Checkpoint: Virginia Tech Transportation Safety Index (VTTSI)

Jason Cusati
Computer Science
Virginia Tech
Blacksburg, VA, USA
djjay@vt.edu
(LEADER)

Cheng-Shun Chuang
Computer Science
Virginia Tech
Alexandria, VA, USA
cchengshun@vt.edu

Victory Uhunmwangho
Computer Science
Virginia Tech
Blacksburg, VA, USA
victoryu@vt.edu

***Abstract*—This project presents the development of the Virginia Tech Transportation Safety Index (VTTSI)—a cloud-based framework for real-time, data-driven safety assessment of urban intersections. Unlike traditional crash-based analyses that depend on multi-year historical data, VTTSI fuses multimodal traffic inputs—including Basic Safety Messages (BSM), Personal Safety Messages (PSM), vehicle and vulnerable road-user counts, speed distributions, and sensor-detected safety events—to compute exposure-adjusted risk indices refreshed every 15 minutes. The system produces three interpretable metrics: a Vehicle Safety Index, a Vulnerable Road User (VRU) Index, and a Combined Safety Index scaled from 0 to 100. The architecture integrates a FastAPI-based backend with PostgreSQL/PostGIS storage and a Streamlit frontend deployed on Google Cloud Run, delivering interactive geospatial dashboards for vehicles and VRUs. The framework employs severity weighting, Empirical Bayes stabilization, and multi-criteria decision-making (EDAS–CODAS–SAW) for robust index estimation. By transforming high-frequency V2X data into actionable safety intelligence, VTTSI enables transportation agencies, fleet managers, and autonomous-vehicle systems to identify emerging risks, support Vision Zero objectives, and prioritize interventions dynamically. This real-time approach bridges the gap between static crash records and the proactive safety management required for modern connected mobility systems.**

I. Introduction

A. Motivation

Roadway safety analysis is a critical task for transportation planning and engineering. Crash data collected by agencies such as VDOT provide a foundation for constructing quantitative indicators of roadway risk. However, the ability to act on safety risks in real time remains limited. While technology companies such as Google, Apple, and INRIX provide traffic flow visualizations, these services focus on travel time estimation rather than proactive safety assessment. Their models often lag actual conditions by several minutes, and the underlying training data can be weeks or even months old, reducing their utility for immediate decision-making. This gap highlights the need for a system that integrates crash history, traffic exposure, and real-time dynamics into a unified safety index.

B. Problem Statement

Raw crash counts alone are insufficient for assessing roadway risk, as they do not account for exposure, severity, or temporal context. Existing commercial tools provide aggregate congestion information but fail to generate actionable insights about safety risk at specific intersections or road segments. Furthermore, archived safety scoring methods often rely on multi-year data, producing results that are stale and less useful for real-time safety management [1], [2]. A data-driven approach is needed that combines crash severity, vehicle and vulnerable road user volumes, and dynamic speed variance into an interpretable 0–100 index, refreshed in 15-minute intervals.

C. Significance

Developing such a Safety Index has direct implications for both public agencies and specialized driver populations. For VDOT, an exposure-adjusted and severity-weighted index provides a transparent and evidence-based tool to prioritize interventions, support Vision Zero goals, and engage the public with clear safety metrics. For technology adopters such as autonomous vehicles, advanced driver assistance systems (ADAS), and large delivery fleets, the index can provide real-time safety context, enabling routing decisions that avoid high-risk intersections during periods of heightened danger. By delivering actionable, fresh, and population-specific insights, the proposed system bridges the gap between static crash datasets and the real-time decision engines required for modern mobility systems.

II. Literature Review: Traffic Safety, Real-Time Analysis, and Machine Learning in ITS

A. Traffic Safety Assessment

Traditional traffic safety assessment methods rely heavily on multi-year crash data to identify high-risk sites. For example, state and federal agencies typically require 3–5 years of crash records for hotspot analysis and project evaluation [3]. This retrospective approach often results in delays: safety concerns are only addressed after enough collisions have occurred. Researchers have identified this as a major limitation, leading to a “wait-for-crashes” paradigm that is slow to adapt to emerging hazards [4].

Recent studies call for a shift to proactive safety management, leveraging surrogate safety measures (e.g., hard braking, time-to-collision) and live traffic data to assess risk in real time. Hossain *et al.* describe how real-time monitoring of traffic conditions can detect precursors to crashes and enable interventions before collisions occur [4]. This vision is especially relevant for supporting Vision Zero goals, which require timely and actionable insights rather than archived reports.

B. Real-Time Traffic Analysis

Real-time traffic analysis provides the foundation for proactive safety by monitoring and responding to live traffic streams. Recent work shows how machine learning models trained on traffic flow and weather data can predict imminent crash risk. Zhang *et al.* (2022) demonstrated statewide crash risk prediction using random forests, SVM, and XGBoost on live traffic feeds combined with crash records, showing strong potential for real-time deployment [5]. Yeole *et al.* (2021) applied LSTM networks to predict crash risk at intersections, outperforming traditional regression approaches.

For emergency responders, real-time analytics are critical. Connected vehicle (CV) technologies can provide situational awareness and warnings to approaching drivers, reducing secondary collisions. Seattle’s recent pilot (dCAAMP) integrates live 911 dispatch data with CV systems, creating geofenced “emergency zones” that alert nearby vehicles and navigation apps in real time [6]. Such systems directly address the gap in providing responders with immediate safety context.

C. Machine Learning in Intelligent Transportation Systems

Machine learning (ML) has rapidly become central to modern Intelligent Transportation Systems (ITS). A comprehensive review of 191 studies (2018–2023) shows ML’s effectiveness in leveraging heterogeneous data sources for safety prediction [7]. Ensemble methods and deep learning models can capture complex spatiotemporal patterns in traffic streams that classical methods often miss.

Zhao *et al.* [8] introduced an AdaBoost-based risk predictor in vehicular ad-hoc networks (VANETs) with SMOTE oversampling, achieving an AUC of 0.77 in real time. More recently, Hou *et al.* [9] developed a deep transfer learning model for predicting traffic conflicts (near crashes) using trajectory data, enabling proactive warnings before actual collisions occur. Reinforcement learning has also been applied: Gong *et al.* [10] trained adaptive traffic signal controllers to minimize both delay and conflicts, demonstrating ML’s ability to balance efficiency and safety in real time.

ML also addresses data imbalance through synthetic data generation. GAN-based augmentation has been shown to improve detection of rare crash events, reducing false negatives [7]. Combined with connected vehicle streams, these ML models pave the way for ubiquitous, low-latency safety scoring systems that inform drivers, responders, and infrastructure controllers in real time.

D. Summary

The literature indicates a clear paradigm shift: from retrospective, crash-based safety scoring to proactive, real-time, ML-driven systems. This transition addresses the limitations of stale insights and equips both connected vehicles and emergency responders with immediate risk information. Future work lies in scaling these models to statewide networks, fusing multiple live data streams, and ensuring robustness against imbalanced and sparse data environments.

III. Literature Review: Flow-First Strategies to Reduce Incidents

A. Safety index merging methods

The development of a Safety Index for urban unsignalized intersections can be informed by prior research in systemic safety analysis, Empirical Bayes evaluation, severity weighting, and network screening. Four key references provide complementary insights into this process.

Montella et al. [11] develop and validate a procedure to rank unsignalized *urban* intersections using a composite Safety Index (SI) derived from road safety inspections. The SI combines two components: (i) *Exposure*—the extent to which road users are exposed to hazards; and (ii) a *Risk Index*—scores for safety issues that increase crash likelihood. The SI includes separate vehicle and pedestrian components and is computed from a structured checklist of detailed and general safety issues assessed in the field. Validation against Empirical Bayes (EB) safety estimates and potential-for-improvement (PFI) on a large urban sample showed strong agreement, supporting the SI’s utility for proactive screening and prioritization.

Persaud and Lyon [12] highlight the importance of the Empirical Bayes (EB) approach in before–after safety studies. EB adjusts observed crash frequencies with expected crash counts derived from Safety Performance Functions (SPFs), mitigating regression-to-the-mean bias. In a Safety Index context, EB-adjusted frequencies offer a statistically rigorous baseline for benchmarking or validating index outputs.

Schultz [1] emphasizes incorporating crash severity via weights (e.g., fatal and injury crashes weighted higher than property-damage-only). Embedding severity weighting into a Safety Index ensures locations with rarer but more severe outcomes are prioritized appropriately.

The Maricopa Association of Governments (MAG) [2] describe practical network screening using measures such as crash rate, excess crash frequency, and equivalent property damage only (EPDO). These measures, paired with exposure (e.g., traffic volumes) and site characteristics, provide scalable screening inputs to a composite Safety Index.

B. Synthesis & Gaps

Across engineering controls (C/AV smoothing, shockwave-aware signals, V2X intersections) and behavioral interventions (anti-tailgating and interaction-aware guidance), the unifying mechanism is *variance reduction*: shrink unexpected speed and lane-choice changes so drivers face fewer surprises. Open questions suitable for your project include: (i) how to prioritize corridors where small flow improvements most reduce rear-end risk; (ii) how to couple behaviorally informed messages with algorithmic flow control; and (iii) how to detect and preempt patterns like early braking waves and last-second exits using roadside sensing plus nudges (signing, lane guidance, advisory speeds).

IV. Dataset Description

The proposed Real-Time Safety Index will be constructed using *Smart Intersection datasets* collected from connected infrastructure and sensor systems at Virginia intersections (e.g., Glebe & Potomac). These datasets capture multimodal road user activity, vehicle kinematics, and safety events at fine temporal granularity, enabling dynamic risk monitoring.

A. Basic Safety Messages (BSM)

The *BSM dataset* contains vehicle-generated data packets compliant with SAE J2735 standards. Each record includes:

- Vehicle position (latitude/longitude).
- Speed, heading, acceleration, and yaw rate.
- Brake and stability control status.
- Vehicle size and type classification.

This stream allows identification of aggressive maneuvers, hard braking, and near-conflict trajectories in real time.

B. Personal Safety Messages (PSM)

The *PSM dataset* contains messages broadcasted from vulnerable road users (VRUs), such as pedestrians and cyclists. Each record provides:

- User position and movement type (walking, cycling).
- Device-reported activity state.
- Dynamic attributes such as acceleration and yaw rate.

By integrating PSM with BSM, the system detects potential conflicts between vehicles and VRUs.

C. Vehicle and VRU Counts

Dedicated counting datasets report aggregated flows:

- **Vehicle Count Dataset**: Records classified vehicle counts (passenger vehicles, buses, bicycles, etc.) by movement (e.g., left turn, through, right turn) and 15-minute time intervals.
- **VRU Count Dataset**: Provides pedestrian and bicycle volumes by approach and time interval.

These enable normalization of safety event rates by exposure, ensuring fair risk comparisons across intersections.

D. Speed Distribution Data

The *Speed Distribution Dataset* provides the number of vehicles by approach within defined speed bins (e.g., 0–5 mph, 6–10 mph). This distribution supports:

- Identification of speeding patterns.
- Exposure-adjusted severity modeling, since crash likelihood and outcome severity increase with higher speeds.

E. Safety Event Detection

The *Safety Event Dataset* logs events automatically detected by intersection sensors and video analytics, including:

- Intersection conflicts (IC).
- Lane change violations (LCV).
- Red-light running and near-miss events.
- Detection area and object classification (vehicle, pedestrian, cyclist).

These events are direct indicators of safety-critical interactions, forming the core of the real-time Safety Index.

F. Data Integration

- **Event-level fusion**: BSM and PSM are correlated with detected safety events to identify specific vehicle–VRU or vehicle–vehicle conflicts.
- **Exposure adjustment**: Safety event frequencies are normalized by vehicle and VRU counts, as well as speed distribution.
- **Temporal granularity**: Aggregation can be performed in near real-time (e.g., every 5–15 minutes) to generate live safety scores.

G. Data Limitations

Potential limitations include:

- Sensor blind spots and misclassification of VRUs.
- Limited adoption of PSM by pedestrians/cyclists carrying enabled devices.
- Data packet loss or latency in BSM/PSM transmission.
- Short-term variability in counts requiring smoothing for stable index trends.

Despite these challenges, the Smart Intersection dataset provides a rich, high-resolution foundation for constructing a **real-time, exposure-adjusted Safety Index** capable of capturing emerging risks at urban intersections.

V. Methodology

A. Real-Time Safety Index (RT-SI)

We compute a per-intersection, per-15-minute Safety Index that is severity-weighted, exposure-adjusted, stabilized by Empirical Bayes (EB), and adjusted by real-time operating factors.

a) Historical severity-weighted crash rate.:

$$r_i \;=\; \frac{\sum_{s\in\{\text{Fatal,Injury,PDO}\}} w_s\, C_{i,s}}{E_i}, \tag{1}$$

where $C_{i,s}$ are crash counts of severity s, w_s are severity weights (e.g., w_F=10, w_I=3, w_P=1), and E_i is the exposure (e.g., VMT or entering vehicle volume).

b) Empirical Bayes stabilization.:

$$\hat{r}_i \;=\; \alpha_i\, r_i + (1-\alpha_i)\, r_0, \qquad \alpha_i = \frac{E_i}{E_i+\lambda}, \tag{2}$$

where r_0 is the pooled mean rate and λ a shrinkage parameter.

c) Real-time uplift factors.:

$$F_{i,t}^{\text{spd}} = \min\left(1,\ k_1 \frac{v_i^{\text{FF}} - \bar{v}_{i,t}}{v_i^{\text{FF}}}\right), \tag{3}$$

$$F_{i,t}^{\text{var}} = \min\left(1,\ k_2 \frac{\sigma_{v,i,t}}{\bar{v}_{i,t} + \varepsilon}\right), \tag{4}$$

$$F_{i,t}^{\text{conf}} = \min\left(1,\ k_3 \frac{\text{turningVol}_{i,t} \cdot V_{i,t}^{\text{vru}}}{\text{scale}}\right). \tag{5}$$

We combine these into an uplift factor:

$$U_{i,t} = 1 + \beta_1 F_{i,t}^{\text{spd}} + \beta_2 F_{i,t}^{\text{var}} + \beta_3 F_{i,t}^{\text{conf}}. \tag{6}$$

d) Sub-indices.:

$$G_{i,t} = \min\left(1,\ k_4 \frac{V_{i,t}^{\text{vru}}}{V_{i,t}^{\text{veh}} + \varepsilon}\right), \quad \text{VRU}_{i,t} = \gamma \cdot \hat{r}_i \cdot U_{i,t} \cdot G_{i,t}, \tag{7}$$

$$H_{i,t} = \min\left(1,\ k_5 \frac{V_{i,t}^{\text{veh}}}{\text{capacity}_i}\right), \quad \text{VEH}_{i,t} = \gamma \cdot \hat{r}_i \cdot U_{i,t} \cdot H_{i,t}. \tag{8}$$

e) Combined and scaled index.:

$$\text{COMB}_{i,t} \ = \ \omega_{\text{vru}} \text{VRU}_{i,t} + \omega_{\text{veh}} \text{VEH}_{i,t}, \quad \omega_{\text{vru}} + \omega_{\text{veh}} = 1, \tag{9}$$

$$\text{SI}_{i,t}^{\text{RT}} \ = \ 100 \times \frac{\text{COMB}_{i,t} - \min}{\max - \min}. \tag{10}$$

—

B. EDAS (Evaluation Based on Distance from Average Solution)

Given a decision matrix x_{ij} for site i and criterion j, with weights w_j, define

$$\text{SP}_i = \sum_j w_j \,\text{PDA}_{ij}, \quad \text{SN}_i = \sum_j w_j \,\text{NDA}_{ij}, \tag{11}$$

$$\text{NSP}_i = \frac{\text{SP}_i}{\max \text{SP}}, \quad \text{NSN}_i = 1 - \frac{\text{SN}_i}{\max \text{SN}}, \tag{12}$$

$$\text{EDAS}_i = \tfrac{1}{2}(\text{NSP}_i + \text{NSN}_i). \tag{13}$$

—

C. CODAS (Combinative Distance-Based Assessment)

Define the negative ideal point $\mathbf{x}^-$. Then compute Euclidean and Taxicab distances:

$$E_i = \sqrt{\sum_j (x_{ij} - x_j^-)^2}, \tag{14}$$

$$T_i = \sum_j |x_{ij} - x_j^-|. \tag{15}$$

The CODAS score ranks alternatives by lexicographic order of (E_i, T_i), with a threshold function applied to E_i if required.

—

D. SAW (Simple Additive Weighting)

Normalize each criterion, weight by w_j, and sum:

$$\text{SAW}_i = \sum_j w_j \,\tilde{x}_{ij}. \tag{16}$$

—

E. Hybrid MCDM Index

We combine the three MCDM methods into a hybrid index:

$$\text{SI}_i^{\text{MCDM}} = W_E \,\text{EDAS}_i + W_C \,\text{CODAS}_i + W_S \,\text{SAW}_i, \tag{17}$$

with method weights $W_E + W_C + W_S = 1$.

—

F. Blended Final Index

To harmonize real-time safety with long-term prioritization, we define

$$\text{SI}_{i,t}^{\text{Final}} = \alpha \cdot \text{SI}_{i,t}^{\text{RT}} + (1 - \alpha) \cdot \widetilde{\text{SI}}_i^{\text{MCDM}}, \tag{18}$$

where α tunes the emphasis (e.g., $\alpha = 0.7$ for driver-facing dashboards).

—

G. Parameter Summary

Parameter	Meaning
w_s	Severity weights (e.g., Fatal=10, Injury=3, PDO=1)
λ	EB shrinkage strength
$k_1..k_5$	Scaling constants for speed drop, variance, conflicts, VRU/veh ratios
$\beta_1..\beta_3$	Weights for real-time uplift factors
$\omega_{\text{vru}}, \omega_{\text{veh}}$	Weights for VRU vs Vehicle sub-indices
W_E, W_C, W_S	Method weights for EDAS, CODAS, SAW in Hybrid index
α	Blend factor for Final index (RT vs. MCDM)

TABLE I
KEY TUNABLE PARAMETERS OF THE SAFETY INDEX FORMULAS.

H. Weight Determination Framework

We employ a combination of policy anchors and data-driven learning to determine weights at different levels of the Safety Index. This subsection documents the formulas and reproducibility protocol.

1) Severity Weights w_s: We assign severity weights using relative societal cost (policy-driven), e.g. Fatal = 10, Injury = 3, PDO = 1. These values can be validated by regressing crash costs or KSI probability against severity, but the policy basis ensures interpretability.

2) Empirical Bayes Shrinkage Parameter λ: We set the shrinkage factor as

$$\alpha_i = \frac{E_i}{E_i + \lambda}, \tag{19}$$

with λ tuned via cross-validation (see Protocol below) to minimize predictive log-loss for future crash/near-miss outcomes.

3) Real-Time Uplift Factors (k, β): Scaling constants $k_1..k_5$ and multipliers $\beta_1..\beta_3$ are estimated using a regularized GLM (or monotonic gradient boosting) predicting next-interval safety outcomes from Δv, coefficient of variation of speed, and VRU conflict exposure. Parameters are regularized to avoid over-sensitivity.

4) VRU vs. Vehicle Blend ω: We set ω_{vru} as a policy knob (e.g., 0.6 in pedestrian-heavy cores, 0.5 elsewhere). Optionally, ω can be tuned as a multi-objective optimization problem balancing predictive fit for VRU and vehicle incident probabilities.

5) MCDM Criterion Weights w_j: We apply objective weighting using a blend of CRITIC and Entropy methods.

a) CRITIC (Criteria Importance Through Intercriteria Correlation).: Let σ_j be the standard deviation of criterion j, and ρ_{jk} the correlation between criteria j and k. Then:

$$C_j = \sigma_j \cdot \left(\sum_{k=1}^{m} (1 - \rho_{jk}) \right), \quad w_j^{\text{CRITIC}} = \frac{C_j}{\sum_{j=1}^{m} C_j}. \tag{20}$$

b) Entropy Weighting.: Let $p_{ij} = x_{ij} / \sum_i x_{ij}$ for normalized criterion values. Then

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln(p_{ij}), \quad d_j = 1 - e_j, \quad w_j^{\text{ENT}} = \frac{d_j}{\sum_{j=1}^{m} d_j}. \tag{21}$$

c) Blended Criterion Weights.: We combine CRITIC and Entropy as:

$$w_j = \tfrac{1}{2}\big(w_j^{\text{CRITIC}} + w_j^{\text{ENT}}\big). \tag{22}$$

6) Hybrid Method Weights W_E, W_C, W_S: We determine the aggregation weights for EDAS, CODAS, and SAW via a constrained regression (stacking).

a) Optimization.: Let y_i be the observed outcome (e.g., crash probability in the following period), and $\hat{e}_i, \hat{c}_i, \hat{s}_i$ the standardized scores from EDAS, CODAS, and SAW respectively. We solve:

$$\min_{W_E, W_C, W_S} \sum_i \Big(y_i - (W_E \hat{e}_i + W_C \hat{c}_i + W_S \hat{s}_i) \Big)^2 \tag{23}$$

$$\text{s.t. } W_E, W_C, W_S \geq 0, \quad W_E + W_C + W_S = 1. \tag{24}$$

7) Final Blend α: We blend the real-time and MCDM indices:

$$\text{SI}_{i,t}^{\text{Final}} = \alpha \cdot \text{SI}_{i,t}^{\text{RT}} + (1 - \alpha) \cdot \widetilde{\text{SI}}_i^{\text{MCDM}}, \tag{25}$$

with α chosen by audience (0.6–0.8 for driver-facing, 0.3–0.5 for planning).

8) Cross-Validation Protocol (Reproducibility):

1) Split data temporally into K folds (rolling origin evaluation).
2) For each fold:
   a) Estimate λ by minimizing out-of-fold log-loss for EB-stabilized crash rates.
   b) Fit uplift factor GLM to estimate β and k parameters.
   c) Compute CRITIC and Entropy weights; average to get w_j.
   d) Stack EDAS, CODAS, and SAW via constrained regression to get W_E, W_C, W_S.
3) Aggregate results; report mean λ, β, k, w_j, and $W.$ across folds.
4) Apply isotonic regression to calibrate raw Safety Index values against observed incident probabilities.

9) Guardrails: To ensure robustness and fairness:

- Monotonicity constraints: increased variance or VRU exposure must not reduce risk.
- Clipping: per-interval change capped at ± 15 points unless a major event is detected.
- Equity: increasing VRU flow should never decrease VRU index.

VI. V2X-Derived Inputs for the Safety Index

This section enumerates candidate input parameters available from Basic Safety Messages (BSM), Personal Safety Messages (PSM), and Signal Phase And Timing (SPaT), plus static MapData attributes. For each parameter, we provide a safety rationale, expected direction of effect on a 0–100 risk score (higher = more risk), and an initial weighting heuristic to be refined empirically. Sources: VCC Public API (message structures and fields) :contentReference[oaicite:0]index=0, BSM JSON examples (speed/position/brakes/size) :contentReference[oaicite:1]index=1 :contentReference[oaicite:2]index=2, SPaT DTOs/metadata and signal group end-times :contentReference[oaicite:3]index=3 :contentReference[oaicite:4]index=4, SPaT endpoints/websocket streaming :contentReference[oaicite:5]index=5 :contentReference[oaicite:6]index=6, MapData (speed limits, lane geometry) :contentReference[oaicite:7]index=7 :contentReference[oaicite:8]index=8 :contentReference[oaicite:9]index=9; prior severity weighting/EB notes :contentReference[oaicite:10]index=10; adaptive signal effects and congestion modeling evidence :contentReference[oaicite:11]index=11 :contentReference[oaicite:12]index=12.

A. Parameter Catalog (BSM/PSM/SPaT/MapData)

B. Weighting and Calibration Strategy

We recommend a two-layer scheme: (i) **Historical Prior** built from EB-stabilized severity-weighted crash rates by intersection/approach and Time-of-Day/Day-of-Week (TOD/DOW), consistent with your prior plan (Fatal=10, Injury=3, PDO=1). :contentReference[oaicite:29]index=29 (ii) **Real-Time Uplift** from V2X: additive (or multiplicative) adjustments from BSM/PSM/SPaT signals that capture deviations from the historical norm. Adaptive SPaT coordination can reduce delays/emissions (a proxy for risk) and should thus *lower* the uplift when detected. :contentReference[oaicite:30]index=30

A generic instantaneous score for an intersection approach a in 15-minute bin t:

$$S_{a,t} = \underbrace{\theta_{0,a,\mathrm{TOD,DOW}}}_{\text{EB prior}} + \sum_{j\in\mathcal{B}} \theta_j^{(B)} z_{j,a,t}^{(B)} + \sum_{k\in\mathcal{P}} \theta_k^{(P)} z_{k,a,t}^{(P)} + \sum_{m\in\mathcal{S}} \theta_m^{(S)} z_{m,a,t}^{(S)}$$

where z terms are standardized (by local historical TOD/DOW means) BSM, PSM, and SPaT features, and θ are weights learned via regularized regression or Bayesian updating.

C. Feature Engineering Notes (per message family)

a) BSM (vehicles).: Use per-vehicle speed, heading, position, brake flags, and size from the JSON representation to aggregate approach-level features (mean/variance of speed, hard-brake rate, heading change rate). :contentReference[oaicite:31]index=31 :contentReference[oaicite:32]index=32

b) PSM (VRUs).: Use VRU positions/speeds/trajectories to estimate VRU flow and conflict overlap windows with turning movements during permissive phases (PSM DTOs are supported in the VCC set). :contentReference[oaicite:33]index=33

c) SPaT (signals).: From per-group phases and min/max end timestamps, derive remaining time, phase volatility, and permissive/protected status; consume via the WebSocket stream for near real-time updates. :contentReference[oaicite:34]index=34 :contentReference[oaicite:35]index=35 :contentReference[oaicite:36]index=36

d) MapData (static context).: Use lane geometry and regulatory speed limits as context features and to determine applicable limits for speed deviation. :contentReference[oaicite:37]index=37 :contentReference[oaicite:38]index=38

D. Narrative for Methods Section (drop-in)

Data sources. We fuse vehicle BSM, VRU PSM, SPaT streams, and static MapData via the Virginia Connected Corridor (VCC) Public API, which exposes J2735 messages in JSON and via WebSockets for streaming SPaT. :contentReference[oaicite:39]index=39 :contentReference[oaicite:40]index=40

Historical prior. For each intersection/approach and 15-min TOD/DOW bin, we compute a severity-weighted crash rate (Fatal=10, Injury=3, PDO=1) stabilized with Empirical Bayes to obtain θ_0. :contentReference[oaicite:41]index=41

Real-time uplift. We derive standardized features from BSM (speed level/variance, hard-brake rate, weaving), PSM (VRU density/speed/path alignment), and SPaT (phase state, time-to-change, volatility). Features that increase instantaneous conflict probability (e.g., high speed variance, permissive turn with high VRU flow, short remaining green) raise the index; features that decrease conflicts (e.g., protected turns, coordinated/adaptive SPaT) lower it. :contentReference[oaicite:42]index=42 :contentReference[oaicite:43]index=43

Estimation. We fit θ with regularized regression on matched 15-min bins, targeting leading indicators (surrogate conflicts, hard-brakes) and lagging outcomes (crash/severity), with MapData features as controls. We recommend periodic recalibration and shrinkage to preserve stability.

VII. System Design

The system adopts a cloud-based client–server architecture with a clear separation between the backend services, frontend visualization layer, and database management components. This modular design ensures scalability, maintainability, and seamless integration with cloud services.

A. Backend

The backend is implemented using **FastAPI**, selected for its asynchronous I/O support and lightweight performance suitable for cloud deployment. The backend is currently hosted on **Google Cloud Run**, allowing automatic scaling and containerized execution.

Backend API Endpoint: https://cs6604-trafficsafety-180117512369.europe-west1.run.app/

The backend is responsible for:

- Establishing connections with a relational database (PostgreSQL or MySQL, to be deployed on GCP).
- Fetching raw crash, traffic, and SPaT datasets.
- Processing data to compute the **Safety Index**, severity-weighted rates, and related performance metrics.
- Providing RESTful API endpoints that return structured results in JSON format for visualization.

B. Database

The database layer will be deployed on **Google Cloud SQL** as the central repository for structured crash, traffic, and geospatial data. The database will store preprocessed and raw datasets, supporting both analytical queries and backend computations.

- **PostgreSQL** is preferred for its reliability, scalability, and built-in geospatial support through the **PostGIS** extension, enabling efficient spatial queries.
- **Cloud Integration:** The database will connect securely to the FastAPI backend via private VPC access to minimize latency and improve data privacy.

C. Frontend System Design

The frontend is developed using **Streamlit**, which provides a rapid, Python-native framework for building interactive dashboards and analytical interfaces. It is also deployed on **Google Cloud Run** to enable cloud scalability and ease of integration with backend services.

Frontend Web Application: https://safety-index-frontend-180117512369.europe-west1.run.app/

The Streamlit interface offers:

- **Interactive Dashboards:** Display Safety Index values, severity statistics, and crash trends using dynamic charts and tables.

- **Geospatial Visualization:** Integrate mapping modules (e.g., `streamlit-folium` or `pydeck`) to visualize intersections with risk levels and Safety Index overlays.
- **Filtering and Selection Tools:** Allow users to explore crash data by region, severity, and time interval.
- **Real-Time Data Updates:** Dynamically refresh charts and maps based on backend API responses.

D. System Integration Workflow

The end-to-end workflow of the cloud system is as follows:

1) Users access the **Streamlit** web app hosted on Google Cloud Run.
2) Streamlit sends HTTPS requests to the **FastAPI** backend service.
3) The backend processes incoming queries, retrieves relevant records from the **Cloud SQL (PostgreSQL)** database, and computes Safety Index metrics.
4) Processed results are returned as JSON objects and visualized on the Streamlit interface through charts, tables, and maps.

E. Progress Note

The current project status (https://github.com/Ulrixon/cs6604-trafficsafetyGitHub: Ulrixon/cs6604-trafficsafety) is summarized as follows:

- **Frontend:** The Streamlit dashboard has been deployed successfully on Google Cloud Run. It features an interactive UI for displaying Safety Index data, maps, and analytical charts.
- **Backend:** The FastAPI backend service is fully operational on Google Cloud Run, exposing endpoints for Safety Index computation and data retrieval.
- **Integration:** Initial connectivity between the Streamlit frontend and FastAPI backend is established, enabling live API calls and dynamic content rendering.
- **Database:** The next major milestone involves deploying a **PostgreSQL (PostGIS)** instance on **Google Cloud SQL**, integrating it with the backend for persistent data storage and spatial analysis capabilities.
- **Next Steps:** Future enhancements include database integration, real-time data synchronization, API authentication, and automated CI/CD pipelines for versioned deployments.

VIII. Feedback Question Answer

This section addresses the key feedback points received on the Virginia Tech Transportation Safety Index (VTTSI) proposal and summarizes our responses and progress since the initial submission.

A. 1. Safety Index Formula Validation and Justification

The reviewer noted that the Safety Index formulation appeared ad hoc and requested theoretical grounding and sensitivity analysis. **Response:** The index now follows a reproducible, data-driven weighting framework grounded in established transportation safety literature. Severity weights (w_s) are policy-anchored (Fatal = 10, Injury = 3, PDO = 1) while exposure adjustment and Empirical Bayes (EB) shrinkage stabilize variability. Real-time uplift factors ($F^{spd}, F^{var}, F^{conf}$) are learned through regularized generalized linear modeling using BSM/PSM features. A formal **sensitivity analysis** will test robustness of α, β, and k parameters across intersections to quantify how weight perturbations influence index stability. The team will also compare this hybrid model against machine-learning alternatives (e.g., random-forest regression) to ensure empirical validity.

B. 2. Data Integration Challenges

The feedback emphasized clarifying real-time fusion between BSM, PSM, count, and event data and handling sensor limitations. **Response:** The updated design specifies an event-level fusion:

- BSM and PSM messages are synchronized with safety-event timestamps within 15-minute rolling windows.
- Vehicle/VRU counts and speed distributions are aggregated in near-real time, refreshed every 5–15 minutes.
- Sensor dropout and message latency are mitigated through smoothing, missing-data imputation, and EB stabilization.

Data are ingested via the FastAPI backend on Google Cloud Run, which will connect to a future Cloud SQL (PostGIS) database for persistent storage and spatial joins. This real-time integration supports continuous Safety Index computation and API delivery.

C. 3. Index Interpretability

The reviewer requested clearer meaning for the 0–100 scale and actionable thresholds. **Response:** The Safety Index is now categorized as:

Score Range	Risk Level / Recommended Action
0–30	Low — Normal conditions, no action needed.
30–60	Moderate — Monitor; review signal timing or volume tr
60–80	High — Investigate contributing factors; issue advisory r
80–100	Critical — Immediate engineering or enforcement reviev

Each index output will include percentile ranks and confidence intervals derived from historical distributions to convey uncertainty and contextual meaning.

D. 4. User Case and Experience

The system interface is now designed around two primary user groups that directly experience and influence intersection safety conditions: **vehicle users** and **vulnerable road users (VRUs)** such as pedestrians and cyclists. Tailoring the visualization and interaction modes to these groups ensures that the Safety Index delivers actionable, context-specific insights.

- **Vehicle Users:** The Streamlit dashboard presents vehicle-oriented views highlighting congestion levels, red-light compliance, and intersection risk severity. Drivers and fleet operators can monitor Safety Index values across

intersections, supported by color-coded heatmaps and live updates every 15 minutes. When index values exceed predefined thresholds (e.g., > 80, Critical Risk), the interface can be extend to generate alerts or notifications advising users to reroute or reduce speed near affected intersections in the future.
- **Vulnerable Road Users (VRUs):** The interface includes a VRU-focused view emphasizing pedestrian and cyclist exposure, crosswalk activity, and real-time risk conditions. For example, pedestrians can visualize high-risk crossings (Safety Index > 60) and time windows of increased vehicle turning volume. The interface can be extended to mobile or wearable alerts in future work, warning VRUs when approaching intersections under high-risk conditions.
- **Decision Support Integration:** By differentiating between vehicle and VRU perspectives, the Safety Index enables more personalized and actionable decision-making. Vehicle users can plan safer routes or adjust driving behavior, while VRUs can choose safer crossing times and routes. Transportation engineers can still access aggregated views across both user types for planning and evaluation.

This user-centered adaptation strengthens the project's practical relevance by connecting Safety Index outputs directly to the end-users who experience roadway conditions, improving engagement, comprehension, and proactive safety awareness.

E. 5. Validation Plan

The reviewer highlighted the need for a comprehensive validation framework. **Response:** The team will employ four complementary validation strategies:

1) **Face Validity:** Solicit expert feedback from transportation professionals to ensure results align with intuitive risk perceptions.
2) **Predictive Validity:** Test whether high-index intervals predict subsequent crash or near-miss occurrences using held-out data.
3) **Concurrent Validity:** Compare Safety Index values with existing measures such as crash rate, EPDO, and VDOT safety rankings.
4) **Robustness Testing:** Evaluate index stability under simulated sensor loss, latency, and noise perturbations.

These analyses will establish the reliability, interpretability, and policy relevance of the Safety Index.

F. 6. Implementation Progress Summary

Following the feedback, major milestones have been completed:

- **Backend:** FastAPI server deployed on Google Cloud Run (https://cs6604-trafficsafety-180117512369.europe-west1.run.app/).
- **Frontend:** Streamlit dashboard deployed on Google Cloud Run (https://safety-index-frontend-180117512369.europe-west1.run.app/).
- **Integration:** Functional API–UI linkage established for live JSON responses.
- **Database:** Cloud SQL (PostgreSQL + PostGIS) instance under setup to enable geospatial persistence and query optimization.

These updates collectively address the reviewer's methodological, interpretability, and implementation concerns, advancing the project toward a fully cloud-deployed, data-validated Safety Index system.

References

[1] G. Schultz, "Prioritizing safety funding using severity weighted risk scores," *Accident Analysis & Prevention*, 2025.

[2] Maricopa Association of Governments (MAG), "Network screening methodology for intersections," Maricopa Association of Governments, Phoenix, AZ, Tech. Rep., 2010, Accessed: 2025-09-18. [Online]. Available: https://azmag.gov/Portals/0/Documents/TSC_2010-07-20_MAG-Network-Screening-Methodology-for-Intersections-May-2010_.pdf.

[3] Federal Highway Administration, "Incorporating data-driven safety analysis in traffic impact analyses: A how-to guide," *FHWA-SA-19-026, US Department of Transportation*, 2019. [Online]. Available: https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa19026.pdf.

[4] M. Hossain, M. Abdel-Aty, M. A. Quddus, Y. Muromachi, and S. N. Sadeek, "Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements," *Accident Analysis & Prevention*, vol. 124, pp. 66–84, 2019. DOI: 10.1016/j.aap.2018.12.022.

[5] D. Wu, L. Xing, Y. Li, Y. D. Wong, J. J. Lee, and C. Dong, "A framework for real-time traffic risk prediction incorporating cost-sensitive learning and dynamic thresholds," *Accident Analysis & Prevention*, vol. 218, p. 108 087, 2025. DOI: 10.1016/j.aap.2025.108087.

[6] E. Bancroft, *Seattle keeps emergency scenes clear by alerting smart vehicles of incidents in near real-time*, Seattle DOT Blog (May 28, 2025), 2025. [Online]. Available: https://sdotblog.seattle.gov/2025/05/28/connected-vehicles-technology-pilot-program/.

[7] N. Behboudi, S. Moosavi, and R. Ramnath, "Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques," *arXiv preprint arXiv:2406.13968*, 2024.

[8] H. Zhao, H. Yu, D. Li, T. Mao, and H. Zhu, "Vehicle accident risk prediction based on adaboost-so in vanets," *IEEE Access*, vol. 7, pp. 14 549–14 557, 2019. DOI: 10.1109/ACCESS.2019.2894176.

[9] Q. Hou, Y. Yang, J. Liang, X. Huo, and J. Leng, "A deep transfer learning approach for real-time traffic conflict prediction with trajectory data," *Accident Analysis & Prevention*, vol. 214, p. 107 966, 2025. DOI: 10.1016/j.aap.2025.107966.

Research Part C: Emerging Technologies, vol. 166, p. 104 757, 2024. DOI: 10.1016/j.trc.2024.104757.

[21] J. W. Lee, H. Wang, K. Jang, A. Hayat, M. Bunting, and et al., "Traffic control via connected and automated vehicles: An open-road field experiment with 100 cavs," *arXiv e-prints*, 2024, Art. no. arXiv:2402.17043. DOI: 10.48550/arXiv.2402.17043. [Online]. Available: https://arxiv.org/abs/2402.17043.

[22] M. Oh, J. Shaw, and J. Dong-O'Brien, "Enhancing work zone safety: A psychological and behavioral analysis of anti-tailgating strategies using dynamic message signs and fixed warning signs," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 106, pp. 150–168, 2024. DOI: 10.1016/j.trf.2024.08.007.

[23] M. A. Arman, C. M. J. Tampère, et al., "Empirical study of lane-changing maneuvers in a weaving area based on reconstructed trajectories of floating car data," *Transportation Research Record*, 2023. DOI: 10.1177/03611981231179474.

[24] A. Shimojo et al., "How impressions of other drivers affect one's behavior when merging lanes," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 89, pp. 236–248, 2022.

[25] C. Gutierrez-Osorio and C. A. Pedraza, "Modern data sources and techniques for analysis and forecast of road accidents: A review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 4, pp. 432–446, 2020. DOI: 10.1016/j.jtte.2020.05.002.

APPENDIX A
TASK ASSIGNMENT FOR EACH MEMBER

Task	**Assigned To**
Define objectives & KPIs (delay, conflict rate, speed variance)	Victory Uhunmwangho
Draft system architecture & data flow diagram	Victory Uhunmwangho
Set repo structure, naming, meeting cadence	Victory Uhunmwangho
Ingest all CSVs; standardize timestamps/IDs	Cheng-Shun Chuang
Basic data QC (missing rates, ranges, anomalies)	Cheng-Shun Chuang
Appendix C: Data Dictionary (columns, dtypes, units)	Cheng-Shun Chuang
Engineer features (vehicle/VRU volumes, speed stats, platoons)	Cheng-Shun Chuang
Link safety events to BSM/PSM (time/space bins)	Cheng-Shun Chuang
Define train/val/test time splits	Jason Cusati
Implement baselines (exposure, severity/EPDO)	Jason Cusati
Draft & freeze Safety Index spec (factors, weights, 0–100 scaling)	Jason Cusati
Create API stubs for `/index` & `/kpi`	Cheng-Shun Chuang
Compute TTC/PET + hard-brake flags from BSM/PSM	Jason Cusati
Validate surrogate-safety metrics on held-out day	Jason Cusati
Expand API with `/conflicts` & `/meta`	Cheng-Shun Chuang
Build dashboard v1 (map, trends, 15-min heatmap)	Victory Uhunmwangho
Draft Appendix D: Methods (index formulas, TTC/PET defs, KPI math, stats tests)	Victory Uhunmwangho
Hook dashboard to API endpoints	Cheng-Shun Chuang
Sensitivity analysis of index weights/normalizers	Jason Cusati
Ablations (drop factors; re-score)	Jason Cusati
Harden data pipeline (edge cases, missing-data strategies)	Cheng-Shun Chuang

[10] Y. Gong, M. Abdel-Aty, J. Yuan, and Q. Cai, "Multi-objective reinforcement learning approach for improving safety at intersections with adaptive traffic signal control," *Accident Analysis & Prevention*, vol. 144, p. 105 655, 2020. DOI: 10.1016/j.aap.2020.105655.
[11] A. Montella, C. Guida, J. Mosca, J. Lee, and M. Abdel-Aty, "Systemic approach to improve safety of urban unsignalized intersections: Development and validation of a safety index," *Accident Analysis & Prevention*, vol. 141, p. 105 523, 2020. DOI: 10.1016/j.aap.2020.105523.
[12] B. Persaud and C. Lyon, "Empirical bayes before–after safety studies: Lessons learned from two decades of experience and future directions," *Accident Analysis & Prevention*, vol. 39, no. 3, pp. 546–555, 2007. DOI: 10.1016/j.aap.2006.09.009. [Online]. Available: https://doi.org/10.1016/j.aap.2006.09.009.
[13] Federal Highway Administration, "Bicycle and pedestrian data collection in united states communities: Quantifying use, surveying users, and documenting facility extent," U.S. Department of Transportation, Washington, DC, Tech. Rep., 2016, Accessed: 2025-09-18. [Online]. Available: https://www.fhwa.dot.gov/environment/bicycle_pedestrian/publications/bikeped_data_collection/.
[14] Y. Gao and D. Levinson, "Lane changing and congestion are mutually reinforcing?" *Communications in Transportation Research*, vol. 3, p. 100 101, 2023. DOI: 10.1016/j.commtr.2023.100101.
[15] S. Mohammadi, R. Arvin, A. J. Khattak, and S. Chakraborty, "Social interactions in driving: How driver–driver interactions can lead to complex traffic dynamics," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 80, pp. 203–217, 2021. DOI: 10.1016/j.trf.2021.04.002.
[16] N. Xie and H. Wang, "Distributed adaptive traffic signal control based on shockwave theory," *Transportation Research Part C: Emerging Technologies*, vol. 173, p. 105 052, 2025. DOI: 10.1016/j.trc.2025.105052.
[17] I. Mutambik, D.-Q. Luong, Y. Zhou, B. Alotaibi, and M. S. Aziz, "Iot-enabled adaptive traffic management: A multiagent framework for urban mobility optimisation," *Sensors*, vol. 25, no. 13, p. 4126, 2025. DOI: 10.3390/s25134126.
[18] M. Khanmohamadi and M. Guerrieri, "Smart intersections and connected autonomous vehicles for sustainable smart cities: A brief review," *Sustainability*, vol. 17, no. 7, p. 3254, 2025. DOI: 10.3390/su17073254.
[19] S. Wang, M. Shang, M. W. Levin, and R. Stern, "A general approach to smoothing nonlinear mixed traffic via control of autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103 967, 2023. DOI: 10.1016/j.trc.2022.103967.
[20] M. Wu, B. Wang, Y. Yin, and J. P. Lynch, "Participatory traffic control: Leveraging connected and automated vehicles to enhance network efficiency," *Transportation*

Task
Write Ethics/Privacy/Accessibility section
Add safety guardrails (min ped clearance, safe defaults)
Dashboard v2 (legends, tooltips, export PNG/CSV)
Integrate figures/tables into main document
Results dry-run; fix narrative gaps
Repo cleanup; README "how to reproduce"
Final proof (numbering, captions, references)
Sanity re-compute of headline KPIs
Package submission (PDF + Appendices + code link + data dict)

Appendix B
Schedule for project

Appendix C
Acknowledgment

This proposal used AI for assistance in generating content and will assist in writing codes.

TABLE II
CANDIDATE INPUTS FOR THE SAFETY INDEX DERIVED FROM V2X MESSAGES. DIRECTION: ↑ INCREASES RISK SCORE; ↓ DECREASES RISK.

Parameter	**Source**	**Safety Rationale (Proof-Backed)**	**Dir.**	**Initial Weighting Guidance**
Instantaneous speed (m/s)	BSM (coreData.speed)	Higher speeds elevate kinetic energy and crash severity; baseline exposure normalizes by typical conditions. BSM exposes per-vehicle speed for aggregation over 15-min bins. :contentReference[oaicite:13]index=13	↑	Start modest (e.g., 0.5–1.0) and scale by deviation from posted limit (MapData).
Speed deviation from limit	BSM + MapData (speedLimits)	Risk increases as speed exceeds posted/regulatory limits per lane. :contentReference[oaicite:14]index=14	↑	Weight by z-score relative to historical TOD/DOW distribution.
Speed variance (within 15-min)	BSM	High variance indicates turbulence/instability and elevated conflict potential.	↑	Use variance-to-mean ratio; calibrate via EB shrinkage to reduce noise.
Hard braking frequency	BSM (brake flags)	Frequent braking spikes proxy for conflicts/near-misses; BSM JSON includes brake indicators. :contentReference[oaicite:15]index=15	↑	Per 100 BSMs; cap influence to prevent outlier domination.
Heading change rate / weaving	BSM (heading + positions)	Rapid heading oscillations imply lane changes/merging conflicts. :contentReference[oaicite:16]index=16	↑	Derive as deg/s; weight increases with density.
Vehicle density proxy	BSM counts per segment	BSM volume within a segment/time bin captures instantaneous exposure.	↑	Normalize by segment length; sigmoid scaling to avoid saturation.
Vehicle class proxy (size)	BSM (size.width/length)	Larger vehicles yield higher severities in conflicts; BSM size fields available. :contentReference[oaicite:17]index=17	↑	Apply multiplier to speed-related risk terms.
Elevation/grade context	BSM (elevation)	Steep grades correlate with speed control issues/braking risk. :contentReference[oaicite:18]index=18	↑	Low weight; increase in wet/cold weather (future WX fuse-in).
VRU speed	PSM	Higher pedestrian/cyclist speed at crossings increases conflict exposure time-window overlap. (PSM DTOs present in VCC) :contentReference[oaicite:19]index=19	↑	Weight higher during permissive turns (from SPaT).
VRU count/density	PSM counts per approach	More VRUs ⇒ higher exposure for turning vehicles.	↑	Use square-root scaling to avoid overwhelming the index.
VRU path alignment	PSM trajectories	Trajectories aligned with turning paths increase conflict probability.	↑	Multiply by concurrent permissive vehicle phase.
Signal phase state	SPaT (per signalGroup phase)	Red/permissive/protected states alter conflict structure; SPaT provides group phases and end times. :contentReference[oaicite:20]index=20	↑ / ↓	Protected lefts reduce risk (down); permissive lefts increase (up).
Time-to-phase change	SPaT (min/max end timestamps)	Short remaining greens induce last-	↑	Weight increases nonlinearly as remaining time

TABLE IV
PROJECT SCHEDULE (10 WEEKS)

Week	Task
1	Literature Review & Requirements Finalization: Review Safety Index and ITS literature; finalize project scope, datasets, and technical stack (FastAPI, PostgreSQL, React).
2	Dataset Acquisition & Preprocessing: Collect Smart Intersection datasets (BSM, PSM, vehicle counts, VRU counts, speed distribution, safety events); clean, normalize, and integrate into PostgreSQL/PostGIS.
3	Safety Index Model Design: Define VRU Index, Vehicle Index, and Combined Index formulas; establish severity weighting and exposure-adjusted normalization.
4	Model Prototyping & Validation: Implement safety index calculations; validate with benchmarks (e.g., Empirical Bayes–adjusted SPFs).
5	Backend Development: Build FastAPI services; implement ingestion, aggregation, and REST APIs for index computation.
6	Frontend Development: Develop React-based visualization; implement interactive map, dashboards, and real-time updates.
7	Backend–Frontend Integration: Connect React UI to FastAPI APIs; test real-time data flows and visualization updates.
8	System Testing & Optimization: Perform unit tests, performance tests, and debugging; address latency, blind spots, and variability smoothing.
9	Refinement & Documentation: Adjust index weighting parameters; prepare technical documentation, user guide, and API reference.
10	Finalization & Presentation: Prepare final demo and walkthrough; deliver project report, presentation, and recommendations.