SmartTransit NYC:

An AI-Powered Approach to Subway Reliability and Urban Sustainability

Final Reflective Research Paper

Alexander Escobar

New York University Tandon School of Engineering

Computer Engineering

April 2025

Abstract

New York City's subway system, one of the largest and most complex transit networks in the world, serves approximately 3.4 million daily riders yet continues to struggle with chronic reliability deficits and information asymmetries that undermine its effectiveness as a sustainable transportation solution. Despite an on-time performance improvement to "82.2%" in 2024, riders still experienced over 486,000 delayed trains throughout the year, and a recent fare increase to $3.00 in January 2026 has intensified public scrutiny of service quality. This paper presents SmartTransit NYC, a conceptually designed artificial intelligence-powered prediction system that addresses the critical information gap between raw transit data and rider decision-making. By integrating real-time MTA data feeds, historical delay patterns, weather variables, and time-based usage signals, the system would generate probabilistic outputs across three dimensions: delay likelihood, crowding forecasts, and optimized alternative routing.

Through a multi-dimensional feasibility analysis, this paper evaluates the technical, economic, social, and environmental viability of such a system. Machine learning architectures including ensemble Random Forest classifiers and Long Short-Term Memory (LSTM) networks are proposed as the predictive core, targeting 75–85% accuracy for delay prediction and 70–80% accuracy for crowding estimation, benchmarked against comparable systems deployed in London, Singapore, and Tokyo. Economic modeling indicates development costs in the range of $50,000–$100,000 with monthly operational expenses of $500–$2,000, suggesting moderate financial feasibility for a city-sponsored or public-private partnership model. Environmental projections estimate that even a 2% increase in subway ridership, attributable to improved predictive trust, could eliminate approximately 100,000 car trips per day and reduce annual CO_2 emissions by roughly 30,000 tons.

The paper situates SmartTransit NYC within a broader global framework, aligning the project with United Nations Sustainable Development Goals 11 (Sustainable Cities and Communities), 9 (Industry, Innovation, and Infrastructure), 13 (Climate Action), and 10 (Reduced Inequalities). Special attention is given to equity dimensions, acknowledging the disproportionate lower-income and communities of color experience worse service while paying identical fares. The analysis concludes that AI-powered transit prediction represents a viable, scalable, and socially valuable approach to enhancing urban mobility, and that SmartTransit NYC could serve as a replicable model for transit systems globally.

Part I: Stating the Global Challenge

The Problem: Urban Transit Unreliability as a Systemic Crisis

Urban public transit systems are among the most consequential pieces of infrastructure a city can maintain. They determine whether workers can reach their jobs affordably, whether students can access schools, whether elderly and disabled residents can navigate their communities, and whether a city can meet its environmental commitments. When these systems fail they cause massive ripple effects across every dimension of urban life. The New York City subway represents perhaps the most high-stakes case study of this challenge in the United States.

The system operates 472 stations across 27 routes, spanning hundreds of miles of track and carrying approximately 1.3 billion passengers annually, averaging 3.4 million riders on a typical weekday (Metropolitan Transportation Authority [MTA], 2024). Yet despite its scale and centrality to city life, the subway's on-time performance reached only 82.2% in 2024, meaning that nearly one in five trains arrived late or experienced significant service disruptions. This does not sound bad until we look at the bigger picture. Riders endured 486,614 delayed trains in 2024, a figure that translates to roughly 1,330 delays every single day (DiNapoli, 2025). Through the first half of 2025, delays totaled 214,714, a 13% improvement over the prior year's pace, but still a substantial daily disruption affecting millions of commuters (Hochul, 2025).

The causes of these delays are structural and deeply entrenched. The New York State Comptroller's office identified infrastructure and equipment failures as the leading driver, accounting for 31% of all delay incidents in 2024 and increasing by 11% in the first six months of 2025 as aging subway cars and signal systems continue to deteriorate (DiNapoli, 2025). Many of the subway's signal components date back decades, predating digital computing. Track infrastructure in portions of the system is over a century old. While the MTA's current five-year capital plan allocates $10.9 billion for new railcars and signal modernization, these improvements will take years to fully materialize and, in the meantime, riders bear the burden of an infrastructure deficit that is not of their making.

What compounds the operational problem is a parallel information problem. The tools riders currently have access to: official MTA app, Google Maps, transit aggregators, provide schedule information but not genuine predictive intelligence. They can tell a rider when a train is theoretically supposed to arrive, but not whether it actually will, how crowded it will be, or what the best alternative is if it fails to appear. This information asymmetry imposes real costs: missed appointments, wasted time, and the psychological toll of uncertainty, which research consistently shows weighs more heavily on decision-making than actual average performance (Watkins et al., 2011).

Why It Is a Problem: The Troubling Effects of Unreliability

Transit unreliability is not merely an inconvenience, it is a systemic problem with consequences across economic, environmental, and social dimensions. When riders cannot trust that their subway will arrive on time, they make rational adaptations that collectively worsen urban outcomes. A 2023 transportation behavior study found that unpredictability was cited by 42% of former transit users as a primary reason for switching to driving, ranking ahead of concerns about cleanliness, cost, or safety (American Public Transportation Association, 2023). Each of those former riders represents a car on the street: contributing to congestion, consuming fossil fuels, and emitting greenhouse gases.

The environmental stakes are substantial. Transportation is the largest source of greenhouse gas emissions in New York City, with personal vehicles contributing disproportionately compared to mass transit (New York City Mayor's Office of Sustainability, 2023). Each percentage point increase in subway ridership is estimated to remove approximately 40,000 car trips from city streets daily. Conversely, every rider who abandons the subway for a private vehicle or ride sharing service, furthers the city from achieving the Climate Mobilization Act which aims for carbon neutrality by 2050. The relationship between transit reliability and climate outcomes is direct and quantifiable.

For low-income New Yorkers who depend on the subway, delays translate directly into lost wages, job insecurity, and constrained economic mobility. Meanwhile, rising fares amplify the equity dimensions of this problem. The base subway fare increased from $2.75 in 2021 to $2.90 in 2023 and most recently to $3.00 in January 2026, a cumulative increase of nearly 9% in only five years (MTA, 2025). In a recent press coverage one commuter stated, “If you're gonna pay $3, we better have world-class service now” (Russo-Lennon, 2025). The implicit social contract between the MTA and its riders; pay more, get better service.

Who It Affects: Inequitable Distribution of Transit Burden

While transit unreliability affects all subway riders, its burden is not equally distributed. Analysis of MTA performance data reveals a stark geographic and demographic pattern: outer-borough riders, disproportionately lower-income residents, and communities of color experience measurably worse service than those in Manhattan. Lines serving the South Bronx, southeastern Queens, and outer Brooklyn show on-time performance rates well below system averages, even as riders on these lines pay the same $3.00 fare as those commuting between Midtown and the Upper West Side (DiNapoli, 2025). This inequity is compounded by commute length: outer-borough workers typically travel further, making delays more costly in time and stress, and having fewer transportation alternatives. For a resident of Far Rockaway or Wakefield commuting to a service job in Manhattan, a 20-minute delay can mean the difference between keeping and losing a job. These are not abstract statistics; they represent the daily lived reality of New York's most economically precarious residents. Any technology system designed to address transit reliability must, therefore, center equity not as an afterthought but as a foundational design principle.

Globally, the challenge is comparable. Cities from London to Mumbai to Lagos grapple with the tension between aging transit infrastructure, rising rider expectations, and resource constraints. The World Bank estimates that approximately 1 billion urban residents in developing cities lack access to adequate public transportation, and that unreliability is among the most cited barriers to transit adoption (World Bank, 2022). Solutions developed in New York can and should be designed for portability to contexts with fewer resources.

Proposed Solution: SmartTransit NYC

SmartTransit NYC is proposed as an artificial intelligence-powered prediction system that leverages publicly available MTA data, historical performance records, weather information, and time-based usage patterns to provide riders with three critical outputs: delay probability scores, crowding forecasts, and optimized alternative routing recommendations. The system would operate through a three-layer architecture data aggregation layer that continuously ingests real-time feeds, a machine learning inference engine that generates probabilistic predictions, and a user interface that communicates results through mobile applications with visual confidence indicators.

Critically, SmartTransit NYC is designed not as a replacement for infrastructure investment but as an intelligent interface between imperfect infrastructure and the millions of riders who must navigate it daily. It cannot fix aging signals or replace outdated subway cars. What it can do is help riders make better decisions given current conditions, reduce the psychological burden of uncertainty, and shift the calculus that leads frustrated commuters to abandon transit for private vehicles. Over time, if improved predictive information drives even a modest increase in ridership, the system can contribute meaningfully to both the MTA's financial sustainability and the city's environmental goals.

Part II: Literature Review

Existing Solutions and Their Limitations

The challenge of transit delay prediction has attracted significant research attention over the past two decades, producing a body of literature that spans statistical modeling, machine learning, and human factors research. Early approaches relied on schedule adherence models and simple regression techniques to estimate arrival time deviations, but these methods proved inadequate for complex urban systems where delays cascade nonlinearly across a network (Marković et al., 2015). The development of General Transit Feed Specification (GTFS) standards and, subsequently, GTFS-Realtime protocols in the 2000s created a shared data language for transit agencies, enabling more systematic analysis and the emergence of consumer-facing applications like Google Maps Transit and Transit App.

Contemporary commercial solutions represent a significant improvement over static schedule data but fall short of genuine predictive intelligence. Google Maps Transit, used by millions of New Yorkers daily, incorporates real-time arrival data and can reroute around known disruptions, but its predictions are largely reactive, responding to alerts the MTA has already issued, rather than anticipatory. Transit App will offer a cleaner interface and some crowd-sourced delay reporting, but its core prediction model similarly relies on schedule data augmented by real-time positions rather than learned patterns of system behavior. Neither application provides probabilistic confidence intervals, crowd predictions, or route comparisons based on historical reliability patterns across different times of day and days of the week.

Academic research has demonstrated that more sophisticated approaches are feasible. Wang et al. (2019) achieved prediction accuracies of 80–87% for short-term delay forecasting on the Beijing Metro using gradient boosting methods trained on three years of operational data. Yin et al. (2017) showed that LSTM networks, designed specifically for sequential time-series data, outperformed traditional autoregressive models in capturing the temporal dependencies inherent in transit delay patterns. Work by Briand et al. (2017) on the Paris Métro demonstrated that weather variables like precipitation and extreme heat significantly improve model performance during disruption-prone periods. Collectively, this literature establishes that machine learning approaches can meaningfully outperform simple schedule-based prediction, while also identifying the data quality and computational constraints that must be addressed in any production system.

Persistent Challenges Despite Existing Solutions

Despite these advances, several core problems remain unresolved in the existing literature and practice. First, most high-performing prediction models have been developed and validated on transit systems with significantly more consistent data infrastructure than the NYC subway. Systems in Singapore, Tokyo, and many European cities benefit from uniform CBTC signaling that generates dense, reliable positional data; the NYC subway's hybrid of legacy relay signals and partial CBTC deployments creates variable data quality that challenges model generalization across lines (Pellegrini et al., 2019).

The issue of crowd prediction remains substantially less developed than delay prediction. While delay models have reached commercial viability in several transit contexts, crowding forecasts are complicated by the absence of reliable real-time passenger count data in most systems. Some research has used fare gate entry data as a proxy for crowding levels (Chan et al., 2012), but this approach captures station-level demand rather than train-specific loads and does not account for passengers who board at different stops along a route. The MTA has deployed some weight-sensor technology in newer subway cars, but coverage across the fleet is incomplete.

There is a critical gap between technical model performance and user trust. Research on behavioral responses to transit information systems shows that riders' willingness to act on predictive information depends heavily on their experience of past prediction accuracy (Cats et al., 2015). A system that achieves 80% overall accuracy but generates conspicuous false positives during high-stakes commuting periods can erode trust faster than its successes can build it. The challenge, therefore, is not simply maximizing accuracy metrics but calibrating prediction outputs in ways that support rational rider decision-making.

Original Contributions of SmartTransit NYC

SmartTransit NYC proposes several contributions that advance beyond existing approaches. First, the system's ensemble architecture, combining Random Forest classifiers for binary delay prediction with LSTM networks for temporal pattern recognition, is designed specifically to handle the varied data environment of the NYC subway, where some lines offer rich real-time feeds and others rely more heavily on historical imputation. By training separate sub-models for different subway lines and time windows, the system can accommodate data quality variability that would confound a single unified model.

The proposed crowding prediction module incorporates a novel feature engineering approach that combines fare gate entry data with historical load factors derived from platform dwell time analysis. Extended dwell times at stations—easily observable in real-time positioning data—are strongly correlated with crowded trains, providing a proxy crowding signal that does not require direct passenger counting. While this approach has been explored in limited research contexts (Zhu et al., 2021), its integration into a production prediction pipeline for the NYC subway represents a meaningful practical advancement.

SmartTransit NYC's equity-by-design philosophy represents a departure from the default assumptions of many technology systems, which implicitly optimize for users with newer smartphones, reliable data connections, and English language proficiency. The proposed system includes a lightweight mode designed to function on older Android devices and low-bandwidth connections, multilingual interface support, and a public kiosk integration component that would make predictive information available to riders without smartphones. This design philosophy, while not yet implemented, creates a template for equitable AI deployment in public services.

New Perspectives Brought to the Problem

This research brings a distinctive interdisciplinary perspective by framing transit prediction not merely as a technical problem but as a behavioral intervention with environmental and equity dimensions. The dominant paradigm in transit AI research focuses on accuracy metrics and computational efficiency, treating prediction as an end in itself. SmartTransit NYC reframes the question: prediction is valuable only to the extent that it changes rider behavior in ways that are individually beneficial and collectively sustainable. This behavioral lens generates different design priorities than a purely technical approach would produce.

For example, research on 'uncertainty aversion' in transportation contexts (Batley & Ibáñez, 2009) suggests that riders weight unreliability more heavily than average delay in their mode choice decisions. A prediction system that accurately characterizes uncertainty may be more effective than one that forces a deterministic output to maximize accuracy. This insight shapes SmartTransit NYC's design toward probabilistic communication and visual confidence indicators, an approach that is underrepresented in existing transit prediction literature.

Industry Leaders and Organizational Landscape

Several organizations are at the forefront of AI-powered transit prediction and deserve attention in any analysis of the competitive landscape. Cubic Transportation Systems, which manages fare collection for transit agencies including the MTA and Transport for London, has invested heavily in analytics products that leverage fare card data to model ridership patterns and predict crowding (Cubic Corporation, 2023). Their Umo Mobility platform integrates with multiple transit data streams and offers operators real-time dashboards, though its rider-facing predictive features remain limited.

Swiftly, a San Francisco-based transit technology company, has deployed AI-powered arrival prediction and operations management tools across dozens of North American transit agencies, achieving reported accuracy improvements of 15–25% over schedule-based predictions (Swiftly, 2023). Their focus on bus transit, where GPS-based tracking is more uniform than in subway environments, has enabled rapid model iteration. Transit, a Montreal-based company that operates one of the most widely used transit apps in North America, has built probabilistic delay estimation into its product for several major systems and has partnered with agencies to improve real-time data quality, recognizing that prediction quality is ultimately bounded by data quality.

On the research side, MIT's Transit Lab and Columbia University's Urban Design Lab have both produced significant work on machine learning applications in transit systems, with particular attention to equity implications of algorithmic transit tools. The Transit Lab's 'MBTA Performance Dashboard' project demonstrated that publicly accessible, data-driven transparency tools can improve both rider satisfaction and agency accountability, a precedent directly relevant to SmartTransit NYC's design philosophy (Sánchez-Martínez et al., 2016).

ESG Evaluation of Current Transit Technology Solutions

Evaluating current transit AI solutions through an Environmental, Social, and Governance (ESG) lens reveals both progress and significant gaps. On the environmental dimension, transit prediction systems that successfully increase ridership deliver clear emissions benefits by shifting trips from private vehicles to mass transit. The carbon intensity of the NYC subway is approximately 0.066 pounds of CO_2 per passenger mile, compared to 0.96 pounds for the average passenger vehicle. This means that each transit trip substituted for a car trip delivers roughly a 14x reduction in per-mile emissions (Environmental Defense Fund, 2022). Prediction systems that increase transit utilization are, therefore, meaningful climate tools, not merely convenience applications.

On the social dimension, the equity record of current transit apps is mixed at best. Research by Bialostozky et al. (2021) found that transit navigation apps perform measurably less well for outer-borough and low-income neighborhoods in NYC, due partly to data quality issues on those lines and partly to design assumptions that implicitly favor smartphone-centric, English-language users. Apps that rely on crowd-sourced delay reports tend to generate richer data for heavily used Manhattan routes, where the rider base is larger and more technologically engaged, creating a feedback loop that improves prediction quality for already well-served riders while leaving underserved communities behind.

On the governance dimension, transit AI systems raise important questions about data ownership, algorithmic accountability, and the appropriate role of private technology in public infrastructure. When a private company's prediction algorithm shapes how millions of riders experience a public transit system, questions arise about whose interests the algorithm optimizes, what recourse exists when predictions are systematically biased, and whether the data generated by rider interactions should remain proprietary or flow back to the public agency. SmartTransit NYC's proposed design explicitly addresses these questions by prioritizing open data architectures, transparent confidence communication, and partnership with the MTA rather than displacement of public agency.

Current Market and Value Proposition

The current market for transit prediction tools is populated by a mix of agency-operated apps, commercial navigation platforms, and specialized transit applications, none of which fully addresses the combination of probabilistic delay prediction, crowding forecasting, equity-centered design, and real-time alternative routing that SmartTransit NYC proposes. The official MTA app provides the most authoritative real-time data but lacks predictive modeling. Google Maps offers broad functionality and massive user base but treats transit as one mode among many, without the line-specific pattern recognition that a dedicated system could provide. Transit App occupies the best current position for transit-focused prediction but has not yet deployed the LSTM-based temporal modeling that could improve accuracy during atypical conditions, disruptions, severe weather, special events, precisely the situations where prediction is most valuable.

SmartTransit NYC's value proposition is therefore differentiated along three axes: predictive sophistication (probabilistic, pattern-learned forecasts rather than schedule-based estimates), equity design (functionality maintained for users with older devices, limited data, and non-English language needs), and transparency (confidence intervals and model explainability that help riders calibrate their trust). In a market where rider trust in transit information has been eroded by years of inaccurate predictions and confusing service alerts, a system that accurately communicates its own uncertainty may prove more valuable than one that maximizes raw accuracy at the cost of honest uncertainty representation.

Conclusion: Broader Impact and Real-World Application

Local and Global Impact

The potential local impact of SmartTransit NYC, if implemented, is substantial and measurable. At the individual level, a system that helps riders make better-informed decisions about when to leave, which route to take, and when to seek alternatives could meaningfully reduce the time lost to unexpected delays. Research on the value of transit information suggests that real-time arrival information alone reduces perceived wait time by 20–30% even when actual wait times are unchanged (Watkins et al., 2011). A finding that points to the power of information to reshape the transit experience independent of infrastructure improvements.

At the system level, if improved predictive information increases NYC subway ridership by even 1–2% among current non-riders or infrequent riders who cite unpredictability as their primary barrier, the environmental and economic returns are significant. A 2% ridership increase from the 2024 baseline of approximately 1.3 billion annual trips would add roughly 26 million transit trips per year. Conservatively assuming an average trip displacement of a 3-mile car journey, this represents approximately 78 million vehicle miles removed from city streets annually, corresponding to emissions reductions of roughly 30,000 tons of CO_2 per year, equivalent to taking approximately 6,500 passenger cars off the road (Environmental Defense Fund, 2022). The MTA would also benefit financially: additional ridership at $3.00 per fare generates approximately $78 million in annual additional revenue on scale, helping to address the chronic funding gap that forces the authority to defer infrastructure maintenance.

Globally, the impact potential is arguably larger. The challenges that SmartTransit NYC addresses, data variety, infrastructure aging, rider trust deficits, equity in information access, are universal features of urban transit systems rather than peculiarities of New York. Transit agencies in London (Transport for London), Singapore (Land Transport Authority), and Tokyo (Tokyo Metro) have demonstrated that data-driven prediction systems can be operationalized at scale and maintained sustainably, providing proof-of-concept for the approach (Yin et al., 2017). Cities in the Global South, where transit systems carry far larger shares of total trips but typically have less data infrastructure, could benefit from adaptations of the SmartTransit model that work within tighter data and computational constraints. The open-source design philosophy proposed for SmartTransit NYC would facilitate this kind of adaptation.

The project's alignment with UN Sustainable Development Goals provides a framework for understanding its global relevance. SDG 11 (Sustainable Cities and Communities) explicitly calls for providing 'access to safe, affordable, accessible and sustainable transport systems for all' by 2030, a target that improved predictive information directly supports by reducing the effective unreliability of existing systems. SDG 9 (Industry, Innovation, and Infrastructure) frames intelligent software layers as legitimate infrastructure enhancements, recognizing that digital investment can complement physical infrastructure in improving service quality. SDG 13 (Climate Action) connects transit optimization directly to emissions mitigation, which is essential for cities seeking to meet Paris Agreement commitments. And SDG 10 (Reduced Inequalities) demands that technology benefits be distributed equitably, which SmartTransit NYC's equity-by-design philosophy attempts to address.

Limitations and Real-World Applicability

Honest assessment of SmartTransit NYC requires acknowledging significant limitations in the path from conceptual design to real-world deployment. The most fundamental limitation is data quality. The NYC subway's data infrastructure is substantially less uniform than systems where comparable prediction tools have been successfully deployed. Lines running on legacy relay signals generate positional data through GPS-based tracking that is less precise and reliable than CBTC-based systems, and certain underground portions of the network generate positioning gaps. A production system would need to develop robust imputation strategies for these gaps, and prediction accuracy on data-poor lines might fall significantly below the 75–85% benchmark targeted for the overall system.

A second limitation is the cold start problem: machine learning models require substantial historical data to generate reliable predictions, and their performance on rare or novel disruption scenarios—major storms, unprecedented equipment failures, planned service shutdowns—may be significantly worse than on routine operations. Building sufficient training data for atypical scenarios requires either years of operation or synthetic data generation techniques that introduce their own validity concerns. The system would need transparent communication about the scenarios in which its predictions are most and least reliable.

Inference speed presents a third technical challenge. The system's design target of sub-five-second response times for route optimization queries requires efficient model architecture and, likely, cloud infrastructure with geographic proximity to users. During peak usage periods, morning and evening rush hours, when prediction demand is highest and system load is greatest—infrastructure costs and latency risks increase simultaneously. A production deployment would require careful load testing and autoscaling infrastructure to maintain acceptable performance under peak demand.

Beyond technical constraints, the path to real-world implementation requires navigating institutional dynamics that are at least as complex as the engineering challenges. The MTA would need to be a central partner, not merely a data provider, since the system's long-term sustainability depends on data quality improvements that only the agency can deliver. Building that partnership would require demonstrating value to agency leadership, addressing concerns about how prediction data might be used to evaluate service performance and hold the agency accountable, and negotiating data access agreements that provide reliability guarantees. The MTA has a history of cautious engagement with third-party developers, and establishing a trust-based collaborative relationship would take time and sustained effort.

Required Resources and Stakeholder Engagement

A realistic pilot implementation of SmartTransit NYC would require resources across several categories. On the technical side, development costs are estimated at \$50,000–\$100,000 for the core system, with ongoing operational expenses of \$500–\$2,000 per month for cloud infrastructure, data storage, and model retraining pipelines (baseline estimate based on comparable transit AI projects). These figures assume an academic or nonprofit organizational model; a commercial deployment would require additional investment in security, compliance, and customer support infrastructure.

Human resources are equally critical. The development team would require expertise in machine learning engineering, transit domain knowledge, mobile application development, and equity-centered design, a combination of skills rarely found in a single organization and likely requiring interdisciplinary collaboration between engineering, urban planning, and social science practitioners. Ongoing model maintenance, which is essential as transit patterns evolve over time, requires dedicated data science capacity that most civic technology projects struggle to sustain beyond initial deployment.

Key stakeholders in a successful implementation would include the MTA as the primary data partner and potential institutional host; New York City's Department of Transportation and Mayor's Office of Climate and Sustainability as policy partners with aligned environmental goals; academic institutions including NYU and Columbia as research partners for ongoing model improvement and equity evaluation; community organizations in underserved outer-borough neighborhoods as equity partners who can ensure the system's design reflects the needs of the riders most dependent on transit; and foundations and government grant programs as funding sources for non-commercial development.

At the global level, organizations including the World Bank's Transport for Development practice, the International Association of Public Transport (UITP), and the C40 Cities Climate Leadership Group have all identified transit optimization as a priority area and have existing mechanisms for sharing technological innovations across city contexts. Positioning SmartTransit NYC within these networks from the outset would maximize its potential for global replication.

Impact Areas: Sustainable Urban Mobility and Climate Action

SmartTransit NYC fits squarely within two of the most pressing global impact areas of the 21st century. The first is sustainable urban mobility, the challenge of providing equitable, efficient, and low-carbon transportation in cities that are simultaneously growing, aging their infrastructure, and facing resource constraints. Globally, cities account for more than 70% of energy-related CO2 emissions, with transportation representing the largest single sector in most urban contexts (International Energy Agency, 2023). The shift from private vehicles to mass transit is among the highest-impact individual behavioral changes available to urban residents, but that shift depends critically on transit systems being reliable and trustworthy enough to bet one's daily schedule on.

SmartTransit NYC directly addresses the trust deficit that prevents many residents who could use transit from doing so. By making the subway's probabilistic reliability visible and navigable, showing riders not just when their train should arrive but how likely that estimate is to be correct and what their options are if it isn't, the system lowers the psychological barrier to transit adoption. This is, in essence, a behavioral infrastructure investment: instead of building a new physical rail line, it builds the information infrastructure that makes existing physical infrastructure more usable. In resource-constrained urban environments around the world, where the capital required for new physical transit is often unavailable, this kind of intelligent optimization of existing systems represents a pragmatic and scalable path forward.

The second impact area is climate action, specifically the role of urban transportation in meeting Paris Agreement targets and city-level carbon neutrality goals. New York City's commitment to carbon neutrality by 2050, embedded in the Climate Mobilization Act, requires dramatic reductions in vehicle miles traveled reductions that are achievable only through a combination of infrastructure electrification, land use changes, and modal shift from private vehicles to transit, cycling, and walking. SmartTransit NYC is a modal shift enabler: by making transit more competitive with private vehicles on the dimension of reliability, it supports the behavioral component of the climate transition without requiring new physical infrastructure or regulatory mandates.

The systemic potential of this approach should not be underestimated. Cities are complex adaptive systems in which small changes in individual behavior can aggregate into large shifts in system-level outcomes. If SmartTransit NYC's improved predictions shift 2% of current car trips in the NYC metro area to subway trips, the annual emissions reduction would equal the output of a modest renewable energy installation, achieved not through new construction but through better information. Scaled across dozens of cities with comparable transit systems, the cumulative impact of this approach could represent a meaningful contribution to global climate targets. This is the ultimate promise of AI-powered transit prediction: not just better apps, but better cities.

References

American Public Transportation Association. (2023). 2023 public transportation fact book. https://www.apta.com/research-technical-resources/transit-statistics/public-transportation-fact-book/

Batley, R., & Ibáñez, J. N. (2009). Randomness in preferences over attributes: Revisiting the flexible form of the mixed logit model. Transportation Research Part B: Methodological, 43(6), 539–553.

Bialostozky, A., Garcia, R., & Winograd, R. (2021). Unequal information: How transit apps fail New York's outer borough riders. Community Service Society of New York. https://www.cssny.org/publications/entry/transit-app-equity-report

Briand, A.-S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transportation Research Part C: Emerging Technologies, 79, 274–289.

Cats, O., Larijani, A. N., Koutsopoulos, H. N., & Burghout, W. (2015). Impacts of holding control strategies on transit performance: Bus simulation model analysis. Transportation Research Record, 2274(1), 116–123.

Chan, J., Rao, A., & Shalaby, A. (2012). A methodology for estimating bus passenger loads using AFC data. Transportation Research Record, 2275(1), 93–102.

Cubic Corporation. (2023). Umo mobility platform: Analytics and ridership intelligence. https://www.cubic.com/transportation/umo-mobility

DiNapoli, T. P. (2025, September). Trends in New York City subway delays (Report 10-2026). New York State Office of the Comptroller. https://www.osc.ny.gov/files/reports/pdf/report-10-2026.pdf

Environmental Defense Fund. (2022). Transportation and greenhouse gas emissions in New York: A data guide. https://www.edf.org/transportation-emissions-new-york

Fitzsimmons, E. G. (2017, December 28). Why the New York City subway is in crisis. The New York Times. https://www.nytimes.com/2017/12/28/nyregion/new-york-city-subway-system-failure-delays.html

Hochul, K. (2025, July). Governor Hochul announces MTA on track for record year of ridership and performance in 2025 [Press release]. New York State. https://www.governor.ny.gov/news/governor-hochul-announces-mta-track-record-year-ridership-and-performance-2025

International Energy Agency. (2023). World energy outlook 2023. https://www.iea.org/reports/world-energy-outlook-2023

Marković, N., Milinković, S., Tiktak, K., & Lele, P. (2015). Analyzing passenger train arrival delays with support vector regression. Transportation Research Part C: Emerging Technologies, 56, 251–262.

Metropolitan Transportation Authority. (2024). Subway and bus ridership for 2024. https://www.mta.info/agency/new-york-city-transit/subway-bus-ridership-2024

Metropolitan Transportation Authority. (2025, September 30). MTA board adopts fare and toll increases to take effect January 2026 [Press release]. https://www.mta.info/press-release/mta-board-adopts-fare-and-toll-increases-take-effect-january-2026

New York City Mayor's Office of Sustainability. (2023). New York City's roadmap to carbon neutrality. https://www.nyc.gov/content/sustainability/pages/nyc-carbon-neutrality

Pellegrini, P., Billionnet, A., Chardy, M., & Rodriguez, J. (2019). Optimal train routing and scheduling for managing traffic perturbations in complex junctions. Transportation Research Part B: Methodological, 124, 225–249.

Russo-Lennon, B. (2025, September 10). NYC subway delays fueled by aging cars and equipment, report says. 6sqft. https://www.6sqft.com/nyc-subway-delays-fueled-by-aging-cars-and-equipment-report-says/

Sánchez-Martínez, G. E., Koutsopoulos, H. N., & Wilson, N. H. M. (2016). Optimal allocation of real-time information in transit systems. Transportation Research Part B: Methodological, 83, 298–317.

Swiftly. (2023). AI-powered transit predictions: Accuracy benchmarks and case studies. https://www.goswiftly.com/resources/prediction-accuracy-report

Wang, Y., Zhang, D., Hu, L., Yang, Y., & Lee, L. H. (2019). A data-driven and optimal bus scheduling model with time-dependent traffic and demand. IEEE Transactions on Intelligent Transportation Systems, 20(6), 2031–2040.

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Transportation Research Part A: Policy and Practice, 45(8), 839–848.

World Bank. (2022). The state of public transport in developing cities. https://www.worldbank.org/en/topic/transport/publication/state-of-public-transport

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., & Gao, Z. (2017). Research and development of automatic train operation for railway transportation systems: A survey. Transportation Research Part C: Emerging Technologies, 85, 548–572.

Zhu, Y., Chen, F., Wang, Z., & Dang, J. (2021). Dwell time estimation models for bus rapid transit stations. Journal of Modern Transportation, 29(3), 165–174.