
| RESEARCH ARTICLE

Windows Are Strategy: Generative Artificial Intelligence for Time-Window Design in Big-and-Bulky Delivery

Manish Patel

Independent Researcher, USA

Corresponding Author: Manish Patel, **E-mail:** manish.jk.patel@gmail.com

| ABSTRACT

The design of delivery windows is a strategic gateway in the field of large-scale logistics, with poorly planned time slots creating systemic inefficiencies. The architecture is a generative-plus-optimization architecture that retasks windows, based on post hoc operational considerations, to strategic considerations. The interface joins an AI-based generative element, which produces the window menus and communications to the customers, with a solid operations-research interior, which finds the optimal solution to the issue of vehicle routing at scale. A pyramidal assessment system including replay harnesses, shadow arms, and controlled split validations ensures a long-term performance in terms of promise accuracy, route density, and environmental measures. The variants of operational deployment are geographic differentiation of urban and rural settings, seasonally versatile modifications, and tiers of service that incorporate window accuracy and installation offers. The governance approach, based on modern AI risk frameworks, provides fair results and compliance with the regulations. The effect of strategic window design in terms of the improvement of operational efficiency, environmental sustainability, and customer experience is empirically proven in the context of the specific domain of large-volume delivery.

| KEYWORDS

Time-Window Design, Generative Artificial Intelligence, Vehicle Routing, Big-And-Bulky Delivery, Last-Mile Optimization

| ARTICLE INFORMATION

ACCEPTED: 20 October 2025

PUBLISHED: 04 November 2025

DOI: 10.32996/jcsts.2025.7.11.19

1. Introduction

The high-volume delivery industry presents operational complexities that are not equal to the traditional parcel logistics, requiring specialisation of crew logistics, supporting services, installation practices, and the presence of customers. The scientific study of logistics, as we know it, has continuously pointed to the fact that larger-scale deliveries require unique solutions in terms of fleet configuration, load management methods, and time management, and, thus, completely transforms classic patterns of scheduling [7].

Traditionally, the delivery windows have been treated as tactical considerations that are introduced to the design of core networks, but not as strategic resources. The result of this subordinate framing has resulted in great inefficiencies within distribution networks; in cases where windows do not have a strategic and dynamic governance structure, operations are characterized by high levels of miles driven, lower promise accuracy, and high levels of unsuccessful attempts to deliver. The existing literature on the vehicle routing problem with time windows proves that time constraints are central to routing strategies, and the strategic potential is underutilised [5].

In addition to operational indicators, the window strategy has a significant impact on environmental sustainability, satisfaction, and financial effectiveness. Analytic reports by the World Economic Forum point out that under current operational conditions, last-mile delivery emissions might increase by 32% in the top 100 urban areas worldwide by 2030, a pattern that can be attributed to the increasing popularity of e-commerce alongside the stagnant window allocation processes. Organized actions

Copyright: © 2025 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

with the involvement of dynamic window design can instead minimize emissions by up to 30 percent and, at the same time, deliveries by 25 percent, aligning the environmental and economic goals [5].

It creates a reference framework which reconceptualises window design as a product option and not a routing constraint. This approach combines a context-sensitive generative artificial intelligence layer, which generates window menus and customer communications, on top of an operations-research basis, which solves the underlying vehicle routing problems. This type of integration promotes responsive window design that adapts to geographic factors, seasonal changes, and network capacity changes in real time [6].

Applications within the industry highlight the effectiveness of optimisation-based programs. A reported case study of the route-optimisation system of a large parcel delivery enterprise documented 100 million miles of reduction, as well as around 10 million gallons of fuel usage, and still maintained its service level and supported the growing volume of deliveries [6].

By making window design a strategic aspect and not a mere tactical consideration, the delivery operations are able to enhance the route density, assure accuracy, customer experience, mileage efficiency, and emissions reduction. This change is a paradigm shift in the way the idea of a delivery window is conceptualised/managed with implications across both day-to-day routing choices and long-term sustainability requirements in the highly-specialized field of large-volume logistics [5], [7].

2. Evolution of Delivery Window Optimization

The history of the delivery window optimization is a complex development of primitive practices to new sophisticated computer-based algorithms. The systems of early delivery utilized simple methods of slotting of appointments, with customer requests being fulfilled on the basis of mere availability checks and not in terms of geographical proximity or efficiency of routes. This produced discontinuous delivery schedules that led to inefficient routes and overuse of resources. The 1987 paradigm shift in Operations Research developed systematic methods to address vehicle routing issues subject to time window constraints, and suggested intuitive approaches that cleverly located customer visits, obeying capacity constraints and time constraints. It was shown that in computational experiments, such optimization-based methods were always better than the traditional booking methods in terms of distance traveled, vehicles needed, and quality of the schedule [1].

The mathematical model of the vehicle routing problem with time windows (VRPTW) put in place a mathematical model that is still used to base current delivery planning models. This model summarizes critical realities of operations such as fleet capacity, service duration, time window preferences, and the aim of ensuring the minimization of transportation costs with service commitments. According to the records in econometric research archives, classification of algorithms under this optimization paradigm has led to the development of such algorithmic categories as construction heuristics, improvement heuristics, and exact mathematical programming methods. The flexibility of the model has allowed practitioners to respond to progressively more subtle operational conditions by adding context-related constraints. Econometric analysis indicated that there was a considerable performance difference in problem instances that had varied temporal and geographical distributions, where appropriate solution methods have to be directed towards a particular delivery environment [2].

Although algorithmic innovations have increased scheduling possibilities, modern documentation of implementation accepts the existence of inherent challenges of optimization-only methods. The current routing libraries use several solution strategies, among them local search algorithms with time-specific operators and metaheuristic methods exploring solution spaces of complexity. According to these technical resources, time windows density and width have a strong impact on solution quality and computational requirements; tight windows have exponentially more complicating effects on the problems, and may require more vehicles altogether. Guides to implementation note that optimization is at its best at solving routing problems with a set of time windows, but the strategic planning of such time windows is a different problem that has a strong influence on routing performance. Best practices are increasingly suggesting incorporating optimization as a part of a larger decision-making model, which starts with strategically designing the window [3].

Era	Approach	Key Features
Early	Appointment Slotting	Simple availability checks
1987+	VRPTW Optimization	Heuristics, capacity constraints
Modern	Optimization Libraries	Local search, metaheuristics
Strategic	Proactive Design	Dynamic window generation

Table 1: Evolution of Delivery Window Optimization [1, 2, 3]

This awareness has led to a paradigm shift in delivery management, where window accommodation is a reaction to a situation and window design is a proactive, strategically planned reaction to an opportunity. With continuous re-optimization, advanced delivery systems have now dynamically produced window offerings based on the current route commitments, geographical clustering opportunities and capacity constraints, as the bookings keep on accumulating. This methodology allows delivery operations to achieve a trade-off between competing goals: providing convenient customer options and maintaining a rate of route density necessary to achieve operational and environmental efficiency [3].

3. Generative-Plus-Optimization Architecture

The time-window design architecture in production-grade is a big-and-bulky delivery, which incorporates specialized components that generate an efficient and reliable delivery schedule. It is a framework that integrates both generative artificial intelligence and optimization to tackle the complex problem of window design. Documentation on open-source routing shows that vehicle routing problems with time windows represent a specialized category of optimization problems that are best tackled using a specialized solution methodology. These are issues to do with minimum-cost paths of vehicle fleets to serve customers with particular time constraints, capacity limitations, and service needs- exactly the situation with big-and-bulky delivery [3].

The generative layer serves as a window menu design and customer communication intelligence center. This is a component that uses language models that have been trained using historical data on delivery to generate the right time windows and customer guidance. According to the optimization documentation, window design needs to be a compromise between accuracy and flexibility--it should be specific enough to allow customers to plan, but flexible enough to allow the operation to be efficient. The generative layer examines historical performance based on region, season, and category of product to suggest window structures that maximize operational density and customer convenience to adjust according to network conditions, current commitments, and predicted demand [3].

A formal body of knowledge is the starting point for working with rules and constraints that govern window design. Discussion points in the developer community emphasize the need to keep the parameter boundaries of the time window setup clear, documenting the need to make these windows too small, which results in highly constrained routing tasks that may require extra vehicles; and too large, which results in low efficiency with long wait times. The knowledge base sets operational parameters such as service duration estimates by product category, geographical travel time patterns, and capacity restrictions with vehicle types. Exchanges at the community level focus on keeping these parameters in an easily readable form instead of inputting them directly into the optimization models [4].

The optimization core solves the underlying vehicle routing problem, including applicable operational constraints. Technical description Techniques of the modern solutions use algorithms with specific objectives to tackle time-limited routing issues, such as, but not limited to, insertion heuristics, local search methods, and metaheuristics. The optimization engine analyzes the feasibility of routing of various window designs and gives feedback on the implications of proposed structures on operation, and balances between solution quality and the computation efficiency [3].

Cyclical window contraction is one of the most important patterns of work. In developer discussions, document implementation strategies that use broad windows that become narrower in the approach to the delivery date due to eventualities like achieving capacity level, or crossing planning horizon boundaries. This gradual enhancement balances sustaining the flexibility in early booking stages, with offering rising schedule assurance as the procedures near implementation [4].

Component	Function	Integration
Generative Layer	Window menu design	Proposes window structures
Knowledge Repository	Operating rules	Ground all components
Optimization Core	Route solution	Feasibility assessment
Dynamic Narrowing	Window refinement	Booking-execution bridge

Table 2: Generative-Plus-Optimization Architecture [3, 4, 10]

Patterns of integration integrate components using common interfaces that are consistent with regulatory frameworks. Legal interpretation of the EU Artificial Intelligence Act indicates the demands of the AI systems within the context of service delivery, such as the transparency demands, the necessity to control data, and human control measures. Such considerations influence

the integration requirements, including the extensive logging, explainability mechanisms, and the ability to review it by humans [10].

4. Evaluation Framework and Metrics

The extensive system of evaluation of the delivery operations based on the requirements of effective window design becomes the basis of the continuation of improvement and the reduction of the risk of implementation. This is an organized method that puts in place more than one layer of testing with an escalating exposure to operations. The analysis of the last-mile ecosystem interventions provided by the World Economic Forum defines that the successful optimization of the window has to be measured in terms of the three dimensions: the efficiency of its operations, the environmental impact, and the customer experience. The study refers to the delivery window strategies as the essential part of the larger interventions of the last-mile that can help not only achieve the goal of sustainability but also reduce costs when applied with other interventions, such as vehicle electrification and load pooling [5].

The replay harness is the first evaluation layer that evaluates the design of candidate windows and compares it with historical demand patterns without exposure to customers. This is a simulated mechanism that operates on archived order data by running the data through proposed window designs to evaluate the variants of the designs under consistent conditions. According to the World Economic Forum, there is a need to focus on forward-looking simulation strategies that can be used to consider the future increase in demand for urban delivery. By doing so, organizations can determine the extent to which various window structures would have performed versus historical order concerning metrics such as route density and promise reliability, finding the best structure to use in certain regions and in specific time frames [5].

The second level of evaluation is a shadow arm implementation, which applies the presented designs to real-world settings of production and is not seen by customers. This method works on parallel processing of orders that come in with current and candidate systems and compares them directly under equal working conditions. The AI Risk Management Framework of the National Institute of Standards and Technology also reveals that shadow implementation is one of the essential risk mitigation measures addressing service delivery systems. The framework highlights that the testing period must include a range of operating conditions, such as seasonal changes and abnormal demand conditions, so that issues that could create problems in the delivery of customer experience can be identified before they are experienced [8].

Controlled split validation is the last stage of evaluation that exposes a low percentage of customer traffic to new designs with pre-established rollback thresholds. This stepwise approach is advised in the NIST framework, which focuses on a gradual implementation process with clear measurement criteria and pre-set triggering events of interventions. The organisations are supposed to set specific metrics that are monitored, baseline performance, acceptable range, and automation alerting in case the metrics are near thresholds. This systematic methodology has been used to control deployment risks along with certifying performance in the real customer interactions [8].

The key measures run across various dimensions that reflect the overall effect of the window design strategies, promise accuracy measures, delivery completion within promised windows, route density measures, the amount of stops in the distance covered, missed handover measures, failed delivery caused by customer unavailability, and environmental impact measures, including distance covered by the vehicle and the nature of the vehicle. In their analysis, the authors of the Forum highlight the importance of the consideration of both operational and wider ecosystem effects when an evaluation of innovations in delivery is to be made [5].

Layer	Purpose	Implementation
Replay Harness	Historical testing	Simulates with archived data
Shadow Arm	Production testing	Parallel processing
Controlled Split	Customer exposure	Gradual with rollbacks
Key Metrics	Performance tracking	Promise accuracy, density

Table 3: Evaluation Framework [5, 8]

A comparative study on the use of the static and adaptive co-designed windows has shown that the traditional methods create demand imbalance over time slots, which reduces the density of the route, and the adaptive methods dynamically fill the window availability according to the network capacity and the available commitment during the delivery day [5].

5. Applications and Implementation Strategies

Locally specific playbooks of rural and urban delivery environments are a model implementation method of window design systems. The set of recommendations of the National Institute of Standards and Technology regarding the AI Risk Management Framework highlights the importance of providing mapping capabilities in AI systems implemented under different geographical conditions that reflect unique operating peculiarities. The framework emphasizes contextual analysis at the mapping stage, which makes organizations analyze the performance of AI systems in various deployment contexts, such as population density differences. This context-based design is in line with governance recommendations highlighting technical and social-technical aspects of AI implementation. The way systems can change in response to different operational settings should be recorded by the organizations, so that there is consistency in performance metrics, whether in a geographical deployment situation or not, and at the same time maintain a fair level of service provision [8].

Peak-season slotting strategies are strategies that overcome the challenges posed by high-volume seasons such as holidays, promotional events, and seasonal demand surges. Analysis of emerging agentic artificial intelligence capabilities in the industry is plagued by the fact that such systems are becoming more and more capable of handling a complex scheduling situation without human intervention. Studies have shown that agentic scheduling systems will change up to simple automation and transform to advanced decision-making systems like predictive rescheduling, proactive capacity allocation, and exception handling. Such development will eliminate the manual intervention at peak periods of demand when the conventional system normally demands a lot of human control. It has been observed that scheduling agents would take into consideration historical performance statistics as well as real-time operational indicators in the dynamically adjusting of delivery windows under high-demand periods, which may also redefine the way operations can manage seasonal changes in capacity [9].

Tiers of service levels that combine window accuracy with installation services and accessories offer opportunities to strategically differentiate and to optimize resource allocation. The NIST framework is able to define service differentiation as an implementation area that needs to be governed carefully to bring about fair results. Organizations are advised to keep a full-fledged tracking of results in the different types of customers when the AI systems are involved in service availability or premium accessibility. This strategy is consistent with the principles of trustworthiness, especially the consideration of fairness, which promotes the consideration of whether AI-enabled services form a different treatment of different groups of the population. It is also implemented that organizations make documentation of design decisions on service differentiation, keep measurement protocols on equity in service allocation, and put in place review processes on exceptional cases [8].

Regulation in accordance with the established AI risk models is a critical implementation factor. The European Union Artificial Intelligence Act provides legal analysis that covers elements that are applicable to delivery scheduling applications. The systems that affect the availability of the services might have certain compliance standards associated with transparency, data management, and human control. The Act resolves the case of information asymmetry generated by AI systems between service providers and service recipients by creating transparency conditions that ensure that customers are aware of how the algorithmic systems impact the service options provided to them [10].

The considerations for the implementation roadmap relate to the gradual deployment strategies of the complex AI systems. The NIST framework would provide a step-by-step approach that focuses on the constant measurement process through the deployment lifecycle, suggesting a thorough mapping of contextual factors, defining baseline performance metrics, deploying proper monitoring capabilities, and continually comparing the performance results with objectives [8].

Strategy	Focus	Benefits
Regional Playbooks	Urban/rural differences	Context-specific solutions
Peak-Season Slotting	High-volume periods	Reduced manual intervention
Service-Level Tiers	Window-service bundling	Strategic differentiation
Governance Controls	Regulatory compliance	Ethical implementation

Table 4: Implementation Strategies [8, 9, 10]

6. Conclusion

Reconfiguring delivery windows not as a tactical limitation to strategic resources is a paradigm shift in the time and customer-experience practice of the voluminous logistics operations. With the combination of generative AI and optimisation models, organisations can simultaneously achieve two seemingly incompatible goals of providing convenience to customers and maintaining operational efficiency and environmental stewardship. This strategy uses windows as design considerations and not

mere enhancements to routing as an afterthought, therefore, allowing it to respond dynamically to geographic, seasonal, and network-capacity variations. The architecture proposed provides an extensive framework extending to the window design, knowledge management, optimisation, and evaluation with governance controls that are in line with recognised risk frameworks. With increasing pressure on the delivery networks to reduce emissions without failing to fulfill rising expectations of customers, strategic window design is becoming an important tool to achieve the economic, environmental, and experiential objectives. The next step in work should be a continuous improvement of window strategies by systematic experimentation, evidence-based implementation, and gradual development into autonomous systems with dynamic optimization over the entire delivery ecosystem.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Marius M. Solomon, "Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints," INFORMS, 2010. [Online]. Available: https://www.iro.umontreal.ca/~difft6751/paper_solomon.pdf
- [2] Marius M. Solomon, "Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints," EconPapers, 1987. [Online]. Available: https://econpapers.repec.org/article/inmoropre/v_3a35_3ay_3a1987_3ai_3a2_3ap_3a254-265.htm
- [3] Google OR-Tools, "Vehicle Routing Problem with Time Windows". [Online]. Available: <https://developers.google.com/optimization/routing/vrptw>
- [4] GitHub, "Vehicle Routing Problem with Time Windows change example," 2022. [Online]. Available: <https://github.com/google/or-tools/discussions/3385>
- [5] World Economic Forum, "The Future of the Last-Mile Ecosystem," 2020. [Online]. Available: https://www3.weforum.org/docs/WEF_Future_of_the_last_mile_ecosystem.pdf
- [6] BSR, "Looking Under the Hood: ORION Technology Adoption at UPS," 2016. [Online]. Available: <https://www.bsr.org/en/case-studies/center-for-technology-and-sustainability-orion-technology-ups>
- [7] Bahram Alidaee et al., "The Last-Mile Delivery of Heavy, Bulky, Oversized Products: Literature Review and Research Agenda," MDPI, 2023. [Online]. Available: <https://www.mdpi.com/2305-6290/7/4/98>
- [8] NIST, "AI Risk Management Framework". [Online]. Available: <https://www.nist.gov/itl/ai-risk-management-framework>
- [9] STAMFORD, "Gartner Predicts Agentic AI Will Autonomously Resolve 80% of Common Customer Service Issues Without Human Intervention by 2029," Gartner, 2025. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/2025-03-05-gartner-predicts-agentic-ai-will-autonomously-resolve-80-percent-of-common-customer-service-issues-without-human-intervention-by-20290>
- [10] Clifford Chance, "The European Union Artificial Intelligence Act: Overview of key rules and requirements," 2024. [Online]. Available: <https://www.cliffordchance.com/content/dam/cliffordchance/PDFDocuments/the-eu-ai-act-overview.pdf>