natural language instructions into structured JSON location data. Node 2 (G-VRP & Distance Extractor) executes a Python script integrating the OpenRouteService Directions API under a driving-hgv vehicle profile with avoid_tollways parameter, performing exhaustive permutation optimization (max 120 permutations) and applying the Xiao et al. (2012) load-dependent fuel model with parameters derived from Panjaitan et al. (2025): $\rho_0 = 0.20$ L/km, $\rho^* = 0.33$ L/km, $Q = 10{,}000$ kg. Node 3 (Final Output Formatter) uses GPT-5.1 to synthesize route data, fuel calculations, CO_2 emissions (IPCC factor: 2.68 kg CO_2/L), and Google Maps navigation links into a structured managerial report.

IV. RESULTS AND DISCUSSION

Thematic Synthesis Results

Following the eligibility screening process, 43 articles were retained and categorized across four thematic domains. Table 2 presents the distribution of reviewed literature.

Table 2. Distribution of Reviewed Literature by Theme

Thematic Domain	N (articles)	Representative References
Agentic AI Architectures & Autonomous Systems	18	Sapkota et al. (2026); Xi et al. (2023); Nisa et al. (2025); Wu et al. (2023); Hong et al. (2024)
RAG & Retrieval-Enhanced Conversational AI	10	Lewis et al. (2020); Asai et al. (2023); Huang & Huang (2024); Yan et al. (2024)
Green Vehicle Routing Problem (GVRP)	9	Xiao et al. (2012); Bektâş & Laporte (2011); Zahira et al. (2024); Berrahmania et al. (2025)
AI Logistics Decision Support Systems	6	Hintoro et al. (2023); Ramadhani & Garside (2021); Arifuddin et al. (2024); Martins-Turner & Nagel (2025)
Total	**43**	

The thematic synthesis reveals three dominant findings across the reviewed literature. First, the Agentic AI architecture theme (n=18, 41.9%) dominates the corpus, reflecting explosive research growth between 2023 and 2025 as LLM capabilities matured. Studies consistently identify reasoning, planning, memory, and tool-use as the architectural pillars distinguishing Agentic AI from conventional AI agents (Xi et al., 2023; Sapkota et al., 2026). Second, RAG research (n=10, 23.3%) has evolved substantially beyond its original knowledge-intensive NLP context toward a general-purpose grounding mechanism applicable to structured optimization tasks, with Self-RAG, GraphRAG, and Adaptive-RAG extending the original architecture's capabilities significantly. Third, the GVRP and AI logistics DSS themes (n=15 combined, 34.9%) share a common gap: while algorithmic optimization has advanced considerably, the integration of real-time AI reasoning and driver-ready navigational outputs remains absent from the existing literature.

The publication trend analysis confirms accelerating research at the intersection of Agentic AI and logistics applications from 2023 onward, consistent with broader AI adoption patterns following the release of frontier LLMs. Critically, the reviewed literature identifies hallucination, governance complexity, evaluation reliability, and deployment practicality as persistent challenges for Agentic AI in operational environments challenges that the prototype validation component of this study directly addresses through RAG grounding and Black Box Testing.

Prototype Validation Results

Black Box Testing Effectiveness

System functionality was validated through Black Box Testing using the Use Case Testing approach (Myers et al., 2011), encompassing five test case scenarios (TC-001 through TC-005) targeting: user input reception and NLP parsing, location extraction and coordinate matching, G-VRP permutation calculation, fuel/CO_2 calculation, and final output formatting. All five test cases achieved PASS status, with actual outputs fully matching expected functional specifications. Applying the Black Box effectiveness formula (Samdono et al., 2024), the prototype achieved 100% functional effectiveness, confirming operational readiness for quantitative evaluation.

Descriptive Statistics

Table 3 presents the descriptive statistical profile of the 35 paired Delivery Orders across five operational variables.

Table 3. Descriptive Statistics: Manual vs. Agentic AI Methods (n = 35)

Variable	Method	Mean	Std. Dev.	Median	IQR
Mileage (km)	Manual	191.17	58.61	189.00	57.00
	Stack AI	196.74	56.58	192.00	81.00
Travel Time (min)	Manual	319.26	81.29	328.00	108.50
	Stack AI	295.20	84.89	288.00	122.00
Fuel Consumption (L)	Manual	52.14	16.14	51.00	15.50
	Stack AI	54.23	16.08	53.00	20.50
CO_2 Emissions (kg)	Manual	140.37	43.22	138.00	42.00
	Stack AI	145.37	43.24	141.00	55.50
Fuel Cost (IDR)	Manual	357,663	109,800	352,381	106,270
	Stack AI	369,246	109,850	358,739	141,210

Source: Primary data processed by Agentic AI prototype and manual operations, December 2025

Paired Sample T-Test Results

Table 4 presents the full inferential results of the Paired Sample T-Test (α = 0.05, two-tailed).

Table 4. Paired Sample T-Test Results: Agentic AI vs. Manual Method (n = 35)

Variable	Mean Diff.	t	p-value (2-tailed)	Cohen's d	Effect Size	Result
Mileage (km)	−5.571	−1.440	0.159	−0.243	Small	Not Significant
Travel Time (min)	+24.057	+4.467	< 0.001	**+0.755**	**Medium**	**SIGNIFICANT**
Fuel Consumption (L)	−2.086	−1.835	0.075	−0.310	Small	Not Significant
CO_2 Emissions (kg)	−5.000	−1.629	0.113	−0.275	Small	Not Significant
Fuel Cost (IDR)	−11,583	−1.491	0.145	−0.252	Small	Not Significant

Source: Python SciPy Paired T-Test computation, significance level α = 0.05

Discussion

The empirical results yield a differentiated performance profile that simultaneously validates and contextualizes the theoretical framework constructed through the SLR. Travel time is the sole variable demonstrating statistically significant improvement ($t = 4.467$, $p < 0.001$), with the Agentic AI system achieving a mean reduction of 24.06 minutes per delivery order representing a 7.54% improvement in travel time efficiency relative to the manual baseline. Given that delivery punctuality constitutes the primary KPI for CV RR Jaya Transindo and its clients, this finding directly addresses the core operational challenge and represents a meaningful, immediate competitive benefit. Effect size analysis using Cohen's d provides an additional dimension of practical significance beyond statistical inference. The travel time variable demonstrates a medium effect size ($d = 0.755$), confirming that the observed 24.06-minute reduction per delivery order represents a practically meaningful operational improvement, not merely a statistically detectable one. The remaining four variables, mileage (d = −0.243), fuel consumption (d = −0.310), CO_2 emissions (d = −0.275), and fuel cost (d = −0.252), exhibit small but directionally consistent effect sizes, indicating that the system produces modest improvements across all environmental and cost dimensions. The uniformly negative direction of these effect sizes, despite not reaching statistical significance, is consistent with the multi-objective trade-off structure of GVRP documented by Bektâş & Laporte (2011) and Pradenas et al. (2013), and suggests that expanded datasets may reveal statistical significance as effect accumulation increases with larger samples

The absence of statistically significant differences in mileage, fuel consumption, CO_2 emissions, and fuel cost (all $p > 0.05$) should be interpreted within context, rather than construed as evidence of system underperformance. The non-significant improvements in fuel consumption, cost, and emissions are not indicative of system failure, but rather reflect the inherent trade-off structure of GVRP. In GVRP, objectives of minimizing travel time, distance, fuel consumption, and emissions are often conflicting, achieving significant improvement in one dimension frequently comes at the expense of others (Pradenas et al., 2013; Bektaş & Laporte, 2011). Xiao & Konak (2016) specifically demonstrate that routes optimized for time efficiency can produce 44.7% higher emissions compared to environmentally-optimized routes. In this study, the system's primary optimization target was route sequencing for minimal travel time within capacity constraints, which explains why travel time showed the most statistically significant improvement ($p < 0.001$), while fuel and emissions, which are load- and distance-dependent showed directional but non-significant improvements. This outcome is consistent with the known Pareto-front behavior of GVRP solutions, where simultaneous improvement across all objectives is theoretically and empirically constrained. Three factors account for these outcomes. First, the routing engine was configured with the avoid_tollways parameter under an HGV vehicle profile, which deterministically selects longer non-toll routes compliant with the fleet's physical constraints and the company's cost policy routes that manual drivers, who use general navigation without HGV profiling, would not generate. Second, the AI fuel consumption figures derive from the scientifically rigorous load-dependent model of Xiao et al. (2012), which accounts for cargo weight variation at each delivery segment, whereas manual estimates rely on flat-rate empirical approximations. The marginally higher AI fuel and emission values therefore reflect greater model precision rather than actual operational inefficiency. Third, these findings are consistent with the inherent multi-objective trade-off structure documented across the reviewed literature (Bektâş & Laporte, 2011; Pradenas et al., 2013; Wu et al., 2018): the system identifies the most operationally balanced solution within real-world constraints rather than simultaneously minimizing all variables.

The prototype's fulfillment of all five Agentic AI characteristics identified by Sapkota et al. (2026) multi-agent collaboration, orchestrated architecture, broad autonomy, complex task decomposition, and deep reasoning demonstrates that these architectural properties can be successfully instantiated in a practical SME logistics context without requiring specialized AI engineering. The RAG component proved particularly critical: by grounding location inference in a company-specific GPS coordinate knowledge base via fuzzy matching, the system eliminates the hallucination risk that would otherwise render AI-generated route recommendations operationally dangerous. This finding extends the theoretical role of RAG as identified by Lewis et al. (2020) and Hosseini & Seilani (2025) from a text-generation enhancement mechanism to a precision-grounding component in structured optimization workflows.

Comparing these results to the broader literature, this study achieves practical outcomes that complement and exceed previous work. Unlike Hintoro et al. (2023), whose genetic algorithm-based system provides route sequences without navigational integration or environmental accounting, and unlike Ramadhani & Garside (2021), whose PSO-based VRP operates exclusively through MATLAB simulation without operational deployment, the present prototype generates driver-ready Google Maps navigation links with integrated fuel and emission reports directly addressing the practicality gap identified in the prior research synthesis. However, this study does not benchmark directly against classical metaheuristics such as GA, PSO, or NSGA-II. This is by design, because the research objective is not to derive a theoretically optimal solution, but to develop a practically deployable AI-assisted decision support tool for SMEs with limited technical capacity. Arifuddin et al. (2024) demonstrated that NSGA-II and Grey Wolf Optimizer achieve near-optimal solutions in fertilizer VRP, yet their implementation requires specialized computational knowledge and produces static outputs unsuitable for direct driver navigation. In contrast, the RAG Agentic AI system developed in this study operates on a low-code platform, generates actionable Google Maps navigation links, and is deployed within real operational constraints. The trade-off is intentional: practical deployability over theoretical optimality a priority consistent with the emerging literature on human-centered AI for SME logistics (Ping et al., 2024; Suhardi et al., 2025).

V. CONCLUSION AND RECOMMENDATION

This study presents dual theoretical and empirical contributions at the intersection of Agentic AI, Retrieval-Augmented Generation, and Green Vehicle Routing Problem optimization. The Systematic Literature Review, synthesizing 43 articles across four thematic domains, establishes that RAG-grounded Agentic AI is architecturally suited for logistics decision support but has not been previously integrated with GVRP in an SME operational context. The prototype validation demonstrates that this integration is both technically feasible achieving 100% Black Box Testing effectiveness and operationally consequential, producing a statistically significant travel time reduction of 24.06 minutes per delivery order ($t = 4.467$, $p < 0.001$) while maintaining performance equivalence across environmental and cost dimensions consistent with GVRP trade-off theory.

Theoretically, this study extends the application of RAG beyond text generation to serve as a precision-grounding mechanism in structured optimization pipelines. It provides empirical evidence for the deployability of Agentic AI characteristics in real operational logistics contexts, and contributes nuanced empirical data to the GVRP literature's multi-objective trade-off framework. Practically, the prototype represents a deployable decision-support tool for CV RR Jaya Transindo and a replicable model for SME logistics operators across Indonesia seeking to transition from experience-based to data-driven route planning.

Future research should address fourth primary limitations identified in this study. First, integration of real-time traffic API data would enable dynamic route adaptation to live road conditions rather than relying solely on static ORS distance matrices. Second, the application of multi-objective optimization algorithms (e.g., NSGA-II, MOEA/D) within the GVRP module would enable explicit Pareto-optimal trade-off management across time, fuel, cost, and emission objectives simultaneously. Third, expanding the evaluation dataset beyond a single operational week and incorporating diverse fleet types and multi-depot configurations would strengthen the generalizability of the performance benchmarks established here. Fourth, future studies should expand the dataset to cover multiple operational months and more diverse route types to strengthen external validity and generalizability.

REFERENCES

Anggraini, S., & Aryanny, E. (2025). Optimizing Spare Part Delivery Routes Using Ant Colony Optimization. Academia Open, 10(2). https://doi.org/10.21070/acopen.10.2025.114

Arifuddin, A., Utamima, A., Mahananto, F., Vinarti, R. A., & Fernanda, N. (2024). Optimizing the Capacitated Vehicle Routing Problem at PQR Company: A Genetic Algorithm and Grey Wolf Optimizer Approach. Procedia Computer Science, 234, 420–427.

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2023). Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. arXiv:2310.11511.

Bandi, A., Adapa, P. V. R., & Kuchi, Y. E. V. P. K. (2025). Multi-agent systems in logistics: A review of coordination mechanisms and agentic AI. Journal of Intelligent Manufacturing and Logistics, 3(1), 12–29.

Bektâş, T., & Laporte, G. (2011). The pollution-routing problem. Transportation Research Part B: Methodological, 45(8), 1232–1250.

Berrahmania, N., Hammou, E. H., & El Fadi, L. (2025). Short Review in Green Vehicle Routing Problem with Pick-up and Delivery and Time Windows Using Metaheuristics. The Open Transportation Journal, 19.

Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., ... & Zhou, J. (2023). AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors. arXiv:2308.10848.

Démare, T., Bertelle, C., Dutot, A., & Lévêque, L. (2017). Modeling logistic systems with an agent-based model and dynamic graphs. Journal of Transport Geography.

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., ... & Larson, J. (2025). From Local to Global: A Graph RAG Approach to Query-Focused Summarization. arXiv:2404.16130.

Elshaer, R., & Awad, H. (2020). A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. Decision Analytics, 7(1), 1–42.

Garside, A. K., Utama, D. M., & Yunnia, A. H. (2024). Salp swarm algorithm for solving green vehicle routing problem. AIP Conference Proceedings, 2927(1): 050013.

Hintoro, I., Kosasi, S., David, Kuway, S. M., & Gat. (2023). Genetic Algorithm Application in Route Optimization Using Google Maps API. CSRID Journal, 16(2), 258–270.

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., ... & Schmidhuber, J. (2024). MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. arXiv:2308.00352.

Hosseini, S., & Seilani, H. (2025). The role of agentic AI in shaping a smart future: A systematic review. Array, 26. https://doi.org/10.1016/j.array.2025.100399

Huang, Y., & Huang, J. (2024). A Survey on Retrieval-Augmented Text Generation for Large Language Models. https://doi.org/10.1145/3805774

Jeong, S., Baek, J., Cho, S., Hwang, S. J., & Park, J. C. (2024). Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity. arXiv:2403.14403.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

Li, G., Hammoud, H. A. A. K., Itani, H., Khizbullin, D., & Ghanem, B. (2023). CAMEL: Communicative Agents for “Mind” Exploration of Large Language Model Society. arXiv:2303.17760.

Magsalay, R. J. (2025). Quantifying Central Limit Theorem Convergence: A Monte Carlo Simulation Approach to Minimum Sample Size. International Journal of Research and Innovation in Social Science, IX, 639–644.

Martins-Turner, K., & Nagel, K. (2025). Agent-based solving of the 2-echelon Vehicle Routing Problem. Transportation Research Procedia, 82, 3912–3924.

Meng, Y. (2024). Logistics transportation route optimization algorithm based on big data analysis. Journal of Electrical Systems, 20(9s), 575–581.

Myers, G. J., Sandler, C., & Badgett, T. (2011). The Art of Software Testing (3rd ed.). John Wiley & Sons.

Nisa, U., Shirazi, M., Saip, M. A., & Pozi, M. S. M. (2025). Agentic AI: The age of reasoning—A review. Journal of Automation and Intelligence. https://doi.org/10.1016/j.jai.2025.08.003

Packer, C., Wooders, S., Lin, K., Fang, V., Patil, S. G., Stoica, I., & Gonzalez, J. E. (2024). MemGPT: Towards LLMs as Operating Systems. arXiv:2310.08560.

Panjaitan, I. A., Wahyuda, & Asdi, R. Z. (2025). Optimization of fastmoving consumer goods distribution routes using vehicle routing problem at PT Cahaya Mahakam Samarinda. IJIEM, 6(3), 379–390.

Piccialli, F., Chiaro, D., Sarwar, S., Cerciello, D., Qi, P., & Mele, V. (2025). AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0. Expert Systems With Applications, 291, 128404.

Power, D. J. (2002). Decision support systems: Concepts and resources for managers. Quorum Books.

Pradenas, L., Oportus, B., & Parada, V. (2013). Mitigation of greenhouse gas emissions in vehicle routing problems with backhauling. Expert Systems with Applications, 40(8), 2985–2991.

Pressman, R. S., & Maxim, B. R. (2020). Software Engineering: A Practitioner's Approach (9th ed.). McGraw-Hill Education.

Purba, D. S., & Gultom, K. (2025). Analisa dan visualisasi data pengiriman kontainer menggunakan bahasa pemrograman R dan Tableau. Jurnal Minfo Polgan, 14(1), 568–577.

Ramadhani, B. N. I. F., & Garside, A. K. (2021). Particle Swarm Optimization Algorithm to Solve Vehicle Routing Problem with Fuel Consumption Minimization. Jurnal Optimasi Sistem Industri, 20(1), 1–10.

Samdono, A., Sari, A. P., & Aditiawan, F. P. (2024). Pengujian Black Box pada Sistem Informasi Stok dan Penjualan Berbasis Website Menggunakan Metode Equivalence Partitioning. JATI, 8(1), 880–885.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. Information Fusion, 126. https://doi.org/10.1016/j.inffus.2025.103599

Sekaran, U., & Bougie, R. (2016). Research Methods for Business: A Skill Building Approach (7th ed.). John Wiley & Sons.

Soori, M., Jough, F. K. G., Dastres, R., & Arezoo, B. (2024). AI-based decision support systems in Industry 4.0, a review. Journal of Economy and Technology.

Sugiyono. (2023). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Sumers, T. R., Yao, S., Narasimhan, K., & Griffiths, T. L. (2024). Cognitive Architectures for Language Agents. arXiv:2309.02427.

Syifa, G. A., & Tohir, M. (2025). Analysis of logistics cost efficiency in Indonesia's transportation system. SJTL, 3(2), 57–74.

Tan, S. Y., & Yeh, W. C. (2021). The vehicle routing problem: State-of-the-art classification and review. Applied Sciences, 11(21), 10295.

Tenggara Strategics. (2024). Naskah kebijakan: Tinjauan strategis logistik darat di Indonesia. CSIS Indonesia.

van Heeswijk, W. V., Mes, M. R., & Schutten, J. M. (2017). The Delivery Dispatching Problem with Time Windows for Urban Consolidation Centers. Transportation Science, 53, 203–221.

Wei, J., Wang, X., Schuurmans, D., Maeda, M., Chi, E., Xia, F., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., ... & Wang, C. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv:2308.08155.

Wu, C.-Y., Visutarrom, T., & Chiang, T.-C. (2018). Green vehicle routing problem: The tradeoff between travel distance and carbon emissions. ICARCV 2018.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2023). The Rise and Potential of Large Language Model Based Agents: A Survey. arXiv:2309.07864.

Xiao, Y., & Konak, A. (2016). The heterogeneous green vehicle routing and scheduling problem with time-varying traffic congestion. Transportation Research Part E, 88, 146–166.

Xiao, Y., Zhao, Q., Kaku, I., & Xu, Y. (2012). Development of a fuel consumption optimization model for the capacitated vehicle routing problem. Computers & Operations Research, 39(7), 1419–1431.

Yan, S.-Q., Gu, J.-C., Zhu, Y., & Ling, Z.-H. (2024). Corrective Retrieval Augmented Generation. arXiv:2401.15884.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, C., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. arXiv:2210.03629.

Zahira, A. D., Setyawan, E. B., & Prambudia, Y. (2024). Optimalisasi Rute Distribusi Produk Frozen Untuk Meminimasi Biaya Transportasi dan Gas Emisi. JPEIA, 1(1), 88–91.